Corticotropin-Releasing Factor: An Ancient Peptide Family Related to the Secretin Peptide Superfamily

Ola M. Michalec 1, Belinda S. W. Chang 1, Nathan R. Lovejoy 2 and David A. Lovejoy 1*

1 Department of Cell and Systems Biology, University of Toronto - St. George Campus, Toronto, ON, Canada, 2 Department of Biological Sciences, University of Toronto Scarborough, Toronto, ON, Canada

Corticotropin-releasing factor (CRF) is the hypothalamic releasing peptide that regulates the hypothalamic-pituitary-adrenal/inter-renal (HPA/I) axis in vertebrates. Over the last 25 years, there has been considerable discussion on its paralogs genes, urotensin-I/urocortin-1, and urocortins-2 and-3 and their subsequent role in the vertebrate stress response. Phylogenetically, the CRF family of peptides also belong to the diverse assemblage of Secretin- and Calcitonin-based peptides as evidenced by comparative-based studies of both their ligand and G-protein-coupled receptor (GPCR) structures. Despite this, the common origin of this large assemblage of peptides has not been ascertained. An unusual peptide, teneurin-C-terminal associated peptide (TCAP), reported in 2004, comprises the distal extracellular tip of the teneurin transmembrane proteins. Further studies indicated that this teneurin region binds to the latrophilin family of GPCRs. Initially thought to be a member of the Secretin GPCR family, evidence indicates that the latrophilins are a member of the Adhesion family of GPCRs and are related to the common ancestor of both Adhesion and Secretin GPCR families. In this study, we posit that TCAP may be a distantly related ancestor of the CRF-Calcitonin-Secretin peptide family and evolved near the base of metazoan phylogeny.

Keywords: secretin superfamily, TCAP, teneurin, adhesion GPCRs, evolution, metabolism

INTRODUCTION

Corticotropin-releasing factor (CRF) is the critical hypothalamic releasing factor that regulates the hypothalamus-pituitary-adrenal/inter-renal (HPA/I) axis in vertebrates, yet after some 40 years after its discovery, numerous questions still exist regarding when, why, and how this peptide evolved. We hypothesize that due to the high level of primary structure similarity among CRF paralogs and related peptide lineages (e.g., calcitonin, secretin) there was likely an ancestor peptide common to this cluster. We further suggest that the “teneurin C-terminal associated peptides” (TCAPs) represent an extant candidate lineage related to the hypothetical common ancestor.

The discovery of CRF in the early 1980s (1) occurred around the same time as the discovery of other peptides of similar structure [sauvagine (2); urotensin-I (3)]. Later, Vale and his laboratory characterized a mammalian version of sauvagine/urotensin-I in rat brain that they termed, urocortin (4). Further phylogenetic studies suggested that mammalian urocortin, amphibian sauvagine, and fish urotensin-I were orthologs of the same gene (5). In 2001, the structures of two novel related peptides were reported by the Vale laboratory who named the peptides, urocortin 2
and 3 (6, 7) and by Hsu and Hsieh (8) who termed the peptides as “stressespin” and “stressespin-related peptide.” These novel CRF family homologs were subsequently established to be a separate paralogs lineage of CRF and the urotensin-I/sauvagine/urocortin grouping (5, 9–13). In parallel to studies of vertebrate CRF isoforms, the presence of related peptides were reported in insects and arthropods (12, 14–16). Therefore, the high degree of structural similarity among CRF-like peptides in both deuterostome (e.g., chordates) and protostomes (e.g., arthropods) indicated that an ancestral peptide with CRF family primary structure attributes was present before the bifurcation of these metazoan lineages. Importantly, this ancestral peptide appeared to exist in a physiologically mature form indicative of a distant lineage that likely radiated as other ancestral peptides with distinct but overlapping functions. The identity of these hypothetical ancestral peptides has remained elusive, however, it is plausible that these lineages led to the evolution and expansion of the secretin and calcitonin family of peptides (11, 12).

The Secretin superfamily of peptides is a diverse assemblage of peptide lineages with overlapping functions utilizing structurally related receptors. The nomenclature describing the phylogeny of the secretin grouping of peptides and receptors is confusing. In order to clarify this, we have used the term “secretin family” to denote those peptides that are thought to be part of a direct monophyletic clade (e.g., secretin, PACAP, VIP, and glucagon paralogs). For the inclusion of the wider group which include CRF and calcitonin, we have referred to this as the “Secretin paralogs.” Interestingly, some derived phylogenetically younger Adhesion GPCR paralogs. For the inclusion of the wider group which include CRF and calcitonin receptors are included within the Secretin superfamily of ligands, suggesting a common evolutionary origin. Several recent studies established that the distal region of the extracellular domain of the teneurin transmembrane proteins binds ADGRL with high affinity and activates the receptor. Silva et al. (22) first discovered that teneurin-2, expressed on pre-synaptic dendritic branches, binds LPHN-1 expressed on pre-synaptic nerve terminals to form a trans-synaptic complex. Similar trans-cellular interactions were observed between teneurins-2 and 4 and all three LPHNs (23) and between teneurin-1 and LPHN-3 (24). A C-terminal fragment of teneurin-2, named Lasso, triggered an increase in cytosolic Ca2+ in Nb2a cells overexpressing LPHN-1 and in pre-synaptic nerve terminals of hippocampal cells (22). This distal region of the teneurin extracellular domain contains a peptide-like sequence termed “teneurin C-terminal associated peptide” (TCAP). The TCAPs are a family of four bioactive peptides that are 40–41 amino acids in length and are located at the C-terminus of each of the teneurin transmembrane proteins (25, 26). TCAPs possess a cleavage motif at the N-terminus and an amidation motif at the C-terminus (27) and may be autolytically cleaved from teneurins upon binding with LPHN (28, 29). TCAP shares about 20% sequence similarity with CRF and calcitonin, members of the Secretin superfamily of ligands, suggesting a common evolutionary origin (27). Moreover, our laboratory has recently identified that teneurin C-terminal associated peptide (TCAP)-1 is likely an endogenous ligand that interacts with the HBD of LPHN (30).

Therefore, as TCAP binds to an Adhesion GPCR and shares sequence similarity to CRF and calcitonin, ligands that bind to Secretin GPCR receptors that are classified as being most closely related to ancestral Adhesion GPCRs, this prompted the investigation of TCAP as a progenitor of the Secretin superfamily. The hypothesis that the teneurin-TCAP system is an ancient system that arose prior to the emergence of the Metazoa as a result of a horizontal gene transfer (HGT) event from a prokaryote to a choanoflagellate ancestor has previously been raised (31–33). However, the TCAP family has not been previously examined. Thus, TCAP may be associated with an early evolving lineage of peptides that is a sister lineage to the CRF, calcitonin, and secretin families of peptides (11, 34). We therefore examined the phylogenetic relationships of these peptides using TCAP as an outgroup.
MATERIALS AND METHODS

Collection of Sequences

Peptide sequences of Secretin GPCR ligands, including CRF, calcitonin and secretin families, and Adhesion GPCR ligands, including TCAP 1–4, as well as reference groups including neuropeptide Y (NPY) and insulin were collected among a range of extant protostomes and deuterostomes, using the GenBank genome sequence analysis program on the NCBI website. The peptides were organized by organism, phylum, class, and order and were tabulated and their accession numbers were recorded (Table 1). Sequences were divided into pre-pro-peptides (or pro-peptides for TCAP) and mature peptides, after which they were imported to MEGA 6.0 for analysis (38). Downloaded from http://www.megasoftware.net/.

Sequence Alignments

Peptide sequences were aligned using the MUSCLE algorithm (39). The alignment was examined, reviewed for duplicate sequences using pairwise distances (d = 0.0 was identical) and excess sequence was cut at both 5’ and 3’ ends, as these fragments did not contribute to the alignment. Modifications to the alignment were made to ensure that the characteristic residue motifs were conserved. This included highly conserved cysteine (C), tryptophan (W), arginine (R), and lysine (K) residues throughout as well as motifs characteristic of each family. For the CRF family this was the 5’ leucine (L), serine (S), and the 3’ asparagine (N) motif that is conserved throughout the entire family, the “TCV” or “TCXV” motif that is conserved among the calcitonin family and the “PELAD” motif that is conserved among the TCAP family.

Phylogenetic Analysis

Phylogenetic tree construction and statistical analyses were carried out in MEGA 6.0 (38). A multi-step approach was undertaken in order to understand the relationship of each family relative to TCAP prior to conducting a comprehensive analysis of all of the families.

Maximum Likelihood (ML) Method

The amino acid substitution model and the rate among sites were both chosen based on the model that resulted in the greatest log likelihood, the lowest Akaike Information Criterion (AIC) and the lowest Bayesian Information Criterion (BIC), parameters calculated by MEGA 6.0. To ensure the most accurate analysis, these parameters were calculated for each constructed tree. The model that maximized the log likelihood was used for analysis. A partial deletion of sequences with too many gaps/missing data was applied with a cutoff of 95%, so sites that were not found in at least 95% of sequences were not used toward the analysis. The applied heuristic method was Nearest-Neighbor Interchange (NNI), so the initial trees were obtained using the NJ method to a matrix of pairwise distances estimated using a JTT model. Reliability of the tree was tested using 1,000 bootstrap replicates.

Pre-propeptide and Mature Peptide Analysis

Two sets of analyses were performed. The first involved Secretin superfamily pre-pro-peptides, which are composed of a signal, cryptic, and mature peptide and TCAP pro-peptides, as TCAP does not possess a signal peptide. Given the functional importance, bioactivity, and high level of conservation throughout evolution, a second separate analysis was performed on mature peptides of both Secretin superfamily and TCAP family members.

For analysis involving Secretin superfamily pre-propeptides and TCAP family pro-peptides, a total of 181 amino acid sequences were used, with a total of 44 positions in the final dataset after all positions with <95% site coverage were eliminated.

Mature Peptide Analysis

For analysis involving Secretin superfamily mature peptides and TCAP mature peptides, a multi-step analysis was undertaken in order to elucidate the relationships of each family with respect to one another and TCAP. As insulin has a tertiary structure where the peptide folds and the two mature chains are connected by sets of disulfide bonds from the cysteine residues (40), the mature peptide had to be divided into A and B chains for the purpose of this analysis. Due to the high sequence conservation of NPY that may have resulted in the odd placement of the NPY reference group in the pre-pro-peptide analysis and given that the NPY mature peptide is even so more highly conserved, it was not included as a reference group in the analysis of mature peptides. For insulin and the calcitonin family, analysis involved 72 amino acid sequences, with a total of 14 positions in the final dataset. For insulin, calcitonin, and TCAP, analysis involved 95 amino acid sequences, with a total of 14 positions in the data set. For insulin, calcitonin, CRF and TCAP families, analysis involved 135 amino acid sequences leaving 12 positions in the final data set. Lastly, for insulin, calcitonin, CRF, secretin, and TCAP families, analysis included 179 amino acid sequences leaving 15 positions in the final dataset.

RESULTS

Sequence Analysis of TCAP Paralogs and Orthologs

TCAP paralogs, those that diverged as a result of a genome duplication event, demonstrated a high degree of conservation (Figure 1). When TCAP 1–4 are aligned in mouse, residues Q2, L4, G7, V9, Q10, G11, Y12, G14, V17, V20, E21, Q22, Y23, E25, L26, D28, S29, N32, I33, F35, R37, Q38, and E40 are all conserved among the four paralogs (Figure 1). Similarly, TCAP orthologs, those that arose as a result of a species divergence, demonstrate a high degree of conservation among vertebrates (Figure 2). When mammalian, bird, amphibian, and fish TCAP 1–4 sequences are aligned residues L3, G7, V9, G11, Y12, G14, L18, Q22, E25, L26, D28, N32, R37 are conserved among TCAP-1 orthologs (Figure 2A). Among TCAP-2 orthologs, residues Q2, L3, G7, G11, Y12, E13, G14, Y15, V16, V17, L18, P19, V20, Q22, Y23, P24, E25, L26, A27, D28, S29, S30, N32, I33, F35, Q38, N39, E40, M41 are conserved (Figure 2B). Among TCAP-3 orthologs, residues Q2, L3, L4, S5, K8, V9, G11, Y12, D13, G14, Y15, V17, L18, S19, V20, E21, Q22, Y23, E25, L26, D28, S29, N32, F35, R37, Q38, E40, I41 are conserved (Figure 2C). Lastly, among TCAP-4...
### Figure 1

Multiple sequence alignment of the TCAP family of peptides in mouse. The mature peptide sequences were aligned using MUSCLE (Multiple Sequence Comparison by Log-Expectation). Dark gray boxes indicate amino acid identity and light gray boxes indicate a functional replacement.

| Peptide          | Sequence                                      |
|------------------|-----------------------------------------------|
| Mouse TCAP1      | QQLLGTRVQGVDGFVLSVEQYELSDANNHHFMRQSEI        |
| Mouse TCAP2      | QQLNTGRVQGVEGYVLPVEQYPEDSNNIQFLRQNEI         |
| Mouse TCAP3      | QLLSAQKVQGVDGFVLSVEQYELSDANNHHFMRQSEI        |
| Mouse TCAP4      | QQLNTRVQGVDGFVLSVEQYELSDANNHHFMRQSEI         |

### Figure 2

Multiple sequence alignment of TCAP family members among various species. (A) TCAP-1; (B) TCAP-2; (C) TCAP-3; (D) TCAP-4. The mature peptide sequences were aligned using MUSCLE (Multiple Sequence Comparison by Log-Expectation). Dark gray boxes indicate amino acid identity and light gray boxes indicate a functional replacement.

| Peptide          | Sequence                                      |
|------------------|-----------------------------------------------|
| Mouse TCAP2      | QQLNTGRVQGEGGYYVLPLEQYPULASDSNNIQFLRQNEI     |
| Human TCAP2      | QQLSTGRVQGEGGYYVLPLEQYPULASDSNNIQFLRQNEI     |
| Chicken TCAP2    | QQLNTGRVQGEGGYYVLPLEQYPULASDSNNIQFLRQNEI     |
| Clawedfrog TCAP2 | QQLNTGRVQGEGGYYVLPLEQYPULASDSNNIQFLRQNEI     |
| Eleph. shark TCAP2 | QQLAGGRRVPGDFVLPVEQYPULASDSNNIQFLRQNEI         |
| Zebrafish TCAP2  | QQLAMG狭VAGEGGYYVLPLEQYPULASDSNNIQFLRQNEI     |
| Zebrafish TCAP2A | QQLAMG狭VAGEGGYYVLPLEQYPULASDSNNIQFLRQNEI     |
| Zebrafish TCAP2B | QQLAMG狭VAGEGGYYVLPLEQYPULASDSNNIQFLRQNEI     |

Orthologs, Q1, Q2, L4, G7, R8, V9, Q10, G11, Y12, G14, F15, V20, Q22, P24, E25, L26, D28, N31, N32, H34, F35, R37, Q38, E40, M41. Overall, TCAP-2 orthologs (Figure 2B) are the most highly conserved and TCAP-1 orthologs (Figure 2A) are the least highly conserved.
conserved. Also, a characteristic “PELAD” motif at positions 24–28 from the N-terminus is conserved among the TCAP paralogs and orthologs. The high level of conservation of the “PELAD” motif suggests that it possesses a functional attribute, such as a receptor binding or activation site (27). This family of peptides contains the “PELAD” motif at residues 24–28 from the N-terminus. Therefore, both TCAP orthologs and paralogs demonstrate a high degree of conservation among vertebrates.

**Evolutionary Analysis of Pre-propeptides and Mature Peptides of Secretin Superfamily and TCAP Family Members**

Phylogenetic analysis of CRF, calcitonin, and secretin pre-propeptide families and TCAP family preproteins revealed that each family formed a distinct group. TCAP, CRF, and secretin families form distinct clades and insulin forms a sister group with the calcitonin family (Figure 3). Also, CRF and calcitonin are closely related sister lineages and they, in turn, form a sister lineage to the secretin family. TCAP, the putative progenitor, is most distantly related to these families relative to their relationships to one another.

A separate analysis was performed with mature peptide sequences of the Secretin superfamily and TCAP mature peptides due to their high conservation and functional importance throughout evolution. Phylogenetic analysis of calcitonin mature peptides, insulin A and B mature chains and TCAP demonstrated that calcitonin and insulin families are sister lineages (Figure 4). Insulin A chains were more closely related to the calcitonin family than insulin B chains (Figure 4). Phylogenetic analysis of calcitonin, insulin A and B chains, CRF, and TCAP mature peptides confirmed that calcitonin and insulin families were sister lineages and that CRF formed a separate group to these two families (Figure 5). Lastly, phylogenetic analysis of calcitonin, insulin A and B chains, CRF, secretin, and TCAP mature peptides revealed that calcitonin and insulin families were sister lineages and that both CRF and secretin formed separate groups from these two families (Figure 6). Therefore, the multi-step mature peptide analysis confirmed that insulin and calcitonin are sister lineages, that form distinct groups from CRF and secretin families and in turn, that the TCAP family is a distinct clade from Secretin superfamily members.

**DISCUSSION**

In this study, the TCAP family is presented as a putative progenitor of the Secretin superfamily of ligands for the first time. The evolutionary relationships among the receptors of these peptides are well-established (18, 19). However, the relationships among members of the Secretin superfamily of ligands as well as a progenitor for this family of peptides have not been elucidated. We considered TCAP as a putative progenitor of the Secretin superfamily for the following reasons. First, evolutionary relationships among the receptors of these ligands demonstrate that Secretin GPCRs derived from Adhesion GPCRs (19) and as TCAP-1 binds to LPHN, an Adhesion GPCR with a HBD characteristic of Secretin GPCRs (17). It is possible that a similar course of evolution occurred for the ligands. Second, the sequence similarity that TCAP shares with CRF and calcitonin (27), both Secretin superfamily members whose receptors are the most closely related to Adhesion GPCRs, suggests that these peptides may have evolved from TCAP, a candidate progenitor peptide.

The teneurin-TCAP system is well-established as being evolutionarily ancient. Evidence suggests that this system arose before the Metazoa evolved about 1 billion years ago and prior to the emergence of the Secretin superfamily that arose around the time of the protostome-deuterostome divergence, about 600 million years ago. As a result, although the TCAP sequence shows some amino acid similarity with the Secretin superfamily, there are a number of differences indicating that the two lineages are evolutionarily divergent. Indeed, we could not determine any significant binding or activation capacity of TCAP with any members of the Secretin GPCRs [(11, 34); Lovejoy, unpublished observations]. In contrast, TCAP binds to the latrophilin HBD and activates this receptor [(30); Reid et al., submitted]. As proposed by Zhang et al. (33), the teneurin-TCAP system likely evolved from a polymorphic proteinaceous toxin (PPT) gene that arose as a result of a HGT event from a prokaryote to a choanoflagellate, a primitive unicellular organism. Importantly, the teneurin gene has been identified in the choanoflagellate, *Monosiga brevicollis* (32). Choanoflagellates are thought to be a progenitor to the Metazoa (42). This supports the hypothesis that a choanoflagellate may have engulped a prokaryote containing the PPT gene, which became integrated into its genome and lost its toxic role over time (32, 33). With respect to structural evidence, the teneurins share characteristics of PPTs: the same type II orientation, rearrangement hotspot (RHS) domains and close similarity to the C-terminal domain to the histidine-asparagine-histidine (HNH) bacterial toxin of the glycine-histidine-histidine (GHH) clade (33, 43). The GHH domain may be an ancestor of TCAP that lost its toxic role and functioned as an intracellular signaling molecule (33). Additionally, the C-terminal region of the *M. brevicollis* teneurin protein contains tyrosine-aspartate (YD) repeats characteristic of proteobacteria and most of the extracellular domain is encoded on one large 6,829 base pair exon characteristic of prokaryotic genomes and of HGT (32). Therefore, evidence suggests that the teneurin-TCAP system is ancient as it evolved as a result of a HGT event prior to the emergence of the Metazoa.

Moreover, with respect to the course of evolution of the receptors, evidence demonstrates that Adhesion GPCRs evolved prior to Secretin GPCRs and that Secretin GPCRs are derived from Adhesion GPCRs. Adhesion GPCR genes have been identified in the genome of amphioxus, *Branchiostoma floridae*, the choanoflagellate, *M. brevicollis*, and the sea anemone, *Nematostella vectensis* (18), meaning that these lineages were present prior to the protostome-deuterostome divergence. On the other hand, Secretin GPCRs have not been identified in these species and therefore, receptor lineages of the Secretin superfamily likely expanded and radiated around the time of the bifurcation of protostomes and deuterostomes. Also, Nordström et al. (18) demonstrated Secretin GPCRs evolved from Adhesion
FIGURE 3 | Phylogenetic analysis of CRF, calcitonin, insulin, and secretin family pre-propeptides with TCAP propeptides (rooted to TCAP). Each family is highlighted with a different color: CRF (red), calcitonin (orange), secretin (purple), TCAP (blue). Analysis was conducted using the maximum likelihood method based on the JTT+G matrix-based model (lnL = −11224.5064; +G, parameter = 1.3976) (41). Initial trees for the heuristic search were obtained by applying the NJ method to a matrix of pairwise distances estimated using a JTT model. Branch lengths represent the number of substitutions per site, with the tree shown to scale. Bootstrap analysis involved 1,000 replicates. CRF family: CRF, corticotropin-releasing factor; TCN, teleocortin; UCN, urocortin; UCN2, urocortin 2; UCN3, urocortin 3; UI, urotensin; SVG, sauvagine; DH, diuretic hormone; Calcitonin family: CALC, calcitonin; CGRP1, calcitonin-gene-related peptide 1; CGRP2, calcitonin-gene-related peptide 2; AM, amylin; ADM, adrenomedullin; ADM2, adrenomedullin 2; Secretin family: SCT, secretin; GHRH, growth hormone releasing hormone; GIP, gastric inhibitory peptide; GCG, glucagon; PACAP, pituitary adenylate cyclase-activating peptide; VIP, vasoactive intestinal peptide; Reference groups: NPY, neuropeptide Y; INS, insulin; Outgroup: TCAP, teneurin C-terminal associated peptide. The scale bar indicates the level of magnification for the tree.

GCPRs using phylogenetic analysis. Therefore, evidence that the teneurin-TCAP system arose prior to the emergence of the Metazoa as well as the characterization of Adhesion GCPRs but not Secretin GCPRs prior to the protostome-deuterostome divergence suggests that the teneurin-TCAP system predates members of the Secretin superfamily. We suggest that if the
ligands for these receptors underwent a similar course in evolution, the TCAP family may be a putative progenitor to the Secretin superfamily.

In light of the evidence to suggest that the teneurin-TCAP system evolved prior to the emergence of the Metazoa, the previously established relationship that Secretin GPCRs derived from Adhesion GPCRs [(Nordstrom et al., 2009); (19)], the evidence that TCAP binds to LPHN, an Adhesion GPCR with a HBD characteristic of Secretin GPCRs (17) and given the sequence similarity that TCAP shares with Secretin superfamily members, CRF, and calcitonin (27), a phylogenetic investigation using TCAP as a putative progenitor of the Secretin superfamily was undertaken. A putative progenitor of the Secretin superfamily of ligands has not been previously established. Sequence analysis of TCAP family members demonstrated a highly conserved peptide and
FIGURE 5 | Phylogenetic analysis of insulin, calcitonin, CRF and TCAP mature peptides (rooted to TCAP). The trees are represented as (A) original tree with the appropriate scale (B) magnified and rooted to TCAP. Each family is highlighted with a different color: CRF (red), calcitonin (orange), insulin (green), TCAP (blue).

Analysis was conducted using the maximum likelihood method based on the Dayhoff matrix-based model (lnL = −1019.5552; +G, parameter = 6.6766) (41). Initial trees for the heuristic search were obtained by applying the NJ method to a matrix of pairwise distances estimated using a JTT model. Branch lengths represent the number of substitutions per site, with the tree shown to scale. Bootstrap analysis involved 1,000 replicates. Calcitonin family: CALC, calcitonin; CGRP1, calcitonin-gene-related peptide 1; CGRP2, calcitonin-gene-related peptide 2; AM, amylin; ADM, adrenomedullin; ADM2, adrenomedullin 2; Insulin: INSa, insulin A chain; INSb, insulin B chain; CRF family: CRF, corticotropin-releasing factor; TCN, teleocortin; UCN, urocortin; UCN2, urocortin 2; UCN3, urocortin 3; UI, urotensin; SVG, sauvagine; DH, diuretic hormone; Outgroup: TCAP, teneurin C-terminal associated peptide. The scale bar indicates the level of magnification for the tree.
FIGURE 6 | Phylogenetic analysis of insulin, calcitonin, CRF, secretin, and TCAP mature peptides. The trees are represented as (A) unrooted and (B) rooted to TCAP. Each family is highlighted with a different color: CRF (red), calcitonin (orange), insulin (green), secretin (purple), and TCAP (blue). Analysis was conducted using the maximum likelihood method based on the Whelan and Goldman model ($lnL = -1781.0007; +G$, parameter $= 23.5912$) (41). Initial trees for the heuristic search were obtained by applying the NJ method to a matrix of pairwise distances estimated using a JTT model. Branch lengths represent the number of substitutions per site, with the tree shown to scale. Bootstrap analysis involved 1,000 replicates. Calcitonin family: CALC, calcitonin; CGRP1, calcitonin-gene-related peptide 1; CGRP2, calcitonin-gene-related peptide 2; AM, amylin; ADM, adrenomedulin; ADM2, adrenomedulin 2; INSa, insulin A chain; INSb, insulin B chain; CRF family: CRF, corticotropin-releasing factor; TCN, teleocortin; UCN, urocortin; UCN2, urocortin 2; UCN3, urocortin 3; UI, urotensin; SVG, sauvagine; DH, diuretic hormone; Secretin family: SCT, secretin; GHRH, growth hormone releasing hormone; GIP, gastric inhibitory peptide; GCG, glucagon; PACAP, pituitary adenylate cyclase-activating peptide; VIP, vasoactive intestinal peptide; Outgroup: TCAP, teneurin C-terminal associated peptide. The scale bar indicates the level of magnification for the tree.
| Hormone | Organism                          | Phylum            | Class       | Order     | Accession number             |
|---------|-----------------------------------|-------------------|-------------|-----------|------------------------------|
| CRF     | Xenopus laevis (African clawed frog) | Chordata          | Amphibia    | Anura     | NM_001172210                 |
| CRF     | Rana sylvatica (Wood frog)        | Chordata          | Amphibia    | Anura     | HG330608                     |
| CRF     | Gallus gallus (Chicken)           | Chordata          | Aves        | Galliformes | NM_001123303                 |
| CRF     | Columbia livia (Rock pigeon)      | Chordata          | Aves        | Columbiformes | XM_05506466                  |
| CRF     | Danio rerio (Zebrafish)           | Chordata          | Actinopterygi | Cypriniformes | JN859047                    |
| CRF     | Salmo salar (Atlantic salmon)     | Chordata          | Actinopterygi | Salmoniformes | NM_001141590                 |
| CRF     | Mus musculus (House mouse)        | Chordata          | Mammalia    | Rodentia  | NM_205769 XM_357335          |
| CRF     | Canis lupus familiaris (Dog)      | Chordata          | Mammalia    | Carnivora | NM_00104278 XM_544106       |
| CRF     | Petromyzon marinus (Sea lamprey)  | Chordata          | Cephalaspidomorphi | Petromyzontiformes | Endsin et al. (35) |
| CRF     | Latimeria chalumnae (Coelacanth)  | Chordata          | Sarcopterygi | Coelacanthiformes | XM_006009123                 |
| CRF     | Chrysemys picta bellii (Western painted turtle) | Chordata | Reptilia | Testudines | XM_005288086                 |
| CRF     | Asterias rubens (Common starfish) | Echinodermata     | Asteroidea  | Forcipulatida | Semmens et al. (35)       |
| CRF     | Ciona intestinalis (Sea tunicate) | Urochordata       | Asciacea    | Enterogona | Lovejoy and Barsyte-Lovejoy, (37) |
| TCN     | Oryzias latipes (Japanese rice fish) | Chordata          | Actinopterygi | Beloniformes | BAR80710.1                |
| SVG     | Pachymedusa dacnicolor (Mexican tree frog) | Chordata | Amphibia | Anura | FR846380                     |
| SVG     | Physomedusa sauagvi               | Chordata          | Amphibia    | Anura     | AAY21509.1                   |
| UCN     | Pseudopodicoces humilis (Tibetan ground-tit) | Chordata | Aves | Passeriformes | XM_05532743                 |
| UCN     | Melopsittacus undulatus (Budgerigar) | Chordata          | Aves        | Psittaciformes | XM_005140144                |
| UCN     | Mus musculus (House mouse)        | Chordata          | Mammalia    | Rodentia  | NM_021290                    |
| UCN     | Canis lupus familiaris (Dog)      | Chordata          | Mammalia    | Carnivora | XM_648667                    |
| UCN     | Alligator mississippiensis (American alligator) | Chordata | Reptilia | Crocodylia | XM_006273603                 |
| UCN2    | Sus scrofa (Pig)                  | Chordata          | Mammalia    | Artiodactyla | XM_00312185                  |
| UCN2    | Bos taurus (Cattle)               | Chordata          | Mammalia    | Artiodactyla | XM_005223016                 |
| UCN2    | Mus musculus (House mouse)        | Chordata          | Mammalia    | Rodentia  | NM_145077 XM_910603          |
| UCN2    | Canis lupus familiaris (Dog)      | Chordata          | Mammalia    | Carnivora | XM_005632587                 |
| UCN2    | Oryzias latipes (Japanese rice fish) | Chordata          | Actinopterygi | Beloniformes | NP_001121991.1             |
| UCN3    | Anas platyrhynchos (Mallard)      | Chordata          | Aves        | Anseriformes | XM_005014655                 |
| UCN3    | Mus musculus (House mouse)        | Chordata          | Mammalia    | Rodentia  | NM_031250                    |
| UCN3    | Bos taurus (Cattle)               | Chordata          | Mammalia    | Artiodactyla | XM_001076527                 |
| UCN3    | Alligator sinensis (Chinese alligator) | Chordata | Reptilia | Crocodylia | XM_006035307                 |
| UCN3    | Petromyzon marinus (Sea lamprey)  | Chordata          | Cephalaspidomorphi | Petromyzontiformes | Endsin et al. (35) |
| UCN3    | Oryzias latipes (Japanese rice fish) | Chordata          | Actinopterygi | Beloniformes | NP_001121992.1             |
| UCN     | Oncorhynchus mykiss (Rainbow trout) | Chordata          | Actinopterygi | Salmoniformes | XM_001124343                 |
| UCN     | Danio rerio (Zebrafish)           | Chordata          | Actinopterygi | Cypriniformes | XM_001030180 XM_687090     |
| UCN     | Carassius auratus (Goldfish)      | Chordata          | Actinopterygi | Cypriniformes | AF129115                     |
| UCN     | Petromyzon marinus (Sea lamprey)  | Chordata          | Cephalaspidomorphi | Petromyzontiformes | Endsin et al. (35) |
| DH 31   | Drosophila melanogaster (Fruit fly) | Arthropoda       | Insecta | Diptera | NM_078790                    |
| DH 40   | Bombyx mori (Domestoc silkworm)   | Arthropoda       | Insecta | Lepidoptera | AB298934                    |
| DH 31   | Rhodius prolivus (Assassin bug)   | Arthropoda       | Insecta | Hemiptera | HM007016                     |
| DH      | Balanus amphitrite (Barnacle)     | Arthropoda       | Maxillopodite | Sessilia | JQ864196                     |

**CALCITONIN FAMILY**

| CALC   | Gallus gallus (Chicken)          | Chordata          | Aves        | Galliformes | X03012                     |
| CALC   | Bos taurus (Cattle)              | Chordata          | Mammalia    | Artiodactyla | AB462435                   |
| CALC   | Mus musculus (House mouse)       | Chordata          | Mammalia    | Rodentia   | X97991                     |
| CALC   | Alligator sinensis (Chinese alligator) | Chordata | Reptilia | Crocodylia | XM_006018232                |
| CALC   | Chrysemys picta bellii (Western painted turtle) | Chordata | Reptilia | Testudines | XM_005303304                |
| CALC   | Ciona intestinalis (Sea tunicate) | Urochordata       | Asciacea    | Enterogona | AB485672                   |
| CALC   | Equus caballus (Horse)           | Chordata          | Mammalia    | Perissodactyla | AF249307                  |
| CALC   | Chinchilla lanigera (Long-tailed chinchilla) | Chordata | Mammalia | Rodentia   | XP_005380394.1             |

(Continued)
| Hormone | Organism | Phylum | Class | Order | Accession number |
|---------|----------|--------|-------|-------|------------------|
| CALC    | Spermophilus tridecemlineatus (Squirrel) | Chordata | Mammalia | Rodentia | XM_005326775 |
| CGRP1   | Gallus gallus (Chicken) | Chordata | Aves | Galliformes | NM_001113706 |
| CGRP1   | Danio rerio (Zebrafish) | Chordata | Actinopterygii | Cypriniformes | NM_001002471 |
| CGRP1   | Salmo salar (Atlantic salmon) | Chordata | Actinopterygii | Salmoniformes | NM_001146580 |
| CGRP1   | Bos taurus (Cattle) | Chordata | Mammalia | Artiodactyla | NM_001076340 |
| CGRP2   | Canis lupus familiaris (Dog) | Chordata | Mammalia | Carnivora | NM_001003266 |
| CGRP2   | Callorhinus milii (Elephant shark) | Chordata | Chondrichthyes | Chimaeriformes | XM_007887527 |
| CGRP2   | Bos taurus (Cattle) | Chordata | Mammalia | Artiodactyla | NM_001134662 |
| CGRP2   | Canis lupus familiaris (Dog) | Chordata | Mammalia | Carnivora | NM_001003266 |
| CGRP2   | Mus musculus (House mouse) | Chordata | Mammalia | Rodentia | NM_054084 |
| CGRP2   | Sus scrofa (Pig) | Chordata | Mammalia | Artiodactyla | NM_001120473 |
| CGRP2   | Oryzias latipes (Japanese rice fish) | Chordata | Actinopterygii | Beloniformes | NM_001104894 |
| Amylin  | Columba livia (Rock pigeon) | Chordata | Aves | Columbiformes | NM_005504096 |
| Amylin  | Callorhinus milii (Elephant shark) | Chordata | Chondrichthyes | Chimaeriformes | XM_007887527 |
| Amylin  | Latimeria chalumnae (Coelacanth) | Chordata | Sarcopterygii | Coelacanthiformes | XM_005998862 |
| Amylin  | Bos taurus (Cattle) | Chordata | Mammalia | Artiodactyla | NM_001195038 |
| Amylin  | Mus musculus (House mouse) | Chordata | Mammalia | Rodentia | NM_010491 |
| Amylin  | Sus scrofa (Pig) | Chordata | Mammalia | Artiodactyla | NM_001120473 |
| Amylin  | Alligator mississippiensis (American alligator) | Chordata | Reptilia | Crocodylia | XM_006270879 |
| ADM     | Xenopus tropicalis (Western clawed frog) | Chordata | Amphibia | Anura | NM_002938340 |
| ADM     | Columba livia (Rock pigeon) | Chordata | Aves | Columbiformes | NM_005493377 |
| ADM     | Carassius auratus (Goldfish) | Chordata | Actinopterygii | Cypriniformes | EU000533 |
| ADM     | Latimeria chalumnae (Coelacanth) | Chordata | Sarcopterygii | Coelacanthiformes | XM_005998862 |
| ADM     | Sus scrofa (Pig) | Chordata | Mammalia | Artiodactyla | D14875 |
| ADM     | Bos taurus (Cattle) | Chordata | Mammalia | Artiodactyla | NM_173888 |
| ADM     | Canis lupus familiaris (Dog) | Chordata | Mammalia | Carnivora | AB191461 |
| ADM     | Mus musculus (House mouse) | Chordata | Mammalia | Rodentia | NM_009627 |
| ADM2    | Xenopus tropicalis (Western clawed frog) | Chordata | Amphibia | Anura | NM_002938340 |
| ADM2    | Galus gallus (Chicken) | Chordata | Aves | Galliformes | NM_001040464 |
| ADM2    | Latimeria chalumnae (Coelacanth) | Chordata | Sarcopterygii | Coelacanthiformes | NM_006120149 |
| ADM2    | Alligator mississippiensis (American alligator) | Chordata | Reptilia | Crocodylia | XM_006275255 |
| GHRH    | Xenopus laevis (African clawed frog) | Chordata | Amphibia | Anura | NM_001096728 |
| GHRH    | Galus gallus (Chicken) | Chordata | Aves | Galliformes | NM_001040464 |
| GHRH    | Callorhinus milii (Elephant shark) | Chordata | Chondrichthyes | Chimaeriformes | XM_007885752 |
| GHRH    | Danio rerio (Zebrafish) | Chordata | Actinopterygii | Cypriniformes | NM_001080092 |
| GHRH    | Bos taurus (Cattle) | Chordata | Mammalia | Artiodactyla | NM_1738325 |
| GHRH    | Canis lupus familiaris (Dog) | Chordata | Mammalia | Carnivora | NM_001290112 |
| GHRH    | Mus musculus (House mouse) | Chordata | Mammalia | Rodentia | NM_0102026 |
| GHRH    | Sus scrofa (Pig) | Chordata | Mammalia | Artiodactyla | NM_001195118 |
| GHRH    | Anolis carolinensis (Green anole) | Chordata | Reptilia | Squamata | XM_003225928 |
| GHRH    | Chrysemys picta bellii (Western painted turtle) | Chordata | Reptilia | Testudines | XM_005295322 |
| GiP     | Xenopus laevis (African clawed frog) | Chordata | Amphibia | Anura | NM_001097922 |
| GiP     | Galus gallus (Chicken) | Chordata | Aves | Galliformes | NM_001080104 |
| GiP     | Danio rerio (Zebrafish) | Chordata | Actinopterygii | Cypriniformes | NM_001080509 |
| GiP     | Bos taurus (Cattle) | Chordata | Mammalia | Artiodactyla | NM_001166605 |
| GiP     | Mus musculus (House mouse) | Chordata | Mammalia | Rodentia | NM_008119 |

(Continued)
| Hormone | Organism | Phylum | Class | Order | Accession number |
|---------|----------|--------|-------|-------|------------------|
| GIP     | Sus scrofa (Pig) | Chordata | Mammalia | Artiodactyla | NM_001287408 |
| GIP     | Alligator mississippiensis (American alligator) | Chordata | Reptilia | Crocodylia | XM_006277990 |
| GIP     | Chelonia mydas (Green sea turtle) | Chordata | Reptilia | Testudines | XM_007091197 |
| GCG     | Gallus gallus (Chicken) | Chordata | Aves | Galliformes | NM_205260 |
| GCG     | Callorhinus milii (Elephant shark) | Chordata | Chondrichthyes | Chimaeriformes | XM_007889848 |
| GCG     | Bos taurus (Cattle) | Chordata | Mammalia | Artiodactyla | NM_173916 |
| GCG     | Sus scrofa (Pig) | Chordata | Mammalia | Artiodactyla | NM_214324 |
| GCG     | Ovis aries (Sheep) | Chordata | Mammalia | Cetartiodactyla | XM_004004659 |
| GCG     | Alligator mississippiensis (American alligator) | Chordata | Reptilia | Crocodylia | XM_006277994 |
| GCG     | Poecilia formosa (Amazon molly) | Chordata | Actinopterygii | Cyprinodontiformes | XM_007546594 |
| GCG     | Latimeria chalumnae (Coelacanth) | Chordata | Sarcopterygii | Coelacanthiformes | XM_006004345 |
| PACAP   | Xenopus laevis (African clawed frog) | Chordata | Amphibia | Anura | AF187877 |
| PACAP   | Gallus gallus (Chicken) | Chordata | Aves | Galliformes | MY956323 |
| PACAP   | Chthamalus glareolus (Grass carp) | Chordata | Actinopterygii | Cypriniformes | EF592488 |
| SCT     | Taeniopogon nigritus (Zebra finch) | Chordata | Aves | Passeriformes | NM_001256233 |
| SCT     | Mus musculus (House mouse) | Chordata | Mammalia | Rodentia | X73580 |
| SCT     | Sus scrofa (Pig) | Chordata | Mammalia | Artiodactyla | XM_003122391 |
| SCT     | Equus caballus (Horse) | Chordata | Mammalia | Perissodactyla | XM_003362642 |
| SCT     | Chelonia mydas (Green sea turtle) | Chordata | Reptilia | Testudines | XM_007091197 |
| VIP     | Xenopus laevis (African clawed frog) | Chordata | Amphibia | Anura | NM_001085714 |
| VIP     | Columba livia (Rock pigeon) | Chordata | Aves | Columbiformes | XM_005507654 |
| VIP     | Gallus gallus (Chicken) | Chordata | Aves | Galliformes | NM_205366 |
| VIP     | Danio rerio (Zebrafish) | Chordata | Actinopterygii | Cypriniformes | NM_001114553 |
| VIP     | Bos taurus (Cattle) | Chordata | Mammalia | Artiodactyla | AF503910 |
| VIP     | Canis lupus familiaris (Dog) | Chordata | Mammalia | Carnivora | XM_005615524 |
| VIP     | Mus musculus (House mouse) | Chordata | Mammalia | Rodentia | NM_011702 |
| VIP     | Sus scrofa (Pig) | Chordata | Mammalia | Artiodactyla | NM_001195233 |
| VIP     | Alligator mississippiensis (American alligator) | Chordata | Reptilia | Crocodylia | XM_006277994 |

**OTHER**

| Hormone | Organism | Phylum | Class | Order | Accession number |
|---------|----------|--------|-------|-------|------------------|
| NPY     | Mus musculus (House mouse) | Chordata | Mammalia | Rodentia | EDK98613.1 |
| NPY     | Danio rerio (Zebrafish) | Chordata | Actinopterygii | Cypriniformes | AA62071.1 |
| NPY     | Carassius auratus (Goldfish) | Chordata | Actinopterygii | Cypriniformes | AA49186.1 |
| NPY     | Rattus norvegicus (Norway rat) | Chordata | Mammalia | Rodentia | NP_036746.1 |
| NPY     | Gallus gallus (Chicken) | Chordata | Aves | Galliformes | NP_900804.1 |
| NPY     | Xenopus laevis (African clawed frog) | Chordata | Amphibia | Anura | AAH80151.1 |
| NPY     | Ovis aries (Sheep) | Chordata | Mammalia | Cetartiodactyla | NP_01009452.1 |
| NPY     | Bos taurus (Cattle) | Chordata | Mammalia | Artiodactyla | ACH61954.1 |
| NPY     | Homo sapiens (Human) | Chordata | Mammalia | Primates | NP_000896.1 |
| NPY     | Columba livia (Rock pigeon) | Chordata | Mammalia | Rodentia | NP_01269740.1 |
| NPY     | Pseudopodoces humilis (Ground tit) | Chordata | Aves | Passeriformes | XP_005518939.1 |
| NPY     | Callorhinus milii (Elephant shark) | Chordata | Chondrichthyes | Chimaeriformes | AC22970.1 |
| NPY     | Chrysopsitta picta bellii (Western painted turtle) | Chordata | Reptilia | Testudines | XP_005290023.1 |
| INS     | Ciona intestinalis (Sea urchin) | Urochordata | Asciidacea | Enterogona | NP_001123204.1 |
| INS     | Mus musculus (House mouse) | Chordata | Mammalia | Rodentia | AFB4850.1 |
| INS     | Rattus norvegicus (Norway rat) | Chordata | Mammalia | Rodentia | AAA4139.1 |
| INS     | Ictidomys tridactylus (Ground squirrel) | Chordata | Mammalia | Rodentia | AAK72558.1 |
| INS     | Ovis aries (Sheep) | Chordata | Mammalia | Cetartiodactyla | AAB60625.1 |
| INS     | Gallus gallus (Chicken) | Chordata | Aves | Galliformes | NP_900553.1 |

(Continued)
phylogenetic analysis of the Secretin superfamily in relation to TCAP as a putative progenitor revealed relationships among Secretin superfamily members. Calcitonin and insulin families are sister lineages and they are much more closely related to one another than was previously thought. Also, calcitonin and insulin are sister lineages that form distinct lineages to CRF and secretin families. Therefore, placing TCAP as an ancestor of the Secretin superfamily allowed a novel interpretation of evolutionary relationships among Secretin superfamily members.

Sequence Analysis of TCAP Paralogs and Orthologs

Sequence analysis of both TCAP paralogs and orthologs revealed that this family of peptides is highly conserved. The presence of a conserved “PELAD” motif among TCAP orthologs and paralogs, suggests that it may possess a functional attribute, such as a receptor-binding or activation site (27). Also, some characteristic amino acids are retained throughout orthologs and paralogs. Arginine (R) and lysine (K) residues are retained in some parts of the mature peptide and they are often characteristic of the presence of cleavage sites. Glycine (G) and proline (P) are also highly conserved and these amino acids have a tendency to be retained as their secondary structure can break the α-helical structure of peptides. A peptide system with such a large amount of conservation is indicative of great functional importance that may have been selected for. Therefore, the high sequence conservation among TCAP orthologs and paralogs suggests that this peptide system is evolutionarily ancient and may have been strongly selected for throughout evolutionary time.

Evolutionary Analysis of Pre-propeptides and Mature Peptides of Secretin Superfamily and TCAP Family Members

Phylogenetic analysis of Secretin superfamily pre-propeptides (composed of the signal, cryptic, and mature peptide) and TCAP family pre-peptides (composed of the cryptic and mature peptide) was undertaken in order to elucidate the relationships among these peptides. Analysis revealed that calcitonin, CRF,
secretin, and TCAP families formed distinct groups. Despite being chosen to serve as a reference group because it binds to a tyrosine kinase receptor and not a GPCR, insulin formed a group with calcitonin, suggesting that they may be sister lineages (Figure 3). The close relationship between calcitonin and insulin has previously been explored where Wimalawansa (44) suggested that insulin and calcitonin families are closely related. This is supported by phylogenetic analysis of the pre-propeptides and suggests that insulin and calcitonin are sister lineages. When the tree was rooted to TCAP (Figure 3), to establish the assumption that TCAP is the ancestor, CRF, calcitonin, and secretin families formed distinct groups. This evolutionary analysis suggests that the secretin family forms a separate clade that is a sister to CRF and calcitonin families, which, in turn, are sisters to one another. This is consistent with what has been observed with respect to Secretin GPCR evolution, where CRF and calcitonin receptors share the greatest amount of sequence similarity among Secretin GPCRs (17). Therefore, it is possible that a similar evolutionary scheme occurred with respect to the ligands. Thus, analysis of Secretin superfamily pre-propeptides with TCAP prepropeptides suggests that insulin and calcitonin are closely related sister lineages, that calcitonin-insulin and CRF lineages are closely related and that calcitonin-insulin and CRF form a distinct sister lineage to the secretin family.

Subsequently, phylogenetic analysis was performed with the mature peptides of Secretin superfamily members and the TCAP family. The analysis of TCAP family mature peptide sequences with calcitonin and insulin mature sequences (Figure 4) demonstrated that insulin A chains were closely related to mature calcitonin peptides. This suggests that the insulin A mature chain is more closely related to the calcitonin family than the insulin B mature chain, which is different from what was previously suggested by Wimalawansa (44). Subsequent analyses involving CRF, calcitonin, insulin, and TCAP mature peptides (Figure 5) as well as secretin, CRF, calcitonin, insulin, and TCAP mature peptides (Figure 6) confirmed that the insulin A chain was more closely related to the calcitonin family than the insulin B chain. Taken together, insulin and calcitonin are closely related sister groups, which was also observed with the pre-propeptide analysis (Figure 3). Moreover, with respect to relationships among Secretin superfamily members, calcitonin-insulin, and CRF families are more closely related to one another than they are to secretin or TCAP, which is supported by the evolutionary scheme of their receptors, which also appear to be very closely related. Finally, secretin forms a sister lineage to a lineage that comprises both calcitonin-insulin and CRF families. This is consistent with what was observed for analysis of the pre-propeptides (Figure 3).

Considering the evidence with respect to the ancestral origin of the teneurin-TCAP system and in light of the findings presented here, it is possible to present two hypotheses for the evolutionary scheme of these peptides. The first suggests that an ancient TCAP-like peptide may have been the ancestor of the Secretin superfamily and that it evolved prior to the emergence of CRF, calcitonin, and secretin families. This is supported by the identification of TCAP in organisms prior to the protostome-deuterostome divergence, where as members of the Secretin superfamily have not been identified this early in evolution (31, 32, 34). The possibility of a second hypothesis, suggesting that the Secretin superfamily forms a parallel lineage to extant TCAP and that these two lineages evolved from a proto-CRF-calcitonin-secretin-TCAP ancestor that was related to all of these families, cannot be discounted. Due to sequence availability, phylogenetic analysis was performed using extant Secretin superfamily and TCAP sequences. As a result, both of these hypotheses are plausible. Future analysis should be undertaken in order to further investigate whether TCAP is a progenitor of the Secretin superfamily of ligands.

CONCLUSIONS

Taken together, phylogenetic analysis of members of the Secretin superfamily using TCAP as a putative progenitor demonstrated relationships among Secretin superfamily members. First, calcitonin formed a closely related sister lineage to insulin, particularly the insulin A chain with respect to mature peptides, but this was also observed with the pre-propeptides. Also, calcitonin-insulin and CRF families are more closely related to one another than they are to secretin or TCAP, which is supported by the evolutionary scheme of their receptors. Finally, secretin forms a sister lineage to a group that comprises both calcitonin-insulin and CRF. Therefore, given evidence that the teneurin-TCAP system arose as a result of a HGT event prior to the emergence of the Metazoa, as well as the previously established structural similarity of TCAP to calcitonin and CRF, members of the Secretin superfamily, the presented phylogenetic analysis allowed for the elucidation of relationships among members of the Secretin superfamily. To conclude, this is the first time that relationships among this family of peptides were resolved and because a progenitor peptide for the Secretin superfamily has not been elucidated, we present TCAP as a candidate progenitor.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

OM performed all analyses and completed the first draft of the paper. BC and NL provided technical guidance on the construction of the phylogenetic tree. DL oversaw the research program and completed the final draft of the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This research was supported by a grant from the Natural Sciences and Engineering Council of Canada (NSERC)-Canada Graduate Scholarship-Master’s (CGSM) program in addition to funding from the University of Toronto for OM.
REFERENCES

1. Vale W, Spiess J, Rivier C, Rivier J. Characterization of a 41-residue ovine hypothalamic peptide that stimulates secretion of corticotropin and β-endorphin. Science. (1981) 212:1394–7. doi: 10.1126/science.6267999

2. Monticucco PA, Neschen A, Escamper V. Structure of sauvagine, as vasoactive peptide from the skin of a frog. Hoppe-Seyler’s Z Physiol Chem. (1979) 360:1178.

3. Lederis K, Letter A, McMaster D, Moore G, Schlesinger D. Complete amino acid sequence of urotensin I, a hypotensive and corticotropin-releasing neupeptide from Catostomus. Science. (1982) 218:162–4. doi: 10.1126/science.6981844

4. Vaughan J, Donaldson C, Bittencourt J, Perrin MH, Lewis K, Sutton S, et al. Characterization of urotensin, a novel mammalian neupeptide related to fish urotensin-I and to CRF. Nature. (1995) 378:287–92. doi: 10.1038/378287a0

5. Lovejoy DA, Balmant RJ. Evolution and physiology of the corticotropin-releasing factor (CRF) family of neupeptides in vertebrates. Gen Comp Endocrinol. (1999) 155:1–22. doi: 10.1006/gecn.1999.7298

6. Lewis K, Li C, Perrin MH, Blount A, Kunitake K, Vaughan J, Arias CA, et al. Identification of urotensin III, an additional member of the corticotropin-releasing factor (CRF) family with high affinity for the CRF2 receptor. Proc Natl Acad Sci USA. (2001) 98:7570–5. doi: 10.1073/pnas.121165198

7. Reyes TM, Lewis K, Perrin MH, Kunitake K, Vaughan J, Arias CA, et al. Urocortin II: a member of the corticotropin releasing factor (CRF) neupeptide family that is selectively bound by type 2 receptors. Proc Natl Acad Sci USA. (2001) 98:2843–8. doi: 10.1073/pnas.05162398

8. Hsu SY, Hseuh AJW. Human stresscopin and stresscopin-related peptide are selective ligands for the type 2 corticotropin-releasing hormone receptor. Nat Med. (2001) 7:605–11. doi: 10.1038/87936

9. Cardoso JC, Bergqvist CA, Félix RC, Larhammar D. Corticotropin-releasing factor family of peptides. J Mol Biol. (2016) 57:73–86. doi: 10.1038/jme.2016-0051

10. Lovejoy DA. Structural evolution of urotensin-I: retaining ancestral functions before corticotropin-releasing hormone evolution. Gen Comp Endocrinol. (2009) 164:15–9. doi: 10.1016/j.ygcen.2009.04.014

11. Lovejoy DA, De Lannoy L. Evolution and phylogeny of the corticotropin-releasing factor (CRF) family of peptides: expansion and specialization in the vertebrates. J Neurochem. (2013) 54:50–6. doi: 10.1016/j.jchemneu.2013.09.006

12. Lovejoy DA, Jahan S. Phylogeny and Evolution of the corticotropin releasing factor family of peptides. Gen Comp Endocrinol. (2006) 146:1–8. doi: 10.1016/j.ygcen.2005.11.019

13. Lovejoy DA, Chang B, Lovejoy N, Del Castillo J. Molecular evolution and origin of the corticotropin-releasing hormone receptors. J Comp Endocrinol. (2014) 52:6043–79. doi: 10.1350/JME-13-0238

14. Coast GM. Insect diuretic peptides: structures, evolution and actions. Frontiers in Endocrinology | www.frontiersin.org 15 August 2020 | Volume 11 | Article 529

15. Schiöth H, Nordström K, Lovejoy DA, Barsyte-Lovejoy D. Characterization of a diuretic hormone-like peptide from tunicates: insight into the origins of the vertebrate corticotropin-releasing factor (CRF) family. Gen Comp Endocrinol. (2010) 165:330–6. doi: 10.1016/j.ygcen.2009.07.013

16. Michalec et al. CRF, an Ancient Peptide Family
38. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA 6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol.* (2013) 30:2725–9. doi: 10.1093/molbev/mst197
39. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* (2004) 32:1792–7. doi: 10.1093/nar/gkh340
40. Weiss M, Steiner DF, Philipson LH. Insulin biosynthesis, secretion, structure and structure-activity relationships. In: De Groot LJ, Chrousos G, Dungan K, editor. *Endotext.* South Dartmouth, MA: NIH/NLM/NCBI Books (2014). p. 1–6.
41. Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci.* (1992) 8:275–82.
42. Lang BF, O’Kelly C, Nerad T, Gray MW, Burger G. The closest unicellular relatives of animals. *Curr Biol.* (2002) 12:1773–8. doi: 10.1016/S0960-9822(02)01187-9
43. Minet A, Rubin B, Tucker R, Baumgartner S, Chiquet-Ehrismann R. Teneurin-1, a vertebrate homologue of the Drosophila pair-rule gene Ten-m, is a neuronal protein with a novel type of heparin-binding domain. *J Cell Sci.* (1999) 112:2019–32.
44. Wimalawansa SJ. Amylin, calcitonin gene-related peptide, calcitonin and adrenomedullin: a peptide superfamily. *Crit Rev Neurobiol.* (1997) 11:167–239. doi: 10.1615/CritRevNeurobiol.v11.i2-3.40

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Michalec, Chang, Lovejoy and Lovejoy. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.