The secretin G-protein-coupled receptor family: teleost receptors

J C R Cardoso, M S Clark1, F A Viera, P D Bridge1, A Gilles2 and D M Power

Centre of Marine Sciences, Universidade do Algarve, Campus de Gambelas, 8000-810 Faro, Portugal
1British Antarctic Survey, High Cross, Madingley Road, Cambridge, CB3 0ET, UK
2UPRES Biodiversité, Faculté des Sciences St Charles, 3 Place Victor Hugo, case 17, 13331 Marseilles cedex 3, France

Abstract

Twenty-one members of the secretin family (family 2) of G-protein-coupled receptors (GPCRs) were identified via directed cloning and data-mining of the Fugu Genome Consortium database, representing the most comprehensive description of secretin GPCRs in a teleost fish to date. Duplicated genes were identified for many of the family members, namely the receptors for pituitary adenylate cyclase-activating polypeptide (PACAP)/vasoactive intestinal peptide (VIP), calcitonin, calcitonin gene-related peptide (CGRP), growth hormone releasing hormone (GHRH), glucagon receptor/glucagon-like peptide (GLP) and parathyroid hormone-related peptide (PTHrP)/PTH. Mining of other teleost genomes (zebrafish and Tetraodon) revealed that the duplicated genes identified in the Takifugu genome were also present in these fish. Additional database searching of the Escherichia coli, yeast, Drosophila, Caenorhabditis elegans and Ciona genomes revealed that the family 2 of GPCRs were only present in the multicellular organisms. Orthologues of all the human secretin receptors were identified with the exception of secretin itself. Additional database searches in the Fugu Genome Consortium database also failed to reveal a secretin ligand and so it is hypothesised that both the receptor and the ligand evolved after the divergence of teleost/tetrapod lineages. Phylogenetic analysis at both the protein and the DNA level provided strong support for each of the individual receptor family groupings, but weak support between groups, making evolutionary inferences difficult. A more critical analysis of the PACAP/VIP receptor family confirmed previous hypotheses that the vasoactive intestinal peptide receptor (VPAC1R) gene is the ancestral form of the receptor.

Journal of Molecular Endocrinology (2005) 34, 753–765

Introduction

The G-protein-coupled receptors (GPCRs) are one of the largest known groups of proteins, comprising up to 80% of the total number of receptors identified in cells (Bockaert 1991, Bockaert et al. 2002). They are estimated to be between 1000 and 2000 members in vertebrates, which constitute approximately 1% of the genome. Although the different GPCR families share no sequence similarity, they do have certain common features: a central core domain of seven hydrophobic transmembrane (TM) helices connected by three intracellular and three extracellular loops. Two cysteine residues (one each in extracellular loops 1 and 2) are conserved in most GPCRs and form a di-sulphide bridge, which is thought to be important for the packing and the stabilisation of a restricted number of conformations of these seven TMs. GPCRs are involved in signal transduction, binding structurally diverse ligands – such as photons, odorants, biogenic amines, peptides and glycoproteins (Ulrich et al. 1998, Ulloa-Aguirre et al. 1999) – and mediate the extracellular signal by coupling via guanine nucleotide-binding proteins (G-proteins) to various secondary pathways involving ion channels, adenylyl cyclases and phospholipases. They may also couple with other proteins, for example those containing PDZ domains (Bockaert & Pin 1999). The specific properties of the various receptors are conferred by their N-terminal extracellular and C-terminal intracellular domains, and also their intracellular loops. They are among the oldest devices devoted to signal transduction, having been identified in plants, protozoa and the earliest diploblastic metazoa (Vernier et al. 1995, New & Wong 1998, Plakidou-Dymock et al. 1998).

Genomic analyses of GPCRs have largely concentrated on categorisation into clades and families (Attwood & Findlay 1994, Kolakowski 1994, Horn et al. 1998, Fredriksson et al. 2003). This overall classification has changed little over the years and the most commonly used is the A–F (1–5) system (Attwood & Findlay 1994, Kolakowski 1994, Horn et al. 1998). This overall classification has changed little over the years and the most commonly used is the A–F (1–5) system (Attwood & Findlay 1994, Kolakowski 1994, Horn et al. 1998): A, rhodopsin-like; B, secretin-like; C, metabotrophic glutamate/pheromone; D, fungal pheromone; E, cAMP (Dictostelium); F, frizzled/smoothened family). However, there are problems with global-type analyses: not all classes of GPCR
receptors exist in all organisms and some families are currently represented by a very limited number of species. So although a number of in silico phylogenetic studies have been conducted (Josefsson 1999, Graul & Sadee 2001, Fredriksson et al. 2003), it is virtually impossible to make global predictions about the evolutionary origins of the GPCRs. Some analyses have concluded that there is potentially a single common ancestor for all GPCRs (Graul & Sadee 2001, Fredriksson et al. 2003), while others do not disregard multiple origins and indicate that it is not possible to strictly exclude either convergent evolution or combinatorial evolution (Fryxell 1995, Bockaert & Pin 1999, Josefsson 1999) as contributing factors.

Looking at some of the different classifications, the members of one group remain fairly constant; the secretin family (also known as family B, family 2). The family name arose because secretin was the first ligand to be isolated from this group by Bayliss and Starling (1902). These receptors are activated by large peptides such as hormones and neuropeptides and are characterised by the existence of a large N-terminal domain, with at least six highly conserved cysteines that are proposed to be involved in ligand binding. The ligands are well characterised and have been isolated from a whole range of organisms such as tunicates, insects and vertebrates (reviewed in Campbell and Scanes (1992) and Sherwood et al. (2000)). In mammals this family comprises receptors for the peptides: secretin, vasoactive intestinal peptide (VIP), pituitary adenylate cyclase-activating polypeptide (PACAP), corticotrophin releasing factor (CRF), growth hormone releasing hormone (GHRH), parathyroid hormone (PTH), PTH-related peptide (PTHrP), glucagon, glucagon-like peptide (GLP), calcitonin and calcitonin gene-related peptide (CGRP). While the ligands have been isolated and comprehensively studied, the majority of the receptors are poorly characterised with sequences mainly available from the higher vertebrates.

To date our research has focussed on the isolation and characterisation of ligands and receptors from this family in the teleost fish Takifugu rubripes (Power et al. 2000, 2002, Clark et al. 2002, Cardoso et al. 2003a,b, 2004). Most of these were identified either via heterologous hybridisation or degenerate PCR techniques. However, the availability of the Takifugu genome sequence (Aparicio et al. 2002) has enabled comprehensive data-mining for teleost orthologues of this receptor family and characterisation of additional family members. Here we present the results of that data-mining and also the searches for orthologues in other species where genome information is publicly available.

Methods

Database searches were carried out on the Fugu Genome Consortium database (http://www.fugu.mrc.ac.uk, October 2001 release) to search for all members of the secretin GPCRs in the Takifugu genome. Searches were carried out using the human orthologous sequences for the GHRH, vasoactive intestinal polypeptide (VPAC1R and VPAC2R), pituitary adenylate cyclase activating polypeptide (PAC1R), corticotrophin releasing factor (CRF1 and CRF2), glucagon (GLR), GLP1 and GLP2, parathyroid hormone (PTH2 and PTR), glucose-dependent insulinotrophic polypeptide (GIP), secretin (SCTR), calcitonin (CALR) and CGRP receptors via the BLASTn programme using the Blosum62 matrix and an expected value of 10. The genomic organisation of the receptors in the newly identified Takifugu scaffolds was carried out using the manual identification of introns and exons searching for the consensus splice site sequences (AG/GT). The transmembrane domains were characterised using the GPCR information on the PRINTS database (http://www.bioinfo.man.ac.uk/dbbrowser/PRINTS), accession PR00249 GPCRSECRETIN, accessed via SRS (http://srs.hgmp.mrc.ac.uk).

To supplement the secretin family of GPCRs from the publicly annotated Swissprot and Trembl datasets each of the Takifugu receptor genes were used to search for similar genes in the zebrafish, medaka, Tetraodon, Caenorhabditis elegans, Drosophila melanogaster, Ciona (Ciona intestinalis) and Saccharomyces cerevisiae databases (Table 1).

<table>
<thead>
<tr>
<th>Table 1 Database URLs for organisms used in this study</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Organism</strong></td>
</tr>
<tr>
<td>Saccharomyces cerevisiae</td>
</tr>
<tr>
<td>Caenorhabditis elegans</td>
</tr>
<tr>
<td>Ciona intestinalis</td>
</tr>
<tr>
<td>Drosophila melanogaster</td>
</tr>
<tr>
<td>Zebrfish (Danio rerio)</td>
</tr>
<tr>
<td>Medaka (Ory latipes)</td>
</tr>
<tr>
<td>Tetraodon (T. nigroviridis)</td>
</tr>
</tbody>
</table>

Downloaded from Bioscientifica.com at 10/29/2018 11:13:12PM via free access
Twenty-one putative members of the secretin family of GPCRs were identified in Takifugu (Table 2). Data on some of the TM domains was incomplete due to either gaps in the genomic sequence or regions unidentifiable using BLAST sequence similarity searching. Therefore data analysis was restricted to TM regions common to all identified genes (TMs 2, 4, 5 and 6) in Takifugu. Because of the conservative motif in the TM domain, it was difficult to postulate that no exon shuffling had occurred in this gene family (i.e. that TM domain 1 for species 1 was the orthologue of TM domain 1 for species 2) (Fig. 1) (Taylor & Agarwal 1993, Patthy 1999). Considering this potential alignment artefact, we developed the following approach: in a preliminary analysis all the TM domains for each sequence were extracted and separated, and a multiple alignment was constructed using Clustal W (Thompson et al. 1994) with the default parameters (382 TM domains from 98 species). Phylogenetic trees were constructed based on neighbor-joining (NJ) (Saitou & Nei 1987), with Poisson correction distance and pairwise deletion comparison (midpoint rooting). The monophyly of each TM domain was distinguished by high distance values and supported by high bootstrap proportion (BP>69%) (Felsenstein 1985). The internal branch test was used when the number of taxa was higher than the number of sites (Sitnikova 1995).

The concatenation of the TM sequences (TMs 2, 4, 5 and 6) was carried out after the verification of the orthologous relationship of each of the TM domains. These sequences were then subjected to further phylogenetic analysis using three methods: maximum parsimony, neighbor-joining and minimal evolution. The sequences were subjected to principal component analysis via the clustalw package v1.7 and Jalview v1.7-5b option available from the EBI (http://www.ebi.ac.uk/clustalw/).

In-depth phylogenetic analysis concentrated on the PACAP family members (PAC1R, VPAC1R and VPAC2R). Branch lengths differed considerably between different receptor groups, so, as a test, the nucleotide sequences of the designated TM domains were compared with the protein sequences in all members of the VPAC2 and PAC1 receptors. Comparison of branch length differences produced using DNA and protein sequences from the PAC1 and VPAC2 receptors was analysed using both the neighbor-joining (Phylip neighbour programme) and a neighbor-joining consensus tree (Phylip seqboot, neighbor and consense programmes) with 500 bootstrap replicates and also the neighbour-joining method (Saitou & Nei 1987) via the PHYLIO_WIN interface v1.2 (Galtier et al. 1996) Phylowin with 500 bootstrap replicates. The tree was produced using the postscript output from the Phylowin programme. ClustalW was used to calculate percentage identities for the consensus sequences. For the more rigorous cDNA nucleotide sequences were downloaded from the databases, coding sequence extracted using extractseq (EMBOSS) and aligned using clustalw. To minimise long branch length artefacts, the most 5’ and 3’ ends of each sequence were removed, as these are the most evolutionary diverged regions. The sequences were converted to Phylip interleaved format using the EMBOSS (Rice et al. 2000) seqretr programme for analysis using the Phylip package (Felsenstein 1985). Transition/transversion ratios were calculated using TREE-PUZZLE v.5. A DNA distance matrix was calculated using the Phylip (v3-6 (alpha 2)) dnadist programme v3-6a.2-1. Default parameters were used except $D (D=distance)$ was set to the Kimura-2-parameter method and the transition/transversion ratio set to 1:23. Phylogenetic trees were constructed from the nucleotide data using maximum likelihood (Phylip dnaml programme with global rearrangements), neighbor-joining (Phylip neighbor programme) and a neighbor-joining consensus tree (Phylip seqboot, neighbor and consense programmes) with 500 bootstrap replicates. Trees were drawn with Treetool (Olsen et al. 1992), saved in NEWICK format and produced as .gif output using the Phylodendron tree-print programme: http://www.es.ebi.ac.uk/Doc/phylodendron/treeprint-form.html.

Results

Twenty-one members of the secretin family of GPCRs were identified via directed cloning and data-mining of the Fugu Genome Consortium database (Table 2). Orthologues of the vertebrate secretin receptors were identified with the exception of secretin, GIP and GLP2. Duplicated genes were identified for seven of the family members, namely the receptors for PACAP/VIP, calcitonin, CGRP, GHRH and PTHrP/PTH (Fig. 2).

Additionally, the teleost databases of Tetraodon, medaka and zebrafish, plus those of Drosophila, C. elegans, Ciona, E. coli and yeast were searched with the Takifugu VIP/PACAP/PTCR/CRF/CALR receptor coding sequences. No putative members of these receptors were found in the yeast and E. coli databases. Four clones were identified in Drosophila, namely: CG13758-PA, CG8422-PA, CG12370-PA and CG12370-PB; of these, the latter two appear to be alternative splice forms, so only the data from the former were included in the analysis. Each of the Drosophila clones clustered with the CRF receptors. In the C. elegans database, three clones with the accession numbers C13B9+4, ZK643+3 and C18B12+2 were found which shared high sequence similarity to the Takifugu VIP receptor sequences. Unrooted trees suggested that ZK643+3 was the most distantly related of the receptors and therefore the final rooted tree was produced using
Table 2 Identification of teleost secretin family GPCRs by data-mining.

<table>
<thead>
<tr>
<th>Putative receptor name</th>
<th>Scaffold</th>
<th>Status</th>
<th>TM domains characterised</th>
<th>Scaffold</th>
<th>TM domains characterised</th>
<th>Scaffold</th>
<th>TM domains characterised</th>
</tr>
</thead>
<tbody>
<tr>
<td>VPAC1A</td>
<td>S000279</td>
<td>Accession number: AJ296144</td>
<td>TMs 2, 4, 5, 6</td>
<td>002569-2</td>
<td>TMs 2, 4, 5, 6</td>
<td>Not identified</td>
<td></td>
</tr>
<tr>
<td>VPAC1B</td>
<td>S000668</td>
<td>Accession number: AJ296143</td>
<td>TMs 2, 4, 5, 6</td>
<td>00637-2</td>
<td>TMs 2, 4, 5, 6</td>
<td>Not identified</td>
<td></td>
</tr>
<tr>
<td>VPAC2A</td>
<td>S000593</td>
<td>Accession number: AJ408877</td>
<td>TMs 2, 4, 5, 6</td>
<td>027090-1</td>
<td>TMs 2, 4, 5, 6</td>
<td>Not identified</td>
<td></td>
</tr>
<tr>
<td>VPAC2B</td>
<td>S003112</td>
<td>Accession number: AJ408878</td>
<td>TMs 2, 4, 5, 6</td>
<td>016424-1</td>
<td>TMs 2, 4, 5, 6</td>
<td>Not identified</td>
<td></td>
</tr>
<tr>
<td>PAC1A</td>
<td>S0003311</td>
<td>Accession number: AJ490804</td>
<td>TMs 2, 4, 5, 6</td>
<td>005290-1</td>
<td>TMs 2, 4, 5, 6</td>
<td>Not identified</td>
<td></td>
</tr>
<tr>
<td>PAC1B</td>
<td>S003855/S004929</td>
<td>Accession number: AJ494861</td>
<td>TMs 2, 4, 5, 6</td>
<td>004002-2</td>
<td>TMs 2, 4, 5, 6</td>
<td>Not identified</td>
<td></td>
</tr>
<tr>
<td>GHRHA</td>
<td>S003311</td>
<td>Accession number: Q8AXV2</td>
<td>TMs 2, 4, 5, 6</td>
<td>004002-1</td>
<td>TMs 2, 4, 5, 6</td>
<td>Not identified</td>
<td></td>
</tr>
<tr>
<td>GHHRB</td>
<td>S004929</td>
<td>Incomplete</td>
<td>7</td>
<td>Not identified</td>
<td>Accession number: AF32082</td>
<td>Not identified</td>
<td></td>
</tr>
<tr>
<td>PTH2</td>
<td>S002092</td>
<td>Complete</td>
<td>7</td>
<td>005805-1</td>
<td>Accession number: AF132084/ZC163F10</td>
<td>Accession number: AF132085</td>
<td></td>
</tr>
<tr>
<td>PTR1</td>
<td>S000850</td>
<td>Complete</td>
<td>7</td>
<td>014169-1</td>
<td>TMs 2, 4, 5, 6</td>
<td>ZK121K17 TMs 2, 4, 5, 6</td>
<td></td>
</tr>
<tr>
<td>PTR3</td>
<td>S005032</td>
<td>Complete</td>
<td>7</td>
<td>004797-1</td>
<td>TMs 2, 4, 5, 6</td>
<td>Not identified</td>
<td></td>
</tr>
<tr>
<td>GLP/GLR/GIP</td>
<td>S000537</td>
<td>Incomplete</td>
<td>6: TM7 incomplete</td>
<td>286941-1</td>
<td>TMs 2, 4, 5, 6</td>
<td>Not identified</td>
<td></td>
</tr>
<tr>
<td>GLP/GLR/GIP</td>
<td>S000381</td>
<td>Incomplete</td>
<td>6: Frameshift TM5</td>
<td>168681-1</td>
<td>TMs 2, 4, 5, 6</td>
<td>Not identified</td>
<td></td>
</tr>
<tr>
<td>GLP/GLR/GIP</td>
<td>S006614</td>
<td>Incomplete</td>
<td>7</td>
<td>022047-1</td>
<td>TMs 2, 4, 5, 6</td>
<td>Not identified</td>
<td></td>
</tr>
<tr>
<td>GLP/GLR/GIP</td>
<td>S007267</td>
<td>Incomplete</td>
<td>5: TM5 &amp; TM6 incomplete</td>
<td>016099-1</td>
<td>TMs 2, 4, 5, 6</td>
<td>Not identified</td>
<td></td>
</tr>
<tr>
<td>CALR</td>
<td>S012353</td>
<td>Incomplete</td>
<td>7</td>
<td>006036-2</td>
<td>TMs 2, 4, 5, 6</td>
<td>Not identified</td>
<td></td>
</tr>
<tr>
<td>CALR</td>
<td>S012367</td>
<td>Incomplete</td>
<td>6: TM1 not identified</td>
<td>Not identified</td>
<td>ZC77K19 TMs 2, 4, 5, 6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CGRP</td>
<td>S001267</td>
<td>Incomplete</td>
<td>7</td>
<td>Not identified</td>
<td>005865-1</td>
<td>TMs 2, 4, 5, 6</td>
<td></td>
</tr>
<tr>
<td>CGRP</td>
<td>S000611</td>
<td>Incomplete</td>
<td>7</td>
<td>003195-1</td>
<td>TMs 4, 5, 6</td>
<td>Not identified</td>
<td></td>
</tr>
<tr>
<td>CRF1</td>
<td>S000102</td>
<td>Accession number: CAC82924</td>
<td>TMs 2, 4, 5, 6</td>
<td>Not identified</td>
<td>Not identified</td>
<td>Not identified</td>
<td></td>
</tr>
<tr>
<td>CRF2</td>
<td>S000850</td>
<td>Incomplete</td>
<td>7</td>
<td>Not identified</td>
<td>Not identified</td>
<td>Not identified</td>
<td></td>
</tr>
</tbody>
</table>

Scaffold identities are given for all receptors with the identification status of the TM domains (October 2001 data release for Takifugu). In T. rubripes and T. nigroviridis most genes are incomplete due to missing genomic information about the highly divergent 5' and 3' ends which could not be identified using BLAST sequence similarity searching. All T. rubripes scaffolds apart from S007267 contained TM domains 2, 4, 5 and 6; all T. nigroviridis scaffolds apart from S007267 contained TM domains 2, 4, 5 and 6; all T. nigroviridis scaffolds apart from S007267 contained TM domains 2, 4, 5 and 6. Receptor sequences lacking the core TM domains necessary for the analysis (2, 4, 5 and 6) were excluded from the phylogenetic analysis. Accession numbers have been given where the gene has been fully characterised.
ZK643·3 as the outgroup. Both C13B9·4 and C18B12·2 clustered with the CRF genes. The Ciona genome data revealed nine clones with high sequence similarity to the family 2 of GPCRs (Table 3), of which four clones (Cin 0372, 0752, 0005A and 0005B) clustered most strongly with the VIP/GHRH/PTR receptors. Although both the Tetraodon and zebrafish genome databases comprise almost 6× genome coverage each, identification of secretin family GPCR members was limited. Sixteen members were identified in Tetraodon (in ten of which the full complement of TMs 2, 4, 5 and 6 were deduced) and only five were identified in zebrafish (Table 2). Of the zebrafish sequences, it was only possible in three to deduce TMs 2, 4, 5 and 6 and one (ZC163F10) was orthologous to the PTR1 gene already in the protein databases. The medaka genome data so far comprise only 0·9× coverage and all of the contigs were too small to produce meaningful data for this project.

Three different tree methods were used to evaluate the protein TM data. The topology of the three trees was similar and the neighbor-joining tree is shown (Fig. 2). It clearly shows good support for there being three main groups within the secretin family of GPCR (CALR/CGRP, CRF1/CRF2 and VPAC/PAC1/secretin/GHRH/PTH/GLP/GIP). Bootstrap values (not shown) provided strong support for each of the distinct clusters, but relatively poor support linking the different groupings. Principal component analysis showed that the CALR/CGRP and CRF1/CRF2 arms were clearly distinct from the rest of the tree (Fig. 3), as were all the C. elegans and Drosophila receptors. Where the Ciona receptors grouped outside the main families (e.g. Cin0070, Cin0093, etc.), these too showed as separate blocks on the principal component analysis.

Using the protein sequences, some of the family groupings showed very different branch lengths on the neighbor-joining tree, this was particularly true of the VPAC2R and PAC1R sequences (Fig. 4). This potentially indicates that these sequences are undergoing different rates of evolution. The use of DNA sequences, as opposed to protein sequences discriminated the positions of the chicken (Gga) and Xenopus (Xla) PAC1 receptors better. Using DNA sequences, the receptors of these two organisms clustered together which is more logical from an evolutionary point of view than the scattered distribution produced using the protein sequences. In the latter the chicken PAC1 receptor clustered with the fish PAC1A receptors while the Xenopus PAC1 receptor clustered with the Takifugu PAC1B.

These were chosen for a more in-depth analysis, comparing protein and nucleotide conservation rates. ClustalW alignments of VPAC2R genes showed 54% absolute conservation of amino acids and 52% conserved nucleotides in the consensus sequences. Alignments of PAC1R genes showed 77% of amino acids were conserved, while the nucleotide figure was only 62% of nucleotides. Therefore there was a higher rate of mutation at the DNA level in the case of the PAC1 TMs, which was not represented at the protein level. Because we have a particular interest in the evolution of the PACAP receptor family members (PAC1R, VPAC1R, VPAC2R), further analyses were carried out using full-length DNA sequences. The topology of the tree indicates that VPAC1R was the probable ancestral receptor, with a fairly rapid duplication event producing VPAC2R and PAC1R (Fig. 5).

**Discussion**

Data-mining the Takifugu genome data identified 21 members of the secretin GPCR family. Clearly there may be other family members as the genome is not complete, but this study is the most representative description of the secretin receptor family in a teleost fish to date. Orthologues of each of the family members were identified with the exception of the secretin, GIP and GLP2 receptors. Four scaffolds (S000337, S000381, S006614 and S007267) were identified by BLAST sequence similarity searching, and also (in the case of the first three scaffolds) by phylogenetic analysis, as being
members of the GIP/GLR/GLP family. Unfortunately because of the narrow species range in the database, it was not possible to define accurately which particular family members the Takifugu genes were most similar to, although they did tend to group more with the GLR and GLP1 receptors rather than those for GIP and GLP2. Takifugu would be expected to have receptors for glucagon (GLR) and glucagon-like peptide (GLP1 and 2), as these have already been identified in a range of fish species. So far the gastric inhibitory peptide (GIP) has only been identified in mammals (reviewed in Sherwood et al. 2000) and therefore may not be present in fish. However, it should be noted that this was not present in a non-continuous dataset and, where there is ambiguity, analysis is best performed on the complete gene sequences; this was not possible in this instance, as the genome information for each Takifugu gene was incomplete and cDNA sequences were not available. The ends of all of the family 2 GPCR genes are highly divergent and it is often not possible to identify the most 5' and 3' ends using genomic sequence and sequence comparisons.

This whole receptor family was originally named after secretin (Bayliss & Starling 1902). Therefore, it is slightly ironic that so far this is the only receptor still to be identified in Takifugu (and indeed the other databases used). This is in spite of extensive database searching, library screening and degenerate PCR experiments. To further investigate this, screening of the Fugu Genome Consortium database for the ligand was also carried out, with negative results. In fact, the secretin receptor has so far only been characterised in mammals. In contrast, the gene for the ligand, secretin, has been identified in both mammals and birds (Sherwood et al. 2000). In the former group secretin is a potent stimulant of pancreatic secretion. In the latter group, avian and mammalian VIPs are more potent stimulants of the avian pancreas than secretin (Dockray 1975, 1979). Further work aimed at establishing when the secretin receptor first arose in tetrapods will be required. Two principal models exist for the evolution of secretin (Bell 1986, Okubo et al. 1992). One suggests that secretin and VIP shared a common ancestor and secretin arose recently (310 million years ago) by gene duplication, the other suggests that it evolved much earlier from an ancestral glucagon ligand (Campbell & Scanes 1992). Assuming a certain level of co-evolution of ligand and receptor, we propose that the absence of a secretin receptor in Takifugu favours the former model.

Table 3 Ciona intestinalis putative members of the secretin family of GPCRs identified by data-mining. Scaffolds were identified by BLAST searches and homology to the secretin family of GPCRs

<table>
<thead>
<tr>
<th>Organism</th>
<th>Homology</th>
<th>Scaffold</th>
<th>TM domains characterised</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. intestinalis</td>
<td>VPAC, PAC1, GHRH, CRF</td>
<td>2303</td>
<td>TM2 only</td>
</tr>
<tr>
<td>C. intestinalis</td>
<td>PTR</td>
<td>0752</td>
<td>TM2, 4, 5, 6</td>
</tr>
<tr>
<td>C. intestinalis</td>
<td>CGPR/CalR</td>
<td>0093</td>
<td>TM2, 4, 5, 6</td>
</tr>
<tr>
<td>C. intestinalis</td>
<td>VPAC, PAC1, GHRH, PTR</td>
<td>0005A</td>
<td>TM2, 4, 5, 6</td>
</tr>
<tr>
<td>C. intestinalis</td>
<td>VPAC, PAC1, GHRH, PTR, GIP/GLR/GLP2, GLR</td>
<td>0005B</td>
<td>TM2, 4, 5, 6</td>
</tr>
<tr>
<td>C. intestinalis</td>
<td>PTR</td>
<td>0070</td>
<td>TM2, 4, 5, 6</td>
</tr>
<tr>
<td>C. intestinalis</td>
<td>CRF</td>
<td>0372</td>
<td>TM2, 4, 5, 6</td>
</tr>
<tr>
<td>C. intestinalis</td>
<td>CRF</td>
<td>0273</td>
<td>TM2, 4, 5, 6</td>
</tr>
<tr>
<td>C. intestinalis</td>
<td>CRF</td>
<td>0050</td>
<td>TM2, 4, 5, 6</td>
</tr>
</tbody>
</table>

Figure 2 Neighbor-joining tree using the protein sequence from TMs 2, 4, 5 and 6. Accession numbers and abbreviations are as follows. PAC1: human (Hsa), P41586; rat (Rno), P32215; mouse (Mmu), P70205; bovine (Bta), Q29627; goldfish (Cau), O73769. VPAC1: goldfish (Cau), Q60308; frog (Rri), Q9 YHC6; chicken (Gga), Q9 IBG2; mouse (Mmu), P97751; rat (Rno), P30083; human (Hsa), P32241; pig (Ssc), Q28992. VPAC2: human (Hsa), P41587; rat (Rno), P35000; mouse (Mmu), P41588; SCTR: human (Hsa), P47872; rat (Rno), P23811; rabbit (Ocu), O46502. GHRH: rat (Rno), Q02644; mouse (Mmu), P32082; pig (Ssc), P34999; sheep (Oar), Q9 BD10; bovine (Bta), Q9 NIF8; human (Hsa), Q02643; goldfish (Cau), O73768; Fugu (Fru), Q8 AXV2. PH2: zebrfish (Dre), Q9 PW78; rat (Rno) P70555; human (Hsa) P49190. PTHR1: zebrafish (Dre), Q9 PVD3. PTHR2: North American possum (Dma), P25107; rat (Rno), P25961; mouse (Mmu), P41593; pig (Ssc) P50133; human (Hsa) Q03431; dog (Cfa) Q9TU31. GLP2: human (Hsa), Q95838; rat (Rno), Q9 ZOWO. GIP: golden hamster (Mau), Q93218; rat (Rno), Q43219; human (Hsa), P48546. GLP1: rat (Rno), P32301; mouse (Mmu), Q35699; human (Hsa), P43220. GLR: tiger frog (Rti), Q9 PUK1; mouse (Mmu), Q61606; rat (Rno), P30082; human (Hsa), P47871. CGRP: rat (Rno), Q34311; human (Hsa), Q16602; mouse (Mmu), Q9 Q2XH8; flounder (Pol), Q9IBB86. CALR: flounder (Pol), Q9IBB85; pig (Ssc), P25117; mouse (Mmu), Q60755; rat (Rno), P32214; guinea pig (Cpo), Q60893; human (Hsa), P30988; rabbit (Ocu), P79222. CRFR1: Pacific spotted dogfish (Ane), Q98UC2; chicken (Gga), Q90812; Fugu (Fru), QAC82924; toad (Xla), O42602; human (Hsa), P34998; sheep (Oar), Q62772; bovine (Bta), Q9 BGU4; mouse (Mmu), P35347; rat (Rno), P35353. CRFR2: toad (Xla), O42603; catfish (Ani), Q98UC0. Xenopus PAC1 taken from Hu et al. (2000). Chicken PAC1 taken from Peeters et al. (1999).
with the divergence of bony fish from the tetrapods (approximately 405 million years ago) occurring prior to the duplication that gave rise to secretin. It is interesting that in all of the phylogenetic trees generated in the present study each receptor family shows a very clear fish/bird/amphibian–mammalian split. So although most of the receptors evolved prior to the appearance of the fishes 450 million years ago, they may have evolved along distinct lines in the different phyla with the secretin receptor evolving later in response to a specific demand of the physiological adaptations necessary for the colonisation of the terrestrial environment. As these receptors do not appear to be present in the fungal lineage, but are present in the nematode lineage, it is possible to determine that they arose between 1600 and 1200 million years ago (Wang et al. 1999).

It is now well documented that fish contain a large number of duplicated genes (Wittbrodt et al. 1998). It is therefore no great surprise that a number of duplicated secretin family receptors are also present in the Takifugu genome. Orthologues were identified in Tetraodon for the duplicated PTR1, VPAC1A, VPAC1B, GHRH1A, PAC1A and VPAC2B Takifugu receptors. Several of the Tetraodon sequences and two zebrafish sequences also appeared to have Takifugu orthologues in the CALR/ CGRP and GIP/GLR groupings. However, topology within these families was less well defined and therefore accurate determination of duplicates is difficult.

The limited data available from Takifugu and the other fish species, indicates that the secretin family duplications are specific to the teleost lineage. The mechanism by which the piscine ‘extra’ genes were generated still remains an area of contention between two hypotheses. On the one hand, it is suggested that the teleost fish underwent a third round of whole genome duplication (Amores et al. 1998, Wittbrodt et al. 1998, Postlethwait et al. 2000, Taylor et al. 2003, Christoffels et al. 2004, Naruse et al. 2004, Vandepoele et al. 2004) with subsequent gene loss; alternatively, it is suggested that they were subjected to a whole range of smaller chromosomal, segmental or gene-wide duplications at different times through their evolutionary past (Robinson-Rechavi et al. 2001a,b). However, it is probably fair to say that the datasets available at the moment are still too restricted, particularly with regard to the range of fish species for which there is genomic information available. Therefore it is not yet possible to determine accurately with any statistical significance which of the two duplication processes was involved in the evolution of the ‘extra’ genes. However, these data, with their preponderance of additional Takifugu genes do provide further support towards the hypothesis of an extra whole genome duplication having taken place in the teleost ancestor in line with the findings of Amores et al. 1998, Wittbrodt et al. 1998, Postlethwait et al. 2000, Taylor et al. 2003, Christoffels et al. 2004,

Figure 3 Principal component analysis of the sequences in Fig. 2 showing distinct clusters of receptors. 1, CRF group; 2, VIP/PACAP cluster; 3, calcitonin and CGRP receptors. The receptors identified between these three main groups belong to Drosophila, C. elegans and the Ciona genes, which cluster outside the main tree groups.
Naruse et al. 2004 and Vandepoele et al. 2004. While the issue of an extra whole genome duplication event in fish remains a contentious issue, the preponderance of polyploid genomes within the fish certainly complicates genome analysis and a resolution of this issue (Lim et al. 1975, Ferris & Whitt 1977, Schmidtke et al. 1979, Allendorf & Thorgaard 1984, Larhammar & Risinger 1993, 1994). It also raises a further question, as to why these events occurred and the benefits for the organisms involved. It is generally acknowledged that gene duplication has played a significant role in the metazoan radiation (Ohno 1970). Further duplications in the fish could have fuelled their incredible speciation, as they comprise over half of all vertebrate species and, unlike mammals, their genomes (and the ploidy levels) are not constrained by a rigid sex chromosome system. The procession of duplicated genes with differential expression patterns and functional differences (reviewed in Postlethwait et al. 2004) may also, to a certain extent, obviate the need for alternative splicing. The duplications identified here enhance the dataset of duplicated fish genes and therefore will be available for more global analyses of this evolutionary process in fish. The identification of scaffolds will also allow exploitation of the Takifugu genes by groups with specific interests in functional characterisation of this gene set.

Without a doubt, there are problems with global GPCR analyses (similar to those presented here) when trying to determine evolutionary events and timings, the main one being that not all receptors are present in all species (for example, clan D which only comprises the fungal pheromone receptors) (Donnelly 1997) and some groups contain duplicated genes. GPCRs are present in large numbers in the genome (1000–2000 members in vertebrates and approximately 1100 in C. elegans (Bargmann et al. 1998)). Because of these huge numbers, some of the analyses have only been carried out on human data (Fredriksson et al. 2003) and also do not include non-GPCR receptors (Joseffson 1999, Fredriksson et al. 2003). So it is entirely possible that by only using GPCRs in the dataset, artificial groupings are produced, simply because there are no alternatives. In the present work, phylogenetic analysis of the secretin receptor family alone produces three distinct groupings of CALR/CGRP, CRF1/CRF2 and VPAC1,VPAC2,PAC1, secretin,GHRH/ PTR/GLP/GIP. This is mirrored in other broader phylogenetic analyses (Joseffsson 1999, Fredriksson et al. 2003). In the former analysis, the calcitonin, PACAP and glucagon receptors, while belonging to the same clade, branch separately. The more comprehensive analysis of Fredriksson et al. (2003) produces four branches within the same clade,
comprising: (i) VPAC₁R/VPAC₂R/PAC₁R/SCTR/GHRH; (ii) CRF; (iii) CALR; and (iv) GLP₁/GLP₂/GIP/GLR/PTH2/PTR. Global analyses can provide statistical confirmation of the main receptor family groups (e.g. GHRH, VPAC₁, VPAC₂, etc.). However, it must be remembered that such analyses are very restricted as regards species content, which can influence the tree topography. The relatedness between the different families is difficult to prove with any statistical certainty and principal component analysis of secretin receptor family members alone indicates no relationship between these three groupings identified in this analysis. Therefore once the datasets are restricted and multiple species added, the different families is difficult to prove with any statistical certainty and principal component analysis of secretin receptor family members alone indicates no relationship between these three groupings identified in this analysis. Therefore once the datasets are restricted and multiple species added, the differences between the receptor families magnify, indicating possible polyphyly or convergent evolution within the designated clade.

Reducing the content of the tree, as regards the different numbers of families, does not resolve this issue as confidence levels of topology are still weak between the different family members (data not shown) although, as stated above, within families they are very high. This situation does not change whether examining protein or DNA data; however it is important to use DNA sequences when higher resolution is required. This enables a more accurate measure of variation and is particularly useful when considering the evolution of closely related receptor families. This is exemplified by the amino acid/DNA comparisons of the PAC₁R and VPAC₂R receptors. PAC₁R is highly conserved at the protein level, but actually has a similar rate of variability at the DNA level when compared with VPAC₂R. Therefore our more in-depth analysis concentrated only on the PAC₁R, VPAC₁R and VPAC₂R family which share a common set of ligands (VIP and PACAP), but with different affinities depending on the receptor (Harmar et al. 1998). These are found in the brain and gut and share a certain level of common functionality. The most closely related gene family to the PACAP receptors is GHRH and secretin; these have very different functions and localisations, with the GRF ligand being found mainly in the brain and secretin in the pancreas (Sherwood et al. 2000).

The phylogenetic analyses of this family sub-set (Fig. 5) produced a primary branching between VPAC₁R and PAC₂R/VPAC₂R and then divergence of PAC₁R and VPAC₂R. Although the branch lengths are very small and similar, when considered with all the other evidence (presented below) it implies that VPAC₁R is the most ancestral form of this gene. This substantiates the less rigorous phylogenetic analyses and the evolutionary linkage data presented in Cardoso et al. (2004). The PACAP ligand and the receptor are the most conserved at the amino acid level and have been thought in the past to be the more ancient molecules of this grouping, certainly the PACAP ligand (along with a GRF-like peptide) was identified in tunicates (McRory & Sherwood 1997). PACAP has also been shown to be the pivotal molecule in the functional sense, being present in all organs and tissues in which this superfamily is expressed. DNA analysis indicates that the PACAP receptor (PAC₁R) evolved after VPAC₁R. This does not present a problem when considering the pharmacological properties of the receptors and the ligands as VPAC₁R and VPAC₂R both bind PACAP and VIP with equal affinity, whereas PAC₁R has developed a more specialised role, binding PACAP with greater affinity than VIP (Harmar et al. 1998). It will definitely require ligand-binding assays on the receptors from lower vertebrates to help determine the evolutionary origins of this receptor sub-set and to examine the whole question of ligand-receptor evolution. This is supported by chromosomal mapping of the receptors in human and rat. Both VPAC₂R and PAC₁R map to human chromosome 7 and rat chromosome 4, whereas VPAC₁R is located on human chromosome 3 and rat chromosome 8. Using this as a guide, a first duplication would produce the VPAC₁R gene and a common ancestor for VPAC₂R/PAC₁R. A second duplication event acting on the ancestral VPAC₂R/PAC₁R gene

Figure 5  Neighbor-joining tree using complete cDNA sequence (where possible). PAC₁: human (Hsa), D17516; rat (Rno), D16465; mouse (Mmu), D82935; bovine (Bta), D17290; goldfish (Cau), AF048820. VPAC₁: goldfish (Cau), U56391; frog (Rri), AF100644; chicken (Gga), AB029895; mouse (Mmu), D16465; rat (Rno), M86835; human (Hsa), L13288; pig (Ssc), U49434; Takifugu (FruVPAC1A). VPAC₂: human (Hsa), L40764; rat (Rno), Z25885; mouse (Mmu), D28132. For (Ssc), U49434; Takifugu (FruVPAC1A). VPAC₂: human (Hsa), L40764; rat (Rno), Z25885; mouse (Mmu), D28132. For Fugu accession numbers, see Table 2. C. elegans clone C18B122 is used as an outgroup.
would produce the two separate genes (Sreedharan et al. 1993, Cai et al. 1995, Brabet et al. 1996, Mackay et al. 1996, Vaudry et al. 2000). The distance between the VPAC₁R and PAC₁R/VPAC₂R nodes is relatively small, indicating that there was rapid duplication of the original VPAC₁R gene to form the PAC₁R/VPAC₂R ancestor. Timing of these duplication events is difficult because of the presence of the duplicated *Takifugu* genes and also the limited species range, which in most cases is restricted to only mammals and fish (the *Tetraodon* genes are only partial sequences and so could not be included in this analysis). This is exemplified by the VPAC₂R topology which shows long branch lengths for the *Takifugu* genes. These fish genes are very different to their mammalian orthologues with the duplicated genes being as different to each other as they are to their mammalian orthologues (Cardoso et al. 2004), indicating that there has been considerable species-specific evolution of this duplicated gene set in *Takifugu*.

Clearly the relationships between different family 2 GPCR members are complex and sequence similarity comparisons are only the first stage in the characterisation and categorisation process. Increasing the range of vertebrate species included in the phylogenetic analysis and reducing the dataset to clade level certainly adds value and increases the definition of inter-family relationships. Sequence similarity does not necessarily equate to function and it is still to be determined whether the *Takifugu* genes (and the duplicated ones, in particular) have the same range and functionality as the higher vertebrate secretin family receptors. Also it is not strictly accurate to describe the receptors identified in *Drosophila*, *Ciona* and *C. elegans* as ‘ancestral’ molecules purely on the basis of sequence similarity. It is really only with functional determination that more exact relationships can be defined. Indeed this is one of the main problems of wholesale data-mining that sequence similarity may not represent functional equivalence, particularly when comparing such a diverse range of species. Indeed, reviewing the differences between GPCRs from all species, the question is raised whether membership of the club is purely down to possession of physical components, such as the number of TM domains, conserved cysteine residues and N-glycosylation sites. A more comprehensive species collection of GPCR data and functional information will start to answer this question.

### Funding

This work was supported by a FCT pluriannual grant to Centro de Ciências do Mar do Algarve (CCMAR), Portugal. J C R Cardoso was funded by FCT, Portugal (grant BD/19925/99). This paper was produced by MSC within the LATEST core programme.

### References


Cai Y, Xin X, Yamada T, Muramatsu Y, Szpirer C & Matsumoto K 1995 Assignments of the genes for rat pituitary adenylate cyclase activating polypeptide (Adcyap1) and its receptor subtypes (Adcyap1r1, Adcyap1r2, and Adcyap1r3). *Cytochemistry and Cell Genetics* 71:195–196.


Plakidou-Dymock S, Dymock D & Hooley R 1998 A higher plant seven-transmembrane receptor that influences sensitivity to cytokinins. Current Biology 8 315–324.


Received in final form 11 January 2005
Accepted 19 January 2005