PERSPECTIVE

The Evolutionarily Triumphant G-Protein-Coupled Receptor

DIANNE M. PEREZ

Department of Molecular Cardiology, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, Ohio

Received February 6, 2003; accepted February 27, 2003   This article is available online at http://molpharm.aspetjournals.org

Among all of the known proteins in the human genome, the G-protein-coupled receptors (GPCRs) may be the most versatile. They know how to manipulate the evolutionary forces that govern diversity by copying one ancestral gene and turning it into thousands of different proteins that can transduce light, smells, taste, nucleotides, lipids, peptides, and various other chemical compounds. They all seemingly do this by the same mechanism but can selectively and intimately interact with specific ligands and effectors. How did they do this? In this issue of *Molecular Pharmacology*, the tremendous effort of Fredriksson et al. (2003) provides the first overall road map of GPCR ancestry in a single mammalian genome using 342 functional nonolfactory human GPCR sequences. Their results show that there are five main families: Glutamate, Rhodopsin, Adhesion, Frizzled/taste2, and Secretin, forming the GRAFS classification system. From the chromosomal positions of the genes (paralogons) and the finding of "fingerprint" motifs, the authors support the theory that GPCRs in the GRAFS family share a common ancestor and evolved through gene duplication and exon shuffling. Their data also formulated a distinct family of GPCRs called the adhesion family and showed that the two taste receptor groups (TAS1 and TAS2) are not phylogenetically linked. The analysis divided the TAS1 receptors into the glutamate receptor family, whereas the TAS2 receptors grouped with Frizzled.

The GPCRs are so diverse and important to our physiology that they are the most pursued as drug targets. They include about 1000 to 2000 members and comprise >1% of the human genome. Even nematodes cannot live without them or may even require them more; more than a thousand GPCRs make up 5% of the nematode genome (Bargmann, 1998). With the cloning of a plant cytokine GPCR, GCR1 from *Arabidopsis thaliana*, the ancestry of GPCRs is even more ancient and may predate the divergence of plants and animals (Plakidou-Dymock et al., 1998). Although not yet characterized, there is also evidence to suggest that GPCRs exist in protozoa (New and Wong, 1998); certainly, the presence of the cAMP receptors in *Dictyostelium discoideum*, suggests the presence of cAMP receptors in humans (Bankir et al., 2002). Although not linked to G-proteins, bacteriorhodopsin itself suggests that the ancestor of the GPCR-like structure can even reside in bacteria. The phylogenetic analysis of these proteins may provide the tools to understanding the targets, drug design, and classification of the many unknown orphan receptors.

Previous Analyses and Classification Systems

Several classification systems have been proposed that divide the GPCRs based upon their ligands and/or amino acid sequences. In one of the most popular systems, GPCRs were previously classified into six families, termed the A through F classification system of Kolakowski (1994). Family A contained the large rhodopsin-related members, which included the biogenic amine receptors; family B consisted of the glucagon, parathyroid hormone, and calcitonin-related receptors; family C was the metabotropic glutamate receptors; family D was the STE2 yeast pheromone receptors; family E was the STE3 yeast pheromone receptors; and family F was related to the slime mold cAMP receptors. Since 1994, and the sequencing of the human genome, many other GPCRs were cloned or analyzed that did not fit into this classification system based upon homology and phylogenetic analysis. These included the Frizzled family (Wang et al., 1996), the plant GPCR *Arabidopsis thaliana* (Josefsson and Rask, 1997; Plakidou-Dymock et al., 1998), and the vomeronasal (mammalian pheromone) receptors (Dulac and Axel, 1995).

The Classification of Fredriksson et al. (2003)

The first family of the GRAFS classification of Fredriksson et al. (2003) is the glutamate receptors. This consists of the eight metabotrophic glutamate receptors, two GABA receptors and their splice variants, the calcium sensing receptor (CASR), and five of the taste (TAS1) receptors. CASR and

---

ABBREVIATIONS: GPCR, G-protein-coupled receptor; GRAFS, Glutamate, Rhodopsin, Adhesion, Frizzled/taste2, and Secretin; NPY, neuropeptide Y; EGF, epidermal growth factor.

TAS1 are grouped together with GABA, forming the basal branch of the family, suggesting that the glutamate receptors developed later in evolution.

The second family of the GRAFS classification system is the rhodopsin family, which contains by far the most members and is subdivided into four main groups: α, β, γ, and δ. The α-group contains five main branches: prostaglandin, amine, opsin, melatonin, and melanocortin, endothelial differentiation g-protein coupled, cannabinoid, and adenosine (MECA) receptor clusters. Two orphan receptors, GPR57 and GPR58, were identified and show homology to the trace amine receptors, whereas orphan GPR50 was homologous to the melatonins. The MECA group is rather interesting because its members bind very diverse ligands [i.e., peptides, lipids, chemicals (ethanolamides and nucleosides)]. Three orphan GPCRs (GPR3, GPR6, GPR12) show homology to cannabinoids.

The β-group [shown in blue in Fig. 3 of Fredriksson et al. (2003)] of rhodopsin has no main branches and contains many of the peptide hormone receptors, such as tachykinin (TACR), cholecystokinin (CCKs), endothelin (EDNR), vassopressin (AVPR), neuropeptide Y (NPYR), etc. One interesting result was that the NPY5R receptor grouped with the cholecystokinin receptors because it has a large third extracellular loop, unlike the rest of the NPYs. If the loop was removed, the analysis then grouped it with the NPYs. The two orphan receptors, GRP72 and GPR118, grouped near the NPY2.

The γ-group of rhodopsin receptors has three main branches: the somatostatins, opioids and galanin; the melanin-concentrating hormones (MCH); and the chemokines. Two orphan receptors, GPR7 and GPR8, which are homologous to opioid and somatostatin receptors, have recently been shown to bind (picomolar) to the newly discovered hormone, called neuropeptide W (Shimomura et al., 2002), which was not a careful choice of nomenclature (see below). Neuropeptide W exists in two forms, NPW-23 and NPW-30, based upon the number of amino acid residues contained in the peptides. NPW-23 sequence is contained in the N terminus of NPW-30. Although structurally unrelated to any known peptides, including the NPYs (so why call it NPW?), the NPWs, being cloned from the hypothalamus, mediate feeding and the neuroendocrine system. The MCH receptor branch is also involved in feeding behavior. The chemokine receptor branch is interesting because it contains the angiotensin and bradykinin receptors, which at first thought you think should have branched from the peptide hormone receptors or at least the somatostatin branch of the γ-group of rhodopsin.

The δ-group of rhodopsin has four main branches, which include the MAS related receptors, glycoprotein, purin, and the olfactory receptor clusters. The MAS oncogene and its related receptors (MRGs and MRGXs) have recently been found to be highly homologous with six novel genes, called sensory-neuron specific receptors (SNSR1–6; Lembo et al., 2002). However, the homology was quite high (98–99%) to the MRGX receptors and because Fredriksson et al. (2003) could not find this receptor sequence in the human database, it is currently not known whether it represents a novel receptor or (the more likely scenario) polymorphic variants of the MRGX receptors. The glycoprotein receptors include the classic follicle-stimulating hormone and thyrotropin-stimulating hormone receptors but also include the leucine-rich-repeat-containing receptors (LGRs), which contain three or-

phan receptors, LGR4–6. The purin receptor branch contains the formyl peptide receptors, the nucleotide receptors (P2Ys), thrombin, leukotriene, and a large number of orphan receptors. Olfactory receptors probably contain the largest gene superfamily within the GPCRs and may be even bigger than the GPCR family itself. The olfactory cluster represented 460 genes that were considered likely to be unique receptors; 347 were previously identified and cloned (Glusman et al., 2001; Zozulya et al., 2001; Takeda et al., 2002). About 60% of the chemokine genome is considered to be composed of pseudogenes. Fredriksson et al. (2003) do not show the phylogenetic analysis of olfaction because it needs further work, but they do show where it is branched in the δ-group of rhodopsin.

In the third GPCR family are the adhesion receptors, newly classified by Fredriksson et al. (2003). This family was at first compared with the secretin receptors (Ishihara et al., 1991), which secrete bicarbonate and potassium from the pancreas. The secretin receptors now form their own separate but related family. Subsequently, clones followed of a subfamily of GPCRs that exhibited lectin-like and adhesion modules in its N terminus (Hadjantonakis et al., 1997; Balch et al., 1998) and were eventually called the EGF-TM7. They are sometimes referred to as LN-7TM (for long N terminus). All of these members contain epidermal growth factor (EGF) modules and mucin-like domains. Recently, a new family member was identified, called the EGF-TM7-Latrophilin-related protein (ETL; Nechiporuk et al., 2001). Latrophilin and its GPCRs (CL1–3; C-type Lectin or LEC) have been shown to interact with α-latrotoxin, the venom of the black widow spider, to open calcium channels (Hlubek et al., 2000). EGF domains can bind calcium; together with latrophilin coupling to calcium channels, these GPCRs are suggested to be involved in rapid calcium influx. The large N terminus contains EGF modules, a Ser/Thr rich linker, and a Cys-rich proteolysis domain (Nechiporuk et al., 2001). EGF domains are also found in the extracellular portions of other non-GPCRs, such as fibrillin, tenascin, and thrombospondins, and are associated with protein-protein associations. Evolutionarily, this could possibly link the single transmembrane receptors with GPCRs. Another interesting aspect of adhesion receptors is a proteolysis domain near TM1. In both the ETL and CR1 receptors, the cleaved N terminus remains highly tethered to the transmembrane domain (Krasnoperov et al., 1999). The function of the proteolysis has engendered no speculation. On a cautionary note, something not pointed out by Fredriksson et al. (2003) is that the adhesion receptors do indeed have high homology to GPCRs, but no one has yet shown coupling to G-proteins, a prerequisite for GPCR classification. You will not currently find these receptors listed with known GPCRs in the nomenclature literature. These may turn out to be GPCR-like receptors that are mechanistically distinct or may be calcium channels. On the other hand, I am unaware of any highly homologous GPCR-like structure from a higher organism that has not been shown eventually to couple to G-proteins. Initially, the α1-adrenergic receptors were even thought not to couple to G-proteins, because it is difficult to demonstrate a GTP-shift in the binding curves. Because it has now been shown in Fredriksson et al. (2003) that adhesion receptors are evolutionarily linked to the G-protein-coupled secretin receptor family, the likelihood is high that these receptors also couple to G-proteins. The adhesion fam-

ily contains a large number of orphan receptors (i.e., GPR97, GPR110–116).

In the classification scheme of Fredriksson et al. (2003), the Frizzled/TAS2 receptors formed their own family divided into two clusters: one for the Frizzled and the other for the TAS2 receptors. This was a surprise for the authors, but the bootstrap value (774 of 1000), which is the statistical probability that the two branches are really tied or are likely to share a common ancestor, is rather high. At first glance, there is no apparent homology between the two groups, but in Fredriksson et al. (2003), there are consensus sequences or fingerprints that are not shared by the other families. The authors found 13 TAS2 receptors in the human database, two of which were not previously annotated as such.

The last classified family in the Fredriksson et al. (2003) system is the secretin family. Members of this family include secretin, calcitonin, glucagon, corticotropin-releasing hormone, parathyroid receptors, etc. There are no surprises here.

Evidence for the Ancestral Gene

Gene duplication is not a new thought. How else can new DNA or genes be formed? The field received its first notoriety with the landmark publication of Susumu Ohno (1970), which drew evidence from genome sizes from different organisms. His major tenet was that gene duplication allows the second gene to be free of natural selection pressures to accumulate mutations, which sometimes results in a new function. However, most of the duplications would be silenced and be present as pseudogenes. To be most effective, gene duplications would occur on a larger scale, perhaps through polyploidy or whole chromosomal duplications. This work is often contrasted with that of Ed Lewis (1978). This article described a set of linked homeotic genes with similar roles in the development of the bithorax complex of the fruit fly. He also proposed that homeotic genes had evolved through tandem duplication (i.e., the gene was copied and placed next to it in the genome). He rationalized that as the cluster expanded, it allowed for increased segmental specialization. The difference between the two theories of Ohno and Lewis is a single gene in tandem duplication versus whole chromosomal duplication. In effect, both are likely to occur.

In the work of Fredriksson et al. (2003), the use of paralogons is shown as evidence of gene duplication events among the GPCRs. Paralogons are regions on certain sets of chromosomes that contain sets of homologous genes. For example, a large paralogon is located between chromosomes 4 and 5 (Lundin, 1993) (Fig. 1). This paralogon contains GPCRs as well as channels, hormones, and enzymes. Although not drawn to exact chromosomal positions in this figure, certain gene loci down the entire length of chromosome 4 are repeated with homologous genes on chromosome 5. There is even a large expansion of the interleukins on chromosome 5 that could indicate first a gene duplication followed by a tandem repeat. In Fredriksson et al. (2003), chromosome 3 contains the following genes in a tandem array: chemokine receptors, angiotensin (AT1 receptor), and purinergic receptors (P2Y12, P2Y1) can be paired with the X chromosome, which contains the following genes grouped: *P2Y4*, *CXCR3*, *P2Y9*, *P2Y10*, and then the AT2 receptor. The simplest explanation for this "paralogous region" is through whole chro-

mosomal duplication. An alternate explanation (for those that adhere to the nongene duplication theory) is that these regions represent some "functionally important" association, which we have not yet ascertained, and that they have been assembled through chromosomal translocations and maintained through selection pressures. So these genes are somehow "linked", are assembled together on the chromosome through selection pressures, and do not really arise through the duplication of the chromosome. Because of these same selection pressures, other family members are reproduce in the same manner on another chromosome.

If gene duplication is real, this should then create gene clusters, which eventually would soon break off to form smaller clusters and scatter around the genome through chromosomal translocations. The more scattered, the further back in history the gene duplication took place. This seems to have occurred for the GPCRs, which do have many smaller clusters on many different chromosomes and often are located around the chromosomal ends (see Fig. 4 in Fredriksson et al. (2003). However, some well known gene clusters such as the β-globin, *Hox*, and *ParaHox* gene clusters have remained intact for million of years (Brooke et al., 1998; Holland, 1999), so there may be some evolutionary pressures we do not yet understand. Whatever the mechanism, the phylogenetic analysis and patterns of gene duplications seen on the chromosomes suggest that GPCRs are derived from an ancestral source. How long ago this occurred is still debated
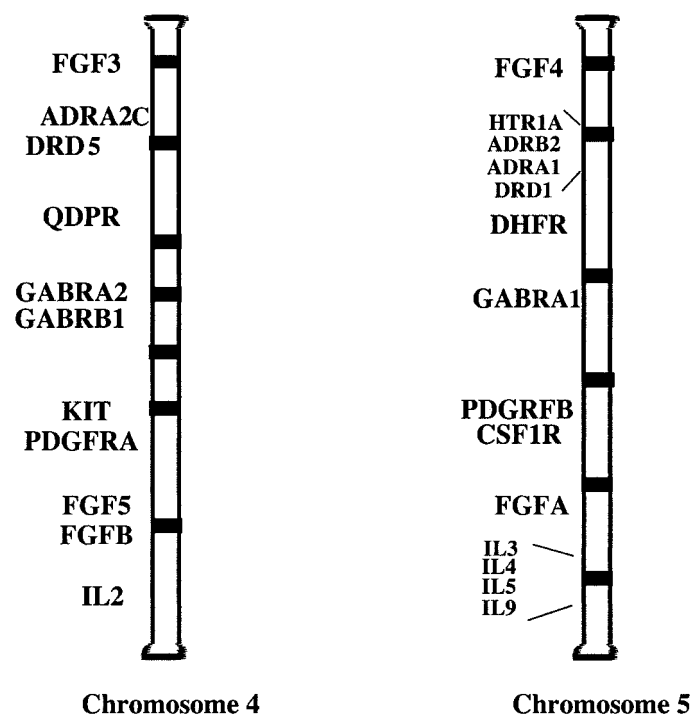

Fig. 1. A possible paralogon between human chromosomes 4 and 5. Data taken from Lundin (1993) and modified to current nomenclature. Various genes of receptors, enzymes, and hormones are duplicated and grouped in similar regions between the two chromosomes. Order and placement of genes is not accurate and is used for illustration only. FGFR, fibroblast growth factor; ADRA2, α_2-adrenergic receptor; DRD5, dopamine receptor; QDPR, quinoid dihydropteridine reductase; GABR, GABA receptors; Kit, feline sarcoma oncogene (PDGFR subfamily); PDGFR, platelet-derived growth factor receptor; IL, interleukins; HTR, serotonin receptor; ADRB2; β_2-adrenergic receptor; ADRA1; α_{1B}-adrenergic receptor; DHFR, dihydrofolate reductase; CSF, colony-stimulating factor (PDGFR subfamily).

and would partly depend on whether the plant and protozoa GPCR-like gene can be shown to couple to G-proteins.

What can pharmacology learn from phylogenetic studies? Although the slime mold cAMP receptors are not considered in these studies because they have not been cloned in humans, the many orphan receptors present in the rhodopsin family could possibly represent this type of receptor. The evidence suggests that these receptors do exist in mammals and may be related to rhodopsin (reviewed in Bankir et al., 2002). In the same context, people who are looking for orphan receptor ligands or a possible orphan receptor target may focus their searches based upon its phylogenetic analysis.

References

Balch SG, McKnight AJ, Seldin MF, and Gordon S (1998) Cloning of a novel C-type lectin expressed by murine macrophages. *J Biol Chem* **273:**18656–18664.
Bankir L, Ahloulay M, Devreotes PN, and Parent CA (2002) Extracellular cAMP inhibits proximal reabsorption: are plasma membrane cAMP receptors involved? *Am J Physiol* **282:**F376–F392.
Bargmann C (1998) Neurobiology of the *Caenorhabditis elegans* genome. *Science (Wash DC)* **282:**2028–2033.
Brooke NM, Garcia-Fernandez J, and Holland PW (1998) The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. *Nature (Lond)* **392:**920–922.
Dulac C and Axel R (1995) A novel family of genes encoding putative pheromone receptors in mammals. *Cell* **83:**195–206.
Fredriksson R, Lagerstrom MC, Lundin L-G, and Schioth HB (2003) The G-protein coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups and fingerprints. *Mol Pharmacol* **63:**1256–1272.
Glusman G, Yanai I, Rubin I, and Lancet D (2001) The complete human olfactory subgenome. *Genome Res* **11:**685–702.
Hadjantonakis AK, Sheward WJ, Harmar AJ, de Galan L, Hoovers JM, and Little PF (1997) Celsr1, a neural-specific gene encoding an unusual seven-pass transmembrane receptor, maps to mouse chromosome 15 and human chromosome 22qter. *Genomics* **45:**97–104.
Hlubek MD, Stuenkel EL, Krasnoperov VG, Petrenko AG, and Holz RW (2000) Calcium-independent receptor for α-latrotoxin and neurexin 1α [corrected] facilitate toxin-induced channel formation: evidence that channel formation results from tethering of toxin to membrane [published erratum appears in *Mol Pharmacol* **57:**832, 2000]. *Mol Pharmacol* **57:**519–528.
Holland PWH (1999) Gene duplication: Past, present and future. *Cell Dev Biol* **10:**541–547.
Ishihara T, Nakamura S, Kaziro Y, Takahashi T, Takahashi K, and Nagata S (1991)

Molecular cloning and expression of a cDNA encoding the secretin receptor. *EMBO (Eur Mol Biol Organ) J* **10:**1635–1641.
Josefsson LG and Rask L (1997) Cloning of a putative G-protein-coupled receptor from *Arabidopsis thaliana*. *Eur J Biochem* **249:**415–420.
Kolakowski LF Jr (1994) GCRDb: a G-protein-coupled receptor database. *Recept Channels* **2:**1–7.
Krasnoperov V, Bittner MA, Holz RW, Chepurny O, and Petrenko AG (1999) Structural requirements for α-latrotoxin binding and α-latrotoxin-stimulated secretion. A study with calcium-independent receptor of α-latrotoxin (CIRL) deletion mutants. *J Biol Chem* **274:**3590–3596.
Lembo PM, Grazzini E, Groblewski T, O'Donnell D, Roy MO, Zhang J, Hoffert C, Cao J, Schmidt R, Pelletier M, et al. (2002) Proenkephalin A gene products activate a new family of sensory neuron-specific GPCRs. *Nat Neurosci* **5:**201–209.
Lewis EB (1978) A gene complex controlling segmentation in Drosophila. *Nature (Lond)* **276:**565–570.
Lundin LG (1993) Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse. *Genomics* **16:**1–19.
Nechiporuk T, Urness LD, and Keating MT (2001) ETL, a novel seven-transmembrane receptor that is developmentally regulated in the heart. ETL is a member of the secretin family and belongs to the epidermal growth factor-seven-transmembrane subfamily. *J Biol Chem* **276:**4150–4157.
New DC and Wong JT (1998) The evidence for G-protein-coupled receptors and heterotrimeric G proteins in protozoa and ancestral metazoa. *Biol Signals Recept* **7:**98–108.
Ohno S (1970) *Evolution by Gene Duplication*. Springer-Verlag, Heidelberg.
Plakidou-Dymock S, Dymock D, and Hooley R (1998) A higher plant seven-transmembrane receptor that influences sensitivity to cytokinins. *Curr Biol* **8:**315–324.
Shimomura Y, Harada M, Goto M, Sugo T, Matsumoto Y, Abe M, Watanabe T, Asami T, Kitada C, Mori M, et al. (2002) Identification of neuropeptide W as the endogenous ligand for orphan G-protein-coupled receptors GPR7 and GPR8. *J Biol Chem* **277:**35826–35832.
Takeda S, Kadowaki S, Haga T, Takaesu H, and Mitaku S (2002) Identification of G protein-coupled receptor genes from the human genome sequence. *FEBS Lett* **520:**97–101.
Wang Y, Macke JP, Abella BS, Andreasson K, Worley P, Gilbert DJ, Copeland NG, Jenkins NA, and Nathans J (1996) A large family of putative transmembrane receptors homologous to the product of the *Drosophila* tissue polarity gene frizzled. *J Biol Chem* **271:**4468–4476.
Zozulya S, Echeverri F, and Nguyen T (2001) The human olfactory receptor repertoire. *Genome Biol* **2:**RESEARCH0018.

Address correspondence to: Dr. Dianne M. Perez, The Department of Molecular Cardiology, NB5, Lerner Research Institute, Cleveland Clinic Foundation, 9500 Euclid Ave., Cleveland, OH 44195. E-mail: perezd@ccf.org