A brief review on the evolution of GPCR: conservation and diversification

Zaichao Zhang1,2†, Jiayan Wu1†, Jun Yu1, Jingfa Xiao1,*

1CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China
2College of Life Science, Graduate University of Chinese Academy of Sciences, Beijing, China
Email: xiaojingfa@big.ac.cn

Received 2012

ABSTRACT

G-protein couple receptors (GPCR) possess diversified functions and they comprise a large protein superfamily in cellular signaling. Numerous identification methods for GPCR have been employed and versatile GPCR types are discussed. Although they share conserved transmembrane structural topology, alignment results of all GPCR show no significant sequence similarities. Each GPCR type distributes diversely in different evolutionary hierarchies of eukaryotes, but it has a distinctive boundary in the era of metazoan. The common ancestor of GPCR metabotropic glutamate receptor includes 7-transmembrane structure and venus flytrap module, which is probably evolved from a compound of bacteriorhodopsin and periplasmic binding protein. Many investigations focus on fine structure shaping and GPCR classification. Here, we briefly discuss evolutionary dynamic mechanism of GPCR from the perspective of classification, diversification and conservation.

Keywords: GPCR; Evolution; Classification; Diversification; Conservation

1. INTRODUCTION

G-protein couple receptors (GPCR) form the largest superfamily of transmembrane proteins in cell signaling mechanism. They vary dramatically in sequence alignment but share an identical structural topology [1]. The primary function of GPCR is signal transduction by sensing molecules from extracellular (e.g. hormones and neurotransmitters) and mediating intracellular signaling through coupling to specific G proteins [2]. They are also essential targets for nearly 50% of all currently used therapeutic drugs [3]. GPCR contain receptors for amines, peptides, amino acids, glycoproteins, prostanooids, phospholipids, fatty acids, nucleosides, nucleotides, Ca2+ ions as well as sensory receptors for different exogenous ligands as odorants, bitter and sweet tastants, pheromones, and photons of light and so forth [4]. Currently thousands of GPCR have been found in human genome, about 350 of them detect hormones, growth factors, and other endogenous ligands, but about 150 of them are still unknown [5].

Studies on GPCR evolution have been done in several eukaryotic species, which provide insights from different perspectives [5-12]. However, our understanding of GPCR evolution is merely based on extant genome sequences since most ancient eukaryotic species ever lived on earth are extinct. With an increasing number of GPCR sequences, they could be concluded into different categories by different classification systems. Because their sequences are dramatically multiform while a barrel structure is shared by all GPCR. Here, we share a specific evolutionary view of GPCR on their classification, diversification and conservation.

2. GPCR REPERTOIRES

2.1. GPCR prediction approaches

Although many GPCR prediction approaches have been proposed during past two decades, a great number of GPCR types are still vexed. The previous common methodology is sequence similarity searching in protein databases (e.g. NCBI, ExPASy, PIR, UniProt), which is mainly based on pairwise sequence alignment such as BLAST and BLAT [13, 14]. But it is difficult to identify GPCR successfully because there is no significant sequence similarities shared. To solve this problem, some statistical and machine learning approaches have been developed for GPCR prediction, such as HMM [15-17], statistical analysis method [18], covariant discriminant algorithm [19, 20], support vector machine method[21, 22], bagging classification tree [23] and SVM-DWT approach [24]. Online tools have been developed as well.
Table 1. Primary classification systems.

<table>
<thead>
<tr>
<th>Leading Author</th>
<th>Year</th>
<th>Methods</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kolakowski L.F. Jr.</td>
<td>1994</td>
<td>Integral component of the design of GCRDb</td>
<td>A-F classification system (A-C for multicellular animals)</td>
</tr>
<tr>
<td>Lars Josefsson</td>
<td>1999</td>
<td>PSI-BLAST searching method</td>
<td>One large clade and two smaller ones</td>
</tr>
<tr>
<td>Richard C. Grau</td>
<td>2001</td>
<td>BLAST and separate position-specific matrices</td>
<td>34 distinct clusters as GPCR groups</td>
</tr>
<tr>
<td>Rachel Karchin</td>
<td>2002</td>
<td>Support vector machines approach</td>
<td>A-E classes with various subclasses</td>
</tr>
<tr>
<td>Lapinsh.M</td>
<td>2002</td>
<td>Alignment-independent extraction of chemicals</td>
<td>Set up a data set of 929 rhodopsin-like receptor</td>
</tr>
<tr>
<td>Robert Fredriksson</td>
<td>2003</td>
<td>Alignment, bootstrapping and Fingerprint</td>
<td>GRAFS in human GPCR</td>
</tr>
<tr>
<td>Huang Ying</td>
<td>2004</td>
<td>Using a bagging classification tree algorithm</td>
<td>An accuracy of 91.1% for sub-family and 82.4% for sub-sub-family</td>
</tr>
<tr>
<td>Thora Bjarna-dottir</td>
<td>2006</td>
<td>BLAST, BLAT, and HMM searches</td>
<td>GPCR varies between the main GRAFS families</td>
</tr>
</tbody>
</table>

For instance, GPCRTree is an online hierarchical classifications webserver [25, 26]. In recent, a domain evaluation model for GPCR classification was also launched [27]. Each method has its own advantages and shortcomings, but HMM method is generalized from a mixture model and has been widely used, compared with other algorithms. The hidden variables that control the mixture component to be selected for each observation are related through a Markov process rather than independent of each other.

2.2. GPCR classifications

Based on different prediction approaches, several classification systems (Table 1) and GPCR databases (Table 2) have been established. The first GPCR database with A-F classification system has been constructed and adopted for almost a decade [1, 28]. With GPCR data accumulation, recently a novel GRAFS classification system [29] has been established and extensively used by latest studies [12, 30]. However, most classification systems still consist of three primary families and other mini-types that are still arguable [31-34]. These three primary families in all are classified mainly based on structure and functional similarity: rhodopsin-like receptor, secretin receptor and metabotropic glutamate receptor. In brief, rhodopsin-like receptor family accounts for 85% GPCR, which plays physiological roles of visual and smell sense, and these receptors distribute widely in mammalian genomes [30]. Rhodopsin-like receptor also represents a widespread protein family that includes hormones, neurotransmitters and light receptors; secretin receptor exists in many mammalians and a few are found in fungi. Receptors in this family mainly act for hormones and neuropeptides; metabotropic glutamate receptor performs a variety of functions in behavioral and mood regulations, as well as in the central and peripheral nervous systems [35]; other GPCR mini-types like fungal mating pheromone, frizzled/smoothened and orphan receptors combine a minority of GPCR, and are charged with their significantly specific duty individually. There are still some GPCR types are disputable. For example, cyclic AMP receptor (cAMP) is recognized as a second messenger and important in many biological processes. Some scientists define them as GPCR class E category [26, 28, 36] while others clarify cAMP as class F [1]. GPCRDB listed eight sequences of cAMP as a main category in Version 10.12.1 while it is no more identified as GPCR in Version 11.3.4. A recent research explains that cAMP receptor family is found in invertebrates and lost in vertebrates [12].

2.3. Signatures of GPCR

GPCR are evolutionary old and evidence shows that specific GPCR signatures can be found in all eukaryotic species [37]. They arrange themselves into a tertiary structure resembling a barrel in cellular membrane with two extracellular terminuses (Figure 1).

For instance, GPCRTree is an online hierarchical classifications webserver [25, 26]. In recent, a domain evaluation model for GPCR classification was also launched [27]. Each method has its own advantages and shortcomings, but HMM method is generalized from a mixture model and has been widely used, compared with other algorithms. The hidden variables that control the mixture component to be selected for each observation are related through a Markov process rather than independent of each other.

2.2. GPCR classifications

Based on different prediction approaches, several classification systems (Table 1) and GPCR databases (Table 2) have been established. The first GPCR database with A-F classification system has been constructed and adopted for almost a decade [1, 28]. With GPCR data accumulation, recently a novel GRAFS classification system [29] has been established and extensively used by latest studies [12, 30]. However, most classification systems still consist of three primary families and other mini-types that are still arguable [31-34]. These three primary families in all are classified mainly based on structure and functional similarity: rhodopsin-like receptor, secretin receptor and metabotropic glutamate receptor. In brief, rhodopsin-like receptor family accounts for 85% GPCR, which plays physiological roles of visual and smell sense, and these receptors distribute widely in mammalian genomes [30]. Rhodopsin-like receptor also represents a widespread protein family that includes hormones, neurotransmitters and light receptors; secretin receptor exists in many mammalians and a few are found in fungi. Receptors in this family mainly act for hormones and neuropeptides; metabotropic glutamate receptor performs a variety of functions in behavioral and mood regulations, as well as in the central and peripheral nervous systems [35]; other GPCR mini-types like fungal mating pheromone, frizzled/smoothened and orphan receptors combine a minority of GPCR, and are charged with their significantly specific duty individually. There are still some GPCR types are disputable. For example, cyclic AMP receptor (cAMP) is recognized as a second messenger and important in many biological processes. Some scientists define them as GPCR class E category [26, 28, 36] while others clarify cAMP as class F [1]. GPCRDB listed eight sequences of cAMP as a main category in Version 10.12.1 while it is no more identified as GPCR in Version 11.3.4. A recent research explains that cAMP receptor family is found in invertebrates and lost in vertebrates [12].

2.3. Signatures of GPCR

GPCR are evolutionary old and evidence shows that specific GPCR signatures can be found in all eukaryotic species [37]. They arrange themselves into a tertiary structure resembling a barrel in cellular membrane with two extracellular terminuses (Figure 1).

![Figure 1. A general structure of GPCR. (sample sequence is metabotropic glutamate receptor d1lx28_sacko metabotropic](image)
The 7-transmembrane is ancient with highly conserved structure as well as a length of 200-300 amino acids. The length and specific sites of both terminuses vary greatly, and N-terminuses of different GPCR contain numerous diversified motifs and domains [4]. Each GPCR family has its own features but it is still not obviously to tell their classifications by observing these features [38]. Most GPCR metabotropic glutamate receptors have longer N-terminus with specific motifs or domains where ligand-binding site is localized to receive signals from extracellular [39]. This is not the case for most rhodopsin-like receptors but some disputing ones like hormone receptor, which has long terminus. Rhodopsin-like receptors also share small molecule ligands, which may reduce the structural constraints for ligands binding and enhance the evolutionary survival of duplicated genes, especially after the appearance of metazoan around 500 million years ago [40].

All GPCR 7-transmembranes contain tm1~7 topology while families with long extracellular N-terminus consist of different motifs or domains including cystein-box, hormone binding domain, Arg-Gly-Asp motif, immunoglobulin mucin like stalk and so forth. Domains on N-terminus of GPCR families would mediate cell-to-cell adhesion or cell migration either by binding to components from extracellular environment or by interacting with membrane proteins from other cells [4]. It is also reported recently that at least 30 GPCR types with long N-terminus containing Ser/Thr-rich motifs found in human genome [41].

3. EVOLUTIONARY INSIGHTS

3.1. Current deductions for GPCR evolution

Many studies have provided insights on GPCR evolution focusing on a certain type of GPCR across different species or on populations within only one species [40]. In 2001, Graul and Sadee presumed that a refined GPCR ancestry evolution may facilitate database annotation for GPCR orphan receptors [42]. Simultaneously, Fredriks-son and Schioth claimed the repertoire of trace amine of GPCR would be one of most ancient GPCR [43]. The first structure signature of GPCR rhodopsin-like receptors in eukaryotic species was found in several protosome around 700 Mya [44]. As for secretin receptor, Cardoso and colleagues put forward a hypothesis that the putative ancestral receptors of this rhodopsin-like receptor is proposed to be more like the deuterostome CAL/CGRP/CRF receptors and evolved into other types ~500 Mya [7]. The ancestor of metabotropic glutamate receptor was proposed to be found in slime molds and sponges [35]. By means of mining GPCR evolutionary data from fossils, Torsten Schoneberg provided several clues that the phylogenetically oldest GPCR might include fungal pheromone receptors, cAMP and glutamate-receptor-like receptors [10-12]. A schematic presentation of GPCR evolution superfamily shows that adhesion and frizzled as well as large rhodopsin family are children of the cAMP. Besides, rhodopsin family is parent to sensory family, taste2 and vomeronasal type1 as well as the nematode chemoreceptor family [11].

3.2. GPCR distribution in eukaryotes

The subfamilies of GPCR consist of various types in different evolutionary period and they have evolved in distinct protein superfamilies since the appearance of metazoan. Protists, thought to be the most ancient eukaryotes, contain all GPCR metabotropic glutamate receptors [45] and part of rhodopsin-like receptors that have longer N-terminus. Longer N-terminus is likely to be more ancient because none of short N-terminus is found in protists. No one could tell exactly what events brought the period of metazoan, but evidence by fossil studies shows that protists and fungi appeared before the appearance of metazoan [46]. However, whole genome duplication event took place after that era, which motivated GPCR expeditiously evolved into more various types acting specific signaling roles. The more advanced a species is, the more diversified GPCR the species might have. This is because GPCR play essential role in

---

The text continues with a table listing GPCR databases and their specific features, followed by further discussion on the evolutionary insights of GPCRs across different species and their potential roles in various biological processes.
advanced species that need more complicated signal connections in versatile cells and tissues. Furthermore, evidences also substantiate that the subfamilies of rhodopsin-like receptor contain 35.5% introns and these in secretin receptor are highly conserved in their position while introns in metabotropic glutamate receptor seldom exist [8]. We make a conclusion that GPCR with longer N-terminus would be more likely the ancestor of GPCR. Besides, subfamilies of rhodopsin-like receptor explosively expanded after the occurrence of metazoan.

3.3. The origin of 7TM and VFTM

Metabotropic glutamate receptors are found in more ancient species than metazoan [12] and this family symbolizes the earliest eukaryotic origin because of longer N-terminus and fewer introns. Numerous evidences have demonstrated that bacteriorhodopsin, an ancient light energy related protein widely presenting in prokaryotes, shares crystal structure and conserved positions with GPCR 7-transmembrane topology albeit sequences alignment of GPCR 7-transmembrane and bacteriorhodopsin is quite low [47-51].

It has already been identified that 7-transmembrane has a similar structural topology with structures in prokaryote genomes such as light-sensitive, proteo-, bacterio- and halorhodopsins [52, 53]. Interestingly, we find that 75% bacteriorhodopsin sequences contain intact seven transmembrane topology and the phylogenetic tree of bacteriorhodopsin and 7-transmembrane shows metabotropic glutamate receptor much closer with bacteriorhodopsin. We infer that the origin of 7-transmembrane is possibly evolved from bacteriorhodopsin topology.

Periplasmic binding proteins (PBP), an important signaling receptor in bacteria, is identified highly resemble with a specific structure entitled venus flytrap module (VFTM) [54]. PBP consists of two large lobes close the bound ligands (possibly cys-riched domains), resembling a similar structure like VFTM [9, 55]. The ligand-binding domain in N-terminus of metabotropic glutamate receptor is homology to PBP in sequence alignment [56]. Functional divergence plays an essential role in characterizing the functions of VFTM, which are also been shaped in the evolution of metabotropic glutamate receptor [9]. The N-terminus of metabotropic glutamate receptor mostly perhaps evolved from ancient PBP and afterwards combined with bacteriorhodopsin via cystein-rich region to form the prototype of metabotropic glutamate receptor.

4. PRESPECTIVES

The recent advance in next generation sequencing and genome sequences analysis methods has greatly reshaped our understanding of GPCR. We aim to describe the repertoire, feature and distribution as well as prototype of GPCR protein superfamily. With the accumulation of eukaryotic genome data, a huge amount of work is being undertaken to annotate and clarify the relationship of GPCR for different species from advanced species to inferior organisms to obtain a comprehensive overview of the entire GPCR evolutionary process. We believe it is essential to understand the particular details affecting the rapidly evolution GPCR subclasses after the appearance of metazoan. Completely understanding GPCR evolution might not only help us predict some potentially important features of GPCR but also bring a horizon for unclassified GPCR in future.

5. ACKNOWLEDGEMENTS

We are deeply appreciated to grant (2012AA020409) from the National Programs for High Technology Research and Development (863 Program), the Ministry of Science and Technology of the People’s Republic of China for supporting this work.

REFERENCES


Z. Zhang et al. / Open Journal of Genetics 2 (2012) 11-17

doi:/10.1016/j.neuropharm.2011.04.015

doi:/10.1128/MCB.00742-10

doi:/10.1007/s11606-010-1482-y

doi:/10.1016/S0163-7258(03)00038-X

doi:/10.1016/j.tips.2007.01.001

doi:/10.1016/j.tips.2007.01.001


doi:/10.1016/j.mce.2010.07.012

doi:/10.1016/S0006-291X(03)00026-3

doi:10.1208/ps030212

doi:/10.1016/j.ympev.2004.12.003

doi:/10.1037/93/evq079

doi:/10.1073/pnas.0812460106

doi:/10.1046/j.1469-8137.2002.00397.x

doi:/10.1021/jm00097a002

doi:/10.1016/0014-5793(94)80274-2

doi:/10.1006/jmbi.1996.0328

doi:/10.1038/35014552

doi:/10.1016/0016-627X(93)80165-8

doi:/10.1016/S0959-440X(98)80128-0


doi:10.1208/ps010202


ligand-binding domain in metabotropic glutamate receptors is related to bacterial periplasmic binding proteins, *Neuron*. 11, 41-52. doi:10.1016/0896-6273(93)90269-W