AAPS PharmSci 1999; 1 (2) article 2 (http://www.aapspharmsci.org).

The Venus Flytrap of Periplasmic Binding Proteins: An Ancient Protein Module Present in Multiple Drug Receptors

Submitted: January 11, 1999; Accepted: May 13, 1999; Published: June 10, 1999

Christian B. Felder[1], Richard C. Graul[2], Alan Y. Lee[2], Hans-Peter Merkle[1] and Wolfgang Sadee[2]

[1]Department of Pharmacy, ETH Zurich, Winterthurerstr. 190, CH-8057 Zurich, Switzerland

[2]Departments of Biopharmaceutical Sciences and Pharmaceutical Chemistry, University of California San Francisco, CA 94143-0446

ABSTRACT Located between the inner and outer membranes of Gram-negative bacteria, periplasmic binding proteins (PBPs) scavenge or sense diverse nutrients in the environment by coupling to transporters or chemotaxis receptors in the inner membrane. Their three-dimensional structures have been deduced in atomic detail with the use of X-ray crystallography, both in the free and liganded state. PBPs consist of two large lobes that close around the bound ligand, resembling a Venus flytrap. This architecture is reiterated in transcriptional regulators, such as the *lac* repressors. In the process of evolution, genes encoding the PBPs have fused with genes for integral membrane proteins. Thus, diverse mammalian receptors contain extracellular ligand binding domains that are homologous to the PBPs; these include glutamate/glycine-gated ion channels such as the NMDA receptor, G protein-coupled receptors, including metabotropic glutamate, GABA-B, calcium sensing, and pheromone receptors, and atrial natriuretic peptide-guanylate cyclase receptors. Many of these receptors are promising drug targets. On the basis of homology to PBPs and a recently resolved crystal structure of the extracellular binding domain of a glutamate receptor ion channel, it is possible to construct three-dimensional models of their ligand binding domains. Together with the extensive information available on the mechanism of ligand binding to PBPs, such models can serve as a guide in drug discovery.

INTRODUCTION Proteins often consist of a modular structure, with each module accounting for one or more of the protein's biological functions [1]. Derived from ancient gene fragments, the genes encoding such modules have inserted themselves into otherwise unrelated genes as an ingenious means to amplify the functional potential of the host protein in the process of evolution. Such gene modules include *sarc* homology domains, Pleckstein homology domains, transducin repeats, phosphotyrosin binding domains, spectrin repeats, and multiple modules on the extracellular side of membrane proteins, such as cadherin, fibronectin, and LDL receptors [1]. These modules, found in numerous proteins relevant to signal transduction, determine ligand-protein or protein-protein interactions, and their origin can often be traced across the most distant phyla, including eukaryotes, archebacteria, and prokaryotes. Even though sequence divergence during evolution may be considerable, it is generally assumed that the protein 3D structure encoded by homologous genes is highly conserved. For example, over 70 mammalian genes are now known to substitute at least in part for dysfunctional mutant genes in yeast, indicating that over 800 million years of evolution, the respective protein 3D structures have been exquisitely conserved. Biologists take advantage of this by inferring structures of distantly related proteins or protein modules from an available crystal structure of one of the members of a gene family. As a result, computer generated homology models of a protein can be readily verified experimentally, for example by site-directed mutagenesis of residues predicted to be in critical locations of the model structure. This is relevant to drug design and development as one can begin to understand how a ligand might bind to the target protein, or one can employ docking strategies to select candidate drugs from large databases of chemical structures.

Correspondence to:
Wolfgang Sadee
Telephone:
Facsimile:
E-mail:

In this study we will discuss here a protein module that is derived from the periplasmic binding proteins (PBPs) of Gram-negative bacteria [2]. Comprising a large and rather diverse gene family, or families, the PBPs serve to scavenge or sense nutrients in the environment to mediate substrate transport or induce chemotaxis towards the nutrient source. As we shall see, the PBP gene module appears in multiple otherwise unrelated gene families throughout prokaryotes and eukaryotes. Specifically, common ancestry with PBPs has already been utilized to construct homology models for the ligand binding domains of two important drug targets, the metabotropic and ionotropic glutamate receptors [3,4] and of the bacterial regulatory protein *lac* repressor [5] which have been recently verified through crystal structures. Moreover, the presence of PBP modules has been proposed for metabotropic GABA receptors, atrial natriuretic peptide/guanylate cyclase receptors, and further transcription factors or repressors. Recognizing the extraordinary recurrence of these gene fragments throughout evolution in diverse gene families, we decided to study the distribution of PBP modules throughout the available protein databases to further our

understanding of how these modules might have spread through multiple genomes and which yet unrecognized protein families might have acquired PBP modules. Indeed, we will show that the distribution of PBP modules is broader than generally recognized.

Understanding these relationships will further help us in defining protein structures by homology to PBPs. This is because a number of PBPs have already been crystallized in the presence and absence of ligands, and their three-dimensional structure and structural changes upon ligand binding are known in detail. Recent crystal structures of PBP-like receptor binding domains have confirmed the conservation of 3D structure during evolution. From this we can learn how proteins bind diverse ligands either in a highly specific fashion or rather non-specifically, and how ligand binding might affect the protein's function. These questions are highly relevant to drug design for targeting receptors where the ligand pocket is a PBP module.

An outstanding feature of the PBP structure is two large polypeptide lobes connected via flexible tethers; ligand binding thus induces a large conformational change such that the two lobes fold together. This has led to the term Venus flytrap, or alternatively, pac-man which has also been used in the literature [2,6,7]. Here, we will describe our findings relative to Venus flytrap structures present in a surprising number of proteins and discuss possible implications for drug design and development.

THEORETICAL APPROACH AND COMPUTATIONAL PROCEDURES

The purpose of this review was to define the gene family (or families) encoding PBPs, and further, to identify gene families related in evolution to the PBPs. Thus, we used the protein sequence of PBPs to search for homologous proteins in the non-redundant protein databases, using BLAST (basic local alignment search tool) [8]. This tool identifies portions of the protein sequence with similarities above a preselected threshold (a probability score representing the sum of all matching fragment, P[S]; we will simply use P here). One commonly selects BLOSUM62 as the mutation matrix, a P value of 10^{-6} to 10^{-10} as a cutoff indicative of possible or probable homology, and the default filter of BLAST to suppress alignments of highly repetitive sequences. Further, we employed recent modifications of BLAST, such as Gapped BLAST and PSI-BLAST, which permit the insertion of gaps or define special mutation matrices in an iterative fashion, respectively, for the aligned sequence fragments [9]. These programs are available at http://www.ncbi.nlm.nih.gov/BLAST/ .

Any single BLAST run is limited by the selection of the starter sequence. Thus, a single run identifies all neighbors of the starter sequence, but it fails to define a complete gene family, or distantly related gene families. This is because even within a given gene family, sequence divergence can be considerable, and alignment scores will drop below the chosen threshold value and disappear in the noise. To circumvent this problem, we performed iterative BLAST runs, beginning with a single starter sequence that identifies all its neighbors in the first BLAST run. Subsequently, each sequence neighbor below the cutoff P value serves again for a BLAST search in the second round. This will identify new sequences with sufficient similarity to be included in the core cluster below the P value cutoff. Any newly identified sequences are then run again in subsequent rounds until the core cluster has converged and no more new sequences are found. This iterative BLAST program, termed INCA (iterative neighborhood cluster analysis) [10] is accessible at http://itsa.ucsf.edu/~gram/home/inca/ . The Java program INCA directs the iterative BLAST runs and provides a summary list which includes a core cluster of all sequences that were connected with each other by at least one alignment below a cutoff probability score (eg, $P = 10^{-6}$), and a list of sequences that align with lower scores (eg, $10^{-6} < P = 10^{-1}$). Moreover, the program records the best scoring sequence pairs that link more distant proteins. This permits one to establish possible evolutionary links among proteins that elude detection by a single BLAST run.

Subjecting large gene families to INCA can result in unwieldy lists that may not converge if the P cutoff value is too high. This can be circumvented by selecting a sufficiently low P cutoff (eg, $P = 10^{-10}$), or by a modified INCA version that performs iterative BLAST runs within a preselected P value range (eg, $10^{-30} = P = 10^{-6}$). In the latter case, not all sequences may be identified that belong to a core cluster. We have applied both of these variations to the analysis of the PBP families.

The INCA approach is limited by the BLAST algorithm in finding evolutionary links among very distant gene families, and by the as yet rather incomplete database of all extant protein sequences. Further, any evolutionary links suggested by the BLAST alignments need to be verified by multiple independent approaches.

Periplasmic Binding Proteins With a Venus Flytrap Structure

Located in the periplasm between the two outer lipid membranes of Gram-negative bacteria, the PBPs bind a spectrum of substrates and mediate their transport into the cells or initiate chemotaxis by activating flagellar motion [2]. Also, PBPs may serve as chaperones for the refolding of denatured proteins [11]. Substrates include a great variety of structures, including monosaccharides, oligosaccharides, amino acids, oligopeptides, oxyanions, cations, and vitamins. Accordingly, the PBP family consists of many proteins with rather diverse sequences. Indeed, BLAST searches reveal the presence of several such families, seemingly without detectable similarity in their primary structures, but with similar tertiary

structures. A comprehensive list of PBPs is accessible in the Entrez program (http://www.ncbi.nlm.nih.gov/Entrez/, enter "periplasmic binding proteins") [12]. Upon binding their cargo, the PBPs form a complex with membrane bound transporters or chemotaxis receptors. The transport complex usually consists of two integral membrane proteins and two subunit ATPases attached on the cytoplasmic surface, each encoded by one or two genes. Their structure resembles that of the multiple drug resistance transporters that secrete substrates from the cells [13]. In contrast, a protein complex involving PBPs commonly transports substrates into the cells. Whether this distinction is universal remains to be seen.

Known in exquisite detail, the structures of the PBPs serve as unique examples of how proteins bind and discriminate small substrates. More than a dozen PBPs have been crystallized and their structures have been solved at the atomic level (see http://www.ncbi.nlm.nih.gov/Entrez/structure.html, enter "periplasmic binding protein" for a list of movable displays of PBP structures). This has revealed a conserved structure common to many of the PBPs, consisting of two globular lobes connected by a hinge region, with a large cleft between the lobes [6, 14-19]. Shown in **Figure 1**, the structure of lysine-arginine-ornithine (LAO) PBP [6, 16] illustrates the two-lobe architecture of PBPs. The hinge region may contain two (as in LAO BP) or more connecting strands, a curious phenomenon that requires that each of the two modules encompass several separate portions of the complete protein chain. Most relevant for the medicinal chemist is the fact that structures are available in the presence and absence of ligands, and this is shown in **Figure 1** comparing the empty and ligand-bound LAO PBP. Therefore, the PBPs provide a unique opportunity to understand ligand binding selectivity, and ligand-induced changes of the binding protein.

Upon substrate binding, the two lobes twist and close, thereby entrapping the ligand, hence the name Venus flytrap [2,6,7]. Such large motion of protein subdomains represent a common theme in protein-substrate interactions, and they also occur with calmodulin, and certain enzymes, such as protein kinases [20,21]. Because the ligand binds at some distance from the hinge region, it is thought that substrate binding to PBPs does not induce this large lobe motion but rather stabilizes the closed form [6]. For leucine-isoleucine-valine (LIV) PBP, three distinct structures have been identified: an unliganded open form, a liganded open form, and a liganded closed form [19].

Given the structural similarities among the PBPs, how is it possible that they bind such diverse ligands, and moreover, display either high selectivity, as is the case for the phosphate and sulfate PBPs, or rather broad substrate selectivity, as seen with LIV PBP, LAO PBP, and peptide PBPs [2] ? The high selectivity for phosphate or sulfate appears to be mediated by a network of H-

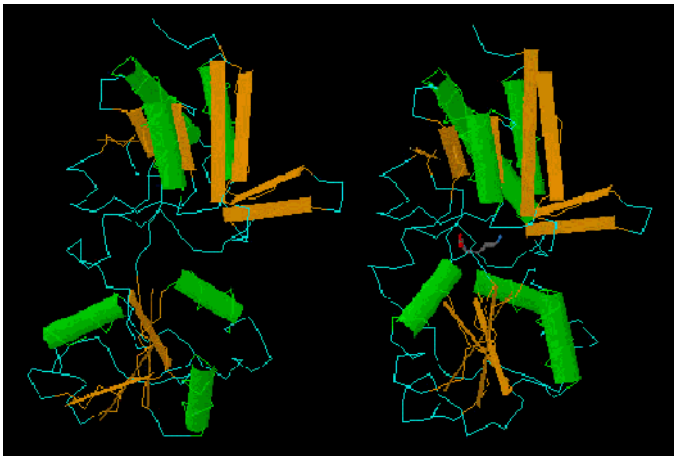


Figure 1. X-ray crystal structures of lysine-arginine-ornithine binding protein (LAO PBP) in the open (http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=2764&form=6&db=t&Dopt=s) and liganded forms (http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?uid=3481&form=6&db=t&Dopt=s)[6,16]. To view the structures, download the "Cn3D v2.0" software; select "Save File" in options, "Cn3D v2.0" in viewers, and "Cn3D Subset" in complexity, then click the "View/Save Structure" button. Alternatively, use "RasMol" software (also a browser plug-in) http://www.umass.edu/microbio/rasmol/getras.htm to view structures directly within your web browser; select "Launch Viewer" in options, "RasMol (PDB)" in viewers, and "Virtual Bond Model" in complexity.

bonds and ion bridges holding the completely dehydrated oxyanion in place [review: [2]]. By stripping any complexed water, PBP binding may facilitate the transit of polar substrates through the lipid bilayer into the cell. A network of hydrogen bonds also plays a role in binding selectivity with respect to carbohydrates and their epimers (eg, D-glucose and D-galactose) [2]. Further, the predominant hydrogen bonds yield only micromolar affinities, thereby, affording rapid dissociation needed for subsequent transport across the membrane. This also applies to the less selective PBPs for large amino acids, eg, LIVBP and LAOBP where the variable side chain is accommodated by a large water filled pocket in the PBP [2,6, 16,17,19]. Upon binding, the amino acid side chain displaces or dislocates water molecules, and thus, can be readily accommodated regardless of its structure. Similarly, oligopeptides are bound to OppA PBPs by extensive hydrogen bonding to the peptide backbone, whereas the side chains protrude into large caverns. By displacing water from these, peptide binding is favored entropically. As a result, these PBPs are selective for oligopeptides versus monomeric amino acids, but they bind oligopeptides non-selectively, independent of amino acid sequence [18].

PBP domains involved in interaction with the transport protein complex or chemotaxis receptor located in the membrane are also well studied [2]. These reside at the C- and N-termini of PBPs, lining the mouth of the two lobes. In some cases the same transporter complex can accept more than one PBP with different substrate specificity, eg, histidine and arginine [22], but more commonly distinct protein subunits recognized the various PBPs. Following

substrate binding, and twisting-closure, the ligand-bound closed form is thought to interact with the membrane complex [6, 20,22]. However, the exact mode of signal transduction or substrate translocation remains to be clarified. This is unfortunate because we shall see that the same protein modules recur in mammalian receptors that are important drug targets. Understanding the precise molecular interactions between the PBP module and the membrane complex that occur upon ligand binding would greatly enhance our understanding of drug action at these receptors. Usage of the PBP module has also caught the attention of biotechnology companies. For example, the maltose PBP serves in a pMAL protein Fusion & Purification System (New England Biolabs Inc.) for high yield protein expression, either in the cytosol or in the periplasmic space (to facilitate formation of disulfide bonds), including purification of the chimera on a maltose affinity column and subsequent cleavage at a Factor Xa cleavage site.

The PBPs are restricted to Gram-negative bacteria where they reside in the periplasmic space. However, a family of apparently homologous binding proteins also exists at the outer cell surface of Gram-positive bacteria [2, 23]. These are surrounded by a single membrane, and therefore, do not contain a periplasmic space. Equivalent to the PBP-dependent transport systems, Gram-positive bacteria possess extracytoplasmic binding lipoproteins, maintained at the membrane surface by embedding their N-terminal glyceride cysteine into the lipid bilayer [23]. For example, MalX is a maltose inducible membrane bound protein of *Streptococcus pneumoniae* that is homologous to the maltose PBP MalE [23]. Similarly, AmiA of *S. pneumoniae* is a homolog of the oligopeptide PBP OppA. Mutations of the AmiA locus confer increased resistance to aminopterin, implying that these proteins are involved in drug transport into the cells [23]. Binding lipoproteins of Gram-positive bacteria can be highly immunogenic, as is the case with a binding lipoprotein of *Mycobacterium tuberculosis* [2]. These results suggest that the PBP protein module has spread throughout diverse prokaryotic species.

Exploring the Protein Sequence Space of the PBPs and Related Families

Since the PBP protein family contains rather diverse sequences, we performed iterative BLAST analyses, using the INCA program [10], to determine which PBPs are related to each other. We selected as the starter sequence LAO PBP as a typical member of the family. To reduce the INCA data output to a reasonable size, we set the probability cutoff at $P = 10^{-10}$. Two sequences scoring with such a low probability are highly likely to be related to each other in evolution. After seven iterations, involving 192 individual BLAST runs, the INCA program converged, ie, no new sequence neighbors were found with $P = 10^{-10}$ after the seventh run. The resultant core cluster contains 329 neighbors, a selection of which is shown in **Table 1** . An additional list in **Table 1** contains sequences outside the core cluster scoring with $10^{-10} < P = P^{-2}$ with at least one core sequence.

Table 1. Neighborhood analysis of the periplasmic binding proteins, using INCA. LAO PBP served as the INCA starter sequence, and the cutoff value was $P \leq 10^{-10}$. INCA converged after 7 rounds of iterative BLAST runs, identifying 329 sequences in the core cluster ($P \leq 10^{-10}$), and 180 sequences with lower similarities ($10^{-10} < P \leq 10^{-2}$). The table contains only selected sequences to minimize redundancy

INCA code	gi code	Sequences producing significant alignments	Species	P value
1	01703391	lysine-arginine-ornithine-binding periplasmic protein precursor (LAOBP)	*Escherichia coli*	1.2e-170
6	02507364	histidine-binding periplasmic protein precursor (HBP)	*Escherichia coli*	1.9e-114
8	00123170	histidine-binding periplasmic protein precursor (HBP)	*Salmonella typhimurium*	3.2e-113
13	02507363	arginine-binding periplasmic protein 2 precursor	*Escherichia coli*	3.2e-35
18	00462726	nopaline-binding periplasmic protein precursor	*Agrobacterium tumefaciens*	2.0e-30
23	00464301	octopine-binding periplasmic protein precursor	*Agrobacterium tumefaciens*	2.2e-25
26	01706860	FLIY protein precursor (sulfate starvation-induced protein 7) (SSI7)	*Escherichia coli*	2.3e-23
26.10	02632624	(Z99105) similar to amino acid ABC transporter (binding protein)	*Bacillus subtilis*	9.7e-25
26.37	00548505	cyclohexadienyl dehydratase precursor (prephenate dehydratase / arogenate dehydratase)	*Pseudomonas aeruginosa*	3.5e-16
28	01731313	probable amino-acid ABC transporter BP in idh-deor intergenic region precursor	*Bacillus subtilis*	4.9e-17
30	00121394	glutamine-binding periplasmic protein precursor (GLNBP)	*Escherichia coli*	1.0e-15
30.4	01652664	(D90907) glutamine-binding periplasmic protein	*Synechocystis sp.*	5.5e-38
30.4.26	00729605	glutamate/aspartate transport system permease protein gltk	*Escherichia coli*	2.9e-28
30.4.37	01154896	(X82596) general amino acid ABC type transporter	*Rhizobium leguminosarum*	1.5e-24
30.4.54	01346150	glutamate transport system permease protein gluc	*Corynebacterium glutamicum*	2.8e-20
33	01731062	probable amino-acid ABC transporter BP in bmru-ansr intergenic region precursor	*Bacillus subtilis*	2.6e-13
33.37	00056286	(Z17238) glutamate receptor subtype delta-1	*Rattus norvegicus*	1.9e-11
33.37.103	00479400	glutamate receptor -	*Drosophila*	1.7e-38

		fruit fly	*melanogaster*	
33.37.106	00292285	(L13267) NMDA receptor subunit	*Homo sapiens*	1.1e-34
33.37.110	01122392	(X94156) NMDA glutamate receptor subunit	*Xenopus laevis*	1.3e-33
33.37.115	00228224	NMDA receptor 1	*Rattus rattus*	1.6e-33
33.37.116	00508971	(U11418) NMDAR1 glutamate receptor subunit	*Rattus norvegicus*	1.6e-33
33.37.142	00508292	(U11419) NMDAR2B glutamate receptor subunit	*Rattus norvegicus*	6.1e-27
33.37.148	00743476	D-MeAsp receptor:isotype=NR2A	*Homo sapiens*	3.6e-26
33.37.175	02708331	(AF038557) ligand gated channel-like protein	*Arabidopsis thaliana*	5.5e-17
33.38	00475542	(U08255) glutamate receptor delta-1 subunit	*Rattus norvegicus*	2.2e-11
33.38.10	00729597	glutamate receptor, ionotropic kainate 1 precursor (GLUR-5) (EAA3)	*Homo sapiens*	1.2e-85
33.38.12	01169965	glutamate receptor, ionotropic kainate 3 precursor (GLUR-7)	*Rattus norvegicus*	1.6e-85
33.38.19	00423447	glutamate receptor GluR6C	*Mus musculus*	2.8e-82
33.38.31	00987866	(X89507) AMPA receptor GluR4/D	*Gallus gallus*	1.4e-73
33.38.39	00478699	glutamate receptor GluR1	*Homo sapiens*	3.7e-73
33.38.46	00987864	(X89509) AMPA receptor GluR3/C	*Gallus gallus*	2.0e-72
33.38.61	00121432	glutamate receptor 2 precursor (GLUR-2) (GLUR-B) (GLUR-K2)	*Mus musculus*	8.6e-72
35	02108229	(U97348) basic surface protein	*Lactobacillus fermentum*	1.8e-11
36	01617328	(X99716) collagen binding protein	*Lactobacillus reuteri*	3.3e-11
37.6	01172055	major cell-binding factor precursor (CBF1) (PEB1)	*Campylobacter jejuni*	3.3e-23
displaying 180 neighbors outside cluster				
33.37.175.152	00543933	extracellular CASR precursor (parathyroid cell calcium-sensing receptor)	*Bos taurus*	1.7e-06
33.37.175.162	01709020	metabotropic glutamate receptor 5 precursor	*Homo sapiens*	2.3e-06
33.37.175.171	01168781	extracellular CASR precursor (parathyroid cell calcium-sensing receptor)	*Homo sapiens*	2.3e-06
33.37.175.179	02495076	metabotropic glutamate receptor 3 precursor	*Homo sapiens*	1.4e-05
33.37.175.185	01362612	guanylate cyclase (EC 4.6.1.2), receptor-type drgc - fruit fly	*Drosophila melanogaster*	3.5e-05
33.37.175.186	01079087	guanylyl cyclase receptor - fruit fly	*Drosophila melanogaster*	3.5e-05
33.37.175.188	00547903	metabotropic glutamate receptor 6 precursor	*Rattus norvegicus*	4.7e-05
33.37.175.190	00400253	metabotropic glutamate receptor 2 precursor	*Rattus norvegicus*	6.3e-05
33.37.175.203	01929419	(Y10369) GABA-BR1a	*Rattus norvegicus*	0.00025
33.37.175.204	01929421	(Y10370) GABA-BR1b	*Rattus norvegicus*	0.00026
33.37.175.205	00121300	net abotropic glutamate receptor 1 precursor	*Rattus norvegicus*	0.00027
33.37.175.216	02133829	natriuretic peptide receptor C NPR-C - Japanese eel	*Anguilla japonica*	0.0083
33.37.175.217	02367478	(AF011424) putative pheromone receptor	*Mus musculus*	0.0092
37.12.17	00728989	virulence sensor protein bvgs precursor	*Bordetella parapertussis*	3.0e-06

As expected, the core cluster contains PBPs for several substrates, including amino acids, nopaline, and octopine. Each individual sequence is sorted such that it is listed with the most similar sequence in the preceding INCA iterations. Thus, one can follow the trail of optimal P scores back to the first core cluster in the first INCA passage, by following the sequence numbers provided before the gi identification code. For most of the PBPs listed in the core cluster, P scores below 10^{-10} support a finding of probable homology. However, this approach failed to identify a number of PBPs, eg, those for taurine, phosphate, maltose, maltodextrin, spermidine, putrescine, sulfate, and sorbitol/mannitol, as clear homologs belonging to the LAO PBP core cluster. This is consistent with previous findings that the PBPs display rather dissimilar sequences. Yet, the INCA run shown in **Table 1** does provide links to the extracellular lipoproteins of Gram-positive bacteria, such as *Bacillus, Corynebacterium, Lactobacillus,* and *Mycobacterium* (eg, refs. [24-26]. A collagen binding protein from *Lactobacillus reuteri* may also be part of a transport system (**sequence 36**, **Table 1**)[27]. Similarly, the major cell binding factor PEB1 (**sequence 37.6**), part of a probable transport system, plays a role in cell adhesion and as antigen [28]. This clearly establishes evolutionary relationships among binding proteins as part of transport systems in rather distantly related prokaryotic phyla.

To investigate these relationships further and identify additional relatives of the PBPs, we again performed INCA with LAO PBP as the starter sequence, but this time with a cutoff value of $P = P^{-6}$. To limit the number of BLAST runs, we modified the INCA program so that individual BLAST runs were done only on sequences scoring between $10^{-30} = P = P^{-6}$. This converged after 9 iterations and yielded 749 sequences in the core cluster, plus 255 more distant neighbors. However, no additional distinct PBPs were discovered, implying that evolutionary relationships between all PBPs are distant at best. This finding is somewhat at odds with results from X-ray crystallography, showing similar 3D structures for diverse PBPs (see Entrez, periplasmic binding proteins such as LAO PBP, structure links, at http://www.ncbi.nlm.nih.gov/Entrez/).

Further BLAST analyses established the presence of several additional PBP families with dissimilar primary

structures. These include oligopeptide PBPs (also containing Ni^{++} PBPs), and ferrichrome, thiamine, spermidine/putrescine, Fe (III) chloride/dicitrate, myo-inositol, C4-dicarboxylate, and carbohydrate PBPs (maltose, maltodextrin, glycerol-3-phosphate, ribose, galactose, xylose).

INCA and single BLAST analyses also revealed a host of high scoring sequences belonging to different protein families. These include enzymes (cyclohexadienyl dehydratase, hyaluronate synthase, phosphorylase), cell binding and antigenic factors, transcriptional factors/repressors, ionotropic glutamate receptors, a number of G protein-coupled receptors (metabotropic glutamate, GABA-B, Ca^{++} sensor, and pheromone receptors), and natriuretic peptide/guanylate cyclase receptors. Several of these relationships have already been noted in the literature. The main purpose of this review is to survey these protein families, and discuss the relevance of these findings for drug discovery and development. We will focus first on the relationship between carbohydrate PBPs and transcriptional repressors, and then on homologies between the amino acid PBPs and mammalian receptor families.

The lac Repressors and Related Bacterial Regulatory Proteins

Whereas sequence similarities between PBPs and *lac* -type repressors are relatively weak, they appear to fold into similar structures. Currently, there are many more protein sequences available than 3D structures. Nevertheless, there may be only a limited number of protein modules that are conserved among diverse protein classes even if the primary sequence identity falls below 20%, into the twilight zone of recognizable homology. It might therefore be possible to deploy search algorithms that identify protein sequences folding into a known 3D structure. Bowie et al. [29] have approached this 'inverse protein folding problem' by matching 1D sequences into established 3D structures while considering replacements or movements in nearby side chains and accounting for insertions and deletions. Applying this to the known structure of the ribose binding protein from *E. coli*, they found that whereas the two highest scoring sequences are other PBPs, the next highest scoring proteins were all members of the *lac* repressor (LacR) family. This result strengthens a finding of common ancestry between PBPs and *lac* repressors. Moreover, using Gapped BLAST analysis, a relationship between carbohydrate PBPs and repressor proteins emerges with a score indicative of probable homology (**Table 2**).

Using primary sequence alignments between arabinose, glucose/galactose, and ribose PBPs, Nichols et al.[5] developed a model of the tertiary structure of the core domain of the *lac* repressor. This bacterial regulatory factor is a tetrameric protein of four identical subunits, each of 37.5 kD. The N-terminal 60 residues are involved in DNA binding while the core domain (residues 60-360) is responsible for sugar binding and subunit interaction [5], and references therein). The C-terminus (349-360) consists of leucine-rich heptad repeat units thought to stabilize the tetrameric structure. The *lac* repressor prevents effective transcription of the *lac* operon containing genes for lactose metabolism.

The homology model of Nichols et al. [5] ignores the C- and N-termini of the *lac* repressor because they *lac* k

Table 2. Single Gapped BLAST run with d-ribose-binding periplasmic protein precursor of *E. coli*. The table contains only selected sequences

gi code	SP code	Sequences producing significant alignments	Species	P value
00132121	RBSB_ECOLI	D-ribose-binding periplasmic protein precursor	*Escherichia coli*	1e-154
01172866	RBSB_HAEIN	D-ribose-binding periplasmic precursor	*Haemophilus influenzae*	1e-113
00548705	RBSB_BACSU	D-ribose-binding protein precursor	*Bacillus subtilis*	1e-48
01346563	MOCB_RHIME	putative rhizopine-binding protein precursor	*Rhizobium meliloti*	9e-28
00118474	DGAL_SALTY	D-galactose-binding periplasmic protein precursor (GBP)	*Salmonella typhimurium*	2e-14
00400962	RBSR_ECOLI	ribose operon repressor	*Escherichia coli*	2e-14
01175039	XYLF_HAEIN	D-xylose-binding periplasmic protein precursor	*Haemophilus influenzae*	9e-13
02612909		(AF015825) lacI repressor-like protein	*Bacillus subtilis*	2e-12
00586265	XYLF_ECOLI	D-xylose-binding periplasmic protein precursor	*Escherichia coli*	3e-12
00585043	DEGA_BACSU	degradation activator	*Bacillus subtilis*	3e-10
00131650	PURR_ECOLI	purine nucleotide synthesis repressor	*Escherichia coli*	7e-09
03023459	CCPA_STAXY	probable catabolite control protein A	*Staphylococcus xylosus*	7e-09
01705847	CHVE_AZOBR	multiple sugar-binding periplasmic receptor CHVE precursor	*Azospirillum brasilense*	1e-07
01173387	SCRR_PEDPE	sucrose operon regulatory protein	*Pediococcus pentosaceus*	1e-07
01168844	CCPA_BACME	glucose-resistance amylase regulator	*Bacillus megaterium*	2e-07
02506561	GNTR_ECOLI	gluconate utilization system GNT-I transcriptional repressor	*Escherichia coli*	2e-06
00125933	LACI_KLEPN	lactose operon repressor	*Klebsiella pneumoniae*	2e-05
00120534	FRUR_ECOLI	fructose repressor (catabolite repressor/activator)	*Escherichia coli*	5e-05

sequence similarity to PBPs. The crude overlay model was energy minimized to alleviate structural incompatibilities. While it is impossible to generate only a single acceptable structure, homology modeling continues to evolve and may eventually lead to unique solutions of the protein folding problem. The optimized *lac* repressor structure contained two lobes connected by two hinge strands, as in the ribose PBP. Results from previous genetic and mutational analysis of *lac* repressor function can be rationalized by this model with regard to folding, substrate binding, and subunit interactions. Whereas the model has 3D features compatible with its amino acid sequence, it accounts only for the core structure of *lac* repressor, and further, ambiguities remain. These center around regions surrounding residues 84 and 282 previously shown to be involved in subunit assembly [5]. Such subunit interactions are of particular interest here because similar subunit assembly of the binding modules in glutamate NMDA receptors may provide clues to their complete quaternary assembly structure.

Subsequently, the crystal structures of the *E. coli lac* repressor *lac* R [30] and the purine repressor PurR [31] have been solved, bound to their respective corepressors/inducers, isopropyl-b-D-thiogalactoside and hypoxanthine. These belong to the *lac* R family of transcriptional regulators of which there are more than 21 members with strong homology among each other. Whereas lactose, fructose, and raffinose repressors represent tetramers, other members appear to be dimers [31]. Confirming homology to the PBPs, in particular the ribose PBP, these structures are consistent with the homology 3D model described above [5]. The tetrameric *lac* R binds to two DNA operator sites and forms bidentate interactions with the *lac* promoter [30]. Its quaternary structure consists of two dyad-symmetric dimers with a deep V-shaped cleft between the two dimers. It appears that binding of the inducer may alter the relative orientation between the two dimers which affects binding to the two operators that are separated by variable DNA spacers. In this fashion, conformational changes induced by ligand binding to the subunits alters the affinity of the repressor to its operators and permits transcription to occur upon repressor release. Whether such a coordinated move would also apply to oligomeric ion channels with homosteric ligand binding domains remains to be seen.

The purine repressor PurR represents an example with a dimeric structure [31]. This study provides exquisite details of the binding interactions between PurR and hypoxanthine, and moreover, a complete topology of the dimerization interface that excludes 2242 A^2 of protein surface. The dimerization region between the two subunits is noncontiguous and largely localized to the C-terminal domain of the monomers. Whether the subunit interactions of PurR and *lac* R resembles those occurring in the oligomeric NMDA receptors with PBP-like binding domains remains to be seen; however, lessons learned from the repressor gene family may well serve as a guide for further study of the quaternary structure of the NMDA receptors.

Search for Gene Fusion Events Among the PBPs and Integral Membrane Protein Permeases in Prokaryotes

The main theme of this paper deals with the possibility that the PBPs have fused with a variety of integral membrane proteins to subserve new functions, for example in receptors. If this indeed had happened during evolution, one might anticipate that fusion could have also occurred between a PBP and its cognate permease membrane protein. This might be readily possible since such genes often reside in each others vicinity within an operon, providing us with a possible example of how such fusion might have occurred, and at what frequency. The INCA approach lends itself well to this search. Assume that a protein contains two modules, such as a PBP module fused with an integral membrane permease. In a single BLAST run, a starter PBP sequence would recognize such a fusion protein, and in subsequent iterative BLASTs, the bimodular fused protein would identify sequences related to both PBPs and membrane permeases. Indeed, shown in **Table 1**, LAO PBP identified a glutamine PBP in the first round with a length of 248 residues (**sequence 30**)[32], and another glutamine binding protein (BP) from *Synechocystis sp.* of more than twice the length with 530 residues (**sequence 30.4**)[33] in the second round. The latter in turn aligned not only with many PBPs, but also with a host of permeases containing 5 predicted TMDs each. Of these, only three examples are shown (**30.4.26**, **30.4.37**, and **30.4.54**), each representing solely the integral membrane portion of the respective permease. To illustrate the relationships between the two modules in the glutamine BP from *Synechocystis sp.*, we performed a Gapped BLAST analysis, and the results are displayed graphically in **Figure 2** . Whereas this glutamine BP contains two fused modules, each of the protein sequences identified by BLAST contains only either the PBP portion (lines on the left) or the permease portion (lines on the right). This indicates that in the genome of the cyanobacterium, *Synechocystis sp.*, which has been fully sequenced [33], such a fusion event has taken place to yield a predicted functional modular transport protein. Clearly, this appears to have been a rare event indeed, the only such example we were able to identify in prokaryotes.

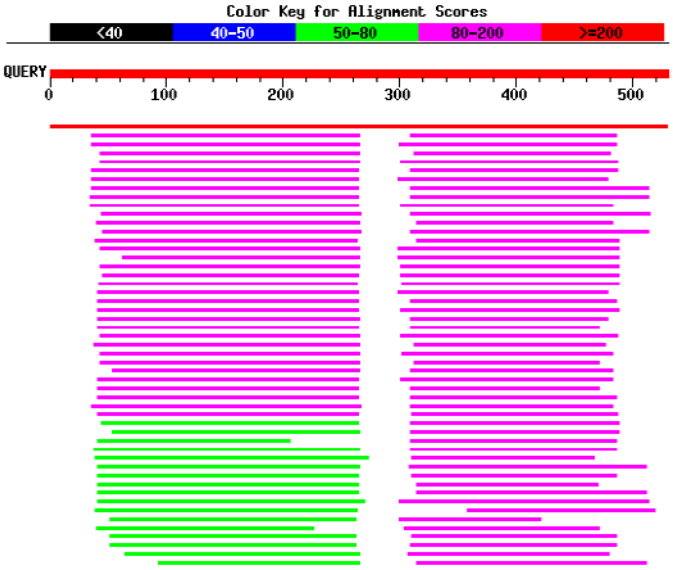


Figure 2. Single Gapped BLAST analysis with glutamine binding protein (BP) from *Synechocystis*. Sequence portions with substantial similarity are shown as horizontal bars, color coded to indicate the respective P-values (refer to Color key for alignment scores at the top of figure). Place the cursor over the colored bars to identify the represented sequences. To view interactive figure, go to: http://www.aapspharmaceutica.com/scientificjournals/pharmsci/journal/venus/fig2.html

It remains to be resolved whether such fusion provided an evolutionary advantage by tethering the PBP portion covalently to the permease, or whether in prokaryotes, the larger size of the gene might have impaired its function because of excessively accumulating mutations. Further, it is unclear whether such fusion events could have played a role in the evolution of the large family of fused ABC transporters in mammals, or secondary active transporters such as the dipeptide transporter, PepT1 [34], which contains a large extracellular loop between TMDs 9 and 10. However, we were unable to establish any relationships between mammalian transporters and the PBPs in our analyses. Nevertheless, this demonstrates that PBPs have fused to integral membrane proteins within prokaryotic genomes. One would assume that any fusion between PBP-like gene modules and integral membrane proteins might have been extremely rare in genomes lacking the operon organization of bacteria. Yet, a number of such fusion genes are apparent in mammalian genomes, indicated in **Table 1** . These include the neurotransmitter and hormone receptors discussed below. It is noteworthy that these fused receptors, even though they represent entirely distinct classes, all trace back in the INCA run to the same binding protein, a probable amino acid ABC transporter BP in the Gram-positive *Bacillus subtilis* (**sequence 33** in **Table 1**), with similarity to glutamine PBPs. Has fusion occurred only once and facilitated further evolution by crossovers between other integral membrane proteins, or did several independent fusion events occur with distinct membrane proteins, serendipitously with a similar PBP-like gene module? At this point, our knowledge of the evolution of integral membrane proteins is insufficient to address these questions; yet, their resolution might shed light on the structure of integral membrane proteins in general.

Ligand Binding Domains of Ionotropic Glutamate Receptors (eg, NMDA Receptors)

As one of the most intensely studied drug receptor targets, the 3D structure of glutamate-gated ion channels, including the NMDA, AMPA and kainate receptor subfamilies, is of major pharmaceutical interest. NMDA receptors play crucial roles in brain development, synaptic plasticity, memory, and excitatory neurotoxicity [4]. Similarly, the related AMPA and kainate receptor channels contribute to many of these physiological events (eg, [35]. The two main subunits of the NMDA receptors bind either glutamate (NR2) or the coactivator glycine (NR1). Whereas NR1 exists in several distinct splice variants, there are at least four NR2 subtypes encoded by different genes (NR2A-D). Unique among other gated channels, the NMDA receptor is a rather non-selective Ca^{++} channel which is blocked by extracellular Mg^{++} at resting membrane potential and opens only upon stimulation with both glutamate and glycine, and simultaneous membrane depolarization. Requiring two glycines and glutamates each for full activation, it is thought that the NMDA receptor consists of a hetero-tetramer. The glycine and glutamate subunits are homologous to each other, having significant sequence similarities.

Taking advantage of the moderate sequence similarity between NMDA subunits and PBPs, Laube et al. [4] have aligned portions of the NR2B and NR1 proteins with the lysine-ornithine-arginine PBP from *Salmonella typhimurium* . The NMDA subunits are proposed to contain large N- and C-terminal domains, three transmembrane domains (M1,3,4), and a reentrant loop (M2). Portions of the extracellular N-terminus and the extracellular loop between M3 and M4 can be aligned with LAO PBP. Thus, if a PBP-like protein module has been introduced into the NMDA subunits, it is split into two domains by insertion of M1-3. One must keep in mind that in PBPs, the two lobes of the complete protein are composed of several segments, with 2 or more hinge strands connecting them. Therefore, gene fragments representing different portions of the lobes might have been shuffled in the process of evolution. On the basis of these observations, a homology model was constructed for the glutamate binding domain from the known crystal structure of LAO PBP [4]. This model accurately predicted the binding pocket for glutamate which was further verified by site-directed mutagenesis. A similar homology model was constructed for the glycine binding subunit NR1. The results show that the guanidinium groups of Arg-493 in the NR2B subunit and of Arg-505 in the NR1 subunit interact ionically with the a-carboxyl groups of glutamate and glycine, respectively, in analogy to the homologous Arg-77 in LAO PBP [4]. Differences in

substrate selectivity between NR1 and NR2B are accounted for by the presence of more bulky aromatic residues in the amino group binding region of the NR1 pocket. Moreover, a crucial Lys-463 exists in NR2B close to the exterior of the binding pocket between the two lobes that interacts strongly with glutamate. This residue may play a role in transmitting the binding signal to the channel complex, and it also appears to serve as an anchor for the binding of antagonist ligands.

Recently, the first PBP-like domain from a ionotropic glutamate receptor was crystallized and its structure determined at atomic resolution. As in other ionotropic glutamate receptors, the binding pocket of the rat GluR2 receptor consists of two lobes formed by the extracellular segments S1 and S2 which are separated by an inserted gene fragment encoding membrane segments M1-3 [36] (**Figure 3**). To facilitate crystallization of the extracellular ligand domain, the amino terminal signal peptide domain and the C-terminus containing membrane segment 4 were truncated; moreover, the inserted membrane segments M1-3 were replaced by a linker sequence, thus, yielding a soluble protein which could be crystallized in the presence of the agonist kainate. This engineered protein consisted of two lobes, each represented by portions of segments S1 and S2 which cross-over between the lobes through two linker sequences, analogous to the hinge region in PBPs (**Figure 3**). The non-desensitizing ligand kainate, bound to the crystal structure, nestles in-between the two lobes, anchored by multiple, primarily polar residues from both lobes.

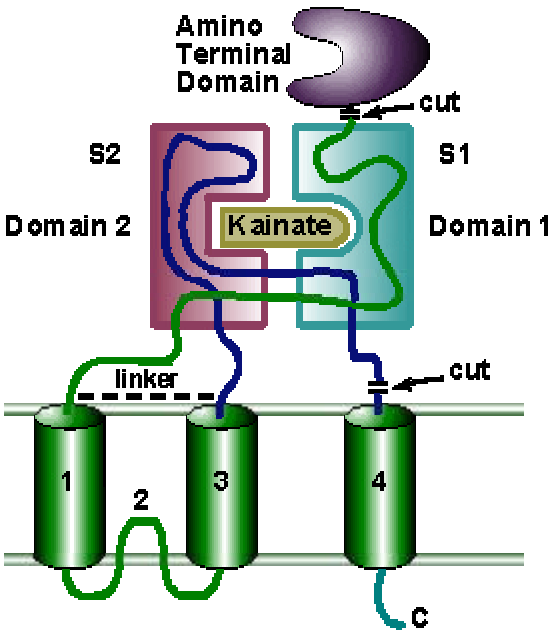


Figure 3. Schematics of the domain structure of the kainate iGluR2 receptor. Domains 1 and 2 each consist of portion of both extracellular domains S1 and S2. The cut sites and the linker position are shown, used to construct the crystallized soluble S1/S2 GluR2-kainate complex (redrawn from Armstrong et al., with permission from *Nature* 395: 914 (1998), MacMillan Magazines Ltd.).

Solving the structure of the kainate receptor binding module afforded the opportunity to compare its architecture to that of the solved structure of glutamine PBP, presumed to be homologous on the basis of weak but significant sequence similarity. The resultant structure overlay is shown in **Figure 4** [36]. When lobe 1 is superimposed on domain 1 of glutamine PBP, lobe 2 needs to rotate 21° to superimpose on lobe 2 of glutamine PBP, and an excellent overlay is attained. Thus, the kainate-occupied GluR2 S1/S2 complex closely resembles the structure of the PBPs, but it assumes a conformation intermediate between the open and ligand-bound closed form of glutamine PBP. This degree of closure may correspond to the open non-desensitized state of GluR2. The authors speculate that different ligands may induce different degrees of closure among the two lobes, and that certain ligands may permit full closure as suggested by fitting a different ligand into the simulated structure. These results can account for qualitatively different effects of distinct ligands at the ionotropic glutamate receptors. Moreover, the crystal structure permits analysis of the mechanisms by which the glutamate receptor can be regulated by redox control, involving an interdomain disulfide bond, and allosteric ligands binding to a site quite distant from the glutamate binding pocket. Even though the overall sequence identity among all ionotropic glutamate receptors is less than 10%, each of the seven residues making direct contact with the ligand are conserved in all structures, and therefore, the structure gleaned from the S1/S2 GluR2 complex is likely to be relevant to all members of this channel family.

Some glutamate channels appear to contain two distinct modules derived from PBPs. Thus, iGluRb1 contains N-terminally a module with similarity to LIV PBP, whereas a second module resembling glutamine PBP or LAO PBP is located in part on the N-terminus close to M1 and in the extracellular loop between M3 and M4 [3]. Therefore, the PBP-like domain appears to exist as a dimer of two lobes each. In evolution, the ionotropic glutamate receptors represent an ancient signaling mechanism that existed before the divergence of plants and animals. Thus, mammalians and plant iGluRs are homologous (see **Table 1**, **sequence 33.37.175** from *Arabidopsis thaliana*). The plant iGluRs display an identical membrane topology and appear to play a role in light-signal transduction [37]. The presence of such neurotransmitter receptor homologs in plants might account for the production of neurotoxic chemicals, not just for the protection against predators, but possibly also a regulatory mechanism for plant growth [37]. Some of these neurotoxins, such as b-N-methyl-amino-L-alanine in chick-peas, have been associated with neurodegenerative diseases in animals.

From these results, one can conjure a picture of the events leading to channel opening. Clearly, ligand binding would be expected to result in a closing of the two lobes of each subunit, a movement which directly

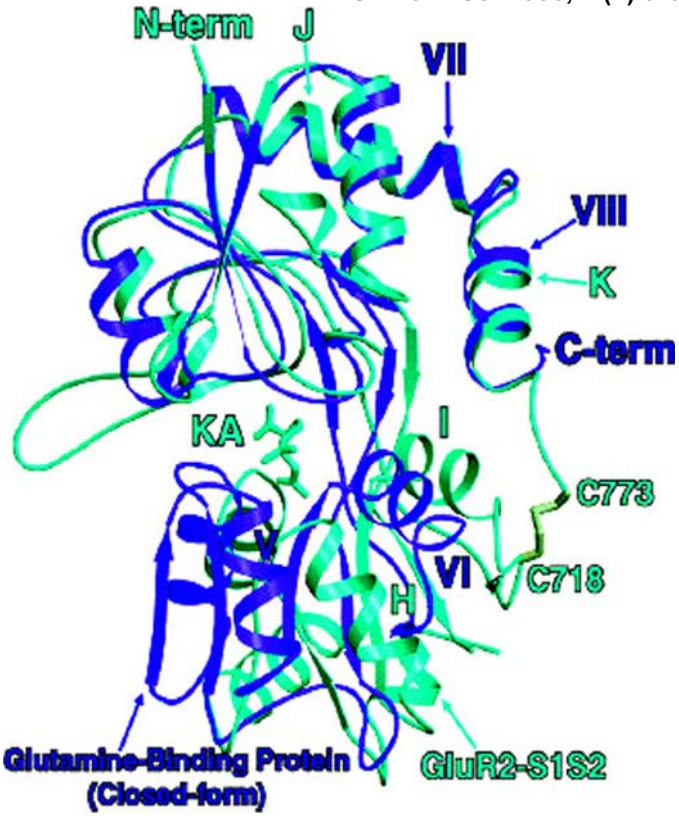


Figure 4. Superposition of the S1/S2 GluR2-kainate complex with the closed form of glutamine PBP (Armstrong et al., 1998). Domains 1 from both GluR2 and Q BP were superimposed first, followed by domains 2. The latter required a rotation of domain 2 such that the two lobes of the kainate-bound form of S1/S2 GluR2 were in a closure state intermediate to those of the open and closed forms of gluatmine PBP. [Reprinted with permission from *Nature* 395: 915 (1998), MacMillan Magazines Ltd.].

impacts upon the transmembrane structure, because of the bipartite composition of the glutamate binding pocket, with three transmembrane segments inserted between them. Further, the insertion of the three membrane segment M1-3 occurs at a position that is close to the first of two interlobe cross-overs in LAO PBP [3] and glutamine PBP [36]. However, the composite movement of the tetrameric channel complex remains to be elucidated. It might be fruitful to study the movements suggested for the tetrameric *lac* R as a guide for understanding channel activation. While there are some differences among the multitude of glutamate channels, it appears that most of them consist of similar structures, and hence might be studied with the same approach. These include the NMDA, AMPA, and kainate receptor families, each of which is recognized by homology searches using BLAST and INCA. Therefore, homology of the ligand binding pocket of these important ion channels with the PBPs can reveal much detail about their structure and function.

BINDING DOMAINS OF G PROTEIN-COUPLED RECEPTORS (GPCRS)

The metabotropic glutamate receptors (mGluRs). Because glutamate can also exert rather slowly evolving actions in neuronal tissues that are incompatible with rapid ion channel activity, a search for distinct classes of glutamate receptors was launched. In 1991, two groups published the sequence of the first member of a receptor gene family (mgr1, mGluR1) that acts as GPCRs [38,39]. Termed the metabotropic glutamate receptors to distinguish them from the ionotropic glutamate regulated ion channels, these proteins share little sequence homology with other GPCRs, but assume the same principal GPCR structure of seven TMDs. At least eight subtypes are known, each encoded by a different gene, and these represent promising drug targets.

Glutaminergic pathways play a role in psychiatric disorders, such as schizophrenia. Thus, NMDA receptor antagonists (phencyclidine) have psychotomimetic properties; however, targeting ionotropic glutamate receptors may not yield any useful drugs because glutamate regulated ion conductances are germane to fast synaptic transmission throughout the CNS. In contrast, metabotropic glutamate receptors modulate synaptic transmission, and a distinct distribution of mGluR subtype expression promises new approaches to therapy. This was recently born out by testing an agonist highly selective for group II mGluRs (mGluR2,3), compound LY354740 [40]. This agents effectively reversed phencyclidine effects in rats, an animal model of schizophrenia, thereby, suggesting a new approach to therapy of this major mental disorder.

The mGluR receptors each contain a large extracellular N-terminus preceding TMD1. Homology between the N-terminal domain of mGluRs, ionotropic glutamate channels, and the PBPs had been noted first by O'Hara et al. [3], but a relationship between metabotropic and ionotropic glutamate receptors had been disputed by Crockcroft et al. [41]. Yet, the ability to link both receptor classes to the PBPs enhances the confidence that an ancestral relationship does exist. In contrast to the glutamate ion channels, the PBP-like domain of the mGluRs is attached as a contiguous module at the N-terminus, and it appears to resemble most the LIV PBP [3]. Using multiple sequence alignments between mGluRs and PBPs, including LIV PBP, O'Hara et al. [3] constructed a homology model for mGluR1, preceding the homology model for the NMDA ligand binding site by six years. On the basis of this model, candidate residues for ligand binding were identified (Arg-201, Phe-412, Asp-474, Ser-165, and Ser-188). The functionality of the two Ser residues predicted to provide hydrogen bonding to the glutamate a-amino acid backbone was verified by site-directed mutagenesis.

These results suggest that the predicted structure of the mGluR binding pocket is compatible with the protein fold

Table 3. Neighborhood analysis (INCA) of the metabotropic glutamate receptors. mGluR1 served as the INCA starter sequence, and the cutoff value was $P \leq 10^{-6}$. The number of iterations was restricted to 3. This INCA format ran BLAST 164 times and identified 895 sequences inside the cluster ($P \leq 10^{-6}$) and 374 sequences with lower similarities ($10^{-6} < P \leq 10^{-2}$). The table contains only selected representative sequences.

INCA code	gi code	Sequences producing significant alignments	Species	P value
1	00121300	metabotropic glutamate receptor 1 precursor	*Rattus norvegicus*	0.0
20	02495079	metabotropic glutamate receptor 8 precursor	*Homo sapiens*	3.0e-171
28	02231438	metabotropic glutamate receptor 6 protein	*Homo sapiens*	1.4e-161
28.66	00311774	guanylate cyclase receptor	*Drosophila melanogaster*	7.4e-07
29	01834427	metabotropic glutamate receptor	*Drosophila melanogaster*	1.5e-141
29.59	01929421	(Y10370) GABA-BR1b	*Rattus norvegicus*	7.0e-14
29.60	01929419	(Y10369) GABA-BR1a	*Rattus norvegicus*	9.1e-14
37	01345672	extracellular CASR precursor (parathyroid cell calcium-sensing receptor)	*Rattus norvegicus*	2.2e-84
37.85	01486476	NMDA receptor	*Xenopus laevis*	1.3e-07
37.91	00472848	(U08107) N-methyl-D-aspartate receptor subunit	*Homo sapiens*	2.5e-07
40	02367613	(AF016182) putative pheromone receptor	*Rattus norvegicus*	1.0e-44
56	02367462	(AF011416) putative pheromone receptor	*Mus musculus*	4.2e-24
56.27	02589206	(AF022250) CASR-RS1	*Mus musculus*	1.5e-78
57	02589210	(AF022252) calcium-sensing receptor related protein 3	*Mus musculus*	6.6e-23
63	02388577	(AC000098) similar to Arabidopsis putative ion-channel	*Arabidopsis thaliana*	3.0e-07
63.5	02147747	N-methyl-D-aspartate receptor type 1	*unclassified vertebrate*	3.1e-19
63.9	00321325	N-methyl-D-aspartate receptor 1D precursor	*Rattus norvegicus*	3.6e-18
63.18	01346142	glutamate receptor 4 precursor (glutamate receptor ionotropic, AMPA 4)	*Homo sapiens*	1.4e-17
63.20	00202874	AMPA selective glutamate receptor	*Rattus norvegicus*	1.4e-17
63.26	02343289	NMDAR1 subunit isoform 4	*Homo sapiens*	3.1e-17
63.33	00743475	D-MeAsp receptor:ISOTYPE=NR1	*Homo sapiens*	2.5e-16
63.44	00940272	(L37836) glutamate receptor subunit 2B	*Tilapia nilotica*	3.4e-14
63.55	00422439	kainate-selective glutamate receptor	*Drosophila melanogaster*	4.6e-13
63.68	00475546	glutamate receptor KA1 subunit	*Rattus norvegicus*	5.7e-12
63.86	00056286	(Z17238) glutamate receptor subtype delta-1	*Rattus norvegicus*	1.4e-10
63.86.171	01731062	probable amino-acid ABC transporter binding protein precursor	*Bacillus subtilis*	3.1e-11
63.130	02133463	ionotropic glutamate receptor	*Caenorhabditis elegans*	2.7e-07
63.133	00465755	hypothetical 110.7 KD protein C06E1	*Caenorhabditis elegans*	2.9e-07
displaying 374 neighbors outside cluster				
37.85.39	02133830	natriuretic peptide receptor type D - Japanese eel	*Anguilla japonica*	1.4e-05
37.91.68	00162668	ANP C-receptor, bovine	*Bos taurus*	0.0042
37.91.69	01703313	atrial natriuretic peptide clearance receptor precursor (ANP-C)	*Bos taurus*	0.0042
63.44.189	00728750	amino-acid binding protein AABA precursor	*Dichelobacter nodosus*	0.0070
63.55.189	01346418	arginine-binding periplasmic protein precursor	*Pasteurella haemolytica*	0.00064
63.68.191	02314331	(AE000623) glutamine ABC transporter, periplasmic glutamine-BP	*Helicobacter pylori*	0.0012
63.86.205	02500690	histidine-binding protein precursor (HBP) (P29)	*Campylobacter jejuni*	0.072
63.86.214	01168523	arginine-binding periplasmic protein precursor	*Haemophilus influenzae*	0.46
63.130.208	02668595	arginine and ornithine binding protein	*Pseudomonas aeruginosa*	8.033
69.133.196	00462726	nopaline-binding periplasmic protein precursor	*Agrobacterium tumefaciens*	0.0013
63.133.206	01617328	collagen binding protein	*Lactobacillus reuteri*	0.017
63.207	00126343	leucine-specific binding protein precursor (LS-BP) (L-BP)	*Salmonella typhimurium*	0.22

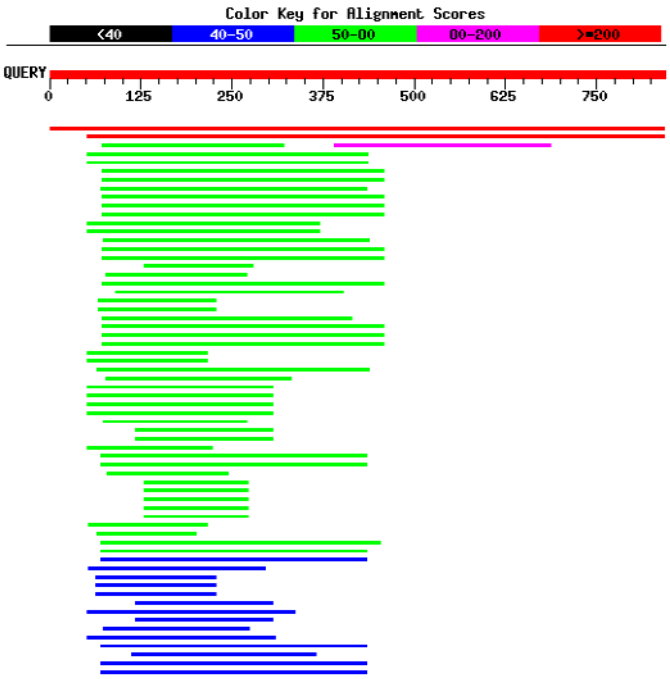


Figure 5. Single Gapped BLAST of GABA-BR1b. See Figure 2 for details. Place the cursor over the colored bars to identify the represented sequences. To view interactive figure, go to: http://www.aapspharmaceutica.com/scientificjournals/pharmsci/journal/venus/fig2.html

of the PBPs, and it further supports a finding of homology between them. To establish the sequence neighborhood of the mGluR's, we performed a limited INCA run with MGR1_RAT as the starter sequence (**Table 3**). This revealed relationships to other protein families, presented in descending order: GABA-B receptors, pheromone receptors, calcium sensing receptors (all GPCRs), NMDA, AMPA, and kainate receptors, guanylate cyclase and atrial natriuretic peptide (ANP) receptors. Lastly, links were also established to the PBPs at rather low scores, owing to the limited number of INCA iterations which failed to identify the highest scoring links (see **Table 1**). Relationships to the ANP receptors had also been noted by O'Hara et al. [3]. We will now briefly discuss each of these receptor families with regard to their relationship with the PBPs.

Metabotropic GABA receptors (GABA-B receptors).

Similar to the actions of glutamate, slow effects of GABA were judged incompatible with GABA action on Cl^- ion channels, but the cloning of metabotropic GABA-B receptors took much longer even though their existence was suspected as early as 1981. Only in 1997 did Kaupman et al. [42] clone the first GABA-selective GPCRs, $GABA_B$ R1a and b. Coupled negatively to adenylyl cyclase, GABA-B receptors modulate synaptic transmission either by presynaptic inhibition of transmitter release or by increasing a K^+ conductance

Table 4. Single Gapped BLAST run with the GABA-BR1 receptor. The table contains only selected representative sequences

gi code	Sequences producing significant alignments	Species	P value
01929421	(Y10370) GABA-BR1b	*Rattus norvegicus*	0.0
01929419	(Y10369) GABA-BR1a	*Rattus norvegicus*	0.0
02388577	(AC000098) similar to Arabidopsis putative ion-channel PID:g2262157	*Arabidopsis thaliana*	2e-13
00113918	atrial natriuretic peptide receptor B precursor	*Rattus norvegicus*	7e-12
00113915	atrial natriuretic peptide receptor A precursor	*Rattus norvegicus*	6e-09
00204270	(M74535) guanylate cyclase	*Rattus norvegicus*	1e-08
00547904	metabotropic glutamate receptor 7 precursor	*Rattus norvegicus*	3e-06
02495078	metabotropic glutamate receptor 7 precursor	*Homo sapiens*	3e-06
02495079	metabotropic glutamate receptor 8 precursor	*Homo sapiens*	3e-05
01168781	extracellular CASR precursor (parathyroid cell calcium-sensing receptor)	*Homo sapiens*	3e-05
00543933	extracellular CASR precursor (parathyroid cell calcium-sensing receptor)	*Bos taurus*	1e-04
00126347	leu/ile/val-binding protein precursor (LIV-BP)	*Citrobacter freundii*	2e-04
00118062	resact receptor precursor (guanylate cyclase)	*Arbacia punctulata*	9e-04
00126343	leucine-specific binding protein precursor (LS-BP) (L-BP)	*Salmonella typhimurium*	0.003
03287973	glutamate receptor 2 precursor	*Homo sapiens*	0.050

leading to long lasting postsynaptic potentials. Baclofen is a GABA-B receptor agonist used to treat spasticity resulting from multiple sclerosis and spinal injury, and other neurological disorders.

The primary amino acid sequence deduced from the cloned cDNA revealed distant relationships to the metabotropic glutamate receptors over their entire length, including both the equally long N-terminus and

the 7-TMD core [42]. Although the primary structures display low sequence identity, hydropathicity profiles attest to a highly similar structural architecture for both GABA-B and metabotropic glutamate receptors. Only a cysteine-rich domain close to TMD1 in the N-terminus of mGluRs is absent in the GABA-B receptors. Strikingly, the N-termini of GABA-B1a and b show homology to the PBPs, and therefore, are likely to fold similarly into two lobes that constitute the binding pocket.

Kaupman et al. [42] noted further sequence similarities to ANP guanylate cyclases (to be discussed below). A single BLAST run reveals that GABA-BR1a is most closely related to ANP receptors, followed by mGluRs, extracellular calcium sensing receptors, and more distantly, ionotropic glutamate receptors (only in the N-terminus) (**Table 4**). In contrast, no significant relationships appear to exist with the large family of ligand gated ion channels which includes the GABA-A, nicotinic acetylcholine, inhibitory glycine, and serotonin 5HT3 receptors [43,44]. **Figure 5** illustrates the limited regions of similarity among the GABA-B receptors and other receptor classes, as detected by a Gapped BLAST analysis. Mostly, the regions of sequence similarities are restricted to the large extracellular N-terminus, thought to be homologous to the PBPs (**Table 4**).

Calcium-sensing G protein-coupled receptors.

Extracellular Ca^{++} (Ca^{++} e) levels are tightly regulated and maintained at ~1 mM, at least 10,000-fold exceeding intracellular resting levels of free Ca^{++} ions. Therefore, the role of Ca^{++} e as a primary signal had long been neglected. Yet, the remarkable stability of Ca^{++} e results from a complex homeostatic system, consisting of Ca^{++} e sensing cells (eg, parathyroid cells) and Ca^{++} e utilizing or translocating tissues (kidneys, intestines, bone) [45]. A principal mechanism maintaining Ca^{++} e homeostasis became clear with the molecular cloning of a Ca^{++} e sensing receptor, CaSR [46,47]. Present in kidney and parathyroid cells, this receptor emerged as yet another variant of GPCR-type receptors, coupled to Gq/G11 and phosphatidylinositol turnover. It regulates multiple phases of Ca^{++} e metabolism, including renal transport and bone resorption, and therefore, may be of interest as a drug target.

Structurally, CaSR closely resembles the mGluR receptors, both in the N-terminus and in the membrane core structure [7, 45]. Clearly, the large N-terminus of both receptors appears to have a common ancestry in the PBPs, and therefore, is expected to fold into similar structures. Because of this homology, one can assume that the Ca^{++} binding domain is in a similar location as that found in the PBPs, including heavy metal ion PBPs. However, some PBPs, such as a galactose PBP, have acquired additional Ca^{++} binding sites that are probably distinct and reside in a different location of the PBP [48]. Because of the remarkable structural similarity between mGluRs and CaSR, Kubo et al. [49] asked whether the mGluRs would also be capable of sensing Ca^{++} e. Indeed, they found that the Gq/G11-coupled mGluRs 1a and 5 were activated upon Ca^{++} addition. These findings may be important to modulation of synaptic glutaminergic neurotransmission, for example in memory processing [50].

Valuable insight into the function of the PBP-fold of the CaSRs can be gathered from numerous polymorphisms of the human CaSR, many of which lead to altered signaling, and thus, inherited disorders of Ca^{++} metabolism [45]. Whereas inactivating mutations of *hCaSR* are associated with familial hypocalciuric hypercalcemia (FHH) [51], activating mutations lead to familial benign hypercalcemia and neonatal hyperparathyroidism [50,52]. A comprehensive listing of the known *hCaSR* polymorphisms is found in the Entrez browser (go to Entrez at http://www.ncbi.nlm.nih.gov/Entrez/protein.html ; enter the gi code 1168781, and go to GenPept Report). Most of the single nucleotide polymorphisms (SNPs) with functional consequences are located in the N-terminus responsible for Ca^{++} e sensing, but a few SNPs also occur in the membrane protein core [50,51].

When probing the sequence neighborhood of hCaSR the nearest neighbor is not the mGluR family, but rather a newly discovered family of pheromone receptors, the best alignments scoring with P= P^{180} ! Only distantly follow the mGluRs and then the NMDA receptors.

Mammalian pheromone receptors.

In many vertebrates, two basic modalities of smell detection exist: the sensing of odorants and perception of pheromones. Odorants are detected by a very large family (hundreds to thousands of genes) of odorant receptors located in the main olfactory epithelium, with neurons projecting via the main olfactory bulb to cortical and neocortical centers. This leads to cognitive and behavioral responses resultant from an awareness of the quality of the smell or fragrance. On the other hand, pheromone perception resides in chemosensory neurons of the vomeronasal organ (VNO) located in the ventral septum and projecting exclusively to specialized structures in the limbic system involved in reproduction and aggression [1, 53]. Thus, pheromones present in urine, sweat, and other bodily secretion trigger stereotyped behavior and neuroendocrine responses. Known mammalian pheromones include extremely diverse structures, such as steroids, prostaglandins, fatty acids, peptides, and proteins. This raises the question of how this spectrum of ligands is recognized by putative pheromone receptors suspected to exist in the vomeronasal organ.

In devising a strategy for cloning the responsible receptors, Dulac and Axel [54] took advantage of the unique distribution of the numerous odorant receptors which are thought to be individually and singly expressed each in a distinct neuron, assuming the same

Table 5. Single Gapped BLAST run with the pheromone receptor GoVN4. The table contains only selected representative sequences

gi code	Sequences producing significant alignments	Species	P value
02367611	(AF016181) putative pheromone receptor	*Rattus norvegicus*	0.0
02576409	(AF011414) putative pheromone receptor	*Mus musculus*	0.0
02367452	(AF011411) putative pheromone receptor	*Mus musculus*	0.0
02367605	(AF016178) putative pheromone receptor	*Rattus norvegicus*	1e-154
02996014	(AF053985) putative pheromone receptor V2R1	*Mus musculus*	1e-130
03130155	(AB008858) pheromone receptor	*Fugu rubripes*	1e-100
01345672	extracellular CASR precursor (parathyroid cell calcium-sensing receptor)	*Rattus norvegicus*	1e-97
01362762	calcium receptor (clone phPCaR-5.2)	*Homo sapiens*	1e-94
02996018	(AF053987) putative pheromone receptor V2R1	*Rattus norvegicus*	6e-88
02996016	(AF053986) putative pheromone receptor V2R2	*Mus musculus*	7e-84
02996028	(AF053992) putative pheromone receptor V2R4	*Rattus norvegicus*	1e-63
02996022	(AF053989) putative pheromone receptor V2R2	*Rattus norvegicus*	2e-59
03024134	metabotropic glutamate receptor 6 precursor	*Homo sapiens*	2e-46
00547904	metabotropic glutamate receptor 7 precursor	*Rattus norvegicus*	4e-46
01709020	metabotropic glutamate receptor 5 precursor	*Homo sapiens*	4e-36

would hold true for the pheromone receptors. Therefore, these investigators constructed cDNA libraries from single VNO neurons and searched for differentially expressed genes. This led to the characterization of a family of putative pheromone receptor genes, representing GPCRs associated with Gia2. However, the spatial distribution of these receptors in VNO neurons indicated that yet another gene family might exist, and indeed, the same approach revealed a novel family of ~100 GPCR genes, associated with Goa (Go-VN receptors) [53]. Its members show sequence similarities with the calcium sensing receptors and mGluRs (**Table 5**), in portions of the membrane core protein, and in a cysteine-rich domain in the N-terminus close to TMD1. However, Herrada and Dulac [53] pointed out that the N-terminus was quite diverse, possibly involving frame shifts, insertion, and deletions to give rise to variable lengths in different Go-VN receptors [53]. Even though no relationship to the PBPs was mentioned, the length of the N-terminus approximates that of CasR and mGluRs (~600 residues in full-length sequences such as Go-VN4,7,13C). Therefore, we asked whether the N-terminus is homologous to the PBPs, and thus, also might fold into the same protein mold.

A BLAST run with the N-terminus of Go-VN4 revealed strong similarities with extracellular calcium sensing receptors ($P \sim 10^{-97}$), and then in decreasing order of similarity, mGluRs, ionotropic glutamate receptors, and guanylate cyclase receptors. Moreover, the similarity between the pheromone and calcium sensing receptors extended over the entire sequence, including most of the large N-terminus (**Figure 6**). This strongly suggests that the similarity was not just restricted to a cysteine-rich domain in the N-terminus of mGluRs adjacent to TMD1 (possibly unrelated to PBPs), but involves the entire N-terminal PBP module present in other receptors.

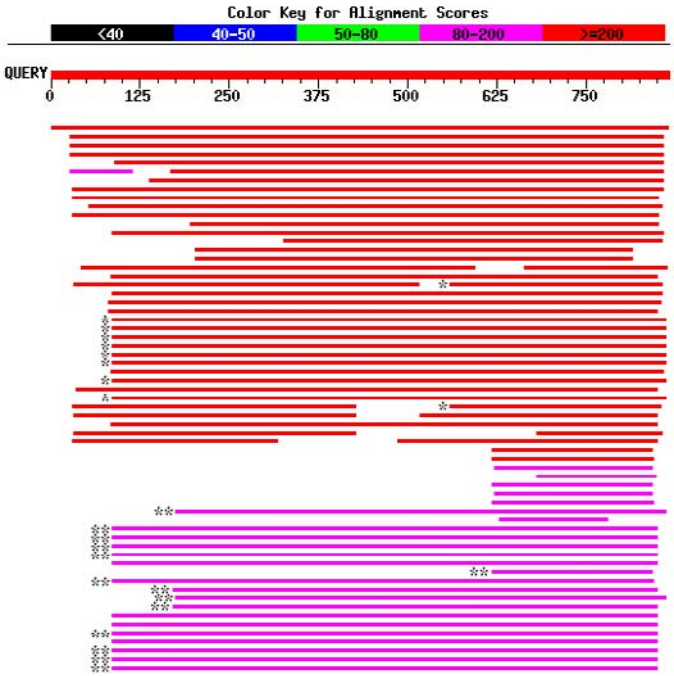


Figure 6. Gapped BLAST of pheromone receptor GoVN4. Indicated by * and ** are the calcium sensing receptor and metabotropic glutamate receptor sequences, respectively. Place the cursor over the colored bars to identify the represented sequences. To view interactive figure, go to: http://www.aapspharmaceutica.com/scientificjournals/pharmsci/journal/venus/fig2.html

Structurally, the Go-VN receptors are of particular interest to drug discovery because these receptors likely recognize different ligands, there are a large number of receptor structures available for analysis, and each ligand binding pocket may fold into the same tertiary structure. Therefore, much can be learned from a detailed phylogenetic-structural analysis about their mechanism of ligand recognition and signal transduction. Moreover, the pheromone receptors appear to exist also in humans, and hence, might play a crucial role in modulating behavior. However, it remains to be determined how pervasive pheromone effects are in humans. This is because human behavior might be more restricted by cognitive components than that of other mammals. Yet, exploiting the possible effect of GoVN receptors on the limbic system might become useful in the treatment of mental or behavioral disorders; on the other hand, the food and cosmetic industry might show an interest in affecting consumer behavior with the judicious choice of pheromones. Much more needs to be learned about this system to exploit therapeutic potentials and prevent misuse.

Dimerization of G protein-coupled receptors

Whereas glutamate ion channels are known to form hetero-oligomers, GPCRs are usually seen as monomeric membrane proteins. Recently, however, evidence has emerged in support of dimerization or oligomerization of GPCRs. Examples include metabotropic glutamate, Ca^{++} -sensing, and GABA-B receptors [55-59], each with a large PBP-like extracellular N-terminal ligand-binding domain. For metabotropic glutamate and Ca^{++} -sensing receptors, homo-dimers are stabilized via disulfide bonds between conserved cysteine residues in their N-terminal domains [55,56]. The functional significance of this homo-dimerization remains to be established, but agonist activation appears to favor the dimeric state. Even though the conserved Cys residues are located in the same position in glutamate and Ca^{++} -sensing receptors, no information is available on any possible heterodimerization between these two receptor families which could lead to interesting pharmacological properties. Further, it remains unknown how dimerization affects agonist activation and signal transduction. In the case of GABA-B receptors, on the other hand, dimerization serves a clear functional significance. Noting that GABA-BR1 receptors when expressed alone transmit agonist signals poorly into the cell, several research groups searched for associated membrane proteins that complex with GABA-BR1 to convey high affinity agonist binding and coupling. This led to the cloning from EST databases of a GABA-B receptor homolog, GABA-BR2, which by itself fails to bind GABA-B ligands, but enhances the agonist binding potency when coexpressed with GABA-BR1 [56-59]. Moreover, GABA-BR2 facilitates trafficking of GABA-BR1 to the cell surface, further contributing to enhanced signal transduction. In this case, no disulfide bonds were noted between the N-terminal domains; rather, a two-hybrid assay revealed that the intracellular C-tail is crucial for dimerization [58]. However, other receptor domains may also contribute. In view of the tendency of the PBP-like N-terminal domain to oligomerize, as seen in ionotropic glutamate receptors, in *lac* Z-repressors, and in certain PBPs, we speculate that dimerization of this domain could play a critical role in activation of these GPCRs. The crystal structure and proposed movements of the *lac* Z could serve as a template for understanding the role of receptor aggregation in these cases.

The Guanylate Cyclase-Atrial Natriuretic Peptide Receptors

The natriuretic peptides and their receptors comprise a complex system responsible for the regulation and control of blood pressure and body fluid homeostasis. Three main peptides and three receptors have been identified: atrial natriuretic peptide (ANP), brain natriuretic peptide (BNP) and C-type natriuretic peptide (CNP); and the atrial natriuretic peptide-A receptor (guanylyl cyclase A, ANP-A receptor), the atrial natriuretic peptide-B receptor (guanylyl cyclase B, ANP-B receptor), and the clearance receptor (ANP-C receptor). The natriuretic peptides form a ring structure of 17 amino acids via a disulfide linkage that is essential for their biological activity [60-62]. Upon binding to their cognate receptors, the ANPs activate the guanylate cyclase activity of the intracellular portion of the ANP-A and -B receptors, as the main mode of signal transduction [63-65]. On the other hand, the clearance receptor (ANP-C) *lac* ks the intracellular guanylate cyclase domain, but it is homologous to the extracellular and transmembrane domains of ANP-A and -B. With a molecular mass of 70 kD, it forms a homodimer and is thought to be involved in the clearance of natriuretic peptides from the circulation [66,67].

The ANP receptors possess a single transmembrane domain. The extracellular putative ligand-binding domains of the ANP-A and -B receptors are 43% identical among each other and approximately 30% identical to the amino acid sequence of the clearance receptor [63,64,66]. Close to the transmembrane domain lies a protein kinase-like domain that may negatively regulate the guanylyl cyclase activity [68], whereas the clearance receptor is down-regulated by natriuretic peptides through the cGMP-dependent pathway via the biologically active ANP receptors [69,70]. The three ANP receptors have different ligand selectivity; for example the ANP-A receptor prefers the ANP over BNP, followed by CNP receptors [71,72]. The ANP receptors are widely distributed in various tissues [63,66,71]. The ANP-A receptor for example is mainly expressed in kidney, lung, adrenal, heart, and adipose tissue.

Previous studies have already suggested that the extracellular domain of the ANP receptor is homologous to those of other receptors with PBP-like domains as the putative ligand binding sites. We have performed a

Table 6. Single Gapped BLAST run with the Atrial natriuretic peptide clearance receptor precursor. The table contains only selected representative sequences

gi code	Sequences producing significant alignments	Species	P value
00113920	atrial natriuretic peptide clearance receptor precursor	*Homo sapiens*	0.0
02133829	natriuretic peptide receptor C NPR-C - Japanese eel	*Anguilla japonica*	1e-151
00113912	atrial natriuretic peptide receptor A precursor (guanylate cyclase)	*Homo sapiens*	1e-54
00113916	atrial natriuretic peptide receptor B precursor (guanylate cyclase)	*Homo sapiens*	4e-49
03059111	(AJ005282) NPR-Bi	*Homo sapiens*	4e-49
00311774	(X72800) guanylate cyclase receptor	*Drosophila melanogaster*	3e-13
00477540	retinal particulate-guanylate cyclase - rat (fragments)	*Rattus norvegicus*	2e-10
03411272	(AF081464) olfactory enterotoxin receptor	*Bos taurus*	1e-07
01706243	retinal guanylyl cyclase 2 precursor (guanylate cyclase 2f, retinal) (retgc-2)	*Homo sapiens*	1e-06
01929419	(Y10369) GABA-BR1a	*Rattus norvegicus*	2e-05
01929421	(Y10370) GABA-BR1b	*Rattus norvegicus*	2e-05
01346321	heat-stable enterotoxin receptor precursor (gc-c)	*Rattus norvegicus*	4e-05
01170947	metabotropic glutamate receptor 5 precursor	*Rattus norvegicus*	1e-04
00548379	glutamate (nmda) receptor subunit zeta 1 precursor (nr1) (nmd-r1)	*Rattus norvegicus*	0.002

single Gapped BLAST run with the ANP-C receptor, to avoid the very large number of hits attributable to the intracellular guanylate cyclase domain (which is absent in ANP-C). A selection of related sequences is shown in **Table 6**. As expected, the GABA-B receptors are most closely related to the ANP receptors, followed by the metabotropic glutamate receptors. Viewing the location of the Gapped BLAST alignments confirms that sequence similarities occur throughout the extracellular domains of these receptors, similar to those shown in **Figure 5** for the GABA-B receptors. P values below 10^{-10} indicate that these similarities are unlikely to have arisen by chance. We conclude that the extracellular portion of the ANP receptors has a common ancestry in the PBP module, and therefore, it should also be possible to construct a 3D homology model for these receptors.

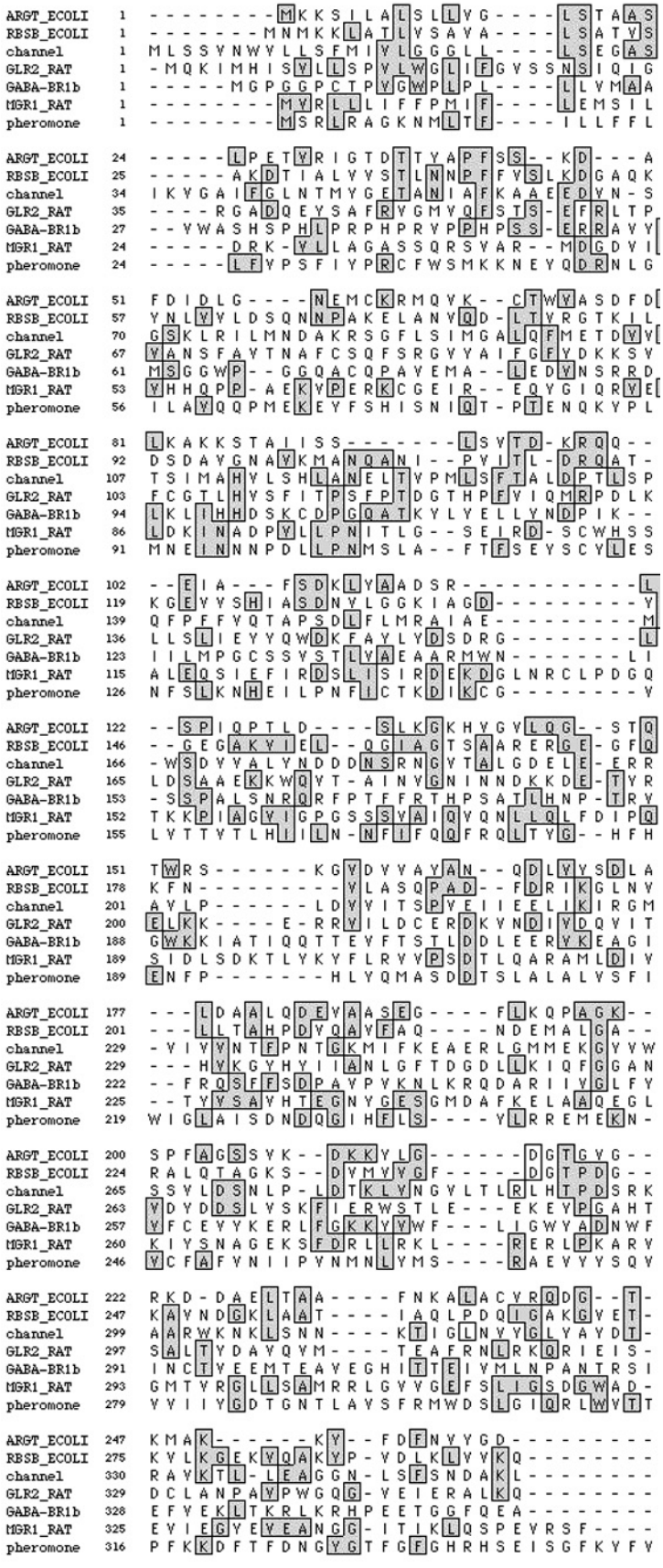
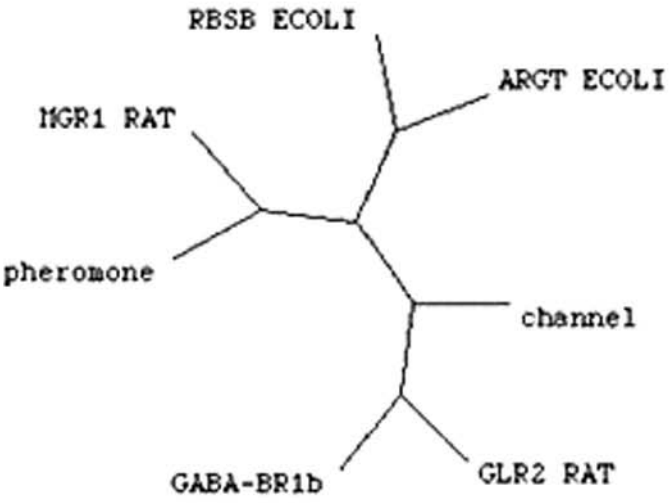


Figure 7. Multiple alignments (ClustalW)[72] of two PBPs and five PBP-like modules of different receptors. SeqVu is used for the display[73]. The boxed letters indicate 28% or more sequence identity among all seven sequences.



List of two PBPs and five PBP-like modules of different receptors. Sequences were identified using the BLAST client INCA

name	gi code	description
ARGT_ECOLI	01703391	lysine-arginine-ornithine-binding periplasmic protein precursor (LAO-binding protein) (*Escherichia coli*)
RBSB_ECOLI	00132121	D-ribose-binding periplasmic protein precursor (*Escherichia coli*)
channel	02708331	ligand gated channel-like protein (*Arabidopsis thaliana*)
GLR2_RAT	03287964	glutamate receptor 2 precursor (glur-2) (glur-b) (glur-k2) (glutamate receptor ionotropic, ampa 2) (*Rattus norvegicus*)
GABA-BR1b	01929421	GABA-B receptor 1b (*Rattus norvegicus*)
MGR1_RAT	00121300	metabotropic glutamate receptor 1 precursor (*Rattus norvegicus*)
pheromone	02367611	putative pheromone receptor (*Rattus norvegicus*)

Figure 8. Evolutionary tree of several selected PBPs and PBP-like modules, using the PROTPARS program of the PHYLIP package[74] and TreeView [75].

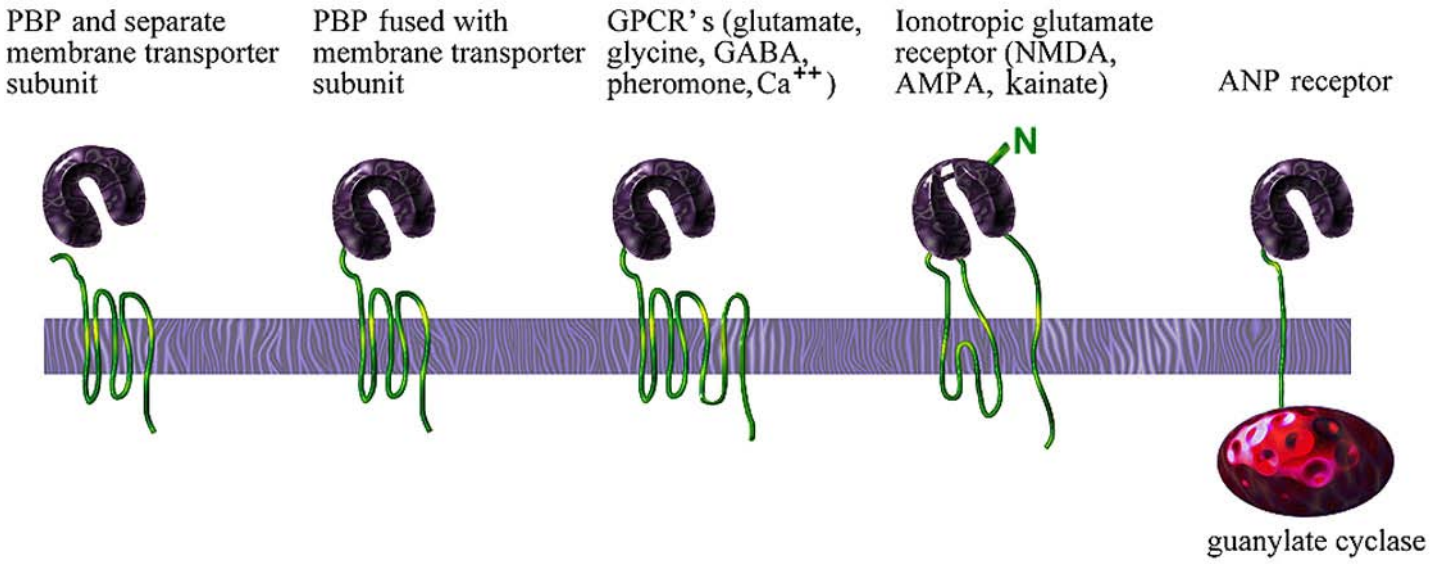


Figure 9. Schematics of modular membrane proteins containing a PBP-like domain.

As the atrial natriuretic peptides play a role in regulation of blood pressure and body fluid homeostasis, their receptors are potential targets for drug design aimed at the treatment of hypertension or other cardiovascular disorders. However, the physiology of the ANP system needs to be studied further, and as yet, no drugs are clinically used that interact with ANP receptors. It should be noted that the ANP receptor family may contain additional members recognizing distinct substrates. Thus, the heat stable enterotoxin receptor listed in **Table 6** similarly consists of a large N-terminal extracellular domain, a single TMD, and an intracellular kinase/guanylyl cyclase domain [73]. The P value of 4×10^{-5} suggests the possibility that the extracellular domain is homologous to the PBPs. The putative ligand for this membrane protein remains to be established.

MULTIPLE ALIGNMENTS AND EVOLUTIONARY TREE OF PBPS AND RELATED PROTEIN MODULES

Our analysis supports and extends the hypothesis that the PBPs have provided a template for the structures of binding modules in numerous receptor families. Even though sequence diversity is great, multiple analyses indicate that these proteins are homologs, and moreover, fold into a highly conserved bilobar structure. Notwithstanding the high sequence diversity, we have attempted to perform multiple alignments with representative members of the protein families discuss here. Shown in **Figure 7**, this can be achieved successfully with a few selected examples, providing an overview of the most highly conserved domains in these structures. Even though the overall sequence identity is small, a reasonable alignment occurs between these very distant structures.

From the INCA results with LAO BP as the starter sequence, one can deduce that many of the receptor binding modules are most closely related to a single precursor (**sequence 33**, **Table 1**), even the structurally distinct ANP receptor. This raises the question whether the metabotropic and ionotropic glutamate receptors might have originated from the same precursor which had arisen from fusion of a PBP with an integral membrane protein. An analysis of possible evolutionary relationships is shown in **8** . This tree exemplifies the common origin of the mammalian receptor modules from a single source cluster of PBPs. Understanding the evolutionary pathways underlying the current diversity in receptor structures with PBP-like modules would provide insight into their structure and function.

CONCLUSION

In conclusion, the periplasmic binding proteins of Gram-negative bacteria provide the blueprint for the receptor binding pocket of numerous and diverse receptors. An overview of the structures discussed here is provided in the schematics shown in **Figure 9** . Availability of several X-ray structures, and the astounding number of cloned sequences provide a unique opportunity to understand these receptors at the molecular level, unravel the dynamic changes occurring upon ligand binding, and predict their tertiary and quaternary structure with a higher degree of confidence than possible for other protein modules. This should pave the way for designing ligands selective for any of the multiple subtypes in any of these receptor families. The PBP-like structures might well become an example where computerized docking to homology models could lead to the *a priori* discovery of novel ligands before laboratory experiments begin to optimize the drug candidates.

ACKNOWLEDGEMENTS

This study was in part supported by research grants from the National Institute of Health, GM37188 and GM43102.

REFERENCES

1. Bork P, Downing AK, Kieffer B, Campbell ID. Structure and distribution of modules in extracellular proteins. Q. Rev. Biophys. 1996;29:119-167.

2. Quiocho FA, Ledvina PS. Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. Mol. Microbiol. 1996;20:17- 25.

3. O'Hara PJ, Sheppard PO, Thogersen H, et al.The ligand-binding domain in metabotropic glutamate receptors is related to bacterial periplasmic binding proteins. Neuron 1993;11:41-52. [PUBMED]

4. Laube B, Hirai H, Sturgess M, Betz H, Kuhse J. Molecular determinants of agonist discrimination by NMDA receptor subunits: analysis of the glutamate binding site on the NR2B subunit. Neuron 1997;18:493-503.

5. Nichols JC, Vyas NK, Quiocho FA, Matthews KS. Model of lactose repressor core based on alignment with sugar-binding proteins is concordant with genetic and chemical data. J. Biol. Chem. 1993;268:17602-17612.

6. Oh BH, Pandit J, Kang CH, Nikaido K, Gokcen S, Ames GF, Kim SH. Three-dimensional structures of the periplasmic lysine/arginine/ornithine-binding protein with and without a ligand. J. Biol. Chem. 1993;268:11348-11355.

7. Conklin BR, Bourne HR. Homeostatic signals. Marriage of the flytrap and the serpent. Nature 1994;367:22.

8. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J. Mol. Biol. 1990;215:403-410.

9. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389-3402.

10. Graul RC, Sadée W. Evolutionary relationships among proteins probed by an iterative neighborhood cluster analysis (INCA). Alignment of bacteriorhodopsins with the yeast sequence YRO2. Pharm. Res. 1997;14:1533-1541.

11. Richarme G, Caldas TD. Chaperone properties of the bacterial periplasmic substrate-binding proteins. J. Biol. Chem. 1997;272:15607-15612.

12. Schuler GD, Epstein JA, Ohkawa H, Kans JA. Entrez: molecular biology database and retrieval system. Methods Enzymol. 1996;266:141-162.

13. Higgins CF. ABC transporters: from microorganisms to man. Annu. Rev. Cell Biol. 1992;8:67-113.

14. Shilton BH, Flocco MM, Nilsson M, Mowbray SL. Conformational changes of three periplasmic receptors for bacterial chemotaxis and transport: the maltose-, glucose/galactose- and ribose-binding proteins. J. Mol. Biol. 1996;264:350-363.

15. Wolf A, Lee KC, Kirsch JF, Ames GFL. Ligand-dependent conformational plasticity of the periplasmic histidine-binding protein HisJ. Involvement in transport specificity. J. Biol. Chem. 1996;271:21243-21250.

16. Oh BH, Ames GF, Kim SH. Structural basis for multiple ligand specificity of the periplasmic lysine-, arginine-, ornithine-binding protein. J. Biol. Chem. 1994;269:26323-26330.

17. Oh BH, Kang CH, De Bondt H, Kim SH, Nikaido K, Joshi AK, Ames GF. The bacterial periplasmic histidine-binding protein structure/function analysis of the ligand-binding site and comparison with related proteins. J. Biol. Chem. 1994;269:4135-4143.

18. Tame JR, Murshudov GN, Dodson EJ, et al.The structural basis of sequence-independent peptide binding by OppA protein. Science 1994;264:1578-1581.

19. Olah GA, Trakhanov S, Trewhella J, Quiocho FA. Leucine/isoleucine/valine-binding protein contracts upon binding of ligand. J. Biol. Chem. 1993;268:16241-16247.

20. Sack JS, Saper MA, Quiocho FA. Periplasmic binding protein structure and function. Refined X-ray structures of the leucine/isoleucine/valine-binding protein and its complex with leucine. J. Mol. Biol. 1989;206:171-191.

21. Kempner ES. Movable lobes and flexible loops in proteins. Structural deformations that control biochemical activity. FEBS Lett. 1993;326:4-10.

22. Higgin CF, Ames GF. Two periplasmic transport proteins which interact with a common membrane receptor show extensive homology: complete nucleotide sequences. Proc. Natl. Acad. Sci. U S A 1981;78:6038-6042.

23. Gilson E, Alloing G, Schmidt T, Claverys JP, Dudler R, Hofnung M. Evidence for high affinity binding-protein dependent transport systems in gram-positive bacteria and in Mycoplasma. Embo J. 1988;7:3971-3974.

24. Yoshida K, Fujimura M, Yanai N, Fujita Y. Cloning and sequencing of a 23-kb region of the Bacillus subtilis genome between the iol and hut operons. DNA Res. 1995;2:295-301.

25. Kronemeyer W, Peekhaus N, Kramer R, Sahm H, Eggeling L. Structure of the gluABCD cluster encoding the glutamate uptake system of Corynebacterium glutamicum. J. Bacteriol. 1995;177:1152-1158.

26. Turner MS, Timms P, Hafner LM, Giffard PM. Identification and characterization of a basic cell surface-located protein from lactobacillus fermentum BR11. J. Bacteriol. 1997;179:3310-3316.

27. Roos S, Aleljung P, Robert N, Lee B, Wadstrom T, Lindberg M, Jonsson H. A collagen binding protein from Lactobacillus reuteri is part of an ABC transporter system? FEMS Microbiol. Lett. 1996;144:33-38.

28. Pei Z, Blaser MJ. PEB1, the major cell-binding factor of Campylobacter jejuni, is a homolog of the binding component in gram-negative nutrient transport systems. J. Biol. Chem. 1993;268:18717-18725.

29. Bowie JU, Luthy R, Eisenberg D. A method to identify protein sequences that fold into a known three-dimensional structure. Science 1991;253:164-170.

30. Friedman AM, Fischmann TO, Steitz TA. Crystal structure of lac repressor core tetramer and its implications for DNA looping. Science 1995;268:1721-1727.

31. Schumacher MA, Choi KY, Zalkin H, Brennan RG. Crystal structure of lacI member, PurR, bound to DNA: minor groove binding by alpha helices. Science 1994;266:763-770.

32. Nohno T, Saito T, Hong JS. Cloning and complete nucleotide sequence of the Escherichia coli glutamine permease operon (glnHPQ). Mol. Gen. Genet. 1986;205:260-269.

33. Kaneko T, Sato S, Kotani H, et al. Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. DNA Res. 1996;3:109-136.

34. Graul RC, Sadée W. Sequence alignments of the H+-dependent oligopeptide transporter family PTR: inferences on structure and function of the intestinal PET1 transporter. Pharm. Res. 1997;14:388-400.

35. Nayak A, Zastrow DJ, Lickteig R, Zahniser NR, Browning MD. Maintenance of late-phase LTP is accompanied by PKA-dependent increase in AMPA receptor synthesis. Nature 1998;394:680-683.

36. Armstrong N, Sun Y, Chen GQ, Gouaux E. Structure of a glutamate-receptor ligand-binding core in complex with kainate. Nature 1998;395:913-917.

37. Masu M, Tanabe Y, Tsuchida K, Shigemoto R, Nakanishi S. Sequence and expression of a metabotropic glutamate receptor. Nature 1991;349:760-765.

38. Houamed KM, Kuijper JL, Gilbert TL, et al. Cloning, expression, and gene structure of a G protein-coupled glutamate receptor from rat brain. Science 1991;252:1318-1321.

39. Moghaddam B, Adams BW. Reversal of phencyclidine effects by a group II metabotropic glutamate receptor agonist in rats. Science 1998;281:1349-1352.

40. Cockcroft VB, Ortells MO, Thomas P, Lunt GG. Homologies and disparities of glutamate receptors: a critical analysis. Neurochem. Int. 1993;23:583-594.

41. Kaupmann K, Huggel K, Heid J, et al. Expression cloning of GABA(B) receptors uncovers similarity to metabotropic glutamate receptors. Nature 1997;386:239-246.

42. Aprison MH, Galvez-Ruano E, Lipkowitz KB. The nicotinic cholinergic receptor: a theoretical model. J. Neurosci. Res. 1996;46:226-230.

43. Smith GB, Olsen RW. Functional domains of GABAA receptors. Trends Pharmacol. Sci. 1995;16:162-168.

44. Brown EM, Vassilev PM, Hebert SC. Calcium ions as extracellular messengers. Cell 1995;83:679-682.

45. Brown EM, Gamba G, Riccardi D, et al. Cloning and characterization of an extracellular Ca2+-sensing receptor from bovine REFthyroid. Nature 1993;366:575-580.

46. Garrett JE, Capuano IV, Hammerland LG, et al. Molecular cloning and functional expression of human REFthyroid calcium receptor cDNAs. J. Biol. Chem. 1995;270:12919-12925.

47. Vyas NK, Vyas MN, Quiocho FA. A novel calcium binding site in the galactose-binding protein of bacterial transport and chemotaxis. Nature 1987;327:635-638.

48. Kubo Y, Miyashita T, Murata Y. Structural basis for a Ca2+-sensing function of the metabotropic glutamate receptors. Science 1998;279:1722-1725.

49. Baron J, Winer KK, Yanovski JA, et al. Mutations in the Ca2+-sensing receptor gene cause autosomal dominant and sporadic hypoREFthyroidism. Hu m. Mol. Genet. 1996;5:601-606.

50. Pollak MR, Brown EM, Chou YH, et al. Mutations in the human Ca2+-sensing receptor gene cause familial hypocalciuric hypercalcemia and neonatal severe hyperREFthyroidism. Cell 1993;75:1297-1303.

51. Pearce SH, Trump D, Wooding C, et al. Calcium-sensing receptor mutations in familial benign hypercalcemia and neonatal hyperREFthyroidism. J. Clin. Invest. 1995;96:2683-2692.

52. Herrada G, Dulac C. A novel family of putative pheromone receptors in mammals with a topographically organized and sexually dimorphic distribution. Cell 1997;90:763-773.

53. Dulac C, Axel R. A novel family of genes encoding putative pheromone receptors in mammals. Cell 1995;83:195-206.

54. Nakao K, Itoh H, Saito Y, Mukoyama M, Ogawa Y. The natriuretic peptide family. Cur. Opin. Nephrol. Hypertension 1996;5:4-11.

55. Romano C, Yang WL, O'Malley KL. Metabotropic glutamate receptor 5 is a disulfide-linked dimer. J. Biol. Chem. 1996;271:28612-28616.

56. Ward DT, Brown EM, Harris HW. Disulfide bonds in the extracellular calcium-polyvalent cation-sensing receptor correlate with dimer formation and its response to divalent cations in vitro. J. Biol. Chem. 1998;273:14476-14483.

57. Jones KA, Borowsky B, Tamm JA, et al. GABABreceptors function as a heteromeric assembly of the subunits GABABR1 and GABABR2. Nature 1998;396:674-679.

58. White JH, Wise A, Main MJ, et al. Heterodimerization is required for the formation of a functional GABAB receptor. Nature 1998;396:679-682.

59. Kaupmann K, Malitschek B, Schuler V, et al. GABAB-receptor subtypes assemble into functional heteromeric complexes. Nature 1998;396:683-687.

60. Nakao K, Ogawa Y, Suga S, Imura H. Molecular biology and biochemistry of the natriuretic peptide system. II: Natriuretic peptide receptors. J. Hypertens. 1992;10:1111-1114.

61. Nakao K, Ogawa Y, Suga S, Imura H. Molecular biology and biochemistry of the natriuretic peptide system. I: Natriuretic peptides. J. Hypertens. 1992;10:907-912.

62. Chang MS, Lowe DG, Lewis M, Hellmiss R, Chen E, Goeddel DV. Differential activation by atrial and brain natriuretic peptides of two different receptor guanylate cyclases. Nature 1989;341:68-72.

63. Schulz S, Singh S, Bellet RA, et al. The primary structure of a plasma membrane guanylate cyclase demonstrates diversity within this new receptor f amily. Cell 1989;58:1155-1162.

64. Lowe DG, Chang MS, Hellmiss R, et al. Human atrial natriuretic peptide receptor defines a new REFdigm for second messenger signal transduction. Embo J. 1989;8:1377-1384.

65. Fuller F, Porter JG, Arfsten AE, et al. Atrial natriuretic peptide clearance receptor. Complete sequence and functional expression of cDNA clones. J. Biol. Chem. 1988;263:9395-9401.

66. Lowe DG, Camerato TR, Goeddel DV. cDNA sequence of the human atrial natriuretic peptide clearance receptor. Nucleic Acids Res. 1990;18:3412.

67. Chinkers M, Garbers DL. The protein kinase domain of the ANP receptor is required for signaling. Science 1989;245:1392-1394.

68. Kishimoto I, Yoshimasa T, Suga S, et al. Natriuretic peptide clearance receptor is transcriptionally down-regulated by b2-adrenergic stimulation in vascular smooth muscle cells. J. Biol. Chem. 1994;269:28300-28308.

69. Kishimoto I, Nakao K, Suga S, et al. Downregulation of C-receptor by natriuretic peptides via ANP-B receptor in vascular smooth muscle cells. Amer. J. Physiol. 1993;265:H1373-1379.

70. Suga S, Nakao K, Hosoda K, et al. Receptor selectivity of natriuretic peptide family, atrial natriuretic peptide, brain natriuretic peptide, and C-type natriuretic peptide. Endocrinology 1992;130:229-239.

71. Koller KJ, Lowe DG, Bennett GL, et al. Selective activation of the B natriuretic peptide receptor by C-type natriuretic peptide (CNP). Science 1991;252:120-123.

72. Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994;22:4673-4680.

73. Gardner J, SeqVu. The Garvan Institute of Medical Research, 384 Victoria Rd., Darlinghurst NSW 2010, Sydney Australia, Sydney, Australia, 1998.

74. Felsenstein J. Inferring phylogenies from protein sequences by parsimony, distance, and likelihood methods. Methods Enzymol. 1996;266:418-427.

75. Page RD. TreeView: An application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 1996;12:357-358.