Construction of a sequence motif characteristic of aminergic G protein–coupled receptors

ENOCH S. HUANG

Pfizer Discovery Technology Center, Cambridge, Massachusetts 02139, USA

(RECEIVED February 14, 2003; FINAL REVISION April 1, 2003; ACCEPTED April 18, 2003)

Abstract

An approach to discover sequence patterns characteristic of ligand classes is described and applied to aminergic G protein–coupled receptors (GPCRs). Putative ligand-binding residue positions were inferred from considering three lines of evidence: conservation in the subfamily absent or underrepresented in the superfamily, any available mutation data, and the physicochemical properties of the ligand. For aminergic GPCRs, the motif is composed of a conserved aspartic acid in the third transmembrane (TM) domain (rhodopsin position 117) and a conserved tryptophan in the seventh TM domain (rhodopsin position 293); the roles of each are readily justified by molecular modeling of ligand-receptor interactions. This minimally defined motif is an appropriate computational tool for identifying additional, potentially novel aminergic GPCRs from a set of experimentally uncharacterized "orphan" GPCRs, complementing existing sequence matching, clustering, and machine-learning techniques. Motif sensitivity stems from the stepwise addition of residues characteristic of an entire class of ligand (and not tailored for any particular biogenic amine). This sensitivity is balanced by careful consideration of residues (evidence drawn from mutation data, correlation of ligand properties to residue properties, and location with respect to the extracellular face), thereby maintaining specificity for the aminergic class. A number of orphan GPCRs assigned to the aminergic class by this motif were later discovered to be a novel subfamily of trace amine GPCRs, as well as the successful classification of the histamine H4 receptor.

Keywords: G protein–coupled receptor; multiple sequence alignment; bioinformatics; motif discovery; protein-ligand interactions; biogenic amines

Among membrane-bound receptors, the seven-transmembrane (7TM) receptors are the most abundant, comprising >700 genes in human beings. They are widely expressed and transduce signals in response to a large diversity of physiologically important molecules, including organic odorants, nucleotides, nucleosides, peptides, lipids, and proteins. They are also catalysts of the GDP/GTP nucleotide exchange on heterotrimeric G proteins and, hence, are also referred to as G protein–coupled receptors (GPCRs). Approximately half of the GPCR superfamily comprises nonsensory receptors, which are potentially attractive novel targets for small molecule intervention. In fact, this receptor superfamily is regarded as the most successful target class in terms of therapeutic benefit and potential sales (Wise et al. 2002).

One particularly important subfamily of GPCRs is represented by the aminergic receptors, with natural ligands that include dopamine, serotonin, acetylcholine, epinephrine, and histamine. Indeed, therapeutic agents currently on the market have as molecular targets members from all five major subtypes of aminergic GPCRs. In the human genome, there are ~150 so-called orphan GPCRs with unknown natural ligands, any of which might become a molecular target for future drugs (Wilson et al. 1998). Because a full understanding of a possible pathophysiological role for an orphan receptor requires knowledge of its cognate ligand, aminergic GPCR identification poses an interesting and important challenge for computational classification methods.

Reprint requests to: Enoch S. Huang, Pfizer Discovery Technology Center, 620 Memorial Drive, Cambridge, MA 02139, USA; e-mail: enoch_huang@cambridge.pfizer.com; fax: (617) 551-3117.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1110/ps.0305603.

 Copyright © 2003 The Protein Society

The approach described here pertains to the discovery of minimal sequence motifs corresponding to conserved ligand-binding residues. Such a motif can be inferred from a high-quality multiple sequence alignment of a protein family, knowledge of its three-dimensional fold, and experimental data relating to ligand-binding function. In the case of GPCRs, these typically involve site-directed mutagenesis data. The motifs are built from residues that are as specific as possible for a GPCR subfamily (Kuipers et al. 1997), especially those that can form plausible intermolecular interactions with the ligand class. As will be discussed, the goal is to balance specificity with sensitivity by including only those positions necessary to capture members of a given ligand-binding class. Although the aim is to discover a motif essential for biogenic amine recognition, there is no reason why it could not also be extended to other ligand classes. Thus, the following section was deliberately written as broadly as possible.

The approach

First, one must construct a multiple sequence alignment of all class A (rhodopsin-like) GPCRs for which ligands are identified. Aligned positions (residues that fall in the same column in the alignment) are either known or presumed to be structurally equivalent. Thus, the primary sequence of rhodopsin and the location of its 7TM helices serve as an important reference sequence in the multiple alignment. In this case, the program hmmalign of the HMMer suite (Eddy 1998, 2001) was used to build a preliminary alignment of GPCR sequences against the 7tm_1 Pfam model (Bateman et al. 2002).

This machine-built alignment was then hand-edited by using the software Pfaat (Johnson et al. 2003). Several sequences failed to align to portions of the Pfam model corresponding to transmembrane segments. In these instances, the conserved residues characteristic of each segment were used as anchors (see Appendix). Alignment was also guided by the principle that residues embedded in transmembrane domains tend to be hydrophobic. Moreover, gap positions in transmembrane domains were disallowed because the membrane bilayer is relatively fixed in thickness, and ~25 amino acids in a helical conformation are required to span it. A final consideration is a conserved disulfide bond present between the start of the third transmembrane domain and the second extracellular loop (corresponding to residues 110 and 187 of rhodopsin).

After constructing the alignment, one systematically marks residues (or residue classes) conserved in a given subfamily of GPCRs (in this case, the aminergic subfamily). If the residue or class of residues is not also characteristic of the superfamily, it is set aside for consideration as a ligand-binding residue. Not all residue types receive equal consideration, however. Specificity of ligand-receptor interactions is often due to electronic interactions, manifested as hydrogen-bond pairs, ionic bonds, and aromatic interactions. Even if aliphatic side-chains actually touch the ligand within a transmembrane pocket, they often are not suitable as part of discrimination motif because hydrophobic residues are commonly seen within the helical regions. Thus, conserved polar, charged, and aromatic amino acids, especially within the transmembrane domains, are evaluated as determinants of ligand-binding specificity for a given subfamily. Regions of the alignment that fall within the intracellular portion of the receptor are not considered. These include the three intracellular loops and the C-terminal domain in their entirety, as well as portions of each of the TM domains. One disqualifies these alignment positions, as they are unlikely to interact directly with the cognate ligand, which is typically presented to the receptor from the extracellular face of the receptor.

The putative role of these amino acids can be supported by properly controlled site-directed mutagenesis experiments that reveal adverse effects on ligand binding and/or signaling. This step is important to distinguish residues conserved in a subfamily due to phylogeny from those that are conserved due to functional constraints. Data of this type are readily available in the literature (Beukers et al. 1999).

If possible, the physicochemical properties of the conserved amino acid (or type) are then matched with the shared physicochemical properties of the ligand type. For example, if an amino acid conserved in the subfamily happens to be positively charged, it would be useful to identify a negatively charged moiety in the ligand (such as a phosphate group) that might directly interact with it. Successful correlation of these data lends additional support to the hypothesis that a given residue or set of residues is responsible for ligand specificity, but is not necessary for a pattern discovery in the general case.

Finally, all implicated positions and their residue identities (or classes) are collected, forming a final set from which to build a discrimination motif for the subfamily for refinement and evaluation for sensitivity and selectivity. One approach is simply to search exhaustively over all combinations of residue (or residue types) to optimize selectivity and sensitivity. Alternatively, one can select the position that is conserved throughout the subfamily and has minimal representation in other subfamilies. If this residue is absent in all other subfamilies, this amino acid may in itself constitute a subfamily motif. However, if this residue or residue class is seen in the same position in other subfamilies, one adds other positions that are also completely conserved in the subfamily but are increasingly common in other subfamilies. After each subsequent addition, the emerging motif is assessed for specificity. This iterative refinement procedure would terminate when a motif is constructed that describes the subfamily of interest without also matching any other

sequence of another subfamily. A stepwise addition of additional conserved positions is desirable to optimize sensitivity of the motif without sacrificing specificity. Avoiding positions not supported by mutagenesis data also minimizes the risk of adding to the motif residues unrelated to ligand binding.

Results

Aminergic subfamily definition

The aminergic subfamily of GPCRs includes receptors for dopamine, serotonin, epinephrine, histamine, and acetylcholine. At the time of this study, only 33 of these had been experimentally characterized and in publicly available databases (see Appendix).

Conserved residues in aminergic GPCRs

Residue numbering of rhodopsin is shown in parentheses. In the alignment, there were 20 residues that were completely conserved in all known aminergic GPCRs:

TM1:	**Asn**	**(55)**
TM2:	**Asp**	**(83)**
EC1:	Trp	(103)
TM3:	**Cys**	**(110)**
TM3:	Asp	(117)
TM3:	Ser	(124)
TM3:	Asp	(134)
TM3:	**Arg**	**(135)**
TM4:	**Trp**	**(161)**
EC2:	**Cys**	**(187)**
TM5:	Phe	(212)
TM5:	**Pro**	**(215)**
TM6:	Phe	(261)
TM6:	Trp	(265)
TM6:	**Pro**	**(267)**
TM7:	Trp	(293)
TM7:	Ser	(299)
TM7:	Asn	(302)
TM7:	**Pro**	**(303)**
TM7:	Tyr	(306)

In bold type are the nine residues mentioned in the Appendix that represent the most conserved residues in each helix, plus the conserved disulfide bond. Because these conserved residues are also highly prevalent throughout the GPCR class A family, they are not considered as subfamily discrimination residues.

Next, the residues that fall within the masked (intracellular) portions of the receptors are disregarded. These include the conserved Asp (rhodopsin, 134), Asn (302), and Tyr (306). The set of residues surviving the second filter include Trp (103), Asp (117), Ser (124), Phe (212), Phe (261), Trp (265), Trp (293), and Ser (299).

Now one sorts these conserved positions, in ascending order, based on the number of nonaminergic GPCR sequences that share the respective amino acid.

TM3:	Asp	(117)	11
TM7:	Trp	(293)	14
TM5:	Phe	(212)	81
TM7:	Ser	(299)	84
EC1:	Trp	(103)	97
TM3:	Ser	(124)	104
TM6:	Trp	(265)	106
TM6:	Phe	(261)	115

The data indicate that the aspartic acid in TM3 is also the most specific for the aminergic receptors (Fig. 1). Its position in rhodopsin is 117, which is 17 residues ahead of the

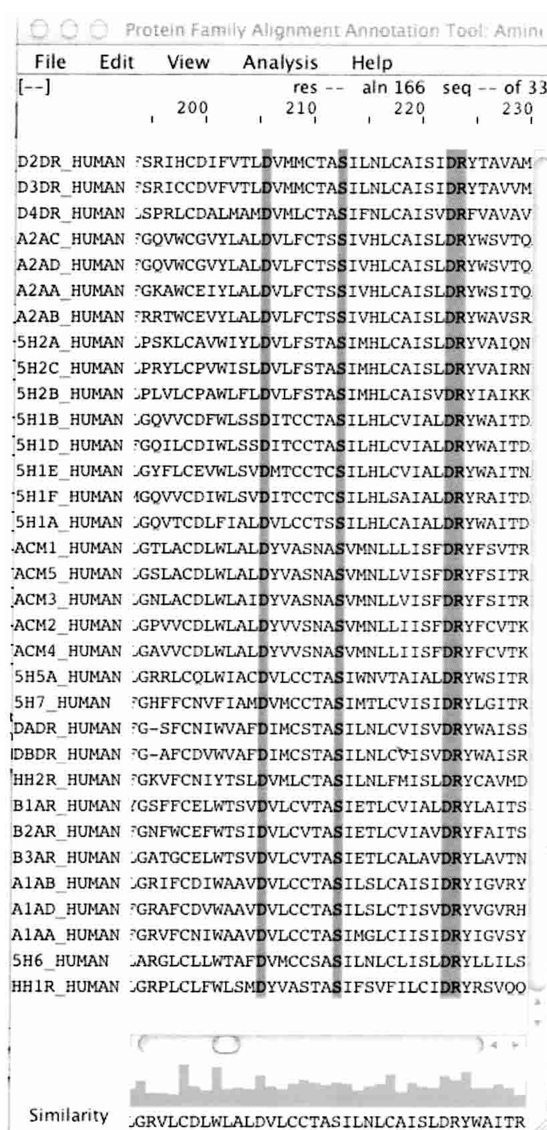

Figure 1. Screenshot of the aligned third transmembrane domain of aminergic GPCRs visualized in Pfaat (Johnson et al. 2003). Conserved positions are highlighted, the first being the aspartic acid found in the motif.

characteristic "DRY" motif at the end of TM3 (Appendix). There is a wealth of published mutation data implicating this role of this residue in recognizing bioamines of all five major subtypes: histamine (Gantz et al. 1992), serotonin (Ho et al. 1992), dopamine (Mansour et al. 1992), epinephrine (Wang et al. 1991), and acetylcholine (Fraser et al. 1989), to cite a few early examples. The consensus model based on these data is that the negatively charged side-chain of aspartic acid participates directly in an ionic interaction with the positively charged amine group conserved in bioamine ligands (for an example, see Donnelly et al. 1994).

Although this interaction may be necessary for specific binding of aminergic GPCRs to their cognate ligands, an aspartic acid at this key position is not sufficient to distinguish aminergic GPCRs from all nonaminergic GPCRs. However, by adding the conserved tryptophan (rhodopsin, 293; 10 positions ahead of the conserved proline in TM7; see Appendix and Fig. 2) to the motif, one can distinguish aminergic GPCRs to the exclusion of all nonaminergic GPCRs. The role of this conserved tryptophan in ligand binding is less well characterized but was implicated in one mutagenesis experiment (Roth et al. 1997). Furthermore, it was modeled to interact with the positively charged amine group via an amine-aromatic interaction (Roth et al. 1997; see also references therein). A simple homology model built with the software SCWRL (Bower et al. 1997) reveals that the aspartate and tryptophan residues are in close spatial proximity (Fig. 4A, B).

This refined motif is a suitable computational tool for classifying orphan GPCR sequences with the aim of predicting novel members of the aminergic family, even those with ligands outside of the five major types. Equally important is the absence of predicted GPCRs that are later shown experimentally to be nonaminergic. Confidence in this motif stems from the sensitivity gained by stepwise addition of residues characteristic of an entire class of ligand (and not tailored for any particular biogenic amine). This sensitivity is balanced by careful consideration of residues (evidence drawn from mutation data, correlation of ligand properties to residue properties, and location with respect to the extracellular face), thereby maintaining specificity for the aminergic class.

Figure 2. Screenshot of the aligned seventh transmembrane domain of aminergic GPCRs visualized in Pfaat (Johnson et al. 2003). Conserved positions are highlighted, the first being the tryptophan found in the motif.

Nonaminergic GPCRs with a conserved aspartic acid

As discussed earlier, the presence of the conserved aspartic acid TM3 is necessary but insufficient to distinguish aminergic GPCRs to the exclusion of others. As observed by MacDonald (2000; see also references therein), the presence of the residue is important for natural ligand recognition and activation by certain peptide GPCRs, specifically opioid, somatostatin, melanin-concentrating hormone, and urotensin-II (accounting for the 11 nonaminergic GPCRs). In the context of opioid GPCRs, the negatively charged aspartate is thought to interact with the positively charged and conserved N terminus of the opioid peptides (Surratt et al. 1994; Befort et al. 1999, Li et al. 1999; Lavecchia et al. 2000). The other three peptiderigic ligands are cyclic (due to a disulfide bridge) and share positively charged amino acids such as K and R in the cyclic part of the ligand (Fig. 3). Mutagenesis experiments and ligand structural data indicate that these positively charged amino acids interact with the conserved aspartate in TM3 (Nehring et al. 1995; Strnad and Hadcock 1995; MacDonald et al. 2000; Flohr et al. 2002).

Recent pairings of ligands to orphan GPCRs

Shortly after the aminergic motif was defined, it was applied to an in-house sequence database of orphan GPCRs and

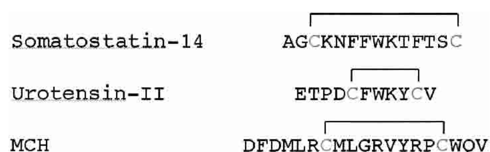

Figure 3. Cyclic peptides that are the natural agonists of GPCRs possessing the conserved aspartic acid but not the tryptophan of the aminergic sequence motif. Disulfide bridges linking the cysteines are schematically drawn.

found eight sequences to be potentially new bioamine GPCRs (Table 1). A few of these correspond to those mentioned in a recent paper that describe trace amine receptors, a novel GPCR subfamily related to, but distinct from, the classic aminergic GPCRs (Borowsky et al. 2001). In addition, a paper by Oda et al. (2000) describes the molecular cloning of a novel histamine receptor (GenBank, AY008280). There are also five orphan GPCRs that possess the conserved aspartate in TM3 but lack the conserved tryptophan in TM7 in Table 2. One (GenBank, AF347063) was recently discovered to be a second melanin concentrating hormone GPCR (Hill et al. 2001), and two others (GenBank, U22491, U22492) have also been paired with peptide ligands (Brezillon et al. 2003). All of these data are consistent with the working hypothesis that a partial match to the motif indicates a peptide ligand rather than an aminergic motif.

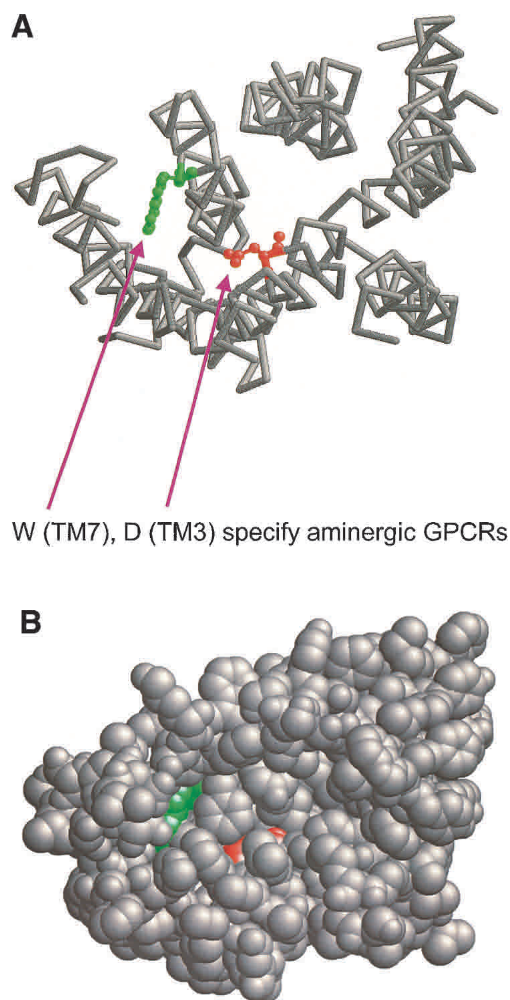

Figure 4. (*A*) Homology model of histamine H4 receptor indicating the proximity of both residues comprising the aminergic motif. (*B*) Space-filling model of the H4 receptor.

Table 1. *Sequences possessing full motif (D in TM3, W in TM7)*

GenBank	Refseq	Comment
AF021818	NM_003967	PNR
AF112460	NM_014626	GPR58
AF112461	NM_014627	Temporarily removed; under review by Refseq
AY008280	NM_021624	HRH4
AF380193	NM_053278	GPR102
AF200627	NM_138327	TAR1
AY183470	NM_175067	TA4
AY183469	NM_175057	TAR3

Discussion

This approach described above is an alternative to and improvement over many effective computational classification schemes, coarsely grouped for the sake of discussion into four categories: pattern discovery and matching, phylogenetic or clustering analysis, traditional sequence comparison (pairwise and profile-based), and other machine-learning approaches (e.g., support vector machines, artificial neural networks). Sampled from the recent literature are examples of GPCR classification techniques: GPCR subtype motifs (Attwood 2001, 2002), hierarchical clustering (Joost and Methner 2002), subfamily profile comparison (Graul and Sadee 2001), support vector machines (Karchin et al. 2002), and an alignment-independent classification based on principal chemical properties of GPCR sequences (Lapinsh et al. 2002).

In one sense, the method described here differs from those listed above primarily in the level of manual inspec-

Table 2. *Sequences possessing partial motif (D in TM3, no W in TM7)*

GenBank	Refseq	Comment
U22491	NM_005285	GPR7
U22492	NM_005286	GPR8
AF411107	NM_080819	GPR78
AF347063	NM_032503	SLT (MCH2)
AJ505757	NM_153442	GPR26

tion, intervention, and curation of the multiple sequence alignment and associated literature information. Underlying this statement, however, is a key pitfall of more automated or unsupervised approaches, that of alignment quality. In phylogenetic as well as pairwise and profile-based sequence analyses, the results ultimately depend on the accuracy of the alignments. Sequence comparisons also have the issue of scoring metrics and cutoffs to discern subfamily membership (and hence ligand-binding function). Phylogenetic or clustering approaches avoid this problem, but may produce groups or subtrees that comprise receptors with mixed ligand types or only other orphan GPCRs (Joost and Methner 2002). Finally, these methods typically involve analysis of full-length sequences, the scoring and/or clustering of which may mask ligand-binding function even assuming perfect alignment. The work by Johnson and Church (2000), which focused on ligand-binding residues in the context of protein family, effectively addressed this issue plaguing subfamily prediction.

In contrast, one of the main advantages of motif-based approaches is that assigning membership to a predefined group is a binary decision, the basis for which is often interpretable in light of protein structure and function. This has practical value in determining whether functional annotation can be safely transferred from pairwise comparisons, for example, using BLAST. For instance, a sequence in GenBank (AF258342) has been annotated as "biogenic amine receptor-like BALGR," presumably because of its sequence homology with known aminergic GPCRs. Indeed, when using BLASTP against NR, the closest match to an experimentally characterized GPCR is the human histamine H2 receptor (data not shown). However, BALGR, which remains an orphan GPCR, does not possess the aminergic motif. Without motif analysis, and in the absence of universally applicable BLAST cutoffs, there is no obvious criterion by which to assign membership to the aminergic class.

Motifs are also a powerful approach because their discovery are not necessarily dependent on prior multiple sequence alignments (compare, e.g., the motif collections of Attwood et al. 1994 and Rigoutsos et al. 1999). The key differences between the excellent and comprehensive work by Attwood and coworkers (2002) on motif-based GPCR classification and the present one relate to specificity, scope, and throughput. In Attwood's hierarchical compendium of GPCR-specific motifs, or fingerprints, the second level under the class A (all rhodopsin-like) category contains all the GPCRs from individual ligand groups (e.g., dopamine, bradykinin, melatonin) with any appropriate subtypes at even lower levels. The motifs for each group are extracted from multiple sequence alignments of the respective sequences and their orthologs from various species. The result is that the motifs, relative to the one described in this study, are much richer in content and thus exquisitely specific for either the ligand type or receptor subtype. However, it is unlikely that Attwood's GPCR fingerprints would be sensitive enough to recognize the aminergic nature of a related, but distinct, novel group of GPCRs such as the trace amine receptors (Borowsky et al. 2001). However, Attwood et al. are able to cover many more GPCR ligand-specific families and subfamilies in a highly automated fashion while providing impressive diagnostic power.

In summary, the motif-based approach described herein offers several features that make it a valuable alternative to and enhancement over current GPCR classification methods, with particular utility toward subfamilies such as the aminergic class, because they comprise a number of chemically related but distinct ligands. Other families that are potentially subjects for future study include the nucleotide and lipid GPCR subfamilies, and even GPCRs activated by families of macromolecules, such as chemokine and complement proteins. This approach, which ideally involves curator inspection and alignment, dovetails with existing and ongoing efforts that provide higher throughput annotation.

Acknowledgments

I wish to thank Nicholas Slepchuk for his careful reading of this document and Nikki Van Noy for her assistance in manuscript preparation and formatting.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

Appendix

Boundaries in rhodopsin

(Adapted from Palczewski et al. 2000, Figure 3.)

Domain	Start residue	End residue
N-terminal domain:	M1	W35
TM1	Q36	V63
IC1	Q64	L72
TM2	N73	H100
EC1	G101	P107
TM3	T108	V137
IC2	V138	E150
TM4	N151	V173
EC2	G174	N200
TM5	E201	V227
IC3	F228	K245
TM6	A246	F276
EC3	T277	P285
TM7	I286	L321
C-terminal domain	C322	A348

Most conserved residues located in rhodopsin family

(Adapted from Ballesteros and Weinstein 1995.)

TM Domain	Residue type	Residue number
TM1	N	55
TM2	D	83
TM3	R	135
TM4	W	161
TM5	P	215
TM6	P	267
TM7	P	303

Note also the disulfide bridge between C110 and C187

List of 33 aminergic GPCRs in SwissProt

(From Bairoch and Apweiler 2000.)

5H1A_HUMAN
5H1B_HUMAN
5H1D_HUMAN
5H1E_HUMAN
5H1F_HUMAN
5H2A_HUMAN
5H2B_HUMAN
5H2C_HUMAN
5H5A_HUMAN
5H6_HUMAN
5H7_HUMAN
A1AA_HUMAN
A1AB_HUMAN
A1AD_HUMAN
A2AA_HUMAN
A2AB_HUMAN
A2AC_HUMAN
A2AD_HUMAN
ACM1_HUMAN
ACM2_HUMAN
ACM3_HUMAN
ACM4_HUMAN
ACM5_HUMAN

References

Attwood, T.K. 2001. A compendium of specific motifs for diagnosing GPCR subtypes. *Trends Pharmacol. Sci.* **22:** 162–165.

Attwood, T.K., Beck, M.E., Bleasby, A.J., and Parry-Smith, D.J. 1994. PRINTS: A database of protein motif fingerprints. *Nucleic Acids Res.* **22:** 3590–3596.

Attwood, T.K., Croning, M.D., and Gaulton, A. 2002. Deriving structural and functional insights from a ligand-based hierarchical classification of G protein–coupled receptors. *Protein Eng.* **15:** 7–12.

Bairoch, A. and Apweiler, R. 2000. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28:** 45–48.

Ballesteros, J.A. and Weinstein, H. 1995. Integrated methods for the construction of three dimensional models and computational probing of structure-function relations in G-protein–coupled receptors. *Meth. Neurosci.* **25:** 366–425.

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M., and Sonnhammer, E.L. 2002. The Pfam protein families database. *Nucleic Acids Res.* **30:** 276–280.

Befort, K., Zilliox, C., Filliol, D., Yue, S., and Kieffer, B.L. 1999. Constitutive activation of the δ opioid receptor by mutations in transmembrane domains III and VII. *J. Biol. Chem.* **274:**18574–18581.

Beukers, M.W., Kristiansen, K., Ijzerman, A.P., and Edvardsen, Ø. 1999. TinyGRAP database: A bioinformatics tool to mine G protein–coupled receptor mutant data. *TiPS* **20:** 475–477.

Borowsky, B., Adham, N., Jones, K.A., Raddatz, R., Artymyshyn, R., Ogozalek, K.L., Durkin, M.M., Lakhlani, P.P., Bonini, J.A., Pathirana, S., et al. 2001. Trace amines: Identification of a family of mammalian G protein–coupled receptors. *Proc. Natl. Acad. Sci.* **98:** 8966–8971.

Bower, M.J., Cohen, F.E., and Dunbrack Jr., R.L. 1997. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: a new homology modeling tool. *J. Mol. Biol.* **267:** 1268–1282.

Brezillon, S., Lannoy, V., Franssen, J.D., Le Poul, E., Dupriez, V., Lucchetti, J., Detheux, M., and Parmentier, M. 2003. Identification of natural ligands for the orphan G protein–coupled receptors GPR7 and GPR8. *J. Biol. Chem.* **278:** 776–783.

Donnelly, D., Findlay, J.B., and Blundell, T.L., 1994. The evolution and structure of aminergic G protein–coupled receptors. *Receptors Channels* **2:** 61–78.

Eddy, S.R. 1998. Profile hidden Markov models. *Bioinformatics* **14:**755–763.

———. 2001. HMMER: Profile hidden Markov models for biological sequence analysis (http://hmmer.wustl.edu/).

Flohr, S., Kurz, M., Kostenis, E., Brkovich, A., Fournier, A., and Klabunde, T. 2002. Identification of nonpeptidic urotensin II receptor antagonists by virtual screening based on a pharmacophore model derived from structure-activity relationships and nuclear magnetic resonance studies on urotensin II. *J. Med. Chem.* **45:** 1799–1805.

Fraser, C.M., Wang, C.D., Robinson, D.A., Gocayne, J.D., and Venter, J.C. 1989. Site-directed mutagenesis of m1 muscarinic acetylcholine receptors: Conserved aspartic acids play important roles in receptor function. *Mol. Pharmacol.* **36:** 840–847.

Gantz, I., DelValle, J., Wang, L.D., Tashiro, T., Munzert, G., Guo, Y.J., Konda, Y., and Yamada, T. 1992. Molecular basis for the interaction of histamine with the histamine H2 receptor. *J. Biol. Chem.* **267:** 20840–20843.

Graul, R.C. and Sadee, W. 2001. Evolutionary relationships among G protein-coupled receptors using a clustered database approach. *AAPS PharmSci.* **3:** E12.

Hill, J., Duckworth, M., Murdock, P., Rennie, G., Sabido-David, C., Ames, R.S., Szekeres, P., Wilson, S., Bergsma, D.J., Gloger, I.S., et al. 2001. Molecular cloning and functional characterization of MCH2, a novel human MCH receptor. *J. Biol. Chem.* **276:** 20125–20129.

Ho, B.Y., Karschin, A., Branchek, T., Davidson, N., and Lester, H.A. 1992. The role of conserved aspartate and serine residues in ligand binding and in function of the 5-HT1A receptor: A site-directed mutation study. *FEBS Lett.* **312:** 259–262.

Johnson, J.M. and Church, G.M. 2000. Predicting ligand-binding function in families of bacterial receptors. *Proc. Natl. Acad. Sci.* **97:** 3965–3970.

Johnson, J.M., Mason, K., Moallemi, C., Xi, H., Somaroo, S., and Huang, E.S. 2003. Protein family annotation in a multiple alignment viewer. *Bioinformatics* **19:** 544–545.

Joost, P. and Methner, A. 2002. Phylogenetic analysis of 277 human G-protein–coupled receptors as a tool for the prediction of orphan receptor ligands. *Genome Biol.* **3:** 1–16.

Karchin, R., Karplus, K., and Haussler, D. 2002. Classifying G-protein–coupled receptors with support vector machines. *Bioinformatics* **18:** 147–159.

Kuipers, W., Oliveira, L., Vriend, G., and Ijzerman, A.P. 1997. Identification of class-determining residues in G protein–coupled receptors by sequence analysis. *Receptors Channels* **5:** 159–174.

Lapinsh, M., Gutcaits, A., Prusis, P., Post, C., Lundstedt, T., and Wikberg, J.E. 2002. Classification of G-protein coupled receptors by alignment-independent extraction of principal chemical properties of primary amino acid sequences. *Protein Sci.* **11:** 795–805.

Lavecchia, A., Greco, G., Novellino, E., Vittorio, F., and Ronsisvalle, G. 2000. Modeling of K-opioid receptor/agonists interactions using pharmacophore-based and docking simulations. *J. Med. Chem.* **43:** 2124–2134.

Li, J.G., Chen, C., Yin, J., Rice, K., Zhang, Y., Matecka, D., de Riel, J.K., DesJarlais, R.L., and Liu-Chen, L.Y. 1999. ASP147 in the third transmembrane helix of the rat μ opioid receptor forms ion-pairing with morphine and naltrexone. *Life Sci.* **65:** 175–185.

Mansour, A., Meng, F., Meador-Woodruff, J.H., Taylor, L.P., Civelli, O., and Akil, H. 1992. Site-directed mutagenesis of the human dopamine D_2 receptor. *Eur. J. Pharmacol.* **227:** 205–214.

MacDonald, D., Murgolo, N., Zhang, R., Durkin, J.P., Yao, X., Strader, C.D., and Graziano, M.P. 2000. Molecular characterization of the melanin-concentrating hormone/receptor complex: identification of critical residues involved in binding and activation. *Mol. Pharmacol.* **58:** 217–225.

Nehring, R.B., Meyerhof, W., and Richter, D. 1995. Aspartic acid residue 124 in the third transmembrane domain of the somatostatin receptor subtype 3 is essential for somatostatin-14 binding. *DNA Cell. Biol.* **14:**939–944.

Oda, T., Morikawa, N., Saito, Y., Masuho, Y., and Matsumoto, S. 2000. Molecular cloning and characterization of a novel type of histamine receptor preferentially expressed in leukocytes. *J. Biol. Chem.* **275:** 36781–36786.

Page, K.M., Curtis, C.A., Jones, P.G., and Hulme, E.C. 1995. The functional role of the binding site aspartate in muscarinic acetylcholine receptors, probed by site directed mutagenesis. *Eur. J. Pharmacol.* **289:** 429–437.

Palczewski, K., Kumasaka, T., Hori, T., Behnke, C.A., Motoshima, H., Fox, B.A., Le Trong, I., Teller, D.C., Okada, T., Stenkamp, R.E., et al. 2000. Crystal structure of rhodopsin: A G protein–coupled receptor. *Science* **289:** 739–745.

Rigoutsos, I., Floratos, A., Ouzounis, C., Gao, Y., and Parida, L. 1999. Dictionary building via unsupervised hierarchical motif discovery in the sequence space of natural proteins. *Proteins* **37:** 264–277.

Roth, B.L., Shoham, M., Choudhary, M.S., and Khan, N. 1997. Identification of conserved aromatic residues essential for agonist binding and second messenger production at 5-hydroxytryptamine 2_A receptors. *Mol. Pharmacol.* **52:** 259–266.

Strnad, J. and Hadcock, J.R. 1995. Identification of a critical aspartate residue in transmembrane domain three necessary for the binding of somatostatin to the somatostatin receptor SSTR2. *Biochem. Biophys. Res. Commun.* **216:** 913–921.

Surratt, C.K., Johnson, P.S., Moriwaki, A., Seidleck, B.K., Blaschak, C.J., Wang, J.B., and Uhl, G.R. 1994. *m*-Opiate receptor: Charged transmembrane domain amino acids are critical for agonist recognition and intrinsic activity. *J. Biol. Chem.* **269:** 20548–20553.

Wang, C.D., Buck, M.A., and Fraser, C.M. 1991. Site-directed mutagenesis of α 2A-adrenergic receptors: Identification of amino acids involved in ligand binding and receptor activation by agonists. *Mol. Pharmacol.* **40:** 168–179.

Wilson, S., Bergsma, D.J., Chambers, J.K., Muir, A.I., Fantom, K.G., Ellis, C., Murdock, P.R., Herrity, N.C., and Stadel, J.M. 1998. Orphan G-protein–coupled receptors: The next generation of drug targets? *Br. J. Pharmacol.* **125:** 1387–1392.

Wise, A., Gearing, K., and Rees, S. 2002. Target validation of G-protein–coupled receptors. *Drug Discovery Today* **7:** 235–246.