Supporting information for Vassilatis *et al.* (2003) *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.0230374100

Supporting Materials and Methods

Database Mining. To identify novel endo-G protein-coupled receptors (GPCRs), the transmembrane (TM) domain one through seven (TM1–TM7) regions of 254 GPCRs were each used as a query in TBLASTN searches ($e = 10$) of the National Center for Biotechnology Information human genome database (August 2001 version). Resulting hits (500,000) were sorted by contig and position, and the hit with the best *e*-value was selected among each group of hits within 1 kb from one another in the same contig. The 50,000 unique hits generated were each used to search (BLASTP) the GenBank (nr) protein database. About 10,000 of these showed similarity to known GPCRs, including almost 2,000 that were identical to a known GPCR and excluded from further consideration. The best 500 hits were then selected, and conceptual translations of a 200-kb segment of DNA containing each were aligned (BLAST2) with the most highly related known GPCR. This identified 23 novel endoGPCR candidates, each of which was then used to search mouse and human EST databases.

In the Hidden Markov Model (HMM) profile-based search, GPCR Class A, B, and C HMM models were downloaded from the Pfam database and used as queries to search the International Protein Index proteome database (HMMSEARCH program; HMMER package). Hits with $E < 0.01$ were evaluated for seven TM domains (HMMTOP program) (1). Full-length coding sequences were predicted by using the EST sequence assembly and the ORF FINDER, GENOMESCAN, GENEWISE, and GENESCAN programs.

Classification. Shared sequence motifs were used to assign endoGPCRs to particular classes. The shared sequence motifs are multiple and scattered throughout the seven TM regions with regular distances. The most common include, but are not limited to, GN in transmembrane domain 1 (TM1), D(E)RY in TM3, and N(D)PXXY in TM7 for Class A; WL(M) in TM3, CW in TM5, and QG and EV in TM7 for Class B; PXVK between TM1

and TM2, NEAK in TM6, and FXPK in TM7 for Class C; TFXXD in TM1, and YFH and VDGD in TM4 for Class F/S.

Phylogenetic Analyses. endoGPCRs of the same class were aligned to the class-specific HMM model (HMMALIGN; HMMER package). Positions not aligned to the HMM model were removed. Neighbor-joining phylogenetic trees were then built (CLUSTALW) from these alignments. Gaps and substitutions were not corrected. Bootstrap consensus trees were plotted by using TREEVIEW (2). Trees were rooted with endoGPCRs that do not fit into any of four known classes. Bootstrap values for nodes near the root of the Class A tree were very low (<10%), reflecting the distant homology of the different families in this class.

RT-PCR. RNA was prepared (Totally RNA kit, Ambion, Austin, TX) from dissected tissues of adult male or female 129S1/SvIMJ mice or from tissues purchased from Pel-Freez Biologicals and then precipitated with LiCl and treated with DNAse (Epicentre Technologies, Madison, WI). To minimize the possibility of DNA contamination during the course of our experiments, we (*i*) performed multiple rounds of DNase treatments on each RNA sample to remove any residual contaminating DNA [RNA preps were digested twice with RNase-Free DNase I (Epicentre, catalogue no. D9902K) for 15 min at 37°C in 33 mM Tris-acetate (pH 7.8)/66 mM potassium acetate/10 mM magnesium acetate/0.5 mM DTT/2 mM $CaCl_2$]; (*ii*) verified the absence of DNA by performing PCR reactions (see below) with material from RT reactions plus or minus RT and either the same (20 ng) or 10-fold higher (200 ng) amounts of RNA; (*iii*) conducted control RT-PCR reactions on each RNA sample by using primers located on either side of a small intron for several different genes (Nurr1, ApoA1, Blue Opsin, and Herculin) for 40 PCR cycles (three more than usual); and (*iv*) used for our experiments only those RNAs that contained no DNA according to the results of these experiments. cDNA was prepared with 5 g of RNA in 40- l reactions with random primers, 40 units of murine Maloney leukemia virus-RT (Roche), and 20 units of RNAse inhibitor (Roche) and then treated with RNAse H (Epicentre) and RNAse A (Ambion) and normalized with 18S RNA primer sets (Ambion).

PCR was performed with cDNA prepared from 2, 20, or 200 ng of RNA in 25- l reactions with 1.25 units of AmpliTaq Gold Polymerase (Applied Biosystems) and 0.25 M of each primer. PCR conditions were 94ºC, 5 min; 37 cycles of 94ºC/0.5 min, 65ºC/0.5 min, 72ºC/1 min; 72ºC, 7 min. PCR products were electrophoresed in 2% agarose gels and visualized and scanned on an ALPHA IMAGER ALPHA EASE program (Alpha Innotech, San Leandro, CA).

***In Situ* Hybridization.** *In situ* hybridization was performed as previously described (3) with minor modifications. Briefly, 10- to 14- m sections were obtained from brains of adult male 129S1/SvIMJ mice. Sections were collected in series at 500- m intervals through the brain (100- m intervals in hypothalamus and amygdala). Sense and antisense 33P-labeled riboprobes were prepared (Ambion Maxiscript) from cDNA fragments flanked by T3 and T7 RNA polymerase promoter sequences. Sections were prehybridized (3), hybridized (3) overnight at 55°C with riboprobes (5×10^6 cpm/slide), treated with RNAse, washed in SSC (final wash in $0.1 \times$ SSC at 70ºC, 30 min). Slides were then dipped in NTB-2 emulsion and developed after 3 weeks. Each sense probe was hybridized to sections through two brains and each antisense probe to sections through one brain. Sections were counterstained with cresyl violet and examined by light- and dark-field microscopy. Images were captured with a Photometric CoolSnap camera and Universal Imaging METAMORPH software (both from Meridian Instruments, Kent, WA).

1. Tusnady, G. E. & Simon, I. (1998) *J. Mol. Biol.* **283,** 489–506.

2. Page, R. D. (1996) *Comput. Appl. Biosci.* **12,** 357–358.

3. Hohmann, J. G., Jureus, A., Teklemichael, D. N., Matsumoto, A. M., Clifton, D. K. & Steiner, R. A. (2002) *Neuroscience* **117,** 105–117.