The Structural Evolution of a P2Y-like G-protein-coupled Receptor*‡

Angela Schulz and Torsten Schöneberg‡

From the Institute of Pharmacology, Medical Faculty, Free University Berlin, 14195 Berlin, Germany and the Institute of Biochemistry, Department of Molecular Biochemistry, Medical Faculty, University of Leipzig, 04103 Leipzig, Germany

Received for publication, April 1, 2003, and in revised form, June 23, 2003
Published, JBC Papers in Press, June 30, 2003, DOI 10.1074/jbc.M303346200

Based on the now available crystallographic data of the G-protein-coupled receptor (GPCR) prototype rhodopsin, many studies have been undertaken to build or verify models of other GPCRs. Here, we mined evolution as an additional source of structural information that may guide GPCR model generation as well as mutagenesis studies. The sequence information of 61 cloned orthologs of a P2Y-like receptor (GPR34) enabled us to identify motifs and residues that are important for maintaining the receptor function. The sequence data were compared with available sequences of 77 rhodopsin orthologs. Under a negative selection mode, only 17% of amino acid residues were preserved during 450 million years of GPR34 evolution. On the contrary, in rhodopsin evolution ~43% residues were absolutely conserved between fish and mammals. Despite major differences in their structural conservation, a comparison of structural data suggests that the global arrangement of the transmembrane core of GPR34 orthologs is similar to rhodopsin. The evolutionary approach was further applied to functionally analyze the relevance of common scaffold residues and motifs found in most of the rhodopsin-like GPCRs. Our analysis indicates that, in contrast to other GPCRs, maintaining the unique function of rhodopsin requires a more stringent network of relevant intramolecular constraints.

Among the different families of transmembrane receptors, G-protein-coupled receptors (GPCRs) form the largest superfamily. Molecular cloning studies and genome data analyses have revealed ~1200–1300 members of the GPCR superfamily in mammalian genomes (1). To predict and understand ligand binding and signal transduction as well as the consequences of disease-causing mutations) within a receptor molecule, detailed information about the native receptor structure in its inactive and active conformations is required. Currently, a high-resolution structure is available only for bovine rhodopsin (2), which now provides the basis to generate models of other GPCRs (3). However, recent studies point out that even with a crystal structure in hand, construction of reliable receptor models still requires time-consuming refinements based on data from mutagenesis, cross-linking, and NMR studies (4).

Herein, we mined evolution as an additional source of structural information that may direct GPCR modeling and mutagenesis studies. This idea has been successfully applied in an early stage of GPCR structure/function analysis. The sequences of >200 different members of the GPCR family were used to predict the approximate arrangements of the seven transmembrane helices (5). To determine more distinct structural determinants that participate in ligand recognition and signal transduction, sequence analysis has to be focused on a single receptor subtype. The structural diversity of a given receptor among different species is the result of a long evolutionary process characterized by a continuous accumulation of mutations. However, the maintenance of vital functions in an organism strictly requires enough structural conservation to ensure the functionality of the receptor protein. Studying the structural diversity of a single GPCR will help to determine evolutionary preserved elements and may disclose the “spatial freedom” of each amino acid position.

For a valid evolutionary analysis, the chosen GPCR should meet the following requirements. First, the GPCR should be “evolutionarily old” to allow for a broad structural variety during evolution. Second, GPCRs with structurally indistinguishable subtypes or pseudogenes should be avoided, because the interpretation of data will be more difficult. Third, the physiological agonist should not be a peptide or a protein because of interpretation of data will be more difficult. Fourth, the coding region of the GPCR gene should contain no or only small introns to allow for an easy amplification from genomic DNA. Among the GPCRs of family 1, only a few receptors meet these requirements. The group of ADP receptor-like GPCRs includes P2Y1, P2Y13, GPR105, GPR87, and GPR34. In an initial study we have shown that GPR34 is an evolutionarily old single-copy gene and has clear structural and functional features that allows its differentiation from other members of the group of ADP-like GPCRs (6). Therefore, GPR34 was chosen to identify structural and functional determinants by an evolutionary approach. Our analysis includes 61 full-length or partial sequences of GPR34 and GPR34-like GPCRs that were cloned from 52 species, including mammals, birds, reptiles, amphibians, teleost fish, and sharks. This large sequence of data enabled us to identify motifs and residues that are important for maintaining the receptor function. The sequence data were compared with sequences of 77 rhodopsin orthologs. Despite the fact that GPR34 orthologs and vertebrate rhodop...
GPR34 Evolution

dopins share less than 7% identical amino acid residues, the gross global structure appears to have been maintained between both receptor groups as judged by analyzing the periodicity of conserved residues and hydrophobicity. During 450 million years of evolution, only 17% of all amino acid residues remained unchanged in GPR34 orthologs. In contrast, in rhodopsin evolution ~43% of residues were absolutely conserved in fish and mammalian orthologs. This indicates that maintaining the unique function of rhodopsin requires a more stringent network of relevant determinants.

EXPERIMENTAL PROCEDURES

Cloning of GPR34-like and ADP-like Receptor Orthologs and Generation of GPR34 Mutants—To identify GPR34 sequence in other vertebrates, genomic DNA samples were prepared from tissue or peripheral mononuclear blood cells of various species or were kindly provided by several other labs (supplemental Table S1, available in the online version of this article). Tissue samples were digested in lyssin buffer (50 mM Tris/HCl, 100 mM EDTA, 100 mM NaCl, 1% SDS, and 0.5 mg/ml proteinase K) and incubated at 55 °C for 18 h. DNA was purified by phenol/chloroform extraction and ethanol precipitation. Based on the sequence of the human, mouse, and carp GPR34 sequences (6), sets of degenerated primer pairs (supplemental Table S3, available in the online version of this article) were tested to evaluate their ability to amplify GPR34-specific sequences. Standard PCR reaction was performed with Taq polymerase under the following conditions (35 cycles): 1 min at 94 °C; 1 min at 56–62 °C; and 2 min at 72 °C. Specific fragments were directly sequenced and/or subcloned into the pCR2.1-TOPO vector (Invitrogen), and at least three different clones were sequenced.

For identification of complete GPR34 open reading frames, a 5′- and 3′-rapid amplification of cDNA ends (RACE) strategy was used. Thus, 1 μg of genomic DNA per reaction was digested with various blunt end cutting enzymes (EcoRV, HincII, MscI, and DraI; New England Biolabs). Then, an ampidated oligonucleotide adapter (Marathon cDNA amplification Kit, Clontech) was linked to the digested DNA. 5′- and 3′RACE PCR reactions were performed with the AP1 primer (Clontech) and species-specific primers using the tagged genomic DNA as a template in the Expand™ high fidelity PCR system (Roche Applied Science). PCR reactions were carried out under the following conditions (35 cycles): 1 min at 94 °C; 1 min at 62 °C; and 3 min at 68 °C. After treatment with Taq polymerase for 10 min at 72 °C, PCR products were subcloned into the pCR2.1-TOPO vector and sequenced with ThermoSequenase and dye-labeled terminator chemistry by an automated sequencer (Applied Biosystems, Foster City, CA).

Comparison of the 5′-untranslated region of human and rodent GPR34 genes revealed high sequence similarity, allowing the use of a primer derived from the sequence upstream of the coding sequence for orthologs and species from mammalian genomic DNA (see supplemental Table S3, available in the online version of this article). Thus, sequences encoding the N termini of GPR34 orthologs were identified from several mammals.

Full-length GPR34 sequences were inserted into the mammalian expression vector pDps. GPR34 mutations (N137A and A265Y) were introduced into the hemagglutinin-tagged version of the human GPR34 expression vector pcDps. GPR34 mutations (N137A and A265Y) were successfully in cloning GPR34 orthologs from sharks, which are also successful in cloning GPR34 orthologs from sharks, which are commonly ancestors of ray-finned fish and tetrapods. This indicates that GPR34 may have existed for at least 450 million years. All attempts to identify GPR34-like sequences in primitive chordate species such as lamprey and hagfish have failed so far, although primer pairs yielded partial sequences of other related GPCR data (not shown). Similarly, sequence analyses of the known genomes of Caenorhabditis elegans, Drosophila melanogaster, and Anopheles gambiae (NCBI data base) as well as cloning attempts from other insects (Spodoptera frugiperda and Melolontha melolontha) revealed no GPR34-like sequences.

Our phylogenetic analysis revealed that, in all tetrapod species investigated, GPR34 coding regions are obviously intronless and lack any subtypes in their genomes. This picture completely changed when phylogenetic analysis was extended to teleost fish. First, up to three structurally distinct GPR34 subtypes were identified in various fish species (e.g. carp) (Fig. 2A). Second, several fish species (e.g. tilapia) present a single
intron in phase one (inserted after codon position one) within the coding region of EL2, which was identified by sequence comparison with other orthologs and intron/exon prediction tools. Sequence comparison, duplicated genes, and appearance/loss of introns provide a valid basis for unveiling the evolutionary history of a protein. Based on our genomic and amino acid sequence analyses, at least two subgroups of GPR34 sequences can be distinguished already in evolutionarily old teleost fish (Anguillidae and Cyprinidae). The first intronless receptor subgroup, which includes e.g. tilapia type 1 and GPR34 receptors from fugu and tetraodon (see Fig. 2A), displays some unique structural features such as a conserved Trp residue at the C-terminal end of IL2 (relative position 4.43) and a second Pro residue in TMD6 (relative position 6.56) (supplemental Fig. S2, available in the on-line version of this article). The second subgroup clusters intron-containing GPR34 genes. Some species present more than one ortholog of the second subgroup (cod and carp). This receptor divergence may be the result of an additional gene or even genome duplication. However, the proposed sub-grouping is probably restricted to the evolutionarily younger Acanthopteryii (e.g. fugu and tilapia) and cannot be applied to more ancient orders such as Anguillidae (eel), Cyprinidae (carp and zebrafish), and Siluridae (catfish). It is interesting to note that in some Acanthopteryii GPR34 is only a single-copy gene (fugu and tetraodon) as judged by PCR and data base search. Tetraodontidae are phylogenetically young, and it is therefore reasonable to assume that the second gene was lost again during evolution. Similarly, all attempts to amplify a second GPR34 subtype from two cartilaginous fish and two sturgeon species failed, indicating the existence of only one GPR34 ortholog at the very beginning of fish evolution. Low stringency genomic PCR yielded only a new and structurally distinguishable intermediate group of GPR34-like/ADP-like receptors (see Fig. 1).

Another interesting marker of GPR34 evolution in fish is the presence of an intron in some orthologs (subgroup 2 in Fig. 2A). The consensus sequence flanking the introns is highly preserved (Fig. 2B). The intron size between species and receptor subtypes varies only marginally. The relative position within

Fig. 1. GPR34 receptors have existed for more than 450 million years. Based on the known GPR34 sequences (6), degenerated primers were designed and used to amplify partial GPR34 sequences from various vertebrate species, including mammals, marsupials, monotremes, birds, reptiles, amphibians, and teleost and cartilage fish (sequences were deposited in GenBank™, see supplemental Table S1 available in the on-line version of this article). Because several ortholog sequences were obtained with primer pairs yielding a GPR34 sequence only between TMD3 and TMD7, phylogenetic analyses were performed on the basis of amino acid sequences between the relative codon positions 3.52 to 7.48 (32). The amino acid sequences of selected GPR34 receptors (for accession numbers see supplemental Table S1 available in the on-line version of this article) and ADP-like receptors such as the human, murine, and fugu P2Y, human and fugu P2Y, human GPR87, and human and murine GPR105 were aligned. The amino acid sequence alignment was made with Clustal X (11), and the PAM matrices were implemented in the software package. Only the sequences depicted here were used to generate the tree. Based on the multiple alignment, a neighbor-joining tree was constructed, and bootstrap values are the result of 1,000 neighbor-joining replicates. Only relevant bootstrap values (>750) that statistically support clades are shown. To verify the structural relation between the GPR34 orthologs and closely related GPCR s, the same data set was analyzed with the PHYLIP software package (12) and yielded essentially similar results (data not shown). The phylogenetic analysis shown here is representative in respect to GPCR group formation because essentially similar results were obtained with the complete set of GPR34 sequence cloned or full-length receptors (supplemental Fig. S1, available in the on-line version of this article). Accession numbers of sequences taken from GenBank™ are as follows: human P2Y (AF013449; murine P2Y, (NM_027571; fugu P2Y (CAAB01004670); human P2Y (NM_023915); human GPR105 (NM_014879); murine P2Y (XM_130886); human platelet-activating receptor (D10202); and bovine rhodopsin (P02699).
the coding sequence is conserved, and almost identical splice acceptor sites indicate a common evolutionary origin. No other obvious introns were found within the coding region between TMD1 and TMD7 as analyzed in cod GPR34 receptors. Gain and loss of spliceosomal introns in a lineage is a unique event that occurs at a specific point in its evolution. Two hypotheses to explain the origin of spliceosomal introns could be proposed. The "intron early" hypothesis would suggest that introns were ancient and were lost during evolution, whereas the "intron late" hypothesis would suggest that introns were inserted into the gene later in evolution (17). The obvious absence of introns in GPR34 genes in cartilaginous fish and evolutionarily old teleost fish (eel, carp, and zebrafish) and a relative uniform intron size suggest that the intron was an evolutionarily new invention after gene duplication. Additional support for an intron late introduction in GPR34 genes comes from the ADP receptor evolution. Some ADP receptors (fugu and tetraodon P2Y₁₂) also contain a spliceosomal intron. However, the position within the coding region and the size are different, excluding a common ancestral origin of the introns found in GPR34 and ADP-like receptors.

Taken together, the evolution of GPR34 probably started in the Cambrian period more than 450 million years ago as suggested by its presence in cartilaginous and ray-finned fish genomes. The evolutionary split between GPR34 and ADP-like receptors is already found in teleost fish, indicating their coexistence for at least 450 million years. Identification of a clear, distinguishable ADP-like receptor from cartilaginous fish may set this evolutionary point even earlier. Furthermore, our data suggest that a GPR34 gene duplication took place in early teleost fish evolution. Then, a single intron was introduced in one of the two GPR34 genes before the divergence of Acan-
thopteryii. GPR34 diversification was further increased by gene or genome duplication in some teleost fish species.

**GPR34 and ADP-like Receptors Have a Common Evolutionary Origin**—While this study was being conducted, P2Y12 and P2Y13 receptors were cloned, and ADP was found to be the agonist for both receptors (18, 19). A high overall amino acid sequence homology already implicated a common ancestral origin of GPR34- and ADP-like receptors (20, 21). Further evidence for a close structural relation between GPR34- and ADP-like receptors came from our genomic PCR experiments. Using the same sets of degenerate primers, we amplified several other GPCR fragments from fish genomic DNA (carp, salmon, pike, zebrafish, sturgeon, dog, and shark), which are closely related to both the GPR34- and the ADP-like receptors. Comparative structural analyses revealed that two carp receptors (carp ADP-like type 1 and type 2; see Fig. 1) and one receptor from salmon and pike (not shown) were more closely related to ADP-like receptors, whereas zebrafish GPR34-like, sturgeon GPR34-like, and dog shark GPR34-like receptors combine structural features of GPR34 and ADP-like receptors, thus forming an intermediate subgroup (see Fig. 1).

To support a common evolutionary origin of GPR34- and ADP-like receptors by functional data, we initiated studies to determine the signal transduction and agonist specificity of the GPR34 receptors. P2Y12 and P2Y13 are coupled to $G_{q/11}$ proteins, which inhibit adenyl cyclase and, therefore, decrease intracellular cAMP levels. According to the current model of GPCR function (22), receptor overexpression can result in a constitutive activation of signaling pathways. Thus, the coupling abilities of several receptors, including “orphan” receptors, have been characterized by overexpression in the absence of an agonist. For example, the wild-type ACCA, an orphan GPCR, has been characterized by overexpression in the absence of an agonist (100%).

$\text{GPR34 Evolution}$

The specificity and sensitivity of chimeric proteins such as $G_{\alpha q}$ can be improved by further modifications (lack of the N-terminal extension and introduction of an N-terminal consensus site for myristoylation), and this modified protein will be referred to henceforth as $G_{\alpha q 4}$ (25). As shown in Fig. 3B, the basal activities of the human P2Y12 and the human GPR34 were increased 6.8- and 3.5-fold, respectively, when compared with GFP-transfected cells. To control whether the high basal activity of the human GPR34 is specific, two mutants (N137A and A265Y) were generated and co-transfected with $G_{\alpha q 4}$. Asn-137 (relative position 3.35) and Ala-265 (relative position 6.34) are fully preserved in GPR34 evolution, and we therefore speculated that mutation of these residues may have an effect on receptor function. Replacement of Asn-137 by an Ala residue resulted in a further increase of basal receptor activity (see Fig. 3B). This indicates that Asn-137 participates in maintaining the ground state of receptor activity. Our results are in concert with previous studies showing that mutation of Asn (relative position 3.35) in Go$\alpha_q$-coupled GPCRs such as the platelet-activating factor (PAF) receptor, the bradykinin B2 receptor, and the AT1 angiotensin II receptor induces constitutive activity (26, 27).

In contrast, mutation of Ala-265 to Tyr almost completely abolished the basal activity of the human GPR34. Plasma membrane expression of A265Y was reduced (59 ± 14%) when compared with those of the wild type GPR34 (100%) and N137A (104 ± 6%) as determined by an indirect cellular enzyme-linked immunosorbent assay. These data indicate that the increased constitutive activity of N137A is indeed caused by disrupting a functionally important constrain and not by an elevation of the receptor number at the cell surface, whereas the loss of basal activity of A265Y is mainly caused by reduction of cell surface expression. It is of interest that Ala-265 (relative position 6.34) is conserved not only in GPR34 receptors but also in many amnergic receptors and glycoprotein hormone receptors, highlighting the general functional importance of this position in GPCRs. It has been proposed (28) and finally proven in the rhodopsin crystal structure that the amino acid residue 6.34 in the TMD6/IL3 transition faces the residue 5.61 in TMD5. Mutagenesis data from the $\alpha_q$-adrenergic receptor and the thyrotropin (TSH) receptor are consistent with this model, showing that introduction of bulky residues at position 6.34 leads to a constitutive receptor activation probably by disturbing the tight packing between TMD5 and TMD6 (28, 29). Although the residues at positions 5.61 and 6.34 in GPR34 receptors, glycoprotein hormone receptors, and many amnergic receptors are identical, the consequence of mutating position 6.34 is different. This finding may indicate structural differences between GPR34 and other family 1 GPCRs. Support for this hypothesis comes from conservation and hydrophobicity analyses that point at differences in the helical character of TMD5 and the relative orientation of TMD6 within the lipid bilayer (see below).

Because ADP had no agonistic effect on the orphan GPR34 receptors, we focused on the high basal activity as a functional parameter to study whether this property is preserved during GPR34 evolution. Thus, mouse, chicken, fugu, two zebrafish, and all three carp GPR34 subtypes were co-transfected with $G_{\alpha q 4}$ and found to be constitutively active (see Fig. 3B). Similarly, the zebrafish GPR34-like receptor, a member of the intermediate receptor group (see Fig. 1), also displayed an elevated basal activity following co-transfection with $G_{\alpha q 4}$. Our data clearly indicate that GPR34-like and ADP-like receptors share $G_{\alpha q}$ coupling and a high basal activity but can be distinguished by their ability to be further activated by ADP. It should be noted, however, that the high basal activity of GPR34-like receptors may be caused by an endogenous agonist that is produced in COS-7 cells and released into the assay medium. Until the endogenous agonist is identified and antagonists are developed, the final outcome of this question remains open.

Based on our phylogenetic and functional data, GPR34-like and ADP-like receptors probably arose from a common GPCR ancestor. After splitting, agonist specificity became different in both groups. Both receptor groups were maintained during evolution, indicating their physiological importance. Furthermore, our data clearly highlight the potential of dissecting...
functionally relevant residues from phylogenetic studies as shown for Asn-137 and Ala-265.

**GPR34 Receptors Show a Purifying Selection Mode**—In the next step, the large set of sequence data from 450 million years of evolution was used to analyze the global diversity of the receptor structure and the mode of gene diversification. As summarized in Table I, GPR34 sequences differ reasonably in their total length, ranging from 369 to 389 amino acids. The overall identity between seven mammalian and seven fish full-length GPR34 orthologs is only 38.2 ± 1.5%. The N and C termini as well as EL2 and IL2 contribute the most to the observed size and sequence differences. For example, major variations in the length of N termini, ranging from 19 to 62 amino acid residues, were already found in mammals (Table I). The obvious lack of conserved determinants and large length differences suggest that the N terminus does not directly participate in GPR34 ligand binding. Similarly, size variations were identified in all loops except IL3 (Table I). These findings are of interest, because evolutionary comparison of mammalian and fish rhodopsin orthologs revealed an overall identity of 77.4 ± 2.2%, and no size differences are present between TMD1 and TMD7 in all rhodopsin sequences analyzed (Table I). Rhodopsin orthologs show no differences in the conservation when loops (mean ~83%) and TMD (mean ~80%) regions were compared between fish and mammalian orthologs. In contrast, GPR34 receptors are significantly less conserved in the loops

Fig. 3. **High basal activity is preserved among GPR34 receptors.** To evaluate signaling specificity of GPR34 and ADP receptors, COS-7 cells were co-transfected with the indicated receptor construct and the chimeric Go protein (see "Experimental Procedures"). IP assays were performed 72 h after transfection as described under "Experimental Procedures." A, plasmids encoding GFP (control), the human GPR34, and the human ADP receptor were co-transfected with Go alpha. IP formation under basal conditions (light gray bars) and in the presence of 10 µM ADP (dark gray bars) was determined. Data are presented as means ± S.E. (cpm/well) of three independent experiments, each carried out in triplicate. B, to verify the constitutive activity of GPR34 and ADP receptors, COS-7 cells were co-transfected with the designated GPR34 ortholog constructs and the chimeric Go alpha plasmid (see "Experimental Procedures"). Basal IP formation is expressed as fold over basal levels of GFP-transfected cells (404 ± 167 cpm/well). Data are presented as means ± S.E. of three independent experiments, each carried out in triplicate.
The sequence information of 61 full-length or partially identified GPR34 orthologs was used to determine global and more distinct structural parameters of the GPR34 receptor subgroup. These data were compared with the sequence information of 77 rhodopsin orthologs (see supplemental Tables S1 and S2 for GenBank™ accession numbers). The second and fourth columns from the left compare segment lengths (minimum/maximum) between GPR34 and rhodopsin, respectively. Results of evolutionary conservation analyses between fish and mammalian orthologs are shown in the third and fifth columns from the left. Numbers given in parentheses indicate the number of orthologs used for analysis. Data are given as means ± S.D.

<table>
<thead>
<tr>
<th>Domain</th>
<th>Length of GPR34 receptors (min/max)</th>
<th>Identity of mammalian vs. fish orthologs %</th>
<th>Length of rhodopsin (min/max)</th>
<th>Identity of mammalian vs. fish orthologs %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full-length receptor</td>
<td>369/389 (15)</td>
<td></td>
<td>348/354 (77)</td>
<td></td>
</tr>
<tr>
<td>N terminus</td>
<td>19/62 (20)</td>
<td></td>
<td>34 (77)</td>
<td></td>
</tr>
<tr>
<td>C terminus</td>
<td>46/80 (15)</td>
<td>38.2 ± 1.5 (7.7)</td>
<td>39/46 (77)</td>
<td>77.4 ± 2.2 (18.47)</td>
</tr>
<tr>
<td>IL1</td>
<td>8/9</td>
<td>50.9 ± 11.7 (21.9)</td>
<td>6</td>
<td>83.5 ± 12.9 (18.47)</td>
</tr>
<tr>
<td>IL2</td>
<td>17/25</td>
<td>56.5 ± 13.2 (20.9)</td>
<td>6</td>
<td>98.2 ± 6.2 (18.47)</td>
</tr>
<tr>
<td>IL3</td>
<td>15/16</td>
<td>58.6 ± 9.5 (25.23)</td>
<td>8</td>
<td>82.0 ± 8.9 (18.47)</td>
</tr>
<tr>
<td>TMD1</td>
<td>29</td>
<td>58.6 ± 3.0 (19.7)</td>
<td>30</td>
<td>71.8 ± 5.6 (18.47)</td>
</tr>
<tr>
<td>TMD2</td>
<td>23</td>
<td>56.5 ± 5.2 (21.9)</td>
<td>30</td>
<td>88.1 ± 5.6 (18.47)</td>
</tr>
<tr>
<td>TMD3</td>
<td>30</td>
<td>71.1 ± 3.2 (21.9)</td>
<td>29</td>
<td>87.0 ± 4.4 (18.47)</td>
</tr>
<tr>
<td>TMD4</td>
<td>23</td>
<td>32.6 ± 6.7 (25.23)</td>
<td>22</td>
<td>78.6 ± 7.5 (18.47)</td>
</tr>
<tr>
<td>TMD5</td>
<td>24</td>
<td>45.3 ± 5.4 (25.23)</td>
<td>24</td>
<td>79.0 ± 6.5 (18.47)</td>
</tr>
<tr>
<td>TMD6</td>
<td>26</td>
<td>62.9 ± 6.4 (25.23)</td>
<td>28</td>
<td>80.4 ± 5.5 (18.47)</td>
</tr>
<tr>
<td>TMD7</td>
<td>24</td>
<td>75.6 ± 4.6 (12.7)</td>
<td>24</td>
<td>76.3 ± 5.8 (18.47)</td>
</tr>
</tbody>
</table>

(mean = -36%) compared with TMDs (mean = -56%). In rhodopsin, IL1 (72%) and TMD1 (68%) display the lowest conservation. In GPR34 orthologs little conservation is found in EL2 (29%), EL3 (27%), IL2 (25%), and TMD4 (33%). It is of interest that the TMD3 and TMD7 of GPR34 receptors are as highly preserved as those found in rhodopsin orthologs.

Positive Darwinian selection may contribute to the observed gene diversification. Therefore, GPR34 sequences were analyzed for positive selection by comparing non-synonymous (dN) and synonymous (dS) substitution rates in the coding region (15). A ratio > 1 indicates a positive Darwinian selection, and a ratio < 1 is indicative of a negative or purifying selection mode.

For analysis, the sequences encoding the receptor proteins from IL2 to the D(N)P motif were used to include also partial sequences of GPR34 receptors in the analysis. In most cases, dN/dS ratios were < 1, indicating a purifying selection mode of GPR34 receptors and rhodopsin (Fig. 4A). However, dN/dS ratios for GPR34 genes (0.221 ± 0.087) were greater than dN/dS ratios of rhodopsin-like sequences (0.068 ± 0.032). To evaluate the significance of this interesting finding, analyses were repeated with sequences from 13 vertebrate species for which sequences (TMDs 1–7) were available from both GPR34 receptors and rhodopsin. As shown in Fig. 4B, essentially the same results were obtained when dN/dS ratios were compared between GPR34 (0.238 ± 0.104) and rhodopsin (0.035 ± 0.013) full-length sequences. The dN/dS ratio difference was highly significant (p < 0.0001, paired two-site t test) and reflects a higher gene diversification in GPR34 orthologs at the genomic level.

In a more detailed analysis, dN/dS ratios for GPR34 genes from fish (n = 24) were compared with those obtained in mammals (n = 25). Analysis revealed significantly (p < 0.001) higher dN/dS ratios within fish GPR34 genes (0.197 ± 0.075) when compared with mammalian GPR34 sequences (0.132 ± 0.044). These results provide additional support for subgroup formation in fish, but negative (purifying) selection is, as in all the other GPR34 sequences tested herein, the mode of evolution.

In sum, most dN/dS ratios were < 1, indicating a purifying selection mode of GPR34 receptors and rhodopsin. In consensus with the finding at the amino acid level, dN/dS ratios for GPR34 genes were significantly greater than the dN/dS ratios
of rhodopsin sequences. It should be noted that the whole range of receptor activity depends upon a global structure of seven TMDs and a spatially defined arrangement of constraints conserved among most GPCRs of family 1. Therefore, a reasonable number of amino acid residues have only a limited evolutionary freedom. Even minor changes can be an evolutionary invention in a receptor’s specificity and signal transduction that is not reflected in the dN/dS ratio.

The Global Architecture of Rhodopsin Is Also Preserved in GPR34—The sequence information of 61 full-length or partially identified GPR34 orthologs enabled us to determine global and more distinct structural parameters of the GPR34 receptor subgroup that are relevant for maintaining the receptor structure and its agonist binding and signal transduction abilities. These data were compared with the sequence information of 77 rhodopsin orthologs together with bovine rhodopsin crystal structure.

First, the hydrophobicity pattern of all putative TMDs of GPR34 orthologs were compared with those of vertebrate rhodopsin orthologs. Despite a low sequence homology, the hydrophobicity pattern of most TMDs showed only minor differences between GPR34 receptors on the basis of the rhodopsin molecule, crystal structure data of the bovine rhodopsin were simply taken as template. The homology-modelling algorithms and refinement procedure implemented in a Deep View Swiss-PDB Viewer 3.7 was used to generate a raw model of the human GPR34. Only the seven TMDs and the residues (red) that are >80% conserved are shown. For clarity, the model was generated to visualize only the relative position of conserved GPR34 residues within the rhodopsin crystal structure. D, for comparison, the structure of the corresponding TMD segments of rhodopsin is presented (2). Only the seven TMDs and the residues (red) that are >90% conserved are shown. The analyses of all other TMD segments are presented in supplemental Fig. S3 (available in the on-line version of this article).

Fig. 5. Structural comparison of the transmembrane domains of GPR34 and rhodopsin orthologs. The structural conservation (A) and hydrophobicity (B) of GPR34 orthologs (red line) and rhodopsin orthologs (black line, gray area) were determined for TMD3, TMD5, and TMD6 as described under “Experimental Procedures.” A family-based numbering scheme was utilized for alignment of GPR34 and rhodopsin data sets (supplemental Table S2, available in the on-line version of this article), preceded by the TMD number and followed by its position within the TMD segment relative to the most conserved and recognizable residue within each TMD, which is denoted 1.50, 2.50, etc. (32). Because of a non-Gaussian distribution of the data, mean values with the 95:5% confidential intervals are shown in hydrophobicity plots. The data are based on 33 (TMD3) and 59 (TMD5 and TMD6) GPR34 and 77 rhodopsin amino acid sequences. C, to evaluate the position of amino acid residues conserved in GPR34 receptors on the basis of the rhodopsin molecule, crystal structure data of the bovine rhodopsin were simply taken as template. The homology-modelling algorithms and refinement procedure implemented in a Deep View Swiss-PDB Viewer 3.7 was used to generate a raw model of the human GPR34. Only the seven TMDs and the residues (red) that are >80% conserved are shown. For clarity, the model was generated to visualize only the relative position of conserved GPR34 residues within the rhodopsin crystal structure. D, for comparison, the structure of the corresponding TMD segments of rhodopsin is presented (2). Only the seven TMDs and the residues (red) that are >90% conserved are shown. The analyses of all other TMD segments are presented in supplemental Fig. S3 (available in the on-line version of this article).
on the rhodopsin structure (see Fig. 5, bundle (see Fig. 5)). Most residues that are less conserved are up on the helix side, which faces the inner cavity of the helical structure. Indeed, residues found to be highly conserved in rhodopsin line up on the helix side, which faces the inner cavity of the helical bundle (see Fig. 5D). Most residues that are less conserved are oriented toward the lipid bilayer in the rhodopsin crystal structure. Interestingly, an almost identical periodic conservation in position and extent is found for TMD6 of GPR34 receptors. Similarly to rhodopsin, the conserved residues pointed to the inner receptor space when the human GPR34 sequence was projected on the rhodopsin structure (see Fig. 5, C and D). Despite significant differences in the hydrophobicity and residue conservation patterns in TMD3, positioning of the highly conserved residues is comparable with the rhodopsin structure (see Fig. 5, C and D). Conservation of residues in TMD5 display, as in TMD6, a strong periodicity, and the hydrophobicity plot almost exactly matches the one of the TMD5 of rhodopsin. However, projection of conserved residues into the three-dimensional model of the rhodopsin crystal structure revealed remarkable differences between GPR34 and rhodopsin. All conserved amino acid residues of the N-terminal half of TMD5 face the lipid bilayer (see Fig. 5, C and D). Careful inspection of the periodicity pattern and its alignment with rhodopsin already indicates a shift in the maxima upstream of the relative position 5.50. In the rhodopsin crystal structure a conserved proline (relative position 5.50) disturbs the α-helical character of TMD5. A proline at this position is missing in all GPR34 receptors. It is therefore reasonable to assume that TMD5 in GPR34 is strictly α-helical, and the conserved residues are probably more inwardly oriented.

**Prediction of Functional Relevant Determinants in GPR34**—In a final step, we focused our sequence analysis on more distinct structural determinants and motifs. During 450 million years of evolution, only 66 (~17%) of all amino acid residues remained unchanged, and 43 (~11%) of the additional positions displayed variations between two amino acid residues (Fig. 6A). In contrast, in rhodopsin evolution ~150 (~43%) residues were absolutely preserved. Raising the number of ortholog sequences also increases the significance of structural and functional predictions made for each amino acid residue. We addressed this point by providing some examples (DRY, disulfide bonds, and structural determinants in the N- and C termini) in the GPR34 receptor protein that may be relevant also for other GPCRs.

In addition to the well preserved cysteine residues that form a disulfide bond between the EL1 and EL2, GPR34 receptors contain two conserved cysteine residues, i.e. one in the N terminus close to TMD1 (see Fig. 6B) and one in the middle of the EL3. Many GPCRs, including receptors for biogenic amines and peptides, harbor this second conserved pair of extracellular cysteine residues bridging the N terminus and EL3 via a disulfide bond. Mutational disruption of this disulfide bond results in a loss of high affinity binding of receptor ligands, suggesting a pivotal role of an N terminus/EL3-connecting disulfide bridge for proper receptor assembly (1). Therefore, it is likely that an N terminus/EL3-connecting disulfide bridge may also provide a guarantee for proper receptor folding. However, in platypus and red perch GPR34 type 2 receptors the Cys residue in EL3 is missing, indicating that the structural function of this second disulfide bond is dispensable or somehow structurally compensated in GPR34 type 1 of these fish species (see supplemental Fig. S2, available in the on-line version of this article). Support for the latter hypothesis comes again from the crystal structure of rhodopsin. The N-terminal segment of rhodopsin is located just above EL3. Specific non-covalent contacts maintain the proper orientation between the rhodopsin N terminus and the extracellular loops so that an additional disulfide bridge, as in other GPCRs, is probably not required.

In addition to common scaffold residues of rhodopsin-like GPCRs, the pattern of strictly conserved residues and the presence of several structural features allow the identification of GPR34 receptors among other members of family 1 GPCRs. For example, the DRY motif located at the TMD3/IL2 transition is a highly conserved triplet of amino acid residues known to play an essential role in GPCR function. The crystal structure of rhodopsin indicates that the Glu residue in the D(F/Y)RY motif forms a salt bridge to the Arg and probably acts as a proton acceptor during receptor activation (30). However, in GPR34 receptors from fugu and tetraodon the acidic Asp residue within this motif is naturally substituted by a His residue without affecting the G-protein coupling of GPR34 (shown for fugu GPR34 in Fig. 3B). The general importance of a salt bridge within the DRY motif is therefore challenged, at least for GPR34 signaling.

The polypeptide chain of most GPCRs is posttranslationally modified, including by N-glycosylation, palmitoylation, and phosphorylation. Potential N-glycosylation sites (NX(S/T)) are usually located within the extracellular N-terminal region but were also found in the extracellular loops. In general, the number and relative positions of potential N-glycosylation sites are not conserved among GPR34 orthologs of different species (see Fig. 6B). Most family 1 GPCRs, including rhodopsin, present a fourth intracellular loop or, better, an α-helix. Here, the C terminus is anchored in the plasma membrane via palmitoylated cysteine residues, which are often found in GPCR C-terminal tails. The importance of palmitoylation or isoprenylation for proper GPR34 function appears to be low, because GPR34 orthologs from platypus, carp type 1, tetraodon, and fugu completely lack cysteine residues in their C termini (see Fig. 6C).

Consensus sites for phosphorylation are usually present in IL3 and the C-terminal domain of GPCRs. Many of the identified GPR34 receptors, including evolutionarily old fish species contain an (E/S)T(S/T)EX(K/R) motif in their C termini (see Fig. 6C). This motif has significant similarities to the consensus sequences of casein kinase substrates (31). An almost similar consensus sequence is found in several other proteins such as the poly-Ig-receptor (ESTTETK), GFAP-1 (ESTTERK), and the dynnein heavy chain (ESTTDWK). As shown above, fish GPR34 receptors can be subdivided in at least two subgroups. Although there is no difference in G-protein coupling (as judged by their basal activity) the (E/S)T(S/T)EX(K/R) motif is missing or only partially preserved in fish type 2 GPR34 receptors (see Fig. 6C). This finding probably indicates differences between both fish GPR34 subgroups in their regulation by kinases.

**Conclusion**—We have demonstrated by analyzing vertebrate rhodopsin orthologs that evolutionary data of the TMD core (conservation and hydrophobicity) are nicely reflected in the rhodopsin x-ray structure. Comparison and projection of the structural data from an evolutionary analysis of GPR34 or-
Fig. 6. Structural conservation of GPR34 orthologs during evolution. A, based on 15 full-length and 44 partial (TMD1 or TMD3 to TMD7) amino acid sequences, conservation of each residue was determined. All amino acid residues fully conserved during vertebrate evolution are highlighted (dark background with white lettering), and amino acid residues that vary only between two residues are shown in gray. B, the N termini of GPR34 orthologs show a high degree of variability in length and fine structure. Only TMD1 is aligned. C, the C termini of GPR34 orthologs are aligned, and a putative phosphorylation site ((E/S)ST-(S/T)E(X)-K/R) that is preserved during evolution is boxed.
Acknowledgments—The Ga\textsubscript{q}	extsubscript{64myr} and Ga\textsubscript{q1y} constructs were kindly provided by Dr. E. Kostenis. We thank Çiğdem Çetindag and Rita Oeltjen for excellent technical assistance and Klaas Zierau for suggestions and critical reading of the knowledge. We are grateful to Drs. J. Wess, W. Goddard III, and an anonymous reviewer for suggestions and critical reading of the manuscript.

REFERENCES