Bidirectional, Iterative Approach to the Structural Delineation of the Functional "Chemoprint" in GPR40 for Agonist Recognition

Irina G. Tikhonova,[†,‡] Chi Shing Sum,[§,||] Susanne Neumann,[§] Craig J. Thomas,[⊥] Bruce M. Raaka,[§] Stefano Costanzi,*,[†] and Marvin C. Gershengorn*,[§]

Laboratory of Biological Modeling, Clinical Endocrinology Branch, Chemical Biology Core Facility, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892

GPR40, free fatty acid receptor 1 (FFAR1), is a member of the GPCR superfamily and a possible target for the treatment of type 2 diabetes. In this work, we conducted a bidirectional iterative investigation, including computational modeling and site-directed mutagenesis, aimed at delineating amino acid residues forming the functional "chemoprint" of GPR40 for agonist recognition. The computational and experimental studies revolved around the recognition of the potent synthetic agonist GW9508. Our experimentally supported model suggested that H137(4.56), R183(5.39), N244(6.55), and R258(7.35) are directly involved in interactions with the ligand. We have proposed a polarized NH−π interaction between H137(4.56) and GW9508 as one of the contributing forces leading to the high potency of GW9508. The modeling approach presented in this work provides a general strategy for the exploration of receptor−ligand interactions in G-protein coupled receptors beginning prior to acquisition of experimental data.

Introduction

G-protein coupled receptors (GPCRs) constitute a superfamily of membrane proteins characterized by a common seven transmembrane helical bundle. GPCR signaling is involved in countless physiological processes and, as a result, GPCRs are the most abundantly targeted biological macromolecules of currently marketed drugs. With the continued advances in pharmacology, structural biology, and molecular modeling, efforts directed toward the investigation of the structure and function of GPCRs[1−4] have been increasingly prevalent. The overarching aims of these studies are the understanding of the structure−function relationships of the receptors and the rational design of new chemicals able to regulate their activities. Such studies have led to identification of potent ligands for a number of receptors that, in most cases, resulted directly from a combination of both experimental and computational tools.[5−10]

GPR40, which has been recently named free fatty acid receptor 1 (FFAR1), is a member of the GPCR superfamily and a possible target for the treatment of type 2 diabetes. It has been shown to be abundantly expressed in the insulin-expressing beta cells of the pancreas and to mediate the majority of the effects of free fatty acids (FFAs) on insulin secretion.[11] Importantly, glucose-stimulated insulin secretion is amplified by FFAs through the activation of GPR40. GPR40 is activated preferentially by unsaturated long chain FFAs found in plasma, such as linoleic and oleic acids (**1a,b**, Scheme 1), with low micromolar potency. The ability to activate GPR40 by compounds based on the 3-(4-{*N*-alkylamino}phenyl)propanoic acid scaffold was discovered by high-throughput screening. Subsequently, the structure−activity relationships of compounds in this series have been explored, leading to the synthesis of analogs endowed with low nanomolar potencies such as GW9508 (**2**, Scheme 1).[12] Another synthetic ligand for GPR40, GW1100, which appears to act as a noncompetitive antagonist, was reported subsequently by the same authors.[13]

Structural analyses of GPCRs via molecular modeling and receptor mutagenesis have proven essential for the understanding of both the pharmacology of small molecule ligands and the ability to engineer these chemical tools to be more potent and efficacious.[14−17] No structural studies on GPR40 and its interactions with ligands have been reported to date. Thus, in this work we conducted a bidirectional iterative investigation, including computational modeling and mutagenesis studies, aimed at delineating the functional "chemoprint" of GPR40, i.e., the amino acid residues involved in agonist recognition. On the basis of sequence analysis and from the computational analysis of the interactions of GPR40 with GW9508, six residues were identified as playing a principal role in the recognition of this ligand. These residues were mutated and the resulting pharmacology was evaluated. In turn, the body of experimental data generated was used to refine the molecular model of the binding cavity.

Throughout the paper, to facilitate the comparison among receptors, we use the GPCR residue indexing system introduced by Ballesteros and Weinstein.[18] Briefly, the most conserved residue in a given transmembrane domain (TM) is assigned the index *X*.50 (where *X* is the TM number), while the remaining residues are numbered relatively to the 50 position.

Results and Discussion

Sequence Analysis. Comparative studies of sequences of homologous proteins provide useful insights into the residues important for function and ligand recognition. In fact, sequence comparison can highlight conserved motifs potentially related to the protein function. Furthermore, if experimental information about the role of specific residues is available for a homologous protein, this knowledge can be applied by analogy to the delineation of the ligand binding site of the protein of interest.[15,19−21] In this context, we performed a detailed analysis

* Corresponding authors. S.C.: phone, 301-451-7353; fax, 301-480-4586; e-mail, stefanoc@mail.nih.gov. M.C.G.: phone, 301-496-4128; fax, 301-496-9943; e-mail, marving@intra.niddk.nih.gov.
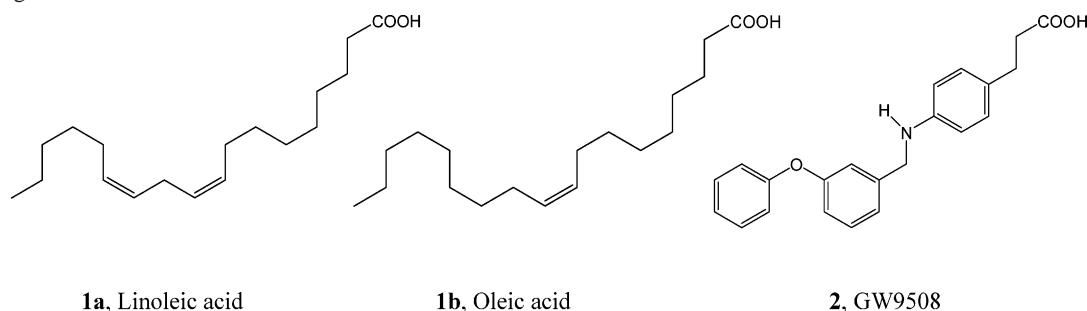† Laboratory of Biological Modeling.
§ Clinical Endocrinology Branch.
‡ Led computational chemistry effort.
|| Led molecular pharmacology effort.
⊥ Chemical Biology Core Facility.

Scheme 1. Agonists for GPR40

1a, Linoleic acid　　　　　　　　　**1b**, Oleic acid　　　　　　　　　**2**, GW9508

of the sequence of GPR40 in relation to those of the phylogenetically closest GPCRs.

A chemogenomic analysis, recently published by Surgand et al., revealed that GPR40 belongs to the same cluster of family A GPCRs to which the nucleotide-activated P2Y receptors belong.[22] Hence, with a BLAST search against the human subset of the SWISS-PROT and TrEMBL databases, we retrieved the closest homologues of GPR40. The retrieved receptors were added to the multiple sequence alignment comprising 68 sequences belonging to the P2Y and to the peptide receptor clusters, reported by Costanzi et al. in the course of a thorough analysis of the P2Y receptors.[21] We expunged all the sequences belonging to the more distantly related peptide receptor branch, thus obtaining a final alignment of 45 sequences that we here designate as the "nucleotide and lipid receptor cluster (NLRC)".

A phylogenetic tree, reflecting the relationships among the 45 receptors, was constructed on the basis of a similarity matrix calculated on the TMs of our new alignment. The cluster includes receptors targeted by phospholipids, lipids, nucleotides, and acid metabolites of the Krebs cycle. In addition, it includes also a family of protease-activated receptors (PARs) and several orphan receptors, whose endogenous ligands are still unknown (Figure 1). Among the orphans is P2Y8, which clusters with the PAR family but does not have an N-terminal region cleavable by proteases. Moreover, GPR17 has recently been found to be activated by uracil nucleotides and cystenyl-leukotrienes.[23]

GPR40 clusters most closely with GPR41, GPR42, GPR43, with which it displays ~33% identity in the TMs. GPR42 is most likely a recent gene duplication of GPR41 and may be a pseudogene.[24] The identity with other NLRC members, calculated on the TMs, ranges from 19% to 27%. Interestingly, GPR120, which also binds long chain FFAs, did not align with the NLRC.

Sequence comparison shows that most of the receptors in the NLRC bear positively charged residues in the extracellular regions of the TM helices, which could attract the anionic part of the ligand. Experimentally, the importance of basic residues has been shown for several members of the NLRC. R3.29 (Ballesteros' residue indexing[18]), H3.33, K/R6.55, and R7.39 proved fundamental for the activation of P2Y$_1$ by nucleotides[15] and for the activation of SUCR1 (GPR91) and OXGR1 (GPR80) by succinate and α-ketoglutarate,[19] while R3.36 proved essential to the binding of nicotinic acid to GPR109A.[21,25,26]

By analogy with other members of the NLRC, we hypothesized that positively charged residues are likely to be relevant to the function of GPR40. Thus, on the basis of sequence comparison, we identified K62(2.60), R183(5.39), R258(7.35), and K259(7.36), all located in the extracellular side of the GPR40 TM helices, as potentially involved in interaction with the carboxyl group of GPR40 ligands. Among these, the residues at positions 2.60, 5.39, and 7.35 are conserved as positively

charged amino acids in all members of the GPR40−43 family, while the residue at position 7.36 is not conserved. Another candidate for the GPR40 functional "chemoprint" is N244(6.55), given that residues located at position 6.55 are often involved in interactions with the ligands in class A GPCRs,[22,27,28] including the P2Y receptors, SUCR1 (GPR91), and OXGR1 (GPR80).[21,25] A further residue of interest is V237(6.48), which is a nonconserved residue in the GPR40−43 family and, as proposed by Surgand et al., could account for the preference of GPR40 for long chain FFAs.[22] It is worth noting that aliphatic residues at position 6.48 are quite rare, while aromatic residues at this position have been proposed to participate in direct interactions with ligands and to act as a conformational switch between the inactive and active state in rhodopsin and other GPCRs,[27,29,30] including members of the NLRC.[21]

Generation of the 3D-Model of GPR40 and Conformational Analysis of the Putative Binding Pocket. A homology model of GPR40 was constructed and optimized on the basis of the ground state of bovine rhodopsin,[31] as explained in the Experimental Section. The sequence alignment between rhodopsin and GPR40 is available in the Supporting Information.

The volume, shape, and physicochemical properties of protein binding sites are determined by the identity and conformation of the residues that form them and are the critical elements for ligand recognition by a receptor. Thus, an extensive conformational analysis of the GPR40 binding site was a crucial step for our studies of the receptor−ligand interactions.

The sequence homology of GPR40 with rhodopsin is only about 16% in the transmembrane domains, and the residues that form the binding site for retinal in rhodopsin are largely not conserved in GPR40. Thus, it is not feasible to predict the orientation of the side chains in the GPR40 binding pocket using homology modeling. When experimental information on specific contacts between the receptor and a ligand are available, ligand-based homology modeling can be used to mold the binding pocket of a receptor around its ligand.[32] In our case, the lack of such experimental information prevented us from applying this methodology.

Therefore, we subjected our homology model of the unoccupied GPR40 to an exhaustive conformational search. A set of 30 residues, located in the upper part of the helical bundle, has been proposed by Surgand et al. to form the GPCR binding cavities on the basis of an analysis of the binding site of retinal in bovine rhodopsin.[22] We selected the corresponding 30 residues of GPR40 and subjected them to de novo torsional sampling with the Monte Carlo multiple minimum (MCMM) method as implemented in MacroModel. In the crystal structure of rhodopsin, the second extracellular loop (EL2) covers the cavity within the helical bundle. We hypothesize that in most GPCRs a flexible EL2 opens up to allow ligands to enter the receptor and closes upon binding to form interactions with the bound ligand.[16] Thus, we removed EL2 from our GPR40 model
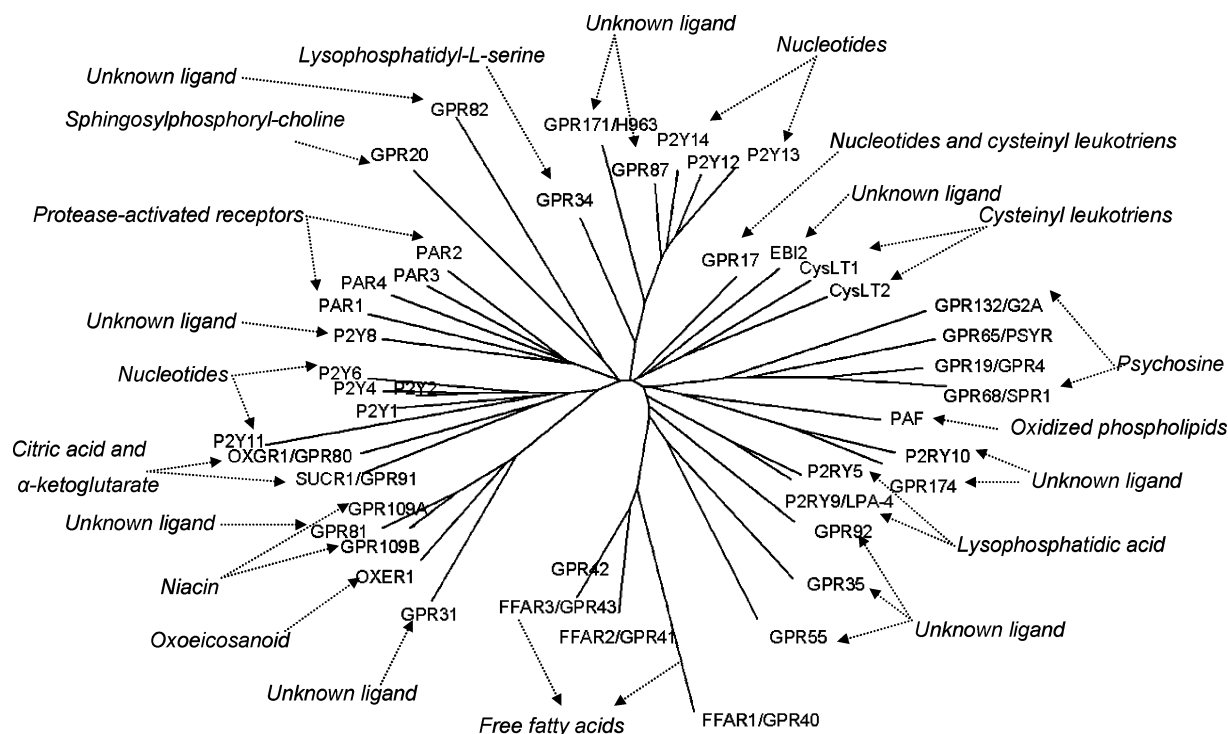
Figure 1. Nucleotide and lipid receptor cluster. The names of sequences are presented according to the IUPHAR nomenclature. When the IUPHAR name is not available, the gene name is shown. Synonyms are shown after the slash. The natural ligands are indicated in italics.
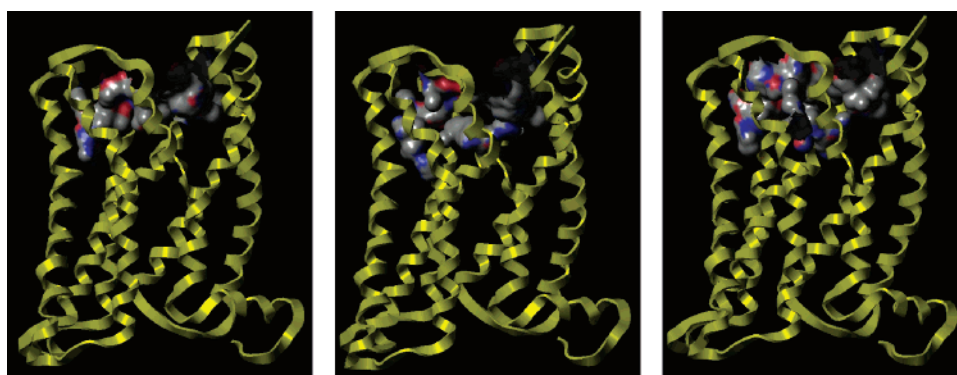
Figure 2. Representative conformations for each of the three clusters of GPR40 conformers. One hundred receptor conformers were divided into 12 groups based on the rmsd of specific residues (see the text), which were subsequently clumped into three clusters on the basis of solvent accessible surface. The backbone of the receptor is represented as a yellow ribbon. The surfaces of the cavities are colored according to H-bonding properties (red, H-bond donors; blue, H-bond acceptors).

prior to the conformational study to simulate the open state of the loop. An ensemble of 100 protein conformations was generated and clustered into 12 groups on the basis of the atomic root-mean-square (rms) displacement of the side chains of four aromatic residues, F87(3.31), H86(3.32), Y91(3.37), and Y240-(6.51). These were chosen because of their central location in the putative binding site and potential to act as a gate for access of the ligand to deeper cavities inside the protein. PROCHEK[33]-analysis of φ, ψ, χ_1, and χ_2 angles of 100 protein conformations did not detect unfavorable side chain conformations.

Subsequently, we subjected the lowest energy conformers from each of the 12 groups to solvent accessible surface analysis. The 12 representative conformations were grouped into three major clusters on the basis of the volume and the shape of the cavities (Figure 2). The first cluster showed a rather shallow cavity with a volume of ~890 Å^3 and a small hollow between TM4 and TM5. The second cluster showed a total volume of ~1060 Å^3 and two deep subcavities between TM4:TM5:TM6 and TM2:TM3:TM7. The third cluster showed three subcavities located between TM4:TM5, TM3:TM6, and TM1:TM7 with a

total volume of ~1350 Å^3. The solvent accessible surfaces of all the models showed a hydrogen bond (H-bond) donor region close to the extracellular side which corresponds to R5.39 and R7.35 (Figure 2, red surface).

An analysis of the solvent accessible surface of the putative binding pocket of the homology model, prior to the conformational analysis, revealed a very shallow cavity with a total volume ~820 Å^3. Thus, the inner cavities of GPCR models are deeply affected by the conformation of the residues that line them. Although originally small when built on the basis of homology to rhodopsin, they can significantly expand to accommodate larger ligands. This observation suggests the importance of a thorough exploration of receptor conformation before performing docking experiments.

Flexible Docking of GW9508 at GPR40. Automatic docking studies were performed using FlexE,[34] which combinatorially joins specific protein conformers to create a larger conformation ensemble. One representative conformer from each of the
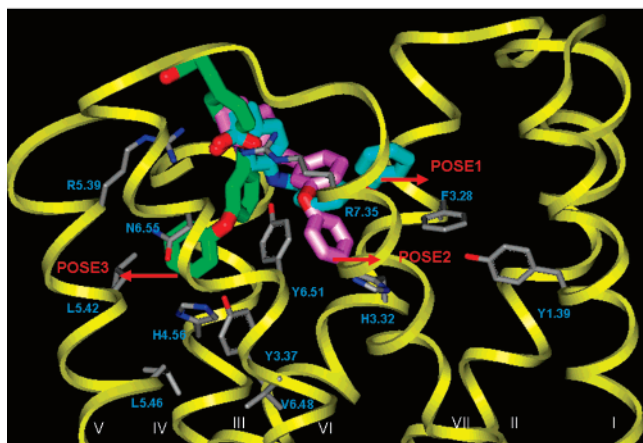
Figure 3. Docking poses for GW9508 generated using FLEXE, which combinatorially joined 12 protein conformations (see the text). The carbon atoms are colored in cyan for pose1, in magenta for pose 2, and in green for pose 3.

Figure 4. Combined molecular interaction fields of carboxyl (red), hydrophobic (white), and aromatic (green) probes with 12 protein conformations. Only low-interaction energy fields are shown.

previously mentioned 12 groups was selected according to the orientation of positively charged residues located in the putative binding pocket. We selected the structures with the side chains oriented toward the inside of the pocket, in order to allow their interaction with the ligand.

We defined the potential binding site as the general GPCR cavity lined by the 30 residues proposed by Surgand et al. and automatically docked the high-affinity synthetic ligand GW9508 (**2**) into our model of GPR40. In all poses, the carboxyl group of the ligand docked near R183(5.39) and R258(7.35). However, the hydrophobic tail showed three different binding modes: oriented upward between TM1, TM2, and TM3 (pose 1); more deeply buried between TM2, TM3 and TM6 (pose 2); and oriented between TM3, TM5 and TM6 (pose3) (Figure 3). Notably, the automatic FlexX docking of GW9508 (**2**) to the GPR40 homology model prior to the conformational search failed to position the ligand hydrophobic tail in the putative binding pocket, confirming the importance of a thorough conformational analysis of the GPCR before the execution of docking experiments

To validate these docking results and to resolve the orientation of the hydrophobic tail, we calculated molecular interaction fields using the program GRID.[35] For this purpose, we took again a representative conformer from each of the 12 groups. The carboxyl, hydrophobic, and aromatic probes within GRID were used to approximate the physicochemical properties of the GPR40 agonists. Consistent with our docking results, the carboxyl probe showed low-energy fields in proximity of R183-(5.39) and R258(7.35) in all models. The hydrophobic probe and the aromatic probe showed low-energy fields in the cavity between TM2, TM3, and TM6, i.e., the region to which the hydrophobic tail bound in pose 2, and in the cavity between TM3, TM5, and TM6, i.e., the region to which the hydrophobic tail bound in pose 3 (Figure 4). In contrast, there were no favorable interactions in the region to which the hydrophobic tail bound in docking pose 1, which we consequently rejected.

To distinguish between the two remaining poses, we identified two His residues in TM3 and TM4 showing direct interactions with the aromatic moiety of the ligand in poses 2 and 3, respectively, and selected them for mutagenesis studies. In particular, for pose 2 we chose H86(3.32), which in our conformational search of GPR40 influenced significantly the size and shape of the cavity between TM3, TM6, and TM7. Notably, position 3.32 is generally occupied by aromatic residues in the NLRC. To investigate pose 3, we chose H137-
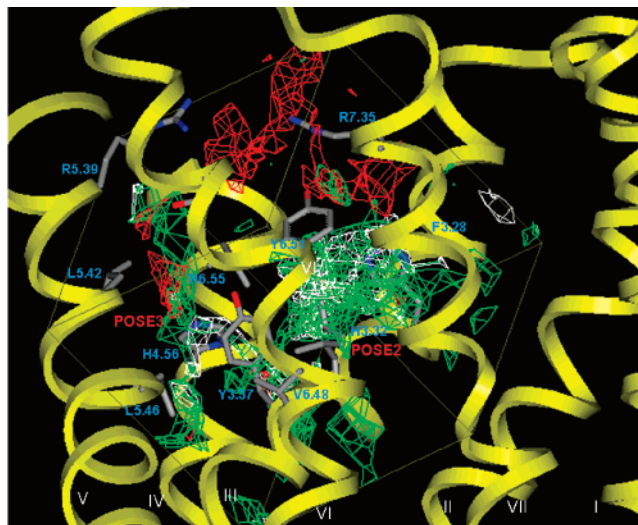
(4.56), which is conserved in the GPR40−43 family and therefore could be part of the ligand recognition motif.

Our model positioned K62(2.60) and K259(7.36) on the external side of the helices, with the side chains pointing away from the putative binding cavity. Thus, we excluded their involvement in the ligand recognition motif, although it seemed plausible after sequence analysis alone.

On the basis of the sequence analysis and molecular modeling, we selected for mutagenesis studies R183(5.39), N244(6.55), and R258(7.35) to explore their possible roles as anchors for the carboxyl moiety of GW9508. Further, mutational examination of H86(3.32) and H137(4.56) was performed to assist in discrimination between poses 2 and 3. We also selected V237-(6.48) as a further validation of our model. A role in the activation of GPR40 by long chain fatty acids has been proposed for this residue,[22] but this residue does not make contacts with GW9508 in our model.

Site-Directed Mutagenesis and Pharmacology. Agonist binding to GPR40 leads to activation of phospholipase C and subsequent increases in intracellular calcium.[11,36,37] Wild type GPR40 and site-specific GPR40 mutants were transiently expressed in HEK-EM 293 cells. After loading with a calcium-sensitive fluorescent dye, cells were treated with the synthetic GPR40 agonist GW9508[12] (**2**) and changes in intracellular calcium levels were assessed using a fluorometric imaging plate reader (FLIPR). For wild-type GPR40, the agonist showed a potency of 223 nM (log EC$_{50}$= −6.65 ± 0.028, n = 18), with a range of 159−1092 nM (Figure 5).

Mutations of residues predicted to anchor the carboxyl moiety of the agonist drastically decreased the agonistic potency (Figure 5). Hence, both R183A(5.39) and R258A(7.35) resulted in a greater than 100-fold decrease in the potency of GW9508 and substantial loss of activity. Mutation of R258(7.35) to Lys, instead of Ala, partially restored potency, suggesting the involvement of hydrogen bonds or electrostatic interactions. Also, mutation of N244(6.55) to Ala resulted in a smaller but still significant decrease in potency (16-fold, EC$_{50}$ = 17.8 μM). Taken together, these data are in agreement with the molecular model described above, in which regions around R183(5.39) and R258(7.35) showed the lowest energy field with a carboxyl probe, reinforcing the idea that the carboxyl group of GW9508 anchors to these residues. Moreover, since the involvement of these residues was predicted from sequence analysis of related
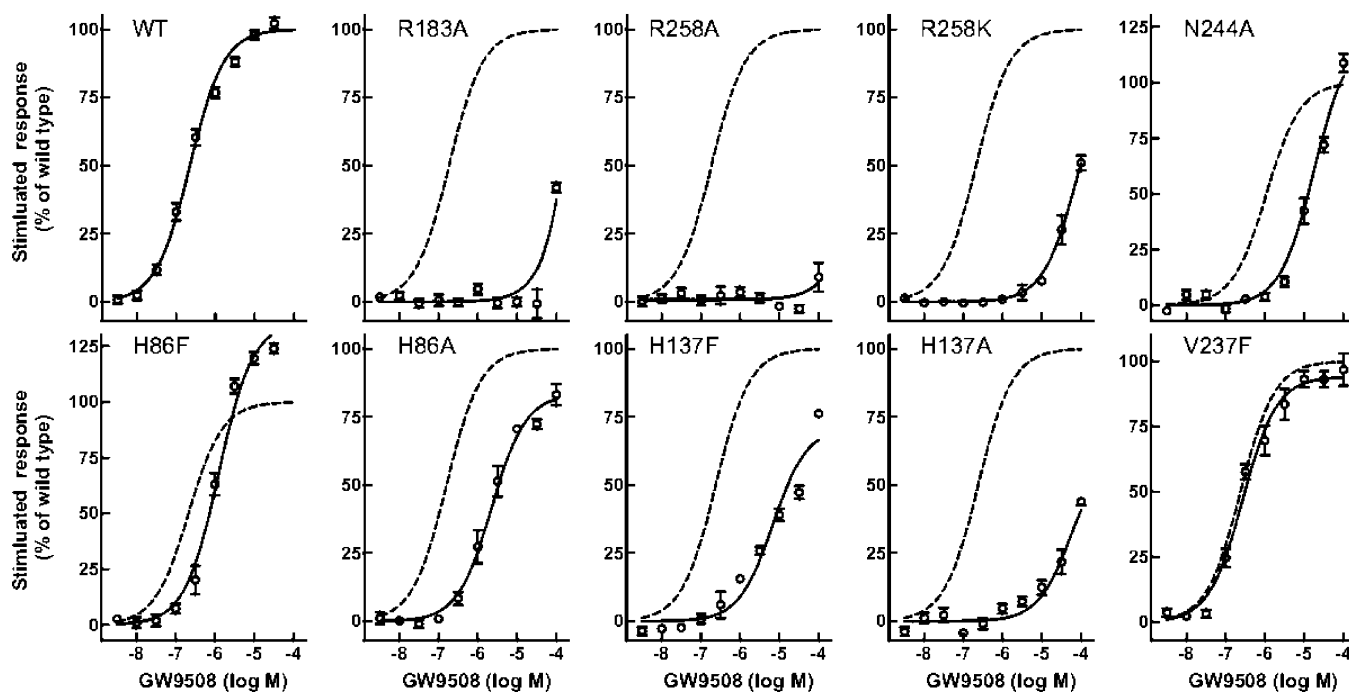
Figure 5. Effect of mutations on the potency of agonist GW9508. The increase in intracellular calcium upon receptor activation was measured and expressed relative to the wild-type receptor as mean percent $\pm$ SEM (see Experimental Section). The mutations are indicated on each graph. Dashed lines show the response of the wild-type receptor performed in parallel with the mutant. The EC_{50} of GW9508 in the wild-type receptor was 223 nM (log $EC_{50} = -6.65 \pm 0.028$, $n = 18$), with a range from 159 to 1092 nM. The decrease in potency of GW9508 following different mutations were: >100 fold, R183A; >100 fold, R258A; >100 fold, R258K; 16 fold, N244A; 5.8 fold, H86F; 14 fold, H86A; 28 fold, H137F; >100 fold, H137A; and 1.1 fold, V237F. Data shown are averages of three or more experiments.

receptors, GPR40 appears to exhibit similar architecture regarding the anchoring of ligand to receptor.

H86(3.32) and H137(4.56) were mutated individually to either Phe or Ala. The conservative substitution to Phe caused a small reduction in potency in H86F but led to a greater change in H137F (5.8-fold vs 28-fold, EC_{50}s: 1.2 μM vs 6.8 μM). The changes were greater with the Ala substitution, which lowered the potency by 14-fold and over 100-fold, respectively (EC_{50}s: 2.2 μM vs > 30 μM). These mutants were designed to distinguish between the two docking poses obtained in the modeling study. The results, therefore, suggest that the agonist is more likely to interact with H137 (pose 3) than with H86 (pose 2) (refer to Figure 6). However, we cannot exclude the possibility of some contribution from H86 as the decrease in potency, although small, was significant.

V237(6.48) has been suggested by others on the basis of sequence analysis as a possible reason for the ligand preference of GPR40 for long-chain free fatty acids.[22] As opposed to Val or Leu, a Phe is present in this position in receptors activated by short-chain free fatty acids (i.e., GPR41 and GPR43). A Phe substitution at this position in GPR40 did not alter the potency of GW9508. This result is consistent with the docking arrangement we have predicted and provides additional support for our model (Figure 3).

The pharmacological properties of the mutants are not due to an altered level of cell surface expression. Except for the H86A mutant, which expressed a bit more poorly (22% below the wild-type), the other mutants were either better expressed or did not show very different expression than the wild-type receptor.
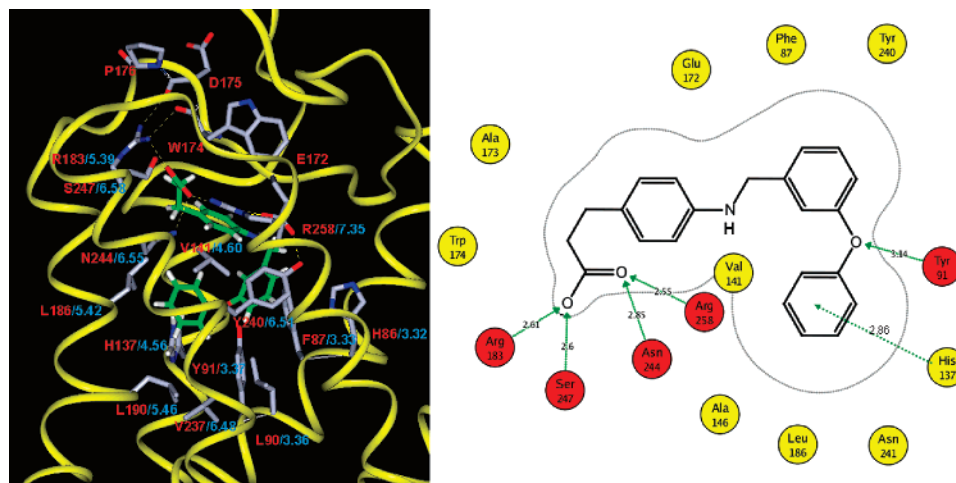
Generation of an Experimentally Supported 3D Model of the GPR40−GW9508 Complex. The replacement of H137-(4.56) with Ala or Phe resulted in a significant reduction of the potency of GW9508 (**2**), while the replacement H86(3.32) with the same residues gave smaller shifts. Hence, H137(4.56) seems
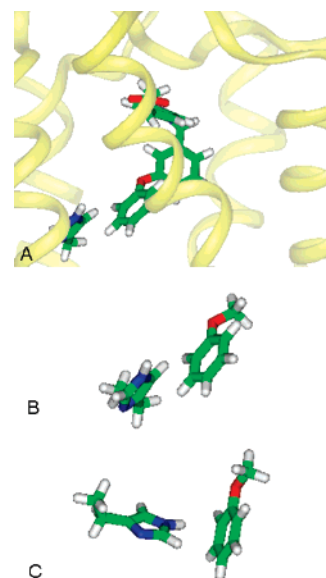
Figure 6. Arrangement of amino−aromatic interactions: A, relative orientation of GW9508 obtained from docking study; B, the simplified system used for QM study; C, result of QM calculation.

directly involved in the ligand binding through aromatic and H-bond interactions. Thus, we chose pose 3, in which H137-(4.56) is in contact with the 3-phenoxy moiety of the ligand, as the most likely binding mode of GW9508 (**2**) and we proceeded to further optimize the model.

EL2 was added to the receptor−ligand complex and the model was optimized using molecular dynamics simulation (see Experimental Section). Subsequently, the ligand and the residues located within a distance of 7 Å were subjected to an MCMM conformational search (see Experimental Section). During optimization, the ligand drifted slightly deeper into the binding pocket.

Figure 7. The experimentally supported binding site of GW9508 in GPR40 (on the left) and the simplified scheme of protein−ligand interactions (on the right).

The calculation of the p*K* values for H137(4.56) in 12 protein conformations using the Poisson−Boltzmann equation and the generalized Born approach (with the program H++ as implemented at the website chekhov.cs.vt.edu/completion/index.php) led to the prediction that the His is unprotonated and, predominantly, in the ε tautomeric form.

In all conformations of the complex generated by MCMM, H137(4.56) was found within a distance of 3−4 Å from the 3-phenoxy moiety of the ligand forming hydrophobic/aromatic interactions. Considering the important role for the imidazole moiety of H137(4.56) suggested by mutagenesis, we hypothesized that the polarized imidazole proton likely forms electrostatic interactions with the π-electron cloud of the 3-phenoxy moiety of the ligand. Moreover, theoretical calculations have suggested that such an interaction can account for up to ∼3 kcal/mol of binding energy,[38] which is consistent with the 28-fold reduction (∼2 kcal/mol) of the GW9508 (**2**) potency in the H137(4.56)F mutant. Molecular mechanics optimization did not yield this interaction, since empirical force fields do not consider explicit π-electrons. Therefore, we performed a quantum mechanical energy minimization on a simplified system constituted only by the His side chain and the 3-phenoxy moiety of the ligand. The missing part of the receptor and the ligand were substituted by capping with methyl groups. To keep the molecular fragments within a distance close to that shown in the complex, we fixed the coordinates of the capping groups (Figure 6). As expected, after minimization a polarized proton−π interaction was observed between the molecular fragments. Finally, we put the fragments back into the complex and subjected the entire model to a final energy minimization with the orientation of the imidazole and the benzyl rings constrained. The final optimized complex is presented in Figure 7.

The ligand is accommodated between TM3, TM4, TM5, and TM6, with the carboxyl group forming a H-bond network with R183(5.39), R258(7.35), N244(6.55), and S247(6.58). Although there are no direct contacts between GW9508 and EL2, the addition of the loop to the model led to the formation of additional receptor−ligand interactions, involving N244(6.55) and S247(6.58), and to changes in the interactions between the ligand and the arginines in the binding site. In particular, after addition of EL2, R183(5.39) showed two H-bond interactions with the oxygen of the backbone of D175 and W174 and one interaction with the ligand, while R258(7.35) showed two

interactions with E172 and two interactions with the ligand. The involvement of all four hydrogens of the R258(7.35) seems critical for the stabilization of the receptor−ligand complex, as suggested by the 100-fold loss of potency exhibited by R258K-(7.35).

The aromatic portion of the ligand lies in a pocket lined by H86(3.32), F87(3.33), L90(3.36), Y91(3.37), H137(4.56), V141-(4.60), L186(5.42), L190(5.46), and Y240(6.51). The 3-phenoxy moiety forms amino−aromatic interactions with H137(4.56) and an H-bond with the hydroxyl group of Y91(3.37). V237(6.48) is not involved in interactions with GW9508, consistent with the neutral effect of its mutation to Phe. Interestingly, L186-(5.42), which faces the 3-phenoxy moiety, is substituted by a Phe in mouse and rat (See Supporting Information). Presumably, this difference could lead to differences in potency of GW9508 (**2**) in various species.

Conclusions

In this work, the GPR40 functional "chemoprint" for agonist recognition was predicted computationally and subsequently validated by site-directed mutagenesis. Our experimentally supported model suggested that H137(4.56), R183(5.39), N244-(6.55), and R258(7.35) are directly involved in interactions with the ligand. The contribution of an amino−aromatic interaction to the binding of GW9508 was also suggested.

Furthermore, an electrostatic interaction between R258(7.35) and E172, located in the second extracellular loop, was detected. This interaction may be crucial to the function of the receptor, as suggested by the significant loss of potency of the R258K-(7.35) mutant. Within the NLRC, acidic residues in EL2 have proven critical for receptor function in P2Y receptors. In particular, mutation of D204 in P2Y$_1$, which corresponds to E172, decreases agonist-promoted activation of the receptor.[39] Furthermore, molecular dynamics in a hydrated lipid bilayer suggested that D179 of the P2Y$_6$ receptor, which also corresponds to E172 in GPR40, and R128(3.29) are engaged in an electrostatic interaction in the model of unoccupied receptor. Subsequent molecular dynamics of the receptor−ligand complex led to the disruption of this electrostatic interaction and of a movement of EL2 toward the extracellular space, which could be associated with activation of the receptor.[16]

These data provide the first structural hypothesis on the interactions of this new potential target for the treatment of type

2 diabetes with synthetic molecules. The coordinates of the complex are supplied as Supporting Information and may be useful for receptor-based drug design of novel GPR40 ligands.

We also provided a general strategy for the exploration of receptor−ligand interactions to be applied to GPCRs for which no previous mutational data are available. Central to our strategy are the removal of EL2, which simulates the opening of the loop, and the subsequent conformational analysis of the inner cavity of the receptor. This procedure allows an exhaustive exploration of the conformational space available to form the binding pocket. Thus, the small interhelical hollow that generally characterizes rhodopsin-based homology models opens up and becomes suitable for automatic docking of ligands larger than retinal. The subsequent reinsertion of EL2 simulates the closing of the loop and allows this crucial extracellular domain to contribute to the formation of the binding pocket.

Experimental Section

Multiple-Sequence Alignment and Phylogenetic Tree. The sequences retrieved from the SWISS-PROT and TrEMBL databases were added to the multiple sequence alignment comprising 68 sequences belonging to the purine receptor and to the peptide clusters, performed by Costanzi et al.[21] using CLUSTALW.[40] The BLOSUM62 matrix[41] was applied, with a gap start penalty of 5 and a gap extend penalty of 0.2.

A pairwise distance matrix was calculated on the TM region of the alignment with the Protdist program of the PHYLIP[42] version 3.65 with the Dayhoff PAM substitution matrix.[43] The resulting pairwise distance matrix was used to generate the phylogenetic tree by the program Neighbor of the PHYLIP version 3.65 applying the neighbor-joining method of Saitou and Nei.[44] The final phylogenetic tree was plotted with the program TreeView.[45]

Generation of 3D Model of GPR40. The homology model was generated using MODELLER[46,47] as implemented in InsightII.[48] The coordinates of bovine rhodopsin (PDB ID 1GZM[31]) were used as 3D-template. One model and five loop refinements were constructed. The disulfide bridge between the second extracellular loop (EL2) and the upper part of TM3 was defined manually. The structure obtained was optimized using the Discover module of Insight II[3] with the CVFF force field. The structure was first minimized until convergence of 0.01 kcal/mol/Å was reached using a conjugated gradient. A short molecular dynamic simulation of 500 ps at 310 K was carried out with an integration time step of 1 fs. During the simulation, distance restraints for the H-bonds between the O atom of *i*-residue and NH atom of the *i* + 4 residue were applied to maintain the helical structure. In the cases where the *i* residue was a Pro, constraints were not applied to the residues *i* − 3 and *i* − 4. The scale factor applied to the attractive and repulsive force of the distance constraints was initially set to 500 kcal/mol/Å and reduced during molecular dynamics to 200, 100, 50, and 0 kcal/ mol/Å every 10 ps. The nonbond cutoff method and the dielectric constant were set up to cell multipole and were distance-dependent. Once the system was equilibrated, the coordinates of 10 snapshots were averaged and submitted again to the previously mentioned minimization protocol, without any restraints.

Conformational Search of Residues in the Binding Site. A side chain conformational search for the 30 residues lining the receptor cavity was performed using torsional sampling (Monte Carlo multiple minimum) implemented in MacroModel.[49] A frozen shell comprising the residues within 3 Å from the cavity was included in the calculation, while the remaining residues were excluded. The following parameters were applied: number of steps = 1000, number of structures to save for each search = 100, and energy window for saving structures = 1000 kJ/mol. The calculations were conducted with MMFFs force field and distance-dependent dielectric constant of 1. Polak−Ribier conjugate gradient with a convergence threshold on the gradient of 0.05 kJ/Å/mol was used for minimization.

GRID Calculations. A 25 Å × 25 Å × 25 Å lattice of points spaced at 0.5 Å was calculated for the cavity formed by the 30 amino acid residues. The probes used were COO− (carboxyl probe), DRY (hydrophobic probe), and C1═ (sp2, aromatic or vinyl carbon). The dielectric constants were set to 4.0 for the macromolecule and 80.0 for the bulk water.

Docking Studies and Final Optimization. To consider protein flexibility, GW9508 was automatically docked with FlexX[50] (Sybyl, Tripos[51]) to an ensemble of protein conformations combinatorially generated by the FlexE[34] module combining 12 distinct conformers. Default parameters were used. Gasteiger−Huckel charges were used for the ligands and MMFFs charges were used for the protein.

After docking the EL2 was added to the complex by superimposition with the initial intact homology model. The resulting complex was optimized in two steps: (1) the TMs and the ligand were fixed and EL2 was minimized and subjected to short dynamic simulation (100 ps) and (2) the whole complex was subjected to 500 ps of molecular dynamics with distance restrains for interhelical H-bonds. The simulations were performed by following the protocol described above. An MCMM conformational search of binding pocket residues and ligand was performed following the protocol described above. The quantum mechanical geometry optimization of the His side chain and 3-phenoxy moiety capped by methyl groups was performed using Spartan.[52] The Hartree−Fock method was used with the 6-31G** basis set. The complex was subjected to a final energy minimization (as described above) constraining the distance between the NH group of H137(4.56) and the 3-phenoxy moiety of the ligand in order to conserve the amino−aromatic interaction.

Synthesis of 3-(4-(3-Phenoxybenzylamino)phenyl)propanoic Acid (GW9508). To 1 g (5 mmol) of 3-phenoxybenzaldehyde and 0.98 g (6 mmol) of 3-(4-aminophenyl)propanoic acid in 50 mL of DCE was added 1 drop of acetic acid and 2.1 g (10 mmol) of sodium triacetoxy borohydride. The resulting solution was allowed to stir for 2 h at room temperature and solvent was removed under reduced pressure. To the resulting slurry was added water (10 mL) and methanol (20 mL), the resulting mixture was allowed to stir for 30 min, and the solvent was removed under reduced pressure.[12] The resulting slurry was purified by column chromatography (1:1 hexanes:ethyl acetate with trace acetic acid) to yield 1.48 g (85%) of 3-(4-(3-phenoxybenzylamino)phenyl)propanoic acid as a clear oil that solidified upon standing. Analysis by C_8 reversed phase LCMS using a linear gradient of H_2O with increasing amounts of CH_3CN (0 → 5 min, 30% → 90% CH_3CN; 5 → 9 min, 90% → 90% CH_3CN at a flow rate of 1 mL/min, t_R 6.46 min) found greater than 98% purity by peak integration. ^{1}H NMR (CDCl₃) δ 2.40 (t, 2H), 2.63 (t, 2H), 4.21 (s, 2H), 6.09 (bs, 1H), 6.44−6.47 (m, 2H), 6.81−6.83 (m, 1H), 6.87−6.89 (m, 2H), 6.94−6.97 (m, 2H), 7.07 (bs, 1H), 7.11−7.18 (m, 2H), 7.29−7.33 (m, 1H), 7.35−7.38 (m, 2H), 12.09 (bs, 1H); MS (TOF); m/z = 348.1600 (M + H⁺) (theoretical 348.1594).

Constructs. Plasmids coding for the human GPR40/FFAR1 were a gift from Dr. Brian O'Dowd (University of Toronto, Canada).[53] The region containing the FFAR1 insert was subcloned into the *Bam*HI and *Xho*I sites of pcDNA3.1/hygro(+) (Invitrogen, Carlsbad, CA). Mutagenesis was carried out using the Quikchange II XL method (Stratagene, LaJolla, CA) following the manufacturer's protocol. All sequences were confirmed by sequencing the full gene including the promoter.

The levels of expression on the plasma membrane have been measured using an epitope tag for the wild-type and the H86A, H86F, H137A, H137F, N244A, R258K mutant receptors (unpublished results; a full paper will follow).

Cell Culture and Transfection. HEK-EM 293 cells were maintained in high-glucose DMEM supplemented with 10% FBS. For transfection, cells were seeded on 100-mm plates (BD Falcon, Franklin Lakes, NJ) to attain 80% confluency the next day. Transfection was carried out with Lipofectamine 2000 (Invitrogen, Carlsbad, CA) according to procedures recommended by the

manufacturer. Cells were reseeded onto 96-well plates at 60×10^3 cells/well 1 day following transfection and then assayed the next day.

Fluorometric Imaging Plate Reader (FLIPR) Analysis. Receptor assay was carried out by measuring the calcium flux in response to the addition of agonist. HEK-EM 293 cells seeded in 96-well plates were loaded with Calcium 3 fluorescent calcium dye (Molecular Devices, Sunnyvale, CA) for 1.5 h at room temperature before stimulation with an agonist. Fluorescent signals were measured on a FLIPR Tetra (Molecular Devices, Sunnyvale CA). Measurements were taken every 0.5 s during the first 5 min of the assay. Ten measurements were made before compound addition followed by 450 after compound addition. Afterward, 10 more measurements were taken for 5 s each. Experiments with each mutant were carried out in parallel with the wild-type receptor in triplicates or quadruplicates.

Data Analysis. Agonist-stimulated response in wild-type and mutant receptors was taken as the maximum − minimum value by subtracting the baseline response in vector-transfected control cells. The dose−response relationship was analyzed by fitting sigmoidal curves to the data sets using GraphPad Prism 4 (San Diego, CA). The set of data obtained with the wild-type receptor and the mutant within the same experiment were analyzed simultaneously to obtain the relative EC_{50} values. The EC_{50} value for the wild-type receptor was obtained as an average value from all experiments. The response was normalized as percentages relative to that of the wild-type receptor, and the data were presented as mean ± SEM of three or more repeat experiments.

Supporting Information Available: Coordinates of the complex GPR40-GW9508 in PDB format. Alignment of the sequences of GPR40 from four different species and bovine rhodopsin. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Fanelli, F.; De Benedetti, P. G. Computational modeling approaches to structure-function analysis of G protein-coupled receptors. *Chem. Rev.* **2005**, *105*, 3297−3351.

(2) Kristiansen, K. Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: Molecular modeling and mutagenesis approaches to receptor structure and function. *Pharmacol. Ther.* **2004**, *103*, 21−80.

(3) Lefkowitz, R. J. Historical review: A brief history and personal retrospective of seven-transmembrane receptors. *Trends Pharmacol. Sci.* **2004**, *25*, 413−422.

(4) Moro, S.; Spalluto, G.; Jacobson, K. A. Techniques Recent developments in computer-aided engineering of GPCR ligands using the human adenosine A3 receptor as an example. *Trends Pharmacol. Sci.* **2005**, *26*, 44−51.

(5) Bock, J. R.; Gough, D. A. Virtual screen for ligands of orphan G protein-coupled receptors. *J. Chem. Inf. Model.* **2005**, *45*, 1402−1414.

(6) Bologa, C. G.; Revankar, C. M.; Young, S. M.; Edwards, B. S.; Arterburn, J. B.; Kiselyov, A. S.; Parker, M. A.; Tkachenko, S. E.; Savchuck, N. P.; Sklar, L. A.; Oprea, T. I.; Prossnitz, E. R. Virtual and biomolecular screening converge on a selective agonist for GPR30. *Nat. Chem. Biol.* **2006**, *2*, 207−212.

(7) Evers, A.; Hessler, G.; Matter, H.; Klabunde, T. Virtual screening of biogenic amine-binding G-protein coupled receptors: Comparative evaluation of protein- and ligand-based virtual screening protocols. *J. Med. Chem.* **2005**, *48*, 5448−5465.

(8) Evers, A.; Klabunde, T. Structure-based drug discovery using GPCR homology modeling: Successful virtual screening for antagonists of the alpha1A adrenergic receptor. *J. Med. Chem.* **2005**, *48*, 1088−1097.

(9) Raitio, K. H.; Salo, O. M.; Nevalainen, T.; Poso, A.; Jarvinen, T. Targeting the cannabinoid CB2 receptor: Mutations, modeling and development of CB2 selective ligands. *Curr. Med. Chem.* **2005**, *12*, 1217−1237.

(10) Schneider, G.; Nettekoven, M. Ligand-based combinatorial design of selective purinergic receptor (A2A) antagonists using self-organizing maps. *J. Comb. Chem.* **2003**, *5*, 233−237.

(11) Itoh, Y.; Kawamata, Y.; Harada, M.; Kobayashi, M.; Fujii, R.; Fukusumi, S.; Ogi, K.; Hosoya, M.; Tanaka, Y.; Uejima, H.; Tanaka, H.; Maruyama, M.; Satoh, R.; Okubo, S.; Kizawa, H.; Komatsu, H.; Matsumura, F.; Noguchi, Y.; Shinohara, T.; Hinuma, S.; Fujisawa, Y.; Fujino, M. Free fatty acids regulate insulin secretion from pancreatic beta cells through GPR40. *Nature* **2003**, *422*, 173−176.

(12) Garrido, D. M.; Corbett, D. F.; Dwornik, K. A.; Goetz, A. S.; Littleton, T. R.; McKeown, S. C.; Mills, W. Y.; Smalley, T. L., Jr.; Briscoe, C. P.; Peat, A. J. Synthesis and activity of small molecule GPR40 agonists. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 1840−1845.

(13) Briscoe, C. P.; Peat, A. J.; McKeown, S. C.; Corbett, D. F.; Goetz, A. S.; Littleton, T. R.; McCoy, D. C.; Kenakin, T. P.; Andrews, J. L.; Ammala, C.; Fornwald, J. A.; Ignar, D. M.; Jenkinson, S. Pharmacological regulation of insulin secretion in MIN6 cells through the fatty acid receptor GPR40: Identification of agonist and antagonist small molecules. *Br. J. Pharmacol.* **2006**, *148*, 619−628.

(14) Jacobson, K. A.; Kim, S. K.; Costanzi, S.; Gao, Z. G. Purine receptors: gpcr structure and agonist design. *Mol. Interv.* **2004**, *4*, 337−347.

(15) Costanzi, S.; Ivavov, A. A.; Tikhonova I. G.; Jacobson, K. A. Structure and function of G protein-coupled receptors studied using sequence analysis, molecular modeling and receptor engineering: Adenosine receptors. In *Frontiers Drug Des. Discovery* **2007**, in press.

(16) Costanzi, S.; Joshi, B. V.; Maddileti, S.; Mamedova, L.; Gonzalez-Moa, M. J.; Marquez, V. E.; Harden, T. K.; Jacobson, K. A. Human P2Y(6) receptor: Molecular modeling leads to the rational design of a novel agonist based on a unique conformational preference. *J. Med. Chem.* **2005**, *48*, 8108−8111.

(17) Patny, A.; Desai, P. V.; Avery, M. A. Homology modeling of G-protein-coupled receptors and implications in drug design. *Curr. Med. Chem.* **2006**, *13*, 1667−1691.

(18) Ballesteros, J. A.; Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure−function relations in G-protein coupled receptors. *Methods Neurosci.* **1995**, *25*, 366−428.

(19) Ballesteros, J. A.; Shi, L.; Javitch, J. A. Structural mimicry in G protein-coupled receptors: Implications of the high-resolution structure of rhodopsin for structure−function analysis of rhodopsin-like receptors. *Mol. Pharmacol.* **2001**, *60*, 1−19.

(20) Shi, L.; Javitch, J. A. A role for information collection, management, and integration in structure−function studies of G-protein coupled receptors. *Curr. Pharm. Des.* **2006**, *12*, 1771−1783.

(21) Costanzi, S.; Mamedova, L.; Gao, Z. G.; Jacobson, K. A. Architecture of P2Y nucleotide receptors: Structural comparison based on sequence analysis, mutagenesis, and homology modeling. *J. Med. Chem.* **2004**, *47*, 5393−5404.

(22) Surgand, J. S.; Rodrigo, J.; Kellenberger, E.; Rognan, D. A chemogenomic analysis of the transmembrane binding cavity of human G-protein-coupled receptors. *Proteins* **2006**, *62*, 509−538.

(23) Ciana, P.; Fumagalli, M.; Trincavelli, M. L.; Verderio, C.; Rosa, P.; Lecca, D.; Ferrario, S.; Parravicini, C.; Capra, V.; Gelosa, P.; Guerrini, U.; Belcredito, S.; Cimino, M.; Sironi, L.; Tremoli, E.; Rovati, G. E.; Martini, C.; Abbracchio, M. P. The orphan receptor GPR17 identified as a new dual uracil nucleotides/cysteinyl-leukotrienes receptor. *EMBO J.* **2006**, *25*, 4615−4627.

(24) Brown, A. J.; Goldsworthy, S. M.; Barnes, A. A.; Eilert, M. M.; Tcheang, L.; Daniels, D.; Muir, A. I.; Wigglesworth, M. J.; Kinghorn, I.; Fraser, N. J.; Pike, N. B.; Strum, J. C.; Steplewski, K. M.; Murdock, P. R.; Holder, J. C.; Marshall, F. H.; Szekeres, P. G.; Wilson, S.; Ignar, D. M.; Foord, S. M.; Wise, A.; Dowell, S. J. The orphan G protein-coupled receptors GPR41 and GPR43 are activated by propionate and other short chain carboxylic acids. *J. Biol. Chem.* **2003**, *278*, 11312−11319.

(25) He, W.; Miao, F. J.; Lin, D. C.; Schwandner, R. T.; Wang, Z.; Gao, J.; Chen, J. L.; Tian, H.; Ling, L. Citric acid cycle intermediates as ligands for orphan G-protein-coupled receptors. *Nature* **2004**, *429*, 188−193.

(26) Tunaru, S.; Lattig, J.; Kero, J.; Krause, G.; Offermanns, S. Characterization of determinants of ligand binding to the nicotinic acid receptor GPR109A (HM74A/PUMA-G). *Mol. Pharmacol.* **2005**, *68*, 1271−1280.

(27) Archer-Lahlou, E.; Tikhonova, I.; Escrieut, C.; Dufresne, M.; Seva, C.; Pradayrol, L.; Moroder, L.; Maigret, B.; Fourmy, D. Modeled structure of a G-protein-coupled receptor: The cholecystokinin-1 receptor. *J. Med. Chem.* **2005**, *48*, 180−191.

(28) Wieland, K.; Zuurmond, H. M.; Krasel, C.; Ijzerman, A. P.; Lohse, M. J. Involvement of Asn-293 in stereospecific agonist recognition and in activation of the beta 2-adrenergic receptor. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 9276−9281.

(29) Colson, A. O.; Perlman, J. H.; Jinsi-Parimoo, A.; Nussenzveig, D. R.; Osman, R.; Gershengorn, M. C. A hydrophobic cluster between transmembrane helices 5 and 6 constrains the thyrotropin-releasing hormone receptor in an inactive conformation. *Mol. Pharmacol.* **1998**, *54*, 968−978.

(30) Crocker, E.; Eilers, M.; Ahuja, S.; Hornak, V.; Hirshfeld, A.; Sheves, M.; Smith, S. O. Location of Trp265 in metarhodopsin II: Implications for the activation mechanism of the visual receptor rhodopsin. *J. Mol. Biol.* **2006**, *357*, 163−172.

(31) Li, J.; Edwards, P. C.; Burghammer, M.; Villa, C.; Schertler, G. F. Structure of bovine rhodopsin in a trigonal crystal form. *J. Mol. Biol.* **2004**, *343*, 1409−1438.

(32) Moro, S.; Deflorian, F.; Bacilieri, M.; Spalluto, G. Ligand-based homology modeling as attractive tool to inspect GPCR structural plasticity. *Curr. Pharm. Des.* **2006**, *12*, 2175−2185.

(33) Laskowski, R. A.; MacArthur, M. W.; Moss, D. S. T. J. M. PROCHECK: A program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **1993**, *26*, 283−291.

(34) Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. FlexE: Efficient molecular docking considering protein structure variations. *J. Mol. Biol.* **2001**, *308*, 377−395.

(35) Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem.* **1985**, *28*, 849−857.

(36) Briscoe, C. P.; Tadayyon, M.; Andrews, J. L.; Benson, W. G.; Chambers, J. K.; Eilert, M. M.; Ellis, C.; Elshourbagy, N. A.; Goetz, A. S.; Minnick, D. T.; Murdock, P. R.; Sauls, H. R., Jr.; Shabon, U.; Spinage, L. D.; Strum, J. C.; Szekeres, P. G.; Tan, K. B.; Way, J. M.; Ignar, D. M.; Wilson, S.; Muir, A. I. The orphan G protein-coupled receptor GPR40 is activated by medium and long chain fatty acids. *J. Biol. Chem.* **2003**, *278*, 11303−11311.

(37) Yonezawa, T.; Katoh, K.; Obara, Y. Existence of GPR40 functioning in a human breast cancer cell line, MCF-7. *Biochem. Biophys. Res. Commun.* **2004**, *314*, 805−809.

(38) Levitt, M.; Perutz, M. F. Aromatic rings act as hydrogen bond acceptors. *J. Mol. Biol.* **1988**, *201*, 751−754.

(39) Hoffmann, C.; Moro, S.; Nicholas, R. A.; Harden, T. K.; Jacobson, K. A. The role of amino acids in extracellular loops of the human P2Y1 receptor in surface expression and activation processes. *J. Biol. Chem.* **1999**, *274*, 14639−14647.

(40) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. 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.

(41) Henikoff, S.; Henikoff, J. G. Amino acid substitution matrices from protein blocks. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 10915−10919.

(42) Felsenstein, J. *PHYLIP Phylogeny Inference Package,* version 3.6; Department of Genome Sciences, University of Washington: Seattle, WA, 2002.

(43) Dayhoff, M. O. *Atlas of Protein Sequence and Structure;* National Biomedical Research Foundation: Washington, DC, 1978; Vol. 5, Suppl. 3, pp 345−352.

(44) Saitou, N.; Nei, M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **1987**, *4*, 406−425.

(45) Page, R. D. M. TREEVIEW: An application to display phylogenetic trees on personal computers. *Comput. Applicat. Biosci.* **1996**, *12*, 357−358.

(46) Sali, A.; Blundell, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779−815.

(47) Sali, A.; Overington, J. P. Derivation of rules for comparative protein modeling from a database of protein structure alignments. *Protein Sci.* **1994**, *3*, 1582−1596.

(48) *InsightII,* Version 2001, Accelrys Inc., 2006.

(49) Mohamadi, F.; Richards, N. G. J.; Guida, W. C.; Liskamp, R.; Lipton, M.; Caufield, C.; Chang, G.; Hendrickson, T.; Still, W. C. Macro-Model-An Integrated Software System for Modeling Organic and Bioorganic Molecules Using Molecular Mechanics. *J. Comput. Chem.* **1990**, *11*, 440−467.

(50) Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.* **2006**, *261*, 470−489.

(51) *SYBYL;* 6.91 ed.; TRIPOS, Assoc., Inc.: St. Louis, MO, 2006.

(52) SPARTAN 5.1; Wavefunction, Inc., Irvine, CA, 2006.

(53) Sawzdargo, M.; George, S. R.; Nguyen, T.; Xu, S.; Kolakowski, L. F.; O'Dowd, B. F. A cluster of four novel human G protein-coupled receptor genes occurring in close proximity to CD22 gene on chromosome 19q13.1. *Biochem. Biophys. Res. Commun.* **1997**, *239*, 543−547.