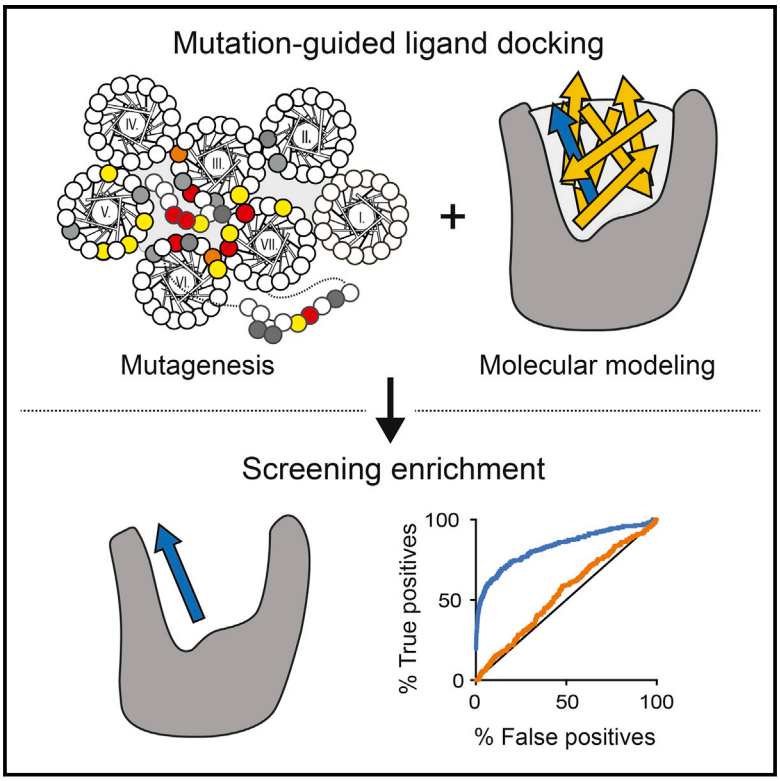
Resource

Structure

Mutation-Guided Unbiased Modeling of the Fat Sensor GPR119 for High-Yield Agonist Screening

Graphical Abstract



Authors

Christoffer Norn, Maria Hauge, Maja S. Engelstoft, ..., Robert M. Jones, Thue W. Schwartz, Thomas M. Frimurer

Correspondence

thomas.frimurer@cpr.ku.dk

In Brief

Norn et al. demonstrate that a data-driven GPCR-ligand docking protocol, based on a library of mutational data, can explain ligand structure-activity relationship. In contrast to standard comparative receptor modeling, the resulting models are of practical use for virtual screening in drug discovery applications.

Highlights

- A comprehensive library of site-directed GPR119 mutations
- Unbiased mutation-guided refinement of receptor-ligand complexes
- Modeling of the prototype agonist AR231453 in complex with GPR119
- Discovery of an inverse GPR119 agonist, AR437948

Norn et al., 2015, Structure *23*, 2377–2386
December 1, 2015
http://dx.doi.org/10.1016/j.str.2015.09.014

Structure
Resource

Mutation-Guided Unbiased Modeling of the Fat Sensor GPR119 for High-Yield Agonist Screening

Christoffer Norn,[2,3] Maria Hauge,[1,2] Maja S. Engelstoft,[1,2,4] Sun Hee Kim,[5] Juerg Lehmann,[5] Robert M. Jones,[5] Thue W. Schwartz,[1,2] and Thomas M. Frimurer[2,3,*]
[1]Laboratory for Molecular Pharmacology, Department of Neuroscience and Pharmacology, University of Copenhagen, Blegdamsvej 3, Copenhagen 2200, Denmark
[2]NNF Center for Basic Metabolic Research, University of Copenhagen, Blegdamsvej 3, Copenhagen 2200, Denmark
[3]NNF Center for Protein Research, University of Copenhagen, Blegdamsvej 3, Copenhagen 2200, Denmark
[4]Danish Diabetes Academy, Sdr Boulevard 29, Odense C 5000, Denmark
[5]Medicinal Chemistry Department, Arena Pharmaceuticals, 6154 Nancy Ridge Drive, San Diego, CA 92121, USA
*Correspondence: thomas.frimurer@cpr.ku.dk
http://dx.doi.org/10.1016/j.str.2015.09.014

SUMMARY

Recent benchmark studies have demonstrated the difficulties in obtaining accurate predictions of ligand binding conformations to comparative models of G-protein-coupled receptors. We have developed a data-driven optimization protocol, which integrates mutational data and structural information from multiple X-ray receptor structures in combination with a fully flexible ligand docking protocol to determine the binding conformation of AR231453, a small-molecule agonist, in the GPR119 receptor. Resulting models converge to one conformation that explains the majority of data from mutation studies and is consistent with the structure-activity relationship for a large number of AR231453 analogs. Another key property of the refined models is their success in separating active ligands from decoys in a large-scale virtual screening. These results demonstrate that mutation-guided receptor modeling can provide predictions of practical value for describing receptor-ligand interactions and drug discovery.

INTRODUCTION

Proteins of the G-protein-coupled receptor (GPCR) superfamily regulate a broad range of physiological processes (Ritter and Hall, 2009) and present attractive targets for drug discovery in multiple therapeutic areas (Wise et al., 2002). Over the past few years more than 20 distinct GPCRs have been crystallized with a number of small-molecule antagonists, agonists, and a G protein (Katritch et al., 2013; Rasmussen et al., 2011; Stevens et al., 2013). Whereas the drug discovery process greatly benefits from analysis of the structures of target receptors and their interactions with ligands (Carlsson et al., 2010, 2011; Katritch et al., 2010a; Kolb et al., 2009), this process has been limited on targets lacking structural information.

Even with the increase in structural information, recent assessment and benchmark studies of GPCR structure modeling and ligand docking methodologies have demonstrated widespread difficulty in predicting native ligand binding conformations to comparative GPCR models (Kufareva et al., 2011, 2014; Michino et al., 2009; Nguyen et al., 2013). This is due to the intrinsic flexibility of the seven transmembrane helices (Deupi and Kobilka, 2010), water-filled pockets (Katritch et al., 2013), and the involvement of extracellular loops in ligand binding (Beuming and Sherman, 2012; De Graaf et al., 2008; Shi and Javitch, 2004). In general, accurate predictions were only achieved when models were built on related template structures (sequence similarity >35%) (Katritch et al., 2010a). With an estimated >360 pharmaceutically relevant GPCRs in the human genome (Vassilatis et al., 2003), it is anticipated that only up to 15% of the GPCR family can be reliably modeled based on the current distinct experimental GPCR structures, signifying the need for improved modeling methodologies for receptor targets lacking closely related template structures.

Realizing these challenges, data-driven methods, which incorporate experimental information such as ligand structure-activity relationships (SAR) (Katritch et al., 2010b), multiple template structures (Kneissl et al., 2009), or mutational data (Michino et al., 2009) have been proposed to guide the modeling and refinement of receptor-ligand complexes.

In this study, we have developed a novel mutation-guided docking protocol that uses the RosettaLigand docking protocol (Kaufmann and Meiler, 2012; Lemmon and Meiler, 2011) with full ligand and receptor flexibility in combination with experimental data to guide the modeling and refinement of receptor-ligand complexes. The protocol integrates (1) structural information from multiple distinct high-resolution X-ray structures to create a knowledge-based ensemble of receptor-ligand complexes that captures the intrinsic flexibility of the known crystal structures, and (2) a large experimental library of mutated receptor variants to guide and enhance the selection of native-like receptor-ligand complexes based on correlations between predicted binding energies and experimentally determined potency shifts for all receptor variants.

We applied this method to refine models of the prototype agonist AR231453 (Figure 1) in complex with the

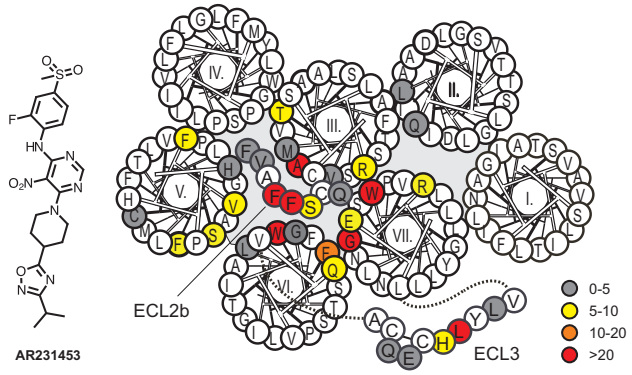


Figure 1. Chemical Structure of the GPR119 Agonist, AR231453, and a Helical Wheel Diagram of the Human GPR119 Receptor

ECL2b contains residues C-terminal to the cysteine in ECL2. Mutated residue positions and the effect on ligand potency as determined in a cAMP assay (fold change) are colored gray <5; 5 ≤ yellow < 10; 10 ≤ orange < 20; ≥20 red.

G-protein-coupled receptor 119 (GPR119), a lipid-responsive class A receptor (Fredriksson et al., 2003). GPR119 has received significant interest as a target for the treatment of type 2 diabetes (Chu et al., 2007, 2008, 2010; Fyfe et al., 2008), with small-molecule agonists (e.g. AR231453) reported to promote pancreatic postprandial insulin and incretin secretion in a glucose-dependent fashion (Semple et al., 2011). In a previous study we applied classical docking methods to elucidate the binding mode of AR231453 to GPR119, but were unable to discriminate between multiple possible binding conformations (Engelstoft et al., 2014).

In this study we find a strong correlation around one binding conformation, which explains the majority of the mutations and is in agreement with SAR for a large number of AR231453 analogs. Interestingly the mutation-guided models, in contrast to simple homology models, proved successful in separating active ligands from decoys in a large-scale retrospective structure-based virtual ligand screening. This method of integrating experimental data in combination with fully flexible receptor-ligand docking simulations may serve as an important tool for structure-function analysis and guide the discovery of novel active chemotypes.

RESULTS

Modeling of the AR231453 GPR119 Receptor Complex

Full-length models of the human GPR119 receptor were generated using a multi-template approach by combining fragment replacement and structural restraints from multiple GPCR X-ray structures (see Supplemental Information). To account for the substantial structural variability of ligand binding pocket conformations in distinct GPCR crystal structures, we produced a number of energetically feasible receptor conformations by minimizing the energy of the backbone and side chains using the Rosetta relax protocol and the membrane force field (Barth et al., 2007; Leaver-Fay et al., 2011). Subsequent clustering based on binding site residues in contact with ligands in the template structures was used to generate a non-redundant low-energy ensemble of 43 receptor models, which were used as input in the docking simulations to broadly sample the binding pocket geometry.

A conformational ensemble of low-energy structures of AR231453 (see Supplemental Information) was docked into the receptor ensemble using the RosettaLigand docking protocol that samples both receptor and ligand flexibility (Kaufmann and Meiler, 2012; Lemmon and Meiler, 2011). A total of 43,000 receptor-ligand complexes were generated to extensively sample receptor-ligand binding conformations. To ensure that receptor flexibility of the binding pocket was sampled within the structural variations observed in the experimental structures, we discarded conformations in which Cα atoms of binding pocket residues deviated by more than four standard deviations from the variation in the corresponding Cα atom positions in 16 distinct experimental GPCR structures (see Supplemental Information and Figure S2). The tolerated structural variability is illustrated in Figure 2A, which shows that some helices, especially the top of transmembrane helix (TM-II), can assume several different orientations, while others are more static.

Docking Simulations Converge to Two Binding Orientations

To analyze the ligand binding space we measured the distance from the sulfonyl moiety to the bottom of the binding pocket, and the distance of the isopropyl group to the loop region in the 43,000 generated receptor-ligand complexes. The ligands in this ensemble assumed widely different binding conformations, as illustrated by the 100 randomly selected ligand orientations shown in gray in Figure 2B.

Ligand interaction energies were applied to reject loosely packed binding conformations expected to be biologically irrelevant. We examined the resulting ensemble by analyzing the ten most frequently occurring binding conformations. This revealed two major ligand binding orientations, one in which the sulfonyl group is in the loop region (SO_2-out, illustrated in Figure 2B) and the other where the ligand is flipped 180° so that the sulfonyl is in the bottom of the binding site (SO_2-in). Within each orientation multiple conformations were seen. The frequency and computed ligand binding energy of the two binding orientations were similar. From the models alone it was difficult to discriminate between the two binding orientations, indicating that flexibility filters alone are not sufficient to reach convergence.

Integration of Mutational Mapping Data Exhibits Preference for One of the Binding Conformations

To further improve convergence of the predicted receptor-ligand complexes, we developed a mutation-guided optimization protocol that correlates the binding affinity associated with a library of GPR119 mutations with the computationally estimated binding energies. The analysis included 33 non-glycine, non-proline, single-point mutations located in the ligand binding pocket (Figure 1 and Table 1).

The ligand binding energy associated with each mutation would ideally be determined by measuring K_D in a ligand binding assay. However, since direct ligand binding assays are currently unavailable for GPR119 we instead used a cyclic AMP (cAMP) assay to determine the potency (EC_{50}). By assuming a constant mutation-independent relationship between K_D and EC_{50}, it is anticipated that potency shifts provide a good surrogate measure of relative binding energies and correlation-based score. However, this relationship might be different when receptor

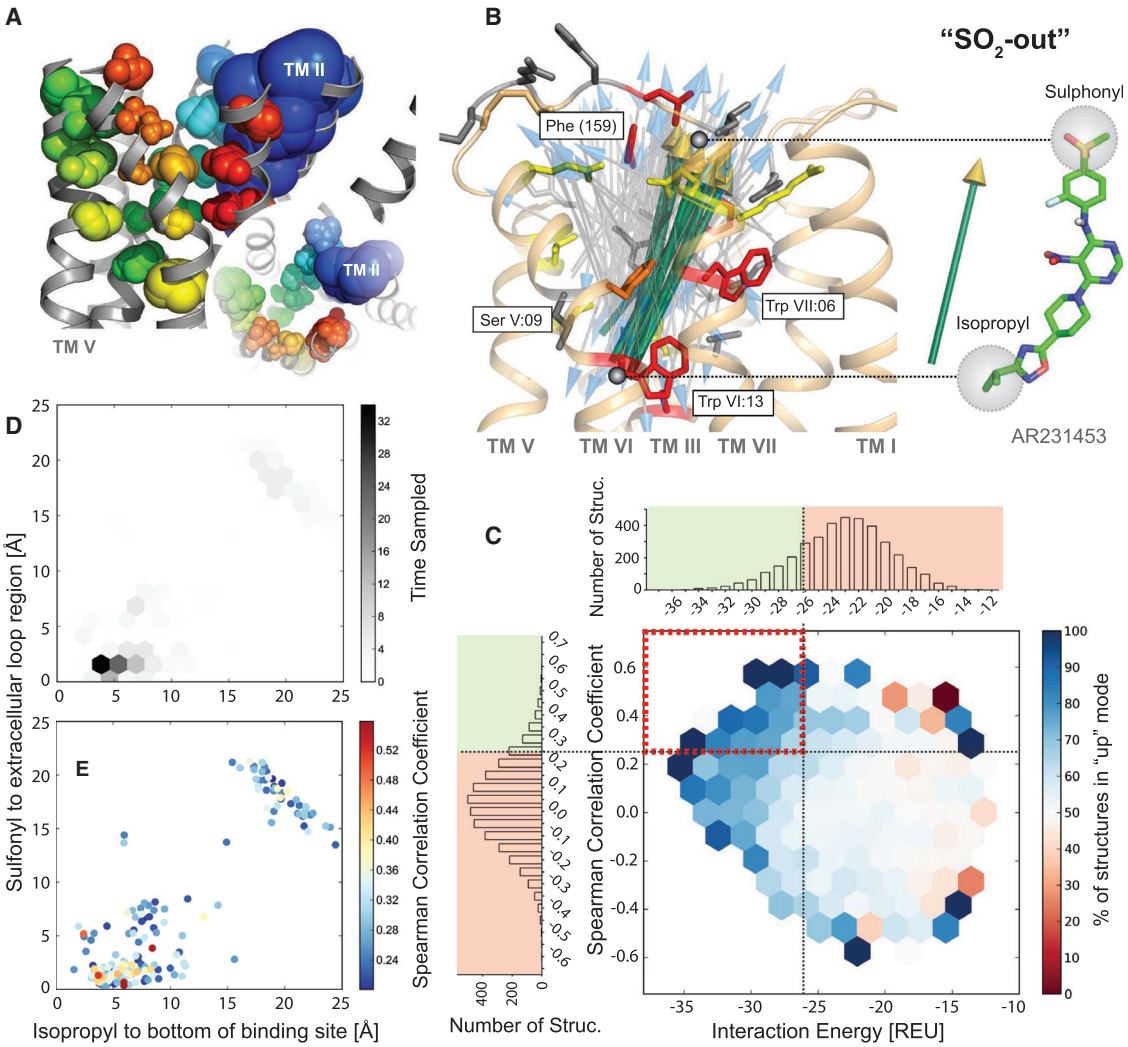


Figure 2. Restricted Sampling of Receptor Flexibility Together with Mutation-Guided Methods Converges to One Major Binding Conformation of AR231453 to GPR119

(A) Biologically irrelevant receptor conformations were removed by filtering the generated GPR119/AR231453 ensemble using residue-specific limits for deviations from known GPCR crystal structures (see Supplemental Information). These limits of flexibility are illustrated as spheres on the β2-adrenergic receptor.

(B) In contrast to the filtered ensemble (green arrows with golden heads), the unfiltered ensemble yielded no strong preference for a particular binding conformation (gray arrows with blue heads). The arrowhead points to the sulfonyl (S atom) of the ligand, while the tail is at the isopropyl moiety. A subset of mutated residues are shown as sticks and colored according to potency shift upon mutation (Figure 1).

(C–E) To further remove biologically irrelevant binding conformations, we mapped the binding space based on the ligand orientation as a function of computed interaction energy (Rosetta Energy Units) and the correlation coefficient between experimental and computational binding energy. The ligand orientation was described as SO_2-out when the sulfonyl group (arrowhead) was pointing to the extracellular side. Structures in the best-scoring quadrant (red dashed square in C) and the SO_2-out were both enriched with respect to frequency of sampling (D), and achieved higher correlation coefficients compared than the SO_2-in conformation (E).

See also Figures S2 and S3.

expression is compromised, and on this behalf we excluded four variants (165Phe$^{5.38}$Leu V:04, 165Phe$^{5.38}$Ala V:04, 174Phe$^{5.49}$Ala V:13, and 154Gln^{ECL2}Ala 11) that decrease expression level to less than 20% of the wild-type receptor (residues are numbered according to the GPCR numbering system of Schwartz et al., 1995 and Ballesteros and Weinstein, 1995). Mutations that alter the functional integrity of the receptor, for example by disallowing the active state, would also lead to misinterpretation. To account for this we discarded one variant (238Trp$^{6.48}$Ala VI:13) with no constitutive activity that could not be activated by the ligand. In addition we discarded one variant (81Arg$^{3.27}$Ala III:03) that was not in the ligand binding pocket. This left us with a set of 28 GPR119 variants. The strongest decrease in receptor activation was observed for the mutated residues located in the extracellular segments of TM-III, -V, -VI, and -VII, as well as in ECL2 and ECL3, as illustrated in Figure 1.

For each of the 43,000 receptor-ligand models we then computed the Spearman correlation coefficient (SCC) between

Table 1. GPR119 Mutagenesis Data for AR231453

	Surface Expression		Basal Activity		Max. Efficacy		AR231453 EC50					
GPR119 Variant	%	SEM	%	SEM	%	SEM	Log	SEM	nM	Potency Shift	n	Assay
Wild-type	100	0.9	37	0.7	100	0.9	−8.73	0.03	1.9		26	A
Wild-type	100	0	56.5	1.8	99.4	1.9	−8.74	0.34	1.9		9	H
Binding Pocket												
61Leu$^{2.60}$Arg II:20*	29	4.4	30	5.1	70	5.2	−8.62	0.61	2.4	1	3	H
65Gln$^{2.64}$Ala II:24*	122	9.4	31	5.3	99	7.1	−8.9	0.24	1.2	1	4	A
81Arg$^{3.28}$Ala III:04	51	2.4	1.9	1.5	41	2.5	−7.94	0.15	11.6	6	4	A
82Met$^{3.29}$Ala III:05*	78	5.8	8.3	1.4	24	1.9	−8.37	0.19	4.3	2	4	A
85Val$^{3.32}$Ala III:08*	115	9.2	29	6	88	7.2	−8.8	0.28	1.6	1	4	A
86Thr$^{3.33}$Ala III:09*	55	12	13	2.1	39	4.5	−7.8	0.36	15.8	8	4	A
86Thr$^{3.33}$Val III:09*	71	14	32	7.2	61	6.2	−8.9	0.63	1.3	1	4	A
89Ala$^{3.36}$Val III:12*	231	43	27	6	90	32	−6.99	0.81	103	54	4	A
162His$^{5.35}$Ala V:01*	58	4.3	20	1.9	73	2.4	−8.75	0.08	1.8	1	3	A
165Phe$^{5.38}$Ala V:04	2	2	8	1.8	28	2.9	−7.84	0.58	14	7	3	H
165Phe$^{5.38}$Leu V:04	7	3	7.6	2.4	27	3.1	−8.34	0.13	4.6	2	3	H
166Val$^{5.39}$Ala V:05*	53	9.1	35	6.6	70	8.5	−8.03	0.61	9.3	5	3	A
170Ser$^{5.43}$Ala V:09*	157	17	57	7.5	108	7.4	−8.64	0.38	2.3	1	4	A
170Ser$^{5.43}$Val V:09*	164	30	12	3.5	77	5	−8.02	0.16	9.7	5	4	A
174Phe$^{5.47}$Ala V:13	20	4.5	26	3	53	4.3	−8.24	0.18	5.7	3	2	H
238Trp$^{6.48}$Ala VI:13	40	2.6	–	–	–	–	–	–	>10,000	>5,000	4	A
241Phe$^{6.51}$Ala VI:16*	72	4.7	11	1.8	27	4.4	−7.47	0.5	34	18	3	A
242Leu$^{6.52}$Ala VI:17*	39	4.9	55	3	103	3.6	−8.52	0.08	3	2	3	H
248Gln$^{6.58}$Ala VI:23*	111	18	31	7.9	87	11	−8.07	0.41	8.6	5	3	A
261Glu$^{7.35}$Ala VII:02*	54	5.1	13	1.3	54	3.3	−7.75	0.16	17.7	9	4	A
262Arg$^{7.36}$Ala VII:03*	50	4.1	−0.1	0.7	18	1.2	−8.01	0.14	9.8	5	4	A
265Trp$^{7.39}$Ala VII:06*	40	2.8	34	2.7	–	–	–	–	>10,000	>5,000	5	A
ECL2a												
Gln11Ala (154)	16	6.3	28	2.7	63	2.8	−8.9	0.24	1.3	1	3	H
ECL2b												
156Ser$^{cys155+1}$Ala*	71	3.5	6.9	1.3	48	2.2	−8.44	0.12	3.7	2	3	A
157Phe$^{cys155+2}$Ala*	69	4.4	6.4	0.5	23	1.7	−6.83	0.18	147	77	5	A
158Phe$^{cys155+3}$Ala*	73	3.6	5.1	0.6	24	1.5	−7.19	0.13	65	34	5	A
160Val$^{cys155+5}$Ala*	66	3.4	5.7	0.9	34	1.4	−8.13	0.11	7.4	4	3	A
161Phe$^{cys155+6}$Ala*	56	2.3	5.6	1	21	1.2	−8.65	0.15	2.2	1	3	A
ECL3												
252Gln$^{6.62}$Ala*	51	6.4	49	8.8	83	9.9	−8.69	0.46	2.1	1	3	H
253Glu$^{6.63}$Ala*	31	5.9	48	4.8	99	5.5	−8.7	0.18	2	1	3	H
255His$^{6.64}$Ala*	43	10.4	33	4.2	76	4.7	−8.76	0.13	1.7	1	3	H
256Leu$^{6.65}$Ala*	44	14.8	54	2	96	6	−7.19	0.35	65	34	4	H
258Leu$^{6.67}$Ala*	48	7.5	44	6.3	96	7.6	−8.52	0.18	3	2	3	H

Surface expression, basal activity, efficacy, potency, and potency fold change for wild-type human GPR119 and receptor mutants. Residues used for correlation analysis are marked by an asterisk. Two assays were used: Alumnia (A) and Hithunter (H). We described data derived from the Alumnia assay in a previous study (Engelstoft et al., 2014).

the computationally predicted and experimentally estimated binding energies for each of the 28 mutated receptor variants. The density of ligand binding orientation as a function of interaction energy and correlation coefficient for the generated complex ensemble is shown in Figure 2C.

We note that despite the exclusion of mutations with decreased expression level or lack of constitutive activity, a strong correlation between experimental potency shifts and computational binding energies cannot be expected due to the inherent approximations of the measure. Therefore, when further analyzing the structures we applied a conservative correlation cutoff of 0.25, while taking the 10% most energetically favorable models and discarding all structures violating the allowed backbone flexibility (Figure 2D). Clustering the resulting 207

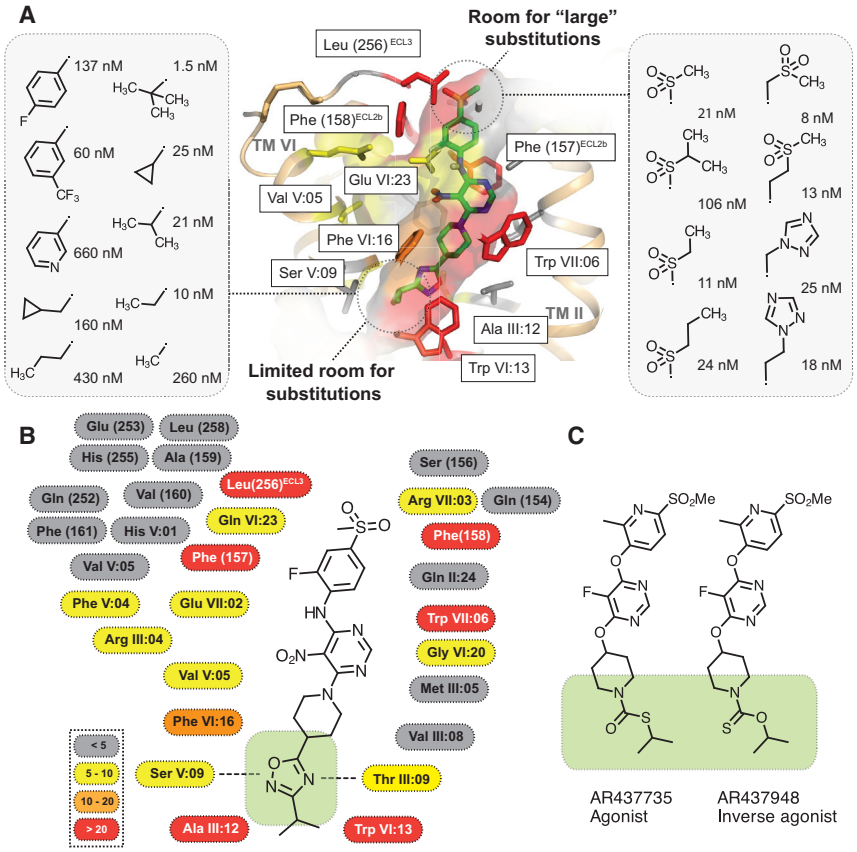


Figure 3. The Mutational Data and SAR Support the Refined Ligand Binding Conformation

(A) The predicted binding mode of AR231453 is consistent with the SAR of ligands. The most likely GPR119/AR231453 complex with residues are colored according to potency shift on AR231453 upon mutation (gray <5; 5 ≤ yellow < 10; 10 ≤ orange < 20; ≥20 red). The sulfonyl group is situated in the extracellular part of the receptor, with room for large substitutions, which only have a minor effect on ligand potency. Conversely, substitutions are not tolerated at the bottom of the binding pocket, which packs tightly around the ligand. The surface of the binding pocket is colored according to the potency shift of the nearest residue.

(B) Ligplot of predicted AR231453 GPR119 interactions (Table 1). AR231453 is predominantly in contact with residues affecting the potency upon mutation. Mutations that did not affect the potency of AR231453 (gray) could in most cases be explained by their remote position relative to AR231453 in the models. The isopropyl-oxadiazole moiety is involved in H-bond interactions with 86Thr$^{3.33}$ III:09, 170Ser$^{5.42}$ V:09.

(C) The chemical moieties responsible for the distinct pharmacology of AR437735 (agonist) and AR437948 (inverse agonist) mimic the common oxadiazole moiety found in many GPR119 agonists (green shaded area), predicted to be involved in different interactions with 86Thr$^{3.33}$ III:09, 170Ser$^{5.42}$ V:09.

See also Figure S4.

structures with a clustering radius of 2.5 Å root-mean-square deviation (RMSD) for the ligand position resulted in 73 clusters. The ten most populated clusters were strongly preferred and numbered approximately half of the structures. Among these, 95% of the models were in the SO_2-out binding conformation (Figure 2B, green arrows), which also had a better binding energy score and higher correlation coefficients with the experimental data than the SO_2-in binding conformation (Figure 2E). Representative low-energy models with high SCC scores from the five most populated binding modes all shared the SO_2-out binding conformation (Figure S3). Corresponding PDB files of the complexes are provided in the Supplemental Information.

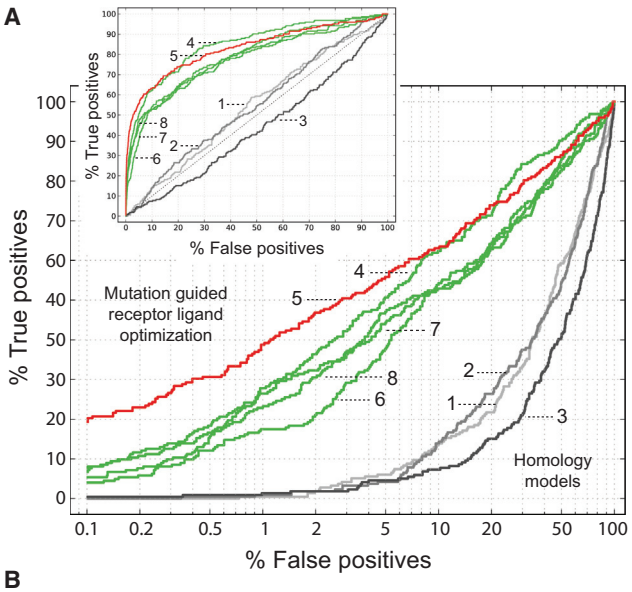
Mutagenesis Data and Ligand SAR Support Predicted Binding Modes of AR231453 in GPR119

Due to the inherent inaccuracy of the force field and the use of potency shifts as a proxy for the actual binding energy, it is unclear whether any of the models would have atomic accuracy, and we do indeed see variations between the different clusters in the SO_2-out conformation. The structures from the largest clusters do, however, generally occupy the same binding pocket surrounded by key residues situated at helices II, V, VI, VII, and the extracellular loops.

Figure 3 shows a representative binding conformation from the largest cluster where the ligand forms very few contacts to residues not affecting ligand potency, but forms contacts to all residues affecting ligand potency by more than 20-fold. The two phenylalanines of ECL2 (157F$^{Cys155+2}$ and 158F$^{Cys155+3}$) and the leucine of ECL3 (L256A) pack tightly around the 2-fluoro-4-methanesulfonyl-phenyl moiety, but leaves the methyl group exposed. This is consistent with ligand SAR previously reported by Semple et al. (2008), who showed that large substitutions can be made to the methyl moiety without dramatic effects on ligand potency (Figure 3). Notably, this SAR further validates the SO_2-out binding conformation in contrast to the SO_2-in, which is incompatible with large substitutions at the SO_2 group.

The 5-nitro-pyrimidine core at the middle of the ligand is loosely packed at the binding pocket in contrast to the 4-methanesulfonylphenyl and oxadiazole termini (see below), which appear to be more tightly bound. Looking across the largest clusters we see some variation in the exact location of this group. In many cases, it is packed against a phenylalanine (241Phe$^{6.51}$ VI:16) and a tryptophan (265Trp$^{7.39}$ VII:06), with the nitro group exposed to a water-filled cavity. In the most common binding conformation, the aniline linker makes H-bond donor interactions with the glutamate (261Glu$^{7.35}$ VII:02), supporting a 9-fold change in ligand potency upon mutation to alanine. In other clusters 261Glu$^{7.35}$ VII:02 is involved in internal residue-residue interactions with the spatial nearby residues 248Gln$^{6.58}$ VI:23 and 262Arg$^{7.36}$ VII:03, which moderately affects potency when mutated to alanine. The loose packing of the 5-nitro-pyrimidine core is also to be expected, as many GPR119 agonists differ greatly in their middle portions (Table S1), while the end

A



B

Initial homology models	AUC	EF(1%)	EF(10%)
1. Raw homology model (A2aAR template)	0.54	0.46	1.38
2. Hybrid GPR119 model	0.54	0.92	1.38
3. Refined model (Rosetta relax)	0.43	1.35	0.72
Best performing mutation guided models			
4. Mutation guided model (cluster 7)	0.85	26.0	6.18
5. Mutation guided model (cluster 2)	0.83	36.8	6.28
6. Mutation guided model (cluster 8)	0.80	17.0	5.25
7. Mutation guided model (cluster 1)	0.78	25.6	5.29
8. Mutation guided model (cluster 1)	0.78	22.9	5.38

Figure 4. Refined Receptor Models Enrichment of Active Compounds in an In Silico Screen

(A) Percentage of recovered actives as function of screened compounds is shown for models from different stages of the protocol. The initial homology models (gray) had a lower docking performance compared with the five best-performing models (red and green) derived from the mutation-guided optimization protocol. Inset shows a normal view of the ROC curves. The false-positive rate as a function of screened compounds is shown on a logarithmic scale to highlight the initial enrichment at 1% of the dataset.

(B) Summary of key characteristics of the corresponding ROC curves and virtual screening performance.

See also Table S1.

termini are often the same—namely a methanesulfonyl at one end, and an aminooxadiazole or carbamate at the other. Because these polar side chains have the ability to both make H-bond interactions with each other and/or a ligand, the models in the largest clusters support the binding of different 5-nitropyrimidine substitutions of, for example, the nitro group and an N/O/CH_2 linker.

The oxadiazole moiety is located at the bottom of the binding pocket surrounded by 170Ser$^{5.31}$ V:09, 86Thr$^{3.33}$ III:09, and 89Ala$^{3.36}$ III:12. The mutational data suggest that there is limited space available, as a mutation of alanine to valine results in a 50-fold decrease of potency. This behavior is also consistent with ligand SAR, which allows limited extension of the isopropyl group, consistent with a small pocket that can be recognized in some of the models, but disallows larger substitutions. Again, ligand SAR favors the SO_2-out conformation over the SO_2-in.

Related Binding Conformation of an Agonist and an Inverse Agonist

In the optimization process of pharmacologically attractive GPR119 agonists, we identified ligands for which single atom substitutions could change the efficacy from agonist (AR437735) to inverse agonist (AR437948) (Figures 3C and S4; Supplemental Information). The models and the similarity of the ligands with AR231453 suggest that they bind in a similar fashion as AR231453 (Figure 3B). The chemical moieties responsible for the distinct pharmacology mimics the common oxadiazole moiety found in many GPR119 agonists (Semple et al., 2011), predicted to be involved in interactions with 86Thr$^{3.33}$ III:09 and 170Ser$^{5.42}$ V:09. That the activity of the ligand can be modulated with a thioester/thionoester substitution is perhaps not entirely surprising, as the energy difference between active and inactive states must be fairly small in GPR119 since it is a constitutively active receptor. It could thus be possible that the minor difference in H-bond acceptor properties between the two ligands is sufficient to stabilize two different conformations. Another possibility is that the ligands adopt different conformations in the receptor. Indeed, a conformational analysis reveals that both ligands are isoenergetic (<0.1 kcal/mol) with respect to having the thioester/thionoester axial or equatorial to the piperidine ring. A similar conformation-based hypothesis has also been proposed by McClure et al. (2011) to describe the agonist response of a related GPR119 agonist.

Mutation-Guided Receptor-Ligand Models Dramatically Improve Virtual Screening Performance Related to Drug Discovery Applications

To further validate the GPR119 models, we evaluated their performance in virtual ligand screening for their ability to discriminate known actives from decoys. We created a library of 223 known GPR119 active ligands (Table S1) extracted from ChEMBL (Gaulton et al., 2012) and 91,712 screening compounds (decoys) with similar molecular property distribution as active compounds extracted from vendor databases (see Supplemental Information). We identified the five best-performing models from the ten largest clusters in terms of active compound scoring with SurflexDock (Jain, 2003). Active and decoy compounds were then docked against those five models. The performance was evaluated based on the percentage of recovered active compounds, the corresponding area under the receiver-operating characteristic (ROC) curve (AUC), and the computed enrichment factors (EF) for the top 1% and 10% ranked compounds. Only suboptimal docking performance and low selectivity of known active compounds (AUC = 0.54, $EF_{1\%}$ = 0.5, $EF_{10\%}$ = 1.4) were achieved for the initial homology model based on the closest related A2aAR template structure (Figure 4). Equally low selectivity (AUC = 0.54, $EF_{1\%}$ = 0.9, $EF_{10\%}$ = 1.4) was found for the hybrid homology model. The ensemble of relaxed models used as input in the docking simulations all performed worse than random and revealed poor docking scores of actives as well as decoys. This might be related to a partial collapse of the ligand binding pocket as a result of the initial receptor optimization in the absence of a ligand, and provides a warning against the use of simple homology models or relaxed homology models in in silico screening.

In contrast, the models developed through the mutation-guided docking protocol demonstrated a dramatic improvement in virtual screening performance compared with the initial homology models and random baseline (Figures 4 and S4). The five best-performing models represented members of some of the largest clusters and were all in the SO_2-out binding orientation. The overall best-performing model demonstrated good virtual screening performance (AUC = 0.85) and early enrichment ($EF_{1\%}$ = 26.0 and $EF_{10\%}$ = 6.18). The second best-performing model (AUC = 0.83), representing the second most populated docking mode, was especially efficient in respect of initial enrichment ($EF_{1\%}$ = 36.8). Good initial enrichment ($EF_{1\%}$ = 25.6) was also achieved for the most populated binding mode. In comparison, the best SO_2-in model had a poor screening performance (AUC = 0.65) and early enrichment ($EF_{1\%}$ = 13.9). Notably, these screening performances are comparable with the virtual screening performance that can be expected from a high-resolution crystal structure (Spitzer and Jain, 2012), suggesting that the mutation-guided models might be valuable in virtual screening applications aimed at identifying novel active chemotypes. Indeed, among the top 1% scoring compounds the virtual screening did not only enrich for agonists similar to AR231453, but also identified active compounds lacking the preferred sulfonyl and oxadiazole moiety. For instance, across all models described in Figure 4, 13% of the hits had both features, 75% had only one of them, and the last 12% had none. In comparison, the full dataset of active compounds contained 8% with both, 54% with only one, and 38% missing both.

DISCUSSION

This study shows that mutagenesis data can be used in a systematic and unbiased way to counter the poor force field performance of GPCR/small-molecule docking, and thus generate models in agreement with ligand SAR and of sufficient quality to support structure-based drug design. In this work we specifically focused on the GPR119 receptor in complex with a prototype agonist.

The major challenges in small-molecule docking arise from minor inaccuracies in structural models due to alignment inaccuracies, the lack of explicit water molecules, and inaccuracies of the force field, combined with the necessity for intensive conformational sampling. These inaccuracies often preclude the use of energies to reliably discriminate between correct and incorrect models (Fleishman and Baker, 2012; Lu et al., 2007). The problem is even greater for GPCRs, where the receptor itself balances between discrete substates of active and inactive conformations (Deupi and Kobilka, 2010; Manglik and Kobilka, 2014; Staus et al., 2014) that, for constitutive active receptors, can be separated by less than 1 kcal/mol. To counter this challenge, ligand structure-activity relationship data (Katritch et al., 2010b) and mutagenesis data (Kufareva et al., 2011; Michino et al., 2009; Nguyen et al., 2013) can be used to increase the accuracy of generated receptor-ligand models by complementing the energy function.

The use of mutagenesis data does, however, pose a problem, as many residues affect ligand binding indirectly and thus do not allow for a simple residue-by-residue interpretation (Jaakola et al., 2010; Kim et al., 2003). This problem is not widely appreciated in the docking literature, where manual selection of specific residues for which residue-ligand "must-have-contact" constraints are defined to reject or accept models from ligand docking ensembles. To avoid this binary approach that is prone to misinterpretation, we developed an unbiased protocol to refine and increase the accuracy of a receptor-ligand ensemble using knowledge-based filters and correlations with a large set of experimental mutation data in combination with a fully flexible receptor-ligand docking protocol. The protocol enables an unbiased selection and refinement of models best suited to describe ligand bound states, which convincingly support the majority of the mutational mapping.

The generated models cluster around one major binding conformation, which is coherent with SAR of GPR119 ligands. Minor extension at the isopropyl-oxadiazole moiety improves binding affinity by filling out a cavity at the ligand-receptor interface, while larger substitutions result in clashes and are detrimental for potency. The other end of the ligand is solvent exposed and allows for larger substitutions. The models demonstrate the ability to enrich a broad selection of diverse high-affinity GPR119 agonists (belonging to several different scaffold classes not employed in model optimization), supporting the protocol's suitability for drug discovery and structure-based drug design applications.

Comparison of the AR231453 binding conformation in GPR119 (Figure 3A) to bound ligands in recently solved GPCR structures suggests that AR231453 occupies a similar position in the transmembrane domain as agonist bound (e.g. UK-432097) A_{2A}AR structures (Xu et al., 2011). Analysis of agonist and antagonist bound A_{2A}AR and β1AR structures suggests that agonists (in contrast to antagonists) makes distinct interactions with polar residues in the lower part of the binding pocket (just above the conserved $Trp^{6.48}$ VI:13) in both A_{2A}AR ($Ser^{7.42}$ VII:09 and $His^{7.43}$ VII:10) (Katritch et al., 2010a; Lebon et al., 2011a; Xu et al., 2011) and β1AR ($Ser^{5.46}$ V:12) (Warne et al., 2011) to promote or stabilize an active receptor state (Dal Ben et al., 2010; Kim et al., 2003; Lebon et al., 2011b). Interestingly, the atom substitutions resulting in the efficacy switch from agonist (AR437735) to inverse agonist (AR437948) occur in the same area as the residues involved in H-bond interactions responsible for triggering the agonist response in A_{2A}AR and β1AR (Figure 3C). At present the extent and duration of activation of the receptor needed for maximum clinical benefit are not known. Chemical modification to this portion of the molecule could prove important for agonist design to identify pharmacologically superior agonists.

To further the use of mutational data in ligand docking, we are currently developing methods that will allow on-the-fly optimization of correlation with experimental data. Likewise, further development of position-specific backbone constraints for homology model construction is expected to be useful, not only for GPCR homology modeling but for flexible backbone modeling of any structure from a family with multiple solved structures.

Conclusion

In the absence of a crystal structure, theoretical modeling and mutagenesis analysis are important tools in characterizing the

structural basis of ligand binding. We have developed an unbiased mutation-guided docking protocol in combination with RosettaLigand to characterize the structural basis for ligand binding to the GPR119 receptor. This protocol uses intrinsic flexibility of distinct GPCR crystal structures and correlations between predicted binding energies and experimentally determined potency shifts for a large number of mutated receptor variants. The generated models of ligand binding to GPR119 are consistent with mutational data and SAR, and are sufficiently accurate to support structure-based design efforts for this target. Although the current mutation-guided simulation protocol was limited to GPR119, the protocol can be translated to other receptors and possibly provide important guidance for GPCR modeling in general.

EXPERIMENTAL PROCEDURES

Experimental Determination of Binding Potency

AR231453 was synthesized as previously described (Semple et al., 2008). Synthesis of AR437735 and AR437948 is described in Supplemental Experimental Procedures. The construction of the GPR119 variants and effect of the mutations on the constitutive and ligand-induced (AR231453) receptor activation was determined with cAMP accumulation assays in transfected COS7 cells as described in Supplemental Experimental Procedures. Surface expression and basal activity of human GPR119 wild-type and receptor mutants along with efficacy and potency of AR231453, as well as the potency fold change between wild-type and mutants, are listed in Table 1.

Construction of Comparative GPR119 Homology Models

We constructed two GPR119 models assembled by fragments from three distinct class A GPCRs including the adenosine A2a receptor (PDB: 3EML), the CXCR4 (PDB: 3ODU) receptor, and the dopamine D3 (PDB: 3PBL) receptor (Figure S1). Models were constructed using dopamine D3 to model ECL3, CXCR4 to model ECL2a, and A2a or CXCR4 to model TM IV, while the rest of the receptor was modeled using A2a. Side chains for all residues were optimized, and the models were subsequently minimized using 200 steps of steepest descent and 300 steps of conjugated gradient-energy minimization as implemented in ICM version 3.7-2 (Molsoft). The initial GPR119 homology models were refined using Rosetta version 3.2.1. Specifically, the seven initial homology models were subjected to 1,000 steps of full-atom structure relaxation using the membrane force field (Barth et al., 2007). During the relax protocol, a disulfide bridge between $78Cys^{3.25}$ III:01 and $155Cys^{ECL2b}$ in the second extracellular loop was specified. The best 300 (top 30%) scored models derived from each of the initial structures were clustered with respect to the RMSD of the residues surrounding the binding pocket, thus selecting 43 major structural low-energy variants.

Unbiased Mutation-Guided Docking of AR231453

A conformational ensemble of AR231453 containing nine conformations within 3 kcal/mol of the global minimum was generated in ICM version 3.7-2 using the Merck Molecular Force Field (MMFF) and a generalized Born implicit solvation model. To extensively sample the ligand binding conformations of AR231453 in a 5-Å docking sphere around $265Trp^{7.39}$ VII:06 in GPR119, we generated 43,000 complexes with the RosettaLigand XML docking protocol (Davis and Baker, 2009; Lemmon and Meiler, 2011; Meiler and Baker, 2006). The resulting models were filtered based on three criteria. First, we discarded the 90% worst-scoring structures with regard to binding energy. Second, models were only accepted if all residues lining the pocket did not deviate by more than four standard deviations from the average Cα coordinate for that position, as observed in experimental GPCR structures. Third, acknowledging that no single residue in a one-sided mutational mapping can be safely interpreted as a ligand-contacting residue, we developed a correlation-based scoring scheme (Figure S3).

The binding energy was estimated by introducing each experimentally assessed point mutation in the 43,000 models, and measuring the energy of the complex and of the receptor and ligand separately. To reduce the amount of residues, which indirectly could affect ligand binding, we included only non-proline and non-glycine residues. Likewise, to ensure that the receptor remained functional, we discarded all receptor variants in which constitutive activity was eliminated, unless they could still be activated by the ligand. We required that the surface expression be at least 20% of the wild-type and that the mutated residues were situated on the inside of the binding pocket or in the loop region. These criteria left us with 28 point mutations in GPR119 (Table 1). The correlation between the computed binding energies and the experimentally determined binding energies was then determined for each of the 43,000 binding conformations. The correlation was measured as the SCC, a rank-based correlation score which, compared with the Pearson correlation coefficient, is less sensitive to outliers. Finally, the models that passed the filters described above and correlated better than 0.25 with the experimental data were clustered (based on ligand binding conformation) in bins of 2.5 Å RMSD using BCL::Cluster to quantitate the general trends in the ensemble (Alexander et al., 2011).

Conformational Analysis of AR437735 and AR437948

The MMFF with a generalized Born solvation model was used to calculate the difference in energy between axial and equatorial conformations of AR437735 and AR437948 using ICM 3.7-3 (Molsoft) with maximum number of conformations = 50, vicinity = 15, thoroughness = 10, and sample rings and Cartesian refinement checked.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and one table and can be found with this article online at http://dx.doi.org/10.1016/j.str.2015.09.014.

ACKNOWLEDGMENTS

We would like to thank Jens Meiler and Elizabeth D. Nguyen (Center for Structural Biology, Vanderbilt University) for valuable discussions on the docking protocol.

Received: June 8, 2015
Revised: September 14, 2015
Accepted: September 15, 2015
Published: October 29, 2015

REFERENCES

Alexander, N., Woetzel, N., and Meiler, J. (2011). Bcl::Cluster: a method for clustering biological molecules coupled with visualization in the Pymol molecular graphics system. In 2011 IEEE 1st International Conference on Computational Advances in Bio and Medical Sciences (ICCABS) (IEEE), pp. 13–18.

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

Barth, P., Schonbrun, J., and Baker, D. (2007). Toward high-resolution prediction and design of transmembrane helical protein structures. Proc. Natl. Acad. Sci. USA *104*, 15682–15687.

Beuming, T., and Sherman, W. (2012). Current assessment of docking into GPCR crystal structures and homology models: successes, challenges, and guidelines. J. Chem. Inf. Model. *52*, 3263–3277.

Carlsson, J., Yoo, L., Gao, Z.-G., Irwin, J.J., Shoichet, B.K., and Jacobson, K.A. (2010). Structure-based discovery of A2A adenosine receptor ligands. J. Med. Chem. *53*, 3748–3755.

Carlsson, J., Coleman, R.G., Setola, V., Irwin, J.J., Fan, H., Schlessinger, A., Sali, A., Roth, B.L., and Shoichet, B.K. (2011). Ligand discovery from a dopamine D3 receptor homology model and crystal structure. Nat. Chem. Biol. *7*, 769–778.

Chu, Z.-L., Jones, R.M., He, H., Carroll, C., Gutierrez, V., Lucman, A., Moloney, M., Gao, H., Mondala, H., Bagnol, D., et al. (2007). A role for beta-cell-expressed G protein-coupled receptor 119 in glycemic control by enhancing glucose-dependent insulin release. Endocrinology *148*, 2601–2609.

Chu, Z.-L., Carroll, C., Alfonso, J., Gutierrez, V., He, H., Lucman, A., Pedraza, M., Mondala, H., Gao, H., Bagnol, D., et al. (2008). A role for intestinal endocrine cell-expressed g protein-coupled receptor 119 in glycemic control by enhancing glucagon-like Peptide-1 and glucose-dependent insulinotropic peptide release. Endocrinology *149*, 2038–2047.

Chu, Z.-L., Carroll, C., Chen, R., Alfonso, J., Gutierrez, V., He, H., Lucman, A., Xing, C., Sebring, K., Zhou, J., et al. (2010). N-oleoyldopamine enhances glucose homeostasis through the activation of GPR119. Mol. Endocrinol. *24*, 161–170.

Dal Ben, D., Lambertucci, C., Marucci, G., Volpini, R., and Cristalli, G. (2010). Adenosine receptor modeling: what does the A2A crystal structure tell us? Curr. Top. Med. Chem. *10*, 993–1018.

Davis, I., and Baker, D. (2009). RosettaLigand docking with full ligand and receptor flexibility. J. Mol. Biol. *385*, 381–392.

Deupi, X., and Kobilka, B.K. (2010). Energy landscapes as a tool to integrate GPCR structure, dynamics, and function. Physiology (Bethesda) *25*, 293–303.

Engelstoft, M.S., Norn, C., Hauge, M., Holliday, N.D., Elster, L., Lehmann, J., Jones, R.M., Frimurer, T.M., and Schwartz, T.W. (2014). Structural basis for constitutive activity and agonist-induced activation of the enteroendocrine fat sensor GPR119. Br. J. Pharmacol. *171*, 5774–5789.

Fleishman, S.J., and Baker, D. (2012). Role of the biomolecular energy gap in protein design, structure, and evolution. Cell *149*, 262–273.

Fredriksson, R., Höglund, P.J., Gloriam, D.E.I., Lagerström, M.C., and Schiöth, H.B. (2003). Seven evolutionarily conserved human rhodopsin G protein-coupled receptors lacking close relatives. FEBS Lett. *554*, 381–388.

Fyfe, M.C., McCormack, J.G., Overton, H.A., Procter, M.J., and Reynet, C. (2008). GPR119 agonists as potential new oral agents for the treatment of type 2 diabetes and obesity. Expert Opin. Drug Discov. *3*, 403–413.

Gaulton, A., Bellis, L.J., Bento, A.P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B., et al. (2012). ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res. *40*, D1100–D1107.

De Graaf, C., Foata, N., Engkvist, O., and Rognan, D. (2008). Molecular modeling of the second extracellular loop of G-protein coupled receptors and its implication on structure-based virtual screening. Proteins *71*, 599–620.

Jaakola, V.-P., Lane, J.R., Lin, J.Y., Katritch, V., Ijzerman, A.P., and Stevens, R.C. (2010). Ligand binding and subtype selectivity of the human A(2A) adenosine receptor: identification and characterization of essential amino acid residues. J. Biol. Chem. *285*, 13032–13044.

Jain, A.N. (2003). Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. J. Med. Chem. *46*, 499–511.

Katritch, V., Jaakola, V.-P., Lane, J.R., Lin, J., Ijzerman, A.P., Yeager, M., Kufareva, I., Stevens, R.C., and Abagyan, R. (2010a). Structure-based discovery of novel chemotypes for adenosine A(2A) receptor antagonists. J. Med. Chem. *53*, 1799–1809.

Katritch, V., Rueda, M., Lam, P.C.-H., Yeager, M., and Abagyan, R. (2010b). GPCR 3D homology models for ligand screening: lessons learned from blind predictions of adenosine A2a receptor complex. Proteins *78*, 197–211.

Katritch, V., Cherezov, V., and Stevens, R.C. (2013). Structure-function of the G protein-coupled receptor superfamily. Annu. Rev. Pharmacol. Toxicol. *53*, 531–556.

Kaufmann, K.W., and Meiler, J. (2012). Using RosettaLigand for small molecule docking into comparative models. PLoS One *7*, e50769.

Kim, S.-K., Gao, Z.-G., Van Rompaey, P., Gross, A.S., Chen, A., Van Calenbergh, S., and Jacobson, K.A. (2003). Modeling the adenosine receptors: comparison of the binding domains of A2A agonists and antagonists. J. Med. Chem. *46*, 4847–4859.

Kneissl, B., Leonhardt, B., Hildebrandt, A., and Tautermann, C.S. (2009). Revisiting automated G-protein coupled receptor modeling: the benefit of additional template structures for a neurokinin-1 receptor model. J. Med. Chem. *52*, 3166–3173.

Kolb, P., Rosenbaum, D.M., Irwin, J.J., Fung, J.J., Kobilka, B.K., and Shoichet, B.K. (2009). Structure-based discovery of beta2-adrenergic receptor ligands. Proc. Natl. Acad. Sci. USA *106*, 6843–6848.

Kufareva, I., Rueda, M., Katritch, V., Stevens, R.C., and Abagyan, R. (2011). Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. Structure *19*, 1108–1126.

Kufareva, I., Katritch, V., Stevens, R.C., and Abagyan, R. (2014). Advances in GPCR Modeling Evaluated by the GPCR Dock 2013 Assessment: Meeting New Challenges. Structure *22*, 1120–1139.

Leaver-Fay, A., Tyka, M., Lewis, S.M., Lange, O.F., Thompson, J., Jacak, R., Kaufman, K., Renfrew, P.D., Smith, C.A., Sheffler, W., et al. (2011). ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. Methods Enzymol. *487*, 545–574.

Lebon, G., Warne, T., Edwards, P.C., Bennett, K., Langmead, C.J., Leslie, A.G.W., and Tate, C.G. (2011a). Agonist-bound adenosine A2A receptor structures reveal common features of GPCR activation. Nature *474*, 521–525.

Lebon, G., Bennett, K., Jazayeri, A., and Tate, C.G. (2011b). Thermostabilisation of an agonist-bound conformation of the human adenosine A(2A) receptor. J. Mol. Biol. *409*, 298–310.

Lemmon, G., and Meiler, J. (2011). Rosetta ligand docking with flexible XML protocols. Methods Mol. Biol. *819*, 143–155.

Lu, Y., Wang, R., Yang, C.-Y., and Wang, S. (2007). Analysis of ligand-bound water molecules in high-resolution crystal structures of protein-ligand complexes. J. Chem. Inf. Model. *47*, 668–675.

Manglik, A., and Kobilka, B. (2014). The role of protein dynamics in GPCR function: insights from the β2AR and rhodopsin. Curr. Opin. Cell Biol. *27*, 136–143.

McClure, K.F., Darout, E., Guimarães, C.R.W., DeNinno, M.P., Mascitti, V., Munchhof, M.J., Robinson, R.P., Kohrt, J., Harris, A.R., Moore, D.E., et al. (2011). Activation of the G-protein-coupled receptor 119: a conformation-based hypothesis for understanding agonist response. J. Med. Chem. *54*, 1948–1952.

Meiler, J., and Baker, D. (2006). ROSETTALIGAND: protein-small molecule docking with full side-chain flexibility. Proteins *65*, 538–548.

Michino, M., Abola, E., Brooks, C.L., Dixon, J.S., Moult, J., and Stevens, R.C. (2009). Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008. Nat. Rev. Drug Discov. *8*, 455–463.

Nguyen, E.D., Norn, C., Frimurer, T.M., and Meiler, J. (2013). Assessment and challenges of ligand docking into comparative models of G-protein coupled receptors. PLoS One *8*, e67302.

Rasmussen, S.G.F., DeVree, B.T., Zou, Y., Kruse, A.C., Chung, K.Y., Kobilka, T.S., Thian, F.S., Chae, P.S., Pardon, E., Calinski, D., et al. (2011). Crystal structure of the β2 adrenergic receptor-Gs protein complex. Nature *477*, 549–555.

Ritter, S.L., and Hall, R.A. (2009). Fine-tuning of GPCR activity by receptor-interacting proteins. Nat. Rev. Mol. Cell Biol. *10*, 819–830.

Schwartz, T.W., Gether, U., Scambye, H.T., and Hjort, S.A. (1995). Molecular mechanism of action of non-peptide ligands for peptide receptors. Curr. Pharm. Des. *1*, 325–342.

Semple, G., Fioravanti, B., Pereira, G., Calderon, I., Uy, J., Choi, K., Xiong, Y., Ren, A., Morgan, M., Dave, V., et al. (2008). Discovery of the first potent and orally efficacious agonist of the orphan G-protein coupled receptor 119. J. Med. Chem. *51*, 5172–5175.

Semple, G., Ren, A., Fioravanti, B., Pereira, G., Calderon, I., Choi, K., Xiong, Y., Shin, Y.-J., Gharbaoui, T., Sage, C.R., et al. (2011). Discovery of fused bicyclic agonists of the orphan G-protein coupled receptor GPR119 with in vivo activity in rodent models of glucose control. Bioorg. Med. Chem. Lett. *21*, 3134–3141.

Shi, L., and Javitch, J.A. (2004). The second extracellular loop of the dopamine D2 receptor lines the binding-site crevice. Proc. Natl. Acad. Sci. USA *101*, 440–445.

Spitzer, R., and Jain, A.N. (2012). Surflex-Dock: docking benchmarks and real-world application. J. Comput. Aided. Mol. Des. *26*, 687–699.

Staus, D.P., Wingler, L.M., Strachan, R.T., Rasmussen, S.G.F., Pardon, E., Ahn, S., Steyaert, J., Kobilka, B.K., and Lefkowitz, R.J. (2014). Regulation of β2-adrenergic receptor function by conformationally selective single-domain intrabodies. Mol. Pharmacol. *85*, 472–481.

Stevens, R.C., Cherezov, V., Katritch, V., Abagyan, R., Kuhn, P., Rosen, H., and Wüthrich, K. (2013). The GPCR Network: a large-scale collaboration to determine human GPCR structure and function. Nat. Rev. Drug Discov. *12*, 25–34.

Vassilatis, D.K., Hohmann, J.G., Zeng, H., Li, F., Ranchalis, J.E., Mortrud, M.T., Brown, A., Rodriguez, S.S., Weller, J.R., Wright, A.C., et al. (2003). The G protein-coupled receptor repertoires of human and mouse. Proc. Natl. Acad. Sci. USA *100*, 4903–4908.

Warne, T., Moukhametzianov, R., Baker, J.G., Nehmé, R., Edwards, P.C., Leslie, A.G.W., Schertler, G.F.X., and Tate, C.G. (2011). The structural basis for agonist and partial agonist action on a β(1)-adrenergic receptor. Nature *469*, 241–244.

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

Xu, F., Wu, H., Katritch, V., Han, G.W., Jacobson, K.A., Gao, Z.-G., Cherezov, V., and Stevens, R.C. (2011). Structure of an agonist-bound human A2A adenosine receptor. Science *332*, 322–327.

