JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

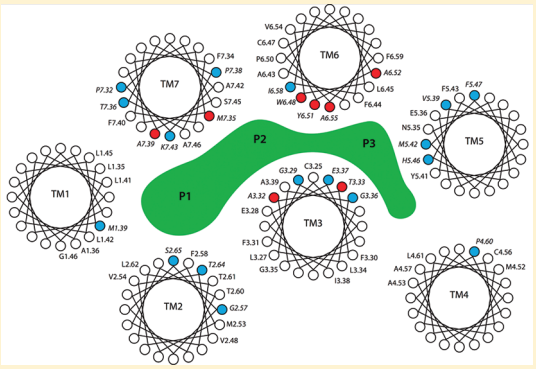
Matching Cavities in G Protein-Coupled Receptors to Infer Ligand-Binding Sites

Praveen K. Madala,[†] David P. Fairlie,[†] and Mikael Bodén[†,‡,§,*]

[†]Institute for Molecular Bioscience, [‡]School of Chemistry and Molecular Biosciences, and [§]School of Information Technology and Electrical Engineering, The University of Queensland, St. Lucia, QLD 4072, Australia

Supporting Information

ABSTRACT: To understand the activity and cross reactivity of ligands and G protein-coupled receptors, we take stock of relevant existing receptor mutation, sequence, and structural data to develop a statistically robust and transparent scoring system. Our method evaluates the viability of binding of *any* ligand for *any* GPCR sequence of amino acids. This enabled us to explore the binding repertoire of both receptors and ligands, relying solely on correlations between carefully identified receptor features and without requiring any chemical information about ligands. This study suggests that sequence similarity at specific binding pockets can predict relative affinity of ligands; enabling recovery of over 80% of known ligands for a withheld receptor and almost 80% of known receptors for a ligand. The method enables qualitative prediction of ligand binding for all nonredundant human G protein-coupled receptors.



INTRODUCTION

G protein-coupled receptors (GPCRs) form a large protein family that plays an important role in many physiological and pathophysiological processes. GPCRs are integral membrane-spanning proteins on the cell surface that transmit signals inside cells in response to external stimuli such as light, odorants, and many endogenous protein and peptide hormones, lipids, biogenic amines, amino acids, nucleosides and nucleotides, and small organic compounds. They mediate many, though not all, of their biological functions through their intracellular interactions with heterotrimeric G proteins. On the basis of sequence homology and functional similarity, GPCRs can be categorized into six broad classes.[1]

The human genome encodes approximately 950 GPCRs,[2−5] and many other species also use GPCRs to mediate signal transduction across cell membranes. Approximately 30% of all drugs in the market exert their actions through binding to human GPCRs.[6,7] Many GPCRs are validated as potential targets for therapeutics, but either lack small molecule lead compounds[8] or those that are known have yet to be transformed into clinical candidates.[3] To effectively exploit the functional diversity of GPCRs for drug discovery and development, there is a need to better understand how ligands act on receptors with limited binding and functional information.

Knowledge of protein structures is important for understanding receptor function. GPCRs are membrane-spanning proteins with low water solubility and are often difficult to crystallize. As a consequence, rational drug design for GPCRs is extremely challenging. To date about a dozen different GPCRs (including bovine rhodopsin, 2-adrenergic receptor, 1-adrenergic receptor, adenosine A2A receptor, C-X-C chemokine receptor-4, dopamine D3 receptor, histamine H1 receptor, muscarinic acetylcholine receptor, and micro- and kappa-opioid receptors)[9−17] have been resolved structurally. All known GPCR structures contain seven transmembrane (TM) helices connected by three extracellular and three intracellular loop regions. The TM helices also have similar orientations within the membrane. Most resolved GPCR structures have very low structural variation within the 7TM region compared to extracellular and intracellular components. Intracellular residues are more conserved than extracellular residues (26% and 6%, respectively, over all structures published pre-2012). These variations near the extracellular boundaries of the TM helices help in identifying the selective endogenous ligands for their biological activity. Generally, 50−60% of residues are identical between the ligand pockets of a GPCR subtype. Katrich and Abagyan[18] showed that the selectivity profile of different adenosine receptor antagonists is based on the similarity profile of the residues within the ligand-binding pocket. While GPCRs are flexible in the transmembrane regions to accommodate a variety of signals, multiple crystal structures with different agonists and antagonists have demonstrated somewhat limited conformational rearrangement of their binding pockets.[19,20] This encouraged us to develop homology models of different peptide activated GPCRs and to identify common ligand binding sites.

To characterize novel GPCRs computationally, several groups have developed methods for categorizing them into

Received: December 5, 2011
Published: April 18, 2012

 © 2012 American Chemical Society dx.doi.org/10.1021/ci2005498 | *J. Chem. Inf. Model.* 2012, 52, 1401−1410

known classes.[21−25] Van der Horst and colleagues inferred receptor similarity on the basis of ligand sharing where receptors had 20 or more known ligands. Surgand et al. identified a set of 30 key residues for ligand binding. They also identified subgroups of receptors with particular cavity delimiting properties.

To explicitly predict receptor−ligand binding, several groups have aimed to model properties of the receptor while requiring information about ligands and ligand substructures.[26−31] Bondensgaard and colleagues[31] did not develop a specific predictor but suggested, on the basis of sequence analysis and experimental analyses, that there are "privileged structures" that recognize a conserved but relatively generic and complementary binding pocket of a subset of class-A GPCRs. Joost and Methner[27] used phylogenetic distances between receptor sequences to group and predict their ligands. Both studies demonstrate the significance of using receptor sequence similarity as a basis for identifying applicable ligands, also attested to by Gloriam et al[30] who took stock of specific regions relevant for binding. Sanders and colleagues[29] also built a system for identifying specific residues involved in ligand binding on the basis of multiple alignment of many GPCRs. None of these studies use three-dimensional information coupled with experimental data to qualify the importance of such key residues.

Jacob and colleagues[28] developed a discriminative system that predicts whether a specified receptor and ligand pair binds or not. They demonstrated by the selective use of input features that GPCR class contributes to prediction accuracy, as does amino acid identity at key sites. The ability to predict the optimal ligand for any given receptor, or the optimal receptor for any given ligand, was not tested. Weill and Rognan[26] developed a combined receptor−ligand fingerprint that merged information about the receptor (including pharmacophoric properties of residues in a consensus binding cavity) with that of the ligand (including molecular descriptors such as SHED, MACCS, and DistFP). Both studies used discriminative support-vector machines making biological interpretation of decisions hard and forcing the artificial separation of reference training data into distinct positive and negative cases (referred to as "decoys" by Nicholls and Jain[32]). It is unclear to what extent ligand descriptors contributed to the prediction performance that was reported in these studies because the added details might have resulted in the classifier overfitting the relatively limited sample size.

To understand the cross reactivity and activity of all the ligands and all class-A GPCRs in a transparent and biologically interpretable manner, the present study takes stock of relevant receptor mutation data and existing sequence and structure data to build a statistically robust scoring system. It evaluates the viability of binding of *any* ligand target for *any* GPCR amino acid sequence. We are able to explore the binding repertoire of both receptors and ligands, relying solely on correlations between carefully identified receptor features and without requiring chemical information about ligands. When detailed structural information is available, there are several tools that can approach the problem on that basis.[33,34] However, with very few structures of GPCRs available, to explore binding events of ligands for novel receptors, we cannot rely upon structural information.

From analyses of crystal structures and ligand specific mutational data of class-A GPCRs, we first illustrate that small molecule ligands primarily interact with the extracellular half of the transmembrane helices along with the extracellular loops of the receptor. We find that this region of the receptor has a number of "hot spots" that will help either activate or antagonize the receptor. Identifying residues by statistical or information−theoretic means, without considering available structural and mutational data,[35] led to poor generalization in our trials.

We note that similarity between pockets can be measured by comparing physiochemical properties of the amino acids at the ligand-binding site when such knowledge is available. For example, Asp 3.32 of biogenic amine receptors interacts with a particular class of compounds with a positively charged amine nitrogen. We carefully crafted an amino acid profile—gleaning both structure and sequence conservation—that reflects the statistical properties of individual residues for the receptor family that may be involved in ligand binding.

To identify lead molecules for a novel class-A GPCR of interest, we used ligand specific mutational data of 22 different GPCRs to characterize common ligand-binding sites with three different pockets. This information is used to develop a sequence alignment template and a scoring method for novel GPCRs on the basis of their alignment with the template, measuring the similarity at the pockets relevant to ligand binding. Using the GLIDA GPCR-ligand database,[36] we verified that the score could be used on novel GPCRs to rank familiar GPCRs with similar affinity for ligands. Apart from indirectly increasing our understanding of GPCR binding, the method can dramatically reduce the number of ligands that need to be screened to find new agonists and antagonists for novel GPCRs. The method draws its power from reducing the problem to that of identifying similar pairs of GPCRs, where the known affinity for ligands of one is statistically transferred to the other.

RESULTS

Class-A GPCRs Have Conserved Ligand Binding Residues. This section aims to demonstrate that there are similar conserved ligand-binding sites in the 7TM region of class-A GPCRs. We integrated several sources of information, including sequence, structure, ligand binding, and effects of mutations. GPCRDB[37] provides literature references that allowed us to establish which residues are critical for ligand binding. We first identified and verified (by inspecting the scientific literature) 140 individual residues in 22 class-A GPCRs that, if mutated, changed ligand function or affinity (see Supplementary Table 1 of the Supporting Information for residues for which experimental evidence is qualified).

Second, we introduced two crystal structures (bovine rhodopsin and human B2-adrenergic receptor) and built homology models for each peptide-activated receptor using Modeller,[38] refined using molecular dynamics simulations to remove internal side chain clashes within transmembrane regions, and energy minimized in Insight II. Structures of ligands were docked into each receptor using GOLD.[39] From about 30 docked conformations, we manually inspected known contact residues (Supplementary Table 1 of the Supporting Information).

To further generalize the sites found to be critical for ligand binding across receptors, we constructed a reference alignment of the class-A GPCRs in which binding was modulated, using Clustal and several standard refinements (Methods and Materials). We also mapped bovine rhodopsin onto the alignment to estimate the spatial arrangement of columns.

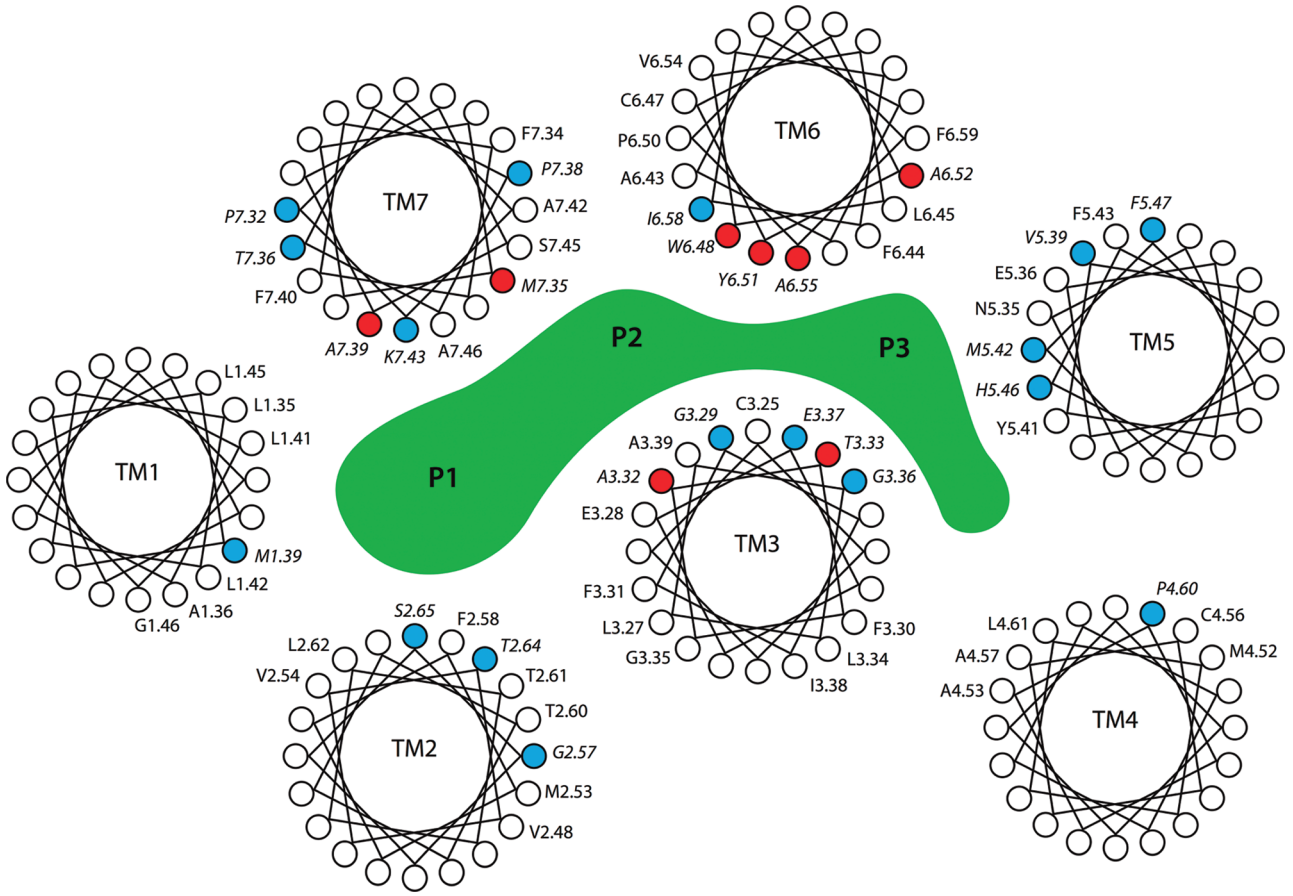


Figure 1. Residues important for class-A GPCR ligand affinity or activity using the structure of Bovine rhodopsin receptor. Specific residues (placed in helical wheels corresponding to the seven transmembrane helices labeled TM1−TM7) are highlighted in red when mutations are documented to affect five or more of the 20 peptide-activating GPCRs, and in blue for 2−4 GPCRs (Supplementary Table 3 of the Supporting Information). The marked area (green) surrounded by aforementioned residues identifies the extended region of binding pockets, labeled P1−P3.

We manually ensured that the alignment captured relevant sites across all GPCRs, as indicated by mutant data, transmembrane domains, and other structural and functional properties shared by multiple members of the receptor family. Mutational data of ligand-binding and contact residues were superimposed on the alignment to identify columns of particular interest (Supplementary Table 2 of the Supporting Information). Mutational data suggested that residues form at least three major binding pockets toward the extracellular side of the transmembrane domains (Figure 1).

From the locations of mutated residues in the alignment, we identified three geometric binding "centers" that are positioned in the cavities formed by the transmembrane domains (TM1, TM2, TM3, and TM7; TM3, TM5, and TM6; and TM3, TM4, and TM5, respectively). By first identifying all residues that lie within 12 Å of each center and then removing those that face the lipid bilayer, we defined three binding pockets—or hot spots. These pockets contain residues that are likely to directly or indirectly contribute to the binding of a broad range of ligands. We also inspected the human B2-adrenergic receptor structure against the reference alignment with mutational data and recovered the same three major binding pockets due to the structural similarities (Supplementary Figure 1 of the Supporting Information).

Importantly, with the reference alignment, we can map any novel GPCR sequence to probe the amino acids at the defined binding pockets and compare them with those of other GPCRs for which more information is readily available (as discussed next).

Sequence Similarity at Common Sites Reflects Recognition of Similar Ligand Scaffolds in the 7TM Region. All GPCRs bind to multiple ligands. When the number of features describing both receptors and ligands far outnumbers the number of examples of GPCR—ligand associations, we face a major challenge. How do we identify the specific features and combinations thereof that enable binding as they may differ between ligands for the same GPCR and between GPCRs for the same ligand? This study borrows ideas from machine learning to develop an approach that circumvents this issue. We define a decision of which ligand binds to a GPCR, not in terms of features, but in terms of other examples. A decision is based on comparisons between GPCRs, therefore only implicitly on their features.

At the core of our approach is thus a GPCR−GPCR similarity metric, based on our definition of class-A binding pockets. We used a standard substitution scoring system (BLOSUM62) to score the alignment of two GPCR sequences at the columns associated with the binding pockets. A novel GPCR is first aligned to the reference alignment (see Methods and Materials for details).

To evaluate whether matching GPCRs this way indeed reflects ligand binding, we constructed two matrices. The first matrix contains the *sum of the substitution scores* for pairs of GPCRs, involving one of the 23 GPCRs in the reference

dx.doi.org/10.1021/ci2005498 | J. Chem. Inf. Model. 2012, 52, 1401−1410

Table 1. AUC for Predicting Receptor Binding for a Set of 37 Representative Ligands That Collectively Binds to 105 GPCRs[a]

		Receptors (387)				
			Correct			
	Ligand	AUC	Top 10%	1% FP	Total	Top 1
L000063	2-methylthio-ATP	0.79	1	0	4	0
L000092	ACTH	1.00	3	3	3	1
L000147	Bromocriptine	0.91	10	4	18	1
L000195	Clozapine	0.76	13	5	26	1
L000237	Beta-chemokine exodus-3	0.96	3	0	3	0
L000299	DEFB4 protein, human	0.44	0	0	2	0
L000345	CCL20 protein, human	0.97	2	0	2	0
L000382	MC148, CC chemokine	0.96	3	0	3	0
L000384	CCL8 chemokine	0.91	4	1	5	1
L000385	CCL7 chemokine	0.91	4	3	5	1
L000448	NPY	0.98	4	1	4	0
L000453	Octreotide	0.98	4	0	4	0
L000479	Dinoprostone	0.93	8	8	9	1
L000528	S 20098	0.85	3	3	4	1
L000553	CXCL12 chemokine	0.91	3	1	4	0
L000619	Uridine Triphosphate	0.97	4	0	4	0
L000624	CFLTKRGRQVC	0.95	5	4	6	1
L000645	ALX40−4C	0.83	1	0	3	0
L000690	Gonadotropin-releasing hormone	0.97	2	1	2	0
L000734	Diprenorphine	1.00	3	3	3	0
L000988	Adenosine Triphosphate	0.83	1	0	4	0
L001326	Nedocromil	0.66	2	0	4	0
L001445	Dronedarone	0.99	9	4	9	0
L001453	S 20098	0.78	2	0	3	0
L004107		0.94	3	3	4	1
L006256	Terazosine	0.99	5	3	5	0
L006334		0.98	4	0	4	0
L009321		0.99	2	0	2	0
L010797		0.99	2	0	2	0
L012713		0.99	4	4	4	0
L016272		0.80	2	2	3	1
L017048		0.99	3	3	3	0
L018522		0.40	0	0	2	0
L021445		0.96	2	0	2	0
L022270		0.99	3	3	3	0
L023819		0.79	2	0	3	0
L024424	[Ala31,Aib32]-NPY (Porcine)	0.98	4	2	4	0
Mean		0.89	0.78 (TP rate)	0.34 (TP rate)		0.27

[a]The final row shows the average AUC and the average true positive rate in the top 10% (top 39 receptors), at less than 1% false positives, and the top 1 of predictions.

alignment, and one of all 387 class-A GPCRs we found in UniProt (that shared less than 80% sequence similarity). The second matrix contains the *number of ligands* shared by each corresponding pair of GPCRs, according to GLIDA.[36] We added to this the number of ligands that are shared indirectly via a "proxy" receptor.

For each of the 387 GPCRs, all 23 reference GPCRs were (a) ranked on the basis of their similarity score with the query receptor (the first matrix) and (b) ranked on the basis of the number of ligands they share with the query receptor (the second matrix). If the sequence similarity score reflects ligand-binding affinity, a stronger similarity score should correlate with a greater number of shared ligands. Thus, for each of the 387 GPCRs, we determined if the two ranks are in agreement; we summed the non-negative differences in ranks over the 23 GPCRs. This sum is compared to 100 "random" sums (when sequence similarity is computed over the same alignment columns but with randomly permuted rows, i.e., reference GPCR identity). This empirical test demonstrates that the rank based on sequence similarity at the binding pockets is correlated with the ligand sharing rank. With $p < 0.05$, we reject the null hypothesis that the ranks are random. When the score is based on all columns in the alignment, we largely fail to detect ligand binding. If columns include the residues in the seven transmembrane domains only, the scores improve, but we are still unable to reject that ranks are random ($p > 0.10$). Decisions about ligand binding should thus be made on the basis of the sites of the binding pockets. With these pockets subject to restricted structural variation, this observation applies not only to bovine rhodopsin and B2-adrenergic receptors, but potentially all class-A GPCRs.

GPCR Ligand Scaffold Binding Is Predicted by Matching Novel with Known GPCRs. To predict a scaffold that binds to a novel GPCR, we take stock of the sequence and

dx.doi.org/10.1021/ci2005498 | J. Chem. Inf. Model. 2012, 52, 1401−1410

Table 2. AUC for predicting ligand binding, given a receptor, for all ligands and the set of 37 representative ligands that collectively binds to the maximum number of GPCRs[a]

	All Ligands				37 Ligands			
		Correct				Correct		
Receptor	AUC	Top 10%	Total	Top 1	AUC	Top 10%	Total	Top 1
AG2R	0.48	0	13	0	0.38	0	1	0
BKRB1	1.00	173	173	1	1.00	1	1	1
BKRB2	1.00	173	174	1	1.00	1	1	1
NMBR	0.85	0	49	0	0.96	1	1	0
GASR	1.00	304	304	0	0.97	1	1	0
EDNRA	0.99	618	926	0	0.96	2	2	0
EDNRB	0.99	618	922	0	0.96	2	2	0
MC4R	1.00	11	11	1	1.00	1	1	1
NK1R	0.97	335	344	1	0.69	1	2	1
NK2R	1.00	428	428	1	1.00	1	1	1
NPY2R	0.95	15	16	0	1.00	2	2	1
CCR1	1.00	15	15	1	1.00	3	3	1
CCR2	1.00	13	13	0	1.00	4	4	1
CCR3	1.00	16	16	1	1.00	3	3	1
CCR5	1.00	18	18	1	0.99	3	3	1
CXCR4	1.00	8	8	1	0.95	2	3	1
OPRK	0.98	122	125	0	1.00	1	1	1
V1AR	1.00	301	301	1	1.00	1	1	1
V2R	1.00	213	213	1	1.00	1	1	1
Mean	0.91	TP rate 0.83		0.54	0.92	TP rate 0.84		0.66

[a]Additionally, the number of true positives in the top 10% of predictions, the total number of ligands, and the correctness of the top 1 prediction are given. Because of space constraints, only the peptide-activating receptors in the reference alignment are shown. The final row shows the means of each column computed over all GPCRs with at least one known ligand.

ligand binding diversity of the broader group of human class-A GPCRs. The reference alignment is extended to encompass the wide variety of class-A members using Clustal profile alignment. As quantified and validated below, using the extended alignment we can accurately predict bound ligands (agonists and antagonists) for most GPCRs for which there is at least one to be validated.

The prediction is based on the following steps (see Methods and Materials for details):

- The sequence of a novel (yet unseen) GPCR is aligned to the extended alignment.
- With positions of pockets mapped onto the query sequence, its similarity with each receptor is determined. (This is the sum of substitution scores from BLOSUM62 over all positions in the three pockets.)
- For each ligand, a label "yes" or "no" is assigned to each scored receptor to indicate whether the ligand is known or not known to bind. The Wilcoxon Ranksum test is used to test if receptors labeled "yes" are scored more highly than those labeled with "no". Accordingly, a *p*-value is assigned to each ligand to indicate statistical support.
- Ligands are sorted according to this *p*-value (in ascending order). The topmost ligands are predicted to bind to the query GPCR with a certainty indicated by the *p*-value.

For testing purposes, we withheld a query sequence from the extended alignment in the same way as leave-one-out cross validation. Because the prediction method assigns a *p*-value to each possible receptor−ligand pair, we investigated three different prediction scenarios: predicting ligands given receptor, predicting receptor given ligand, and classifying whether a specified receptor−ligand pair binds or not.

We first asked, given a ligand, with what accuracy could bound receptors be predicted? According to GLIDA, 105 of the 387 GPCRs in our data set have one or more known ligands (agonists or antagonists; peptides or small molecules).

To estimate the ligand prediction accuracy robustly without introducing bias and across as many GPCRs as possible, we determined the accuracy for the minimal subset of ligands that bind to the maximum number of *different* GPCRs. Choosing ligands this way avoids overestimating the accuracy that would result from using all ligands, many of which will be binding to the same receptors (e.g., chemokines are frequently bound to and thus can be "guessed"; see Methods and Materials for details).

Table 1 lists the area-under-ROC curve (AUC) for 37 ligands that collectively (but nonexclusively) bind to 105 receptors. The AUC was determined from the set of 387 GPCRs. The average receptor prediction accuracy had an AUC of 0.89 (where 1.0 represents perfect; 0.50 is chance performance). For the average ligand, screening the top 10% of receptors (387/10) recovered 78% of all its known receptors. (When computing these metrics over all ligands, the average AUC goes up to 0.90, and the 10% recovery rate rises to 0.83.) When inspecting only the top 1% of receptor predictions, 26% of all known receptors were recovered. We also report the true positive rate at specific levels of false positives as recommended by Jain and Nicholls.[32] By only accepting 1% false positives (at less than 387/100 incorrect predictions), 34% of all known positives were recovered, on average. The corresponding numbers at 0.5%, 2%, and 5% accepted false positives are 19%, 49%, and 71% of known receptors, respectively.

Second, we asked, given a receptor, with what accuracy can its ligands be predicted? Because of the scarcity of data, the assessment is crude. In our assessment, we distinguished

dx.doi.org/10.1021/ci2005498 | J. Chem. Inf. Model. 2012, 52, 1401−1410

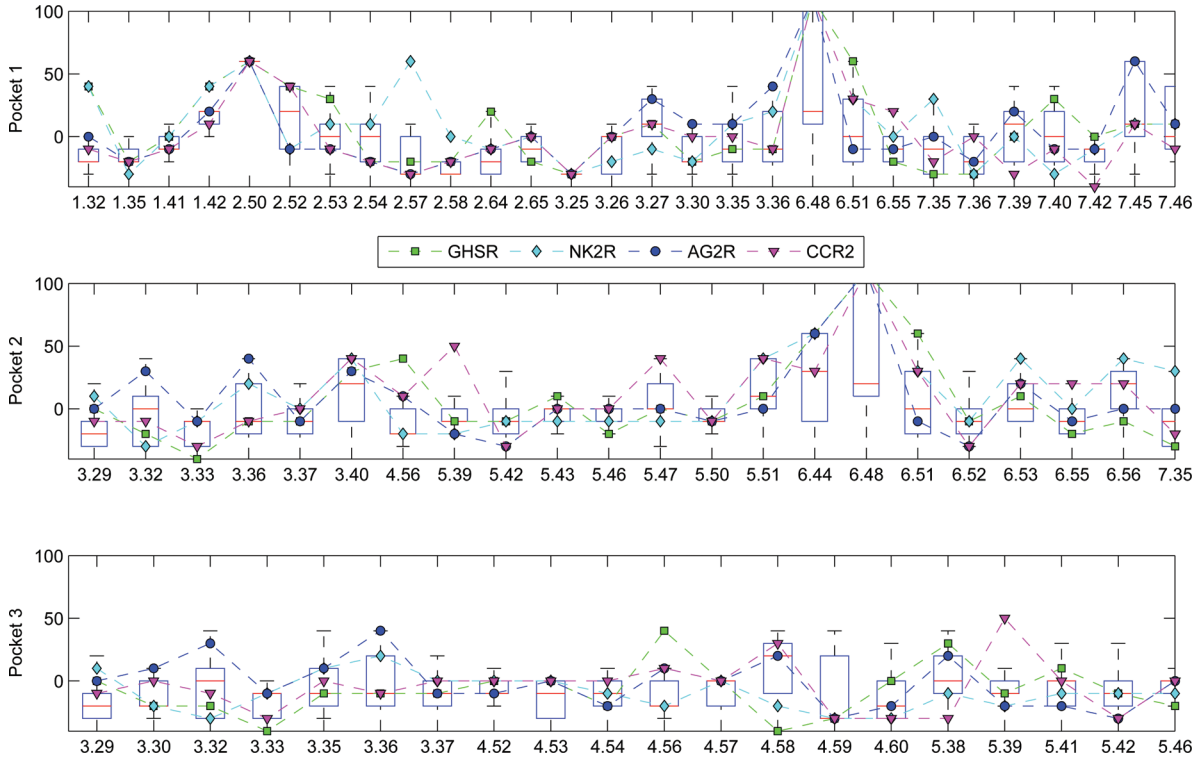


Figure 2. Boxplot of residue similarity scores at binding pockets for MC4R. For each position the median, 25th, and 75th percentiles and nonoutlier extreme scores comparing MC4R with each human GPCR are indicated. Scores comparing MC4R with four receptors that share affinity for spiroindane scaffold are explicitly plotted, suggesting sites at which similarity may be important for binding, e.g., H6.52 in GHSR and F6.51 in NK2R. (Shown scores are multiplied by 10).

between the case of 37 ligands and the case of all ligands. We only explored the performance for the 105 receptors to which at least one of these ligands binds. The average prediction AUC is 0.92 and 0.91, respectively. For the average receptor, screening the top 4 of the 37 ligands (10%) recovered 84% of binding, screening the top 10% of all ligands recovers 83% of all known ligands bound (Table 2).

Jacob et al[28] used 80 GPCRs and about 4000 ligand interactions involving them. From the same ligands, they randomly selected the same number of "negative" receptor interactions, assuming that a pair was not binding if it is not listed in GLIDA. They trained support-vector machines using various input kernels to classify any given receptor−ligand pair as "binding" or "non-binding". They reported that if a receptor was completely withheld from training (similar to our setting), the best accuracy was achieved with a "binding pocket" kernel for the receptor and a "2D Tanimoto" kernel for the ligand. With this setting, with a support-vector machine threshold set according to the expected number of positives, 78.1% of pairs were correctly classified.

We extended our set of receptors to encompass the additional 18 receptors. We had to replace a few of the original GPCRs to keep maximum sequence similarity at 80%. We were unable to exactly reproduce the same set of ligands because our scoring system required that a ligand bind to more than one receptor. In our set, there were 2320 true interactions and 1210 false interactions. By setting the p-value threshold of our scoring system optimally, 93.8% of the 3530 interactions were correctly classified. With the proviso of the difference in ligand coverage, this test indicated a substantial improvement to the accuracy reported by Jacob et al.

Having established that the scoring system performs reliably, we provide an extensive resource documenting the top 10 ligands for each of the 387 receptors (Supplementary Table 3 of the Supporting Information). Each prediction treated the receptor as if it was an orphan receptor. The highest scoring receptor for each ligand prediction is presented, which enables the user to quickly determine on what grounds the prediction is made. We additionally provide sequence similarity measures, both pocket-aggregate and pocket-specific, to allow detailed exploration of the basis for inferring binding. The list of predictions is hyperlinked to GLIDA and UniProt to facilitate continued investigation and exploration of annotations in these rich resources.

To illustrate how the model operates on the basis of receptor sequence information and how that could give insights to ligand binding, we examined more closely the predictions for MC4R (melanocortin-4 receptor). Receptors are (as described previously) ranked on the basis of their similarity at the transmembrane domains making up the three binding pockets. Not surprisingly, and as viewed in Supplementary Table 3 of the Supporting Information, MC3R (melanocortin-3 receptor) is most similar and its ligand, which is shared by number two ranked ACTHR (adrenocorticotropic hormone receptor), is transferred to MC4R (p = 0.025). More interesting is the appearance of AGTR2 (angiotensin II receptor) and NK2R receptor (neurokinin-2), both associated with spiroindanepiperidine[31] with MC4R. Figure 2 shows the similarity scores that make up the receptor scores used for MC4R predictions for four receptors with affinity for this molecular scaffold. We note in particular that experimentally verified mutations for MC4R, NK2R, and GHSR (growth hormone secretagogue receptor), namely, F6.51, F6.52 (for both MC4R[31] and NK2R[40]), and

dx.doi.org/10.1021/ci2005498 | J. Chem. Inf. Model. 2012, 52, 1401−1410

H6.52 (for GHSR[41]), are singled out and their detrimental effects for binding are thus supported from the inspection of the similarity scores.

The current scoring system is only based on similarities between receptors. It is thus not obvious how it could be sensitive in predicting binding ligands and how ligands can be used to predict applicable receptors, except of course that protein sequence is well-known to determine three-dimensional structure. Bondensgaard and colleagues[31] noted that ligand libraries can be constructed with limited knowledge of the endogenous ligand. To understand limitations of the method better, we posed the following problem.

We added a fictional ligand with receptors on the basis of those known for the spiroindoline piperidine scaffold,[31] namely GHSR, MC4R, OXYR (oxytocin receptor), and NK1R (neurokinin-1 receptor). During testing, we withheld each of the receptors in turn, recording the *p*-value assigned to the binding of the spiroindoline piperidine ligand. Overall, the top predictions (all with the same *p*-value) were SSR1 (somatostatin-1 receptor), SSR3 (somatostatin-3 receptor), MTLR (a growth hormone secretagogue like receptor), O5AN1 (olfactory II family 5), GPR75 (unclassified class-A receptor), V1BR (vasopressin-1 receptor), and OXYR. From the literature, it is known that SSR2 has a potent receptor agonist containing a spiroindoline piperidine scaffold. These results gave us confidence that our method is able to identify ligands for GPCRs of interest.

DISCUSSION

A computational method that can successfully identify candidate ligands for orphan GPCRs can reduce the degree of expensive and time-consuming experimental high throughput screening. However, to infer that ligands bind to novel GPCRs is nontrivial. Not only are data prohibitively sparse and case-specific, but binding events are also intrinsically multifaceted, and of course, resolved structures are only available in very few cases.

Our scoring system is potentially able to predict any ligand known to bind to other GPCRs, for an entirely novel GPCR for which no known ligand exists, represented only by the GPCR sequence. The approach we have taken avoids explicit identification of the large number of features that factor into successful binding by making qualitative decisions relative to other binding events. This circumvents the discriminative approach imposed by many machine-learning algorithms. Inference is thus made more robust, and it is transparent to the user. Indeed, having access to examples of actual receptor−ligand pairs that are statistically similar to that of a query allows the expert to manually inspect possible determinants.

The similarity scoring system is crucially based on sequence features at sites deemed relevant by cross-referencing observations from a structurally informed sequence alignment, extensive GPCR mutagenesis data, ligand-docking simulations, and binding pockets discerned from a representative structure. The simplification imposed on the problem by the score enables robust tests, bringing impressive generalization. Notably, when individual receptor−ligand predictions are evaluated, there is no information about what specific ligand-binding pockets are used. Ligand-specific properties can only be discerned by inspecting a group of receptors also bound by the same ligand. This is an issue that could be overcome by, not unlike how we dealt with receptors, deconstructing ligand moieties to measure their relative similarity and by performing a complementary statistical test of their receptor-binding enrichment. We think that this conceptually new approach to computational predictions of ligand binding may be extendable beyond GPCRs to be a promising comparative method for patterning proteins of like classes, searching for ligand-binding hot spots, and for inferring ligand function. Considerable further development is needed to realize the promise that is seeded here in this extensive investigation of multiple GPCR sequences.

METHODS AND MATERIALS

Residues Involved in Ligand Binding. Peptide-activating receptors with specific mutational data that have confirmed impact of ligand-related activity and binding affinity in the literature were obtained from GPCRDB[23,37] (Supplementary Table 1 of the Supporting Information). The amino acid sequences of 20 human GPCRs were retrieved from the UniProt database (angiotensin-2 type-1 receptor, P30556; neuromidin preferring bombesin receptor, P28336; bradykinin-1 receptor, P46663; bradykinin-2 receptor, P30411; complementary-5 anaphylatoxin chemotactic receptor, P21730; chemokine receptor-1, P32246; chemokine receptor-2, P41597; chemokine receptor-3, P51677; chemokine receptor-5, P51681; cholecystokinin type-B receptor, P32239; endothelin-A receptor, P25101; endothelin-B receptor, P24530; melanocortin-4 receptor, P32245; neurokinin-1 receptor, P25103; neurokinin-2 receptor, P21452; neuropeptide-Y receptor 2, P49146; kappa opioid receptor, P41145; protease activated receptor-2, P55085; vasopressin-1 receptor, P37288; and vasopressin-2 receptor, P30518).

The sequences were aligned using Clustal[42] multiple alignment program following the recommendations of Bissantz and colleagues.[43] The seven transmembrane regions were identified on the basis of the conserved residues in each transmembrane region.[44] The receptors were aligned with the bovine rhodopsin receptor, P02699, with associated crystal structure PDB 1U19, (and as an additional test for robustness) the human B2-adrenergic receptor, P07550, with associated crystal structure PDB 2RH1, using the PAM-250 matrix, which aligns the sequence on the basis of conservation of charged, bulk aliphatic residues, and aromatic residues. The alignment was refined manually by taking the transmembrane regions and structurally conserved regions into consideration. Approximate C and N termini of each transmembrane region were identified from the data obtained from GPCRDB,[37] and hydrophilic residues such as Arg/Lys/Asp/Glu were placed such that they occupy the junction between the transmembrane and cytoplasmic regions at the level of phospholipid head groups.[45]

Residues important for ligand affinity or activity of the each receptor were highlighted (see highlighted cells in Supplementary Table 2 of the Supporting Information), and the relative position of these residues were further plotted on the crystal structure of the Bovine rhodopsin crystal structure. (The same procedure was completed for the human B2-adrenergic receptor to confirm that the same residues were identified.)

Reference Alignment Annotated with Binding Pockets. As explained in the previous section, the "reference" alignment includes 22 human GPCRs and Bovine rhodopsin, referred to as X_1, X_2, ..., X_{23}. We specifically identified 176 columns of broad functional interest, corresponding to regions of the each of the seven transmembrane domains (1.30−1.57, 2.40−2.65, 3.25−3.51, 4.40−4.67, 5.38−5.54, 6.36−6.59, 7.32−7.57; indexed $i \in \{1, ..., 176\}$). The PDB structure for

rhodopsin supplies approximate coordinates for most columns in the alignment (x_i, y_i, z_i).

Within the columns of the reference alignment, three ligand binding pocket centers were first identified by inspecting 1U19A, namely, $c_1 = (45, 14, 15)$, $c_2 = (35, 7, 17)$, and $c_3 = (34, 14, 21)$. Second, all columns with residues that lie within 12 Å of each center were tentatively and nonexclusively categorized as belonging to pocket 1, 2, and 3. Next, we removed columns corresponding to residues in 1U19 that were facing the lipid bilayer. The final set of columns was manually extended to encompass columns that had substantial support from mutational data to be essential to binding at least two ligands. The resulting (overlapping) subsets of columns in the transmembrane alignment are referred to as π_1, π_2, and π_3.

We refer to 364 human GPCRs in UniProt with less than 80% sequence identity and not overlapping with the 23 reference GPCRs as X_{24}, ..., X_{387}. We refer to residues in any sequence X, in a column i as $r_i(X)$. We define $R_i = a$ to represent the observation $a \in \mathrm{A} \cup \sim$ in a column $i \in \{1, ..., 176\}$ in the 387 sequence alignment, where A is the set of amino acids and $\sim$ represents a gap (a deletion).

GPCR Similarity Metrics. For any pair of GPCRs X and X', we can score how similar they are in regard to a pocket π.

$$\mathrm{SEQ}_\pi(X, X') = \sum_{i \in \pi} S(r_i(X), r_i(X'))$$

where S is the BLOSUM62 substitution matrix score. (We tried several alternative substitution matrices including BLOSUM80 with near identical accuracy.) The full similarity score is simply the sum of all pocket specific scores.

$$\mathrm{SEQ}(X, X') = \sum_i \mathrm{SEQ}_{\pi_i}(X, X')$$

The GLIDA database[36] provides interaction data between GPCRs and their ligands. We inspected all ligands (agonists and antagonists) in GLIDA (24076 ligands as on 26 April 2010) to see if they bind to any of the 387 GPCRs. In particular, we tracked all ligands that are shared between at least two GPCRs (5939 ligands). A total of 105 of the 387 GPCRs in our set have one or more known ligands.

To assess whether the similarity score for a novel GPCR X' indicates ligand binding, we first scored X' against all GPCRs in the reference alignment on the basis of sequence similarity, $\mathrm{SEQ}(X, X')$. The scores were sorted in descending order and given a rank

$$\mathrm{rank}^{\mathrm{SEQ}}_{X'}(X) \in \{1, ..., 23\}$$

Second, we created a ligand-sharing table from GLIDA, with each entry containing the number of ligands for a pair of receptors X and X'. We define $\mathbf{GL}(X)$ to be the set of ligands that are known to bind to X, and $\mathbf{GL}(X,X') = \mathbf{GL}(X) \cap \mathbf{GL}(X')$ is the set of ligands that are known to bind to both X and X'.

$$\mathrm{LIG}(X, X') = |\mathbf{GL}(X, X')| + |\mathbf{GL}(X, X'') \cap \mathbf{GL}(X', X'')|$$

For a novel GPCR X', we then ranked all GPCRs in the reference alignment on the basis of ligands shared, $\mathrm{LIG}(X, X')$. Again we sorted scores in descending order and gave each GPCR a rank

$$\mathrm{rank}^{\mathrm{LIG}}_{X'}(X) \in \{1, ..., 23\}$$

GPCRs that have no shared ligands were excluded.

The two ranks of the 23 GPCRs in the reference alignment are sourced independently and should thus not display any correlation unless there is a biological reason. We quantified the correlation the following way.

$$\mathrm{RDev}(X') = \sum_{X \in B_{X'}} \max(\mathrm{rank}^{\mathrm{SEQ}}_{X'}(X) - \mathit{rank}^{\mathrm{LIG}}_{X'}(X), 0)$$

where $B_{X'} \in \{X; |\mathbf{GL}(X, X')| > 0\}$. That is, for each core GPCR X that share at least one ligand with X', add the difference between the sequence-based and the ligand-based rank. We used the z-test to standardize and compare outcomes of using the actual sequences and 100 within-column random permutations of the 22 peptide-activating GPCRs and Bovine rhodopsin.

Receptor−ligand Prediction Method and Metrics of Accuracy. Our method predicts ligands of a novel GPCR X' by scoring X' against all GPCRs in the extended alignment, $s_X = \mathrm{SEQ}(X, X')$. For a ligand L, the method associates with each score s_X a label *positive* if $L \in \mathbf{GL}(X)$ or *negative* if $L \notin \mathbf{GL}(X)$. It then performs a one-tailed Wilcoxon Ranksum test (aka the Mann−Whitney U-test) to see if positive-labeled scores are statistically greater than negative-labeled scores. The method does this only for the top 10% of GPCR scores to avoid overestimating the statistical support for promiscuous ligands when their values are only slightly skewed toward the higher end.

The ligand predictor is able to identify any one of the 5939 ligands that are associated with at least two GPCRs in the extended alignment. However, the distribution is very skewed with a few GPCRs binding to a large proportion of ligands. Evaluating the ability of a predictor on the complete set of ligands will thus provide a distorted view of the accuracy over the full spectrum of GPCRs. As a remedy, for each GPCR, we include only the ligand that binds to the greatest number of other GPCRs. In total, we end up with a set of 37 ligands, binding to 105 of the 387 GPCRs according to GLIDA (see Table 1 for their identities).

The output of our ligand predictor is a continuous value for each ligand L, with a smaller (p-) value indicating greater support. We thus need to set a threshold θ at which the p-value for L is *positive* if it is at or below θ and *negative* otherwise.

To test the accuracy of the method, we predicted what ligands a specified GPCR X binds to and what GPCRs that a given ligand L binds to. For each GPCR X, we simply scored each of the ligands. For each ligand, we scored each of the 387 GPCRs against all others as if it was a novel GPCR (withholding all information about it). We associated with each GPCR−ligand pair a p-value indicating the support for their binding. We assessed how reliable this indicator is across the full spectrum of GPCRs and range of ligands.

We define the true positive rate as $TPR_\theta = TP_\theta/P_L$, and the false positive rate as $FPR_\theta = FP_\theta/N_X$, where P_L is the number of known binders, N_L is the complement of P_L (the number of nonbinders), TP_θ is the number of *positives* at θ that are binders, and FP_θ is the number of *positives* that are not. Both TPR and FPR will range between 0 and 1 and increase monotonically with a decreasing θ, forming a so-called receiver-operating characteristic (ROC) curve that captures the accuracy of the prediction at all possible thresholds. We specifically report the area under the ROC curve (AUC) as it summarizes the accuracy for a given ligand (Table 1) or a given receptor (Table 2), with 1 being perfect and 0.5 is chance performance.

■ ASSOCIATED CONTENT

Supporting Information

Supplementary Table 1 contains ligand specific mutations for peptide-activating G protein-coupled receptors and literature references for each. Supplementary Table 2 (MS Excel worksheet) contains the reference alignment of peptide-activating GPCRs with important residues highlighted. Supplementary Table 3 (HTML; also available at bioinf.scmb.uq.edu.au/gpcr) contains the top ligand predictions for each of the 387 receptors in our data set and is hyperlinked to GLIDA and UniProt for easy reference. Supplementary Figure 1 illustrates how critical residues map onto the B2-adrenergic receptor transmembrane helices. This information is available free of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: m.boden@uq.edu.au.

Author Contributions

P.K.M., D.P.F., and M.B. conceived the study. Collection and analysis of receptor data was done by P.K.M. The statistical method was designed and implemented by M.B. Predictions and performance analyses were completed by M.B. The following authors wrote the paper: P.K.M., D.P.F., and M.B.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

M.B. acknowledges support from the Australian Research Council (ARC) Centre of Excellence in Bioinformatics. D.P.F. acknowledges support from the ARC for a Federation Fellowship (FF668733) and a GPCR research grant (DP1093245), and from the Australian National Health and Medical Research Council for a Senior Principal Research Fellowship (1027639) and GPCR research grants (456060, 569595, 1030169). Dr. Martin Stoermer is also thanked for assistance in assessing peptide activated GPCR−ligand pairs that underpinned the initial idea for this research and which will be reported elsewhere.

■ REFERENCES

(1) Pierce, K. L.; Premont, R. T.; Lefkowitz, R. J. Seven-transmembrane receptors. *Nat. Rev. Mol. Cell Biol.* **2002**, *3*, 639−650.

(2) Foord, S. M.; Bonner, T. I.; Neubig, R. R.; Rosser, E. M.; Pin, J. P.; Davenport, A. P.; Spedding, M.; Harmar, A. J. International Union of Pharmacology. XLVI. G protein-coupled receptor list. *Pharmacol. Rev.* **2005**, *57*, 279−288.

(3) Blakeney, J. S.; Reid, R. C.; Le, G. T.; Fairlie, D. P. Nonpeptidic ligands for peptide-activated G protein-coupled receptors. *Chem. Rev.* **2007**, *107*, 2960−3041.

(4) Tyndall, J. D.; Pfeiffer, B.; Abbenante, G.; Fairlie, D. P. Over one hundred peptide-activated G protein-coupled receptors recognize ligands with turn structure. *Chem. Rev.* **2005**, *105*, 793−826.

(5) Ruiz-Gómez, G; Pfeiffer, T. J.; Abbenante, B; Fairlie, G DP: Update of: Over one hundred peptide-activated G protein-coupled receptors recognize ligands with turn structure. *Chem. Rev.* **2010**, *110*, 1−41.

(6) Harris, C. J.; Stevens, A. P. Chemogenomics: Structuring the drug discovery process to gene families. *Drug Discovery Today* **2006**, *11*, 880−888.

(7) Cheng, A. C.; Coleman, R. G.; Smyth, K. T.; Cao, Q.; Soulard, P.; Caffrey, D. R.; Salzberg, A. C.; Huang, E. S. Structure-based maximal affinity model predicts small-molecule druggability. *Nat. Biotechnol.* **2007**, *25*, 71−75.

(8) Klabunde, T.; Hessler, G. Drug design strategies for targeting G-protein-coupled receptors. *Chembiochem* **2002**, *3*, 928−944.

(9) Tebben, A. J.; Schnur, D. M. Beyond rhodopsin: G protein-coupled receptor structure and modeling incorporating the beta2-adrenergic and adenosine A(2A) crystal structures. *Methods Mol. Biol.* **2011**, *672*, 359−386.

(10) Wu, B.; Chien, E. Y.; Mol, C. D.; Fenalti, G.; Liu, W.; Katritch, V.; Abagyan, R.; Brooun, A.; Wells, P.; Bi, F. C.; Hamel, D. J.; Kuhn, P.; Handel, T. M.; Cherezov, V.; Stevens, R. C. Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* **2010**, *330*, 1066−1071.

(11) Chien, E. Y.; Liu, W.; Zhao, Q.; Katritch, V.; Han, G. W.; Hanson, M. A.; Shi, L.; Newman, A. H.; Javitch, J. A.; Cherezov, V.; Stevens, R. C. Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* **2010**, *330*, 1091−1095.

(12) Shimamura, T.; Shiroishi, M.; Weyand, S.; Tsujimoto, H.; Winter, G.; Katritch, V.; Abagyan, R.; Cherezov, V.; Liu, W.; Han, G. W.; Kobayashi, T.; Stevens, R. C.; Iwata, S. Structure of the human histamine H1 receptor complex with doxepin. *Nature* **2011**, *475*, 65−70.

(13) Kruse, A. C.; Hu, J.; Pan, A. C.; Arlow, D. H.; Rosenbaum, D. M.; Rosemond, E.; Green, H. F.; Liu, T.; Chae, P. S.; Dror, R. O.; Shaw, D. E.; Weis, W. I.; Wess, J.; Kobilka, B. K. Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature* **2012**, *482*, 552−556.

(14) Haga, K.; Kruse, A. C.; Asada, H.; Yurugi-Kobayashi, T.; Shiroishi, M.; Zhang, C.; Weis, W. I.; Okada, T.; Kobilka, B. K.; Haga, T.; Kobayashi, T. Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature* **2012**, *482*, 547−51.

(15) Manglik, A.; Kruse, A. C.; Kobilka, T. S.; Thian, F. S.; Mathiesen, J. M.; Sunahara, R. K.; Pardo, L.; Weis, W. I.; Kobilka, B. K.; Granier, S.: Crystal structure of the micro-opioid receptor bound to a morphinan antagonist. *Nature* **2012**. March 21. DOI: 10.1038/nature10954. [Epub ahead of print.]

(16) Wu, H.; Wacker, D.; Mileni, M.; Katritch, V.; Han, G. W.; Vardy, E.; Liu, W.; Thompson, A. A.; Huang, X. P.; Carroll, F. I.; Mascarella, S. W.; Westkaemper, R. B.; Mosier, P. D.; Roth, B. L.; Cherezov, V.; Stevens, R. C.: Structure of the human kappa-opioid receptor in complex with JDTic. *Nature* **2012**. March 21. DOI: 10.1038/nature10939. [Epub ahead of print.]

(17) Hanson, M. A.; Roth, C. B.; Jo, E.; Griffith, M. T.; Scott, F. L.; Reinhart, G.; Desale, H.; Clemons, B.; Cahalan, S. M.; Schuerer, S. C.; Sanna, M. G.; Han, G. W.; Kuhn, P.; Rosen, H.; Stevens, R. C. Crystal structure of a lipid G protein-coupled receptor. *Science* **2012**, *335*, 851−855.

(18) Katritch, V.; Abagyan, R. GPCR agonist binding revealed by modeling and crystallography. *Trends Pharmacol. Sci.* **2011**, *32*, 637−643.

(19) Wacker, D.; Fenalti, G.; Brown, M. A.; Katritch, V.; Abagyan, R.; Cherezov, V.; Stevens, R. C. Conserved binding mode of human beta2 adrenergic receptor inverse agonists and antagonist revealed by X-ray crystallography. *J. Am. Chem. Soc.* **2010**, *132*, 11443−11445.

(20) Katritch, V.; Cherezov, V.; Stevens, R. C. Diversity and modularity of G protein-coupled receptor structures. *Trends Pharmacol. Sci.* **2012**, *33*, 17−27.

(21) Bhasin, M.; Raghava, G. P. GPCRpred: an SVM-based method for prediction of families and subfamilies of G-protein coupled receptors. *Nucleic Acids Res.* **2004**, *32*, W383−W389.

(22) Davies, M. N.; Flower, D. R. In silico identification of novel G protein coupled receptors. *Methods Mol. Biol.* **2009**, *528*, 25−36.

(23) Peng, Z. L.; Yang, J. Y.; Chen, X. An improved classification of G-protein-coupled receptors using sequence-derived features. *BMC Bioinf.* **2010**, *11*, 420.

(24) 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.

(25) van der Horst, E.; Peironcely, J. E.; Ijzerman, A. P.; Beukers, M. W.; Lane, J. R.; van Vlijmen, H. W.; Emmerich, M. T.; Okuno, Y.;

Bender, A. A novel chemogenomics analysis of G protein-coupled receptors (GPCRs) and their ligands: A potential strategy for receptor de-orphanization. *BMC Bioinf.* **2010**, *11*, 316.
(26) Weill, N.; Rognan, D. Development and validation of a novel protein-ligand fingerprint to mine chemogenomic space: application to G protein-coupled receptors and their ligands. *J. Chem. Inf. Model.* **2009**, *49*, 1049−1062.
(27) Joost, P.; Methner, A.: Phylogenetic analysis of 277 human G-protein-coupled receptors as a tool for the prediction of orphan receptor ligands. *Genome Biol.* **2002**, *3*, RESEARCH0063.
(28) Jacob, L.; Hoffmann, B.; Stoven, V.; Vert, J. P. Virtual screening of GPCRs: an in silico chemogenomics approach. *BMC Bioinf.* **2008**, *9*, 363.
(29) Sanders, M. P.; Fleuren, W. W.; Verhoeven, S.; van den Beld, S.; Alkema, W.; de Vlieg, J.; Klomp, J. P. ss-TEA: Entropy based identification of receptor specific ligand binding residues from a multiple sequence alignment of class A GPCRs. *BMC Bioinf.* **2011**, *12*, 332.
(30) Gloriam, D. E.; Foord, S. M.; Blaney, F. E.; Garland, S. L. Definition of the G protein-coupled receptor transmembrane bundle binding pocket and calculation of receptor similarities for drug design. *J. Med. Chem.* **2009**, *52*, 4429−4442.
(31) Bondensgaard, K.; Ankersen, M.; Thogersen, H.; Hansen, B. S.; Wulff, B. S.; Bywater, R. P. Recognition of privileged structures by G-protein coupled receptors. *J. Med. Chem.* **2004**, *47*, 888−899.
(32) Jain, A. N.; Nicholls, A. Recommendations for evaluation of computational methods. *J. Comput.-Aided Mol. Des.* **2008**, *22*, 133−139.
(33) Konc, J.; Janezic, D. ProBiS algorithm for detection of structurally similar protein binding sites by local structural alignment. *Bioinformatics* **2010**, *26*, 1160−1168.
(34) Shulman-Peleg, A.; Shatsky, M.; Nussinov, R.; Wolfson, H. J. MultiBind and MAPPIS: webservers for multiple alignment of protein 3D-binding sites and their interactions. *Nucleic Acids Res.* **2008**, *36*, W260−W264.
(35) Wichard, J. D.; Ter Laak, A.; Krause, G.; Heinrich, N.; Kuhne, R.; Kleinau, G. Chemogenomic analysis of G-protein coupled receptors and their ligands deciphers locks and keys governing diverse aspects of signalling. *PLoS ONE* **2011**, *6*, e16811.
(36) Okuno, Y.; Tamon, A.; Yabuuchi, H.; Niijima, S.; Minowa, Y.; Tonomura, K.; Kunimoto, R.; Feng, C. GLIDA: GPCR-ligand database for chemical genomics drug discovery—Database and tools update. *Nucleic Acids Res.* **2008**, *36*, D907−D912.
(37) Vroling, B.; Sanders, M.; Baakman, C.; Borrmann, A.; Verhoeven, S.; Klomp, J.; Oliveira, L.; de Vlieg, J.; Vriend, G. GPCRDB: Information system for G protein-coupled receptors. *Nucleic Acids Res.* **2011**, *39*, D309−D319.
(38) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M. Y.; Pieper, U.; Sali, A.: Comparative protein structure modeling using Modeller. *Curr. Protoc. Bioinf.* **2006**, *Chapter 5*, Unit 5.6.
(39) Jones, G.; Willett, P.; Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.* **1995**, *245*, 43−53.
(40) Greenfeder, S.; Cheewatrakoolpong, B.; Billah, M.; Egan, R. W.; Keene, E.; Murgolo, N. J.; Anthes, J. C. The neurokinin-1 and neurokinin-2 receptor binding sites of MDL103,392 differ. *Bioorg. Med. Chem.* **1999**, *7*, 2867−2876.
(41) Feighner, S. D.; Howard, A. D.; Prendergast, K.; Palyha, O. C.; Hreniuk, D. L.; Nargund, R.; Underwood, D.; Tata, J. R.; Dean, D. C.; Tan, C. P.; McKee, K. K.; Woods, J. W.; Patchett, A. A.; Smith, R. G.; Van der Ploeg, L. H. Structural requirements for the activation of the human growth hormone secretagogue receptor by peptide and nonpeptide secretagogues. *Mol. Endocrinol.* **1998**, *12*, 137−145.
(42) 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.
(43) Bissantz, C.; Bernard, P.; Hibert, M.; Rognan, D. Protein-based virtual screening of chemical databases. II. Are homology models of G-protein coupled receptors suitable targets? *Proteins* **2003**, *50*, 5−25.
(44) Ballesteros, J.; Palczewski, K. G protein-coupled receptor drug discovery: Implications from the crystal structure of rhodopsin. *Curr. Opin. Drug Discovery Dev.* **2001**, *4*, 561−574.
(45) Visiers, I.; Ballesteros, J. A.; Weinstein, H. Three-dimensional representations of G protein-coupled receptor structures and mechanisms. *Methods Enzymol.* **2002**, *343*, 329−371.