Published in final edited form as:

J Mol Graph Model. 2019 January ; 86: 235–246. doi:10.1016/j.jmglm.2018.10.016.

GPCR homology model template selection benchmarking: Global versus local similarity measures

Paige N. Castleman¹, Chandler Sears¹, Judith A. Cole², Daniel L. Baker¹, and Abby L. Parrill¹

¹The University of Memphis, Department of Chemistry and Computational Research on Materials Institute (CROMIUM)

²The University of Memphis, Department of Biological Sciences

Abstract

G protein-coupled receptors (GPCR) are integral membrane proteins of considerable interest as targets for drug development. GPCR ligand interaction studies often have a starting point with either crystal structures or comparative models. The majority of GPCR do not have experimentally-characterized 3-dimensional structures, so comparative modeling is a good structure-based starting point. Comparative modeling is a widely used method for generating models of proteins with unknown structures by analogy to crystallized proteins that are expected to exhibit structural conservation. Traditionally, comparative modeling template selection is based on global sequence identity and shared function. However high sequence identity localized to the ligand binding pocket may produce better models to examine protein-ligand interactions. This *in silico* benchmark study examined the performance of a global versus local similarity measure applied to comparative modeling template selection for 6 previously crystallized, class A GCPR (CXCR4, FFAR1, NOP, P2Y12, OPRK, and M1) with the long-term goal of optimizing GPCR ligand identification efforts. Comparative models were generated from templates selected using both global and local similarity measures. Similarity to reference crystal structures was reflected in RMSD values between atom positions throughout the structure or localized to the ligand binding pocket. Overall, models deviated from the reference crystal structure to a similar degree regardless of whether the template was selected using a global or local similarity measure. Ligand docking simulations were performed to assess relative performance in predicting protein-ligand complex structures and interaction networks. Calculated RMSD values between ligand poses from docking simulations and crystal structures indicate that models based on locally selected templates give docked poses that better mimic crystallographic ligand positions than those based on globally-selected templates in five of the six benchmark cases. However, protein model refinement strategies in advance of ligand docking applications are clearly essential as the average RMSD between crystallographic poses and poses docked into local template models was 9.7 Å and typically less than half of the ligand interaction sites are shared between the docked and

Corresponding Author: Abby L. Parrill.

Present/Permanent Address: 108 Scates Hall, University of Memphis, Memphis, TN 38152

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

crystallographic poses. These data support the utilization of local similarity measures to guide template selection in protocols using comparative models to investigate ligand-receptor interactions.

Keywords

Comparative Modeling; Comparative Protein Modeling; Homology Modeling; Ligand Identification; GPCR; Deorphanization; Template Selection; G Protein-Coupled Receptor

INTRODUCTION

G protein-coupled receptors (GPCR) are 7-transmembrane (TM), integral membrane proteins that play critical roles in cell signaling¹ and consequently are common targets of pharmaceutical agents. GPCR are divided into five subfamilies, *Rhodopsin*, *Adhesion*, *Secretin*, *Glutamate*, and *Frizzled*.² All GPCR share common characteristics of an extracellular amino-terminus, 7 transmembrane alpha-helical domains, 3 extracellular loops, 3 intracellular loops, and an intracellular carboxy-terminus.² For GPCR, the transmembrane domains have conserved residues and motifs throughout the majority of GPCR subfamilies.² The Ballesteros-Weinstein numbering system highlights correspondences between GPCR sequences.² In this system, the most conserved residue within each transmembrane domain is indexed as the TM.50 residue where *TM* is replaced by the number of the transmembrane domain.² For example, the conserved residue in the first transmembrane domain is an aspartic acid which is termed D1.50 residue. Each amino acid position from this most conserved residue closer to the amino-terminus decreases the index number (e.g. TM.49 then TM.48 and so on). Likewise, each amino acid position from the most conserved residue closer to the carboxy-terminus increases the index number (TM.51 then TM.52 and so on). While the transmembrane regions of the different GPCR subfamilies do have conserved residues, each GPCR subfamily has distinct characteristics within the TM regions that outline the ligand binding pocket.² These well-known distinctions allow different members of the family to discriminate between signaling molecules.²

The critical role of GPCR in cell signaling processes has led to considerable research into GPCR interactions with agonists (activating ligands), antagonists (blocking ligands) and inverse agonists (deactivating ligands).² Only 50 unique members of the GPCR family are represented by experimentally-determined atomic resolution structures in the Protein Data Bank³ as of June 15, 2018 (<http://gpdb.org/structure/statistics>). This is less than 10% of all human GPCR family members. Structure-based studies are becoming increasingly popular with the rise of reverse pharmacology, in which ligand discovery efforts are guided by three-dimensional structures of the biomolecular target.⁴ For GPCR with unresolved three-dimensional structures, computational methods are cost-effective starting points. Comparative modeling, commonly known as homology modeling, is a popular computational tool that generates models of proteins with unknown three-dimensional (3D) structure by analogy to a protein expected to have similar structure.⁵ An expectation that two proteins will share similar 3D structure is generally justified by high primary sequence similarity and shared function. This combination indicates that divergent evolution has not

dramatically altered the structures.⁵ Generally, a percent identity of greater than 25%–30% is acknowledged to produce “good” homology models.^{6–8} Once a comparative model is built, it is typically refined and validated before being utilized to make predictions.⁵ Typical refinements include sampling of loop conformations, particularly that of extracellular loop 2 (EL2) which shows high length and structural variability among the known members of the family, as well as molecular dynamics to generate an ensemble of candidate structures to use as docking targets.^{9,10} Comparative modeling continues to provide more accurate models than template-free modeling methods.^{9,10} Comparative models are built by a three step process in advance of any refinements: template selection, sequence alignment, and automated comparative modeling.⁵ This work focuses on the outcome of these three steps, and seeks to identify whether a localized similarity measure provides an advantage over a global similarity measure prior to any further refinements. This focus is based on the observation that refinement methods provide consistent, but small, improvements over initial models.^{9,10} Thus a better initial model will produce a better outcome after refinement.

Ligand interaction studies of GPCR require homology models that most accurately reflect the geometry in and around the ligand binding pocket. Model accuracy and ability to predict ligand binding geometry differ with different template selection methods. We predict that a template with a high sequence identity localized within the ligand binding pocket could be used to develop better models of the ligand binding pocket than templates selected based on global sequence similarity. This is based on the premise that many of the crystallized GPCR might give reasonable models of the TM domain due to conservation of backbone structure throughout the GPCR superfamily (Figure 1). However, these crystallized GPCR may not have the same level of accuracy within the binding pocket due to the well-known primary sequence differences within the binding pocket of different subfamilies.

“CoINPocket” analysis is a local similarity score developed by Ngo et. al¹¹ that could prove to be a new basis for homology model template selection optimized for investigations of ligand binding. This local similarity score was developed to aid in surrogate ligand identification for orphan GPCR by enhancing GPCR binding pocket comparisons through addition of ligand contact strengths from the GPCR Pocketome.¹² By weighting residue contributions to a score based on interaction strengths within the ligand binding pocket of a representative set of GPCR, screening for pharmacological neighbors of a relatively poorly-characterized GPCR without the use of existing ligand-receptor interactions becomes possible.¹¹ The ligand binding pockets of 27 class A GPCRs were analyzed and 61 residues surrounding those binding pockets were characterized. Eight of these residues were found to have high interaction strengths across 70% or more of the class A GPCR Pocketome¹² entries.¹¹ Pairs of human GPCR were given local similarity scores weighted to emphasize similarities at sites of high interaction strengths.¹¹ While this method did have success in identifying the first small molecule ligands of the GPR37L1 orphan receptor by searching for pharmacological neighbors, we propose that this local similarity score developed by Ngo et. al¹¹ can also be used in template selection for homology modeling because it is implied that GPCR with pharmacological similarities also have structurally similar binding sites.

This *in silico* benchmark study compared the performance of two similarity measures in template selection: global sequence identity and a local similarity score. We deliberately

avoid multi-template modeling in this benchmark to allow simple head-to-head performance comparisons. Performance was assessed based on both the accuracy of the resulting model GPCR structure and subsequent ligand binding geometry prediction from docking experiments. Homology models for 6 crystallized class A GPCR (chemokine receptor type 4 (CXCR4), free fatty acid receptor 1 (FFAR1), nociceptin opioid receptor (NOP), P2Y purinoceptor 12 (P2Y12), kappa opioid receptor (OPRK), and muscarinic receptor 1 (M1) were developed using templates selected by both global and local similarity measures. The long-term goal of this effort is to optimize comparative modeling of GPCR so that ligand identification, ligand binding mechanism of action, and other structure-based ligand interaction studies guided by GPCR models more closely reflect native ligand binding modes. Overall, our results show that homology models constructed from a locally selected template yielded better models of the ligand binding pocket and more accurate ligand position than those constructed from a globally selected template. However, our results also highlight the necessity of utilizing structure refinement or sampling strategies after comparative model construction as the docking results into models are not satisfactorily close to the crystallographic reference poses.

METHODOLOGY

Template Selection:

This study compares two similarity measures for template selection, a local similarity measure and a global similarity measure. From this point forward, template selection based on the local similarity score will be referred to as the local method and template selection based on the global similarity score as the global method. The local similarity measure used to determine the local template was the “CoINPocket” score developed by Ngo et al.¹¹ The global template was selected based on global percent amino acid identity, the global similarity measure, using the GPCRdb receptor similarity search tool.¹³ For each target GPCR, all crystallized human GPCRs were examined to select a template with the highest local similarity score and another template with the highest global similarity score (excluding crystal structures of the target GPCR and those of closely related GPCR known to recognize the same endogenous ligand). G protein coupled receptors with two templates selected using these different similarity measures that showed the largest difference in global similarity scores or local similarity scores were chosen for the study. Two comparative models were then constructed for each GPCR based on the selected templates.

Homology Model Construction:

Selected template sequences were aligned to the sequence of the target GPCR to be modeled using two procedures. The first procedure started with automatic alignment in MOE 2016.0802.¹⁴ After the automatic alignment, a manual adjustment was performed. Any automatic alignment should be manually checked as the quality of the sequence alignment impacts the quality of the resulting comparative model.⁸ The manual alignment adjustment was performed by ensuring alignment of TM.50 residues and shifting gaps from TM segments into the intra- and extracellular loops where greater structural differences both in terms of amino acid number and identity are observed to occur. The alignment generated by this procedure is designated the structure-independent alignment. Alternatively, the

structure-based alignment available in the GPCR database (<http://gpcrd.org>) was utilized, and is designated the structure-based alignment. Homology models were generated from both alignments using the automated homology modeling interface in MOE.¹⁴

Homology Model Analysis:

In general, homology model accuracy depends upon the alignment step of the homology model construction process.⁸ After model construction, the model quality and accuracy was assessed. Ramachandran plots were analyzed for outliers as a first pass assessment of protein structure quality. Two additional validation processes were used in this study. First, the quality of GPCR models constructed from each template was assessed by superposing each homology model onto its respective crystal structures followed by calculation of the root-mean-square deviation (RMSD). Global and local alpha carbon and all atom RMSD values were calculated for each homology model. Local RMSD values reflected superposition and RMSD values for either the 8 sites involved in ligand binding in 70% or more class A GPCR Pocketome¹² entries¹¹ (PRMSD₈, **bold** in Table 1) or all 61 residues in the ligand binding pocket (PRMSD₆₁, Table 1). Second, ligands present in the crystallographic reference structures were docked into the homology models for comparison to crystallographic ligand positions. Model validation for the purpose of studying ligand interactions was assessed through calculation of all atom docked pose RMSD relative to the crystallographic ligand position.

Ligand Docking and Pose Selection:

Ligand docking was performed in MOE.¹⁴ Ligands from the reference crystal structures were docked back into their respective crystal structure and into their respective models as well as into an additional crystal structure of the target receptor as a cross-docking method when available. Two-dimensional representations of all ligands used are shown in Figure 2.

Protein structures were prepared for docking by checking for proper residue protonation states at physiological pH (7.4 pH), fixing any chain breaks and correcting charges. Ligands were prepared for docking by ensuring proper protonation/charge for physiological pH and performing an energy minimization using the AMBER10:EHT forcefield. After the proteins were prepped, the ligand binding site was identified using the “Site Finder” function in MOE.¹⁴ The Site Finder function is a geometric method to identify open cavities in a protein structure using an adaptation of the Alpha Spheres method reported by Edelsbrunner, et al.¹⁵ Two variations of ligand docking were performed—rigid receptor docking and induced fit docking. Both docking methods were performed to initially retain and rescore the top 400 docking poses for each analysis to produce a final top 5 poses. The top 5 final ligand poses for each docking simulation were evaluated visually to select a single top pose.

Pose selection was performed because the top 5 poses usually exhibit score differences within the expected margin of error. Automated selection of best ligand pose from the top five ranked poses was based on the proportion of ligand hydrogen bond interaction sites involved in strong hydrogen bonds. The pose with the highest proportion of strongly complemented hydrogen bonding sites was selected for further analyses. If there was a tie

for the best pose, the pose with the better dock score was selected. The use of an automated pose selection method eliminates bias that could occur from manual pose selections.

RESULTS AND DISCUSSION

GPCR Selection:

Homology modeling templates for 36 GPCR were identified based on both local and global similarity measures (Table S2). Only one-third of these receptors had a single candidate template that was ranked most similar by both measures. The receptors chosen as benchmark targets were GPCR for which different templates ranked most similar by the local and global similarity measures, and that had either a large difference in protein sequence identity (global similarity) or a large difference in local similarity score between the templates selected by both methods. The GPCR chosen were selected from the Class A GPCR subfamily as more Class A GPCR have been crystallized. Therefore, only two of the top four receptors chosen by these criteria were included, as all were members of the opioid receptor family. Four additional non-opioid receptors were selected to create a benchmark set containing a total of six receptors (CXCR4, FFAR₁, NOP, P2Y₁₂, M₁ and OPRK, Table 2). Amino acid sequences for the six GPCR were downloaded from GenBank¹⁶ and the reference and template structures were downloaded from the RCSB Protein Databank (Table 3).³

Homology Model Analysis:

Homology models based on both local and global templates for each receptor were generated. Ramachandran plots were calculated to assess protein geometry. The majority of points in the phi-psi (Ramachandran) plot fell within core and allowed regions. Only two outliers that fell outside this region were found within a 4.5 Å radius of the bound ligand when models were superimposed on the corresponding crystallographic reference structure. One of these outliers involved residue C200 in the NOP model based on the local template from the structure-based alignment. The other involved D217 in the OPRK model based on the local template from the structure-independent alignment. Both of the outliers near the ligand binding pockets were found in extracellular loop 2, a region of GPCR structures that shows high variability in length and structure. This would be a typical region that would be further sampled and refined in most studies focused on modeling GPCR ligand interaction. Since this study aimed to evaluate unmanipulated homology models built from the application of local versus global template selection methods and the number of outliers was small, optimization of the geometry around these two outliers was not performed.

Homology model quality was first assessed by comparison of model structures to crystallographic reference structures. To provide context for homology model quality assessments, different crystal structures of the target proteins were compared for the five benchmark receptors represented in the PDB with multiple structures (Table 5). Note that in cases where more than two crystal structures of a single GPCR were available, the structure with the greatest difference from the reference structure has been included to illustrate the experimental structural variability observed. These comparisons show that alpha carbon RMSD values between two crystal structures of the same receptor range from 0.63 to 3.32

Å. This range of values indicates the experimentally observed variability for a given GPCR, and sets an appropriate expectation for homology model RMSD values. The structural differences in the case of the OPRK crystal structures reflect different receptor activation states, and are more pronounced at the intracellular end of the structures, distant from the ligand binding pocket (Supplemental Figure 1A). In contrast, the difference between the P2Y12 crystal structures is due to dramatically different ligand structures and ligand binding modes that result in differences that are more pronounced in the vicinity of the ligand binding pocket (Supplemental Figure 1B). The extracellular end of TM6 in the reference crystal structure is displaced relative to the position of the corresponding helical segment in most known GPCR structures due to ligand position between TM6 and TM7. This ligand pose represents the greatest challenge to reproduce among the benchmark set using either alternative crystal structures of FFAR1 (cross-docking) or comparative models of FFAR1 due to this helical displacement. Any model with global alpha carbon RMSD values compared to reference crystal structures up to 3.3 Å should be considered to fall within the observed experimental variation based on the data in Table 5. RMSD values an additional 2 Å beyond this value, or 5.3 Å, will be used to designate acceptable models. Both global and ligand binding pocket weighted superpositions of the homology models to their respective reference crystal structures were prepared. Results from the comparisons of global and binding pocket weighted RMSD calculations are shown in Table 4.

The results in Table 4 demonstrate that models with RMSD values that fall within the experimentally-observed variability between different crystal structure of the same receptor were obtained in five cases. NOP models generated using structure-based alignments from templates chosen using both local and global similarity measures had RMSD values less than the 3.3 Å RMSD observed between two different crystal structures of OPRK. Additionally, three of four models of M1 also had RMSD values relative to the reference structure below this threshold. In addition to the two targets for which models fell within the experimentally-observed variability, models of three additional targets were obtained that had RMSD values no more than 2 Å higher than the experimental variability. All models generated for CXCR4 and OPRK meet the 5.3 Å threshold for models within an acceptable range above the experimental variability. In the case of the P2Y12 models, only those generated using the template selected using local similarity measures met this threshold. The only target for which no acceptable models were generated was the FFAR1 receptor, for which the reference crystal structure has already been described as having structural differences relative to the majority of known GPCR crystal structures in the position of the extracellular end of TM6 due to the ligand protrusion between TM6 and TM7. We had expected that global structure comparisons (RMSD) would provide lower RMSD values for models based on templates selected using the global sequence similarity measure, but the results demonstrate this is not consistently the case. We had also expected that structure comparisons localized to the binding pocket (PRMSD₆₁ AND PRMSD₈) would provide lower values for models based on templates selected using the local sequence similarity measure, but the results demonstrate that this is also not the case.

Figure 3 shows a representation of the 61 and 8 residues of the P2Y12 ligand binding pocket that were used to calculate the PRMSD₆₁ and PRMSD₈ values and a superposition of the three crystal and four P2Y12 model structures used or generated in this study. P2Y12 was

selected as it has relatively high experimental structural variability (RMSD between two different crystal structures shown using red and teal ribbons in Figure 3B was 2.65 Å), and a relatively high difference in RMSD between the local (3.75 and 3.97 Å by the structure-independent and structure-based alignments) and global (5.48 and 5.33 Å by the structure-independent and structure-based alignments) template models when superposed on the reference crystal structure. The two crystal structures show good structural similarity throughout the TM region, with differences localized to the intracellular end of TM6 (better illustrated in Figure S1B). The local (grey and yellow in Figure 3B) and global (blue and orange in Figure 3B) template models also show considerable structural similarity to the reference crystal structure throughout the TM region, although the deviations between the global template models and the crystal structures are more apparent. Superpositions of the other 5 receptors used in this study can be found in Figure S2.

Ligand Docking:

As an initial test of the ligand fit within the ligand binding pockets in each crystallographic and model structure, the crystallographic pose was transferred into each structure and the induced fit score (allowing limited sidechain flexibility in the vicinity of the ligand) was calculated (Table 5). Unsurprisingly, the crystallographic pose generally shows the lowest scores in the crystallographic protein structure it was originally from. In two of five cases, the ligand pose in an alternate crystal structure scored within 0.5 kcal/mol of that in the original reference crystal structure. These are also the two cases in which the RMSD between the two crystal structures was less than 1 Å (NOP and P2Y12, Table 4). Notably, for the second alternate crystal structure of P2Y12, the score is notably worse (−7.9 kcal/mol compared to −12.7 kcal/mol) as might be expected from the substantially larger difference in protein structure relative to the reference (2.65 Å, Table 4). Only one model produced a score for the crystallographic ligand pose within 0.5 kcal/mol of that in the reference, the CXCR4 global template model based on the structure-independent alignment. This model structure was not the closest to the crystallographic reference structure, either globally or within the ligand binding pocket, according to the metrics in Table 4. As previously noted, the FFAR1 ligand occupies an uncommon position protruding between TM6 and TM7 with a concomitant unusual TM6 geometry. This ligand pose scores poorly (positive scores) in all other structures, with the worst scores observed for the models produced from structure-dependent alignments.

Models were subsequently assessed as docking targets. Ligands and protein structures were prepared and docked with induced fit and rigid docking methods as described. Ligand poses were selected automatically based on proportion of hydrogen bonding groups complemented by strong hydrogen bonds to the docking target without reference to the crystal structures to eliminate bias.

The ligand docking protocol was evaluated for success by calculating the RMSD between the crystallographic ligand position and the best docked pose into the reference crystal structure (Table 6). Figure 4 shows superpositions of induced fit docked and crystallographic poses for three examples which span the range of RMSD values between the induced fit docked pose and the reference crystallographic pose, OPRK (RMSD=0.65 Å), FFAR1

(RMSD=2.42 Å), and CXCR4 (RMSD=7.71 Å). This figure shows that even with an RMSD of 2.42 Å (panels B and E) considerable overlap in ligand position and orientation exists. At the extremes, an RMSD of 7.71 Å can indicate reasonable volume overlap as in panels C and F with clear orientation problems, or could indicate a difference in ligand position rather than orientation (example not shown). The similarity between the receptor residues contacted by the ligand in the crystallographic and docked poses is reflected by the Tanimoto coefficient calculated from the lists of first neighbors to the ligand in the residue interaction network calculated using the RING 2.0 server.¹⁷ The considerable volume overlap among the poses is reflected in similarity scores ranging from 0.83 to 0.94. Ligand position RMSD values averaged 2.6 and 2.7 for the set of six receptors for rigid and induced fit docking, respectively. The similarity between receptor residues contacted by the ligand in the crystallographic reference and docked poses was consistently higher from the induced fit docking protocol, thus only induced fit results will be reported when docking into model structures. The highest RMSD values were observed for ITD docking into the CXCR4 crystal structure, highlighting ITD as a challenging case with respect to orientation within the binding site, although not in correct localization in the binding pocket, as reflected in the high Tanimoto coefficient (0.88) for ligand contact sites.

The ligands from the reference crystal structures were each docked using the induced fit protocol into three to six target structures, depending on whether additional crystal structures of the reference protein were available and whether the structure-independent alignment differed from the structure-based alignment from GPCRdb. The targets included a second crystal structure (when available) that did not have the reference ligand bound (such as PDB entry 5DHH for NOP bound to (5S,7S)-7-[4-(2,6-dichlorophenyl)piperidin-1-yl]methyl]-1-methyl-6,7,8,9-tetrahydro-5H-benzo[7]annulen-5-ol instead of ONN), as well as the local and global template models derived from both alignment procedures. Several of the benchmark GPCR had more than two crystal structures available. In these cases, the crystal structure exhibiting the largest RMSD relative to the reference structure was selected to represent a scenario in which the only available structure is not ideal for one reason or another. In the case of the P2Y12 receptor, an additional best case scenario cross-docking target bound to a very similar ligand (6AD instead of 6AT differing by one phosphate group) was also utilized for contrast. Comparisons between the docked and reference poses are shown in Table 7. Pose comparisons include all-atom RMSD between ligand atom positions and the Tanimoto similarity coefficient between the first neighbor nodes of the ligand in the residue interaction network generated by the RING server.¹⁷

The cross-docking results, in which the ligand was docked into a crystal structure of the target protein that did not have that ligand bound, represent the typical ‘gold standard’ for the study of protein-ligand interactions when crystallographic complexes with the ligand of interest are not available. Table 7 demonstrates that the induced fit poses in the cross-docking results show RMSD values ranging from 3.70 to 13.47 Å for these targets. The average RMSD (7.2 Å) for the cross-docking results is substantially higher than the average RMSD observed when docking back into the same crystal structure the ligand came from. This result is not surprising in the context of our current understanding that complex formation is an induced fit process, which may involve larger-scale changes in protein structure than current induced fit docking protocols can capture. However, this result does

underscore the importance of including some refinement or sampling process either on the protein structure prior to docking or on the complex structures obtained from docking even when docking into a crystal structure. The average RMSD value for induced fit docking into the local and global template models derived from the structure-independent alignments over the same five targets for which cross-docking was possible are higher (9.7 and 12.7 Å, respectively) than the average RMSD from cross-docking (7.2 Å), with the global template models giving substantially greater deviation from the crystallographic pose than the local template models. The same trend was observed over all six benchmark proteins, with average ligand pose RMSD values for global template models 3 Å higher for the global template models than the local template models generated using either alignment method. Figure 5 illustrates the docking poses obtained for OPRK relative to the reference crystal structure, due to the large range in ligand pose RMSD values (from 5.05 to 16.11 Å) and Tanimoto coefficients reflecting ligand interactions (from 0.58 to 0.07). The greatest difference from the crystallographic pose is observed when JDC is docked into the model constructed from the global template using the structure-based alignment, with an RMSD of 16.11 Å, Tanimoto coefficient of 0.07, and a ligand position that is clearly far above the ligand binding pocket (Figure 5F). The smallest difference from the crystallographic pose is observed when JDC is docked into the model constructed from the local template using the structure-independent alignment, with an RMSD of 5.05 Å, and a Tanimoto coefficient of 0.51, and a ligand occupying a similar volume with the central isopropyl substituent pointed in the opposite direction (Figure 5B). These results strongly support our premise that studies of ligand interactions should use a template selection process biased toward similarities within the ligand binding pocket and clearly illustrate that homology models of GPCR generated without refinement, sidechain sampling within the binding pocket, or conformational sampling of extracellular loop 2 are not sufficient to produce desirably accurate ligand orientations within the binding pocket. Localization of ligands within the binding pocket was substantially better, as reflected by Tanimoto coefficients between residues interacting with the ligand typically higher than 0.3 and as high as 0.6.

CONCLUSIONS

The protein structure similarity measures (RMSD, RMSD₆₁, RMSD₈) reported in table 4 indicate that multiple templates are capable of producing GPCR models that fall within 2 Å of the 3.3 Å observed variability between different crystal structures of the same GPCR. Templates chosen based on the global similarity measure (percent sequence identity) did not consistently produce models with greater structural similarity to the reference crystal structure, suggesting that sequence identity should not be the sole metric considered in choosing comparative modeling templates, either for single-template or multiple-template modeling strategies. Ligand docking into the reference crystal structure resulted in ligand poses that shared highly similar interaction networks to the experimental complex, and ligand pose RMSD values averaged 2.7 Å. Ligand docking into alternative crystal structures and models gave substantially lower interaction network similarities and substantially higher RMSD values, indicating the importance of protein structure refinement and/or sampling strategies to be integrated into studies of protein-ligand interactions and ligand discovery pipelines. Nevertheless, given the consistent but modest structural improvements that stem

from such refinements,^{9,10} it is important to design comparative modeling approaches that provide higher performing starting structures than current standard approaches as starting points for the subsequent refinements. The models constructed using templates selected using the GPCR “CoINPocket”¹¹ local similarity score, weighted to emphasize similarities at sites consistently observed to strongly interact with GPCR ligands, showed a performance advantage for ligand docking studies over models constructed using the traditional global similarity measure, percent sequence identity. The performance advantage was reflected in both metrics used to assess ligand docking performance in Table 7. First, average Tanimoto coefficients between ligand interaction sites in model and reference crystal structures of 0.39 for local template models and 0.26 for global template models illustrate docking into local models gives better ligand localization within the binding pocket of models construction from templates with greater local similarity to the target GPCR. Second, average ligand pose RMSD relative to the reference crystallographic pose was 3 Å lower when docking into models constructed from local rather than global templates.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

Research reported in this publication was supported by the National Institute of Mental Health of the National Institutes of Health under Award Number R15MH109034. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

MOE 2016.0802 was used throughout this work and was provided as a courtesy of the Chemical Computing Group.

Funding: This work was supported by the National Institute of Health [grant R15 MH109034].

References

1. Pierce KL; Premont RT; Lefkowitz RJ Signalling: Seven-transmembrane receptors. *Nature Reviews Molecular Cell Biology* 2002, 3, 639–650. [PubMed: 12209124]
2. Gacasan SB; Baker DL; Parrill AL G protein-coupled receptors: the evolution of structural insight. *biophysics* 2017, Vol. 4, Pages 491–527 2017.
3. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE The Protein Data Bank. *Nucleic Acids Res.* 2000, 28, 235–242. [PubMed: 10592235]
4. Shaikh MF Reverse Pharmacology: Fast Track Path of Drug Discovery. *Pharmacy & Pharmacology International Journal* 2016, 4, 1–2.
5. Vyas VK; Ukawala RD; Ghate M; Chintla C Homology Modeling a Fast Tool for Drug Discovery: Current Perspectives. *Indian J Pharm Sci* 2012, 74, 1–17. [PubMed: 23204616]
6. Forrest LR; Tang CL; Honig B On the accuracy of homology modeling and sequence alignment methods applied to membrane proteins. *Biophys. J* 2006, 91, 508–517. [PubMed: 16648166]
7. Beuming T; Sherman W Current assessment of docking into GPCR crystal structures and homology models: successes, challenges, and guidelines. *Journal Of Chemical Information And Modeling* 2012, 52, 3263–3277. [PubMed: 23121495]
8. Nayeem A; Sitkoff D; Krystek J, Stanley A comparative study of available software for high-accuracy homology modeling: from sequence alignments to structural models. *Protein science : a publication of the Protein Society* 2006, 15, 808–824. [PubMed: 16600967]
9. Moulton J; Fidelis K; Kryzhanovskiy A; Schwede T; Tramontano A Critical assessment of methods of protein structure prediction: Progress and new directions in round XI. *Proteins: Structure, Function, and Bioinformatics* 2016, 84, 4–14.

10. Kufareva I; Katritch V; Stevens RC; Abagyan R Advances in GPCR modeling evaluated by the GPCR Dock 2013 assessment: meeting new challenges. *Structure* 2014, 22, 1120–1139. [PubMed: 25066135]
11. Ngo Tony; Ilatovskiy Andrey V; Stewart Alastair G; Coleman James L J; Mcrobb Fiona M; Peter Riek R; Graham Robert M; Abagyan Ruben; Kufareva Irina; Smith Nicola J Orphan receptor ligand discovery by pickpocketing pharmacological neighbors. *Nature Chemical Biology* 2017, 13, 235–242. [PubMed: 27992882]
12. Kufareva I, Ilatovskiy AV, Abagyan R Pocketome: an encyclopedia of small-molecule binding sites in 4D. *Nucleic Acids Research*, 40, D540.
13. Isberg V, Mordalski S, Munk C, Rataj K, Harpsøe K, Hauser AS, Vroiling B, Bojarski AJ, Vriend G, Gloriam DE GPCRdb: an information system for G protein-coupled receptors. *Nucleic Acids Res.* 2016, 44, D364.
14. Molecular Operating Environment (MOE), 2016 0802 Chemical Computing Group ULC. **2016**, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A, 2R7.
15. Edelsbrunner H; Facello M; Fu P; Liang J Measuring Proteins and Voids in Proteins In Proceedings of the 28th Hawaii International Conference on Systems Science; 1 1995; Vol. 5, pp 264.
16. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL GenBank. *Nucleic Acids Res.* 2005, 33, D38.
17. Piovesan D; Minervini G; Tosatto SCE The RING 2.0 web server for high quality residue interaction networks. *Nucleic Acids Res.* 2016, 44, 367.
18. Wang C; Jiang Y; Ma J; Wu H; Wacker D; Katritch V; Han GW; Liu W; Huang X; Vardy E; McCorvy JD; Gao X; Zhou XE; Melcher K; Zhang C; Bai F; Yang H; Yang L; Jiang H; Roth BL; Cherezov V; Stevens RC; Xu HE Structural basis for molecular recognition at serotonin receptors. *Science* 2013, 340, 610–614. [PubMed: 23519210]
19. Shimamura T; Shiroishi M; Weyand S; Tsujimoto H; Winter G; Katritch V; Abagyan R; Cherezov V; Liu W; Han GW; Kobayashi T; Stevens RC; Iwata S Structure of the human histamine H1 receptor complex with doxepin. *Nature* 2011, 475, 65–70. [PubMed: 21697825]
20. Zhang H; Unal H; Gati C; Han GW; Liu W; Zatsepin NA; James D; Wang D; Nelson G; Weierstall U; Sawaya MR; Xu Q; Messerschmidt M; Williams GJ; Boutet S; Yefanov OM; White TA; Wang C; Ishchenko A; Tirupula KC; Desnoyer R; Coe J; Conrad CE; Fromme P; Stevens RC; Katritch V; Karnik SS; Cherezov V Structure of the Angiotensin receptor revealed by serial femtosecond crystallography. *Cell* 2015, 161, 833–844. [PubMed: 25913193]
21. Zhang H; Han GW; Batyuk A; Ishchenko A; White KL; Patel N; Sadybekov A; Zamyatny B; Rudd MT; Hollenstein K; Tolstikova A; White TA; Hunter MS; Weierstall U; Liu W; Babaoglu K; Moore EL; Katz RD; Shipman JM; Garcia-Calvo M; Sharma S; Sheth P; Soisson SM; Stevens RC; Katritch V; Cherezov V Structural basis for selectivity and diversity in angiotensin II receptors. *Nature* 2017, 544, 327–332. [PubMed: 28379944]
22. Caruz A; Samsom M; Alonso JM; Alcamí J; Baleux F; Virelizier JL; Parmentier M; Arenzana-Seisdedos F Genomic organization and promoter characterization of human CXCR4 gene. *FEBS Lett.* 1998, 426, 271–278. [PubMed: 9599023]
23. Wu B; Chien EYT; Mol CD; Fenalti G; Liu W; Katritch V; Abagyan R; Brooun A; Wells P; Bi FC; Hamel DJ; Kuhn P; Handel TM; Cherezov V; Stevens RC Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* 2010, 330, 1066–1071. [PubMed: 20929726]
24. Qin L; Kufareva I; Holden LG; Wang C; Zheng Y; Zhao C; Fenalti G; Wu H; Han GW; Cherezov V; Abagyan R; Stevens RC; Handel TM Structural biology. Crystal structure of the chemokine receptor CXCR4 in complex with a viral chemokine. *Science* 2015, 347, 1117–1122. [PubMed: 25612609]
25. Strausberg RL; Feingold EA; Grouse LH; Derge JG; Klausner RD; Collins FS; Wagner L; Shenmen CM; Schuler GD; Altschul SF; Zeeberg B; Buetow KH; Schaefer CF; Bhat NK; Hopkins RF; Jordan H; Moore T; Max SI; Wang J; Hsieh F; Diatchenko L; Marusina K; Farmer AA; Rubin GM; Hong L; Stapleton M; Soares MB; Bonaldo MF; Casavant TL; Scheetz TE; Brownstein MJ; Ustin TB; Toshikiyuki S; Carninci P; Prange C; Raha SS; Loquellano NA; Peters GJ; Abramson RD; Mullahy SJ; Bosak SA; McEwan PJ; McKernan KJ; Malek JA; Gunaratne PH; Richards S; Worley KC; Hale S; Garcia AM; Gay LJ; Hulyk SW; Villalon DK; Muzny DM; Sodergren EJ; Lu

- X; Gibbs RA; Fahey J; Helton E; Ketteman M; Madan A; Rodrigues S; Sanchez A; Whiting M; Madan A; Young AC; Shevchenko Y; Bouffard GG; Blakesley RW; Touchman JW; Green ED; Dickson MC; Rodriguez AC; Grimwood J; Schmutz J; Myers RM; Butterfield YSN; Krzywinski MI; Skalska U; Smailus DE; Schnerch A; Schein JE; Jones SJM; Marra MA Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc. Natl. Acad. Sci. U. S. A* 2002, 99, 16899–16903. [PubMed: 12477932]
26. Srivastava A; Yano J; Hirozane Y; Kefala G; Gruswitz F; Snell G; Lane W; Ivetac A; Aertgeerts K; Nguyen J; Jennings A; Okada K High-resolution structure of the human GPR40 receptor bound to allosteric agonist TAK-875. *Nature* 2014, 513, 124–127. [PubMed: 25043059]
 27. Ho JD; Chau B; Rodgers L; Lu F; Wilbur KL; Otto KA; Chen Y; Song M; Riley JP; Yang H; Reynolds NA; Kahl SD; Lewis AP; Groshong C; Madsen RE; Connors K; Lineswala JP; Gheyi T; Saflor MD; Lee MR; Benach J; Baker KA; Montrose-Rafizadeh C; Genin MJ; Miller AR; Hamdouchi C Structural basis for GPR40 allosteric agonism and incretin stimulation. *Nat Commun* 2018, 9, 1645. [PubMed: 29695780]
 28. Allard WJ; Sigal IS; Dixon RA Sequence of the gene encoding the human M1 muscarinic acetylcholine receptor. *Nucleic Acids Res.* 1987, 15, 10604. [PubMed: 3697105]
 29. Thal DM; Sun B; Feng D; Nawaratne V; Leach K; Felder CC; Bures MG; Evans DA; Weis WI; Bachhawat P; Kobilka TS; Sexton PM; Kobilka BK; Christopoulos A Crystal structures of the M1 and M4 muscarinic acetylcholine receptors. *Nature* 2016, 531, 335–340. [PubMed: 26958838]
 30. Chien EYT; Liu W; Zhao Q; Katritch V; Han GW; Hanson MA; Shi L; Newman AH; Javitch JA; Cherezov V; Stevens RC Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist. *Science* 2010, 330, 1091–1095. [PubMed: 21097933]
 31. Matsushima A; Nishimura H; Matsuyama Y; Liu X; Costa T; Shimohigashi Y Specific affinity-labeling of the nociceptin ORL1 receptor using a thiol-activated Cys(Npys)-containing peptide ligand. *Biopolymers* 2016, 106, 460–469. [PubMed: 27271345]
 32. Thompson AA; Liu W; Chun E; Katritch V; Wu H; Vardy E; Huang X; Trapella C; Guerrini R; Calo G; Roth BL; Cherezov V; Stevens RC Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic. *Nature* 2012, 485, 395–399. [PubMed: 22596163]
 33. Miller RL; Thompson AA; Trapella C; Guerrini R; Malfacini D; Patel N; Han GW; Cherezov V; Caló G; Katritch V; Stevens RC The Importance of Ligand-Receptor Conformational Pairs in Stabilization: Spotlight on the N/OFQ G Protein-Coupled Receptor. *Structure* 2015, 23, 2291–2299. [PubMed: 26526853]
 34. Zhang C; Srinivasan Y; Arlow DH; Fung JJ; Palmer D; Zheng Y; Green HF; Pandey A; Dror RO; Shaw DE; Weis WI; Coughlin SR; Kobilka BK High-resolution crystal structure of human protease-activated receptor 1. *Nature* 2012, 492, 387–392. [PubMed: 23222541]
 35. Wu H; Wacker D; Mileni M; Katritch V; Han GW; Vardy E; Liu W; Thompson AA; Huang X; Carroll FI; Mascarella SW; Westkaemper RB; Mosier PD; Roth BL; Cherezov V; Stevens RC Structure of the human κ -opioid receptor in complex with JDTic. *Nature* 2012, 485, 327–332. [PubMed: 22437504]
 36. Che T; Majumdar S; Zaidi SA; Ondachi P; McCorvy JD; Wang S; Mosier PD; Uprety R; Vardy E; Krumm BE; Han GW; Lee M; Pardon E; Steyaert J; Huang X; Strachan RT; Tribo AR; Pasternak GW; Carroll FI; Stevens RC; Cherezov V; Katritch V; Wacker D; Roth BL Structure of the Nanobody-Stabilized Active State of the Kappa Opioid Receptor. *Cell* 2018, 172, 67e15.
 37. Hollopeter G; Jantzen HM; Vincent D; Li G; England L; Ramakrishnan V; Yang RB; Nurden P; Nurden A; Julius D; Conley PB Identification of the platelet ADP receptor targeted by antithrombotic drugs. *Nature* 2001, 409, 202–207. [PubMed: 11196645]
 38. Zhang J; Zhang K; Gao Z; Paoletta S; Zhang D; Han GW; Li T; Ma L; Zhang W; Müller CE; Yang H; Jiang H; Cherezov V; Katritch V; Jacobson KA; Stevens RC; Wu B; Zhao Q Agonist-bound structure of the human P2Y₁₂ receptor. *Nature* 2014, 509, 119–122. [PubMed: 24784220]
 39. Zhang K; Zhang J; Gao Z; Zhang D; Zhu L; Han GW; Moss SM; Paoletta S; Kiselev E; Lu W; Fenalti G; Zhang W; Müller CE; Yang H; Jiang H; Cherezov V; Katritch V; Jacobson KA; Stevens RC; Wu B; Zhao Q Structure of the human P2Y₁₂ receptor in complex with an antithrombotic drug. *Nature* 2014, 509, 115–118. [PubMed: 24670650]

40. Zhang J; Zhang K; Gao Z; Paoletta S; Zhang D; Han GW; Li T; Ma L; Zhang W; Müller CE; Yang H; Jiang H; Cherezov V; Katritch V; Jacobson KA; Stevens RC; Wu B; Zhao Q Agonist-bound structure of the human P2Y₁₂ receptor. *Nature* 2014, 509, 119–122. [PubMed: 24784220]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Highlights:

- Local and global similarity metrics suggest different GPCR modeling templates.
- Comparative GPCR models constructed from local templates yield more accurate ligand docking results.
- Refinement strategies will still be necessary components of comparative modeling protocols.

Figure 1.

Superposition of 50 representative GPCR structures. Fusion partners were deleted prior to superposition. Backbone structures in the TM region show strong structural conservation. TM segments 1 and 4 are labeled, TM segments 5–7 are behind TM segments 1–4 in this view. Specific GPCR structures used are shown in Table S1.

Figure 2:

Names and structures of ligands present in reference crystal structures and cross-docking target crystal structures. Ligands present in reference crystal structures were used in docking studies and are shown in the ionization state used in docking studies. The first line of text under each ligand indicates the ligand abbreviation used in the PDB, followed by the receptor name and the four-character PDB entry code.

Figure 3:

Panel A) Reference crystal structure of P2Y12 (red ribbons) with the 53 unique of 61 residues for the PRMSD61 shown in grey and the 8 residues common to the PRMSD61 and PRMSD8 shown in teal. Panel B) Superposition of P2Y12 reference (red ribbons and alternate crystal structures (teal and pink ribbons for 4NTJ and 4PXZ, respectively), local and global template models from structure-independent alignment (grey and blue ribbons, respectively), local and global template model from structure-dependent alignment (yellow and orange ribbons, respectively).

Figure 4:

Superposition of docked and crystallographic poses for three examples: OPRK (Panels A and D), FFAR1 (Panels B and E), and CXCR4 (Panels C and F). Reference crystal structures are shown in red ribbons. Induced fit causes slight sidechain movements. Yellow ribbons resulted from induced fit docking. A-C) View perpendicular to helical axes. TM 6 and TM 7 are hidden to allow viewing into the binding pocket. D-F) View from extracellular space.

Figure 5.

Superposition of docked and crystallographic poses for JDC in OPRK. OPRK reference (4DJH) and cross-docking target crystal structures (6B73) are shown with red and teal ribbons. OPRK local and global template homology models from structure-independent alignments are shown with grey and blue ribbons. OPRK local and global template homology models from structure-based alignments are shown with yellow and orange ribbons. Ligand colors are matched to the ribbon color of the docking target. TM 6 and 7 are

hidden to allow view into binding pocket. A,B,C, E, F: View perpendicular to helical axes.
D: View from extracellular space.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1:

61 residues used for the PRMSD61 calculations. Eight residues used for the PRMSD8 calculations shown in **bold**.

TM 1	TM 2	TM 3	TM 4	TM 5	TM 6	TM 7
1×28	2×53	3×21	4×56	5×33	6×44	7×24
1×35	2×56	3×26	4×57	5×35	6×48	7×27
1×39	2×57	3×28	4×58	5×36	6×51	7×30
	2×59	3×29	4×59	5×37	6×52	7×31
	2×60	3×30	4×60	5×39	6×54	7×33
	2×61	3×32	4×61	5×40	6×55	7×34
	2×63	3×33		5×43	6×58	7×35
	2×64	3×34		5×44	6×59	7×37
		3×36		5×46	6×61	7×38
		3×37		*5×461	6×62	7×39
		3×40		5×47		7×41
						7×42

* A closely related numbering system to the Ballesteros-Weinstein system is used by GPCRdb.org, utilizing 'x' in place of the period. This system also recognizes that a subset of GPCR have single amino acid insertions with certain TM relative to the majority of GPCR. In these cases, a '1' is appended to the number of the preceding amino acid number. One specific example is the 5.461 residue that a subset of GPCR contain that falls between the more conserved 5.46 and 5.47 residues.

Table 2:

Selected Template Structures for 6 crystallized GPCR based on both local and global similarity measures used in this study due to their substantial differences in sequence identity and/or local similarity measure as compared to the other crystallized Class A GPCR.

Receptor	Local Template	Local Similarity Measure	GPCRdb % Identity	Global Template	Local Similarity Measure	GPCRdb % Identity	Local Similarity Measure	% Identity
CXCR4	AT ₂ R	1.72	22	AT ₁ R	1.28	29	0.44	7
FFAR₁	P2Y ₁₂	1.42	12	AT ₁ R	-0.40	16	1.82	4
NOP	M ₄ R	1.22	14	AT ₁ R	0.03	23	1.20	9
P2Y₁₂	PAR1	1.78	13	5-HT _{1B}	0.01	19	1.76	6
M₁	H ₁ R	2.58	22	D ₃	1.14	22	1.44	0
OPRK	H ₁ R	1.93	14	AT ₁ R	0.09	25	1.84	11

Table 3:

GenBank accession numbers and PDB ID numbers for GPCRs used in this study.

	GenBank Accession Number	PDB ID	
5HT1B	N/A	4IAQ ¹⁸	
H1R	N/A	3RZE ¹⁹	
AT1R	N/A	4YAY ²⁰	
AT2R	N/A	5UNH ²¹	
CXCR4	CAA12166.1 ²²	3OE6 ²³	*4RWS ²⁴
FFAR1	AAI20945.1 ²⁵	4PHU ²⁶	*5KW2 ²⁷
M1	CAA68560.1 ²⁸	5CXV ²⁹	
D3	N/A	3PBL ³⁰	
M4R	N/A	5DSG ²⁹	
NOP	NP_872588.1 ³¹	4EA3 ³²	*5DHH ³³
PAR1	N/A	3VW7 ³⁴	
OPRK	AAC50158.1	4DJH ³⁵	*6B73 ³⁶
P2Y12	Q9H244.1 ³⁷	4PY0 ³⁸	*4NTJ ³⁹ *4PXZ ⁴⁰

N/A: Not Applicable - Models for these receptors were not constructed, so native sequences were not needed.

* Crystal structures used for the cross docking experiments.

Table 4:

Global and Binding Pocket Weighted Alpha Carbon and All Atom RMSDs (Å) calculated from the compared structure against the reference crystal. Lowest RMSD for each set of models in each column are shown in **bold**.

Reference Structure	Compared Structure	Global RMSD (Å)		PRMSD ₆₁ (Å)		PRMSD ₈ (Å)	
		Alpha Carbon	All Atom	Alpha Carbon	All Atom	Alpha Carbon	All Atom
CXCR4 (3OE6)	PDB entry 4RWS	1.33	2.11	1.19	1.70	0.53	0.89
	Local Template ^a	4.33	5.11	1.75	2.36	1.16	1.61
	Global Template ^a	3.48	4.28	1.98	2.57	1.37	2.22
FFAR1 (4PHU)	PDB entry 5KW2	2.13	2.73	1.68	2.10	1.04	1.47
	Local Template ⁱ	5.78	6.53	3.11	3.71	1.22	1.92
	Global Template ⁱ	6.82	7.35	3.39	4.21	1.54	2.55
	Local Template ^s	5.66	6.85	2.90	3.89	1.31	1.68
	Global Template ^s	5.16	5.93	3.07	3.85	1.43	1.88
NOP (4EA3)	PDB entry 5DHH	0.70	1.17	0.31	0.58	0.21	0.64
	Local Template ⁱ	4.17	4.76	3.22	4.04	1.51	2.24
	Global Template ⁱ	3.45	3.92	1.78	2.25	1.23	1.84
	Local Template ^s	3.04	3.62	2.93	3.32	1.38	2.28
	Global Template ^s	3.12	3.87	2.14	2.70	1.01	1.69
P2Y12 (4PY0)	PDB entry 4NTJ	2.65	3.29	2.51	3.34	1.07	1.50
	PDB entry 4PXZ	0.63	1.06	0.33	0.54	3.15	2.83
	Local Template ⁱ	3.75	4.52	3.08	4.29	1.49	1.94
	Global Template ⁱ	5.48	6.09	3.53	4.94	1.00	1.60
	Local Template ^s	3.97	4.74	3.48	4.45	1.48	1.61
	Global Template ^s	5.33	6.05	4.80	6.17	1.06	2.07
M1 (5CXV)	Local Template ⁱ	3.01	3.78	1.81	2.72	0.90	1.34
	Global Template ⁱ	3.68	4.37	2.36	3.23	1.65	2.00
	Local Template ^s	3.19	3.95	1.54	2.47	1.16	1.61
	Global Template ^s	2.78	3.64	1.99	2.58	1.45	2.33
OPRK (4DJH)	PDB entry 6B73	3.32	3.79	1.53	2.01	0.92	1.28
	Local Template ⁱ	4.59	4.98	2.62	3.60	1.18	1.70
	Global Template ⁱ	4.14	4.46	2.44	2.96	1.10	1.64
	Local Template ^s	4.73	5.04	4.72	5.05	1.33	1.75

Reference Structure	Compared Structure	Global RMSD (Å)		PRMSD ₆₁ (Å)		PRMSD ₈ (Å)	
		Alpha Carbon	All Atom	Alpha Carbon	All Atom	Alpha Carbon	All Atom
	Global Template ^s	3.86	4.40	3.86	4.41	1.21	1.73

^a: model generated from indicated template using structure-independent alignment which didn't differ from the structure-based alignment

ⁱ: model generated from indicated template using structure-independent alignment.

^s: model generated from indicated template using structure-based alignment from GPCRdb.

Table 5:

Crystallographic pose scoring in reference crystal structure, cross-docking target crystal structures, and comparative model. Ligand structures and names are provided in Figure 2. Scores for ligand poses within 0.5 kcal of the reference are shown in bold. Positive scores are highlighted with grey backgrounds.

Receptor	Ligand	Crystallographic Pose Scores (kcal/mol)					
		Reference	Alternate	Global ⁱ	Local ⁱ	Global ^s	Local ^s
CXCR4	ITD	-6.1	-5.3	-6.2	-5.3	N/A	N/A
FFAR1	MK6	-9.9	4.8	18.0	2.0	121.5	86.7
NOP	0NN	-9.0	-8.8	-7.4	-6.3	-7.6	1.2
P2Y12 [*]	6AT	-12.7	-7.9/ -12.3	-1.0	-7.4	2.1	-6.7
M1	0HK	-9.1	N/A	-6.0	-7.0	-5.4	-5.2
OPRK	JDC	-10.6	-4.8	-7.8	-8.7	-7.9	-8.0

^{*} 6AT pose scores in alternate crystal structures listed first for 4NTJ and then for 4PXZ

Table 6:

RMSD between crystallographic and rigid or induced fit ligand pose in the reference crystal structure. Ligand structures and names are provided in Figure 2. Lowest RMSD for each row are highlighted in **bold**. Rank-ordered pose number (#) selected for analysis indicated.

Receptor	PDB Entry	Ligand	Rigid Pose RMSD (Å)/#	Rigid Pose Tanimoto Coefficient	Induced Fit Pose RMSD (Å)/#	Induced Fit Pose Tanimoto Coefficient ^a
CXCR4	3OE6	ITD	8.87/3	0.66	7.71/2	0.88
FFAR1	4PHU	MK6	1.70/1	0.83	2.42/3	0.83
NOP	4EA3	0NN	1.20/3	0.88	1.25/1	0.91
P2Y12	4PY0	6AT	1.06/1	0.94	0.95/1	0.94
M1	5CXV	0HK	1.69/1	0.89	2.98/3	0.93
OPRK	4DJH	JDC	1.12/1	0.88	0.65/1	0.89

^a: Tanimoto coefficient indicates the similarity between first-neighbors to the ligand in the residue interaction network calculated for the crystallographic and docked complexes using the RING 2.0 server.

Table 7:

Comparison of docked and crystallographic reference poses. Ligand structures and names are in Figure 2. Lowest RMSD and highest Tanimoto coefficient among models for a given receptor shown in **bold**. Rank-ordered pose number (#) selected for analysis indicated.

Receptor	Docked Ligand	Docking Target (original ligand)	RMSD (Å)/#	Tanimoto Coefficient ^b
CXCR4	ITD	PDB entry 4RWS (VMIP-II)	8.77/2	0.65
	ITD	Local Template ^a	8.20/5	0.17
	ITD	Global Template ^a	11.70/2	0.10
FFAR1	MK6	PDB entry 5KW2 (6XQ)	13.47/3	0.49
	MK6	Local Template ⁱ	13.81/1	0.35
	MK6	Global Template ⁱ	19.13/4	0.31
	MK6	Local Template ^s	15.38/2	0.41
	MK6	Global Template ^s	16.12/2	0.49
NOP	0NN	PDB entry 5DHH (DGW)	3.70/2	0.74
	0NN	Local Template ⁱ	11.08/1	0.37
	0NN	Global Template ⁱ	13.72/1	0.32
	0NN	Local Template ^s	8.97/3	0.61
	0NN	Global Template ^s	13.71/3	0.22
P2Y12	6AT	PDB entry 4NTJ (AZJ)	8.09/1	0.46
	6AT	PDB entry 4PXZ (6AD)	2.18/2	0.91
	6AT	Local Template ⁱ	10.73/1	0.32
	6AT	Global Template ⁱ	8.47/3	0.31
	6AT	Local Template ^s	10.23/5	0.32
	6AT	Global Template ^s	7.32/1	0.29
M1 [*]	0HK	Local Template ⁱ	13.78/5	0.16
	0HK	Global Template ⁱ	16.11/3	0.12
	0HK	Local Template ^s	6.11/1	0.60
	0HK	Global Template ^s	18.18/2	0.10
OPRK	JDC	PDB entry 6B73 (CVV)	9.29/2	0.52
	JDC	Local Template ⁱ	5.05/1	0.51
	JDC	Global Template ⁱ	10.68/1	0.58
	JDC	Local Template ^s	7.11/1	0.52
	JDC	Global Template ^s	16.11/1	0.07

* No available crystal structure for cross-docking experiment.

^a: model generated from indicated template using structure-independent alignment which didn't differ from the structure-based alignment

^b: Tanimoto coefficient indicates the similarity between first-neighbors to the ligand in the residue interaction network calculated using the RING 2.0 server.

ⁱ: model generated from indicated template using structure-independent alignment.

^s: model generated from indicated template using structure-based alignment from GPCRdb.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript