PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Critical analysis of the successes and failures of homology models of G protein-coupled receptors

Supriyo Bhattacharya, Alfonso Ramon Lam, Hubert Li, Gouthaman Balaraman, Michiel Jacobus Maria Niesen, and Nagarajan Vaidehi*

Division of Immunology, Beckman Research Institute of the City of Hope, Duarte, California 91010

ABSTRACT

We present a critical assessment of the performance of our homology model refinement method for G protein-coupled receptors (GPCRs), called LITICon that led to top ranking structures in a recent structure prediction assessment GPCRDOCK2010. GPCRs form the largest class of drug targets for which only a few crystal structures are currently available. Therefore, accurate homology models are essential for drug design in these receptors. We submitted five models each for human chemokine CXCR4 (bound to small molecule IT1t and peptide CVX15) and dopamine D3DR (bound to small molecule eticlopride) before the crystal structures were published. Our models in both CXCR4/IT1t and D3/eticlopride assessments were ranked first and second, respectively, by ligand RMSD to the crystal structures. For both receptors, we developed two types of protein models: homology models based on known GPCR crystal structures, and ab initio models based on the prediction method MembStruk. The homology-based models compared better to the crystal structures than the ab initio models. However, a robust refinement procedure for obtaining high accuracy structures is needed. We demonstrate that optimization of the helical tilt, rotation, and translation is vital for GPCR homology model refinement. As a proof of concept, our in-house refinement program LITiCon captured the distinct orientation of TM2 in CXCR4, which differs from that of adrenoreceptors. These findings would be critical for refining GPCR homology models in future.

Proteins 2013; 81:729–739.

Key words: G protein-coupled receptor; structure prediction; homology model refinement; GPCRDOCK2010, CXCR4; dopamine receptor; rigid body optimization.

INTRODUCTION

G protein-coupled receptors (GPCRs) are membrane bound proteins that transduce signal from outside to inside the cell thus leading to a variety of biological processes. They form the largest super family of drug targets and therefore their three-dimensional structural and dynamic information are of great interest to researchers and the pharmaceutical industry to design effective drugs targeting GPCRs. In the past 5 years, there have been several breakthroughs in solving GPCR structures both in the form of inactive and active conformations of the receptors.[1–6] However, the reported crystal structures represent a small fraction of the total number of GPCR species in the mammalian genome. As the number of available crystal structures increases, homology or comparative modeling methods can be used to derive structural models of other GPCRs. However, the accuracy of homology models needs to be improved to the level of accuracy of high-resolution crystal structures before using the models for drug design. There is a dire need for computational methods that can "refine" the homology models to capture the nuances of high-resolution structures. In this context, prior to the publication of the crystal structures of the human chemokine receptor CXCR4 and dopamine receptor D3DR, a structure prediction assessment named GPCRDOCK2010 was announced.[7] The objective was to assess the status of current computational methods for predicting the structures of GPCRs and especially the ligand binding sites. As part of this assessment we submitted 15 receptor models: five CXCR4

Additional Supporting Information may be found in the online version of this article.

Supriyo Bhattacharya and Alfonso Ramon Lam contributed equally to this work.

*Correspondence to: Nagarajan Vaidehi, Division of Immunology, Beckman Research Institute of the City of Hope, 1500 Duarte Road, Duarte, CA 91010.
E-mail: NVaidehi@coh.org
Received 24 May 2012; Revised 17 September 2012; Accepted 21 September 2012
Published online 8 October 2012 in Wiley Online Library (wileyonlinelibrary.com).
DOI: 10.1002/prot.24195

models with a small molecule antagonist IT1t bound, five CXCR4 models with a peptide CVX15 bound, and five models of the dopamine D3DR with the antagonist eticlopride bound. For small molecule bound D3DR and CXCR4, our predicted top ranking receptor–ligand complexes showed close agreement with the crystal structures in terms of ligand binding pocket and ligand pose prediction. Although the predicted binding pose for the peptide CVX15 was ranked second among the contestants, its agreement with the crystal structure was substantially less compared to the small molecule.

Three out of five top ranking receptor models submitted for CXCR4 and D3DR were modeled using the homology modeling package MODELLER.[8] The homology models were refined subsequently using our in-house homology model refinement method LITiCon.[1,9,10] While D3DR dopamine receptor is close in sequence identity to the β_1- and β_2-adrenergic receptors (denoted as β_1-AR and β_2-AR hereafter), CXCR4 did not have a closely related homologous template that can be used for homology modeling.

In this article, we demonstrate that optimization of the helical tilt, rotation, translation, and gyration of the seven transmembrane (TM) helices lead to substantial refinement of the GPCR homology models even with remote homology templates. We present a critical analysis of our predicted receptor models and the binding site of the antagonists as well as provide insights into how the modeling can be done for GPCRs that have little homology to known structures. During this work, we also developed a more detailed procedure "GPCRCompare" for comparing two GPCR structures. This procedure compares the relative TM helical tilt, rotation, gyration, and translation of all the seven TM helices. The procedures described in this article will be very useful for drug design in academia and pharmaceutical industry.

METHODS

GPCR structural comparison

To compare the helical orientations of two GPCR structures, the two structures are aligned using the Cα coordinates of the TM regions. As the loops are more structurally divergent than the TMs, the loops are not included in the alignment. The relative orientation of each TM helix with respect to another is composed of three independent components: translation of the helix, tilt about the center of mass, and rotation about the helical axis. We calculate these individual components for each TM in one receptor structure relative to the corresponding TM in the other one (Supporting Information Fig. S5).

CXCR4 structural models

Of the top five structural models of CXCR4 that we submitted to GPCRDOCK 2010 assessment, three were homology-based models and two were derived using an ab initio method called MembStruk.[11] The details of the MembStruk method are discussed elsewhere.[11] In brief, the MembStruk method starts with a template of the 7 TM domains based on β_2-AR crystal structure. It then optimizes the packing of the TM domains by refining the rotational orientation of each TM helix over $\pm 180^\circ$. The top scoring energy minima for each TM helix are combined to identify the best scoring combination of rotation angles.

The three homology-based models were derived using MODELLER[8] and with three different sets of templates. One of the models was built using both A_{2A} adenosine (PDB ID: 3EML) and β_2 AR (PDB ID: 2RH1) as templates,[4,2] and the remaining two models were built using β_1-AR (PDB ID: 2VT4) and β_2-AR as templates, respectively.[2,5] In the β_1-AR-based model, TM2 and TM4 were optimized using the LITiCon method. TM2 was rotated in increments of 5° from -40° to 40°. At each rotated conformation, the side chains were optimized using SCWRL3.0,[12] and the total energy of the receptor was minimized using conjugate gradient minimization. The Van der Waals (vdW) energy showed a minimum at a rotation angle of -15° as shown in Figure 5(a). Hence we selected this conformation for TM2. A similar procedure was followed for TM4 as well. For TM4, the vdW energy was flat near the initial structure, while the potential energy showed minima at -10° and 35° as shown in Figure 5(c). We selected the TM4 conformation corresponding to a 35° rotation, which was the global minimum in energy for TM4 rotations. We docked the IT1t ligand to all the homology models with different templates and selected the β_1-AR-based homology model for submission, because it gave the best ligand docked poses satisfying the criteria used for selecting the best ligand pose as detailed in the ligand docking section below. The best structure was then minimized using the minimizer in NAMD and CHARMM force field.[13] The major difference between the MembStruk and LITiCon procedure for rotational optimization is that LITiCon performs simultaneous optimization of rotational orientations of all the seven helices, while the MembStruk method performs this optimization serially for each helix.

D3DR structural models

Of the five D3DR structural models, the first three models (by ranking) were derived through homology methods using β_2-AR as template.[2] Since the extracellular loop 2 (ECL2) loop of D3DR shows little homology to β_2-AR, this loop was modeled using the ab initio loop modeling procedure in MODELLER[8] without using any structural template and an imposed disulfide constraint between C103 and C181 (ECL2). Each of the structures

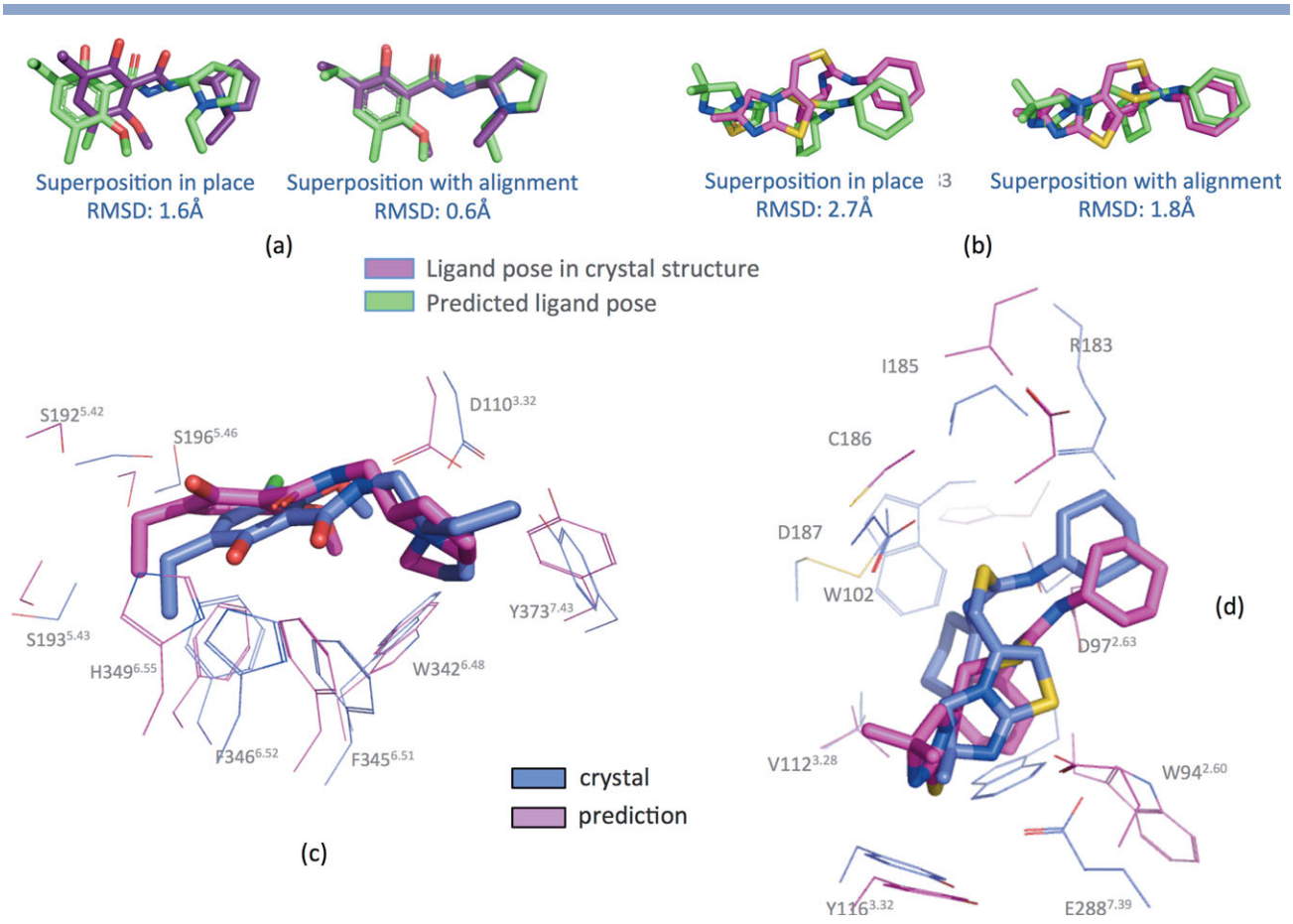

Figure 1

Comparison between the predicted and crystal structure ligand poses superposed both in place (individual receptor structures aligned) and structurally aligned; (**a**) eticlopride; (**b**) IT1t. (**c**) and (**d**): Comparison between the predicted and crystal structure binding sites of (c) eticlopride in D3DR; (d) IT1t in CXCR4; Residues within 5 Å of the ligands are displayed. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

was minimized using the DREIDING force field,[14] before subjecting to the docking protocol. The LITiCon rigid body optimization was not performed in the case of D3DR because of the high homology with β_2-AR.

Docking of small molecule ligands

The small molecules IT1t for CXCR4 and eticlopride for D3DR were built and optimized using LigPrep module in Maestro. Several ligand conformations were generated outside the protein and docked using Glide XP (Schrödinger). Ten docked conformations were retained for each ligand conformation. The docked conformations were clustered using Xcluster (Schrödinger) with a cluster distance cut-off of 1.5 A. The criteria to select the poses were based on the mutation data available for cyclam and noncyclam compounds in CXCR4.[15–18] For selecting the eticlopride docked poses, the mutation data on D2DR[19] (eticlopride binding not affected by any of the serine mutations on TM5) were considered.

RESULTS AND DISCUSSION

Comparison of the predicted binding site of the small molecule IT1t in CXCR4 and eticlopride in D3DR to their crystal structures

Figure 1(a,b) show the root mean square deviation in coordinates referred to as CRMSD hereafter of the predicted ligand poses (eticlopride and IT1t) to those in the crystal structures. The ligand CRMSDs were calculated by aligning the predicted receptor models to the corresponding crystal structures. This gives the ligand CRMSD "in place," which indicates how close the predicted ligand pose is to the crystal structure pose. We also calculated the ligand CRMSDs by aligning the predicted ligand structure to the ligand pose from the crystal structure ("superposition with alignment"). This gives the accuracy of the internal structures of the predicted ligand poses. Both IT1t in CXCR4 and eticlopride in D3DR show close agreement with the placement of the ligand in the receptor (CRMSDs

Table I
Percentage of Receptor–Ligand Interactions Captured in Our Predicted Binding Poses

	Receptor	CXCR4	D3DR
Residues within 3 Å	Number of residues in crystal structure	2	1
	% captured in homology model	50%	100%
Residues within 4 Å	Number of residues in crystal structure	12	14
	% captured in homology model	42%	86%
Residues within 5 Å	Number of residues in crystal structure	16	20
	% captured in homology model	50%	90%
Residues within 6 Å	Number of residues in crystal structure	22	27
	% captured in homology model	59%	85%

for ligand heavy atoms: 1.6 Å for eticlopride and 2.7 Å for IT1t). For all the comparison to the crystal structures, we used the PDB ID: 3ODU for CXCR4 and 3PBL for D3DR. The predicted ligand conformations show internal CRMSD of 0.6 Å for eticlopride and 1.8 Å for IT1t, as shown by alignment of the ligands to the crystal poses. Figure 1(c,d) show the orientations of the ligands within their respective binding cavities as compared to their crystal structures. Our predicted ligand binding pose captures 50% and 90% of all the protein ligand contacts within 5 Å of the ligand in the CXCR4/IT1t and D3DR/eticlopride crystal structures respectively. Although, we missed about 50% of the binding site residues in case of CXCR4/IT1t, many of these missed residues are in the EC loops. Among the residues, in the TM regions, our prediction rate was much higher. We predicted 70% of the TM residues that are within 5 Å of IT1t. Many of the loop residues were not identified due to the lack of accuracy in our loop predictions. However, most of the key residues that govern the binding of small molecule antagonists (such as IT1t) in CXCR4 (as observed from mutation studies) are located in the TM domains. This in conjunction with the challenges associated with the CXCR4 target indicate that the level of accuracy in the predicted models is well suited for predicting site directed mutagenesis candidates as well as for studying the structure–activity relationships. The number of residues was calculated using PyMol,[20] unlike in the GPCRDOCK2010 assessment where a soft cut-off was used.[7]

Detailed predictions of the binding site residues

Table I shows the number of residues within a certain cut-off distance from the ligand in CXCR4 and D3DR predicted structural models. It is clear that about 50% of the protein ligands are captured in the CXCR4/IT1t structure and about 90% of the contacts are captured in D3DR/eticlopride. There are two possible protonatable nitrogens in the compound IT1t. Our model captures the tight contact of one of these amine groups with the $D97^{2.63}$ (2.7 Å) and the weaker contact with $E288^{7.39}$ as seen in the crystal structure. In the predicted ligand pose, the imidothiazole group is positioned in the same location within the binding site as the crystal structure but rotated leading to a larger CRMSD for this part of the ligand.

To evaluate the accuracy of predicting the binding sites, we calculated the CRMSDs of the amino acids within 5 Å of the ligand in the predicted receptor structures to those in the crystal structures. The CRMSDs of the predicted models of CXCR4 and D3DR are shown in Table II. In the CXCR4 predicted model, with the exception of $W94^{2.60}$, all the residues in the TM helices show low CRMSDs to the crystal structure (<3 Å), while the loop residues show larger deviations. From Figure 1(d), the backbone of $W94^{2.60}$ is situated close to that of the crystal structure. However, the side chain of $W94^{2.60}$ shows a large deviation from that of the crystal structure, which contributes to the large CRMSD of this residue. In D3DR, all binding site residues are part of TM domains. All of these residues show low CRMSDs to the crystal structure thus highlighting the accuracy of our prediction.

Pitfalls of docking small ligands to GPCRs

Clearly between the two receptor targets, prediction of ligand binding to CXCR4 was more challenging, as there were no prior site directed mutagenesis studies on the IT1t or related ligands. The binding pocket of CXCR4 consists of four charged residues ($D97^{2.63}$, $D171^{4.60}$, $D262^{6.58}$, and $E288^{7.39}$), all of which show strong reduction of antagonist binding upon mutation to neutral residues.[16,18] Docking of IT1t showed two possible energetically favorable binding sites: one located in the region between $D171^{4.60}$, $D262^{6.58}$, and $E288^{7.39}$ and the other located among the residues $D97^{2.63}$, $D262^{6.58}$, and $E288^{7.39}$. The literature on the effect of site directed mutagenesis on binding of cyclam-based compounds leads to ambiguous positioning of the IT1t ligand.[16–18] Thus determining the principal salt bridge for IT1t using preexisting knowledge from other small molecules is nontrivial. Based on our previous knowledge of antagonist binding in chemokine receptors,[21–24] we hypothesized that the nitrogen atom on the imidothiazole group (one on the same side as the sulfur atom) could form a hydro-

Table II
CRMSD of Residues in the Binding Site for CXCR4/IT1t and D3DR/Eticlopride

CXCR4/residue	CRMSD (Å)	D3DR/residue	CRMSD (Å)
$W94^{2.60}$	5.5	$D110^{3.32}$	1.2
$D97^{2.63}$	2.9	$S192^{5.42}$	2.2
$V112^{3.28}$	2.5	$S193^{5.43}$	1.3
$Y116^{3.32}$	1.6	$S196^{5.46}$	1.7
$E288^{7.39}$	2.3	$W342^{6.48}$	1.3
$W102^{ECL1}$	6.3	$F345^{6.51}$	1.1
$R183^{ECL2}$	10.9	$F346^{6.52}$	1.1
$I185^{ECL2}$	4.1	$H349^{6.55}$	2.9
$C186^{ECL2}$	5.6	$Y373^{7.43}$	1.1
$D187^{ECL2}$	8.2		

gen bond with D97$^{3.63}$ and E288$^{7.39}$. As D97$^{2.63}$ was important in noncyclam antagonist binding in CXCR4[18], and E288$^{7.39}$ was important for antagonist binding in many chemokine receptors,[21] we selected those ligand poses which satisfied the criteria of contacting D97$^{2.63}$ and E288$^{7.39}$ and obtained 48 ligand poses. We selected the final ligand pose by visualizing each of the 48 poses and using criteria such as how well the ligand fits into the binding cavity, and whether there is any unusual internal strain in the ligand pose. In contrast, predicting the eticlopride pose in D3DR involved clustering the docked poses into distinct clusters based on CRMSD and then selecting the top pose by binding energy from each cluster. However, the final ligand poses were selected by visualization in conjunction with the binding energy.

Based on our docking predictions, we found that while docking methods were efficient in sampling accurate ligand poses (one close to the crystal structure poses), binding energy of the docked poses were not accurate enough to score the best structure at the top. The inaccuracies in the binding energy calculation could be contributed by the structural defects of the homology models as well as our approximate method of estimating energy, which involves limited minimization and crude approximation of desolvation. At the end, experimental data from binding site mutations were useful in identifying the correct ligand pose.

Analysis of receptor models and crystal structures

The CRMSD in coordinates of the prescribed set of atoms when comparing two GPCR structures gives an averaged quantitative comparison while missing the differences in the structural features such as the seven helical rotations, tilt, translation, and gyration that are important for the function of the GPCR. Therefore, we derived a geometrical method, "GPCRCompare" to compare two GPCR structures in terms of these helical properties. Each of the seven helical TM regions will be compared in terms of the helical rotation, tilt, and translation. We have analyzed the merits and pitfalls of our submitted CXCR4 and D3DR models by comparing them to the crystal structures. We have also compared the crystal structures of D3DR and CXCR4 to the crystal structure of β_2-AR (PDB ID: 2RH1) in terms of the seven helical properties. More importantly, we have demonstrated how optimization of these helical translation rotation, tilt, and gyration leads to refinement of the homology models for two GPCRs with low sequence homology.

TM helical rotation and tilts capture most of the differences between GPCR crystal structures

Analysis of the structural differences in the seven TM helices for the CXCR4 and D3DR crystal structures with respect to β_2-AR in the inactive state are detailed in the next section. Summarizing these results, it is seen that the maximum difference in structures of CXCR4 and β_2-AR stems from helical rotation and tilt of TM helices 5 and 6. To demonstrate this point clearly, we took the TM5 and TM6 of CXCR4 crystal structure and performed the helical translation, rotation, and tilt differences between CXCR4 and β_2-AR crystal structures in TM5 and TM6 and got structures that are close to β_2-AR. Figure 2 demonstrates the result of these operations on the receptor models. It is seen that after successive rigid body transformations, TM5 and TM6 of CXCR4 show substantial improvement resulting in less than 0.5 Å CRMSD to β_2-AR at the end of these helical movements. Thus we find that the helical tilts and rotations contribute substantially to the structural diversities of TM5 and TM6 and therefore optimization of helical rotations, tilt, and translation would lead to substantial refinement in GPCR homology models.

Comparison of all the predicted models to crystal structures and pitfalls of the structure prediction methods

For the GPCRDOCK2010 challenge, we used both homology modeling technique[8] and the ab initio GPCR structure prediction method MembStruk[11] to generate the receptor structural models. Among the five predicted models for CXCR4, models 1, 2, and 5 (by our ranking) were homology models, and structures 3 and 4 were predicted using MembStruk. Among D3DR models, structures 1–3 were homology models, and structures 4 and 5 were using MembStruk. The CXCR4 homology model was further subjected to rigid body optimization of TM2 and TM4 using the software LITiCon.[1] No LITiCon optimization was performed with the D3DR models because D3DR has a sufficiently high homology with the template β_2-AR. We have compared all the receptor structural models to their respective crystal structures. Supporting Information Table S2 summarizes the major differences in the orientations of TM helices of the CXCR4 and D3DR models to the respective crystal structures. Both the homology and MembStruk CXCR4 models show large RMS deviations from the crystal structures in the TM regions. However, the MembStruk model shows a larger CRMSD (6 Å) compared to the homology model (4.9 Å). Comparing the CRMSDs of individual TM domains of the homology model, TM1 shows the lowest deviation with the crystal structure (3.4 Å), while TM5 and TM6 show the worst deviations (6 Å). The RMS deviation of TM5 is due to a misalignment of the sequence in the intracellular region, and most of the CRMSD in TM6 in the homology model comes from the intracellular domain. TMs 1, 4, 5, and 7 in the CXCR4 homology model show large CRMSDs when the TMs are individually aligned to the corresponding TMs in the crystal structure. For these TMs, the differences with the

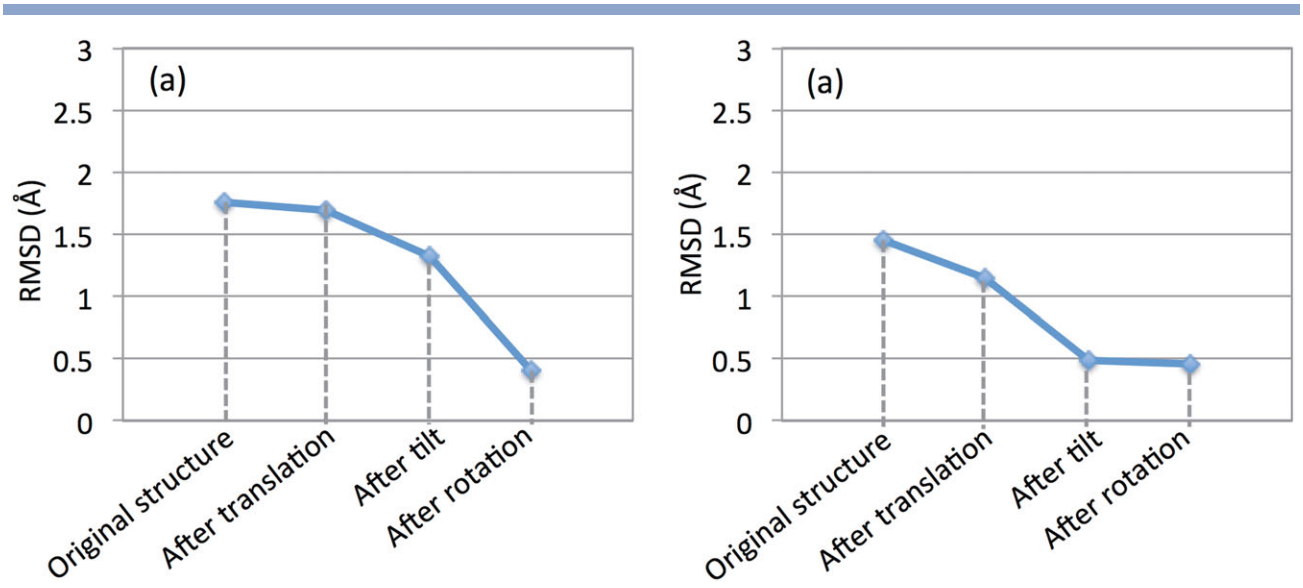

Figure 2

Improvement in CRMSD of TM helices 5 and 6 of CXCR4, to those of β_2-AR after successive rigid body transformations (**a**) TM5; (**b**) TM6. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

crystal structure are contributed by both packing and internal deformations in the helices. In contrast, the structural deviations of TMs 2, 3, and 6 are mostly due to packing differences, as their internal structures are close to those in the crystal structures.

In the MembStruk model, all TM regions show larger RMS deviation with the crystal structure compared to the homology model, with the largest deviations being observed in TM2 and TM6. Both these TMs show large internal deformations (high "aligned" CRMSDs in Supporting Information Table S2). Hence their structural deviations are contributed by helical kinks as well as packing differences. The optimization of the helical degrees of freedom in the MembStruk models did not yield rotation angles close to the crystal structure, due to the inadequacy of the energy functions used in structure prediction or refinement to identify the best energy structure as the one close to the crystal structure. This has been observed previously for globular proteins (Critical Assessment of techniques for Structure Prediction http://predictioncenter.org/) and hence more comprehensive energy functions that take into account membrane solvation are required. Moreover, the internal structures of TMs 1, 2, 4, 6, and 7 are different than the ones in the crystal structure (as reflected by their high "aligned" CRMSDs), which also contribute to the structural differences with the crystal conformation.

Among the D3DR models, both homology and MembStruk models are closer in CRMSD to the D3DR crystal structure (1.5 and 2.7 Å, respectively) in comparison to the CXCR4 models. In the homology model, TM4 shows a 16° rotation relative to the crystal structure. In the MembStruk model, the largest deviations are shown by TM1, TM2, and TM4 (3–3.2 Å). The CRMSD in TM1 is contributed by a 2.3 Å translation and a 45° rotation. TM2 shows high internal deformation, while TM4 shows a 2.5 Å translation. For the remainder of this section, we will discuss the homology models of the two receptors. Thus we observe that the homology model with subsequent rigid body optimization would produce better results than ab initio methods.

Figure 3 shows the CRMSD for each residue of CXCR4 and D3DR from their respective crystal structures. In both models, the TM regions show much closer agreement with the crystal structures compared to the loop regions. In the CXCR4 model, the extracellular loop 2 (ECL2) proximal to TM4 shows higher CRMSD (~15 Å), while the part near TM5 shows a relatively lower CRMSD. Overall, our loop predictions place the EC loops covering the entire extracellular surface of the TMs. In the crystal structure, the ECL2 is located closer to TMs 2 and 3 while ECL3 is located outward between TMs 6 and 7. This opens up a cavity in the EC surface of the crystal structure with direct access to the ligand binding pocket.

Comparison of the location of the helical kinks and the kink angles in each TM calculated using the software PROKINK[25] are shown in Supporting Information Table S1. The kink angles are compared between the models and the corresponding crystal structures. Each of the TM regions in the CXCR4 crystal structure shows moderate to large kinks. TM7 shows the largest kink (45°) and TM3 shows the smallest kink (19°). The CXCR4 homology model shows good overall agreement in terms of kink prediction with the exception of TM1. TM1 in the crystal structure shows a 24° kink near G52$^{1.46}$. This kink

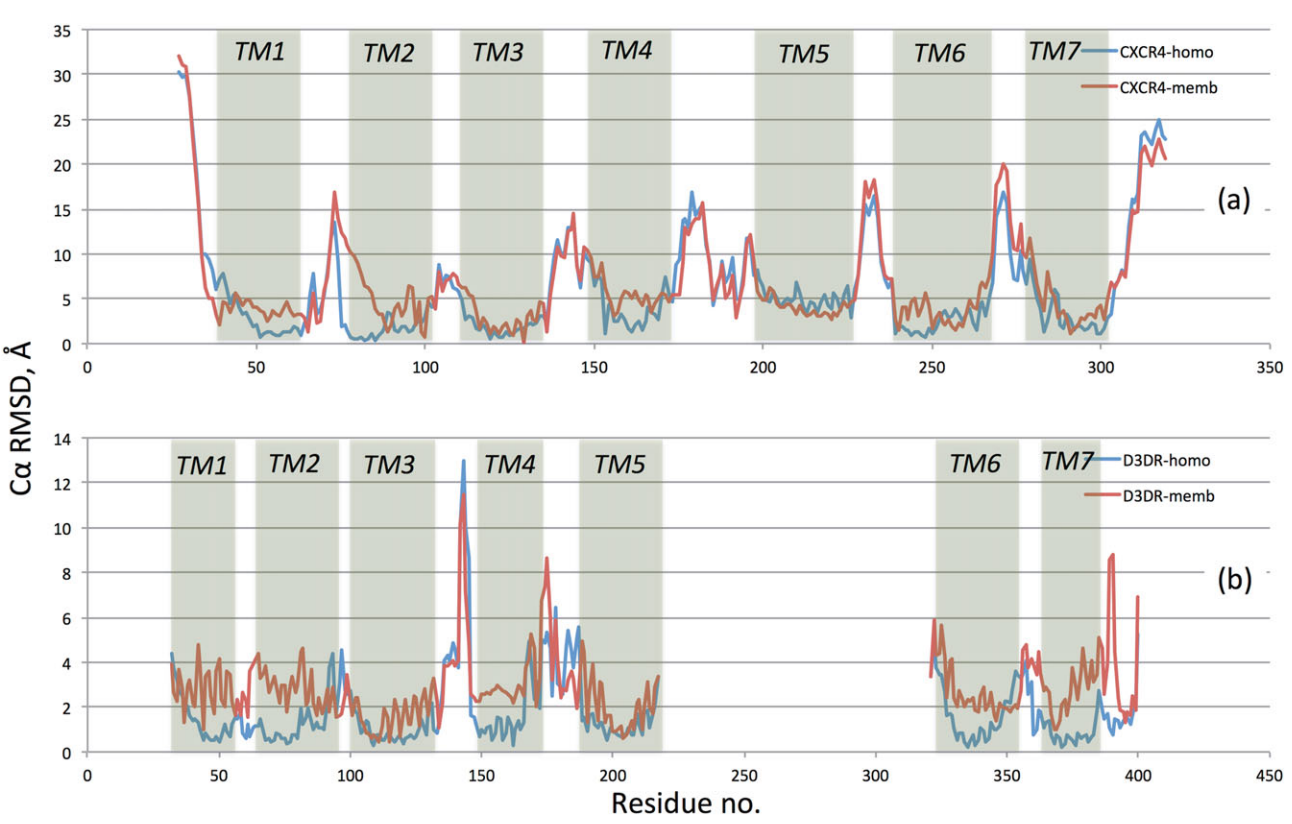

Figure 3

CRMSD of the C_α atoms of each residue in our best models of CXCR4 and D3DR compared to their respective crystal structures. For CXCR4, we used the PDB ID: 3ODU and for D3DR PDB ID: 3PBL. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

is not observed in the homology model. Also TM6 in the homology model shows a smaller kink (23°) compared to the one in the crystal structure (40°). In the D3DR model, all TMs except TM1 and TM3 show significant kinks. TM2 shows the largest kink near $V82^{2.57}$. For D3DR all kinks observed in the crystal structure were correctly predicted in our best model (Supporting Information Table S1).

The helical tilt, rotation, and translation are shown schematically in Supporting Information Figure S5. Supporting Information Figures S1 and S2 show the detailed structural comparison of the predicted models of CXCR4 and D3DR with their respective crystal structures in terms of helical rotation, tilt, and translation for each TM helix. The signs of the rotation and tilt angles in Supporting Information Figures S1, S2, and S3 represent the direction of movement relative to the protein core while viewed from the EC side. A positive tilt indicates tilting toward the protein core (termed "inward" henceforth), whereas negative tilt shows motion away from the protein core ("outward"). Likewise, positive rotation represents anticlockwise motion and vice versa. Figure 4 shows a schematic representation of the major differences between the TM helices of the crystal structures. Among the TM helices of CXCR4 model, TM5 shows the largest deviation from its crystal structure, with a 3.5 Å translation, 17° tilt, and a 58° rotation. The structural differences of TM6 with crystal structure come mainly from helical rigid body deviations, whereas for TM4, TM5, and TM7, these differences stem from both changes in the rigid body degrees of freedom and internal helical deformations. For the D3DR model, the TM helices show closer agreement with the crystal structure. All the helices show negligible translation (<1 Å) and tilt (<5°), except TM1.

Merits and current pitfalls of LITiCon, the helix rigid body optimization procedure for GPCR homology models

At present, homology modeling is the most widely used technique for predicting the structures of GPCRs. However, modeling GPCRs using homology to known crystal structures poses a unique challenge, as not many high-resolution crystal structures are available for distinct receptor classes. We demonstrate how the LITiCon

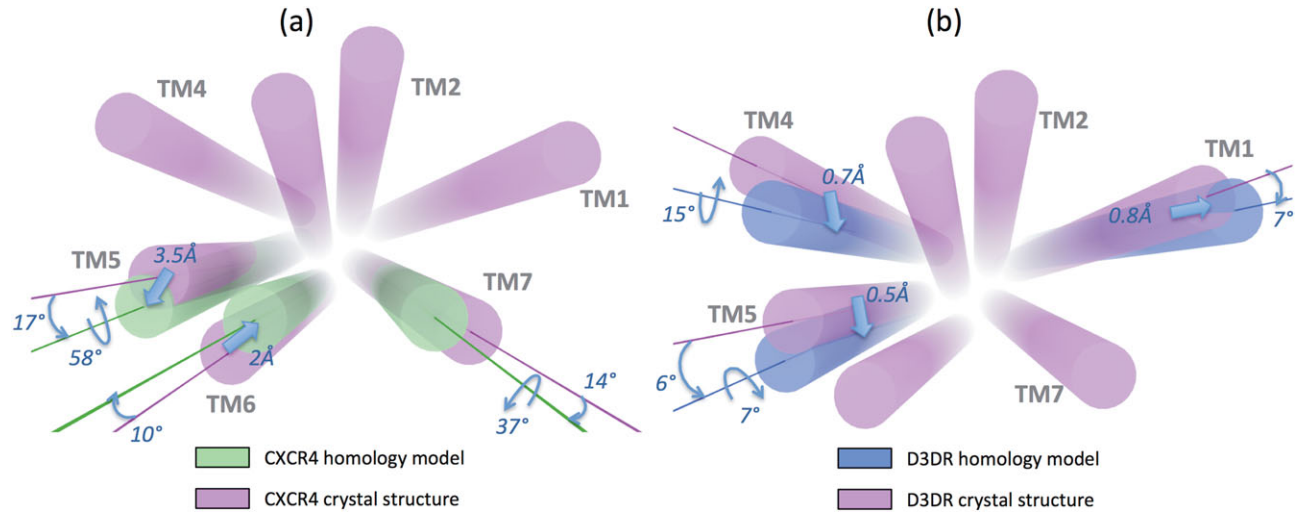

Figure 4

Comparison of rigid body structural differences (translation, tilt, and axial rotation) between the homology models and the corresponding crystal structures: (**a**) CXCR4; (**b**) D3DR; helices are represented by cylinders; block arrows represent center of mass translations, curved arrows represent tilts and rotations. Figures are schematic, not to actual scale. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

method optimizes the helical rotations of the predicted models of CXCR4 compared to its crystal structure in the GPCRDOCK2010 assessment.

In CXCR4, the TM helices TM2 and TM4 are shown to be important for binding nonpeptide antagonists from mutation studies.[16–18] Particularly W94$^{2.60}$ and D97$^{2.63}$ on TM2 and D171$^{4.60}$ on TM4 have been implicated in the binding of noncyclam antagonist.[18] In the CXCR4 homology model based on β_2-AR template, the TXP motif on the EC side of TM2 shows a different kink compared to the β_2-AR crystal structure, which places W94$^{2.60}$ facing the membrane instead of the binding pocket in the homology structure of CXCR4. We used the LITiCon method[1] to optimize the rotational orientation of TM2. Figure 5(a) shows the van der Waals energy of the CXCR4 model for different rotational orientations of TM2. A rotation of 20° relative to the original state was found to be the minimum in energy and was selected as the refined model. A comparison of the unrefined homology model, the refined TM2 model, and the TM2 structure from the CXCR4 crystal structure shown in Figure 5(b) reveals the improvement that is achieved using the rigid body optimization in LITiCon. The TM2 rotational orientation of the predicted model is 3° different in the axial rotation compared to the crystal structure. These subtle structural improvements are important for optimizing the binding pocket as seen in our predicted binding pocket. Our predicted ligand binding pocket was the closest to the crystal structures among all the models submitted to GPCRDOCK2010.

LITiCon optimization was also performed on TM4 of the CXCR4 homology model. The purpose of this was to optimize the position of D171$^{4.60}$ that was facing outward in the original homology model [Fig. 5(d)]. Figure 5(c) shows the plot of the potential energy of TM4 as function of axial rotations, and Figure 5(d) shows the comparison of the original and refined TM4 conformations to the crystal structure. The potential energy landscape shows two distinct minima and we selected the global minimum (rotation angle of 40°) as the refined conformation. In the refined homology model, D171$^{4.60}$ has moved inward into the binding cavity thus improving the agreement with the crystal structure. It should be noted that for these predicted homology models we performed rigid body optimization for some of the helices and not all the seven TM helices. However, before we move toward performing optimization of rigid body orientations for all the seven helices, one needs to develop a robust scoring function that identifies native like structure as low energy conformations.

However, due to the difference in helical kinks, the internal structure of TM4 is different in the homology model compared to the crystal structure, as shown in Figure 5(d). Therefore rigid body optimization would be effective provided we are able to predict the internal structures of the TM helices accurately. Since homology models mimic the kinks and deformations of the template structures rather than the target, efficient computational protocols are needed to optimize the internal deformations of the TM helices such as kinks and bulges.[26] Kink optimization combined with rigid body refinement techniques like LITiCon could lead to a robust platform for homology model refinement.

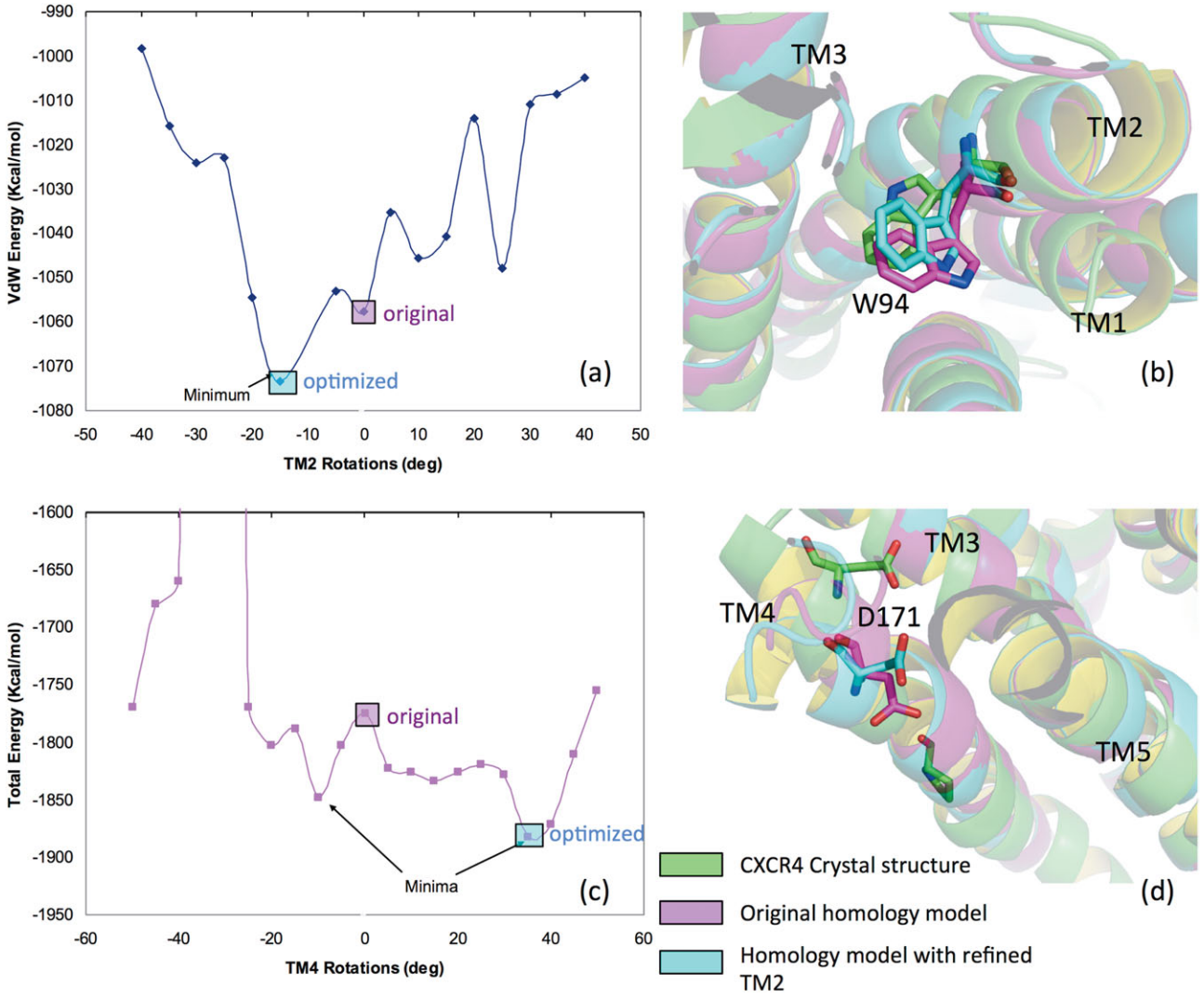

Figure 5

Optimization of TM2 and TM4 orientations of CXCR4 using LITiCon conformational sampling. (**a** and **c**) Plots of potential energy of TM2 and TM4 as functions of helical rotation; the minima in energy are highlighted. The original and optimized conformations are marked using same color code as (**b**) and (**d**). (b) and (d) Overlay of original homology model with the optimized structure and the crystal structure of CXCR4 with IT1t bound.

Comparison of the crystal structures of CXCR4 and D3DR to the crystal structure of β_2-AR in terms of rigid body movements

Table III shows the CRMSD of the C_α atoms of the seven helices of D3DR (PDB ID: 3PBL) and CXCR4 crystal structures (PDB ID: 3ODU) from the β_2-AR crystal structure (PDB ID: 2RH1). For each receptor, the first column shows the CRMSD for each helix when the entire receptor structures are aligned (called "in place" in Table I), and the second column shows the CRMSD when each helix is individually aligned to the corresponding helix in β_2-AR. The "in place" CRMSD gives the overall difference in CRMSD between the two receptors, while the "aligned" CRMSD shows whether the structural differences are due to internal deformations or rigid body orientations. The CXCR4 helices have a higher CRMSD to β_2-AR compared to those of D3DR as anticipated, since their sequences have less than 20% similarity. The TM helices 1, 4, 5, and 6 in CXCR4 show large CRMSD from β_2-AR (>2 Å) with TM1 and TM4 showing the highest CRMSD. TM5 and TM6 of CXCR4 when aligned to those of β_2-AR (Table I, "aligned") show CRMSDs below 2 Å implying that the structural differences in these helices arise mainly from rigid body orientations. In contrast, TM1 and TM4 do not show a sizable reduction in CRMSD on alignment with β_2-AR. Thus their structural differences arise from both rigid body orientations and internal deformations, namely kinks and bulges. Among

Table III
CRMSDs in Å of the Cα Atoms of the TM Helices of CXCR4 and D3DR with Respect to β_2-AR Crystal Structure

	D3DR		CXCR4
TM region	``in place''	``aligned''	``in place''
TM1	1.7	0.7	4.5
TM2	1.0	0.7	2.0
TM3	1.0	0.6	1.5
TM4	2.0	1.4	3.4
TM5	1.8	1.5	2.3
TM6	2.3	0.8	2.9
TM7	1.3	0.7	1.5

CRMSD "*in place*" denotes the CRMSD of a TM helix when the TM domains of the entire receptor are aligned to those of β_2-AR. CRMSD "*aligned*" denotes the CRMSD of the C_α atoms of the TM helix aligned to the corresponding TM region of β_2-AR.

the TM helices in D3DR, only TM4 and TM6 show CRMSDs greater than 2 Å. All D3DR helices show "aligned" CRMSDs less than 2 Å (Table III). Therefore, these differences are caused by rigid body orientations and not by structural deformations.

Further analysis of the rigid body movements such as helical tilt, rotation, and translation of the TM helices of CXCR4 and D3DR with respect to β_2-AR characterizes the structural changes more accurately (Supporting Information Fig. S4). The structural differences of TM1 in CXCR4 are in helical translation (1.4 Å), an inward tilt (17°), and a small rotation (10°). TM2 shows a translation (0.9 Å) and a rotation (20°), but no tilt. TM4 shows translation of 0.5 Å and tilt (12°), with a large rotation (25°). The rigid body differences of TM5 are caused by a large rotation (30°) and relatively minor translation and tilt. TM6 shows mainly translation (0.9 Å) and little tilt (9°), but no rotation. In D3DR, TM4 shows a small rotation (10°), while TM6 shows a small translation (0.6 Å), tilt (5°), and rotation (8°). A schematic representation of these major structural changes is shown in Supporting Information Figure S5.

CONCLUSIONS

Our predicted D3DR and CXCR4 structures with bound ligands were ranked at the top in the GPCRDOCK2010 assessment. Here we have shown that optimization of TM helical tilt, rotation, translation, and gyration of the homology models using the LITiCon method captures the structural differences between the template structure and the homology model, leading to structure refinement for GPCRs. TM2 in chemokine receptors were predicted to have a distinctly different conformation from the biogenic amine receptors. Our predicted structural model has TM2 rotational orientation within 3° of the crystal structure.

The CRMSD is a simplistic measure of comparison of two GPCR structures that lacks a detailed comparison of the structural nuances critical to refinement of homology models. We have developed a geometrical method, GPCRCompare, to compare two GPCR structures in terms of helical tilt, translation, rotation, and gyration degrees of freedom. We have demonstrated the use of this method by comparing crystal structures of D3DR and CXCR4 with β_2-AR. TM1, TM4, TM5, and TM6 were the most structurally diverse between CXCR4 and β_2-AR, while TM3 was the most conserved. TM4 also showed large internal deformations, whereas the structural differences of TM5 and TM6 were mainly from rigid body orientations and not from internal deformations. We have also demonstrated that optimizing the rigid body orientations (i.e., translation, tilt, and rotation) of TM5 and TM6 of CXCR4 leads to improved CRMSD with the β_2-AR crystal structure. Hence rigid body optimization along with kink prediction could be useful in refining crude homology models. We have also discussed the structural comparison of our predicted homology models to the crystal structures and discussed the merits and failures of our prediction scheme.

The ligand docking methods are able to predict the ligand docked poses but the energy functions are not adequate to pick the correct pose. We had to use prior knowledge from other chemokine receptors in selecting the best ligand pose. In summary, refinement of homology models based on rigid body degrees of freedom is critical to deriving accurate homology models for GPCRs.

ACKNOWLEDGMENTS

This work was partially supported by Boehringer-Ingelheim, Biberach, Germany.

REFERENCES

1. Palczewski ,K, Kumasaka T, Hori T, Behnke CA, Motoshima H, Fox BA, Le Trong I, Teller DC, Okada T, Stenkamp RE, Yamamoto M, Miyano M. Crystal structure of rhodopsin: a G-protein-coupled receptor. Science 2000;289:739–745.
2. Cherezov V, Rosenbaum DM, Hanson MA, Rasmussen SGF, Thian FS, Kobilka TS, Choi H-J, Kuhn P, Weis WI, Kobilka BK, Stevens RC. High-resolution crystal structure of an engineered human β2-adrenergic G protein- coupled receptor. Science 2007;318:1258–1265.
3. Scheerer P. Park JH, Hildebrand PW, Kim YJ, Krauss N, Choe HW, Hofmann KP, Ernst OP. Crystal structure of opsin in its G-protein-interacting conformation. Nature 2008;455:497–502.
4. Jaakola V-P, Griffith MT, Hanson MA, Cherezov V, Chien EYT, Lane JR, Ijzerman AP, Stevens RC. The 2.6 Angstrom crystal structure of a human A2A adenosine receptor bound to an antagonist. Science 2008;322:1211–1217.
5. Warne T, Serrano-Vega MJ, Baker JG, Moukhametzianov R, Edwards PC, Henderson R, Leslie AG, Tate CG, Schertler GF. Structure of a beta1-adrenergic G-protein-coupled receptor. Nature 2008;454:486–91.

6. Rasmussen SG, Devree BT, Zou Y, Kruse AC, Chung KY, Kobilka TS, Thian FS, Chae PS, Pardon E, Calinski D, Mathiesen JM, Shah ST, Lyons JA, Caffrey M, Gellman SH, Steyaert J, Skiniotis G, Weis WI, Sunahara RK, Kobilka BK. Crystal structure of the β_2 adrenergic receptor-Gs protein complex. Nature 2011;477:549–555.
7. Kufareva I, Rueda M, Katritch V, Stevens RC, Abagyan R. Status of GPCR modeling and docking as reflected by community-wide GPCR Dock 2010 assessment. Structure 2010;19:1108–1126.
8. Eswar N, Eramian D, Webb B, Shen MY, Sali A. Protein structure modeling with MODELLER. Methods Mol Biol 2008;426:145–159.
9. Bhattacharya S, Hall S, Li H, Vaidehi N. Ligand-stabilized conformational states of human β2 adrenergic receptor: insight into G-protein-coupled receptor activation. Biophys J 2008;94:2027–2042.
10. Bhattacharya S, Vaidehi N. Computational mapping of the conformational transitions in agonist selective pathways of a G-protein coupled receptor. J Am Chem Soc 2010;132:5205–5214.
11. Trabanino RJ, Hall SE, Vaidehi N, Floriano WB, Kam VWT, Goddard W. A first principles prediction of the structure and function of G-protein-coupled receptors: validation for bovine rhodopsin. Biophys J 2004;86:1904–1921.
12. Canutescu AA, Shelenkov AA, Dunbrack RL, Jr. A graph-theory algorithm for rapid protein side-chain prediction. Protein Sci 2003;12:2001–2014.
13. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Jr., Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wi—rkiewicz-Kuczera J, Yin D, Karplus M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998;102:3586–3616.
14. Mayo SL, Olafson BD, Goddard IIIWA. DREIDING-a generic force field for molecular simulations. J Phys Chem 1990;94:8897–8909.
15. Gerlach L, Skerlj R, Bridger G, Schwartz T. Molecular interactions of cyclam and bicyclam non-peptide antagonists with the CXCR4 chemokine receptor. J Biol Chem 2001;276:14153–14160.
16. Rosenkilde MM, Gerlach LO, Jakobsen JS, Skerlj RT, Bridger GJ, Schwartz TW. Molecular mechanism of AMD3100 antagonism in the CXCR4 receptor. J Biol Chem 2004;279:3033–3041.
17. Rosenkilde MM, Gerlach LO, Hatse S, Skerlj RT, Schols D, Bridger GJ, Schwartz TW. Molecular mechanism of action of monocyclam versus bicyclam non-peptide antagonists in the CXCR4 chemokine receptor. J Biol Chem 2007;282:27354–27365.
18. Wong RSY, Bodart V, Metz M, Labrecque J, Bridger G, Fricker SP. Comparison of the potential multiple binding modes of bicyclam, monocylam, and noncyclam small-molecule CXC chemokine receptor 4 inhibitors. Mol Pharmacol 2008;74:1485–1495.
19. Cox BA, Henningsen RA, Spanoyannis A, Neve RL, Neve KA. Contributions of conserved serine residues to the interactions of ligands with dopamine D2 receptors. J Neurochem 2006;59:627–635.
20. DeLano WL. The PyMOL molecular graphics system. Available at:http://www.pymol.org. [Accessed in 2012].
21. Vaidehi N, Pease J, Horuk R. Modeling small molecule compound binding to G-protein coupled receptors. Methods Enzymol 2009;460:263–288.
22. Vaidehi N, Schyler S, Trabanino RJ, Floriano WB, Abrol R, Sharma S, Kochanny M, Koovakkat S, Dunning L, Liang M, Fox JM, Lopes de Mendonça F, Pease JE, Goddard IIIWA, Horuk R. Predictions of CCR1 chemokine receptor structure and BX 471 antagonist binding followed by experimental validation. J Biol Chem 2006;281:27613–27620.
23. Lam AR, Bhattacharya S, Patel K, Hall SE, Mao A, Vaidehi N. The importance of receptor flexibility in binding of cyclam compounds to the chemokine receptor CXCR4. J Chem Inf Model 2011;6:432–43.
24. Nedjai B, Li H, Stroke IL, Wise EL, Webb ML, Merritt JR, Henderson I, Klon AE, Cole AG, Horuk R, Vaidehi N, Pease JE. Small-molecule chemokine mimetics suggest a molecular basis for the observation that CXCL10 and CXCL11 are allosteric ligands of CXCR3. Br J Pharmacol 2012;166:912–923.
25. Visiers I, Braunheim BB, Weinstein H. Prokink: a protocol for numerical evaluation of helix distortions by proline. Protein Eng 2000;13:603–606.
26. Hall SE, Roberts K, Vaidehi N. Position of helical kinks in membrane protein crystal structures and the accuracy of computational prediction. J Mol Graph Model 2009;27:944–950