

Using molecular dynamics for the refinement of atomistic models of GPCRs by homology modeling

Cecylia S. Lupala, Bahareh Rasaeifar, Patricia Gomez-Gutierrez and Juan J. Perez*

Department of Chemical Engineering (ETSEIB), Universitat Politècnica de Catalunya, Av. Diagonal, 647. 08028 Barcelona, Spain

Communicated by Ramaswamy H. Sarma

(Received 4 April 2017; accepted 12 July 2017)

Despite GPCRs sharing a common seven helix bundle, analysis of the diverse crystallographic structures available reveal specific features that might be relevant for ligand design. Despite the number of crystallographic structures of GPCRs steadily increasing, there are still challenges that hamper the availability of new structures. In the absence of a crystallographic structure, homology modeling remains one of the important techniques for constructing 3D models of proteins. In the present study we investigated the use of molecular dynamics simulations for the refinement of GPCRs models constructed by homology modeling. Specifically, we investigated the relevance of template selection, ligand inclusion as well as the length of the simulation on the quality of the GPCRs models constructed. For this purpose we chose the crystallographic structure of the rat muscarinic M3 receptor as reference and constructed diverse atomistic models by homology modeling, using different templates. Specifically, templates used in the present work include the human muscarinic M2; the more distant human histamine H1 and the even more distant bovine rhodopsin as shown in the GPCRs phylogenetic tree. We also investigated the use or not of a ligand in the refinement process. Hence, we conducted the refinement process of the M3 model using the M2 muscarinic as template with tiotropium or NMS docked in the orthosteric site and compared with the results obtained with a model refined without any ligand bound.

Keywords: homology modeling; molecular dynamics; GPCRs; template selection; model refinement

1. Introduction

G-protein-coupled receptors (GPCRs) belong to one of the largest families of integral proteins in eukaryotic organisms, being responsible for the primary mechanism for signal transduction from the cell external environment to its cytoplasm. These receptors are activated by a plethora of stimuli including different kinds of ligands, like protein hormones, lipids, peptides, biogenic amines, nucleotides, or ions and other exogenous stimuli like light, odor, or taste. Agonists bind to receptor active conformation that subsequently binds to a heterotrimeric G-protein in the cytosol, resulting in inhibition or stimulation of the production of second messengers that eventually elicit a cellular response (Katrach, Cherezov, & Stevens, 2013; Kobilka, 2013). Interestingly, most of these receptors exhibit a basal activity in the absence of stimuli, suggesting that part of the receptors on the cell surface are activated. Upon binding to these receptors, ligands can act as agonists, antagonists, or inverse agonists. The differential pharmacological profile exerted by GPCR ligands can be explained assuming that receptors in a cell are in an ensemble of conformations, so that agonists preferably bind to the active conformation, whereas inverse agonists show more affinity for the

resting conformation. Furthermore, antagonists exhibit similar affinity for both conformations and thus, do not disturb the ensemble population, preventing both agonists and inverse agonists to bind to the receptor.

Careful sequence alignments made using the transmembrane domain common to all GPCRs, permits to classify this family of receptor into five main classes: class A or rhodopsin-like, class B or secretin-like, class C or metabotropic glutamate/pheromone, taste/frizzled-like receptors and adhesion receptors (Schiöth & Fredriksson, 2005). Abnormalities of signaling by GPCRs are at the root of disorders that affect most tissues and organs in our body. Consequently, GPCRs are attractive targets for therapeutic intervention with more than 30% of the current drugs on the market (Overington, Al-Lazikani, & Hopkins, 2006).

Early sequence analysis studies permitted to conclude that GPCRs share a common architecture consisting of a bundle of seven transmembrane α -helices (TM1–TM7) connected by three intracellular loops (ICL1–ICL3) and three extracellular loops (ECL1–ECL3). This common architecture was subsequently confirmed by X-ray diffraction studies (Salon, Lodowski, & Palczewski, 2011). Presently, high-resolution structures of more than

*Corresponding author. Email: juan.jesus.perez@upc.edu

25 members of the rhodopsin-like family have been solved bound to diverse ligands including biogenic amines, peptides, nucleosides, and a sphingolipids, providing unprecedented insights into the structural and functional diversity of these proteins (Shonberg, Kling, Gmeiner, & Löber, 2015). Binding of agonists, inverse agonists, and competitive antagonists to GPCRs occurs at the orthosteric site, a pocket located on the extracellular side of the helix bundle that is highly conserved among the members of a subfamily, although binding can be modulated by residues located at the extracellular loops that flank the entrance. Moreover, analysis of the crystallographic structures available suggests that GPCRs are very plastic and that different ligands can affect the intrinsic dynamics of a receptor and activate distinct G protein-mediated signaling pathways, a feature defined as ‘functional selectivity’ (Rajagopal, Rajagopal, & Lefkowitz, 2010). Interestingly, while residues defining the orthosteric binding site are well conserved within the different GPCR subfamilies, those of the extracellular loops are remarkably diverse within a subfamily, providing a good possibility to design selective allosteric ligands (Rajagopal et al., 2010).

Despite the number of crystallographic structures of GPCRs available having increased steadily in the last few years, there are still challenges that hamper the availability of novel structures including their low-expression yields, low receptor stability after detergent extraction from native membranes, and high conformational heterogeneity (Ostermeier & Michel, 1997). In the absence of crystallographic structures of many members of the GPCR family, homology modeling can be used to construct atomic resolution models that can be used: for virtual screening studies; to accurately predict ligand binding pockets; to study drug-receptor recognition as well as the clarification of experimental data (Cutolo, Basdevant, Bernadat, Bachelierie, & Ha-Duong, 2017; Gandhimathi & Sowdhamini, 2016; Simms et al., 2009). However, reliable modeling of GPCR-ligand complexes for the vast majority of receptors with unknown structure remains to be one of the most challenging goals for computer-aided drug design (Spyrakakis & Cavasotto, 2015).

The publication of the bovine rhodopsin structure at atomic resolution provided for the first time a template to construct realistic models of GPCRs by homology modeling (Palczewski et al., 2000). The structure has been widely used as template to construct models of diverse members of the rhodopsin-like family (Fanelli & De Benedetti, 2011). Although sequence identity with other GPCRs is small (about 20% on average), conservation of specific features among all the members of the rhodopsin-like family provides support for considering that there is a high structural similarity among members of the family and consequently, for its use as template (Bissantz, Logean, & Rognan, 2004). Despite its

successful use as template, its suitability as template to accurately predict the structure of other GPCRs was soon questioned (Archer, Maigret, Escricut, Pradayrol, & Fourmy, 2003). A critical point regards its low sequence identity with other GPCRs and also, the high diversity of ligand-binding features observed among different GPCRs. The subsequent publication of the crystallographic structures of diverse ligand-infused GPCR complexes a few years later, made these drawbacks more evident as inferred from the results of the diverse community-wide GPCR-Dock assessment competitions (Kufareva, Katritch, Participants of GPCR Dock 2013, Stevens, & Abagyan, 2014). In these studies it was established a 35–40% sequence identity between target and template as an empirical cutoff for reliable homology-based prediction of ligand-receptor interactions. Presently, modeling approaches achieve close-to-experimental accuracy for small rigid orthosteric ligands when templates with high sequence identity are used (Carlsson et al., 2011). However, there are still many issues that need to be addressed to improve the quality of the models constructed by homology modeling. Diverse reports published recently address issues like template selection (Costanzi et al., 2016; Latek, Pasznik, Carlomagno, & Filipek, 2013; Mobarec, Sanchez, & Filizola, 2009; Rataj, Witek, Mordalski, Kosciolk, & Bojarski, 2014; Worth, Kleinau, & Krause, 2009; Zhu & Li, 2012); inclusion of knowledge-based constraints into the modeling process (Rodríguez, Ranganathan, & Carlsson, 2014); modeling of the extracellular loops (Goldfeld, Zhu, Beuming, & Friesner, 2013) and on structure refinement (Thomas et al., 2014).

In the present study, we investigated the use of molecular dynamics (MD) simulations for the refinement of GPCRs models constructed by homology modeling (Nowroozi & Shahlaei, 2017). Specifically, we investigated the relevance of template selection, ligand inclusion as well as the length of the simulation on the quality of the GPCRs models constructed. For this purpose, we chose the crystallographic structure of the rat muscarinic M3 receptor as reference (Kruse et al., 2012) and constructed diverse atomistic models by homology modeling, using different templates. Specifically, templates used in the present work include the human muscarinic M2 (Haga et al., 2012); the more distant human histamine H1 (Shimamura et al., 2011) and the even more distant bovine rhodopsin (Palczewski et al., 2000) as shown in the GPCRs phylogenetic tree. We also investigated the use or not of a ligand in the refinement process. To that end we conducted the refinement process of the M3 model using the M2 muscarinic as template with tiotropium or N-methylscopolamine (NMS) docked in the orthosteric site and compared with the results obtained with a model refined without any ligand bound. Preliminary results of this work have been

already reported (Lupala, Rasaeifar, Gomez-Gutierrez, & Perez, 2015; Martinez-Archundia, Cordomi, Garriga, & Perez, 2012).

2. Methods

The procedure to construct atomistic models of GPCRs by homology modeling involves the selection of a template; target-template sequence alignment; construction of a crude model and verification; and model refinement (Cavasotto & Palomba, 2015; Nayeem, Sitkoff, & Krystek, 2006). As mentioned above, homology models of the rat muscarinic M3 receptor were constructed from diverse templates using comparative modeling by means of the Molecular Operating Environment (MOE) package (2013.08; Chemical Computing Group ULC, Montreal, QC, Canada). The sequence of the rat M3 muscarinic receptor was retrieved from the Uniprot database (ID: P08484), whereas for the diverse templates, we used the sequences of the corresponding crystallographic structures, taking care of removing those residues corresponding to the T4 Lysozyme inserted between TM5 and TM6 from the respective sequences, if applicable (PDB ID: 3UON for the human muscarinic M2; PDB ID: 3RZE for the human histamine H1 and PDB ID: 1GZM for bovine rhodopsin).

2.1. Model construction

Multiple sequence alignment of the sequences was carried out using the MOE-Align tool (2013.08; Chemical Computing Group ULC, Montreal, QC, Canada). Once the alignment was completed and checked for unwanted gaps within the TM region, we proceeded to construct diverse atomistic models for each template. In order to simplify the construction of the atomistic models of the rat M3 receptor without compromising their accuracy, the long intracellular loop (ICL3) was substituted by a stretch of a few residues. Specifically, in the case of both the M2 muscarinic and the histamine H1 receptors, since both crystallographic structures correspond to fusion proteins of the receptor with the T4-lysozyme inserted in the ICL3, we simply removed the sequence of lysozyme and joined the residues left. Thus, in the case of the M2 muscarinic receptor four residues from the C-terminus of TM5 and seven residues from the N-terminus of TM6 and five residues from the C-terminus of TM5 and three residues from the N-terminus of TM6 in the case of the histamine H1 receptor. In the case of rhodopsin, we simply considered all residues of the corresponding ICL3. Accordingly, diverse crude homology models for each of the templates were constructed by threading the sequence of the target receptor to each of the crystallographic structures with the subsequent incorporation of alternative sidechain conformations using an extensive rotamer

library generated from a high-resolution structural database embedded in MOE. Once hydrogens were added using the protonate3D method (Labute, 2008), the crude models were energy minimized using a contact energy function to relieve any serious steric strains. The diverse models generated for each template were scored according to their *rmsd* to the average structure using the C $^{\alpha}$ atoms for the calculation. The model with the highest score was considered as a crude model and used for further refinement. Finally, the stereochemical quality of the models constructed was assessed by the distribution of the backbone dihedral angles in the Ramachandran map.

2.2. Molecular docking

Before proceeding the refinement process, the antagonist tiotropium (**1**) (Figure 1) co-crystallized in the rat muscarinic M3 receptor (PDB ID: 4DAJ) was docked onto the orthosteric site of each of the three crude models, using the GLIDE algorithm (Friesner et al., 2004). The compound was docked in multiple orientations and multiple conformations. The resulted poses were rank ordered by their binding/docking score using the glide-score function. The best poses were analyzed visually and validated in accordance to the information available on the involvement of specific residues in binding from diverse site directed mutagenesis studies (Blüml, Mutschler, & Wess, 1994; Han et al., 2005). Crude homology models with tiotropium bound were used as starting models for refinement. In addition, in order to understand the effect of using a ligand bound in the refinement process, two additional starting models for the M2 template were constructed: one bound to the antagonist NMS (**2**) (Figure 1) and another without any ligand bound. Henceforth, four protein-ligand complexes of the M3 muscarinic receptor and one without ligand were refined using molecular dynamic simulations.

2.3. Model refinement

Models were refined using MD simulations (Grossfield, 2011; Sadiq et al., 2013). For each model, the starting structure was embedded into a box of 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipids and water molecules previously equilibrated according to the procedure described elsewhere (Cordomi, Edholm, & Perez, 2007). The box had an initial size of $10.3 \times 8.0 \times 10.2 \text{ nm}^3$ (XYZ), organized in such a way that the bilayer plane was oriented on the XY plane. Before protein insertion, the box contained 256 lipids (corresponding to an area per lipid of 0.64 nm^2) and circa 17,000 water molecules. The protein was placed in the center of the box, and the overlapping molecules were removed. Specifically, all water molecules with oxygen atoms closer than 0.40 nm to a non-hydrogen atom of the protein, as well as all lipid molecules with at least one

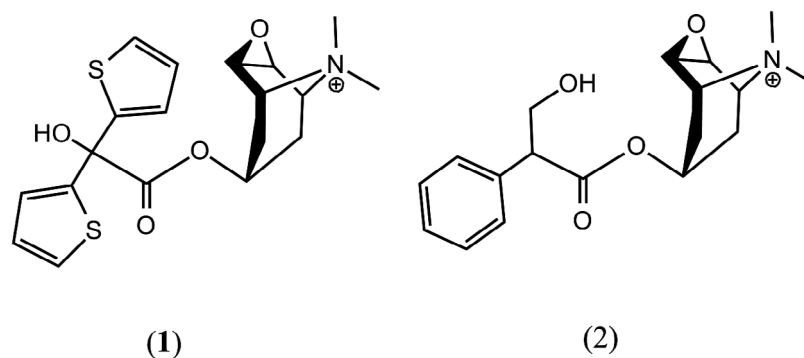


Figure 1. Chemical structures of tiotropium (1) and N-methylscopolamine (NMS) (2).

atom closer than 0.25 nm to a non-hydrogen atom of the protein were removed. This resulted in a final system containing 188 lipids and circa 14,655 water molecules. Removal of these atoms introduced small voids between the protein and water or lipid molecules that disappeared during the first part of the MD simulation, in which a progressive adjustment of the lipid bilayer and water molecules to the protein takes place. Next, 105 randomly selected water molecules were replaced by 45 sodium and 60 chloride ions, providing a neutral system with a concentration approximately 0.2 M on sodium chloride. This concentration is fairly similar to that found in biological organisms, although they exhibit different intra- and extracellular ion concentrations.

For each model, a 500 ns MD simulation was carried out at constant pressure using the GROMACS package 4.6 (Van Der Spoel et al., 2005). The all-atom OPLS force field (Kaminski, Friesner, Tirado-Rives, & Jorgensen, 2001) currently implemented in GROMACS, was used to describe all molecules of the systems, except for water that was modeled using the TIP3P model (Jorgensen, Chandrasekhar, Madura, Impey, & Klein, 1983). The systems were subjected to periodic boundary conditions in the three coordinate directions. The temperature was kept constant at 300 K using separate thermostats for the protein, water, ions and lipid molecules. The time constant for the thermostats was set to 0.1 ps except for water, for which a smaller value of 0.01 ps was used. The pressure in the three coordinate directions was kept at 0.1 MPa by independent Berendsen barostats using a time constant of 1.0 ps. The equations of motion were integrated using the leapfrog algorithm with a time step of 2 fs. All bonds involving hydrogen atoms within the protein and lipid molecules were kept frozen using the LINCS algorithm (Miyamoto & Kollman, 1992). The bonds and the angle of water molecules were fixed using the analytical SETTLE method. Lennard-Jones interactions were computed using a cutoff of 1.0 nm. The electrostatic interactions were treated either using the PME technique (Essmann et al., 1995).

3. Results and Discussion

As mentioned above, the goal of the present study was to investigate the performance of molecular dynamics simulations used in the refinement process. For this purpose, we carried out a ‘semi-blind’ homology modeling study of the rat M3 muscarinic receptor whose crystallographic structure is available, using the human M2 muscarinic receptor with and without an antagonist bound to it, the human H1 histamine receptor and bovine rhodopsin as templates.

Sequence alignment of the rat muscarinic M3 receptor with the human muscarinic M2, the human histamine H1, and bovine rhodopsin shows sequence identities of 60, 34 and 17%, respectively (see Figure 2). These values are larger if only the transmembrane (TM) regions are considered, with values of 79, 41, and 23%, respectively. Interestingly, the best sequence identity score is found in the TM3 segment, whereas the poorer is found in the TM1 segment in all the cases (Venkatakrishnan et al., 2013). As in the rest of the muscarinic receptor subtypes, the intracellular loop ICL3 of M3 is very long and was omitted for modeling purposes. The deletion of this large segment bears no consequences to the overall structure of the GPCR or the orthosteric binding pocket as in the muscarinic M3 receptor-phage T4 lysozyme fused protein used for crystallographic studies where, receptor ability to bind agonist or antagonist ligands is not modified.

As mentioned above, crude models were constructed by threading the sequence of the target receptor on the crystallographic structure of the templates with the subsequent incorporation of the corresponding side chains using a library of conformers. For each template alternative models were generated using template backbone coordinates with alternative side chain conformations using an extended rotamer library implemented in the MOE package (2013.08; Chemical Computing Group ULC, Montreal, QC, Canada). Following this process a number of independent models, based on loop and side chain placements were scored using a contact energy function. Among these, the model with the highest score

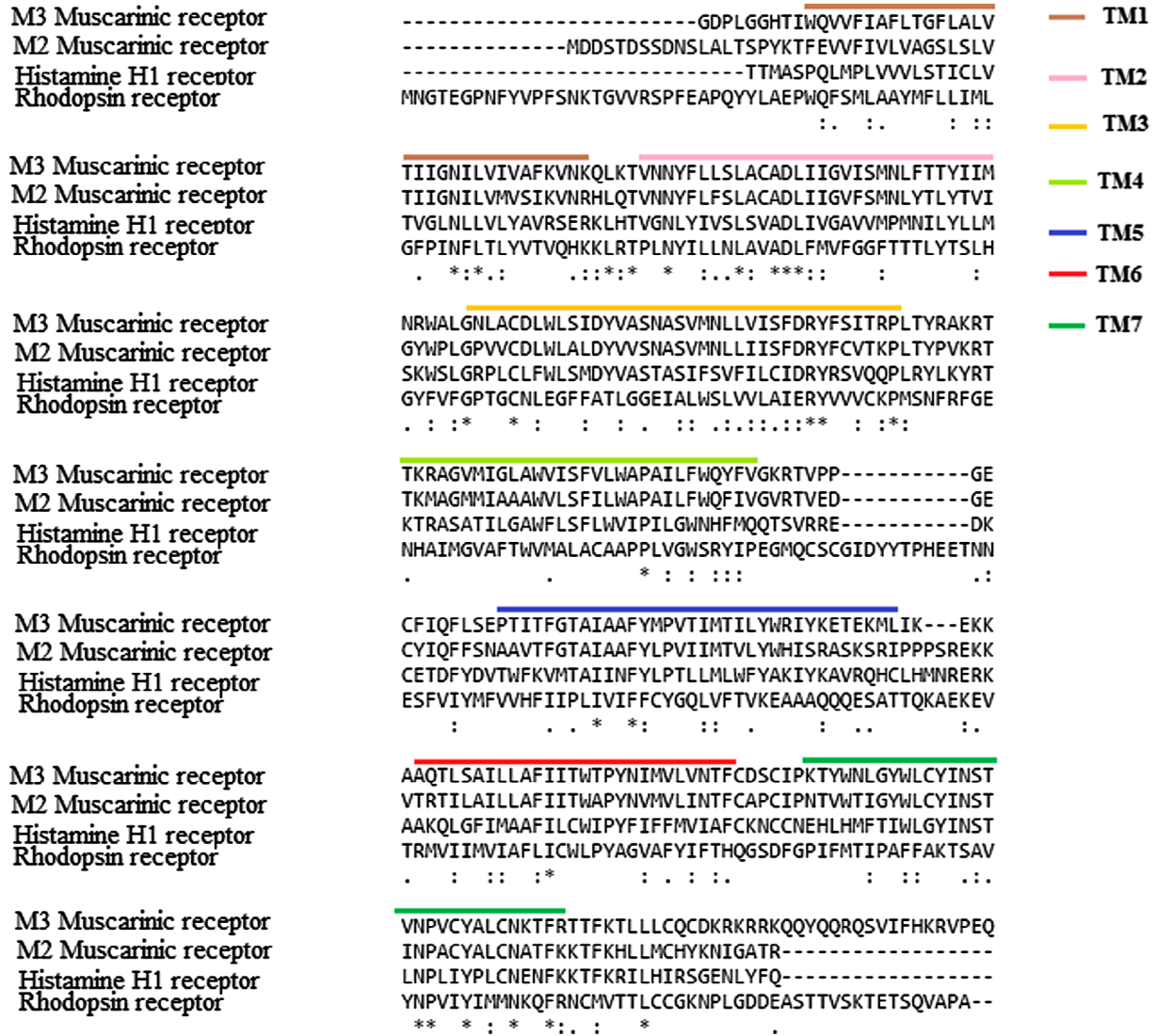


Figure 2. Sequence alignment of the rat muscarinic M3 receptor (target) and the three templates (human muscarinic M2, human histamine H1 and bovine rhodopsin) used in the present study. The transmembrane regions are shown in colors as TM1-TM7.

was selected for refinement process. These crude models presented similar distances to the target M3 crystallographic structure than the corresponding crystallographic structures of the diverse templates. Thus, the muscarinic M2 receptor exhibits a root-mean square deviation (*rmsd*) 1.6 Å; 1.9 Å in the case of the histamine H1 receptor; and 2.3 Å in the case of rhodopsin (using the backbone C α for the measure).

An important drawback of the crude models constructed by homology modeling regards transmembrane helix lengths. Residues involved in the diverse TMs, both in the crude models and in the M3 crystallographic structure are listed in Table 1. As can be seen, for some of the helices differences between the crude model and the M3 target structure are notorious. Thus, in the crude model

generated using M2 as template there are not large differences. Important exceptions are TM3 that is seven residues longer than the corresponding helix in the M3 crystallographic structure and TM6 that is five residues longer. In contrast, crude models generated using the histamine H1 and rhodopsin as templates show larger deviations. Specifically, in the case of the H1 receptor TM1 is five residues shorter; TM4 is eight residues longer and TM6 is eleven residues longer, whereas the rest of the helix segments have similar lengths. In the case of rhodopsin TM4 is six residues shorter, whereas TM5 is five residue longer and TM2 and TM6 exhibit a three residue difference. Taking into account that a α -helix contains 3.6 residues per turn, differences in helix length can sum up to two turns.

Table 1. Amino acid segments involved in the diverse TM helices. The first column regards the crystallographic M3 receptor, whereas the rest of the columns correspond to the diverse homology models. In this case there are two entries for each TM helix: the upper corresponds to the starting structure whereas the lower corresponds to the segment after refinement.

Helix	M3 Crystal structure	M3 _{M2}	M3 _{M2} NMS	M3 _{M2} with Tiotropium	M3 _{HRH1} with Tiotropium	M3 _{RHO} with Tiotropium
TM1	W66 - V95	W66 - V95	W66 - V95	W66 - V95	F70 - K92	I64 - F93
		W66 - V95	W66 - V95	W66 - V95	F70 - F93	V65 - F93
TM2	V102 - M131	V102 - M131	V102 - M131	V102 - M131	V102 - I130	I103 - I130
		V102 - M131	V102 - M131	V102 - M131	V102 - I130	T101 - Y128
TM3	L139 - T171	G137 - Y176	G137 - Y176	G137 - Y176	G137 - Y176	N138 - I170
		N138 - I170	N138 - T171	N138 - T171	A140 - T171	L139 - Y167
TM4	T182 - V211	T182 - V211	T182 - V211	T182 - V211	T182 - A203	T182 - L205
		T182 - V211	T182 - V211	T182 - V211	K183 - Q208	R180 - W200
TM5	P229 - K256	P229 - K256	P229 - K256	P229 - K256	P229 - E257	E228 - K260
		P229 - K256	P229 - K256	P229 - K256	P229 - E257	P229 - E259
TM6	L492 - T514	K488 - T515	K488 - T515	K488 - T515	I484 - C517	A489 - T514
		A490 - T515	Q491 - T515	Q491 - T515	K488 - C517	I484 - T514
TM7	K522 - C546	K522 - C546	K522 - C546	K522 - C546	K522 - C546	K522 - C546
		K522 - C546	K522 - C546	K522 - C546	K522 - C546	K522 - C546

3.1. The refinement process

The refinement process was aimed at relaxing the crude model to capture specific structural features of the target protein, not present in the template. As described in the methods section, refinement was carried out using MD simulations of a system consisting of the target protein embedded in a POCP lipid bilayer. The final model of the target receptors were produced from the average structures computed of the last 50 ns of the respective production runs, followed by energy minimization of the protein structure using an effective dielectric constant of 2 to mimic the protein environment. This process was carried out simply to eliminate possible crashes between atoms found in the average structure.

In order for the refinement process to be robust, simulation times require to be long enough to ensure that the system is equilibrated. System equilibration can be monitored by the time evolution of the *rmsd* of the successive trajectory snapshots in regard to the starting structure. Figure 3(a) shows the *rmsd* time evolution using the α -carbons of all residues for each of the simulations performed in the present work. We also carried out the same exercise considering only those residues involved in the transmembrane domain (Figure 3(b)). As can be seen, equilibration takes more than 200 ns of simulation time. Comparison of the *rmsd* time evolution between the whole protein and the transmembrane domain suggests that there is a large contribution of the loops to the high *rmsd* values observed. Moreover, analysis of Figure 3(a) and (b) clearly shows that during the refinement process, the initial structures suffer a reorganization that is larger for those templates that are more distant from the target receptor in the phylogenetic tree, as expected. Thus, models of M3 constructed from

the M2 muscarinic receptor exhibit a *rmsd* of ~ 2.3 Å, the model constructed from the histamine H1 receptor exhibit a *rmsd* of ~ 2.7 Å and larger for the model constructed from rhodopsin that reaches ~ 3.0 Å. Another interesting point inferred from the analysis of Figure 3(a) and (b) is that those models refined with a ligand bound exhibit shorter equilibration times. This can be clearly seen by looking at the *rmsd* time evolution of the three models constructed using the M2 receptor as template. The faster equilibration must be due to the smaller flexibility of the system expected when a ligand is bound to the protein (Lin, Gether, & Kobilka, 1996). These results agree with the use of ligands to model active sites for in silico screening (Evers, Gohlke, & Klebe, 2003).

The root-mean-square fluctuations (*rmsf*) of the protein residues in the refinement process are shown in Figure 4(a) and (b). Specifically, Figure 4(a) shows the fluctuations per residue for each of the models along the whole trajectory, whereas Figure 4(b) shows the fluctuations in the last 200 ns. Inspection of Figure 4(a) and (b) reveals that the largest fluctuations are associated with the loops, as expected. Furthermore, comparison of the two plots points to the large fluctuations that occur before equilibration, suggesting that they are basically the result of loop rearrangement rather than actual domain fluctuations. For example, the EC1 and IC3 loops in rhodopsin show large fluctuations at the beginning of the refinement process that disappears after 200 ns. Similar behavior is shown by loops EC1 and EC2 in the model constructed from the H1 histamine receptor. The evolution of the fluctuations during the refinement process provides important information to compute the average structure of the model.

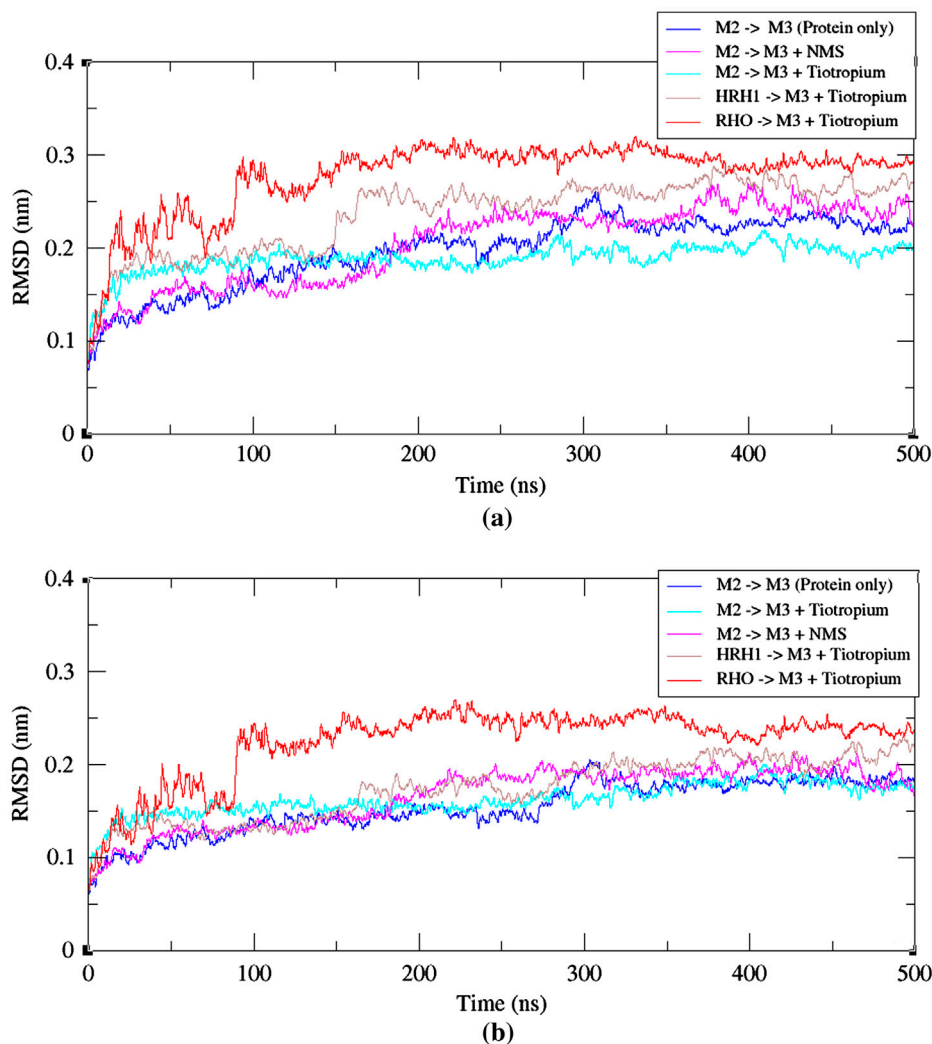


Figure 3. Time evolution of the *rmsd* of the diverse models studied in the present work from their respective starting structures along the refinement process. (a) *rmsd* computed using the $C\alpha$ of all the residues of the protein; (b) *rmsd* computed using the $C\alpha$ of the residues of the transmembrane domain.

In regard to the transmembrane domain, it is very encouraging to observe that the refinement process is able to correct the length of the TM helices. Table 1 lists the amino acids involved in the diverse TM helices both in the crystallographic M3 receptor and in the diverse models. In this latter case for each TM helix there are two entries: the upper corresponds to the starting model, whereas the lower corresponds to the TM helix after the refinement process. As shown in Table 1, in the case of the models constructed using the M2 as template, both TM3 and TM6 are longer than in the crystallographic structure of the target receptor, however during the refinement process the length of the helices is shortened, providing a model that is only one residue longer than in the crystallographic structure. In the case of the model constructed from the histamine H1 receptor TM1, TM3

and TM6 are longer than the crystallographic structure of the target receptor, whereas TM4 is shorter. As can be seen from Table 1, except for the TM1 the refinement process is capable to modify helix length accordingly. In the case of the model constructed from rhodopsin TM4 is shorter and TM6 longer than the target crystallographic structure. Unfortunately, in this case the refinement process is not capable to get the right length for these two segments.

A key point for assessing the performance of the refinement process is to monitor the improvement of the model from the starting structure. Figure 5 displays the time evolution of the *rmsd* of the models compared to the crystallographic structure of the muscarinic M3 receptor, using the C^α of the residues that belong to transmembrane region. Inspection of Figure 5

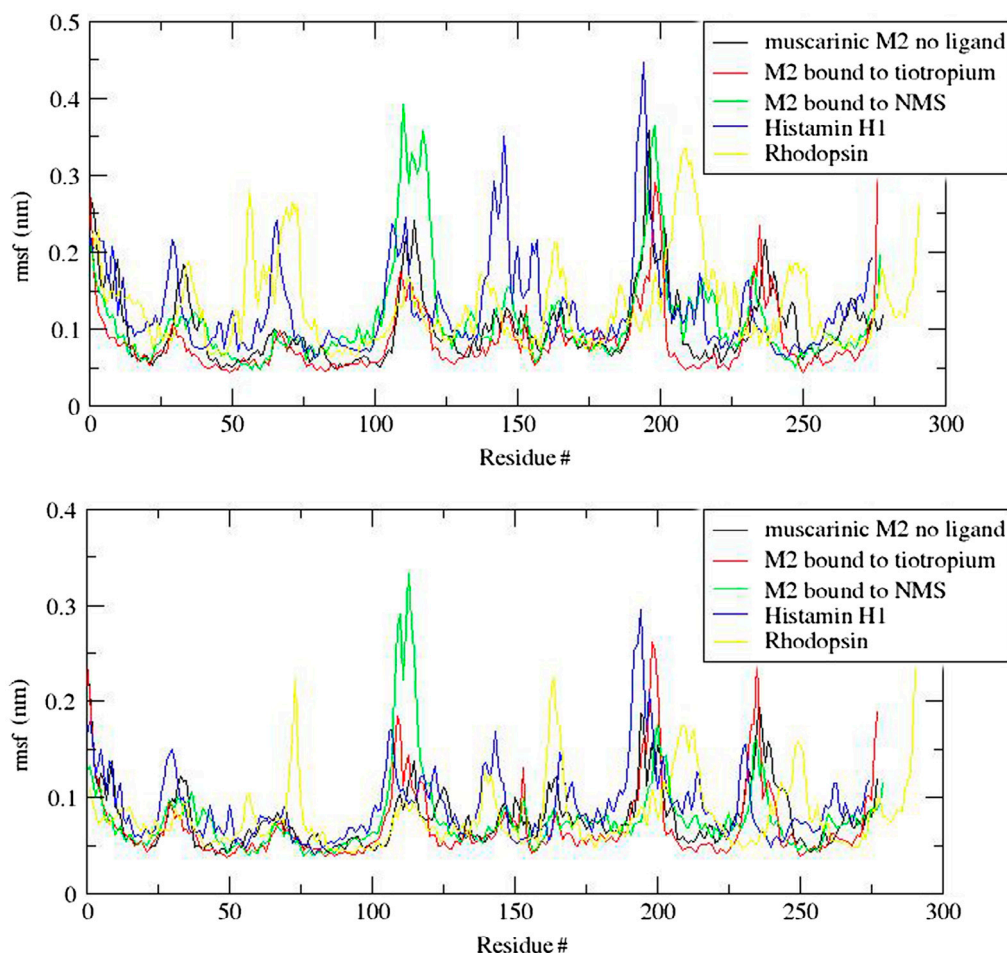


Figure 4. Root mean square fluctuations (*rmsf*) per residue of the diverse models computed along the refinement process. (a) *rmsf* computed along the whole refinements process. (b) *rmsf* obtained during the last 200 ns of the refinement process.

suggests that in general, the *rmsd* decreases during the refinement process with a consequent net improvement of the structures. Specifically, after the 500 ns the model constructed using rhodopsin as template shows an improvement from 4.8 Å to 4.3 Å; for the model built from the histamine there is not a net improvement; whereas, for the models constructed from the muscarinic M2 there is an improvement from 2.5 Å to 2.3 Å for the case that the model was constructed without ligand and smaller, from 2.0 Å to 1.8 Å for the model refined using NMS as ligand and, from 1.8 Å to 1.5 Å for the model constructed with tiotropium. In contrast, the *rmsd* using the whole protein (data not shown) does not show a net improvement in the refinement process due to the fluctuations of the loops.

In regard to the loops, the benefits of the refinement process are not so obvious. Thus, present results once more suggest that the accuracy of the loops obtained after the refinement process depends on the distance between the template and the target in the phylogenetic

tree. However, the large fluctuations experienced by loops during the MD simulation tend to produce average structures that despite capturing the features of the target structure are not very accurate. Specifically, present results suggest that when the initial structure is close to the target, loops are worsened after the refinement process. In contrast, when the initial structure is far from the target, loops are improved after the refinement process. Figures 6 and 7 show the superimposition of the ECL2 and the ICL2 loops, respectively, for the diverse models studied in the present study before and after the refinement process. Inspection of Figure 6(a) shows for those models constructed using the M2 muscarinic receptor the ECL2 loop is well superimposed to that of the target M3 muscarinic receptor, whereas in the case of the models constructed using the histamine H1 and rhodopsin receptors large deviations are obvious. However, after refinement (Figure 6(b)) models constructed using the M2 muscarinic receptor show the loops with the features of the target receptor but with small deviations. In

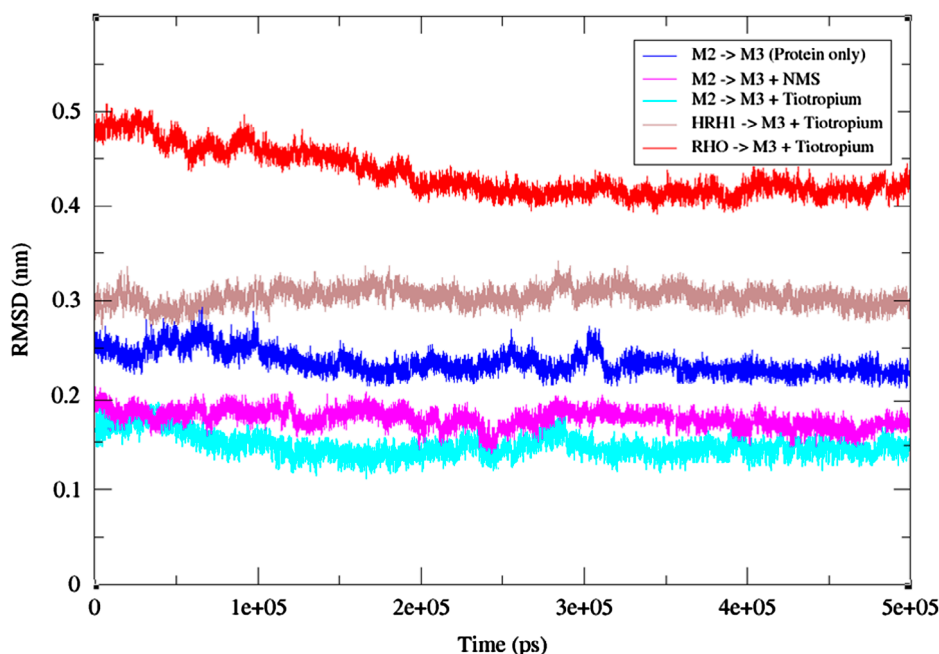


Figure 5. Time evolution of the *rmsd* of the diverse homology models from the target crystallographic structure computed using the *Ca* of the residues of the transmembrane domain.

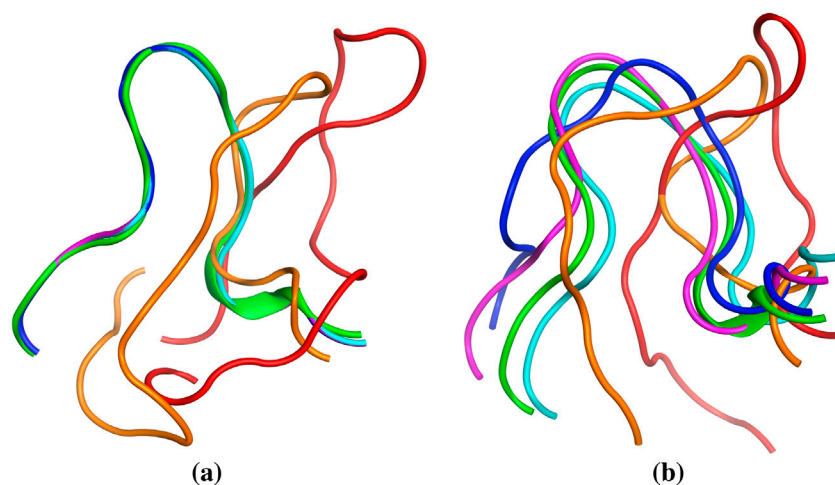


Figure 6. Superposition of the ECL2 loop corresponding to the five models studied in the present work. (a) Before the refinement process. (b) After the refinement process. In Green is the rat M3 crystal structure; in cyan the M2 template-tiotropium model; in magenta the M2 template-NMS model; in dark blue the M2 template without ligand model; in orange the histamine template-tiotropium model; and in red the rhodopsin template-tiotropium model.

contrast, the loops of the models constructed using the histamine H1 and rhodopsin show a modest improvement in regard to the starting structure. The same trends can be deduced from the analysis of the ICL2 (Figure 7(a) and (b)). Specifically, for the models constructed using the M2 muscarinic receptor as template, the loop of the initial models exhibits a helical structure that superimposes well in the target structure, but shows

deviations after the refinement process. In contrast, in the models constructed using the histamine H1 and rhodopsin the loop does not exhibit any secondary structure in the initial models, but it is captured after refinement. In view of these results, it can be concluded that the large fluctuations suffered by the loops in the refinement process worsen the structure of the loops when the template and the target are close although the structural

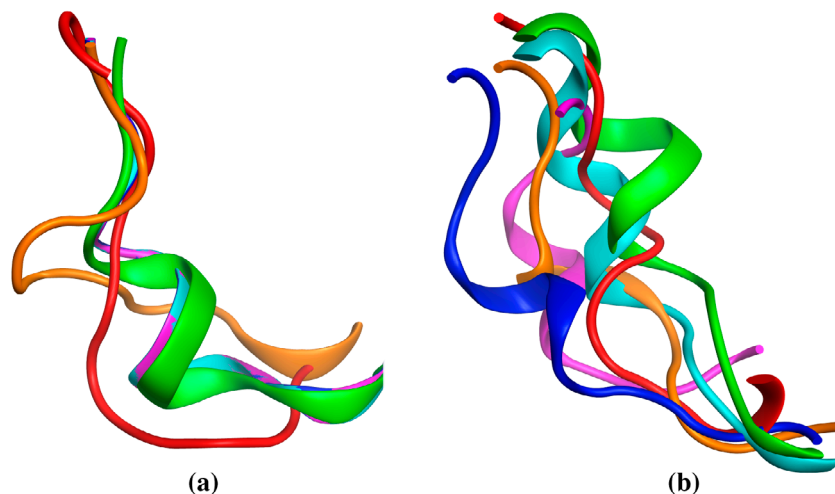


Figure 7. Superposition of the ICL2 loop corresponding to the five models studied in the present work. (a) Before the refinement process. (b) After the refinement process. In Green is the rat M3 crystal structure; in cyan the M2 template-tiotropium model; in magenta the M2 template-NMS model; in dark blue the M2 template without ligand model; in orange the histamine template -tiotropium model; and in red the rhodopsin template-tiotropium model.

features are preserved. In contrast, when the template is far from the target the structures are improved, although longer sampling times are presumably required.

3.2. Analysis of the homology models constructed

As expected, the models constructed using the M2 muscarinic receptor as template, are closer to the target structure. Values of the *rmsd* of the diverse models constructed using the human muscarinic M2 receptor (using backbone C α) are: 1.9 Å for the model refined with no ligand bound; 2.4 Å for the model constructed with tiotropium bound and 2.5 Å for the model constructed with NMS bound. However, the relative high *rmsd* values are due to the loops. The models are able to capture small rearrangements of the TM1 N-terminus segment; TM5 C-terminal segment, as well as TM7 middle segment toward the interior of the helix bundle, shown in the comparison of the M2 and M3 crystallographic structures (Kruse et al., 2012). Despite these rearrangements, there are certain features associated with the loops that are present in the starting models and are lost during the refinement process. Specifically, the pronounced outward bend at the extracellular end in TM4, found in the crystallographic structures of the M2 and M3 muscarinic receptors (Kow & Nathanson, 2012) is preserved in the models constructed with tiotropium or NMS, but lost in the model constructed without any ligand.

The model of the muscarinic M3 constructed using the histamine H1 receptor bound to tiotropium yields an *rmsd* of 3.2 Å compared to the reference crystallographic structure. Comparison of the two structures reveals

differences between the modeled and the crystallographic structure, being the largest difference found at the C-terminus of TM4. In this model both the ECL2 and ICL2 loops do not accommodate well to the space shown by the models constructed using the muscarinic M2 receptor, although it can be seen certain improvement in regard to the initial structures (see Figures 6 and 7). Regarding the length of the TM segments, the refinement process reduces slightly the differences found in the starting structure (see Table 1), but are not completely resolved. Specifically, TM1 is six residues shorter; TM2 and TM3 are one residue shorter; TM4 is in the final model four residues shorter catching up partly to the eight residues shorter of the starting model. Contrastingly, TM5 remains one residue longer while TM6 is seven residues longer correcting in part the eleven residues of the starting structure and TM8 remains one residue shorter.

Finally, the model of the muscarinic M3 receptor constructed using the rhodopsin bound to tiotropium yields a *rmsd* with the crystallographic structure of 4.79 Å similar to the starting model, suggesting that little improvement has been achieved during the refinement process. Comparison of the two structures reveals large differences between the modeled and the crystallographic structure. On the one hand, like in the model constructed using the histamine H1 receptor both the ECL2 and ICL2 loops do not accommodate well to the space shown by the models constructed using the muscarinic M2 receptor and only a small improvement can be seen in regard to the initial structures (see Figures 6 and 7). In regard to the length of the TM segments the refinement process reduces slightly the differences found in

the starting structure (see Table 1), but differences are still remarkable. Specifically, TM2 is one residue shorter; TM3 is four residues shorter; TM4 is thirteen residues shorter worsening the starting model; TM5 remains one residue longer and TM6 is nine residues longer.

Finally, in order to understand the utility of the models generated for docking studies, we docked tiotropium onto the diverse models before and after being subjected to the refinement process. The crystallographic structure of tiotropium bound to the muscarinic M3 receptor shows the ligand sitting in a pocket conformed by helices TM3, TM5, TM6 and TM7. It is oriented in such a way that the ligand shows the tiophene groups in the proximity of TM5, whereas the quaternary nitrogen is flanked by TM3 and TM7. The ligand exhibits diverse interactions with the neighboring residues: the quaternary nitrogen sits close to Asp¹⁴⁷; Asn⁵⁰⁷ forms two polar interactions with the carbonyl and with the hydroxyl groups of tiotropium, respectively; the epoxy group exhibits an interaction with the sidechain of Ser¹⁵² and one of the tiophene groups sits at interacting distance with Trp⁵⁰⁴ (Kruse et al., 2012). All the models described in the present work are capable to reproduce the pose tiotropium adopts when bound to the M3 receptor. However, in regard to the diverse interactions with the sidechains of the neighboring residues, most of the initial structures fulfill only part the set of ligand-receptor interactions found in the crystal, however after refinement all the structures fulfill all the features of the crystallographic structure.

4. Conclusions

The aim of the present work was to assess the performance of molecular dynamics simulations of the receptor embedded in a lipid bilayer for the purpose of refining homology models of GPCRs. Specifically, several homology models of the rat muscarinic M3 receptor whose crystallographic structure is known were constructed using diverse templates located at diverse distances on the phylogenetic tree. Specifically, templates used in the present work include the human muscarinic M2 receptor; the human histamine receptor H1 and bovine rhodopsin. We also investigated the effect of using a ligand bound to the orthosteric site on the performance of the refining process. For this purpose we also constructed models of the rat muscarinic M3 receptor using the human M2 muscarinic receptor without ligand and with tiotropium and NMS bound. The study reveals that these systems achieve equilibration after more than 250 ns of simulation, although if only the transmembrane region is considered equilibration is achieved much faster. Present results show that the use of molecular dynamics improves the quality of the homology models mostly on the transmembrane region. Specifically, the

refinement process permits the correction of the length of the helices and improves the accuracy of the helix bundle. Despite the constructed models capturing most of the features of the M3 receptor, the distance in the phylogenetic tree affects their quality, being the most accurate models those constructed using a template close to the target receptor. Accordingly, the refinement process as performed in this work is not powerful enough to correct the starting model when the chosen template is distant from the target. Moreover, inclusion of a ligand on the modeling makes the refinement more robust since equilibration is observed faster.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Archer, E., Maigret, B., Escrieut, C., Pradayrol, L., & Fourmy, D. (2003). Rhodopsin crystal: New template yielding realistic models of G-protein-coupled receptors? *Trends in Pharmacological Sciences*, 24, 36–40. doi:10.1016/S0165-6147(02)00009-3
- Bissantz, C., Logean, A., & Rognan, D. (2004). High-throughput modeling of human G-protein coupled receptors: Amino acid sequence alignment, three-dimensional model building, and receptor library screening. *Journal of Chemical Information and Computer Sciences*, 44, 1162–1176. doi:10.1021/ci034181a
- Blüml, K., Mutschler, E., & Wess, J. (1994). Functional role in ligand binding and receptor activation of an asparagine residue present in the sixth transmembrane domain of all muscarinic acetylcholine receptors. *Journal of Biological Chemistry*, 269, 18870–18876.
- Carlsson, J., Coleman, R. G., Setola, V., Irwin, J. J., Fan, H., Schlessinger, A., ... Shoichet, B. K. (2011). Ligand discovery from a dopamine D3 receptor homology model and crystal structure. *Nature Chemical Biology*, 7, 769–778. doi:10.1038/nchembio.662
- Cavasotto, C. N., & Palomba, D. (2015). Expanding the horizons of G protein-coupled receptor structure-based ligand discovery and optimization using homology models. *Chemical Communications*, 51, 13576–13594. doi:10.1039/c5cc05050b
- Cordomi, A., Edholm, O., & Perez, J. J. (2007). Effect of different treatments of long-range interactions and sampling conditions in molecular dynamic simulations of rhodopsin embedded in a dipalmitoyl phosphatidylcholine bilayer. *Journal of Computational Chemistry*, 28, 1017–1030. doi:10.1002/jcc.20579
- Costanzi, S., Skorski, M., Deplano, A., Habermehl, B., Mendoza, M., Wang, K., ... Gao, J. (2016). Homology modeling of a class A GPCR in the inactive conformation: A quantitative analysis of the correlation between model/template sequence identity and model accuracy. *Journal of Molecular Graphics and Modeling*, 70, 140–152. doi:10.1016/j.jmgl.2016.10.004
- Cutolo, P., Basdevant, N., Bernadat, G., Bachelier, F., & Ha-Duong, T. (2017). Interaction of chemokine receptor CXCR4 in monomeric and dimeric state with its endogenous ligand CXCL12: Coarse-grained simulations identify

- differences. *Journal of Biomolecular Structure and Dynamics*, 35, 399–412. doi:10.1080/07391102.2016.1145142
- Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., & Pedersen, L. G. (1995). A smooth particle mesh Ewald method. *The Journal of Chemical Physics*, 103, 8577–8593. doi:10.1063/1.470117
- Evers, A., Gohlke, H., & Klebe, G. (2003). Ligand-supported homology modelling of protein binding-sites using knowledge-based potentials. *Journal of Molecular Biology*, 334, 327–345. doi:10.1016/j.jmb.2003.09.032
- Fanelli, F., & De Benedetti, P. G. (2011). Update 1 of: Computational modeling approaches to structure function analysis of G protein-coupled receptors. *Chemical Reviews*, 111, PR438–PR535. doi: 10.1021/cr100437t
- Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., ... Shenkin, P. S. (2004). Glide: A new approach for rapid accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of Medicinal Chemistry*, 47, 1739–1749. doi:10.1021/jm0306430
- Gandhimathi, A., & Sowdhamini, R. (2016). Molecular modelling of human 5-hydroxytryptamine receptor (5-HT_{2A}) and virtual screening studies towards the identification of agonist and antagonist molecules. *Journal of Biomolecular Structure and Dynamics*, 34, 952–970. doi:10.1080/07391102.2015.1062802
- Goldfeld, D. A., Zhu, K., Beuming, T., & Friesner, R. A. (2013). Loop prediction for a GPCR homology model: Algorithms and results. *Proteins*, 81, 214–228. doi:10.1002/prot.24178
- Grossfield, A. (2011). Recent progress in the study of G protein-coupled receptors with molecular dynamics computer simulations. *Biochimica Biophysica Acta*, 1808, 1868–1878. doi:10.1016/j.bbamem.2011.03.010
- Haga, K., Kruse, A. C., Asada, H., Yurugi-Kobayashi, T., Shi-roishi, M., Zhang, C., ... Kobayashi, T. (2012). Structure of the human M2 muscarinic acetylcholine receptor bound to an antagonist. *Nature*, 482, 547–551. doi:10.1038/nature10753
- Han, S. J., Hamdan, F. F., Kim, S. K., Jacobson, K. A., Bloodworth, L. M., Li, B., & Wess, J. (2005). Identification of an agonist-induced conformational change occurring adjacent to the ligand-binding pocket of the M3 muscarinic acetylcholine receptor. *Journal Biological Chemistry*, 280, 34849–34858. doi:10.1074/jbc.M506711200
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*, 79, 926–935. doi:10.1063/1.445869
- Kaminski, G., Friesner, R. A., Tirado-Rives, J., & Jorgensen, W. L. (2001). Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *Journal of Physical Chemistry B*, 105, 6474–6487. doi:10.1021/jp003919d
- Katritch, V., Cherezov, V., & Stevens, R. C. (2013). Structure-function of the G protein-coupled receptor superfamily. *Annual Review of Pharmacology and Toxicology*, 53, 531–556. doi:10.1146/annurev-pharmtox-032112-135923
- Kobilka, B. (2013). The structural basis of G-protein-coupled receptor signaling (Nobel Lecture). *Angewandte Chemie International Edition*, 52, 6380–6388. doi:10.1002/anie.201302116
- Kow, R. L., & Nathanson, N. M. (2012). Muscarinic receptors become crystal clear. *Nature*, 482, 480–481. doi:10.1038/482480a
- Kruse, A. C., Hu, J., Pan, A. C., Arlow, D. H., Rosenbaum, D. M., Rosemond, E., ... Kobilka, B. K. (2012). Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature*, 482, 552–556. doi:10.1038/nature10867
- Kufareva, I., Katritch, V., Participants of GPCR Dock 2013, Stevens, R. C., & Abagyan, R. (2014). Advances in GPCR modeling evaluated by the GPCR dock 2013 assessment: Meeting new challenges. *Structure*, 22, 1120–1139. doi:10.1016/j.str.2014.06.012
- Labute, P. (2008). Protonate3D: Assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins*, 75, 187–205.
- Latek, D., Pasznik, P., Carlomagno, T., & Filipek, S. (2013). Towards improved quality of GPCR models by usage of multiple templates and profile-profile comparison. *PLoS ONE*, 8, e56742. doi:10.1371/journal.pone.0056742
- Lin, S., Gether, U., & Kobilka, B. K. (1996). Ligand stabilization of the β_2 adrenergic receptor: Effect of DTT on receptor conformation monitored by circular dichroism and fluorescence spectroscopy. *Biochemistry*, 35, 14445–14451. doi:10.1021/bi961619+
- Lupala, C. S., Rasaeifar, B., Gomez-Gutierrez, P., & Perez, J. J. (2015). Effect of template selection on the construction of atomistic models of GPCRs by homology modeling. *Journal of Biomolecular Structure and Dynamics*, 33, 127–128. doi:10.1080/07391102.2015.1032830
- Martinez-Archundia, M., Cordomi, A., Garriga, P., & Perez, J. J. (2012). Molecular modeling of the M3 acetylcholine muscarinic receptor and its binding site. *Journal of Biomedicine and Biotechnology*, 1–12. Article ID 789741. doi: 10.1155/2012/789741
- Miyamoto, S., & Kollman, P. A. (1992). Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models. *Journal of Computational Chemistry*, 13, 952–962. doi:10.1002/jcc.540130805
- Mobarec, J. C., Sanchez, R., & Filizola, M. (2009). Modern homology modeling of G-protein coupled receptors: Which structural template to use? *Journal of Medicinal Chemistry*, 52, 5207–5216. doi:10.1021/jm9005252
- Nayeem, A., Sitkoff, D., & Krystek, S., Jr. (2006). A comparative study of available software for high-accuracy homology modeling: From sequence alignments to structural models. *Protein Science*, 15, 808–824. doi:10.1110/ps.051892906
- Nowroozi, A., & Shahlaci, M. (2017). A coupling of homology modeling with multiple molecular dynamics simulation for identifying representative conformation of GPCR structures: A case study on human bombesin receptor subtype-3. *Journal of Biomolecular Structure and Dynamics*, 35, 250–272. doi:10.1080/07391102.2016.1140593
- Ostermeier, C., & Michel, H. (1997). Crystallization of membrane proteins. *Current Opinion on Structural Biology*, 7, 697–701. doi:10.1016/S0959-440X(97)80080-2
- Overington, J. P., Al-Lazikani, B., & Hopkins, A. L. (2006). How many drug targets are there? *Nature Reviews on Drug Discovery*, 5, 993–996. doi:10.1038/nrd2199
- Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Moto-shima, H., Fox, B. A., & Miyano, M. (2000). Crystal structure of rhodopsin: A G protein-coupled receptor. *Science*, 289, 739–745. doi:10.1126/science.289.5480.739
- Rajagopal, S., Rajagopal, K., & Lefkowitz, R. J. (2010). Teaching old receptors new tricks: Biasing seven-transmembrane receptors. *Nature Reviews on Drug Discovery*, 9, 373–386. doi:10.1038/nrd3024

- Rataj, K., Witek, J., Mordalski, S., Kosciolk, T., & Bojarski, A. J. (2014). Impact of template choice on homology model efficiency in virtual screening. *Journal of Chemical Information and Computer Modeling*, 54, 1661–1668. doi:10.1021/ci500001f
- Rodríguez, D., Ranganathan, A., & Carlsson, J. (2014). Strategies for improved modeling of GPCR-drug complexes: Blind predictions of serotonin receptors bound to ergotamine. *Journal of Chemical Information and Computer Modeling*, 54, 2004–2021. doi:10.1021/ci5002235
- Sadiq, S. K., Guixà-González, R., Dainese, E., Pastor, M., De Fabritiis, G., & Selent, J. (2013). Molecular modeling and simulation of membrane lipid-mediated effects on GPCRs. *Current Medicinal Chemistry*, 20, 22–38.
- Salon, J. A., Lodowski, D. T., & Palczewski, K. (2011). The significance of G protein-coupled receptor crystallography for drug discovery. *Pharmacology Reviews*, 63, 901–937. doi:10.1124/pr.110.003350
- Schiöth, H. B., & Fredriksson, R. (2005). The GRAFS classification system of G-protein coupled receptors in comparative perspective. *General and Comparative Endocrinology*, 142, 94–101. doi:10.1016/j.ygcen.2004.12.018
- Shimamura, T., Shiroishi, M., Weyand, S., Tsujimoto, H., Winter, G., Katriitch, V., ... Iwata, S. (2011). Structure of the human histamine H1 receptor complex with doxepin. *Nature*, 475, 65–70. doi:10.1038/nature10236
- Shonberg, J., Kling, R. C., Gmeiner, P., & Löber, S. (2015). GPCR crystal structures: Medicinal chemistry in the pocket. *Bioorganic and Medicinal Chemistry*, 23, 3880–3906. doi:10.1016/j.bmc.2014.12.034
- Simms, J., Hall, N. E., Lam, P. H. C., Miller, L. J., Christopoulos, A., Abagyan, R., & Sexton, P. M. (2009). Homology modeling of GPCRs. *Methods in Molecular Biology*, 552, 97–113. doi:10.1007/978-1-60327-317-6_7
- Spyrakakis, F., & Cavasotto, C. N. (2015). Open challenges in structure-based virtual screening: Receptor modeling, target flexibility consideration and active site water molecules description. *Archives of Biochemistry and Biophysics*, 583, 105–119. doi:10.1016/j.abb.2015.08.002
- Thomas, T., McLean, K. C., McRobb, F. M., Manallack, D. T., Chalmers, D. K., & Yuriev, E. (2014). Homology modeling of human muscarinic acetylcholine receptors. *Journal of Chemical Information and Computer Modeling*, 54, 243–253. doi:10.1021/ci400502u
- Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. (2005). GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry*, 26, 1701–1718. doi:10.1002/jcc.20291
- Venkatakrishnan, A. J., Deupi, X., Lebon, G., Tate, C. G., Schertler, G. F., & Babu, M. M. (2013). Molecular signatures of G-protein-coupled receptors. *Nature*, 494, 186–194. doi:10.1038/nature11896
- Worth, C. L., Kleinau, G., & Krause, G. (2009). Comparative sequence and structural analyses of G-protein-coupled receptor crystal structures and implications for molecular models. *PLoS ONE*, 4, e7011. doi:10.1371/journal.pone.0007011
- Zhu, M., & Li, M. (2012). Revisiting the homology modeling of G-protein coupled receptors: β 1-Adrenoceptor as an example. *Molecular BioSystems*, 8, 1686–1693. doi:10.1039/c2mb05491d