In silico analysis of the binding of agonists and blockers to the β2-adrenergic receptor

Santiago Vilar, Joel Karpiak, Barkin Berk, Stefano Costanzi *

Laboratory of Biological Modeling, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, DHHS, Bethesda, MD 20892, USA

ABSTRACT

Activation of G protein-coupled receptors (GPCRs) is a complex phenomenon. Here, we applied Induced Fit Docking (IFD) in tandem with linear discriminant analysis (LDA) to generate hypotheses on the conformational changes induced to the β2-adrenergic receptor by agonist binding, preliminary to the sequence of events that characterize activation of the receptor. This analysis, corroborated by a follow-up molecular dynamics study, suggested that agonists induce subtle movements to the fifth transmembrane domain (TM5) of the receptor. Furthermore, molecular dynamics also highlighted a correlation between movements of TM5 and the second extracellular loop (EL2), suggesting that freedom of motion of EL2 is required for the agonist-induced TM5 displacement. Importantly, we also showed that the IFD/LDA procedure can be used as a computational means to distinguish agonists from blockers on the basis of the differential conformational changes induced to the receptor. In particular, the two most predictive models obtained are based on the RMSD induced to Ser207 and on the counterclockwise rotation induced to TM5.

Published by Elsevier Inc.

1. Introduction

G protein-coupled receptors (GPCRs) constitute the largest and most ubiquitous superfamily of membrane receptors [1]. Structurally, they are constituted by a single polypeptide chain that spans the plasma membrane with seven transmembrane domains (TMs, numbered from TM1 to TM7) connected by three extracellular and three intracellular loops (ELs and ILs, numbered from EL1 to EL3 and from IL1 to IL3). Being the most common target of pharmacological intervention, GPCRs represent an extremely attractive field for pharmaceutical research. For these reasons, many efforts have been directed toward their structural characterization and, thanks to important technological and scientific breakthroughs, brought to the recent solution of the structure of a few of the members of the superfamily [2–5]. Importantly, these structures provide detailed information directly applicable to drug discovery, allowing for an effective identification of novel ligands through virtual screening [6–14]. Moreover, they also serve as a platform to calibrate and further the accuracy of theoretical structures derived from homology modeling and molecular docking [11–17].

Many pharmacological, biophysical and computational studies have been devoted to the investigation of the mechanisms underlying GPCR activation. These combined efforts led to the hypothesis that local conformational changes induced by agonist binding are propagated through a number of subsequent structural rearrangements that gradually provoke a transition from the inactive to a variety of active conformations. Ultimately, these activated states are responsible for triggering complex G protein dependent or independent signaling pathways [1,18–23]. Of note, GPCR activation has also been proposed to be associated with changes of the dimeric interface of the receptors [24,25].

While GPCR activation is a complex phenomenon, here we describe computational studies intended to generate hypotheses on the immediate conformational changes induced by agonists to the β2-adrenergic receptor (β2-AR), preliminary to the transition of the receptor to the activated state. Specifically, we studied the local conformational changes that agonists and blockers differentially induce to the β2-AR by applying linear discriminant analysis (LDA) [26] in tandem with a docking procedure [27] that takes into account the flexibility of receptor and ligands (Fig. 1). The study was conducted in an unbiased manner, analyzing all possible conformational changes occurring within the domains of the receptor that line the interhelical cavity. In particular, flexibility was granted to three large receptor segments, each of which comprising the exofacial half of two adjacent transmembrane domains and the extracellular loop that connects them (Fig. 2). Importantly, the hypotheses on agonist binding resulting from LDA analysis of flexible docking data were also corroborated by follow-up molecular dynamics simulations, conducted after embedding the receptor in a hydrated lipid bilayer model. Moreover, our study indicated that LDA analysis of flexible docking data, besides being applicable to the generation of structural hypotheses on ligand binding, can also

* Corresponding author.
    E-mail address: stefanoc@mail.nih.gov (S. Costanzi).
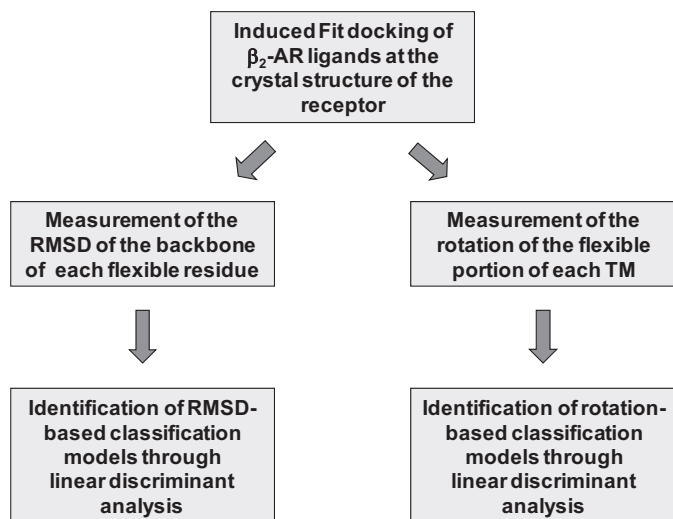
Fig. 1. Flowchart showing the different phases implicated in the development of the IFD/LDA models.

offer an effective computational procedure for the differentiation of ligands into agonists and blockers.

The β_2-AR, and GPCRs in general, can couple to multiple G proteins and can signal through G protein independent pathways, in a ligand-dependent manner [22]. Throughout this paper, whenever we refer to agonists and blockers of the β_2-AR, we refer to compounds capable of stimulating or blocking receptor-mediated G_s activation.
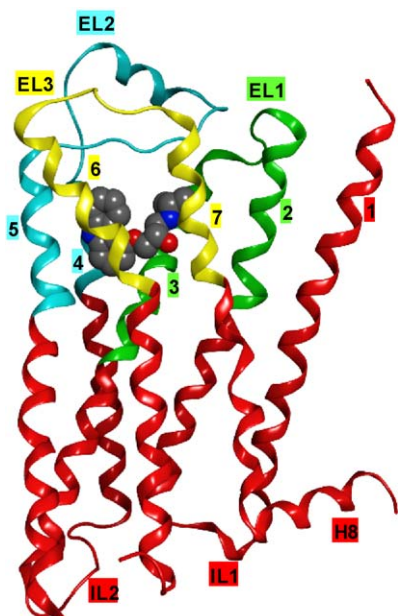
Fig. 2. Ribbon representation of the β_2-AR (PDB ID: 2RH1) indicating the residues treated as flexible in the IFD procedure. Segment 1 (from Met82 to Ile121) is colored in green, segment 2 (from Thr164 to Tyr209) in cyan and segment 3 (from Trp286 to Tyr316) in yellow. The co-crystallized carazolol is represented in space filling mode. Labels indicate the seven TMs, numbered from 1 to 7, extracellular and intracellular loops (ELs and ILs), and the amphipathic helix 8 (H8). The third intracellular loop (IL3) is not represented because it was replaced with T4-lysozyme in the crystal structure. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

2. Results

A previously collected database of 66 β_2-AR ligands [10] – 34 agonists and 32 blockers of the G_s stimulation pathway, see Table S1 of the Supplementary data – was docked by means of Schrödinger's Induced Fit Docking (IFD) [27] procedure at the crystal structure of the β_2-AR obtained in complex with the inverse agonist carazolol (PDB ID: 2RH1) [28,29]. Notably, we define agonists all compounds with intrinsic activity greater than 15% of that measured for the full agonist isoproterenol (including partial agonists and full agonists), while we define blockers all the remaining compounds (including very weak partial agonists, neutral antagonists and inverse agonists). As shown in Fig. 2, in the IFD procedure we allowed flexibility to all the residues contained within three specific sequence segments that form the exofacial half of the receptor and enclose the interhelical binding cavity. Specifically, each segment was composed by the exofacial half of two adjacent TMs and the entire interconnecting loop: TM2-EL1-TM3 (from Met82$^{2.53}$ to Ile121$^{3.40}$), TM4-EL2-TM5 (from Thr164$^{4.56}$ to Tyr209$^{5.48}$), and TM6-EL3-TM7 (from Trp286$^{6.48}$ to Tyr316$^{7.43}$). The N-terminus and the exofacial half of TM1 were not granted flexibility because of their distance from the ligand binding site.

For each docked ligand, the IFD procedure yielded a number of complexes, each with a specific ligand pose and a specific binding pocket conformation. After having detected for each ligand the most populated cluster of docking poses, we eliminated all the spurious poses, defined as those having an RMSD (root mean square deviation) higher than 3 Å from the lowest energy structure of the main cluster. For most of the compounds, the main cluster of poses featured the positively charged amino group interacting with Asp113$^{3.32}$ and the aromatic moiety pointing towards TM5, in line with what was shown for carazolol by the crystal structure [28,29]. However, for 5 ligands (acebutolol, carvedilol, fenoterol, formoterol and T-0509) the main cluster did not show these well-known interactions. Thus, for these ligands we discarded the first cluster and considered, instead, the second most populated cluster.

We subsequently used the ensemble of receptor conformations produced for each of the docked ligands to calculate the average RMSD of each flexible residue and the average rotation of each flexible TM segment, with respect to the crystal structure of the receptor in complex with its inverse agonist carazolol. Notably, the monitoring of the rotation of the TMs was possible because the region to which we granted flexibility in the docking procedure contained uninterrupted TM segments.

2.1. Classification models based on the residue RMSDs

To identify the individual residues most differentially affected by agonists and blockers, we subjected all the calculated RMSD values to linear discriminant analysis (LDA) [26], using the forward stepwise algorithm for variable selection. As a result, we found a group of residues located in TM5 to be the most discriminating. In particular, high RMSD levels for these residues were associated with docking of agonists and low RMSD levels were associated with docking of blockers. This is also graphically evident from the plots shown in Fig. 3, which compare side-by-side the average RMSD of the flexible residues detected for a representative agonist (isoproterenol) and a representative blocker (carazolol).

Details of the three models that yielded the best percentages of good classification are given in Table 1. The most predictive of these models is based on the RMSD of the backbone of Ser207$^{5.46}$. In particular, we found that the probability of a compound to be an agonist ($P_{agonist}$) was related to the average RMSD induced by the compound on the backbone of Ser207$^{5.46}$ through the following

Isoproterenol
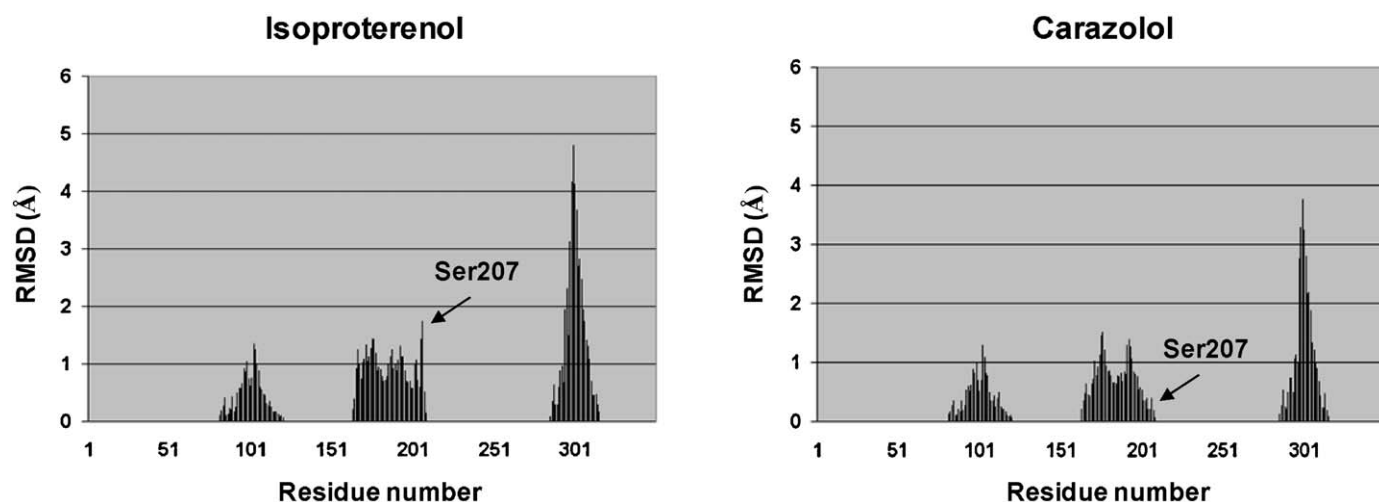
Carazolol

Fig. 3. Plots of the average RMSDs induced to the residues treated as flexible in the IFD procedure by a representative agonist (isoproterenol) and a representative inverse agonist (carazolol). Residues with RMSD values of zero were treated as rigid.

equation:

$$P_{agonist} = 6.754 \times (\text{RMSD of Ser207}^{5.46}) - 8.296 \tag{1}$$

As graphically represented in Fig. 4, according to this equation, a demarcation line corresponding to an average backbone RMSD of Ser207$^{5.46}$ of 1.23 Å classified correctly 82.35% of the agonists and 93.75% of the blockers. The exception were six agonists that induced RMSDs lower than 1.23 Å – namely formoterol, salbutamol, sulfonterol, AH-3021, broxaterol, and T-0509 – and two blockers that induced RMSDs greater than 1.23 Å – namely (SRRR)-Nebivolol and sotalol.

We also analyzed the RMSD of the entire residues, side chain included. The results are in line with those obtained with the backbone RMSD, although the percentages of good classification were slightly lower (data not shown).

2.2. Classification models based on TM rotations

As mentioned, in addition to RMSD values, we also monitored the rotations induced by the different ligands to the various TMs. Our LDA analysis revealed that the rotation of TM5 yielded by far the best discriminatory model (Table 2). Specifically, greater counterclockwise rotations of TM5 were associated with the binding of agonists, while smaller counterclockwise rotations of the same TM were associated with the binding of blockers. In particular, we found the probability of a compound to be an agonist to be related to the rotation that the compound induced on TM5 by the following equation:

$$P_{agonist} = 0.591 \times (\text{TM5 of rotation}) - 3.650 \tag{2}$$

As graphically represented in Fig. 5, according to this equation, a demarcation line corresponding to a counterclockwise rotation of TM5 of 6.18° classified correctly 85.3% of the agonists and 87.5% of the blockers. The exceptions were five agonists that induced coun-

Table 1
The RMSD induced by the ligands on all the residues treated as flexible in the Induced Fit Docking procedure was subjected to a forward stepwise variable selection to identify the models endowed with the highest discriminatory power. The top three resulting LDA models are based on the RMSD of the backbone of Ser207$^{5.46}$, Val206$^{5.45}$ and Phe208$^{5.47}$.

	% CC[a]	CC[b]	MC[c]
Model 1 (RMSD_Ser207$^{5.46}$)			
Agonist	82.35	28	6
Blockers	93.75	30	2
Total	87.88	58	8
$U = 0.499$ $F(1,64) = 64.28809$ $p < 0.000$			
Model 2 (RMSD_Val206$^{5.45}$)			
Agonist	82.35	28	6
Blockers	81.25	26	6
Total	81.82	54	12
$U = 0.543$ $F(1,64) = 53.848$ $p < 0.000$			
Model 3 (RMSD_Phe208$^{5.47}$)			
Agonist	79.41	27	7
Blockers	87.50	28	4
Total	83.33	55	11
$U = 0.595$ $F(1,64) = 43.627$ $p < 0.000$			

[a] % CC is the percentage of correctly classified compounds.

[b] CC is the number of correctly classified compounds.

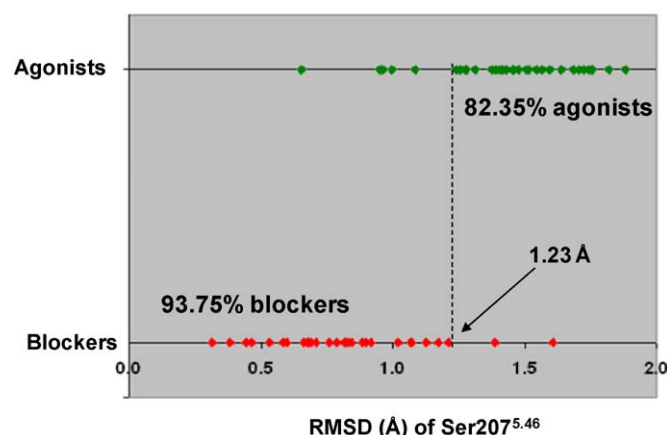[c] MC is the number of misclassified compounds. To test the quality of the derived discriminant function, we used Wilks' statistic (U), Fisher ratio (F) and the significance level (p).

Fig. 4. Plot of the average RMSD induced by the docked ligands to Ser207$^{5.46}$: a demarcation line at 1.23 Å correctly classifies 82.35% of the agonists and 93.75% of the blockers. Each dot represents the average RMSD induced by a particular compound. Agonists, defined as the compounds with an intrinsic activity greater than 15% relatively to isoproterenol, are in green, while blockers, defined as the compounds with an intrinsic activity lower than 15% relatively to isoproterenol, are in red. The misclassified agonists are formoterol, salbutamol, broxaterol, sulfonterol, AH-3021 and T-0509. The misclassified blockers are sotalol and (SRRR)-Nebivolol. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Table 2
LDA models based on the TM rotations measured after Induced Fit Docking. The best model was obtained with the rotation of TM5, with percentages of good classification above 85%.

	% CC[a]	CC[b]	MC[c]
TM2 rotation			
Agonist	47.06	16	18
Blockers	50.00	16	16
Total	48.48	32	34
$U = 0.972\ F(1,64) = 1.873\ p < 0.176$			
TM3 rotation			
Agonist	64.71	22	12
Blockers	68.75	22	10
Total	66.67	44	22
$U = 0.866\ F(1,64) = 9.908\ p < 0.003$			
TM4 rotation			
Agonist	58.82	20	14
Blockers	46.88	15	17
Total	53.03	35	31
$U = 0.999\ F(1,64) = 0.023\ p < 0.881$			
TM5 rotation			
Agonist	85.29	29	5
Blockers	87.50	28	4
Total	86.36	57	9
$U = 0.595\ F(1,64) = 43.552\ p < 0.000$			
TM6 rotation			
Agonist	79.41	27	7
Blockers	59.38	19	13
Total	69.70	46	20
$U = 0.819\ F(1,64) = 14.176\ p < 0.001$			
TM7 rotation			
Agonist	76.47	26	8
Blockers	56.25	18	14
Total	66.67	44	22
$U = 0.925\ F(1,64) = 5.201\ p < 0.026$			

[a] % CC is the percentage of correctly classified compounds.

[b] CC is the number of correctly classified compounds.

[c] MC is the number of misclassified compounds. To test the quality of the derived discriminant function, we used Wilks' statistic (U), Fisher ratio (F) and the significance level (p).
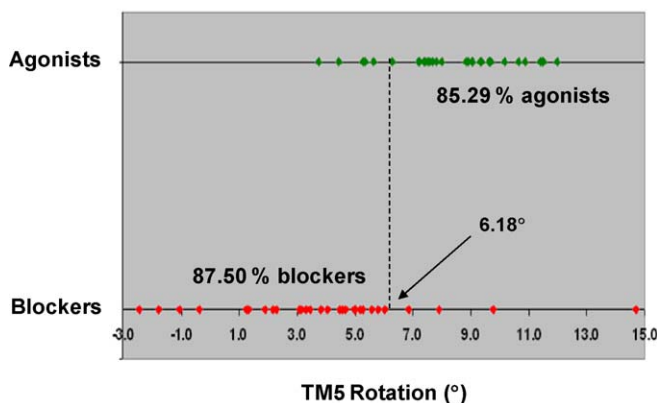
Fig. 5. Plot of the average counterclockwise rotation induced by the docked ligands to TM5: a demarcation line at 6.18° correctly classifies 85.29% of the agonists and 87.50% of the blockers – positives angles indicate counterclockwise rotations, negative angles indicate clockwise rotations. Each dot represents the average RMSD induced by a particular compound. Agonists, defined as the compounds with an intrinsic activity greater than 15% relatively to isoproterenol, are in green, while blockers, defined as the compounds with an intrinsic activity lower than 15% relatively to isoproterenol, are in red. The misclassified agonists are formoterol, salbutamol, sulfonterol, AH-3021 and zinterol. The misclassified blockers are dichloroisoproterenol, H87, xamoterol and (SRRR)-Nebivolol (outlier). Four blockers were predicted to have TM5 clockwise rotation upon the binding (carazolol, propranolol, nadolol and ICI-118551). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

terclockwise rotations smaller than 6.18° – namely formoterol, salbutamol, sulfonterol, AH-3021, and zinterol – and four blockers that induced counterclockwise rotations greater than 6.18° – namely (SRRR)-Nebivolol, dichloroisoproterenol, H87, and xamoterol.

Notably, four compounds endowed with inverse agonistic activity, carazolol, propranolol, nadolol and ICI-118551, were predicted to induce clockwise rotations to TM5 upon binding, indicating a conformational change opposite to the one induced by agonists.

2.3. Molecular dynamics confirms the differential effect of agonists and blockers on TM5 and suggests a correlation between TM5 and EL2 movements

To further explore the effect of agonists and blockers on the conformation of TM5, we embedded the crystal structure of the β_2-AR in a lipid bilayer hydrated model and subjected the whole system to a series molecular dynamics simulations with Desmond [30]. In particular, we simulated the dynamics of the receptor in complex with either the co-crystallized inverse agonist carazolol or the full agonist isoproterenol – the simulations were repeated twice for each system and average results are reported. As indicated by the monitoring of the RMSD of the receptor backbone over the course of the trajectory, each system reached and maintained a stable equilibrium for at least the last 20 ns of the trajectory (Fig. S1 of Supplementary data). As shown in Fig. 6, TM5 underwent greater conformational changes in the simulation of the isoproterenol-bound receptor than in that of the carazolol-bound receptor, thus corroborating the observations derived from our IFD-LDA analysis. Pictures showing receptor–ligand complexes obtained for carazolol and isoproterenol after the energy minimization of the endpoint molecular dynamics structures are given in Fig. 7. Both compounds bridged TM3 and TM5. In particular, they both interacted with Asp113[3.32] through their positively charged amino group and with TM5 through their aromatic moieties. However, the distance between the nitrogen of the positively charged amine and the centroid of the aromatic moiety is generally shorter for agonist than blockers. In particular, in the bound conformation of isoproterenol and carazolol it amounts to 5.1 Å and 6.6 Å, respectively (see Fig. S2 of Supplementary data). Thus, in light of this key difference, the agonist had the effect of pulling TM5 closer to TM3 as a result of the simulation.

It is also evident from Fig. 7 that, during the simulation, isoproterenol dropped more deeply into the binding cavity than carazolol, pulling down with it also the C-terminal portion of EL2. As a result, the simulation of the isoproterenol-bound system produced higher RMSDs not only for TM5 (Fig. 6a) but also for the adjacent EL2 (Fig. 6b), when compared to the simulation of the carazolol-bound system. Thus, to check if the detected EL2 movement was a requisite for that of TM5, we subjected the system to a third molecular dynamics simulation, also in this case performed in duplicate. In this last computational experiment, we simulated again the isoproterenol-bound system; however, this time we applied a harmonic positional restraint to Phe193[EL2.52], which, as outlined by the crystal structures, is the β_2-AR residue that establishes the largest contacts with carazolol. Remarkably, the simulation conducted with the bound isoproterenol, but with a positional restraint applied to Phe193[EL2.52], resulted in significantly lower TM5 movements, giving a response extremely similar to that registered in the simulations with the inverse agonist carazolol (Fig. 6a).

3. Discussion

Our Induced Fit Docking experiments and molecular dynamics simulations results suggest that subtle translations and rotations of
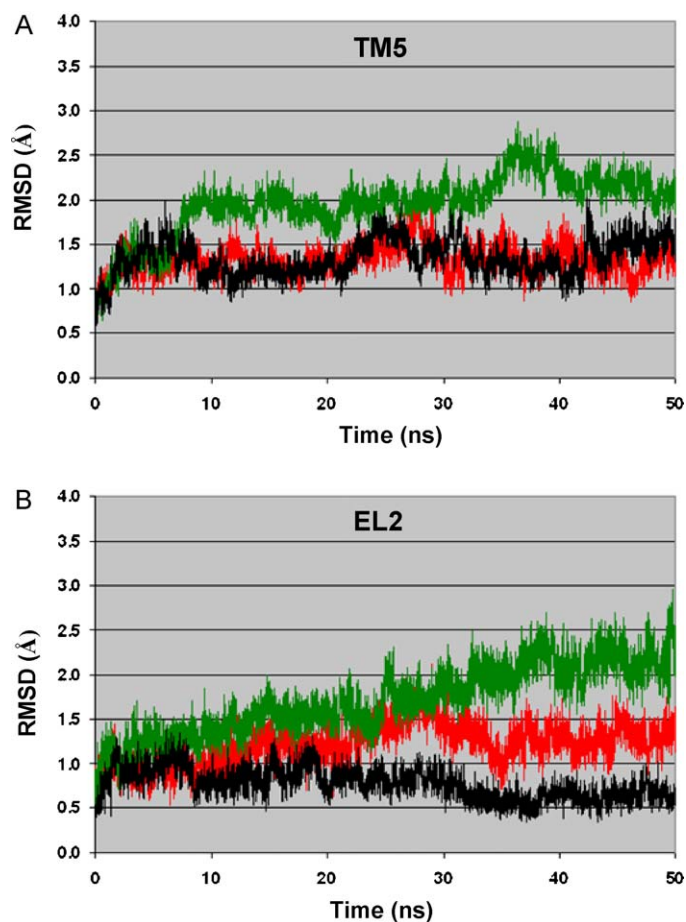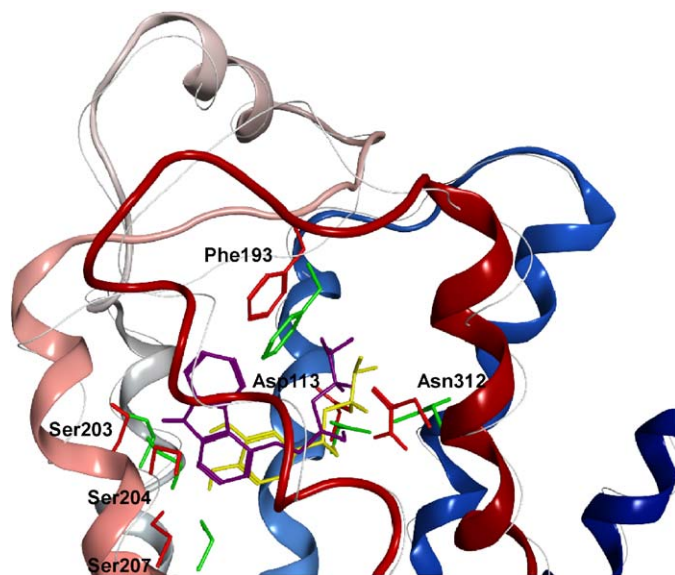
Fig. 6. The evolution of the RMSD of the backbone of TM5 (**a**), from residues Gln197 to Gln229, and EL2 (b), from residues Cys191 to Asn196, over the 50 ns trajectory is shown for each simulation. The red trajectory is relative to the simulation in which the inverse agonist carazolol is bound. The green trajectory is relative to the simulation in which the agonist isoproterenol is bound. The black trajectory is relative to the simulation in which the agonist isoproterenol is bound and a harmonic restraint was applied to Phe193[EL2.52].

Fig. 7. Pictures showing the minimized receptor–ligand complexes for carazolol and isoproterenol after the simulations (50 ns). The β_2-AR–carazolol complex is shown with the residues in red, the ligand in magenta, and the receptor backbone as a thick ribbon colored according to residue position (blue: N-terminus, red: C-terminus). The β_2-AR–isoproterenol complex is shown with the residues in green, the ligand in yellow, and the receptor backbone as a thin white line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

TM5 are an important component of agonist binding. In particular, in our *in silico* studies agonists had the effect of pulling TM5 closer to TM3, due to the shorter distance between positively charged amine and aromatic moiety that characterizes these compounds when compared to blockers.

In line with our results, the crystal structure of the β_2-AR indicates that agonist binding would require a conformational change of TM5 in order to allow the experimentally postulated interactions with residues Ser204[5.43] and Ser207[5.46] [28,29]. Similarly, on the basis of molecular modeling, Goddard and coworkers predicted that a substantial counterclockwise rotation of TM5 upon the binding of an agonist would allow this interaction to occur [31]. Through a systematic mapping of the rotational energy landscape of each of the seven TMs of the β_2-AR, Vaidehi and coworkers also predicted an agonist induced counterclockwise rotation of TM5, viewed from the extracellular side [32,33]. Moreover, Abagyan and coworkers also hypothesized a coupling between the binding of agonists of the β_2-AR and TM5 movements [34].

Our study suggested that EL2 movements are also an important component of agonist binding and are interrelated with TM5 movements. In particular, our molecular dynamics simulations suggested that blocking the freedom of motion of EL2 consequently results in the prevention of agonist-induced TM5 displacements.

In line with our results, the importance of EL2 in GPCR-ligand recognition and receptor activation has been suggested by multiple experimental and computational studies [35–42]. Through molecular dynamics simulations, we previously proposed that agonist binding disrupts a salt bridge that, in the ground state of the P2Y$_6$ receptor, connects an arginine of TM3 to an aspartate of EL2, located at the position corresponding to Phe193[EL2.52] in the β_2-AR [35]. Furthermore, molecular modeling studies of the FFA1 receptor suggested an analogous anchoring of EL2 to the helical bundle in the unliganded state, via a salt bridge between a TM5 arginine and a glutamate residue also located at position EL2.52 [37]. Experimentally, the general role of EL2 in ligand binding was examined in the case of the dopamine D2 receptors, as well [39]. By mutating various EL2 residues to cysteine and examining if disulfide formation inhibited ligand binding, the authors found that the flexibility of the loop is involved in the regulation of ligand activity. Moreover, a recent solid state ^{13}C NMR study on rhodopsin showed a coupling between EL2 and TM5 movement upon retinal isomerization [40]. Finally, Kobilka and coworkers found through NMR spectroscopy that agonists and inverse agonists induced different conformational changes to the extracellular domains of the β_2-AR [43]. In particular, their data suggested that the salt bridge between Lys305[7.32], located at the boundary between TM7 and EL3, and Asp192[EL2.51] is attenuated in the active conformation of the receptor. In light of these data, we monitored the distance between the carboxylate carbon atom of Asp192[EL2.51] and the nitrogen atom of Lys305[7.32] in our molecular dynamics simulations. We detected the disruption of the salt bridge in every simulation. However, in agreement with what was seen by Kobilka and coworkers, we measured higher distances in the simulations with isoproterenol (average: 3.91 Å) compared to those with carazolol (average: 3.56 Å) or with isoproterenol and the positional restraint applied to Phe193[EL2.52] (average: 3.61 Å) – see Supplementary data Fig. S3.

As mentioned, activation is a complex phenomenon, which involves a cascade of events [21,23] and, possibly, even changes in the dimeric or oligomeric state of the receptors [24,25]. With our Induced Fit Docking and our molecular dynamics simulations, we probably have captured only the initial conformational changes

induced to the β_2-AR by the agonist binding. Instead, we did not capture the subsequent conformational changes that lead to the activation of the receptor. The best characterized conformational changes thought to be coupled to β_2-AR activation are a rotamer toggle switch composed of two aromatic TM6 residues, namely Trp286[6.48] and Phe290[6.52], and an ionic lock between Arg131[3.50] in TM3 and Glu268[6.30] in TM6, the disruption of which supposedly causes the two TMs to move further apart, thus opening a G protein docking site [44,45]. We did not detect the movement of Trp286[6.48] as shown by the monitoring of the evolution of the χ_1 dihedral angle of this residue over the course of the 50 ns simulations (Supplementary data Fig. S4). Moreover, the ionic lock between Arg131[3.50] and Glu268[6.30], already broken in the crystal structure, formed only transiently in the simulations with both the agonist and the blocker. We also did not detect the formation of contacts between the ligands and Asn293[6.55], which has been proposed to establish interactions with the β-hydroxyl group of agonists [46]. However, biochemical experiments conducted by Kobilka and coworkers suggest that interactions involving the β-hydroxyl group take place on a slower time scale than other changes [18]. Longer simulations, in the microsecond time scale, analogous to those performed by Dror et al. for the β_2-AR in complex with carazolol [47] have the potential of capturing further stages of the activation process.

4. Conclusions

Induced Fit Docking (IFD) [27] in tandem with linear discriminant analysis (LDA), as well as molecular dynamics simulations, suggested that agonists induce to TM5 larger conformational changes, when compared to blockers. Furthermore, molecular dynamics also highlighted a correlation between movements of TM5 and EL2, suggesting that freedom of motion of EL2 is required for the subtle displacement of TM5 induced by agonists. This study also suggests that Induced Fit Docking in tandem with linear discriminant analysis can be used as a computational means to distinguish agonists from blockers on the basis of differential conformational changes induced to the receptor. In the case of the β_2-AR, the two most predictive models were based on the RMSD induced to Ser207[5.46] (1) and on the counterclockwise rotation induced to TM5 (2).

Importantly, a key factor at the basis of our analysis was the way we defined the flexible region of the receptor, which was characterized by three uninterrupted segments, each of which composed by the exofacial portion of two adjacent TMs and the entire interconnecting extracellular loop (Fig. 2). Due to its simplicity and the rapidity of the procedure, the Induced Fit Docking/LDA analysis is suitable for the study of receptor–ligand interactions and the production of classification models on a large scale.

5. Materials and methods

5.1. GPCR residue identifiers

To facilitate the comparison among receptors, throughout the article, amino acid residues are designated not only with their residue number in the β_2-AR sequence but also with a superscript extension containing the GPCR residue identifier [48,49]. This system was derived through the analysis of a multiple sequence alignment of Class A GPCRs by Ballesteros and Weinstein, who selected a reference position, usually occupied by the most conserved residue, for each TM. By convention, to this reference position is assigned the identifier X.50, where X is the TM number; the remaining residues in the TM are then numbered relatively to the reference. In the β_2-AR, the reference residues are:

Asn51[1.50], Asp79[2.50], Arg131[3.50], Trp158[4.50], Pro211[5.50], Pro288[6.50], Pro323[7.50]. By extension, we assign the identifier EL2.50 to the conserved cysteine (Cys191[EL2.50] in the β_2-AR) that in most GPCRs is disulfide-bridged with Cys[3.25].

5.2. Ligand dataset

The dataset used in this article was composed of 66 β_2-AR ligands, divided into 34 agonists and 32 blockers, which we collected from the literature in a previous work (note that the previous work contained a list of 67 ligands, but we subsequently found that Halostachine and Mape are actually the same compound) [10]. Compounds with intrinsic activity lower than 15%, measured as the maximal cAMP level produced upon ligand binding relative to the full agonist isoproterenol, were considered as blockers (including weak partial agonists, antagonists, and inverse agonists), while all those with intrinsic activity higher than 15% were considered agonists. Ligands were sketched in Maestro and subjected to energy minimization with the MacroModel engine using the OPLS_2005 forcefield and water as implicit solvent (PRCG method, 10,000 maximal iterations, with a convergence gradient of 0.001) [50,51].

5.3. Preparation of crystal structure for docking

The crystal structure of the β_2-AR in complex with carazolol (PDB ID: 2RH1) [28,29] was preprocessed with the Protein Preparation Workflow in the Maestro user interface of the Schrödinger package [52]. This procedure added hydrogens, which were subsequently minimized with the OPLS_2005 force field and the *impact* molecular mechanics engine while heavy atoms were constrained. Furthermore, it optimized the protonation state of His residues and the orientation of hydroxyl groups, Asn residues, and Gln residues. We gave more details on the protein preparation protocol in a previously published paper [53].

5.4. Induced Fit Docking procedure

Ligands docking of was performed with the Induced Fit Docking procedure implemented in the Schrödinger package [27,54,55]. Specifically, the procedure involved three consecutive steps: (a) a Glide-based docking of the ligands at the receptor, with the standard precision (SP) mode and a vdW scaling factor of 0.5 applied to non-polar atoms of receptor and ligands, defined as those with partial charge lower than 0.25e for the receptor and lower than 0.15e for the ligands; (b) a Prime-based optimization of the binding pocket for all the complexes obtained in step (a); and (c) a Glide-based docking of the ligands at each of the optimized binding pockets obtained in step (b), with the extra precision (XP) mode and a vdW scaling factor of 0.8 for the non-polar atoms of the ligands only (only the complexes within 30 kcal/mol from the best structure were subjected to this final docking run). For steps (a) and (c) the docking grid was centered on Val114[3.33], and was given a size of 26 Å. The ligand was treated as completely flexible in all the steps of the procedure. In step (b), flexibility was granted to three uninterrupted receptor segments, each of which was composed of the exofacial half of two adjacent TMs and the entire interconnecting loop (see Fig. 2). For each ligand, the top scoring 20 complexes were collected and used for remainder of the study.

5.5. Pose clustering

For each ligand, the most frequent binding mode was identified through the Molecular Operating Environment (MOE) [56], using a custom made SVL script. The spurious poses that did not belong to the most populated cluster were then eliminated.

5.6. Calculation of RMSDs

For each of the obtained complexes, the RMSD of the coordinates of the backbone of each flexible residue was measured through the "residue rmsd" script of VMD (version 1.8.3) [57], using as a reference the coordinates of the crystal structure. For each ligand, an average RMSD was then calculated for each flexible residue, taking into account all the accepted complexes relative to that particular ligand. These average residue RMSDs were then used for all subsequent statistical analyses.

5.7. Calculation of TM rotations

For each of the obtained complexes, the rotation of the flexible TM segments around their axis was measured through the TRAJELIX [58] program, using as a reference the coordinates of the crystal structure. For each ligand, an average rotation was then calculated for each TM, taking into account all the accepted complexes relative to that particular ligand. These average TM rotations were then used for all subsequent statistical analyses.

5.8. Statistical methods

The classification models were built through linear discriminant analysis (LDA) [26] carried out with the STATISTICA 8.0 package [59]. In the development of the classification functions, the binary values of 1 and 0 were assigned to agonists and blockers, respectively. The forward stepwise algorithm was used for variable selection. The models were based on the posterior classification probability of the cases, and their quality was tested analyzing Wilks' statistic (U), Fisher ratio (F), significance level (p), and percentage of good classification.

5.9. Molecular dynamics

All molecular dynamics simulations were performed with Desmond [30], as implemented in the Schrödinger package. The starting model for the carazolol-bound system was the published crystal structure, while for the isoproterenol-bound system we used a complex obtained docking the ligand to the crystal structure with Glide in SP mode. The complexes were preprocessed with the Protein Preparation Workflow in the Maestro user interface of the Schrödinger package [52], as specified above. In analogy to what described by Dror et al., residue Glu-122[3.41] was protonated prior to the simulation [47]. Besides receptor and ligands, only the palmitoyl group and the crystallographic water molecules were considered, while the other co-crystallized molecules were deleted. The system was embedded in a fully hydrated POPC (1-palmitoyl-2oleoyl-*sn*-glycero-3-phosphatidylchlorine) bilayer. The TIP3 model was used to describe the water molecules. Water molecules within 2 Å from the receptor were deleted, with the exclusion of the crystallographic ones. Counter-ions were added to neutralize the system, replacing solvent molecules with chlorine ions. The final system was composed of approximately 35,000 atoms. Before the simulation, the models were relaxed through the following procedure: (1) minimization of the full system, restraining the solute from moving through the application of a harmonic force of 50 kcal/mol/Å², performed with a maximum 2000 iterations of a hybrid of the steepest descent and the limited memory Broyden–Fletcher–Goldfarb–Shanno (LBFGS) algorithms, with a convergence threshold of 50.0 kcal/mol/Å²; (2) an unrestrained minimization using the same procedure with a convergence threshold of 5.0 kcal/mol/Å²; (3) a simulation in the NVT ensemble using a Berendsen thermostat, restraining all non-hydrogen solute atoms with a force constant of 50 kcal/mol/Å²,

with a simulation time of 12 ps, a time step of 2 fs, a temperature of 10 K, and velocity resampling every 1 ps; (4) a simulation in the NPT ensemble using a Berendsen thermostat, restraining all non-hydrogen solute atoms with a force constant of 50 kcal/mol/Å², with a simulation time of 24 ps, a time step of 2 fs, a temperature of 300 K, and velocity resampling every 1 ps; (5) a simulation in the NPT ensemble, without restraints, using a Berendsen thermostat with a simulation time of 24 ps, a time step of 2 fs, and a temperature of 300 K, and velocity resampling every 1 ps.

After the initial relaxation protocol, an equilibration consisting of four consecutive runs, each of 1 ns, was carried out in the NPT ensemble using a Berendsen thermostat with a temperature of 310 K applying gradually diminishing harmonic restraints: (1) in the first ns, a restraint of 100 kcal/mol/Å² was applied to receptor, ligand, palmitoyl group, and crystallographic water, while a restraint of 10 kcal/mol/Å² was applied to the oxygen atoms of the bulk water molecules; (2) in the second ns, a restraint of 10 kcal/mol/Å² was applied to receptor backbone, ligand, and crystallographic water, while a restraint of 5 kcal/mol/Å² was applied to the side chains of the receptor and the palmitoyl group and a restraint of 1 kcal/mol/Å² was applied to the oxygen atoms of the bulk water molecules; (3) in the third ns, a restraint of 5 kcal/mol/Å² was applied to receptor backbone, ligand and crystallographic water, while a restraint of 2.5 kcal/mol/Å² was applied to the side chains of the receptor and the palmitoyl group and a restraint of 0.5 kcal/mol/Å² was applied to the oxygen atoms of the bulk water molecules; and (4) in the fourth ns, a restraint of 2.5 kcal/mol/Å² was applied to receptor backbone, ligand and crystallographic water, while a restraint of 1 kcal/mol/Å² was applied to the side chains of the receptor and the palmitoyl group and a restraint of 0.1 kcal/mol/Å² was applied to the oxygen atoms of the bulk water molecules.

For the 50 ns of the production runs, a time step of 2 fs was used with the NPT ensemble using a Berendsen thermostat at 310 K and velocity resampling every 1 ps. In one of the simulations, a harmonic restraint of 500 kcal/mol/Å² was applied to all the atoms of Phe193[EL2.52] to prevent its motion. For statistical purposes, each system was simulated twice, using different random number seeds for the generation of the initial velocities.

5.10. Figures

The three-dimensional figures of the receptor and the receptor–ligand complexes were generated with MOE [56].

Note Added after Initial Review of This Paper

While this manuscript was under review, an article was published by Kobilka and coworkers that unveiled the crystal structure of the β₂-AR in its nanobody stabilized active state, in complex with a full agonist [60]. In striking agreement with our predictions, the authors noted that: "The greatest difference between inactive and active structures in the ligand-binding site is an inward bulge of TM5 centred around Ser 207[5.46], whose Cα position shifts by 2.1 Å".

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmgm.2011.01.005.

References

[1] K. Pierce, R. Premont, R. Lefkowitz, Seven-transmembrane receptors, Nat. Rev. Mol. Cell Biol. 3 (2002) 639–650.
[2] M.A. Hanson, R.C. Stevens, Discovery of new GPCR biology: one receptor structure at a time, Structure 17 (2009) 8–14.
[3] S. Costanzi, J. Siegel, I. Tikhonova, K. Jacobson, Rhodopsin and the others: a historical perspective on structural studies of G protein-coupled receptors, Curr. Pharm. Des. 15 (2009) 3994–4002.
[4] B. Wu, E.Y. Chien, C.D. Mol, G. Fenalti, W. Liu, V. Katritch, R. Abagyan, A. Brooun, P. Wells, F.C. Bi, D.J. Hamel, P. Kuhn, T.M. Handel, V. Cherezov, R.C. Stevens, Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists, Science (2010).
[5] E.Y. Chien, W. Liu, Q. Zhao, V. Katritch, G.W. Han, M.A. Hanson, L. Shi, A.H. Newman, J.A. Javitch, V. Cherezov, R.C. Stevens, Structure of the human dopamine D3 receptor in complex with a D2/D3 selective antagonist, Science 330 (2010) 1091–1095.
[6] C. de Graaf, D. Rognan, Selective structure-based virtual screening for full and partial agonists of the beta2 adrenergic receptor, J. Med. Chem. 51 (2008) 4978–4985.
[7] P. Kolb, D. Rosenbaum, J. Irwin, J. Fung, B. Kobilka, B. Shoichet, Structure-based discovery of beta2-adrenergic receptor ligands, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 6843–6848.
[8] V. Katritch, V. Jaakola, J. Lane, J. Lin, A. Ijzerman, M. Yeager, I. Kufareva, R. Stevens, R. Abagyan, Structure-based discovery of novel chemotypes for adenosine A(2A) receptor antagonists, J. Med. Chem. 53 (2010) 1799–1809.
[9] J. Carlsson, L. Yoo, Z. Gao, J. Irwin, B. Shoichet, K. Jacobson, Structure-based discovery of A2A adenosine receptor ligands, J. Med. Chem. 53 (2010) 3748–3755.
[10] S. Vilar, J. Karpiak, S. Costanzi, Ligand and structure-based models for the prediction of ligand-receptor affinities and virtual screenings: development and application to the beta(2)-adrenergic receptor, J. Comput. Chem. 31 (2010) 707–720.
[11] K. Reynolds, V. Katritch, R. Abagyan, Identifying conformational changes of the beta(2) adrenoceptor that enable accurate prediction of ligand/receptor interactions and screening for GPCR modulators, J. Comput. Aided Mol. Des. 23 (2009) 273–288.
[12] V. Katritch, M. Rueda, P. Lam, M. Yeager, R. Abagyan, GPCR 3D homology models for ligand screening: lessons learned from blind predictions of adenosine A2a receptor complex, Proteins 78 (2010) 197–211.
[13] S. Vilar, G. Ferino, S.S. Phatak, B. Berk, C.N. Cavasotto, S. Costanzi, Docking-based virtual screening for ligands of G protein-coupled receptors: not only crystal structures but also in silico models, J. Mol. Graph. Model. (2010).
[14] S.S. Phatak, E.A. Gatica, C.N. Cavasotto, Ligand-steered modeling docking: a benchmarking study in class A G-protein-coupled receptors, J. Chem. Inf. Model. (2010).
[15] S. Costanzi, On the applicability of GPCR homology models to computer-aided drug discovery: a comparison between in silico and crystal structures of the beta2-adrenergic receptor, J. Med. Chem. 51 (2008) 2907–2914.
[16] M. Michino, E. Abola, G. 2008 Participants, C.R. Brooks, J. Dixon, J. Moult, R. Stevens, Community-wide assessment of GPCR structure modelling and ligand docking: GPCR Dock 2008, Nat. Rev. Drug Discov. 8 (2009) 455–463.
[17] S. Costanzi, Modelling G protein-coupled receptors: a concrete possibility, Chim. Oggi 28 (2010) 26–30.
[18] G. Swaminath, Y. Xiang, T.W. Lee, J. Steenhuis, C. Parnot, B.K. Kobilka, Sequential binding of agonists to the beta(2) adrenoceptor – kinetic evidence for intermediate conformational states, J. Biol. Chem. 279 (2004) 686–691.
[19] J. Bockaert, L. Fagni, A. Dumuis, P. Marin, GPCR interacting proteins (GIP), Pharmacol. Therap. 103 (2004) 203–221.
[20] X.J. Yao, C. Parnot, X. Deupi, V.R.P. Ratnala, G. Swaminath, D. Farrens, B. Kobilka, Coupling ligand structure to specific conformational switches in the beta(2)-adrenoceptor, Nat. Chem. Biol. 2 (2006) 417–422.
[21] B.K. Kobilka, X. Deupi, Conformational complexity of G-protein-coupled receptors, Trends Pharmacol. Sci. 28 (2007) 397–406.
[22] S. Galandrin, G. Oligny-Longpre, M. Bouvier, The evasive nature of drug efficacy: implications for drug discovery, Trends Pharmacol. Sci. 28 (2007) 423–430.
[23] D.M. Rosenbaum, S.G. Rasmussen, B.K. Kobilka, The structure and function of G-protein-coupled receptors, Nature 459 (2009) 356–363.
[24] W. Guo, L. Shi, M. Filizola, H. Weinstein, J. Javitch, Crosstalk in G protein-coupled receptors: changes at the transmembrane homodimer interface determine activation, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 17495–17500.
[25] W. Guo, E. Urizar, M. Kralikova, J. Mobarec, L. Shi, M. Filizola, J. Javitch, Dopamine D2 receptors form higher order oligomers at physiological expression levels, EMBO J. 27 (2008) 2293–2304.
[26] T. Hill, P. Lewicki, Statistical Methods and Applications. A Comprehensive Reference for Science, Industry and Data Mining, StatSoft, Tulsa, 2006.
[27] W. Sherman, T. Day, M.P. Jacobson, R.A. Friesner, R. Farid, Novel procedure for modeling ligand/receptor induced fit effects, J. Med. Chem. 49 (2006) 534–553.
[28] V. Cherezov, D. Rosenbaum, M. Hanson, S. Rasmussen, F. Thian, T. Kobilka, H. Choi, P. Kuhn, W. Weis, B. Kobilka, R. Stevens, High-resolution crystal structure of an engineered human beta2-adrenergic G protein-coupled receptor, Science 318 (2007) 1258–1265.
[29] D. Rosenbaum, V. Cherezov, M. Hanson, S. Rasmussen, F. Thian, T. Kobilka, H. Choi, X. Yao, W. Weis, R. Stevens, B. Kobilka, GPCR engineering yields high-resolution structural insights into beta2-adrenergic receptor function, Science 318 (2007) 1266–1273.
[30] Desmond Molecular Dynamics System, 2.0109, D. E. Shaw Research, New York, NY, 2008; Maestro-Desmond Interoperability Tools, 2.0109, Schrodinger, New York, NY, 2008.
[31] P.L. Freddolino, M.Y.S. Kalani, N. Vaidehi, W.B. Floriano, S.E. Hall, R.J. Trabanino, V.W.T. Kam, W.A. Goddard, Predicted 3D structure for the human beta 2 adrenergic receptor and its binding site for agonists and antagonists, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 2736–2741.
[32] S. Bhattacharya, S.E. Hall, H. Li, N. Vaidehi, Ligand-stabilized conformational states of human beta(2) adrenergic receptor: insight into G-protein-coupled receptor activation, Biophys. J. 94 (2008) 2027–2042.
[33] S. Bhattacharya, N. Vaidehi, Computational mapping of the conformational transitions in agonist selective pathways of a G-protein coupled receptor, J. Am. Chem. Soc. 132 (2010) 5205–5214.
[34] V. Katritch, K.A. Reynolds, V. Cherezov, M.A. Hanson, C.B. Roth, M. Yeager, R. Abagyan, Analysis of full and partial agonists binding to beta(2)-adrenergic receptor suggests a role of transmembrane helix V in agonist-specific conformational changes, J. Mol. Recognit. 22 (2009) 307–318.
[35] S. Costanzi, B.V. Joshi, S. Maddileti, L. Mamedova, M.J. Gonzalez-Moa, V.E. Marquez, T.K. Harden, K.A. Jacobson, Human P2Y(6) receptor: molecular modeling leads to the rational design of a novel agonist based on a unique conformational preference, J. Med. Chem. 48 (2005) 8108–8111.
[36] J.H. Kim, Q.L. Jiang, M. Glashofer, S. Yehle, J. Wess, K.A. Jacobson, Glutamate residues in the second extracellular loop of the human A(2a) adenosine receptor are required for ligand recognition, Mol. Pharmacol. 49 (1996) 683–691.
[37] I.G. Tikhonova, C.S. Sum, S. Neumann, C.J. Thomas, B.M. Raaka, S. Costanzi, M.C. Gershengorn, Bidirectional, iterative approach to the structural delineation of the functional "Chemoprint" in GPR40 for agonist recognition, J. Med. Chem. 50 (2007) 2981–2989.
[38] G Kleinau, M. Claus, H. Jaeschke, S. Mueller, S. Neumann, R. Paschke, G. Krause, Contacts between extracellular loop two and transmembrane helix six determine basal activity of the thyroid-stimulating hormone receptor, J. Biol. Chem. 282 (2007) 518–525.
[39] L. Shi, J.A. Javitch, The second extracellular loop of the dopamine D-2 receptor lines the binding-site crevice, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 440–445.
[40] S. Ahuja, V. Hornak, E.C.Y. Yan, N. Syrett, J.A. Goncalves, A. Hirshfeld, M. Ziliox, T.P. Sakmar, M. Sheves, P.J. Reeves, S.O. Smith, M. Eilers, Helix movement is coupled to displacement of the second extracellular loop in rhodopsin activation, Nat. Struct. Mol. Biol. 16 (2009) 168–175.
[41] M. Olah, K. Jacobson, G. Stiles, Role of the second extracellular loop of adenosine receptors in agonist and antagonist binding. Analysis of chimeric A1/A3 adenosine receptors, J. Biol. Chem. 269 (1994) 24692–24698.
[42] G. Kleinau, M. Claus, H. Jaeschke, S. Mueller, S. Neumann, R. Paschke, G. Krause, Contacts between extracellular loop two and transmembrane helix six determine basal activity of the thyroid-stimulating hormone receptor, J. Biol. Chem. 282 (2007) 518–525.
[43] M.P. Bokoch, Y.Z. Zou, S.G.F. Rasmussen, C.W. Liu, R. Nygaard, D.M. Rosenbaum, J.J. Fung, H.J. Choi, F.S. Thian, T.S. Kobilka, J.D. Puglisi, W.I. Weis, L. Pardo, R.S. Prosser, L. Mueller, B.K. Kobilka, Ligand-specific regulation of the extracellular surface of a G-protein-coupled receptor, Nature 463 (2010) U108–U121.
[44] L. Shi, G. Liapakis, R. Xu, F. Guarnieri, J.A. Ballesteros, J.A. Javitch, Beta2 adrenergic receptor activation. Modulation of the proline kink in transmembrane 6 by a rotamer toggle switch, J. Biol. Chem. 277 (2002) 40989–40996.
[45] J.A. Ballesteros, A.D. Jensen, G. Liapakis, S.G.F. Rasmussen, L. Shi, U. Gether, J.A. Javitch, Activation of the beta(2)-adrenergic receptor involves disruption of an ionic lock between the cytoplasmic ends of transmembrane segments 3 and 6, J. Biol. Chem. 276 (2001) 29171–29177.
[46] K. Wieland, H.M. Zuurmond, C. Krasel, A.P. Ijzerman, M.J. Lohse, Involvement of Asn-293 in stereospecific agonist recognition and in activation of the beta(2)-adrenergic receptor, Proc. Natl. Acad. Sci. U. S. A. 93 (1996) 9276–9281.
[47] R.O. Dror, D.H. Arlow, D.W. Borhani, M.O. Jensen, S. Piana, D.E. Shaw, Identification of two distinct inactive conformations of the beta(2)-adrenergic receptor reconciles structural and biochemical observations, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 4689–4694.
[48] J.A. Ballesteros, H. Weinstein, C.S. Stuart, Integrated methods for the construction of three-dimensional models and computational probing of structure–function relations in G-protein coupled receptors, Methods Neurosci. 25 (1995) 366–428.
[49] A.M. van Rhee, K.A. Jacobson, Molecular architecture of G protein-coupled receptors, Drug Develop. Res. 37 (1996) 1–38.
[50] F. Mohamadi, N.G.J. Richards, W.C. Guida, R. Liskamp, M. Lipton, C. Caufield, G. Chang, T. Hendrickson, W.C. Still, Macromodel – an integrated software system for modeling organic and bioorganic molecules using molecular mechanics, J. Comput. Chem. 11 (1990) 440–467.
[51] MacroModel, 9.8, Schrodinger, LLC, New York, NY, USA. www.schrodinger.com.
[52] Maestro, 8.5, Schrodinger, LLC, New York, NY, USA. www.schrodinger.com.
[53] S. Costanzi, On the applicability of GPCR homology models to computer-aided drug discovery: a comparison between in silico and crystal structures of the beta(2)-adrenergic receptor, J. Med. Chem. 51 (2008) 2907–2914.
[54] Glide, 5.0, Schrodinger, LLC, New York, NY, USA. www.schrodinger.com.
[55] Prime, 2.0, Schrodinger, LLC, New York, NY, USA. www.schrodinger.com.

[56] The molecular operating environment (MOE), 2009.10, Chemical Computing Group Inc., Montreal, Canada. www.chemcomp.com.

[57] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J. Mol. Graph. 14 (1996) 33–38.

[58] M. Mezei, M. Filizola, TRAJELIX: a computational tool for the geometric characterization of protein helices during molecular dynamics simulations, J. Comput. Aided Mol. Des. 20 (2006) 97–107.

[59] Statistica, 8.0, StatSoft Inc., Tulsa, OK, USA. www.statsoft.com.

[60] S.G. Rasmussen, H.J. Choi, J.J. Fung, E. Pardon, P. Casarosa, P.S. Chae, B.T. Devree, D.M. Rosenbaum, F.S. Thian, T.S. Kobilka, A. Schnapp, I. Konetzki, R.K. Sunahara, S.H. Gellman, A. Pautsch, J. Steyaert, W.I. Weis, B.K. Kobilka, Structure of a nanobody-stabilized active state of the beta(2) adrenoceptor, Nature 469 (2011) 175–180.