Journal of Molecular Graphics and Modelling 29 (2011) 809–817

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

In silico analysis of the binding of agonists and blockers to the β_2-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

ARTICLE INFO

Article history:
Received 22 December 2010
Accepted 12 January 2011
Available online 19 January 2011

Keywords:
G protein-coupled receptor
Agonists
Blockers
Induced Fit
Docking
Linear discriminant analysis
Molecular dynamics

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 β_2-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 β_2-adrenergic receptor (β_2-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 β_2-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).

1093-3263/$ – see front matter. Published by Elsevier Inc.
doi:10.1016/j.jmgm.2011.01.005

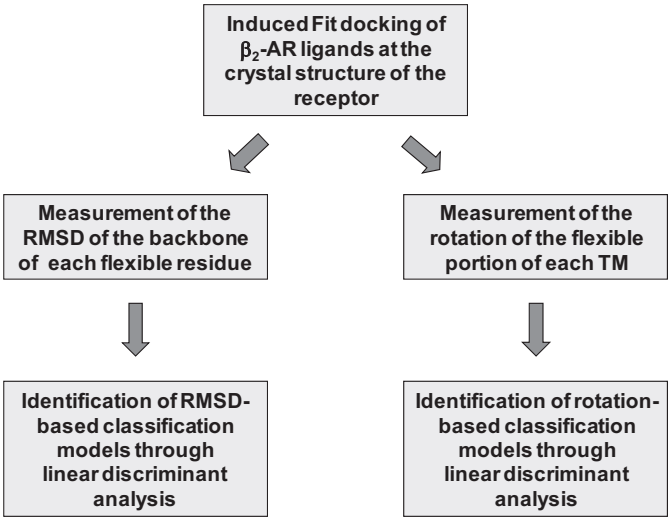

Fig. 1. Flowchart showing the different phases implicated in the development of the IFD/LDA models.

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.

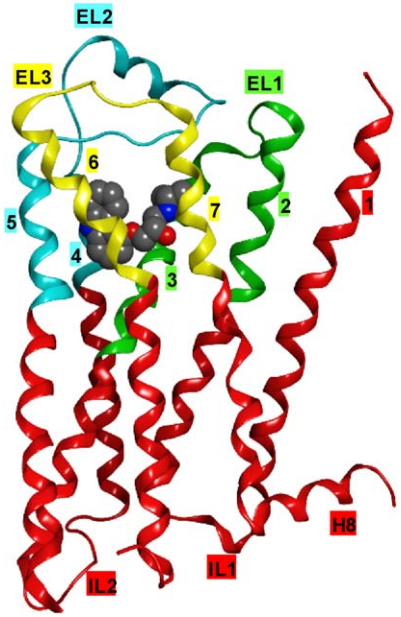

Fig. 2. Ribbon representation of the β_2-AR (PDB ID: 2RH1) indicating the residues treated as flexible in the IFD procedure. Segment 1 (from Met82 to Ile121) is colored in green, segment 2 (from Thr164 to Tyr209) in cyan and segment 3 (from Trp286 to Tyr316) in yellow. The co-crystallized carazolol is 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