

How To Minimize Artifacts in Atomistic Simulations of Membrane Proteins, Whose Crystal Structure Is Heavily Engineered: β_2 -Adrenergic Receptor in the Spotlight

Moutusi Manna,[†] Waldemar Kulig,[†] Matti Javanainen,[†] Joona Tynkkynen,[†] Ulf Hensen,[‡] Daniel J. Müller,[‡] Tomasz Rog,[†] and Ilpo Vattulainen^{*,†,§}

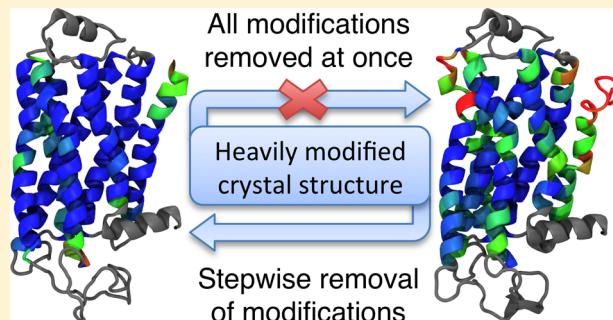
[†]Department of Physics, Tampere University of Technology, P.O. Box 692, FI-33101 Tampere, Finland

[‡]Department of Biosystems Science and Engineering (D-BSSE), ETH-Zürich, 4058 Basel, Switzerland

[§]MEMPHYS-Center for Biomembrane Physics, University of Southern Denmark, Odense, Denmark

Supporting Information

ABSTRACT: Atomistic molecular dynamics (MD) simulations are used extensively to elucidate membrane protein properties. These simulations are based on three-dimensional protein structures that in turn are often based on crystallography. The protein structures resolved in crystallographic studies typically do not correspond to pristine proteins, however. Instead the crystallized proteins are commonly engineered, including structural modifications (mutations, replacement of protein sequences by antibodies, bound ligands, etc.) whose impact on protein structure and dynamics is largely unknown. Here we explore this issue through atomistic MD simulations ($\sim 5 \mu\text{s}$ in total), focusing on the β_2 -adrenergic receptor ($\beta_2\text{AR}$) that is one of the most studied members of the G-protein coupled receptor superfamily. Starting from an inactive-state crystal structure of $\beta_2\text{AR}$, we remove the many modifications in $\beta_2\text{AR}$ systematically one at a time, in six consecutive steps. After each step, we equilibrate the system and simulate it quite extensively. The results of this step-by-step approach highlight that the structural modifications used in crystallization can affect ligand and G-protein binding sites, packing at the transmembrane-helix interface region, and the dynamics of connecting loops in $\beta_2\text{AR}$. When the results of the systematic step-by-step approach are compared to an all-at-once technique where all modifications done on $\beta_2\text{AR}$ are removed instantaneously at the same time, it turns out that the step-by-step method provides results that are superior in terms of maintaining protein structural stability. The results provide compelling evidence that for membrane proteins whose 3D structure is based on structural engineering, the preparation of protein structure for atomistic MD simulations is a delicate and sensitive process. The results show that most valid results are found when the structural modifications are reverted slowly, one at a time.



INTRODUCTION

G-protein coupled receptors (GPCRs) constitute one of the largest families of plasma membrane receptors.^{1,2} They are major contributors to the cellular signal transduction and respond to a wide variety of extracellular stimuli such as light, taste, odor, peptides, neurotransmitters, and hormones. During extracellular signaling, the ligand-induced change in receptor conformation propagates through the receptor's transmembrane region to its cytoplasmic domain. An activated GPCR then couples with its intracellular G-protein (guanine nucleotide-binding protein) to transmit a biochemical signal to the cytosolic side. Importantly, as GPCR signaling regulates major physiological processes, GPCRs are associated with a number of diseases and have therefore become the key target for drug discovery.¹

Although the growing numbers of available crystal structures have significantly advanced our understanding of GPCRs, these static structures are insufficient to describe the dynamic features

that govern GPCR function. Moreover, a matter of concern is the fact that the crystal structures are mostly resolved in non-native detergent media³ and are often highly engineered to overcome the inherent conformational flexibility of receptors.^{4,5} The concept of engineering here refers to structural modifications often used to foster GPCR crystallization.^{4,5} Commonly used techniques include, e.g., introduction of stabilizing mutations to the protein structure, ways to bind antibody fragments, or fusion of T4-lysozyme to flexible dynamic loop regions. Given that such engineering is a common strategy in membrane protein structure determination, there is a growing need to understand the impact of such structural modifications on the relationship between receptor structure and function.⁶

Received: January 26, 2015

Currently, the understanding of potential issues arising from structural modifications in protein structure determination is very limited. This largely stems from inherent difficulties to characterize receptor behavior over nanoscales, where the resolution of most experimental techniques is no longer sufficient to gauge atom-scale changes in protein structure and dynamics. The method of choice to unravel this issue is atomistic molecular dynamics (MD) simulations. Atom-scale simulations have become highly powerful techniques to complement experiments in studies of protein complexes. Recent progress in the field has rendered studies of highly complex proteins possible over time scales where even slow changes in conformational states can be observed.⁷ Together with homology modeling, atomistic MD simulations can contribute to structure prediction, and more generally they can provide substantial insight into rational drug design and receptors' conformational dynamics.^{7–12} As more crystal structures have become available, computational efforts to explore GPCRs and clarify their properties in atomistic detail have become common.

Here we use atomistic MD simulations to assess the impact of structural modifications used in protein structure determination on the structure, structural stability, and dynamics of the β_2 -adrenergic receptor (β_2 AR), which is one of the most studied members of the GPCR superfamily. β_2 AR is expressed in pulmonary and cardiac myocyte tissues and is a therapeutic target for asthma and heart failure.² GPCRs, including β_2 AR, have attracted considerable attention since a great fraction of drug development uses GPCRs as the main target.

To elucidate the effects of structural modifications, we study β_2 AR through two complementary approaches. In the first approach (Method A, "A" standing for "all"), starting from the engineered crystal structure of β_2 AR, we remove all of the numerous structural modifications at the same time, thereby reverting the receptor back to its pristine structure, and then explore the protein behavior in long MD simulations. In the second approach (Method S, "S" standing for "stepwise"), again starting from the engineered crystal structure of β_2 AR, the structural modifications used in experimental structure determination are removed in six successive steps, one at a time. After each of the steps in Method S, β_2 AR is simulated quite extensively to clarify the impact of every structural modification separately. In the end of the procedure, we again end up with the pristine structure where all modifications are removed, and the final model of β_2 AR is simulated to compare its behavior with that of Method A and with the structure predicted by experimental structure determination.¹³

Our results show that, although the stepwise approach (Method S) is time-consuming and computationally more expensive than the more straightforward approach of reverting all structural modifications at once (Method A), Method S produces superior results in terms of structural protein stability. The take-home message of the present work is that the preparation of membrane protein structure for atomistic MD simulations is a very delicate and sensitive process. Great care is required in order to avoid possibly significant artifacts that may arise if the engineered protein structure is reverted for atomistic simulations without a sufficiently rigorous process.

The results of the step-by-step modifications in Method S also provide relevant insight into the effects of crystallization modifications on the intrinsic dynamics of the apo-receptor: this approach shows how structural modifications used for crystallization can affect ligand and G-protein binding sites,

packing at the transmembrane helix interface, and the dynamics of connecting loops.

METHODS AND MODELS

Experimentally Determined Structure of β_2 AR. Due to recent breakthroughs in development of crystallization techniques, structures of the inverse agonist-bound β_2 AR were resolved recently.^{13–15} These structures exhibit the common architecture of GPCRs: seven membrane-spanning helices connected to each other by intracellular and extracellular loops (see the cartoon diagram of β_2 AR in Figure S1 (Supporting Information (SI)). Although β_2 AR and other class-A GPCRs such as rhodopsin share structural homology, they have noticeable differences, too. For instance, as compared to inactive dark-state rhodopsin, inactive β_2 AR exhibits a relatively open structure due to a weaker interaction between the cytoplasmic ends of the transmembrane (TM) helices 3 and 6 and a broken salt-bridge in the highly conserved (E/D)RY motif.¹⁴ These differences may account for the higher basal activity and structural instability of β_2 AR. In 2011, agonist-bound crystal structures of β_2 AR were published: here, the active conformation of the receptor was stabilized either by the intercellular partner G-protein or a G-protein mimetic nano body.^{16,17} The subtle agonist-induced rearrangement in the ligand-binding pocket of β_2 AR move TM5 around Ser207 inward by about 2 Å. This conformational change at the extracellular side leads to a much larger conformational change at the distal G-protein binding site, including an outward 14 Å shift of the cytoplasmic end of TM6. These differences provide the structural basis of the allosteric mechanism of the receptor activation.⁴

In this study, we concentrate on the inactive form of β_2 AR [PDB id: 3D4S], whose crystal structure does not correspond to the pristine receptor.¹³ Rather, the crystal structure is for a receptor that is quite heavily engineered. First, the receptor is cocrystallized with the partially inverse agonist timolol. Second, it has several structural modifications such as a missing intracellular loop 3, a fused T4-lysozyme between TM5 and TM6, and multiple mutations.¹³ Here using atomistic MD simulations for this receptor structure, we explore what are the most efficient and valid ways of removing the structural modifications used in crystallization.

In the first approach (Method A), we simultaneously remove all the changes made for crystallization and then allow the receptor (in its apo-form) to evolve in a lipid bilayer environment. In the second approach (Method S), we first equilibrate the crystal structure in the membrane environment and then remove the many structural modifications one-by-one in a number of successive steps, each followed by hundreds of nanoseconds for equilibration. The findings of the present work are compared with a wealth of available experimental data.

Lipid Membrane Structure. A lipid bilayer in a fluid (liquid-disordered) state, used in the present studies, was initially comprised of 288 DOPC (1,2-dioleoyl-*sn*-glycero-3-phosphocholine) molecules. The lipid membrane structure was equilibrated for 100 ns before embedding the β_2 AR; for details with regard to the force field and the simulations, see the Discussion below. After incorporation of the receptor (see below) and removal of overlapping lipids, the system contained 202 DOPC lipid molecules.

To further validate the β_2 AR model, we performed an additional control simulation where β_2 AR was placed in a bilayer environment containing about 10 mol % cholesterol.

This choice was based on the conditions used in ref 13 to determine the structure of β_2 AR: the study used a ratio of 10% w/w cholesterol/monoolein lipid in the cubic phase, and for thermal stability assays the study used various different concentrations of a cholesterol analogue called cholestryl hemisuccinate (CHS). Meanwhile, it is known that cholesterol modulates membrane physical properties and is required for the proper functioning of many membrane proteins,^{18–21} and it has been suggested or shown to influence the functional state and thus the structural properties of β_2 AR.^{13,15,22,23} Therefore, to account for these features, we carried out a control simulation with a membrane comprised of 304 DOPC and 34 cholesterol molecules that was pre-equilibrated for 100 ns before embedding the protein into the lipid bilayer.

For incorporating the protein into a mixed lipid bilayer, we followed our recently published method, where the idea is to push proteins into a lipid membrane from its side by applying a high lateral pressure on the system.²⁴

Protein Structure: Description of Structural Modifications Made in Experiments for Crystallization. The initial coordinates of human β_2 AR (residues 32–342) that we used in the simulations were obtained from the PDB id 3D4S.¹³ All nonprotein molecules except for the partially inverse agonist timolol and crystallographic water molecules were disregarded from the structure. The X-ray crystal structure¹³ considered here is a complex comprised of β_2 AR together with a T4-lysozyme (T4L): for structure determination, the residues 231–262 of the intracellular loop 3 (ICL3) had been substituted by a T4-lysozyme (T4L). The β_2 AR-T4L complex whose structure had been resolved consisted of seven transmembrane helices (TMs) (TM1 – residues 32–59; TM2 – residues 67–96; TM3 – residues 103–136; TM4 – residues 147–171; TM5 – residues 197–229; TM6 – residues 267–298; and TM7 – residues 305–327), three extracellular loops (ECLs) (ECL1 – residues 97–102; ECL2 – 172–196; and ECL3 – 299–304), three intracellular loops (ICLs) (ICL1 – residues 60–66; ICL2 – residues 137–146; and ICL3 replaced by T4L discussed above), and a short C-terminal helix H8 (residues 330–340) parallel to a membrane surface (Figure S1). The same definitions of the protein segments will be used throughout the manuscript. The receptor's crystal structure has two mutations: (i) E122^{3.41}W on TM3, which in experiments was introduced to enhance receptor stability, and (ii) a N187E mutation on ECL2 introduced in experiments to avoid glycosylation. The superscripts used in describing β_2 AR residues refer to the Ballesteros–Weinstein residue numbering,²⁵ where the first digit ranging between 1 and 7 refers to the transmembrane helix number at which that particular amino acid is located, and the second number after a decimal indicates its position with regard to the most conserved residue of that helix, conventionally assigned to a value of 50. This numbering scheme is used here, too. The following section describes the methods we used to remove the structural modifications used in crystallization in order to model the apo-receptor in its native state.

Modeling Intracellular Loop 3. For ICL3, the sequence of β_2 AR was taken from the UniProt database²⁶ with the access number P07550. We first built an initial conformation of the missing loop attached to the receptor using the MODELER 9.11 program²⁷ and then refined the loop using the same software. ICL3 was modeled as an unstructured loop considering that its sequence shows evidence for intrinsic disorder.²⁸ Out of ~500 models generated, assessment was

made based on DOPE (Discrete Optimized Potential Energy) scoring of MODELER and also based on the loop's structure and its orientation toward the membrane.

Force Fields and Simulation Details. To parametrize the interactions in the simulation systems, the all-atom optimized potentials for liquid simulation (OPLS) force field^{29,30} was used in conjunction with the recently refined parameters for lipid molecules.³¹ The TIP3P model, which is compatible with the OPLS parametrization, was used for water molecules.³² Parameters for timolol were built from the standard OPLS force field, while partial charges were derived using RESP³³ fitting to the electrostatic potential according to the OPLS methodology. All simulations were performed with the GROMACS 4.6.0 software package.³⁴ Simulations were done in the isobaric–isothermal (NpT) ensemble using a time step of 2 fs and 3D periodic boundary conditions. The v-rescale thermostat³⁵ with a time constant of 0.1 ps was employed to maintain the temperature at 310 K. The temperatures of the solute and the solvent were controlled independently. The pressure for the system (1 bar) was controlled semi-isotropically using the Parrinello–Rahman barostat³⁶ with a 1 ps time constant. The LINCS algorithm³⁷ was applied to preserve hydrogen covalent bond lengths. The Lennard–Jones interactions were truncated at a cutoff distance of 1.0 nm. For the long-range electrostatic interactions, the particle mesh Ewald method³⁸ was employed with a real space cutoff of 1.0 nm, β -spline interpolation (order of 6), and a direct sum tolerance of 10^{-6} . In analysis, error bars were estimated through standard error, calculated by dividing the standard deviation of a given data set with the square root of its sample size.²³ We used the g_analyze tool of GROMACS for error estimation.

Methods Investigated To Remove Structural Modifications Used in Experimental Structure Determination. We used two approaches to remove the modifications. These approaches were then used in separate MD simulations, and their results were compared to one another to find how significantly the deformation of the protein structure (compared to experimental structure) depends on the approach used.

Method A: All Changes Made at Once. In this procedure, starting from the engineered crystal structure of the receptor, we applied the commonly used approach, that is, we removed all the structural modifications at the same time, before placing the apo-receptor in a membrane for equilibration. Therefore, we removed the ligand (timolol), removed the T4-lysozyme, and replaced it with the missing ICL3 and removed mutations. We did not attempt to model the unresolved N-terminal (32 residues) and C-terminal (71 residues) parts. After all changes, the receptor was embedded into a pre-equilibrated DOPC bilayer. Subsequently, the system was solvated with 34,516 water molecules. 94 sodium and 102 chloride ions were added to neutralize the net charge of the system and to create a 150 mM NaCl solution. After energy minimization, we simulated the system for 25 ns with position restraints on protein heavy atoms, letting the other components equilibrate first. This was followed by another equilibration period of 25 ns with position restraints focused only on the protein backbone. After simulating for 50 ns, all restraints were released, and the equilibrated system was subjected to a production simulation over a time scale of 1000 ns (referred to as the “Long” simulation). Additionally, in the analysis we also used the trajectories of three shorter (independent) simulations that covered a scale of 200 ns each. The first shorter trajectory

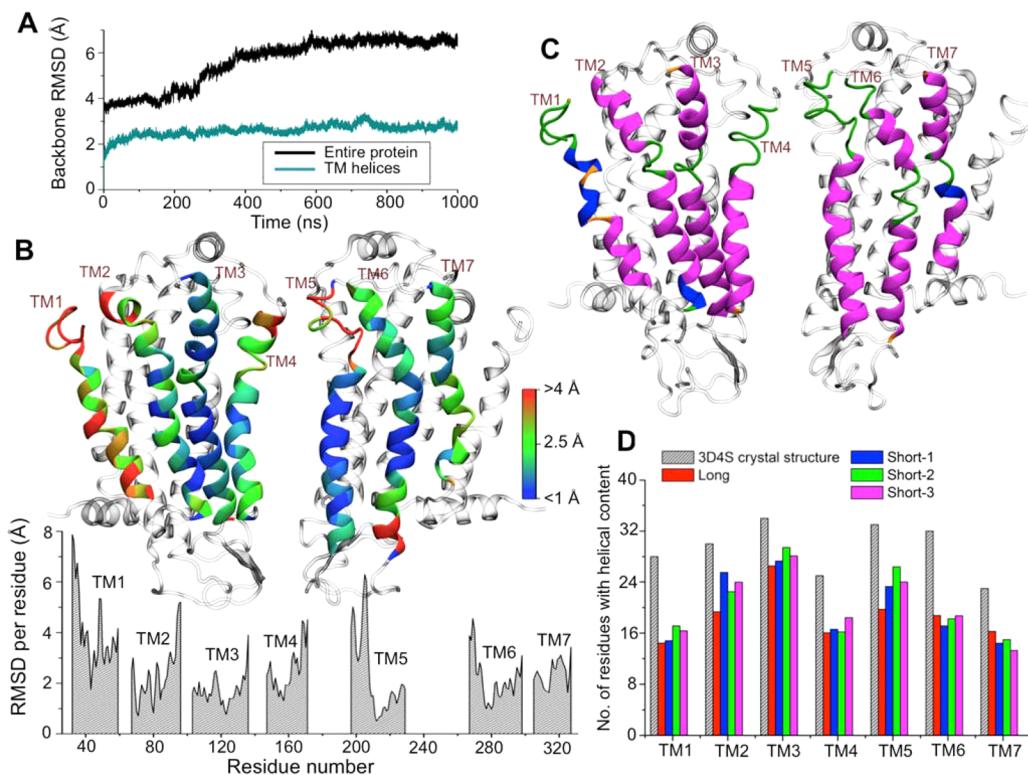


Figure 1. Properties of the apo-receptor model generated by Method A. **A)** Root-mean-square deviation (RMSD) of the protein backbone: entire β_2 AR (black) and the transmembrane region (cyan) (Long simulation). **B)** Per residue RMSD (average over the last 100 ns) of the simulated receptor model compared to the crystal structure, calculated over the backbone atoms of the transmembrane region (Long simulation). The snapshots show the front (TM1-TM4) and back (TM5-TM7) views of the protein's average structure (the last 100 ns), with TM residues colored according to their respective RSMD values and the rest of the protein represented as a transparent gray cartoon. **C)** A representative snapshot of the receptor (from the Long simulation at 900 ns) with TM residues colored according to the secondary structure content (α -helix in magenta, 3₁₀-helix in blue, turn in green, coil in orange): the front (TM1-TM4) and back (TM5-TM7) sides. **D)** The number of residues with helical content (sum of α -, 3₁₀-, and Π -helices) for each TM helix of β_2 AR obtained from different Method A trajectories (averaged over the last 100 ns): Long (red), Short-1 (blue), Short-2 (green), and Short-3 (magenta). These simulation results were compared to that found in the crystal structure 3D4S (gray). The error in the number of residues with helical content was below 0.04.

(referred to as the “Short-1” simulation) covered the first 200 ns of the Long simulation. Then Short-2 and Short-3 were based on independent simulations of the same system and lasted also for 200 ns.

Method S: Stepwise Approach. In the second procedure, the structural modifications used in experimental structure determination were removed in a number of successive steps, one at a time. Each step was followed by simulation of the receptor in a DOPC bilayer. The final structure of the receptor obtained after each step was used as a starting point for the immediate next step.

Below, “EM-PS” refers to energy minimization, followed by equilibration for 50 ns with restraints acting on the protein (using a procedure as described above) followed by a 200 ns production simulation.

- **Step 1.** The β_2 AR-T4L complex, together with its bound ligand and mutations, was placed into a DOPC bilayer. 34496 water molecules were added to the system. The β_2 AR-T4L complex prior to any modifications had a charge of +12. 94 sodium, and 106 chloride ions were added to neutralize the charge of the system and to maintain the physiological-like salt concentration. EM-PS was carried out.

- **Step 2.** The partially inverse agonist timolol was removed from the extracellular ligand-binding cavity of the β_2 AR-T4L complex, followed by EM-PS.

- **Step 3.** T4-lysozyme was removed from the engineered β_2 AR-T4L complex, and the 32 missing residues (residues 231–262) of ICL3 were incorporated in between TM5 and TM6. The resulting receptor is referred to as β_2 AR-ICL3. The details of *de novo* loop modeling are described above. This step required readjustment of the number of neutralizing ions. EM-PS was carried out.

- **Step 4.** W122^{3,41} on TM3 was mutated back to E122^{3,41}. We chose the protonated (neutral) state of the E122 residue, as it faced directly toward membrane hydrophobic interior. EM-PS was carried out.

- **Step 5.** E187^{5,26} on ECL2 was mutated back to N187^{5,26}. The number of counterions was adjusted (as in Method A to 94 Na⁺ and 102 Cl⁻ ions) and EM-PS was carried out.

- **Step 6.** Finally, the native apo-receptor obtained after the above-mentioned changes was simulated for 200 ns in a DOPC bilayer and compared with the results of Method A.

Control Simulations of Method S. For completeness, we carried out two separate tests to confirm that the results for Method S were on a solid basis:

- **Control-1.** We tested the influence of cholesterol: In addition to the simulation in a DOPC bilayer, the β_2 AR model obtained through Method S (after Step 5) was placed in a DOPC bilayer containing 10 mol % cholesterol, followed by EM-PS. The production simulations lasted for 200 ns.

• **Control-2.** Different sequence of steps: The order of steps discussed above (first removing timolol, then T4L, and finally reverting the mutations) was chosen based on the expected degree of significance: Starting from substantial structural changes and moving on structural alterations that were expected to be smaller. However, in general the number of how these steps can be combined is large, and the results may depend to some extent on the order in which the structural modifications in the protein are reverted. To consider the sensitivity of the order of steps in Method S, we also tested a different scenario where the above-mentioned order of steps (Step 2 – Step 5) was reversed. In that case, starting from the final snapshot of Step 1, we first removed mutations; second we replaced the T4-lysozyme with ICL3; and then we removed timolol. Finally, the native apo-receptor was equilibrated in a DOPC bilayer.

As to the final details, the software BODIL 0.8.1³⁹ was used for reverse mutations. The side chain orientation of the back-mutated residues was chosen such that there was no collision with its neighboring residues. Prior experimental knowledge was taken into consideration, where possible. As to the W122^{3,41} to E122^{3,41} mutation, the orientation of E122^{3,41} was chosen to match the one found in another crystal structure of the receptor [PDB id: 2RH1].¹⁵

■ RESULTS AND DISCUSSION

Below we describe and discuss the structural properties of the receptor models we have studied. The Discussion focuses on structural differences arising from different ways to revert the protein modifications used in experimental crystal structure determination.

Method A: Reversal of All Crystallization Modifications at Once Led to Instabilities in Protein Structure. We first used Method A to investigate how the protein structure changes during simulations when *all* modifications done for crystallization assays are reverted at once. Here we discuss the results of the Long simulation, unless stated otherwise. Figure 1A shows the root-mean-square deviation (RMSD) of the protein backbone throughout the simulation. Large RMSD values were observed mainly due to the flexible intercellular and extracellular loops. RMSD for the transmembrane part was significantly smaller and reached a plateau within ~90 ns of simulation. The average RMSD between the crystal structure and the simulation (the last 100 ns), calculated over the backbone atoms of the transmembrane region, was 2.89 ± 0.002 Å.

The corresponding RMSD fluctuations (per residue) from the crystal structure are plotted in Figure 1B. The snapshot shows the average structure (over the last 100 ns) of the protein with the TM residues colored according to the RMSD values. As depicted in Figure 1B, many parts of the protein underwent large deviations from the crystal structure. These include TM1 (especially the N-terminus up to its middle part (RMSD > 5 Å)), TM4, residues 197–207 at the extracellular half of TM5 (RMSD > 4 Å), and residues 315–324 in the middle of TM7 that includes residues from the conserved NPxxY motif. Besides the ends of the TM helices, many residues from their center part located at the helical interface or underneath the ligand-binding cavity deviated considerably from the crystal structure: e.g., residues 115–118 of TM3, residues 282–291 of TM6, residues 315–324 of TM7, etc. We found that these deviations were related to a change occurred in

the secondary structure of the protein during the simulation (Figure S2).

Figure 1C shows a simulation snapshot of β_2 AR with the transmembrane residues colored according to the secondary structure content. Figure 1D represents the average helical content (over the last 100 ns) of each TM helix calculated with DSSP.⁴⁰ As shown in Figures 1C and S2, the first 10 N-terminal residues of the TM1 helix completely unfolded to turn/bend structure. Partial opening of the helix also occurred at residues 42–50 in the middle of TM1. In this part of TM1, we observed a considerable drop in the α -helical content; meanwhile, the percentage of the 3_{10} -helix increased. 3_{10} -helix is the proposed intermediate in the folding/unfolding of α -helices and known to be involved in a helix–coil transition.⁴¹ Altogether, the average helical percentage of TM1 dropped down to 52% compared to the crystal structure (Figure 1D and Table S1). The unwinding of TM helices also occurred in residues 83–86 in the middle of TM2; in residues 162–171 on the extracellular side of TM4 (P168 caused local opening of helical structure); in residues 197–207 in the extracellular-half of TM5 (which is a major part of the ligand-binding pocket); in residues 280–285 in the middle of TM6 and residues 271–274 near the intracellular end of TM6; in residues 315–320 before the NPxxY motif of TM7, etc. (see Figures 1C and S2). As a result, the helical contents decreased to 65% in TM2, 64% in TM4, 60% in TM5, 59% in TM6, and 71% in TM7 of their respective values in the crystal structure (Figure 1D and Table S1).

In agreement with the results obtained from the Long (1 μ s) simulation, a considerable decrease in helical percentages of the TM domains was also observed in the three short trajectories (Short-1, Short-2, and Short-3 (Figure 1D and Table S1)). On average, in the shorter simulations the most noticeable drop in helical content occurred for TM1 (58%), TM4 (68%), TM6 (56%), and TM7 (62%), as compared to the crystal structure.

Summarizing, the β_2 AR model obtained by Method A, when simulated in a DOPC lipid bilayer, exhibited considerable deformation from the crystal structure. The main weakness of Method A was observed to be the significant drop in the helical content of TM domains of the receptor. For example, TM1 and TM6 lost almost half of their helicity compared to the crystal structure. A previous circular dichroism (CD) spectroscopic study characterized very highly helical structure (>80%) of peptides corresponding to the TM domains of a GPCR (called the adenosine A2a receptor) in micelles and lipid vesicles.⁴² The high helical propensity of the TM domains are required for proper folding of GPCRs⁴³ and thus related to their function.

Method S: Step-by-Step Reversal of All Structural Modifications Was Found to Be a Superior Technique Compared to Method A and Resulted in a Stable Receptor Model. The structural instability of the receptor observed in the previous procedure led us to rebuild the model more carefully. Here, in Method S we removed the structural modifications, done originally for crystallization studies, in a stepwise manner. In the following section, we describe the effects that were observed after reversal of each structural change. Finally, when all structural modifications have been reverted, we compare the properties of the receptor found in this manner to the structure found through Method A.

Step 1: Equilibration of the Ligand Bound β_2 AR-T4L Complex in a Lipid Bilayer Environment. For crystallization, membrane proteins are normally extracted from their native lipid membrane environment, and instead detergents or detergent-lipid mixtures are used to solubilize and reconstitute

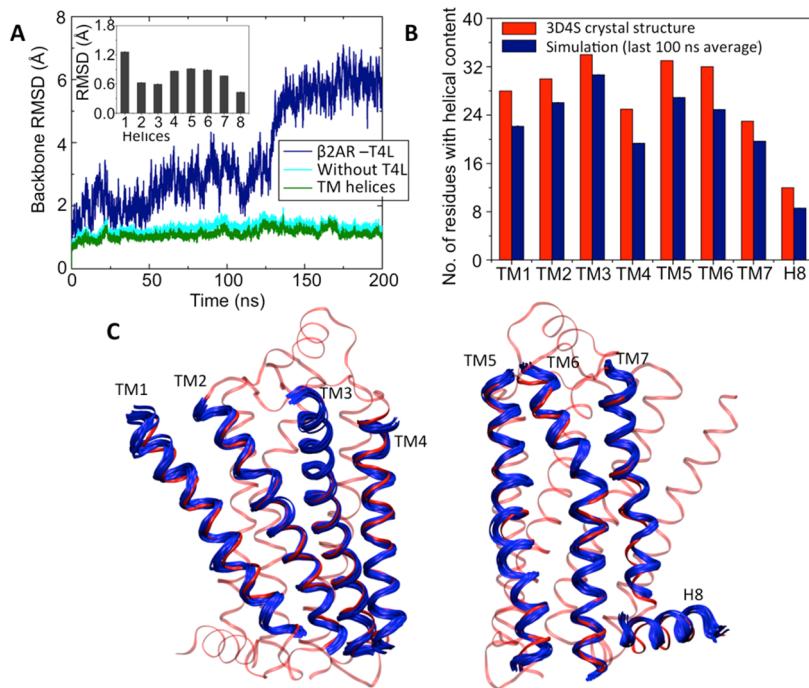


Figure 2. Equilibration of the ligand-bound β_2 AR-T4L complex in a membrane (Method S). A) Time evolution of the backbone root-mean-square deviation (RMSD) of the entire β_2 AR-T4L complex (blue), β_2 AR-T4L without T4L (cyan), and the transmembrane region of the receptor (green). The inset represents the time average (the last 100 ns) for the RMSD of each individual transmembrane helix TM1–7 and helix-8. The error in the average RMSD per helix was less than 0.01 Å. B) The average number of residues with helical content (sum of α -helix, β -helix, and 3_{10} -helix) calculated from the simulation trajectory (blue bars, average over the last 100 ns) and compared to the crystal structure 3D4S (red bars). C) Superimposed snapshots from the last 50 ns of the trajectory (blue ribbons, taken every 5 ns), with respect to the crystal structure (red ribbon). The error in the number of residues with helical content was below 0.04.

membrane proteins in vitro.³ As a starting point for the present work, we allowed the timolol bound β_2 AR-T4L complex (as given by the X-ray structure) to settle down in a lipid bilayer environment. Figure 2A represents the time evolution of the RMSD of the protein backbone. A major contribution to RMSD was observed to originate from the 159-residue long T4L (Figure 2A), which was completely exposed to water. The transmembrane part of the protein was very stable with RMSD less than 1.5 Å and reached a plateau within the first 50 ns of simulation. The protein also retained its overall secondary structure content in a membrane during the entire simulation time span (Figure S3).

We also calculated the average backbone RMSD (inset in Figure 2A) and average helical content (Figure 2B, shown in comparison to the crystal structure) of TM1–7 and the C-terminal short helix H8, separately. These results, along with the superimposed snapshots taken from the last 50 ns of the trajectory (Figure 2C) revealed that from the studied transmembrane segments, TM1 showed the largest deviation (1.26 ± 0.01 Å), especially at its N-terminal end. Though this value is quite small, TM2 (RMSD being 0.63 ± 0.01 Å) and the longest helix TM3 (RMSD 0.60 ± 0.01 Å) were observed to be pretty stable. The mobility of the lysozyme bound to the intracellular ends of TM5 and TM6 had a major contribution to the calculated RMSD of these two helices (0.91 ± 0.01 Å and 0.90 ± 0.01 Å, respectively). The flexible C-terminal helix-8 (RMSD having a value of 0.43 ± 0.01 Å) contributed to the deviation of TM7 (RMSD 0.77 ± 0.01 Å) mainly near the cytoplasmic end. Altogether, these results showed that the protein relaxed in the membrane environment without any significant deviation from the crystal structure.

Step 2: Removal of Timolol Resulted in Fluctuations and a Relative Opening of the Ligand-Binding Cavity, Suggesting That in the Apo-Receptor the Binding Cavity Can Oscillate between Several Conformations.

In the crystal structure, the partially inverse agonist timolol was bound at the main ligand-binding cavity of the receptor located between the extracellular segments of the transmembrane helices 3, 5, 6, and 7 (Figure 3A). With the removal of timolol, the most prominent observation was the penetration of a large number of water molecules into the binding pocket (Figure 3B). Water molecules migrated rapidly into the cavity in order to fill the void generated by the removal of timolol. However, also in the presence of timolol the cavity was highly hydrated (Figure 3A), and we found on average ~ 18 water molecules to be present (at any moment) inside the pocket during the last 100 ns of the trajectory. This level of hydration is in agreement with a previous study, showing the fundamental role of water as a determinant of binding affinity of β -blockers to β_2 AR.⁴⁴ With the removal of timolol, the number of water molecules occupying the cavity increased to ~ 28 (these values are based on calculations where we integrated the area under the number density plot of water molecules that penetrated the binding site). Together with water, a sodium ion was observed to enter deep into the cavity (Figure 3B) and stay there for the rest of the simulation time. The ion binding to the cavity is quite certainly driven by electrostatics, since the cavity has a negatively charged D113, which attracts Na^+ , neutralizing that region.

Next, we made a closer inspection of what happened to the residues at the ligand-binding site (Figure 3C,D). The direct β_2 AR-timolol hydrogen bonding interactions were found to

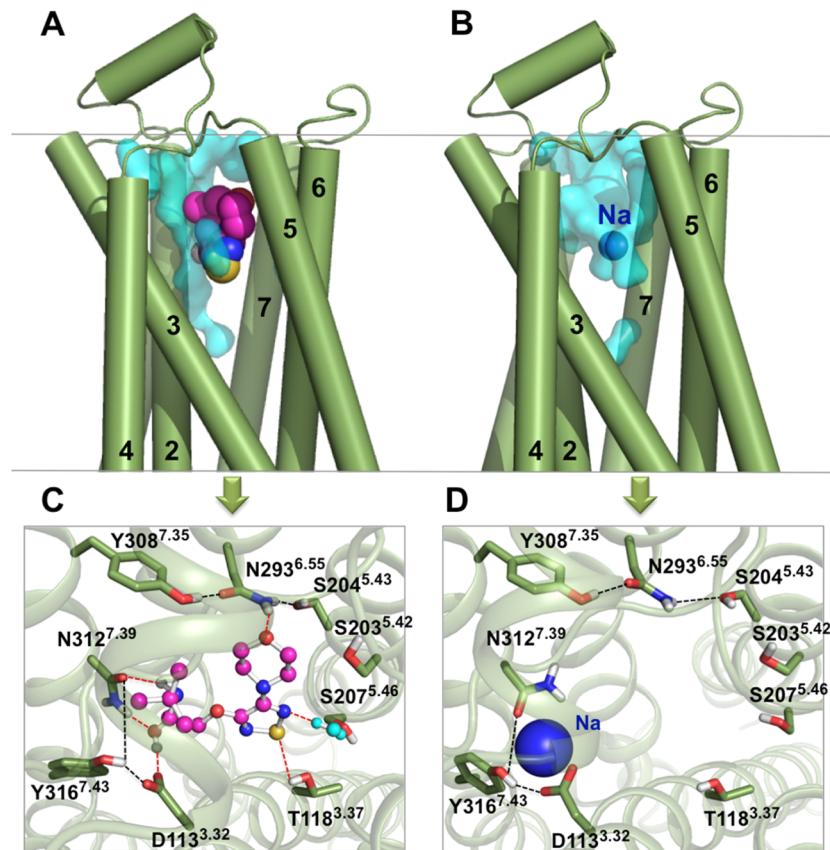


Figure 3. Interactions at the ligand-binding site (Method S). A) A representative simulation snapshot shows timolol at the main ligand-binding cavity of β_2 AR. Timolol is shown as spheres and is colored according to the atom types (magenta for C, blue for N, red for O, yellow for S, and white for H). Nonpolar hydrogens are not displayed for clarity's sake. Protein is represented by a green cartoon with transmembrane helix numbering. Waters inside the cavity are shown as a transparent cyan surface. B) A representative snapshot of the simulation after removal of timolol. More waters occupy the cavity along with a sodium ion (shown as a blue sphere). C–D) Close-up of polar interactions at the ligand-binding site: interactions between timolol and protein are shown by red dotted lines and intraprotein interactions by back-dotted lines. Amino acids participating in the interaction are highlighted as sticks. Waters are not shown for clarity, except one (cyan bead), which made a water-bridge between Ser207^{5.46} and the thiadiazole ring N of timolol (panel C).

take place for (i) D113^{3.32} and N312^{7.39} interacting with the polar tail of timolol, (ii) N293^{6.55} interacting with the timolol head (through the ether moiety of the morpholino ring), and (iii) T118^{3.37} hydrogen bonding with the thiadiazole ring of timolol (Figure 3C). D113^{3.32}, N312^{7.39}, and N293^{6.55} were the most contributing residues, in line with previous experimental and mutagenesis studies.^{13,45–47} In addition to these, we observed a very persistent water-bridging interaction between S207^{5.46} and N of the thiadiazole ring of timolol (Figure 3C). Meanwhile, β_2 AR residues, which frequently involved in nonbonded interactions with timolol are V114 and V117 in TM3, F193 in ECL2, and F289 and F290 as well as W286 in TM6. Other residues contributed less to these interactions (they interacted with timolol (measured as a contact between any heavy atom of timolol and the residue within 0.45 nm) for less than 10% of simulation time). Both in the presence and absence of timolol, the polar triads D113^{3.32}-Y316^{7.43}-N312^{7.39} and S204^{5.43}-N293^{6.55}-Y308^{7.35} remained stabilized through a network of intraprotein hydrogen bonds (Figure 3C,D). These intraprotein interactions helped in stabilizing the transmembrane helix interfaces (e.g., interfaces of TM3 and TM7 in the case of D113^{3.32}-Y316^{7.43}-N312^{7.39} and TM domains 5–6–7 in the case of S204^{5.43}-N293^{6.55}-Y308^{7.35}). As mentioned above, a sodium ion entered the pocket after the removal of timolol and bound near the negatively charged D113^{3.32}. The

ion further stabilized the D113^{3.32}-Y316^{7.43}-N312^{7.39} anchor site, which was previously found to interact with the oxypropanolamine tail of timolol (Figure 3C,D).

With the removal of timolol, the main change we observed at the ligand-binding cavity took place around S207^{5.46} in TM5. Serine residues S203^{5.42}, S204^{5.43}, and S207^{5.46} in TM5 are known to play a key role in agonist binding. Previous studies showed that serine residues of TM5 form strong polar interactions specifically with catecholamine agonists, which resulted in an inward shift of the extracellular part of TM5 around S207^{5.46}, as compared to the antagonist-bound inactive form.^{4,48,49} Here, to measure the displacement of TM5, we calculated the distance between $C\alpha$ atoms of S207^{5.46} and D113^{3.32} (as the position of D113^{3.32} in TM3 did not change noticeably during the simulation). As shown in Figure 4, in the presence of timolol, the ligand-binding site remained very stable, and there was only a small fluctuation in the S207^{5.46}-D113^{3.32} distance around an average value of 12 Å. This value is in quite excellent agreement with the corresponding value (12.07 Å) obtained from the crystal structure of timolol-bound β_2 AR.¹³ However, as we removed the ligand, the pocket started to oscillate between the closed and open forms with the S207^{5.46}-D113^{3.32} distance ranging between 11 and 14.2 Å (Figure 4). Near the end of the simulation, the extracellular part of TM5 drifted away from D113^{3.32} in TM3, resulting in an

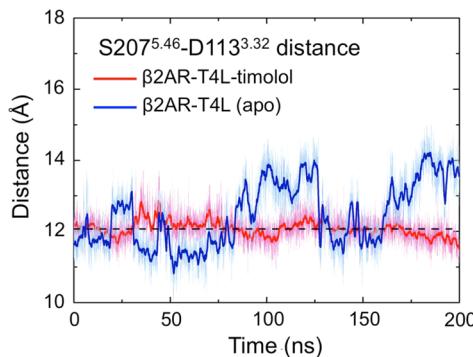


Figure 4. Fluctuation and opening at the ligand-binding site after removal of timolol (Method S). Time evolution for the distance between C α atoms of D133^{3,32} in TM3 and S207^{5,46} in TMS of β_2 AR-T4L, with (light red) and without (light blue) timolol. Corresponding 50-point running averages are shown in dark color. Black dashed line represents the S207^{5,46}-D133^{3,32} distance in the crystal structure (PDB id: 3D4S).

opening of the ligand binding site around S207^{5,46} by $\sim 2 \text{ \AA}$. This outward movement of TMS around S207^{5,46} took place exactly in the opposite direction compared to the case of agonist binding.

Given the above, our results suggest that in the apo-receptor the binding cavity can oscillate between various conformations. The presence of a ligand considerably stabilizes the pocket with favorable receptor–ligand interactions and shifts the equilibrium toward a specific conformation. We will discuss this in detail in future publications.

Step 3: With the Replacement of T4L by ICL3, the Cytoplasmic End of TM6 Moved toward the Core of the Receptor, Further Closing the G-Protein Binding Site. In crystallization, the incorporation of T4L stabilizes the dynamic

region of β_2 AR, namely the cytoplasmic ends of TM5 and TM6, and the intervening ICL3.⁴ The engineered β_2 AR-T4L retains nearly all functions of the native receptor.⁵⁰ Despite that, ICL3 is known to directly interact with the G-protein and thus possibly affect receptor activation.⁵¹ In our study, the third ICL was observed to be the most flexible part of the receptor, exhibiting the highest root-mean-square fluctuation (Figure 5A). As Figure 5B depicts, ICL3 adopted multiple conformations and orientations during the course of the simulation, indicating its highly dynamic nature: starting from a relatively open initial structure (inset I at $t = 0 \text{ ns}$ in Figure 5B), the loop first packed itself more tightly under the extracellular ends of the helix-bundle, near the G-protein's binding site (inset II at $t = 60 \text{ ns}$), but later it again opened up and moved further away from the core of the receptor (inset III at $t = 200 \text{ ns}$). In a recent study,¹¹ a similar structure as in the inset II was pointed to be a “very inactive” conformation, backed up by a view that ICL3 almost completely blocks the G-protein's binding site. During our simulation, the loop remained mostly unstructured. The formation of a short segment of antiparallel β -strands between the residues Q231-K232 and S261-S262 was most frequent with regard to the possible secondary structure elements, complemented by occasional formation of a short helix between the residues 247–250 (Figure 5B). Additional simulations also supported the disordered and highly dynamic nature of ICL3 (see SI, Figures S4 and S5). These simulations also showed that the flexibility of ICL3 did not destabilize the rest of β_2 AR or affect its secondary structure.

The mobility of ICL3 may affect the dynamics of the transmembrane part, particularly helices 5 and 6. In the present work, after replacement of T4L by ICL3, we observed a shift of $\sim 4 \text{ \AA}$ (compared to the crystal structure) of the intracellular end of TM6 toward the core of the heptahelical bundle,

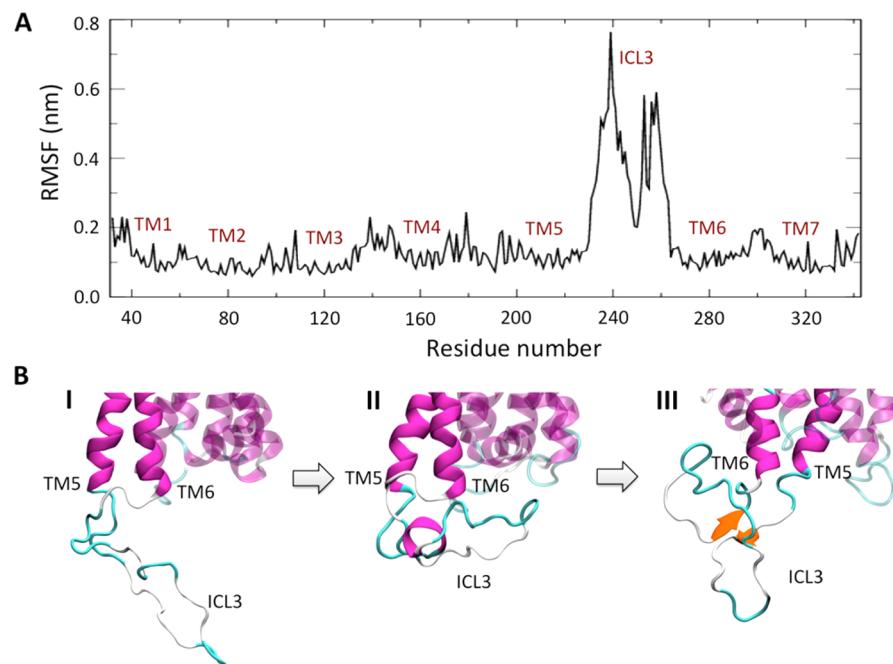


Figure 5. Dynamics of ICL3 (Method S). A) Root-mean-square fluctuation (RMSF) of the backbone atoms per amino acid residues of β_2 AR-ICL3. B) Snapshots showing the various conformations and orientations adopted by ICL3 during the course of the simulation: inset I in the beginning of the simulation (0 ns), inset II at 60 ns , and inset III in the end of the simulation (200 ns). In the snapshots, the secondary structures are colored as magenta for the α -helix, orange for the extended- β structure, cyan for turn, and white for coil.

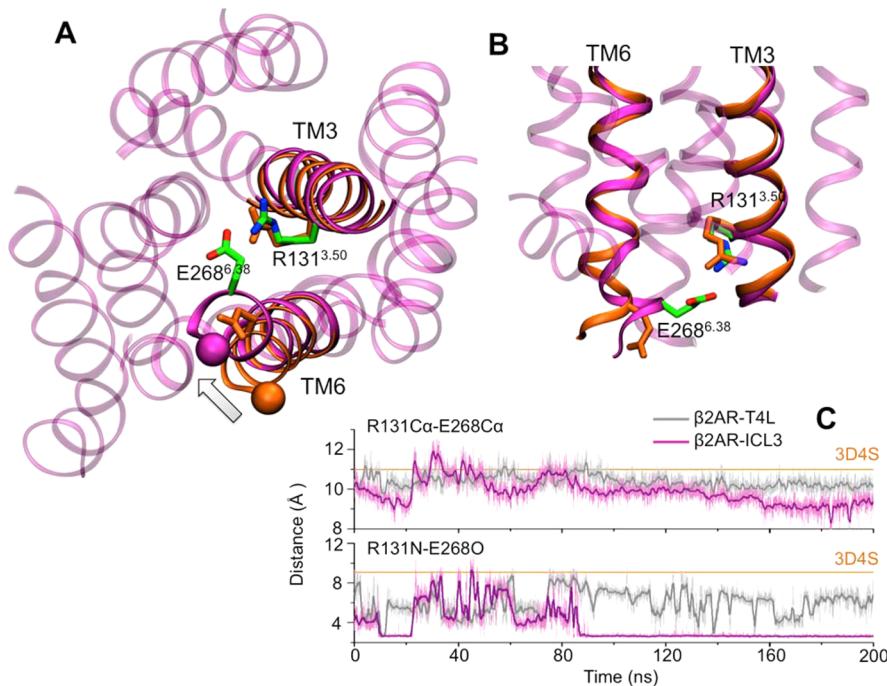


Figure 6. Changes at the G-protein binding site after removal of T4L (Method S). The snapshots of β_2 AR-ICL3 (magenta) represent the displacement of the cytoplasmic end of TM6 toward TM3, compared to the crystal structure (orange). A) View from the G-protein binding site and B) side view. The cytoplasmic half of TM3 and TM6 is highlighted in the snapshot. The residues R131^{3.50} and E268^{6.38} of the simulation model are shown as sticks and are colored according to atom types (green for C, red for O, blue for N). Hydrogen atoms are not displayed for clarity. The corresponding residues of the crystal structure are shown as orange sticks and exhibit a broken salt-bridge. In the snapshot A) the gray arrow shows the shift of the intracellular end of TM6 (Ca atom of K267 shown as a sphere) of β_2 AR-ICL3 (shown in magenta) compared to the crystal structure (shown in orange). C) Upper panel shows the distances between the Ca atoms R131^{3.50} and E268^{6.38} in β_2 AR-T4L (gray) and β_2 AR-ICL3 (magenta). The lower panel shows the minimum distance between the guanidium nitrogen atoms of R131^{3.50} and the carboxyl oxygen atoms of E268^{6.38}, characterizing salt bridge formation. Dark colors refer to corresponding 50-point running averages. The orange lines refer to the corresponding values in the crystal structure [PDB id: 3D4S].

followed by formation of the R131^{3.50}-E268^{6.38} salt-bridge (Figure 6A,B). Such inward movement of TM6 largely closed the G-protein binding site and occurred exactly in the opposite direction than the displacement observed during receptor activation. Figure 6C represents the displacement of the cytoplasmic end of TM6 from TM3, and the time evolution of salt-bridge formation during the simulation time span. Our results show that in β_2 AR-ICL3, the intracellular end of TM6 approached TM3, without noticeable displacement of the latter. As a result, the corresponding Ca-Ca distance of R131^{3.50}-E268^{6.38} decreased by \sim 2 Å compared to the crystal structure (Figure 6). Unlike β_2 AR with the lysozyme (β_2 AR-T4L), the presence of a flexible ICL3 (replacing T4L) allowed the two helices to be in close proximity and also fostered the formation of a salt-bridge between them. For β_2 AR-ICL3, we observed that as the R131^{3.50}-E268^{6.38} salt-bridge formed (in \sim 80 ns), it remained very stable until the end of the simulation (Figure 6C). The salt-bridge formation is in agreement with a previous simulation study^{10,52} and based on the available mutagenesis data⁵³ is known to play a crucial role in β_2 AR activation.

In relation to the conformational change in the intracellular side of the receptor, we also monitored changes that occurred in the extracellular ligand-binding site (using the same approach as described in Step 2). As shown in Figure S6, in β_2 AR-ICL3 the S207^{5.46}-D113^{3.32} distance was \sim 14 Å, which is \sim 2 Å larger than the crystal structure value. As compared to β_2 AR-T4L, in β_2 AR-ICL3 there was no further opening at the ligand-binding pocket or oscillation between the closed and open forms (Figures 4 and S6). Thus, the narrower G-protein binding site

in β_2 AR-ICL3 seems to stabilize the ligand-binding site to a relatively open state, which supports the view that the allosteric network acts between the two distal sites of β_2 AR.

Step 4: Back Mutation of W122^{3.41} to E122^{3.41} Affects the Packing and Interactions at the Transmembrane Helix Interface. In β_2 AR, E122^{3.41} of TM3 is located at the interface between the helices TM3-TM5, facing the membrane exterior. Engineering the helical interface by mutating E122^{3.41} with bulky hydrophobic residues, e.g., tryptophan (as found in rhodopsin) has shown a significant potential to enhance receptor stability while maintaining wild-type functions.⁵⁴ As found in our simulation, W122^{3.41} was sandwiched between V160 in TM4 and P211 in TM5 (Figure 7A). The aromatic ring of W122^{3.41} formed CH- π stacking interactions with the ring of P211. The indole nitrogen of W122^{3.41} formed a direct hydrogen bond with the hydroxyl group of T164 in TM4 (Figure 7A), which further stabilized helix 4. However, we did not observe any polar interactions between W122^{3.41} N and backbone carbonyl of Val206 in TM5, as previously predicted based on homology modeling.⁵⁴

In our studies, back mutation of W122^{3.41} to E122^{3.41} disrupted the above-mentioned packing interactions between the TM segments (Figure 7A). Interaction with P211 was lost. The carboxyl group of E122^{3.41} resides too far to make direct hydrogen bonding with T164. Rather, E122^{3.41} formed only weak water-mediated interactions with T164 in TM4 and with the backbone carbonyl group of V206 in TM5 (Figure 7A).

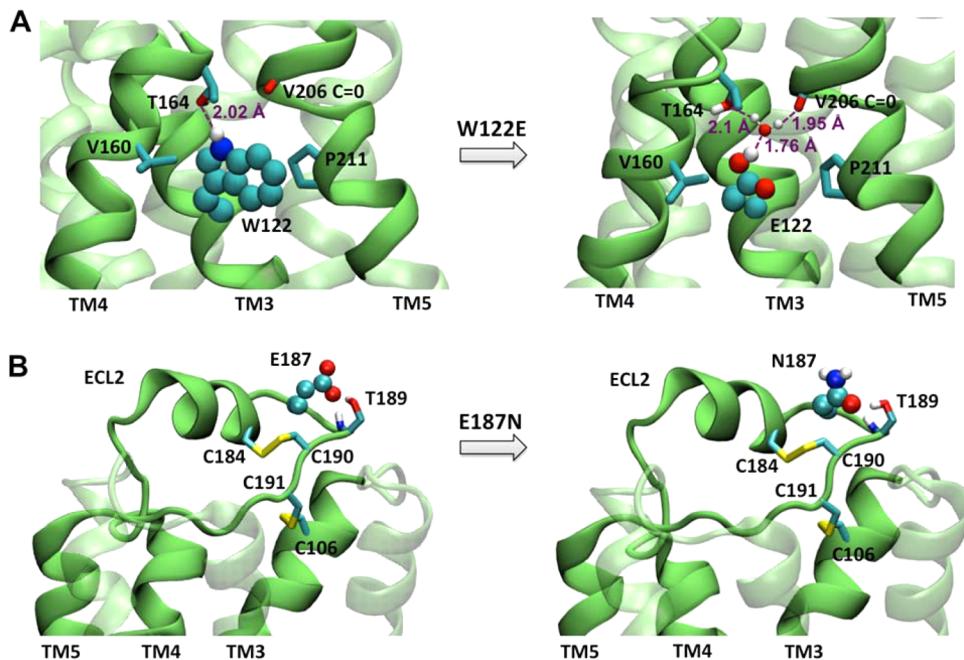


Figure 7. Effects of mutations (Method S). A) Representative simulation snapshots showing the interactions at the TM4-TM3-TM5 interface around W122^{3.41}/E122^{3.41} of TM3. In the figure W122^{3.41}/E122^{3.41} residues are represented as spheres and are colored according to the atom type (cyan for C, blue for N, red for O, yellow for S, and white for H). The neighboring residues are shown as sticks. A water molecule within 5 Å around the 3.41 position is shown as a ball-and-stick representation. Hydrogen bonding interactions are shown with purple dashed lines. B) Representative simulation snapshots showing the interactions around the residue 187 at ECL2 (represented as spheres and colored according to the atom type). Two disulfide bonds of ECL2, C184^{4.76}-C190^{5.29} and C106^{3.25}-C191^{5.30}, are shown as sticks.

Altogether the results point to a conclusion that the loss of packing interactions at the transmembrane helix interface contributes to the inherent structural instability of β_2 AR.

Step 5: Back Mutation of E187^{5.26} to N187^{5.26} Did Not Affect Considerably the Local Structure and Orientation of ECL2. Human β_2 AR has a potential N-glycosylation site located in the second extracellular loop (ECL2) at the residue 187^{5.26,55}. For structure determination in experiments, N187 was mutated to E187 to avoid glycosylation and therefore foster crystal formation. We examined whether the mutation of N187E to E187 had an impact on the structure/orientation of ECL2.

As shown in Figure 7B, the local structure remained largely the same after the reverse mutation of E187^{5.26} back to the native N187^{5.26}. The short helical region of ECL2 remained intact. T189^{5.28} on ECL2, which was previously interacting with the carboxyl oxygen of E187^{5.26}, was now found to interact with the amide oxygen of N187^{5.26}. For both cases, the intraloop disulfide bond between C184^{4.76} and C190^{5.29} stabilized the conformation of ECL2. Another disulfide bond between C191^{5.30} and C106^{3.25} in TM3 maintained the orientation of ECL2 with respect to the transmembrane helix-bundle (Figure 7B). Therefore, the mutation alone had no significant effect on ECL2, which acts as an access point to the binding site.

In human β_2 AR the N-glycosylated residues are known to play an important role for the proper insertion of the receptor in a cell membrane and also in agonist trafficking.^{55,56} However, understanding the effect of glycosylation is beyond the scope of the current study and remains to be discussed elsewhere.

Step 6: The Stepwise Approach Resulted in an Improved Model, Which Reproduced the Properties of the Native Receptor Very Well. Figure 8 shows the key properties of the receptor structure found through Method S, in

which the modifications done for experimental structure determination were removed step by step. These findings should be compared with the results presented in Figure 1 for Method A, where the same structural modifications were corrected simultaneously in a single step. Below we discuss the results of Method S, unless said otherwise.

When we summarize the outcome of Method S, we first find (Figure 8A) that the backbone RMSD profile of the TM region of β_2 AR reached a plateau much faster (within the first 50 ns of simulation) than the entire receptor (~170 ns). The average RMSD (the last 100 ns) of the TM region (backbone atoms only) from the crystal structure was 1.45 ± 0.01 Å, while the corresponding RMSD in Method A was 2.89 ± 0.01 Å.

Figure 8B shows the average RMSDs per TM residues. In Method S, the majority of the protein exhibited a deviation that was less than 1.5 Å from the crystal structure, considerably less than what was found in Method A (Figure 8B). With Method S the N-terminal end of TM1 was the major fluctuating part of the protein (RMSD < 3 Å), although the drift was much smaller than in the model based on Method A (>6 Å) (Figure 8B). TM2 and the longest TM3 were the most stable ones among all the helices. TM4 fluctuated in both models (though the results for RMSD per residue were smaller in the case of Method S). The smallest helix TM4 is considered to be the weakest point (in terms of exhibiting the lowest intramolecular packing calculated by the occluded surface area method in ref 13) of the β_2 AR fold.¹³ Unlike in Method A, in Method S we did not observe any significant deviation at the intracellular end of TM3. Instead the intracellular end of TM6 (residues 267–269) exhibited larger deviations (>3 Å). This occurred due to a shift of TM6 toward the core of the receptor (after removal of T4-lysozyme), in order to form a salt bridge with the cytoplasmic end of TM3 (Figure 6A). Similarly, the fluctuations

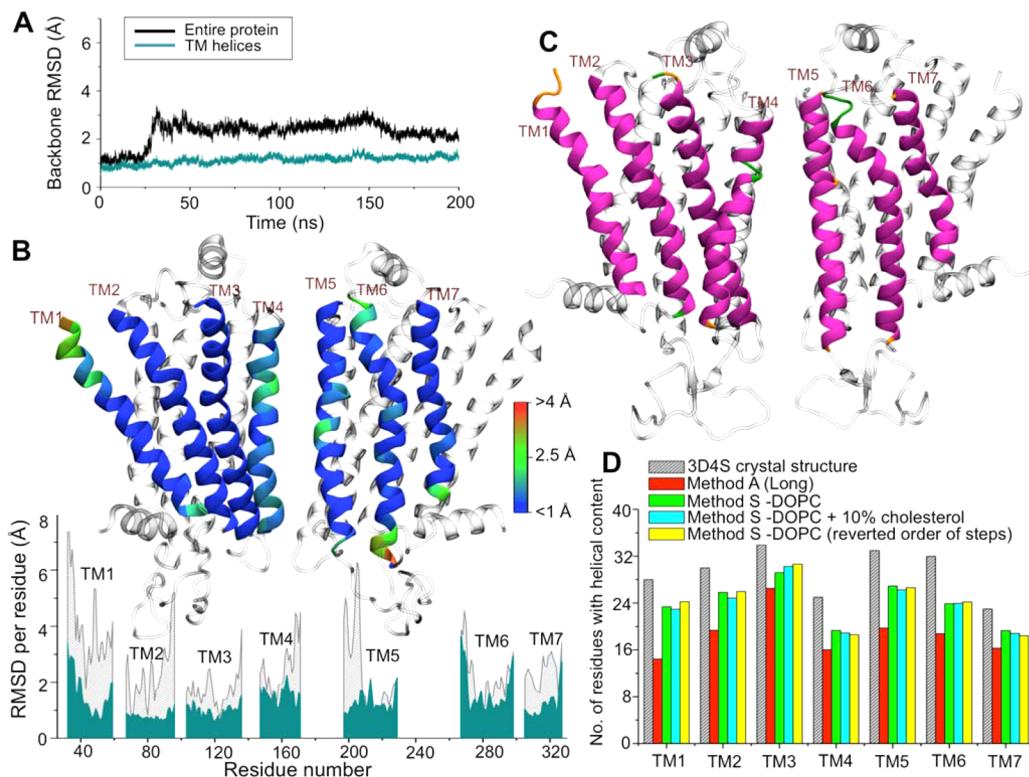


Figure 8. Properties of the native receptor model generated through the stepwise approach (Method S). A) Root-mean-square deviation (RMSD) of the protein backbone: entire β_2 AR (black) and the transmembrane region (cyan), throughout the simulation. B) Per residue RMSD (average over the last 100 ns) of the simulated receptor model compared to the crystal structure, calculated over the backbone atoms of the transmembrane region. In the graph, the cyan and gray areas correspond to the stepwise approach (Method S) and the simultaneous removal of crystallization modifications (Method A), respectively. The snapshots show the front (TM1-TM4) and back (TM5-TM7) views of the protein's average structure (over the last 100 ns), with TM residues colored according to their respective RMSD values, and the rest of the protein represented as a transparent gray cartoon. C) A representative snapshot of the receptor with TM residues colored according to the secondary structure content (α -helix in purple, 3_{10} -helix in blue, turn in green, coil in orange): front (TM1-TM4) and back (TM5-TM7) sides. D) The number of residues with helical content (sum of α -, 3_{10} -, and Π -helices) for each TM helix calculated from different Method S trajectories (averaged over the last 100 ns): in a one-component DOPC bilayer (green), in a DOPC bilayer with 10 mol % cholesterol (cyan, Control-1), and in a DOPC bilayer with the reverted order of steps (yellow; Control-2). The results of Method S were compared to that found in Method A (red, Long simulation) and in the crystal structure 3D4S (gray). The error in the number of residues with helical content was below 0.04.

around S207^{5,46} in TM5 (Figure 4, after removing the timolol ligand) were reflected in higher RMSD values of this region (Figure 8B). The deviation observed in the cytoplasmic end of TM7 (at the conserved NPxxY motif) was consistent with an earlier study.⁷

As shown in Figures 8C and S7, the β_2 AR model obtained by Method S retained its overall secondary structure when simulated in a DOPC lipid bilayer. All TM domains exhibited very high helical propensity: helical content of TM1, TM2, TM3, TM5, and TM7 were above 80%, while that of TM4 and TM6 were ~75% (Figure 8D and Table S2). Proline residues caused local opening of a helix at TM4 (P168) and TM7 (P323). As compared to Method A (Long simulation), the most noticeable enhancement of helical content was observed for TM1 (from 52% in Method A to 84% in Method S), TM2 (from 65% in Method A to 86% in Method S), TM5 (from 60% in Method A to 82% in Method S), TM6 (from 59% in Method A to 75% in Method S), and TM7 (from 71% in Method A to 84% in Method S) (Figure 8D and Table S2). Similar conclusions were drawn from the results by the different sequence of steps in Control-2 (Figure 8D, Figure S8, and Table S2), which indicated that the order of steps did not affect the results of Method S to a significant extent.

The β_2 AR model obtained by Method S was further validated in a DOPC bilayer with 10 mol % cholesterol (Control-1). Cholesterol is the most prevalent sterol of the plasma membrane and known to influence GPCRs structure/function.^{21,22} As shown in Figures 8B and S9, per residue RMSDs of TM residues showed similar behavior in both bilayers. The high helical contents of TM domains as observed in a DOPC bilayer without cholesterol were well reproduced in a DOPC bilayer with 10 mol % cholesterol (Figure 8D and Table S2). To assess the effect of cholesterol on the stability and function of β_2 AR, more thorough studies are needed. Authors are planning to address this issue in the future.

Summarizing, the present model using Method S showed significant improvement in the total average helical content of TM domains of β_2 AR compared to the model prepared with Method A (Figure 8D). For Method S, the helical structure of TM region of the receptor was well preserved both in DOPC bilayer and in more biologically relevant cholesterol-containing bilayer. The stepwise method was able to reproduce the conserved heptahelical architecture of GPCR and provided a much better receptor model than Method A.

CONCLUSIONS

In the present work, we used atomistic molecular dynamics simulations to model the properties of the native β_2 AR receptor in a lipid membrane, starting from a heavily engineered crystal structure. β_2 AR is one of the most well studied GPCRs, and thanks to recent advancements in development of crystallization techniques, several high-resolution structures of this receptor or GPCRs in general are now available in protein structure databases. Like many other GPCRs, the inactive β_2 AR (that is of our interest here) underwent several major structural modifications during its structure determination that were made in experiments to foster crystallization, e.g., replacement of highly dynamic loops with a rigid lysozyme, introduction of stabilizing mutations, etc. In order to overcome the inherent dynamic nature of the receptor, such structural modifications have become general strategies for crystallizing membrane proteins.

In the current study, we explored what could be the most efficient way to get rid of the crystallization modifications without altering the properties of the native β_2 AR receptor substantially. By following the most usual approach, we first removed all changes made originally for crystallization and then simulated the receptor in a lipid membrane environment. In this approach (Method A), we simultaneously modified the ligand-binding site by removing timolol, altered the G-protein binding site by replacing T4L with ICL3, and also modified the transmembrane helical interface by removing a mutation. With all these simultaneous changes, the method led to a receptor structure that deviated quite significantly from the crystal structure. The major drawback found in this method was that it resulted in a considerable drop in the helical content of the transmembrane region of β_2 AR, e.g., on average the helical percentage decreased (compared to helicity in the crystal structure) to 58% in TM1, 56% in TM6, and 62% in TM7. There is reason to stress that heptahelical architecture is an inherent property of GPCRs, and it should be reproduced in the computational model, too.

In an alternative approach (Method S), we removed the modifications (made for crystallization experiments) in a more careful, stepwise manner (one at a time, each phase followed by hundreds of nanoseconds of equilibration). In this method, the instability in the receptor structure was overcome, and the helical characteristic of the transmembrane region was preserved. Without doubt, the results of Method S are quite superior compared to those given by the more straightforward technique, Method A.

The price one has to pay to use Method S is the increased computational cost. In the present study, the simulations with Method S covered a total simulation time scale of about 1.5 μ s. Meanwhile, simulations with Method A could have been done in about 250 ns (though we extended those simulations to a microsecond). Yet, we consider that the accuracy and the reliability of the simulation model and its predictions justify the added computational burden. Further, while one may debate whether the results of Methods S and A can be compared to each other on equal footing due to the different simulation times, it is important to bring out that many membrane protein simulations are performed over times of the order of 100 ns. This is particularly common, e.g., with the CHARMM force field that is often considered the force field of choice for protein simulation. The simulation times used in this work (\sim 5 μ s in

total) should therefore provide a reasonable benchmark for the simulations in the field in general.

Importantly, we did several control simulations to back up the validity of this conclusion. We tested how the order of doing the structural modifications to get back to the pristine protein affecting the results. Our simulations showed that the order of steps does not affect the main results of Method S to a significant degree. We also tested the influence of cholesterol and found that it does not play a significant role in the comparison between Methods A and S. We also tested that the preparation of the loop structures in the model systems did not affect the results much. Finally, the restraints used for the protein were in this work released in a standard way during a time scale of 50 ns. A more gradual release of restraints might improve the results slightly.

Our results also provided a considerable amount of knowledge regarding the effects of individual structural modifications, such as mutations and the use of antibodies in protein structure. The rigid T4L structure was fused between TMS-TM6 in experiment to replace the inherently flexible ICL3. T4-lysozyme also restricted the dynamics of TMS and TM6, which is known to be important for receptor activation and responsible for the open structure of G-protein binding site observed in the crystal structure of β_2 AR. The replacement of T4L with ICL3 relatively closed the G-protein binding site, as the intracellular end of TM6 moved 4 Å toward the core of the helical bundle. Meanwhile, removal of timolol resulted in relative opening of the ligand-binding cavity. The changes observed in the distal ligand- and G-protein binding sites accounted for the allosteric regulation in GPCRs.

The reformation of the R131^{3,50}-E268^{6,38} salt bridge, which was initially broken in the crystal structure of inactive β_2 AR, has been reported in previous computational studies.^{10,52} Dror et al. described the carazolol-bound β_2 AR conformation with an intact R131^{3,50}-E268^{6,38} salt bridge as an alternative inactive conformation of β_2 AR, distinct from the experimentally reported structure.¹⁰ This study, however, did not focus on the change at the ligand-binding site.¹⁰ For comparison, in most of the previous studies, the receptor models used in simulations were modified to some extent. In many of these studies, after removing T4L, instead of incorporating ICL3 the newly exposed ends at the cytoplasmic side of TMS and TM6 were either left open^{7,57,58} or clipped to each other via a peptide bond^{10,52} (termed a "clipped model"). A recent work by Ozcan et al. showed that ICL3 plays a significant role in the intrinsic conformational dynamics of β_2 AR and that the clipped receptor model exhibited very different dynamical behavior than the model with ICL3.¹¹ Additionally, in most of these simulations, the N187^{5,26}E mutation (as found in the crystal structure) was kept unaltered.^{7,10,11,52,58} The purpose of the present work was to model the receptor as closely as possible to its biologically relevant form and also to characterize how much the modifications made in crystallization can actually affect the intrinsic structural properties of the native receptor.

Concluding, our results emphasize how sensitively membrane protein structure and dynamics depend on the preparation of the protein whose properties one would like to explore through atomistic simulations. If the preparation is done in a straightforward manner by cutting corners, then the outcome of the simulations can be partly artificial, not corresponding to the true native state under equilibrium conditions. This threat is quite possible especially in short simulations where the protein has not enough time to relax.

Given that many of the receptor structural changes take place over long time scales (ranging from microseconds to much larger times), and considering that this is particularly the case for receptors such as GPCRs where structural changes typically take place through the entire protein, there is reason to warrant great care in starting atomistic simulations of membrane proteins whose crystal structure determination has required considerable structural modifications. Altogether, we consider that the protocol described here (Method S) may offer a useful strategy for simulating a variety of native state receptors, whose crystal structures suffer from similar modifications.

■ ASSOCIATED CONTENT

Supporting Information

Further results for the protein structures. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00070.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: Ilpo.Vattulainen@tut.fi.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

Reinis Danne is thanked for technical support at the early stage of this work. CSC – Finnish IT Centre for Scientific Computing (Espoo, Finland) is acknowledged for computer resources granted through the Grand Challenge project concept. European Research Council (Advanced Grant project CROWDED-PRO-LIPIDS) and the Academy of Finland Center of Excellence program are thanked for financial support.

■ REFERENCES

- (1) Lefkowitz, R. J. Seven transmembrane receptors - A brief personal retrospective. *Biochim. Biophys. Acta, Biomembr.* **2007**, *1768*, 748–755.
- (2) Lefkowitz, R. J. The superfamily of heptahelical receptors. *Nat. Cell Biol.* **2000**, *2*, E133–E136.
- (3) Seddon, A. M.; Curnow, P.; Booth, P. J. Membrane proteins, lipids and detergents: Not just a soap opera. *Biochim. Biophys. Acta, Biomembr.* **2004**, *1666*, 105–117.
- (4) Kobilka, B. The structural basis of G-protein-coupled receptor signaling (Nobel Lecture). *Angew. Chem., Int. Ed.* **2013**, *52*, 6380–6388.
- (5) Tate, C. G. A crystal clear solution for determining G-protein-coupled receptor structures. *Trends Biochem. Sci.* **2012**, *37*, 343–352.
- (6) Poiry, S.; Cramariuc, O.; Postila, P. A.; Kaszuba, K.; Sarewicz, M.; Osyczka, A.; Vattulainen, I.; Rog, T. Atomistic simulations indicate cardiolipin to have an integral role in the structure of the cytochrome bc₁ complex. *Biochim. Biophys. Acta, Bioenergetics* **2013**, *1827*, 769–778.
- (7) Dror, R. O.; Arlow, D. H.; Maragakis, P.; Mildorf, T. J.; Pan, A. C.; Xu, H.; Borhani, D. W.; Shaw, D. E. Activation mechanism of the β₂-adrenergic receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 18684–18689.
- (8) Li, Y. Y.; Hou, T. J.; Goddard, W. A., III Computational modeling of structure-function of G protein-coupled receptors with applications for drug design. *Curr. Med. Chem.* **2010**, *17*, 1167–1180.
- (9) *G-Protein Coupled Receptors – Modeling and Simulation. Series: Advances in Experimental Medicine and Biology*; Filizola, M. G., Ed.; Springer: 2014; Vol. 796.
- (10) Dror, R. O.; Arlow, D. H.; Borhani, D. W.; Jensen, M. Ø.; Piana, S.; Shaw, D. E. Identification of two distinct inactive conformations of the β₂-adrenergic receptor reconciles structural and biochemical observations. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 4689–4694.
- (11) Ozcan, O.; Uyar, A.; Doruker, P.; Akten, E. D. Effect of intracellular loop 3 on intrinsic dynamics of human β₂-adrenergic receptor. *BMC Struct. Biol.* **2013**, *13*, 29.
- (12) Grossfield, A. Recent progress in the study of G protein-coupled receptors with molecular dynamics computer simulations. *Biochim. Biophys. Acta, Biomembr.* **2011**, *1808*, 1868–1878.
- (13) Hanson, M. A.; Cherezov, V.; Griffith, M. T.; Roth, C. B.; Jaakola, V.-P.; Chien, E. Y. T.; Velasquez, J.; Kuhn, P.; Stevens, R. C. A specific cholesterol binding site is established by the 2.8 Å structure of the human β₂-adrenergic receptor. *Structure* **2008**, *16*, 897–905.
- (14) Rasmussen, S. G. F.; Choi, H.-J.; Rosenbaum, D. M.; Kobilka, T. S.; Thian, F. S.; Edwards, P. C.; Burghammer, M.; Ratnala, V. R. P.; Sanishvili, R.; Fischetti, R. F.; Schertler, G. F. X.; Weis, W. I.; Kobilka, B. K. Crystal structure of the human β₂ adrenergic G-protein-coupled receptor. *Nature* **2007**, *450*, 383–U384.
- (15) Cherezov, V.; Rosenbaum, D. M.; Hanson, M. A.; Rasmussen, S. G. F.; Thian, F. S.; Kobilka, T. S.; Choi, H.-J.; Kuhn, P.; Weis, W. I.; Kobilka, B. K.; Stevens, R. C. High-resolution crystal structure of an engineered human β₂-adrenergic G protein-coupled receptor. *Science* **2007**, *318*, 1258–1265.
- (16) Rasmussen, S. G. F.; Choi, H.-J.; Fung, J. J.; Pardon, E.; Casarosa, P.; Chae, P. S.; DeVree, B. T.; Rosenbaum, D. M.; Thian, F. S.; Kobilka, T. S.; Schnapp, A.; Konetzki, I.; Sunahara, R. K.; Gellman, S. H.; Pautsch, A.; Steyaert, J.; Weis, W. I.; Kobilka, B. K. Structure of a nanobody-stabilized active state of the β(2) adrenoceptor. *Nature* **2011**, *469*, 175–180.
- (17) Rasmussen, S. G. F.; DeVree, B. T.; Zou, Y.; Kruse, A. C.; Chung, K. Y.; Kobilka, T. S.; Thian, F. S.; Chae, P. S.; Pardon, E.; Calinski, D.; Mathiesen, J. M.; Shah, S. T. A.; Lyons, J. A.; Caffrey, M.; Gellman, S. H.; Steyaert, J.; Skiniotis, G.; Weis, W. I.; Sunahara, R. K.; Kobilka, B. K. Crystal structure of the β₂ adrenergic receptor-Gs protein complex. *Nature* **2011**, *477*, 549–U311.
- (18) Rog, T.; Vattulainen, I. Cholesterol, sphingolipids, and glycolipids: What do we know about their role in raft-like membranes? *Chem. Phys. Lipids* **2014**, *184*, 82–104.
- (19) Neuvonen, M.; Manna, M.; Mokkila, S.; Javanainen, M.; Rog, T.; Liu, Z.; Bittman, R.; Vattulainen, I.; Ikonen, E. Enzymatic oxidation of cholesterol: Properties and functional effects of cholestenone in cell membranes. *PLoS One* **2014**, *9*, e103743.
- (20) Manna, M.; Mukhopadhyay, C. Cholesterol driven alteration of the conformation and dynamics of phospholamban in model membranes. *Phys. Chem. Chem. Phys.* **2011**, *13*, 20188–20198.
- (21) Paila, Y. D.; Chattopadhyay, A. The function of G-protein coupled receptors and membrane cholesterol: Specific or general interaction? *Glycoconjugate J.* **2009**, *26*, 711–720.
- (22) Zocher, M.; Zhang, C.; Rasmussen, S. G. F.; Kobilka, B. K.; Müller, D. J. Cholesterol increases kinetic, energetic, and mechanical stability of the human β₂-adrenergic receptor. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 3463–3472.
- (23) Kulig, W.; Tynkkynen, J.; Javanainen, M.; Manna, M.; Rog, T.; Vattulainen, I.; Jungwirth, P. How well does cholestryly hemisuccinate mimic cholesterol in saturated phospholipid bilayers? *J. Mol. Model.* **2014**, *20*, 2121.
- (24) Javanainen, M. Universal method for embedding proteins into complex lipid bilayers for molecular dynamics simulations. *J. Chem. Theory Comput.* **2014**, *10*, 2577–2582.
- (25) Ballesteros, J. A.; Weinstein, H. Integrated methods for the construction of three dimensional models and computational probing of structure function relations in G protein-coupled receptors. *Methods Neurosci.* **1995**, *25*, 366–428.
- (26) The UniProt Consortium. Activities at the universal protein resource (UniProt). *Nucleic Acids Res.* **2014**, *42*, D191–198.
- (27) Sali, A.; Blundell, T. L. Comparative protein modeling by satisfaction of spatial restraints. *J. Mol. Biol.* **1993**, *234*, 779–815.
- (28) Jaakola, V. P.; Prilusky, J.; Sussman, J. L.; Goldman, A. G protein-coupled receptors show unusual patterns of intrinsic unfolding. *Protein Eng., Des. Sel.* **2005**, *18*, 103–110.
- (29) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. Development and testing of the OPLS all-atom force field on conformational

- energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.
- (30) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.
- (31) Maciejewski, A.; Pasenkiewicz-Gierula, M.; Cramariuc, O.; Vattulainen, I.; Rog, T. Refined OPLS all-atom force field for saturated phosphatidylcholine bilayers at full hydration. *J. Phys. Chem. B* **2014**, *118*, 4571–4581.
- (32) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (33) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges - the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.
- (34) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (35) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*.
- (36) Parrinello, M.; Rahman, A. Polymorphic transitions in single-crystals - A new molecular-dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (37) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (38) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
- (39) Lehtonen, J. V.; Still, D. J.; Rantanen, V. V.; Ekholm, J.; Bjorkland, D.; Iftikhar, Z.; Huhtala, M.; Repo, S.; Jussila, A.; Jaakkola, J.; Pentikainen, O.; Nyronen, T.; Salminen, T.; Gyllenberg, M.; Johnson, M. S. BODIL: A molecular modeling environment for structure-function analysis and drug design. *J. Comput.-Aided Mol. Des.* **2004**, *18*, 401–419.
- (40) Kabsch, W.; Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **1983**, *22*, 2577–2637.
- (41) Armen, R.; Alonso, D. O. V.; Daggett, V. The role of alpha-3(10)-, and pi-helix in helix → coil transitions. *Protein Sci.* **2003**, *12*, 1145–1157.
- (42) Lazarova, T.; Brewin, K. A.; Stoeber, K.; Robinson, C. R. Characterization of peptides corresponding to the seven transmembrane domains of human adenosine A2a receptor. *Biochemistry* **2004**, *43*, 12945–12954.
- (43) Corin, K.; Baaske, P.; Ravel, D. B.; Song, J.; Brown, E.; Wang, X.; Geissler, S.; Wienken, C. J.; Jerabek-Willemsen, M.; Duhr, S.; Braun, D.; Zhang, S. A Robust and rapid method of producing soluble, stable, and functional G-protein coupled receptors. *PLoS One* **2011**, *6*, e23036.
- (44) Kaszuba, K.; Rog, T.; Bryl, K.; Vattulainen, I.; Karttunen, M. Molecular dynamics simulations reveal fundamental role of water as factor determining affinity of binding of β-blocker nebivolol to β₂-adrenergic receptor. *J. Phys. Chem. B* **2010**, *114*, 8374–8386.
- (45) Strader, C. D.; Sigal, I. S.; Candelore, M. R.; Rands, E.; Hill, W. S.; Dixon, R. A. F. Conserved aspartic acid residues 79 and 113 of the β-adrenergic-receptor have different roles in receptor function. *J. Biol. Chem.* **1988**, *263*, 10267–10271.
- (46) Suryanarayana, S.; Kobilka, B. K. Amino acid substitutions at position 312 in the 7th hydrophobic segment of the beta2-adrenergic receptor modify ligand-binding specificity. *Mol. Pharmacol.* **1993**, *44*, 111–114.
- (47) Wieland, K.; Zuurmond, H. M.; Krasel, C.; Ijzerman, A. P.; Lohse, M. J. Involvement of Asn-293 in stereospecific agonist recognition and in activation of the β₂-adrenergic receptor. *Proc. Natl. Acad. Sci. U.S.A.* **1996**, *93*, 9276–9281.
- (48) Strader, C. D.; Candelore, M. R.; Hill, W. S.; Sigal, I. S.; Dixon, R. A. F. Identification of 2 serine residues involved in agonist activation of the β-adrenergic-receptor. *J. Biol. Chem.* **1989**, *264*, 13572–13578.
- (49) Katritch, V.; Reynolds, K. A.; Cherezov, V.; Hanson, M. A.; Roth, C. B.; Yeager, M.; Abagyan, R. Analysis of full and partial agonists binding to β₂-adrenergic receptor suggests a role of transmembrane helix V in agonist-specific conformational changes. *J. Mol. Recognit.* **2009**, *22*, 307–318.
- (50) Rosenbaum, D. M.; Cherezov, V.; Hanson, M. A.; Rasmussen, S. G. F.; Thian, F. S.; Kobilka, T. S.; Choi, H.-J.; Yao, X.-J.; Weis, W. I.; Stevens, R. C.; Kobilka, B. K. GPCR engineering yields high-resolution structural insights into β₂-adrenergic receptor function. *Science* **2007**, *318*, 1266–1273.
- (51) Liggett, S. B.; Caron, M. G.; Lefkowitz, R. J.; Hnatowich, M. Coupling of a mutated form of the human β₂-adrenergic receptor to G_i and G_s. *J. Biol. Chem.* **1991**, *266*, 4816–4821.
- (52) Romo, T. D.; Grossfield, A.; Pitman, M. C. Concerted interconversion between ionic lock substates of the β₂ adrenergic receptor revealed by microsecond timescale molecular dynamics. *Biophys. J.* **2010**, *98*, 76–84.
- (53) Ballesteros, J. A.; Jensen, A. D.; Liapakis, G.; Rasmussen, S. G. F.; Shi, L.; Gether, U.; Javitch, J. A. Activation of the β₂-adrenergic receptor involves disruption of an ionic lock between the cytoplasmic ends of transmembrane segments 3 and 6. *J. Biol. Chem.* **2001**, *276*, 29171–29177.
- (54) Roth, C. B.; Hanson, M. A.; Stevens, R. C. Stabilization of the human β₂-adrenergic receptor 4–3–5 helix interface by mutagenesis of Glu-122^{3,41}, a critical residue in GPCR structure. *J. Mol. Biol.* **2008**, *376*, 1305–1319.
- (55) Miallet-Perez, J.; Green, S. A.; Miller, W. E.; Liggett, S. B. A primate-dominant third glycosylation site of the β₂-adrenergic receptor routes receptors to degradation during agonist regulation. *J. Biol. Chem.* **2004**, *279*, 38603–38607.
- (56) Rands, E.; Candelore, M. R.; Cheung, A. H.; Hill, W. S.; Strader, C. D.; Dixon, R. A. F. Mutational analysis of β-adrenergic-receptor glycosylation. *J. Biol. Chem.* **1990**, *265*, 10759–10764.
- (57) Cang, X.; Du, Y.; Mao, Y.; Wang, Y.; Yang, H.; Jiang, H. Mapping the functional binding sites of cholesterol in β₂-adrenergic receptor by long-time molecular dynamics simulations. *J. Phys. Chem. B* **2013**, *117*, 1085–1094.
- (58) Xiao, X.; Zeng, X.; Yuan, Y.; Gao, N.; Guo, Y.; Pu, X.; Li, M. Understanding the conformation transition in the activation pathway of β₂ adrenergic receptor via a targeted molecular dynamics simulation. *Phys. Chem. Chem. Phys.* **2014**, *17*, 2512–2522.