Position of Transmembrane Helix 6 Determines Receptor G Protein Coupling Specificity

Alexander S. Rose,†‡ Matthias Elgeti,† Ulrich Zachariae,§ Helmut Grubmüller,‖ Klaus Peter Hofmann,†┴ Patrick Scheerer,†,# and Peter W. Hildebrand†‡

†Institut für Medizinische Physik und Biophysik, †AG ProteiInformatics, and ‡AG Protein X-ray Crystallography, Charité Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany
§Division of Computational Biology, College of Life Sciences, and Division of Physics, School of Engineering, Physics and Mathematics, University of Dundee, Dow Street, Dundee DD1 5EH, U.K.
‖Department of Theoretical and Computational Biophysics, Max-Planck-Institute for Biophysical Chemistry, D-37077 Göttingen, Germany
┴Centre of Biophysics and Bioinformatics, Humboldt-Universität zu Berlin, Invalidenstrasse 42, D-10115 Berlin, Germany

Supplemental Information Figures

Table S1

<table>
<thead>
<tr>
<th>System</th>
<th>Simulations</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>RhR*</td>
<td>1x 400 ns, 4x 200 ns</td>
<td>1.2 µs</td>
</tr>
<tr>
<td>RhR*·GtaCT&lt;sub&gt;19&lt;/sub&gt;</td>
<td>5x 200 ns</td>
<td>1 µs</td>
</tr>
<tr>
<td>RhR*·GtaCT&lt;sub&gt;11&lt;/sub&gt;</td>
<td>1x 400 ns, 4x 200 ns</td>
<td>1.2 µs</td>
</tr>
<tr>
<td>RhR*·GsaCT&lt;sub&gt;19&lt;/sub&gt;</td>
<td>5x 200 ns</td>
<td>1 µs</td>
</tr>
<tr>
<td>β&lt;sub&gt;2&lt;/sub&gt;AR*</td>
<td>1x 400 ns, 4x 200 ns</td>
<td>1.2 µs</td>
</tr>
<tr>
<td>β&lt;sub&gt;2&lt;/sub&gt;AR*·Gsaβγ</td>
<td>5x 200 ns</td>
<td>1 µs</td>
</tr>
<tr>
<td>β&lt;sub&gt;2&lt;/sub&gt;AR*·GsaCT&lt;sub&gt;19&lt;/sub&gt;</td>
<td>1x 400 ns, 4x 200 ns</td>
<td>1.2 µs</td>
</tr>
<tr>
<td>β&lt;sub&gt;2&lt;/sub&gt;AR*·GsaCT&lt;sub&gt;11&lt;/sub&gt;</td>
<td>10x 500 ns</td>
<td>5 µs</td>
</tr>
<tr>
<td>β&lt;sub&gt;2&lt;/sub&gt;AR*·GiaCT&lt;sub&gt;19&lt;/sub&gt;</td>
<td>3x 600 ns, 8x 200 ns, 10x 100 ns</td>
<td>4.4 µs</td>
</tr>
</tbody>
</table>

Overview of performed MD simulations. Each row lists the system, the length and count of simulations, and the overall per system simulation time. The total simulation time is 17.2 µs.
(A) Superposition of the RhR*-GtαCT crystal structure (orange, PDB entry 3PQR) with a snapshot from the MD simulation of uncomplexed β2AR* (red). Note the similarity of the TM6 tilt at the intracellular side. The grey box denotes the part of TM6 analyzed in B. (B) Time series showing the secondary structure assignment of the intracellular TM6 region of β2AR* as calculated with DSSP (W. Kabsch, C. Sander, Biopolymers 1983, 22, 2577–2637). The observed structure types are α-helix (blue), 3_{10}-helix (grey), turn (yellow), coil (white) or bend (green). The black arrow denotes from which simulation and at what time the β2AR* structure shown in A was taken. The plot shows the data of multiple MD simulations, each between 200 and 400 ns long.
Figure S2

TM6 tilt measured as the TM2-TM6 distances of (A) $\beta_2$AR*-Gsαβγ, (B) $\beta_2$AR*-GsαCT$_{11}$, (C) RhR*-GtαCT$_{19}$, (D) RhR*-GtαCT$_{11}$, (E) $\beta_2$AR*-GtαCT$_{19}$ and (F) RhR*. Shown are the time traces of the TM2-6 distances and the resulting distributions with observed probabilities $p$. The lengths of the individual simulations are given in Table S1. Note that N-terminal truncation to GsαCT$_{11}$ leads to higher observed variability of TM6 tilts compared to $\beta_2$AR*-GsαCT$_{19}$ or $\beta_2$AR*-Gsαβγ, especially within individual simulations. This effect is less articulate for RhR*•GtαCT$_{11}$ compared to GtαCT$_{19}$. 

S3
GsαCT and ICL3 flexibility of various complexes. Shown are the backbone root mean square fluctuations (RMSF) changes of GsαCT relative to β₂AR*·Gsαβγ of (A) β₂AR*·GsαCT₁₁ or (B) β₂AR*·GsαCT₁₉. (C) Tube depiction of β₂AR* with GsαCT₁₁ (blue), with GsαCT₁₉ (green) or with GsαCT from Gsαβγ (yellow). The thickness of the tube corresponds to the RMSF value of the respective amino acids. The RMSF is the standard deviation of atomic positions over simulation time and was calculated with the GROMACS tool g_rmsf from the MD simulations of the respective systems as listed in Table S1. Before the analysis each frame was superposed with the backbone atoms of TM 1-7 of the MD starting structure.
Figure S4

Cation-π interactions between β₂AR* and (A) Gsαβγ, (B) GsαCT₁₉, (C) GsαCT₁₁ and between (D) RhR* and GsαCT₁₉. Each plot shows the data of multiple MD simulations, each between 200 and 500 ns long. See Table 1 for the lengths of the individual simulations. The timeseries show the electrostatic (black) and hydrophobic (red, Lenard-Jones potential) contributions to the interaction energy as calculated by the CaPTURE program (J. P. Gallivan, D. a Dougherty, Proc Natl Acad Sci USA 1999, 96, 9459–64.). Blue lines denote frames where the interaction energy of cation-π interaction between R₃,5₀ and C391 is significant, which occurs continuously in the β₂AR* simulations with the exceptions of simulation 3 in (B) and simulations 3, 4, 6 in (C). The cation-π interaction between RhR* and GsαCT₁₉ seen in simulation 3 (D) exhibits no face-to-face orientation as in β₂AR* but rather an edge-to-face orientation.
Polar interactions between R* and GaCT. Each plot shows the data of multiple MD simulation, each between 100 and 600 ns long. The lengths of the individual simulations are given in Table S1. The timeseries denote one (black), two (blue) or three (green) polar interactions between two residues as observed in the MD simulations of (A) $\beta_2$AR*-Gsaβγ, (B) $\beta_2$AR*-GsaCT$_{19}$mer, (C) $\beta_2$AR*-GsaCT$_{11}$mer, (D) $\beta_2$AR*-GiaCT$_{19}$mer, (E) RhR*-GtaCT$_{19}$mer and (F) RhR*-GtaCT$_{11}$mer. Polar interactions between GaCT and R* were calculated with the GROMACS tool g_hbond using a donor-acceptor distance cutoff at 3.6 Å and a cutoff angle of 30° for the angle given by the acceptor-donor-hydrogen atoms. In (D) only those simulations of $\beta_2$AR*-GiaCT$_{19}$mer are shown where TM6 changes its starting position and moves inward.
(A) Structuring of ICL3 that accompanies TM6 inward movement (see Fig. 2C) in the MD simulation of β2AR*-GiαCT19 used in the Umbrella Sampling. (B) Rotation of TM6 around its axis (at residues 273-280, grey box) observed in the same simulation.
(A) Model of $\beta_2$AR$^*$-Gi$\alpha$ determined by classical and Umbrella Sampling MD simulation data and (B) the crystal structure of $\beta_2$AR$^*$-Gs$\alpha$ (for comparison). Apparently there are no clashes with the membrane (black lines) nor within the complex itself. As in the $\beta_2$AR-Gs holo complex, the arrangement in $\beta_2$AR-Gi does not result in any major clashes (as shown with the inactive GDP bound Gi$t$ in P. Scheerer et al., Nature 2008, 455, 497–502.). Comparison of both complexes, however, reveals a difference in the rotational tilt of the GoCT relative to $\beta_2$AR, resulting in a slightly different orientation of Gi$\alpha$ relative to the receptor. (C) Comparison of the transmembrane helix arrangements in $\beta_2$AR$^*$-Gi$\alpha$CT (red), RhR$^*$-GtaCT (orange) and $\beta_2$AR$^*$-Gs (blue).
Per residue interaction area fraction of the GsαCT\textsubscript{19} peptides observed in MD simulations of (A) β\textsubscript{2}AR\textsuperscript{*}-GsαCT\textsubscript{19}, (B) RhR\textsuperscript{*}-GtαCT\textsubscript{19} and (C) β\textsubscript{2}AR\textsuperscript{*}-GiaCT\textsubscript{19}. The interaction area fraction is calculated as the fraction of solvent accessible surface (SAS, calculated with the GROMACS tool g_sas) of the R\textsuperscript{*}-GaCT complex and the SAS of the free peptide. A fraction of one means a residue is completely accessible whereas completely buried residues have a fraction of zero. Residues with a fraction below 0.35 are denoted buried and those with a fraction above 0.7 are denoted accessible. In (C) only those simulations of β\textsubscript{2}AR\textsuperscript{*}-GiaCT\textsubscript{19mer} are shown where TM6 changes its starting position and moves inward.
Alignment of $\alpha$CT$_{19}$ sequences colored by similarity.
Supplemental Information Methods

Preparations of active receptor (R*) and GαCT structures and complexes

The starting conformations used for MD simulations were prepared based on X-ray structures from co-crystals of β2AR*-Gsαβγ (PDB entry 3SN6) [1] and of RhR*-GαCT (PDB entry 3PQR) [2]. The far C-termini of both R* structures (RhR*: residues 327 to 348, UniProt entry P02699; β2AR*: 342-413, UniProt entry P07550), not resolved in these complexes, were not modeled because they seem not to affect Gt activation [3].

For all simulations of β2AR*, the coordinates from the agonist bound β2AR*-Gsαβγ complex (PDB entry 3SN6), with the T4-lysozyme removed from the N-terminus, were used. A palmitoyl chain was ligated to C341 of R*. Unresolved atoms from the side chains of residues 63, 97-99, 101, 149, 175, 192-195, 267, 269-272, 299, 301-302, 304, 306 and 333 were added applying the standard geometries from the Dunbrack 2002 library [4]. Three stabilizing mutants (M96T, M98T and N187E) in β2AR* were changed back to the wild-type form. The coordinates for the missing residues of the extracellular loop (ECL) 2 (176-178) were taken from the β2AR* structure (PDB entry 3P0G) where ECL 2 is resolved [5]. The conformation of residues 240 to 264 from the intracellular loop (ICL) 3, which are not critical to receptor function [6], were modeled with help of the fragment based loop modeling program SuperLooper [7].

Gsαβγ was prepared as follows. The missing Gsα N-terminal residues 1-8 were modeled using standard geometries before a palmitoyl chain was ligated to C3 and G2 [8]. A geranylgeranyl chain was ligated to residue 68 of the Gsγ-subunit [9], after the missing residues 1-4 of the N-terminus and 63-68 of the C-terminus were added. The mutated residues G72S in Gsα and M1Q in the Gsβ were changed back to the wild-type form. Unresolved atoms from the side chains of residues 24, 35, 58, 59, 94, 118, 136, 139, 188, 189, 191, 194, 195, 201, 216, 240, 300, 322, 369 in Gsα, 1, 42, 129, 130, 172 in Gsβ and 62 in Gsγ were added using standard geometries from the Dunbrack 2002 library [4]. The conformation of the missing residues 60-70, 85-87, 203-204 and 256-262 in Gsα, were again modeled with SuperLooper [7].
For all MD simulations of RhR* two palmitoyl chains were attached to the residues C322 and C323. The coordinates from the double high-affinity K341L, C347V peptide variant in complex with RhR* (PDB entry 3PQR) were used. After back mutation of these two residues, RhR* in complex with native 11-mer GtαCT was obtained (340-350). For simulations of 19-mer GtαCT (332-350) the 11-mer GtαCT was extended N-terminally by 8 amino acids using the geometries of an ideal α-helix. For simulations of 11-mer GsαCT (residues 384-394) and 19-mer GsαCT (residues 376-394) the coordinates from the β2AR*-Gsαβγ complex were used.

**Protonation states and internal water**

The C-termini of GsαCT, GtαCT, GiαCT, RhR* and β2AR* were deprotonated (COO−), whereas the N-termini were fully protonated (NH3+). In RhR*, D83 [10,11], E113 [12], E122 [11] and E134 [13] were protonated. In β2AR*, E122 was protonated, because it is in close contact with the hydrophobic lipid tails in the middle of the lipid bilayer (as suggested by Dror et al. [14]). All other protonation states were defined according to their respective pKa values (provided by GROMACS).

Empty polar water sized internal cavities were filled with water molecules by means of the program DOWSER [15].

**Preparation of the β2AR*-GiαCT complex**

The β2AR*-GiαCT19 complex was created from MD simulations based on the crystal structure complexes of β2AR*-Gsαβγ (PDB entry 3SN6) and RhR*-GtαCT (PDB entry 3PQR). The GiαCT19 starting position within the β2AR* binding crevice was obtained from a superposition with the RhR*-GtαCT complex. The superposition was guided by a sequence alignment of β2AR and RhR* and employed the Ca atoms from TM1-5 and TM7. GiαCT was obtained from GtαCT by changing I338 to V. With this starting position, no contacts are formed with TM6. The peptide is initially attached to ICL2 and TM5, but allowed to move freely within the R* binding crevice. In the simulations where TM6 tilts inwards (Fig. 2C, S2E) the cytoplasmic crevice closes and a tight interaction is formed with TM6 (Fig. 3B, S5D).
**Preparation of the RhR*-GsαCT complex**

The GsαCT starting position for the RhR*-GsαCT complex was obtained following the same sequence alignment and superposition protocol as for β2AR*-GiαCT but with RhR* as the receptor and GsαCT as the peptide target.

**Molecular dynamics protocol**

System preparation and subsequent minimization and equilibration were performed with the GROMACS suite (version 4.5) \[^{[16]}\]. The proteins were inserted into the equilibrated bilayer of dimyristoylphosphatidylcholine (DMPC) using the GROMACS g_membed tool \[^{[17]}\]. Parameters for the DMPC lipids were derived from Berger et al. \[^{[18]}\] and for water from the SPC/E model \[^{[19]}\]. A salt concentration of 0.15 mol/L was obtained by adding Na\(^+\) and Cl\(^-\) ions to the system with the GROMACS tool genion. The AMBER99SB-ILDN force field \[^{[20]}\] was used for proteins and ions. Ligand parameters for the agonist 5-hydroxy-4H-benzoxazin-3-one (a.k.a. BI-167107) of β2AR* were created with the PRODRG2 webserver \[^{[21]}\]. Parameters for the deprotonated all-trans retinal in RhR* were adapted from Mertz et al. \[^{[22]}\]

To obtain clash-free structures suitable for MD simulations, an energy minimization was performed in GROMACS using the steepest descent algorithm until the maximum force went below 1000.0 kJ/mol/nm. In the following equilibration step the energy minimized structure was simulated for 20 ns with all protein backbone atoms restrained to their initial positions. This allows for relaxation at the protein-membrane, protein-water and the membrane-water interfaces so that voids are filled and side chain packing is optimized. For the production MD simulations the position restraints were lifted.

Based on the equilibrated systems, the production runs were started with different initial velocities obtained from Boltzmann distributions at 320 K. For equilibration and the production runs all bonds were constrained using the LINCS algorithm \[^{[23]}\], with the exception of water bonds, which were constrained by the SETTLE algorithm \[^{[24]}\]. The temperature was kept constant by coupling the system to a temperature bath of 320 K, which is high enough to keep the DMPC membrane from entering the gel phase. The temperature coupling was performed using the velocity-rescaling thermostat of Bussi et al. \[^{[25]}\] with a time constant of 0.2 ps. Long range electrostatics
were calculated with the PME method. Berendsen pressure coupling was performed with a time constant of 2.0 ps and semi-isotropic scaling separating scaling in the membrane plane directions from the z-direction (i.e. the membrane plane normal). The integration time step used for all simulations was 0.002 ps.

**Umbrella sampling**

Umbrella sampling (US) facilitates sampling of the conformational space by applying a restraining potential along a transition coordinate. By employing umbrella sampling over a series of windows a range of the transition coordinates can be sampled which would be inaccessible to direct sampling due to energy barriers of the transition coordinate. The resulting series of histograms contains the biased distribution along the transition coordinate. The weighted histogram analysis method is employed to unbias and combine the histograms. From the resulting distribution the potential of mean force can be calculated as $\text{PMF}(c) = -k_B T \ln \langle p(c) \rangle$ for the probability $p$ of the transition coordinate $c$.

Here, the transition coordinate for the free energy calculations of TM6 inward movement and $\beta_2\text{AR}^* \cdot \text{Gi} \cdot \alpha_{\text{CT}}$ interaction was selected from the trajectories of a series of $\beta_2\text{AR}^* \cdot \text{Gi} \cdot \alpha_{\text{CT}}$ MD simulations (Fig. S2E, simulation 8). Along the selected trajectory, umbrella sampling MD simulations were performed with respect to the TM6 inward transition, by applying the umbrella potential to the upper part of TM6, namely to the backbone atoms of residues 265 to 277. We simulated 36 US windows for 200 ns, each. WHAM was then employed to obtain the PMFs from the last 100 ns from each US window and the error was estimated by the standard deviation of a block-wise (three equally sized blocks) analysis.

**TM6 tilt**

The distance between TM2 and 6 ($d_{\text{TM2-6}}$) was used as an indicator of the TM6 tilt. It is measured as the distance between the geometric centers for intracellular sections of TM2 and TM6. For TM2 we used the backbone atom positions of the residues 71-75 (RhR*) and 67-71 ($\beta_2\text{AR}^*$); for TM6 244-248 and 265-269, respectively.

**References**
