Supplementary Information

Large scale investigation of GPCR molecular dynamics data uncovers allosteric sites and lateral gateways

David Aranda-García[1,2], Tomasz Maciej Stepniewski[2,3], Mariona Torrens-Fontanals[1,4], Adrian García-Recio[2], Marta Lopez-Balastegui[1,2], Brian Medel-Lacruz[1,2], Adrián Morales-Pastor[1,2], Alejandro Peralta-García[1], Miguel Dieguez-Eceolaza[1,2], David Sotillo-Nuñez[1], Tianyi Ding[5], Matthäus Drabek[6], Célien Jacquemard[7], Jakub Jakowiecki[8,9], Willem Jespers[10], Mireia Jiménez-Rosés[11,12], Víctor Jun-Yu-Lim[6], Alessandro Nicoli[13], Urszula Orzel[8,9], Aida Shahraki[6], Johanna K. S. Tiemann[14], Vicente Ledesma-Martin[1], Francho Nerín-Fonz[1], Sergio Suárez-Dou[1], Oriol Canal[1], Gáspár Pándy-Szekeres[15,16], Jiafei Mao[17], David E. Gloriam[15], Esther Kellenberger[7], Dorota Latek[8], Ramon Guixà-González[18], Hugo Gutiérrez-de-Terán[19,20], Irina G. Tikhonova[5], Peter W. Hildebrand[21], Marta Filizola[22], M. Madan Babu[23], , Antonella Di Pizio[13,24], Slawomir Filipek[8], Peter Kolb[6], Arnau Cordomi[25], Toni Giorgino[26,27], Maria Marti-Solano[28], Jana Selent[1,2]*

[1]Department of Experimental and Health Sciences, Pompeu Fabra University (DCEXS-UPF), Barcelona, Spain
[2]Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Medical Research Institute (IMIM), Barcelona, Spain
[3]InterAx Biotech AG, PARK InnovAARE 5234 Villigen Switzerland.
[4]Acellera Labs, C Dr Trueta 183, 08005, Barcelona, Spain
[5]School of Pharmacy, Medical Biology Centre, Queen's University Belfast, BT9 7BL Northern Ireland, UK
[6]Pharmaceutical Chemistry, University of Marburg, Marbacher Weg 8, Marburg 35037, Germany
[7]Laboratoire d'Innovation Thérapeutique, University of Strasbourg, Strasbourg, France
[8]Faculty of Chemistry, University of Warsaw, Warsaw, Poland
[9]Biological and Chemical Research Centre, University of Warsaw, Warsaw, Poland
[10]Division of Drug Discovery and Safety, Leiden Academic Centre for Drug Research, Leiden University, Leiden, The Netherlands
[11]Centre of Membrane Proteins and Receptors (COMPARE), University of Birmingham and University of Nottingham, Midlands, UK
[12]Sygnature Discovery Ltd., BioCity, Pennyfoot Street, Nottingham NG1 1GR, UK
[13]Leibniz Institute for Food Systems Biology at the Technical University of Munich, 85354 Freising, Germany
[14]Universität Leipzig, Medizinische Fakultät, Institut für Medizinische Physik und Biophysik, Leipzig, Germany
[15]Department of Drug Design and Pharmacology, University of Copenhagen, Universitetsparken 2, 2100 Copenhagen, Denmark
[16]Medicinal Chemistry Research Group, HUN-REN Research Center for Natural Sciences, Magyar tudósok körútja 2., Budapest H-1117, Hungary

[17]*Beijing National Laboratory for Molecular Sciences (BNLMS) and Center for Physicochemical Analysis and Measurement, Institute of Chemistry Chinese Academy of Science (ICCAS), Beijing, China*
[18]*Department of Biological Chemistry, Institute for Advanced Chemistry of Catalonia (IQAC-CSIC), 08034, Barcelona, Spain*
[19]*Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, 75124Uppsala, Sweden*
[20]*Research Center in Nanomaterials and Nanotechnology (CINN/CSIC) and Health Institute of Principado de Asturias (ISPA), Av. del Hospital Universitario, 33011 Oviedo, Asturias, Spain*
[21]*Institute of Medical Physics and Biophysics, Medical University Leipzig, Leipzig, Sachsen, Germany*
[22]*Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA*
[23]*Department of Structural Biology and Center of Excellence for Data Driven Discovery, St. Jude Children's Research Hospital, Memphis, TN 38105, USA*
[24]*Chemoinformatics and Protein Modelling, School of Life Science, Technical University of Munich, 85354 Freising, Germany*
[25]*Bioinformatics, ESCI - Pompeu Fabra University, Barcelona, Spain*
[26]*Institute of Biophysics (IBF-CNR), National Research Council of Italy, Milano, Italy*
[27]*Department of Biosciences, University of Milan, Italy*
[28]*Department of Pharmacology, University of Cambridge, Cambridge CB2 1PD, UK*

**corresponding author*
Jana Selent: jana.selent@upf.edu

Table of content

Supplementary Notes

Supplementary Note 1. Co-solved insertion recovery in A_{2A} receptors

In experimentally solved (static) structures of the A_2 adenosine receptors, two sites (TM1-TM7$_{EC}$ and TM3-TM4$_{IC}$) present insertions in more than 30% of all structures of this type (12 of 31 and 11 of 31, respectively) (**Figure 3c**, top). Remarkably, both pockets also present very high lipid insertion rates in MD simulations (**Figure 3c**, bottom), even in systems not presenting static insertions in these sites. We thus decided to analyse in detail a case study for this receptor subtype: 6AQF, where both static and MD lipid insertions can be observed for either site (**Supplementary Figure 3**).

The first insertion, occurring in TM1-TM7$_{EC}$, is already discussed in the main text. Therefore, we focus here on the one occurring in TM3-TM4 IC, described previously in A_{2A} receptors as the "intracellular crevice"[1]. The static insertion found here has been reported in many other A_{2A} structures (12 out of 31 in our set), classified as "shallow" in all of them. As in the previous case, recovery rates in MD simulations are very high in all systems (50.72% of all A_{2A} simulated time), regardless of the presence of an original static insertion. In this particular case, the lipid seems to be making intermittent hydrogen bonds with Arg111 and Arg120^{4x41}, something reported previously in other A_{2A} MD studies[2]. This pocket has been previously identified for A_{2A} and A_{1A} receptors as a binding site for cholesterol molecules[3] and seems to have a very high affinity for fatty acids as well[4]. Taken together, these results emphasise the importance of this pocket in the functionality of adenosine receptors.

Supplementary Note 2. Co-solved insertion recovery in Cannabinoid B1 receptors (CB_1R)

To further delve into the biological relevance of lipid insertions, both static and MD ones, we decided to analyse them in another case (**Supplementary Figure 6**). The selected system is a CB_1R (PDB Id: 5XRA; GPCRmd ids: 896, 897) with two significant lipid insertions in TM1-TM7 EC and TM3-TM4 IC. These are the same sites where insertions were found in the previous $A_{2A}R$ case study. However, in this case, MD recovery rates, insertion types, and possible receptor roles are not the same.

Region TM1-TM7 EC presents a lipid insertion with high insertion frequency (36.4%) in MD simulations, but no equivalent appears in the static structure. Nevertheless, this region has been previously reported as the entrance channel for at least two orthosteric ligands (THC, anandamide) in CB_1Rs[5]. Two residues, F174^{2x61} and F177^{2x64}, appear to act as barriers separating this site from the orthosteric binding site, and impending its access by the lipid in our simulations. These residues have been reported by Jakowiecki and Filipek in previous works as the gatekeepers for orthosteric ligands to access the binding site. Therefore, this replicated insertion might be another case of a ligand entrance gateway identified through lipid insertions. A similar insertion was found in another cannabinoid simulation, from the structure 5XR8.

The second insertion, found in site TM3-TM4 IC, is also commonly observed in $A_{2A}R$ and other receptors. Surprisingly, in this case, we observe a "deep" in this site and not a shallow one. This insertion is also recovered from a significantly deep static insertion found in the

5XRA crystal structure, but not in any of the other 4 CB_1R structures we simulated. In MD simulations, recovery of this insertion in CB_1R is very frequent (68.37%). Values in the particular case of 5XRA are even higher, with 92.09% of all frames presenting an insertion here. This is mostly caused because it occurs first during the equilibration, and remains stable through all replicates. Surprisingly, no allosteric site seems to be reported in this region for CB_1R[6], although there are reports for other receptors such as the M2 Muscarinic[7]. Allosteric sites have been found, however, for a neighbouring pocket (between TM1, TM2, and TM4 near its intracellular side) working as an allosteric binding site. It is tempting thus to speculate a potential allosteric function for our identified pocket in CB_1R as well.

Supplementary Note 3. Lateral ligand gateways for β_2-adrenoceptor revealed by lipid insertions

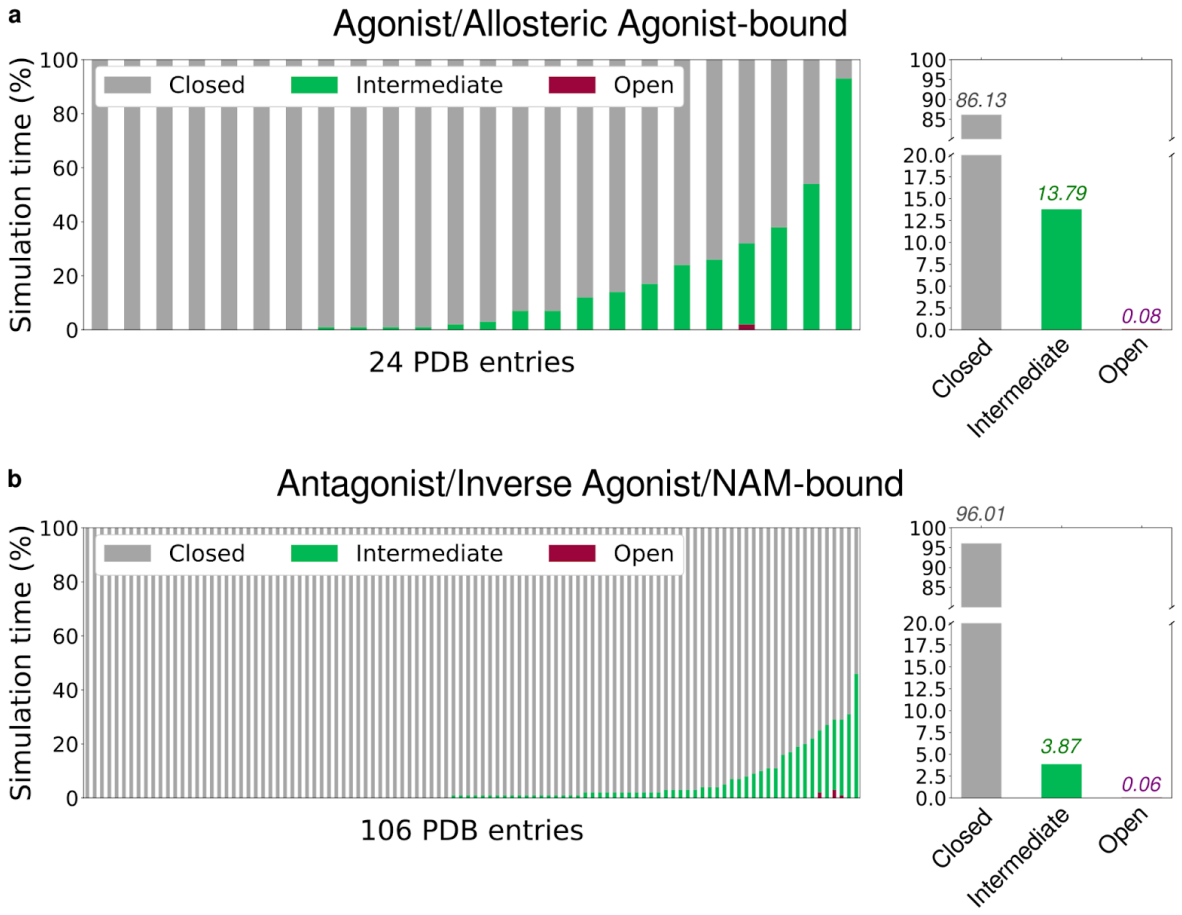
Numerous agonists exhibit affinity for the β_2 adrenergic receptor (B2AR), notable instances being salbutamol, salmeterol, and formoterol, used in asthma treatment. Despite having a similar epinephrine-mimicking saligenin headgroup, salmeterol and formoterol also feature bulky hydrophobic moieties not present in salbutamol. Remarkably, acute pharmacological differences also exist between the latter two with salbutamol, with the former pharmacological effects extending up to 12 hours compared to the latter 4 hours[8]. Previous literature[9] reports salmeterol and formoterol binding mechanisms as being partially from the membrane, instead of directly from the extracellular environment. It is thus tempting to suggest a correlation between both factors.

As we reported in the results, this is also the site where lipid insertions more frequently occur in our simulation dataset. This is also the case for the 15 B2AR systems present in our dataset, where insertions in TM1-TM7 are also highly frequent. Particularly remarkable is the case of system 5X7D (**Supplementary Figure 5**) with very deep insertions occurring in this site, sometimes even reaching the orthosteric site cavity. Similar incidences can also be found in B2AR systems 6PS0 and 6OBA. Such deep insertions however are infrequent, occurring only in one of the 6 replicates in each system.

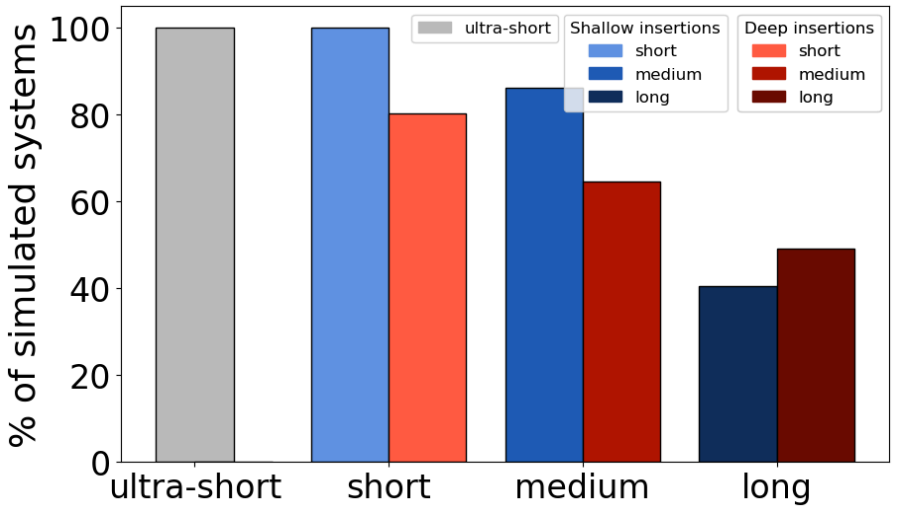
The identified lipid insertion channel seems to be similar in all reported cases, with $Trp313^{7x39}$ being its main gatekeeper residue. Intriguingly, MD simulations run by Szlenk et al. in their work[10] identify this very same channel as the one used by salmeterol to first anchor its hydrophobic moiety and afterwards move into the orthosteric site. A similar case reported in this work is formoterol, which uses this same site as an anchorage point but ligand entry occurs at a higher position.

In summary, despite the inherently hydrophilic nature of the receptor's endogenous ligand, lipid insertions allow us to identify a prominent channel reportedly involved in the binding of relevant agonist drugs.

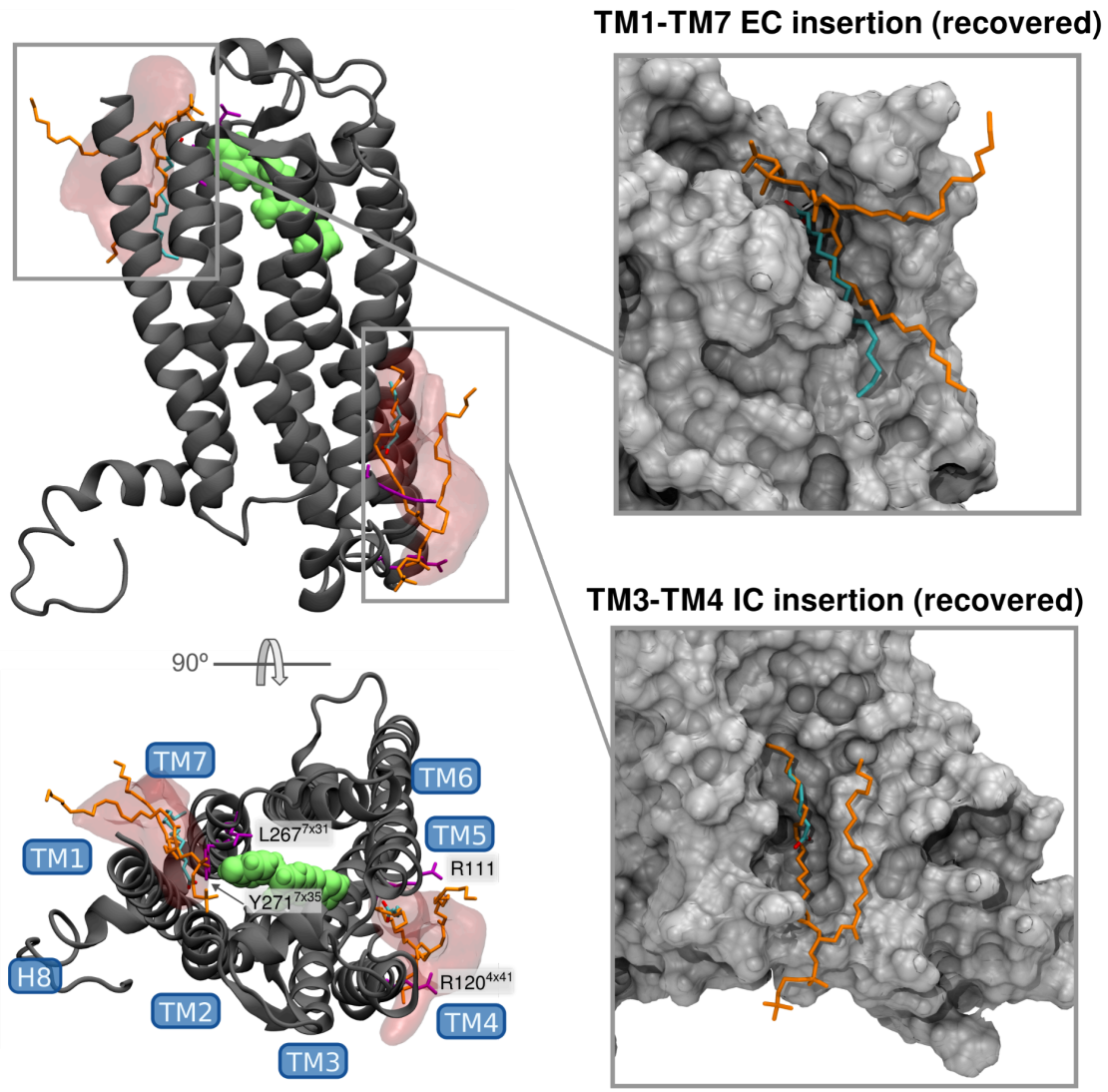
Supplementary figures



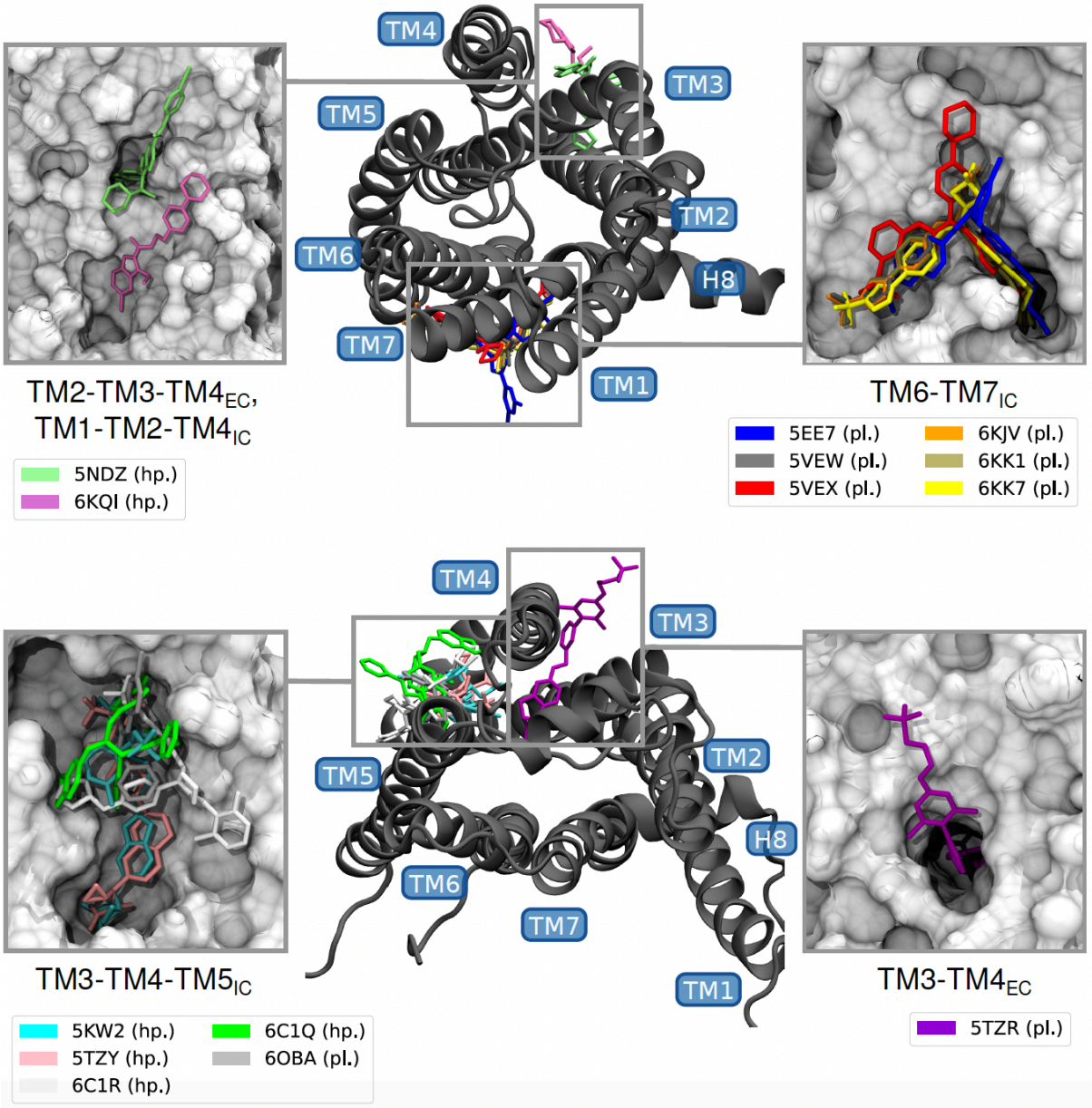
Supplementary Figure 1 | Receptor core flexibility of agonists or allosteric agonists (24 PDBs) versus antagonists, inverse agonists, or NAMs (106 PDBs). Systems started from initially closed (inactive) states belonging to classes A or B1. **a,** Percentage of simulation time spent in each state by initially inactive systems simulated in complex to an agonist or allosteric agonist ligand. Percentages are represented for each system individually (left) or cumulatively for the whole subset (right). **b,** Same concept, applied to systems bound to antagonists, inverse agonists or NAMs.



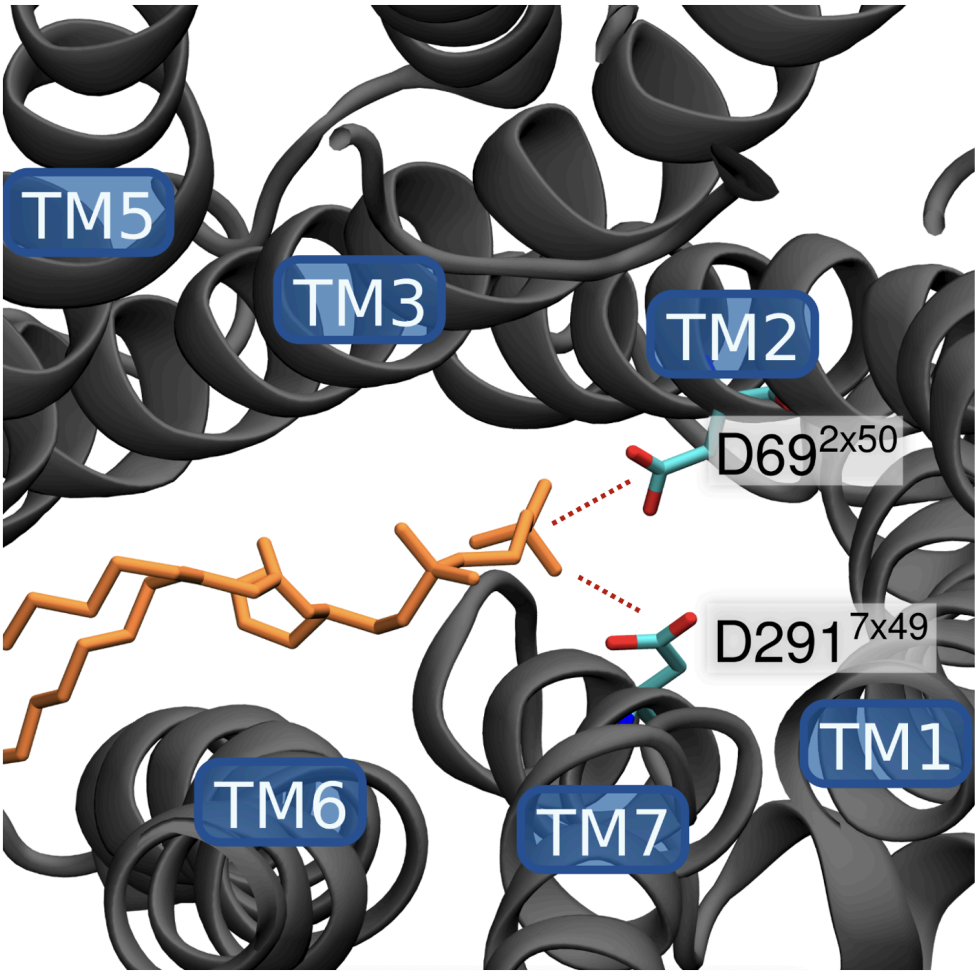
Supplementary Figure 2 | Presence of different types of lipid insertions in the simulated systems. In our simulation dataset (n = 371 simulations), we observe that systems with shallow lipid insertions of short (100%) and medium (86.25%) duration are highly abundant, whereas long-lasting insertions are less present (40.43%). In contrast, deep insertions are less abundant at short (80.32%) and medium (64.69%) durations but more frequently observed at long-lasting time intervals (49.06%) compared to shallow ones. Importantly, the abundance of observed lipid insertions across all receptors and classes suggests a relevant role in GPCR modulation.



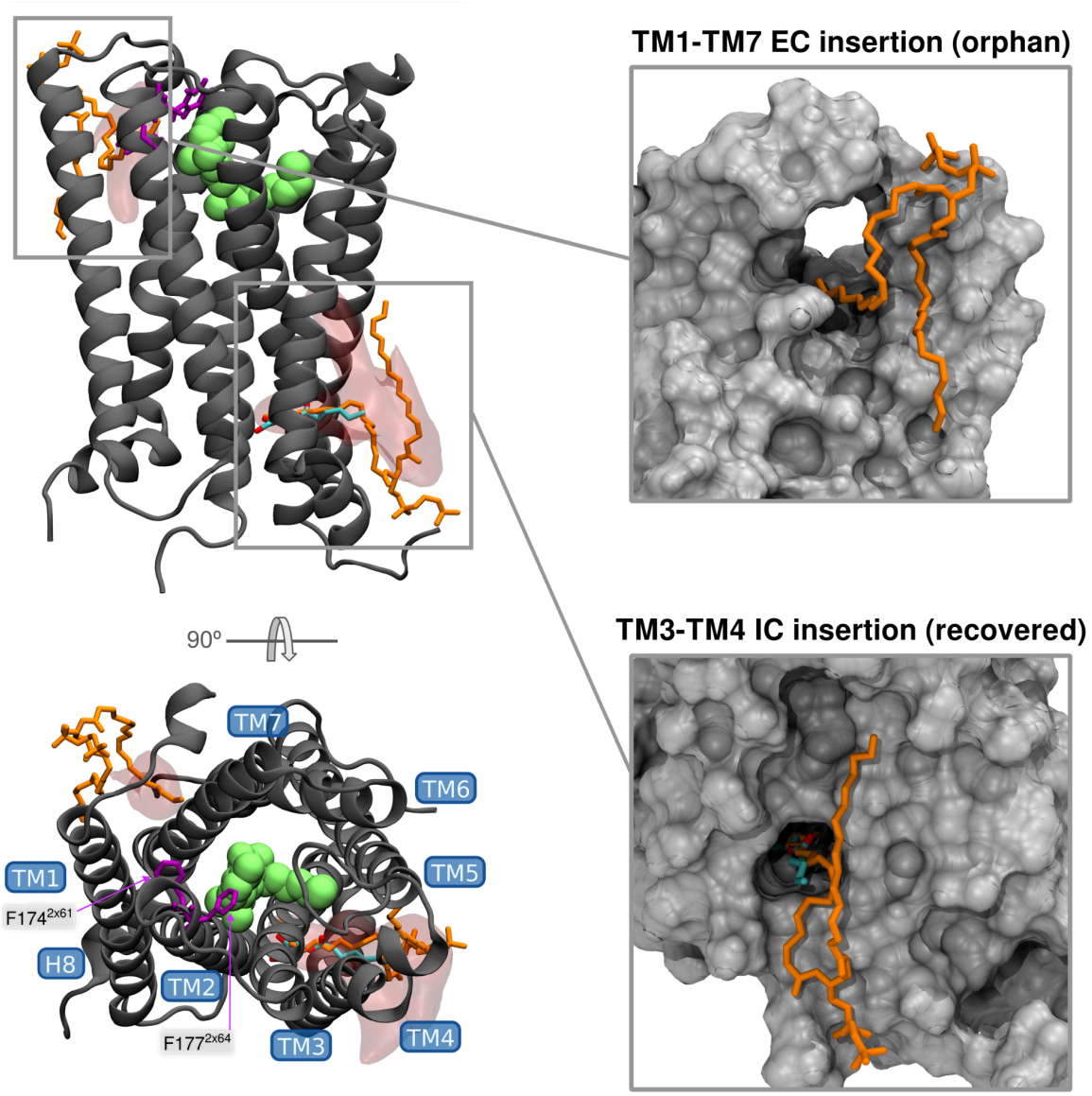
Supplementary Figure 3 | Recovery of co-solved lipid insertions in an adenosine A2A receptor structure in complex with antagonist ZM241385. Both co-solved insertions (cyan, licorice) are replicated by POPC membrane lipids in this system's simulations (orange), with their density maps (red) covering most of the pocket. Gatekeeper residues L267^{7x31} and Y271^{7x35} and lipid-interacting residues R111 and R120^{4x41} are represented with purple licorice. PDB id: 6AQF; GPCRmd id: 754, trajectory id: 15462.



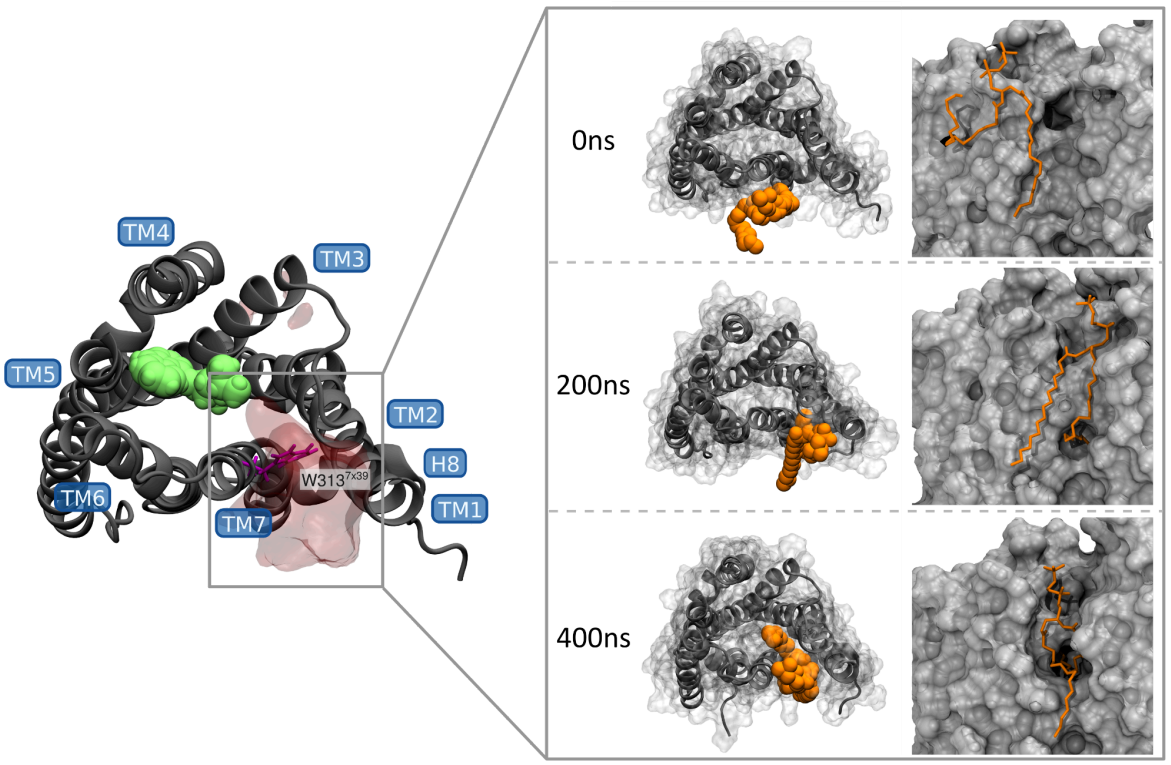
Supplementary Figure 4 | Allosteric modulators and their binding sites. Fourteen allosteric modulators present in our simulated dataset (licorice) are superimposed over a reference structure (grey cartoon). The corresponding PDBs are indicated together with the polarity of the ligand binding pocket (pl: polar, hp: hydrophobic). For visualization purposes, we used two different GPCR structures as references: a GCGR (PDB id: 5EE7) for top panel and a β_2-AR (PDB id: 6OBA) for bottom panel. Ligands of different structures are represented with different colours, identified in the legends with their corresponding PDB id.



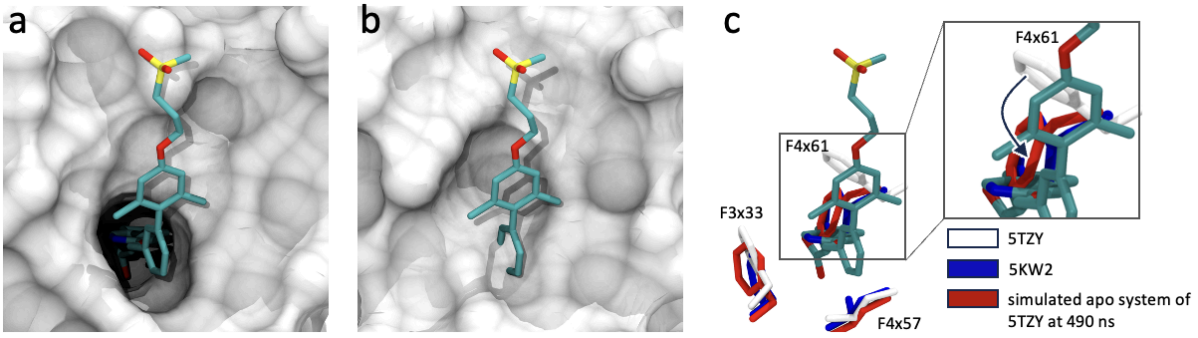
Supplementary Figure 5 | Lipid insertion into the cysteinyl leukotriene receptor 1. A deep and very stable lipid insertion (orange licorice) is observed in all apo form trajectories of the cysteinyl leukotriene receptor 1 (gray cartoon). The insertion occurs at TM5-TM6$_{IC}$, and is stabilised by interactions between the inserted POPC's positively charged choline head and two deprotonated aspartate residues, D69^{2x50} and D291^{7x49} (cyan licorice). PDB id: 6RZ4, GPCRmd id: 738.



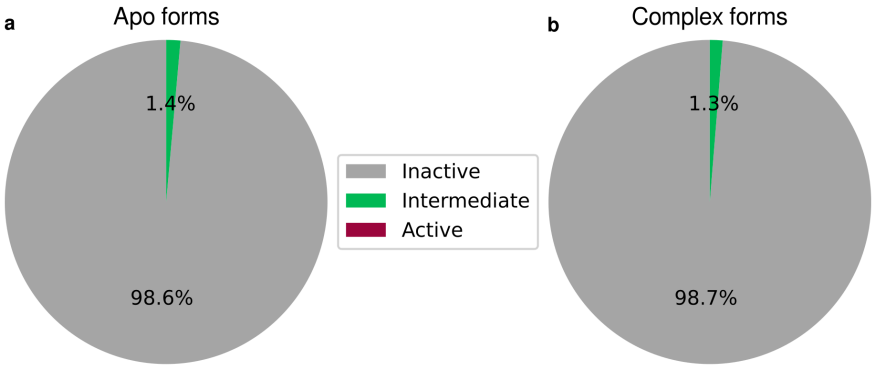
Supplementary Figure 6 | Recovery of co-solved lipid insertions in a CB1 receptor structure in complex with agonist AM11542. Two significant insertions (licorice orange) were detected in this structure's simulations: one between TM3-TM4, recovering an experimental lipid insertion found in the original structure (licorice, cyan), and another one between TM1-TM7, not recovered from any experimental insertion. Gatekeeper residues F174^{2x61} and F177^{2x64} are represented in purple licorice. PDB id: 5XRA; GPCRmd id: 896; trajectory id: 16421.



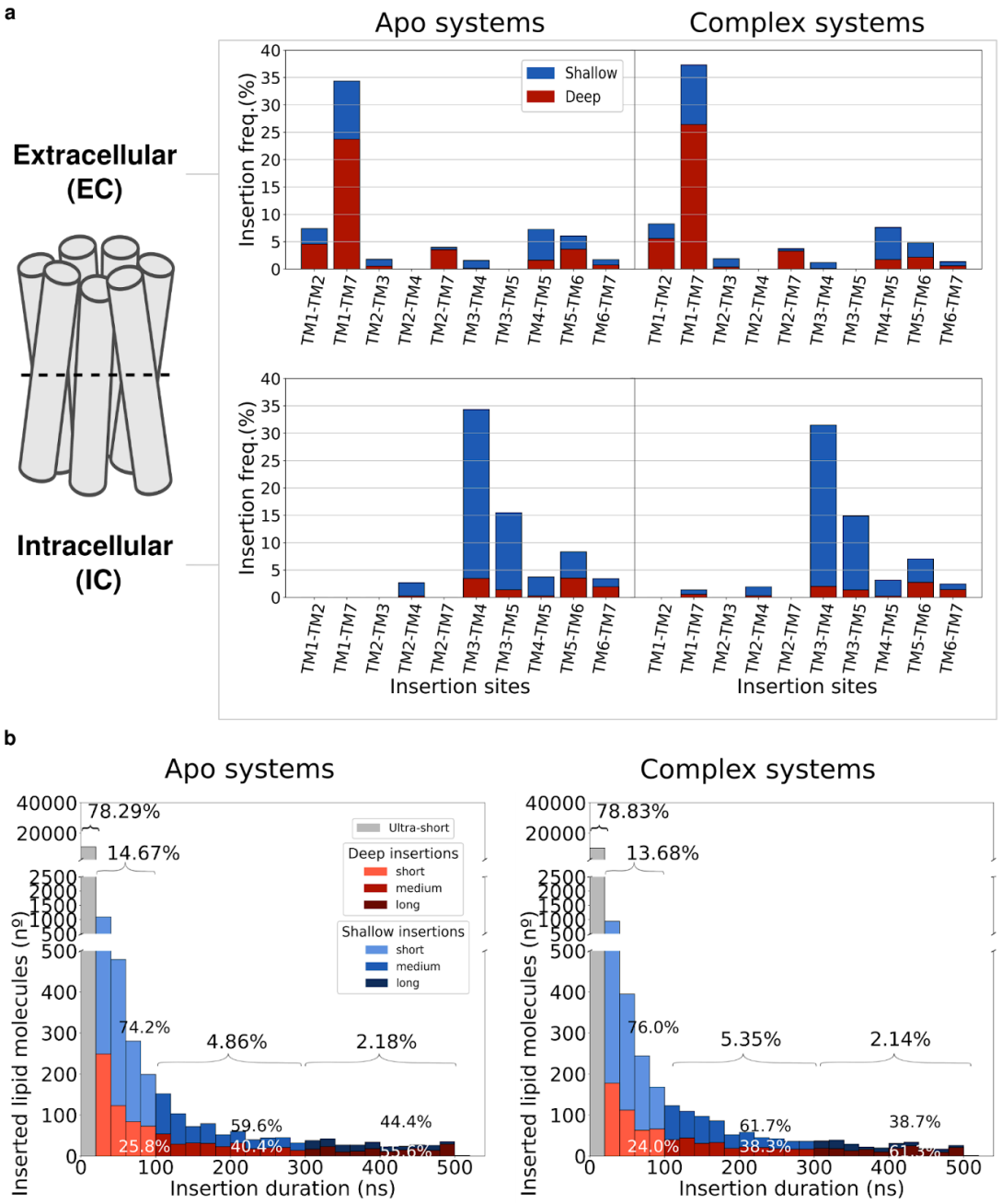
Supplementary Figure 7 | Case study of a deep lipid insertion in TM1-TM7$_{EC}$ in an apo simulation of a β_2-adrenoceptor system. The original crystal structure was co-solved with the inverse agonist carazolol (van-der-waals, lime) occupying the orthosteric binding site. In one of our simulations, a deep lipid insertion occurs (in orange, van-der-waals), managing to reach the orthosteric ligand cavity in the later stages of the simulation (400ns, bottom-right). Gatekeeper residue W313^{7x39} is marked in purple licorice. PDB id: 5X7D; GPCRmd id: 769; trajectory id: 15567.



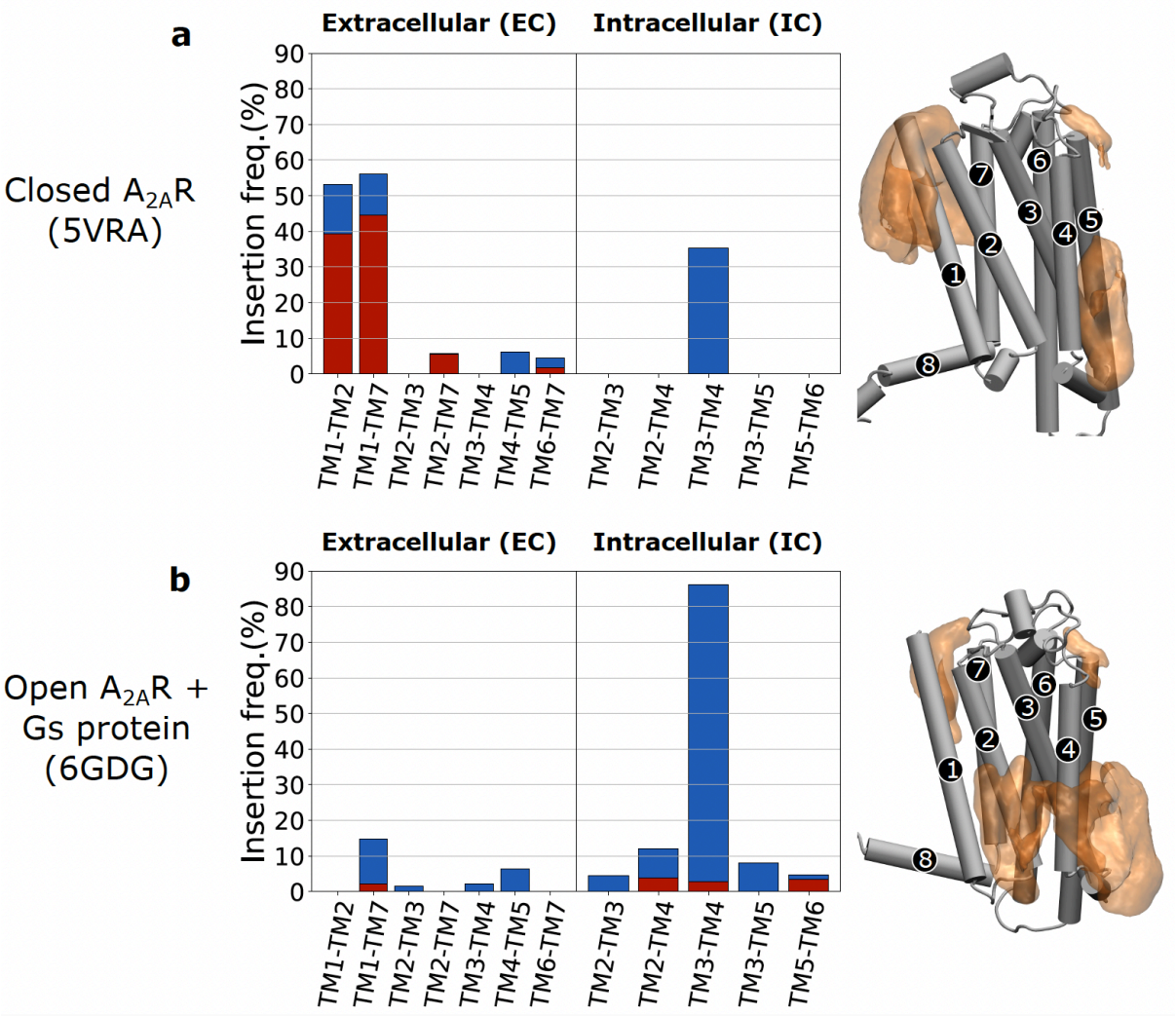
Supplementary Figure 8 | Different conformational states of an allosteric site in the FFA1R. **a**, Experimental FFA1R structure (PDB id: 5TZY): an open pocket state is stabilized by a molecular modulator at TM3-TM4$_{EC}$. **b**, Experimental FFA1R structure (PDB id: 5KW2): a closed pocket at TM3-TM4$_{EC}$ is adopted when the ligand is absent. The molecular modulator from 5TZY is superimposed to indicate the pocket location. **c**, Superimposition of the experimental FFA1R structures (PDB ids: 5TZY and 5KW2) with the simulated apo form (PDB id: 5TZY, GPCRmd id: 928). A snapshot at 490 ns of replicate 2 is shown as a representative. The rotation of the F4x61 closes the pocket in the experimental structure 5KW2 (blue) and the simulated apo form (red) compared to the experimental structure 5TZY (white).



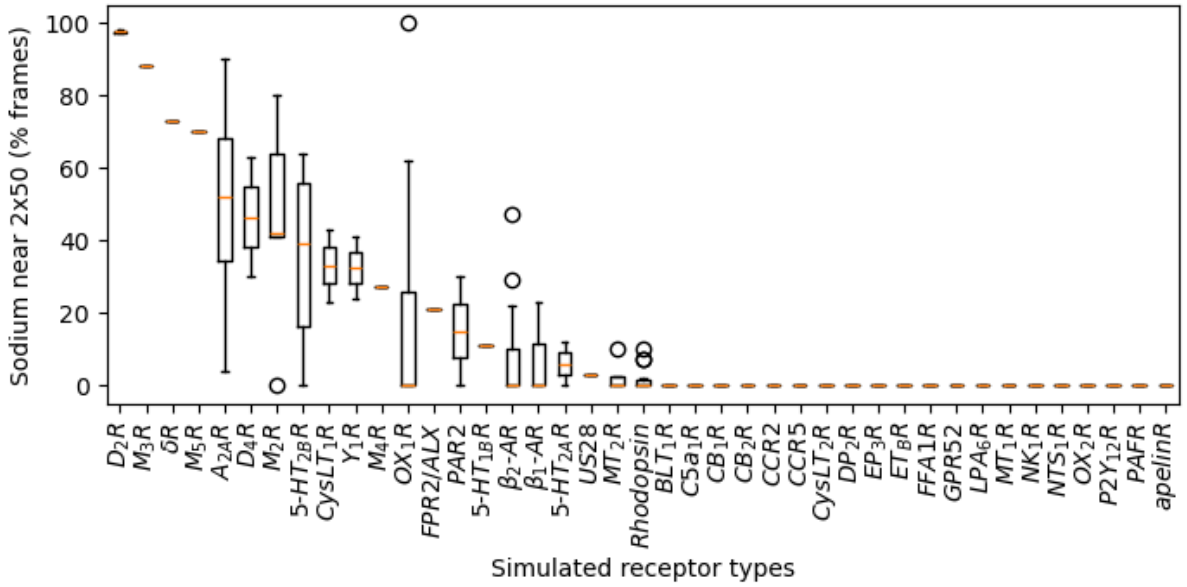
Supplementary Figure 9 | Classification of inactive, intermediate and active states for class A receptors according to the Activation Index A100. **a,** Conformational classification in apo systems (i.e. ligand was removed from original PDB, n = 100 PDB structures) and **b**, complex systems co-solved with an antagonist, inverse agonist or NAM ligand; and identified as inactive by GPCRdb (n = 100 PDB structures). The overall percentage for active, intermediate and inactive states was obtained by computing the activation state for each frame (i.e., each 0.2 ns) in the 3 x 500 ns simulation data across all 100 simulation systems. This data indicates that GPCRs sample at times intermediate states but no active states based on the Activation Index A100.



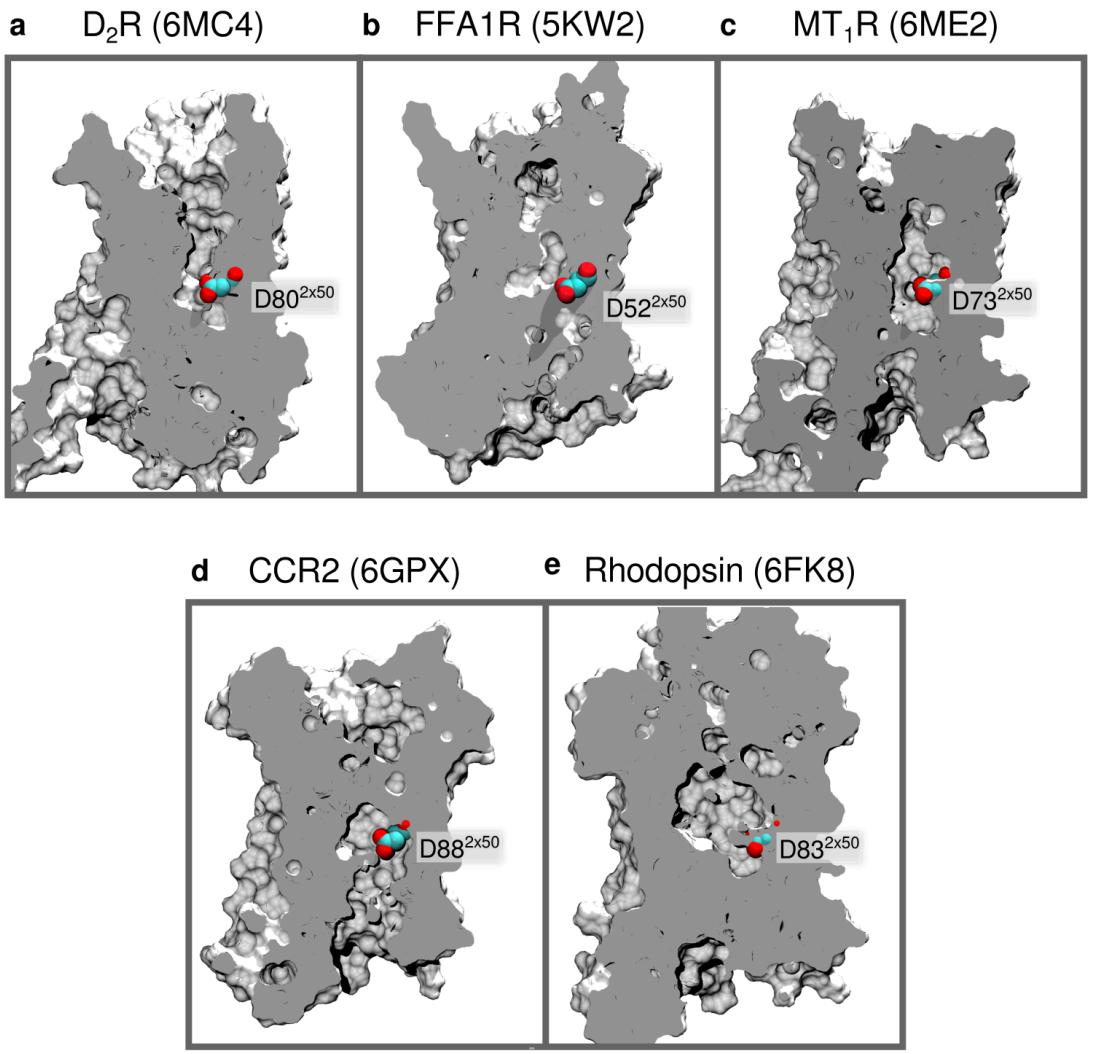
Supplementary Figure 10 | Comparative distribution of lipid insertions between apo and complex form systems. a, Frequency of lipid insertions indicating TM helices and GPCR side (either Extracellular (EC) or Intracellular (IC)). Frequency values represent dataset-wide averages for each pocket in apo (left) and complex (right) systems. Insertions are classified as deep (red) or shallow (blue) (see **Figure 3** for further details on insertion criteria) **b**, Comparative quantification of lipid insertions in the dataset apo and complex simulations. Insertions are classified according to depth (blue vs red) and duration in simulation (color intensity). (n = 190 apo simulations, n = 181 complex simulations).



Supplementary Figure 11 | Lipid insertion for inactive and active adenosine A_{2A} receptors ($A_{2A}R$). **a,** Lipid insertion in a closed (inactive) $A_{2A}R$ (PDB id: 5VRA, GPCRmd id: 1011). Left: Lipid insertion frequencies at specific receptor sites (red: deep, blue: shallow). Right: structural depiction of the $A_{2A}R$, including an occupancy map that highlights lipid insertion sites occurring with at least 5% frequency during the simulation. **b,** Lipid insertion for an open (active) $A_{2A}R$ (PDB id: 6GDG, GPCRmd id: 1967). Left: Lipid insertion frequencies at specific receptor sites (red: deep, blue: shallow). Right: structural depiction of the $A_{2A}R$ including an occupancy map that highlights lipid insertion sites occurring with at least 5% frequency during the simulation.



Supplementary Figure 12 | Percentage of sodium binding at the 2x50 residue in apo systems. Systems are grouped according to their receptor type (x-axis). In the y-axis is represented the percentage of simulated frames, in each individual system, where a sodium ion is found near 5 Å of its 2x50 residue. Only class A systems are included. The center line in boxplots represents the median while the box boundaries extend from the first (25%) and third (75%) quartile, representing the interquartile range (IQR). Boxplot whiskers extend to 1.5*IQR and outliers are represented as circles. The simulated PDB structures belonging to each receptor type (e.g. $A_{2A}R$) in the plot are listed in **Supplementary Data 1**.



Supplementary Figure 13 | Cross section for 5 different GPCR types, showing the accessibility of the allosteric sodium binding site 2x50 residues (van der Waals). **a**, The Dopamine D2 (D_2R) receptor with the highest rate of ion binding (see **Supplementary Figure 12**) shows an open entrance and channel towards D^{2x50}. In contrast, **b,** FFA1R, **c,** MT1R, **d,** CCR2 and **e,** rhodopsin with low or null ion sodium-2x50 binding rates reveal a closed entrance and/or disrupted channel towards D^{2x50}.

Supplementary tables

Supplementary Table 1. Number of transition events observed, transition rate, and the average time it takes for each transition, along with their respective confidence intervals. The differences between the apo and complex forms provide insights into how the presence of a ligand influences the kinetics of GPCR transitions.

State Transition	Form	Transitions	Transition Rate (1/μs)	95% CI for Rate	Transition Time (μs)	95% CI for Transition Time
Inactive to Intermediate	Apo	179	2.0	1.7 - 2.3	0.5	0.4 - 0.6
Inactive to Intermediate	Complex	101	0.85	0.69 - 1.0	1.2	1.0 - 1.4
Inactive to Active	Apo	20	0.13	0.08 - 0.20	7.8	5.1 - 12.8
Inactive to Active	Complex	3	0.02	0.00 - 0.06	52.7	18.0 - 256.0

Supplementary Table 2. Molecules identified in our 2nd simulation round dataset as crystallisation adjuvants, listed by their chemical component (3-letter code) ID and name as they appear in the RCSB database. When searching for static lipid insertions in our dataset, these were the molecules we were looking to find inserted in the receptor.

Ligand code (in rcsb.org)	Molecule name (in rcsb.org)
OLA	Oleic acid
OLB	(2s)-2,3-dihydroxypropyl (9z)-octadec-9-enoate
OLC	(2r)-2,3-dihydroxypropyl (9z)-octadec-9-enoate
PLM	Palmitic acid
LPP	2-(hexadecanoyloxy)-1-[(phosphonooxy)methyl]ethyl hexadecanoate
PEF	Di-palmitoyl-3-sn-phosphatidylethanolamine
2CV	Hega-10
TWT	Docosane
STE	Stearic acid
LMT	Dodecyl-beta-d-maltoside
MPG	[(z)-octadec-9-enyl] (2r)-2,3-bis(oxidanyl)propanoate
DGA	Diacyl glycerol
1WV	(2s)-2,3-dihydroxypropyl (7z)-tetradec-7-enoate
POV	(2s)-3-(hexadecanoyloxy)-2-[(9z)-octadec-9-enoyloxy]propyl 2-(trimethylammonio)ethyl phosphate
PGW	(1r)-2-{[(s)-{[(2s)-2,3-dihydroxypropyl]oxy}(hydroxy)phosphoryl]oxy}-1-[(hexadecanoyloxy)methyl]ethyl (9z)-octadec-9-enoate
PC1	1,2-diacyl-sn-glycero-3-phosphocholine
LDA	Lauryl dimethylamine-n-oxide
J40	[(2r)-1-[oxidanyl-[(2r,3r,5s,6r)-2,3,5,6-tetrakis(oxidanyl)-4-phosphonooxy-cyclohexyl]oxy-phosphoryl]oxy-3-tetradecanoyloxy-propan-2-yl] (5e,8e)-hexadeca-5,8,11,14-tetraenoate

Supplementary Table 3. Structures within our simulation dataset that possess an experimentally solved allosteric ligand, location of the ligand's binding pocket and polarity of the pocket. In addition, we indicate whether a lipid insertion is observed in this pocket during our simulations as well as the GPCRmd and trajectory ID of the simulation where the insertion is observed.

System (PDB id)	Receptor	Ligand type (GPCRdb)	Allosteric site	Site polarity	Insertions in allosteric site?	GPCRmd id (apo form)	Trajectory IDs with insertion
5EE7	GCGR	NAM	TM6-TM7$_{IC}$	Polar	No	815	-
5VEX	GLP-1R	NAM	TM6-TM7$_{IC}$	Polar	No	710	-
5VEW	GLP-1R	NAM	TM6-TM7$_{IC}$	Polar	No	698	-
6KJV	GLP-1R	NAM	TM6-TM7$_{IC}$	Polar	No	746	-
6KK7	GLP-1R	NAM	TM6-TM7$_{IC}$	Polar	No	876	-
6KK1	GLP-1R	NAM	TM6-TM7$_{IC}$	Polar	No	704	-
6OBA	β_2-AR	NAM	TM3-TM4-TM5$_{IC}$ (subpocket)	Polar	No	1003	-
6C1Q	C5a$_1$R	NAM	TM3-TM4-TM5$_{IC}$	Hydrophobic	Yes	1084	18342 18343 18344
6C1R	C5a$_1$R	NAM	TM3-TM4-TM5$_{IC}$	Hydrophobic	Yes	1085	18350
5TZY	FFA1R	PAM	TM3-TM4-TM5$_{IC}$	Hydrophobic	Yes	928	16644 16645 16646
5KW2	FFA1R	Allosteric agonist	TM3-TM4-TM5$_{IC}$	Hydrophobic	Yes	765	15539 15540 15541
5TZR	FFA1R	Allosteric agonist	TM3-TM4$_{EC}$	Polar	Yes	793	15725
5NDZ	PAR2	Allosteric antagonist	TM2-TM3-TM4$_{EC}$	Hydrophobic	Yes	1018	17251 17252
6KQI	CB$_1$R	NAM	TM1-TM2-TM4$_{IC}$	Hydrophobic	No	1136	-

Supplementary Table 4. MD checklist for data reproducibility.

Reliability and reproducibility checklist for molecular dynamics simulations	**Yes**	**N/A**	**Response (Please state where this information can be found in the text)**
1. Convergence of simulations and analysis			
1a. Is an evaluation presented in the text to show that the property being measured has equilibrated in the simulations (*e.g.* time-course analysis)?	☒		Simulation time for ensuring sufficient equilibration and sampling was used according to state-of-the-art protocols (Rodríguez-Espigares I, Torrens-Fontanals M, Tiemann JKS, et al. GPCRmd uncovers the dynamics of the 3D-GPCRome. Nat Methods. 2020 Aug;17(8):777-787) A statement is found in the method section/simulation protocol.
1b. Then, is it described in the text how simulations are split into equilibration and production runs and how much data were analyzed from production runs?	☒		A statement is found in the method section/simulation protocol.
1c. Are there at least 3 simulations per simulation condition with statistical analysis?	☒		Three replicates are used. Statistics are found in Figure 2a, 2e, 5a
1d. Is evidence provided in the text that the simulation results presented are independent of initial configuration?	☒		The simulation time and replicates are sufficient to allow the system to comprehensively explore the conformational space around the experimental starting structure which minimizes the dependence on the initial configuration. This is stated in the method section/simulation protocol.
2. Connection to experiments			
2a. Are calculations provided that can connect to experiments (*e.g.* loss or gain in function from mutagenesis, binding assays, NMR chemical shifts, J-couplings, SAXS curves, interaction distances or FRET distances, structure factors, diffusion coefficients, bulk modulus and other mechanical properties, *etc.*)?	☒		- Recovery of lipid insertions observed in X-ray or cryo-EM into GPCRs: Figure 3c and 3d - Recovery of allosteric sites by lipid insertions: Figure 4a - Opening of ligand gateways by lipid insertions: Figure 4b
3. Method choice			
3a. Is it described in the text what force field and water model are used and why?	☒		In method section/simulation protocol
3b. Do simulations contain membranes, membrane proteins, intrinsically disordered proteins, glycans, nucleic acids, polymers, or cryptic ligand binding?	☒	☐	In method section/simulation protocol
If 3b is **YES**, are enhanced sampling methods used?	☐	☒	
If enhanced sampling methods are used, are the convergence criteria clearly stated?	☐		
If 3b is **YES**, is it explained in the text why or why not enhanced sampling methods are used?	☒		In methods section/simulation protocol and table 2
4. Code and reproducibility			
4a. Is a table provided describing the system setup, such as simulation box dimensions, total number of atoms, total number of water molecules, salt concentration, lipid	☒		The paper describes a large database of simulations. Supplementary Data 1 includes the list of all simulated systems (371 different systems) with links to system composition (lipids, ions, water) and box dimension in the input files.

composition (number of molecules and type)?			Parameters used in equilibration and production common to all the runs are listed in Table 2.
4b. Is it described in the text what simulation and analysis software and which versions are used?	☒		Details provided in the methods section
4c. Are initial coordinate and simulation input files and a coordinate file of the final output provided as supplementary files or in a public repository?	☒		All files are uploaded to the public repository GPCRmd (www.gpcrmd.org)
4d. Is there custom code or custom force field parameters?	☒	☐	Response not needed if **N/A**
If **YES**, are they provided as supplementary profiles or in a public repository?	☒		https://github.com/GPCRmd/simulation_pipeline

References

1. Caliman, A. D., Miao, Y. & McCammon, J. A. Mapping the allosteric sites of the A2A adenosine receptor. *Chem. Biol. Drug Des.* **91**, 5–16 (2018).
2. Bruzzese, A., Dalton, J. A. R. & Giraldo, J. Insights into adenosine A2A receptor activation through cooperative modulation of agonist and allosteric lipid interactions. *PLoS Comput. Biol.* **16**, e1007818 (2020).
3. Tzortzini, E., Corey, R. A. & Kolocouris, A. Comparative Study of Receptor-, Receptor State-, and Membrane-Dependent Cholesterol Binding Sites in A2A and A1 Adenosine Receptors Using Coarse-Grained Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **63**, 928–949 (2023).
4. Hedderich, J. B. *et al.* The pocketome of G-protein-coupled receptors reveals previously untargeted allosteric sites. *Nat. Commun.* **13**, 2567 (2022).
5. Jakowiecki, J. & Filipek, S. Hydrophobic Ligand Entry and Exit Pathways of the CB1 Cannabinoid Receptor. *J. Chem. Inf. Model.* **56**, 2457–2466 (2016).
6. Sabatucci, A., Tortolani, D., Dainese, E. & Maccarrone, M. In silico mapping of allosteric ligand binding sites in type-1 cannabinoid receptor. *Biotechnol. Appl. Biochem.* **65**, 21–28 (2018).
7. Mapping of Allosteric Druggable Sites in Activation‐Associated Conformers of the M2

Muscarinic Receptor - Miao - 2014 - Chemical Biology & Drug Design - Wiley Online Library. https://onlinelibrary.wiley.com/doi/full/10.1111/cbdd.12233.

8. Boulet, L. P. *et al.* A twelve-week comparison of salmeterol and salbutamol in the treatment of mild-to-moderate asthma: a Canadian multicenter study. *J. Allergy Clin. Immunol.* **99**, 13–21 (1997).
9. Szlenk, C. T., Gc, J. B. & Natesan, S. Does the Lipid Bilayer Orchestrate Access and Binding of Ligands to Transmembrane Orthosteric/Allosteric Sites of G Protein-Coupled Receptors? *Mol. Pharmacol.* **96**, 527–541 (2019).
10. Szlenk, C. T., Gc, J. B. & Natesan, S. Membrane-Facilitated Receptor Access and Binding Mechanisms of Long-Acting β2-Adrenergic Receptor Agonists. *Mol. Pharmacol.* **100**, 406–427 (2021).