nature communications

Article

https://doi.org/10.1038/s41467-025-57034-y

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

Received: 15 May 2024

Accepted: 7 February 2025

Published online: 27 February 2025

Check for updates

A list of authors and their affiliations appears at the end of the paper

G protein-coupled receptors (GPCRs) constitute a functionally diverse protein family and are targets for a broad spectrum of pharmaceuticals. Technological progress in X-ray crystallography and cryogenic electron microscopy has enabled extensive, high-resolution structural characterisation of GPCRs in different conformational states. However, as highly dynamic events underlie GPCR signalling, a complete understanding of GPCR functionality requires insights into their conformational dynamics. Here, we present a large dataset of molecular dynamics simulations covering 60% of currently available GPCR structures. Our analysis reveals extensive local "breathing" motions of the receptor on a nano- to microsecond timescale and provides access to numerous previously unexplored receptor conformational states. Furthermore, we reveal that receptor flexibility impacts the shape of allosteric drug binding sites, which frequently adopt partially or completely closed states in the absence of a molecular modulator. We demonstrate that exploring membrane lipid dynamics and their interaction with GPCRs is an efficient approach to expose such hidden allosteric sites and even lateral ligand entrance gateways. The obtained insights and generated dataset on conformations, allosteric sites and lateral entrance gates in GPCRs allows us to better understand the functionality of these receptors and opens new therapeutic avenues for drug-targeting strategies.

G Protein-Coupled Receptors (GPCRs) constitute a diverse family of transmembrane receptors that respond to various stimuli[1], including physical signals such as light or pressure[2], or chemical signals such as hormones, odorants and neurotransmitters. Their key role in cellular signalling has positioned them as crucial targets for pharmaceutical compounds: ~34% of all FDA-approved drugs target GPCRs[3]. Acquiring new structural insights into this therapeutically relevant receptor class can accelerate the discovery of improved and novel drug candidates for a range of diseases (e.g., neurological disorders, diabetes, chronic pain, or inflammation).

Over the past decade, technological advancements in protein engineering, X-ray crystallography and cryogenic electron microscopy (CryoEM) have prompted a steady increase in the number of available protein structures (https://www.rcsb.org/stats/growth/growth-protein), including GPCRs (https://gpcrdb.org/structure/statistics). However, static high-resolution structures offer limited insights into the inherent flexibility of GPCRs, a crucial aspect to fully understand their functionality[4]. In this context, in silico techniques such as molecular dynamics (MD) offer powerful means to exploring GPCR function and regulation at spatial and temporal resolutions not currently achievable with experimental methods[5]. This exploration encompasses aspects such as conformational flexibility[6–9], ligand binding and unbinding[10,11] and allosteric communication networks[12,13]. To encourage a community-wide and systematic study of the structural dynamics of GPCRs, the GPCRmd online web server has been recently developed, featuring tools for

e-mail: jana.selent@upf.edu

data streaming, visualisation, analysis and sharing[14] (https://www.gpcrmd.org/).

To gain deeper understanding of the structural dynamics within the 3D-GPCRome, we have generated an extensive dataset that captures the time-resolved dynamics of 190 GPCR structures, amassing a cumulative simulation time of over half a millisecond. This research endeavour was made possible through the collective efforts of an international team of GPCR experts. To illustrate the power of our large scale, data-driven analysis, we focus on three relevant aspects: the flexibility of GPCRs, their intricate interplay with membrane lipids and the detection of functionally relevant hotspots for receptor modulation (e.g., allosteric sites and lateral entrance gateways).

In this work, we show that MD simulations can effectively capture the inherent flexibility of GPCRs over relatively short timescales, uncovering significant 'breathing' motions at the intracellular side of the receptor across different GPCR types. This provides access to a vast number of receptor conformations previously unseen in experimental structures. Moreover, our analysis indicates that penetrations of membrane lipids into the receptor core, hereafter referred to as 'lipid insertions', are not only frequent in GPCRs but also topographically conserved across receptor subtypes. Importantly, we demonstrate that these insertions act as a valuable marker for membrane-exposed allosteric pockets as well as lateral entrance gates for specific GPCR ligand types. The insights into the plasticity of GPCRs and their interaction with membrane lipids, combined with the disclosure of allosteric sites and entrance gateways, advance our understanding of receptor function and provide valuable information for the development of original therapeutic strategies.

Results

A comprehensive dataset with a large diversity in receptor and ligand types

An international team of GPCR experts selected 190 experimentally solved structures from the GPCRdb database[15–17] and manually curated them for simulation following the GPCRmd standard protocol[14] (see Supplementary Data 1). This effort yielded in 181 ligand-GPCR complex systems, their corresponding 181 apo (ligand-free) systems, as well as 9 additional systems in their apo form only, as no ligand was bound in their original experimental structure. Each system was simulated for 3 × 500 ns, resulting in 1.5 µs of data per system and a total cumulative time of 556.5 µs. This dataset, which marks the second edition of GPCRmd, provides a significant data enrichment through dynamic insights into 33 receptor subtypes, with adenosine, adrenoceptors, opioid, muscarinic, orexin and opsins among the most prevalent (Fig. 1a). This second edition also introduces receptor subtypes unexplored in the first GPCRmd edition[14] such as the prostanoid, leukotriene and melatonin receptors (Fig. 1a). The vast majority of the simulated structures are class A GPCRs (~90% in GPCRmd), which is by far the largest and most therapeutically targeted GPCR class (Fig. 1b, top). Furthermore, the second edition of GPCRmd has significantly diversified the types of simulated ligands, including not only antagonists, (partial) agonists, inverse agonists and negative allosteric modulators (NAMs) but also allosteric (ant)agonists and positive allosteric modulators, as classified by GPCRdb[16] (Fig. 1b, middle). This enriched dynamic dataset offers insights into the modulatory effects of pharmacologically diverse GPCR ligands (Fig. 1b, bottom) and is freely available via www.gpcrmd.org.

GPCRs are flexible molecules with local breathing motions at the nano- to microsecond timescale

Inherent structural flexibility is key for GPCRs to sample a wide range of conformations, including spontaneous activation (i.e., constitutive activity). A structural hallmark of GPCR activation is the movement of transmembrane helix (TM) 6, which opens up an intracellular cavity for transducer coupling[18]. We monitored this TM6 motion as an indicator of receptor flexibility by measuring its distance to TM2 at the intracellular side of the receptor (Fig. 2a and Supplementary Table 1). To discern intracellular opening states, we set a distance threshold based on experimentally solved active (i.e., 'open') and inactive (i.e., 'closed') receptor states. We observe significant and pronounced differences in class A and B1 receptors, while class C and F receptors show no significant or only relatively minor differences, respectively. As a result, classes C and F were excluded from further analyses. Of note, the described conformational states (e.g., intermediate) may not always correspond to low-energy conformations, as our classification relies on structural metrics that do not reflect thermodynamic stability.

For the analysis of receptor flexibility, we focus on a homogenous dataset of inactive class A and B1 GPCR structures bound to an antagonist (90 systems), inverse agonist (9 systems), or NAM (8 systems) for 17 receptor types (Supplementary Data 2). Intriguingly, the analysis reveals unexpected breathing motions of the receptor core on a relatively short timescale (3 × 500 ns per system, Fig. 2b). Despite starting from an initially closed receptor core, apo receptors explore intermediate states (9.07%) and even open receptor cores at times (0.5%) (Fig. 2b, top-left). A notable example of this phenomenon is observed in the adenosine 2 A receptor ($A_{2A}R$, PDB id: 5UIG, Fig. 2c), where the apo form (GPCRmd id: 799 [https://www.gpcrmd.org/view/799/], trajectory 15764) exhibits significant cytosolic TM6 outward movement, indicating that all three conformational states (open, intermediate and closed) are visited during a single 500 ns simulation run. A similar pattern of breathing motion is seen in the apo form of the chemokine CCR2 receptor (CCR2, PDB id: 6GPX, Fig. 2d). It is tempting to speculate that these instances of high receptor flexibility may be linked to basal activity, as both CCR2[19] and $A_{2A}R$[20] are reported to display high basal activity, which in the case of the $A_{2A}R$ is reduced by the presence of antagonists[20].

Importantly, when analysing breathing motions of GPCRs in complex with antagonists, inverse agonists, or NAMs, the conformational sampling is significantly reduced ($P = 3.71 \times 10^{-5}$, two-sided Mann-Whitney U test), with only 3.8% of the simulation time in intermediate states and less than 0.1% in open states (Fig. 2b, bottom-left). This suggests that perturbation of the conformational dynamics through inactive state stabilisation upon ligand binding is a general molecular mechanism of these ligand types across different receptor subtypes. We also investigated whether breathing motions differ when GPCRs are bound to agonists or allosteric agonists (26 PDBs) compared to antagonists, inverse agonists, or NAMs (106 PDBs) (Supplementary Fig. 1). Although this analysis provides only a preliminary perspective due to the underrepresentation of agonist or allosteric agonist complexes, the data suggest that these complexes more frequently sample intermediate states compared to antagonist, inverse agonist, or NAM complexes. This observation aligns with the notion that agonists or allosteric agonists shift the receptor state from a closed, inactive-like conformation toward more intermediate states within 3 × 500 ns simulations. We further speculate that simulations over longer timescales could potentially reveal increased sampling of open (active-like) states for agonists or allosteric agonists.

Monitoring the evolution of breathing motions, we systematically recorded the transitions from closed to intermediate and closed to open states. This allowed us to approximate their respective transition times between different conformations (see 'method' section and Supplementary Table 1). For the apo receptors, we find that state transitions are in the range of 0.5 µs (closed to intermediate) and 7.8 µs (closed to open), respectively (Fig. 2b, right). In contrast, the presence of antagonists, inverse agonists, or NAMs significantly slows down both kinetics to 1.2 µs (closed to intermediate) and 52.7 µs (closed to open), respectively. Of note, these values reflect the average over the studied set of receptors, i.e., transition kinetics can be faster or slower for individual GPCRs.

https://doi.org/10.1038/s41467-025-57034-y

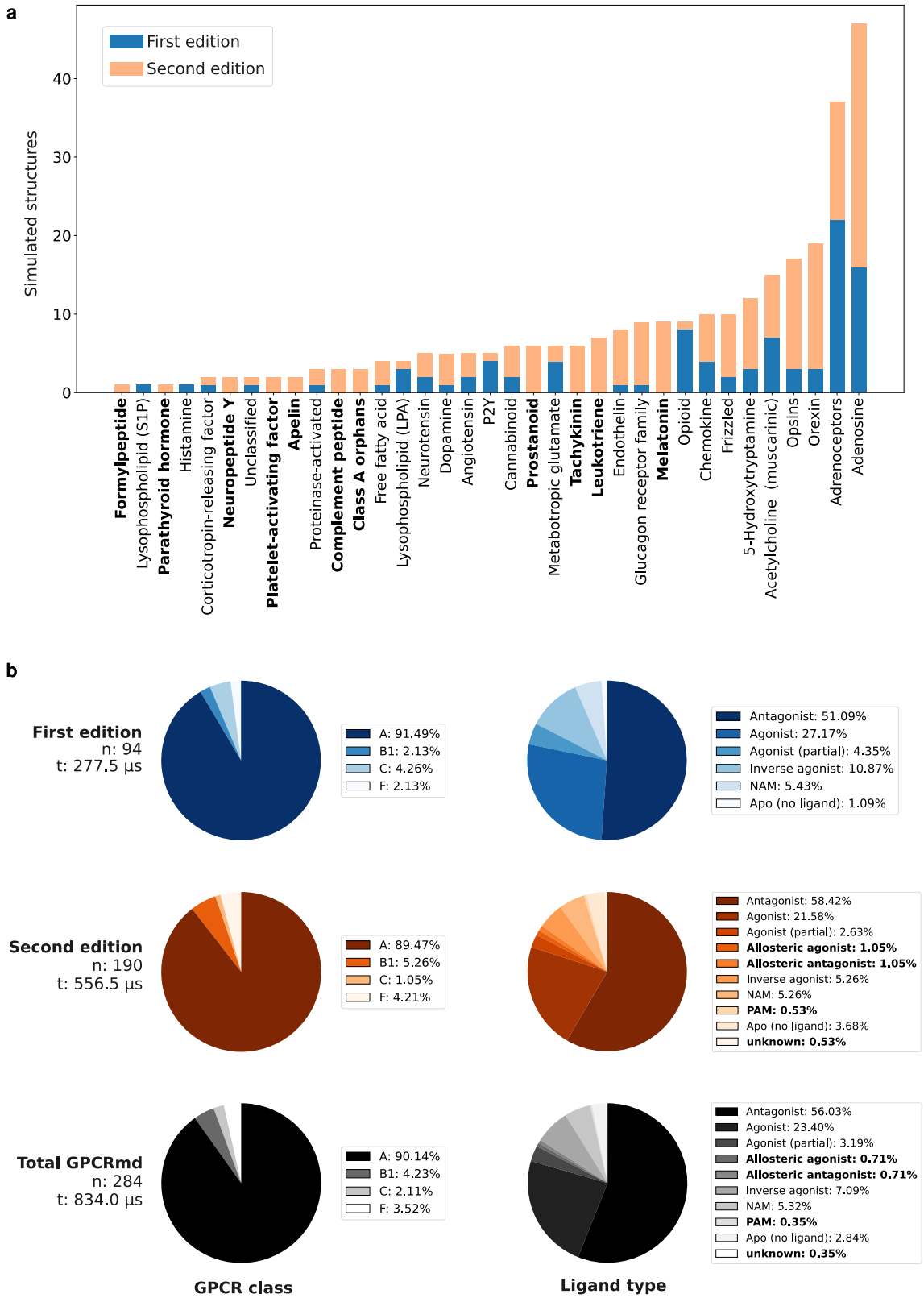


Fig. 1 | Dataset diversity for GPCR class, receptor type and ligand type. Newly introduced features in the second GPCRmd edition are displayed in bold. **a** Receptor type of the structures simulated in the second GPCRmd edition (this publication) versus the first GPCRmd round[14]. **b** GPCR class and ligand type of the first GPCRmd edition (top), the second GPCRmd edition (middle) and the total dataset on GPCRmd (bottom). The reported 'unknown' ligand type refers to cases where a crystallisation adjuvant molecule is found inside the orthosteric pocket[74]. Source data are provided in the Source Data file.

Importantly, the simulated state transitions for our extensive dataset give access to a large number of experimentally unexplored receptor conformations (i.e., diverse inactive-like, intermediate or active-like) along the transition pathways (available at www.gpcrmd.org, and Supplementary Data 2). To make these unexplored conformation available to the researchers, the GPCRmd platform provides a tool to monitor receptor breathing over time and select specific conformations classified as inactive-like, intermediate or active-like as exemplified for the $A_{2A}R$ (PDB id: 5UIG, GPCRmd id: 773 [https://www.gpcrmd.org/view/773], Fig. 2e). Taken together, our results show a high level of conformational flexibility and state transition kinetics within the receptor core for class A and B receptors, particularly in their apo forms. They also confirm the anticipated reduction in GPCR flexibility and transition kinetics in the presence of antagonists, inverse agonists or NAMs.

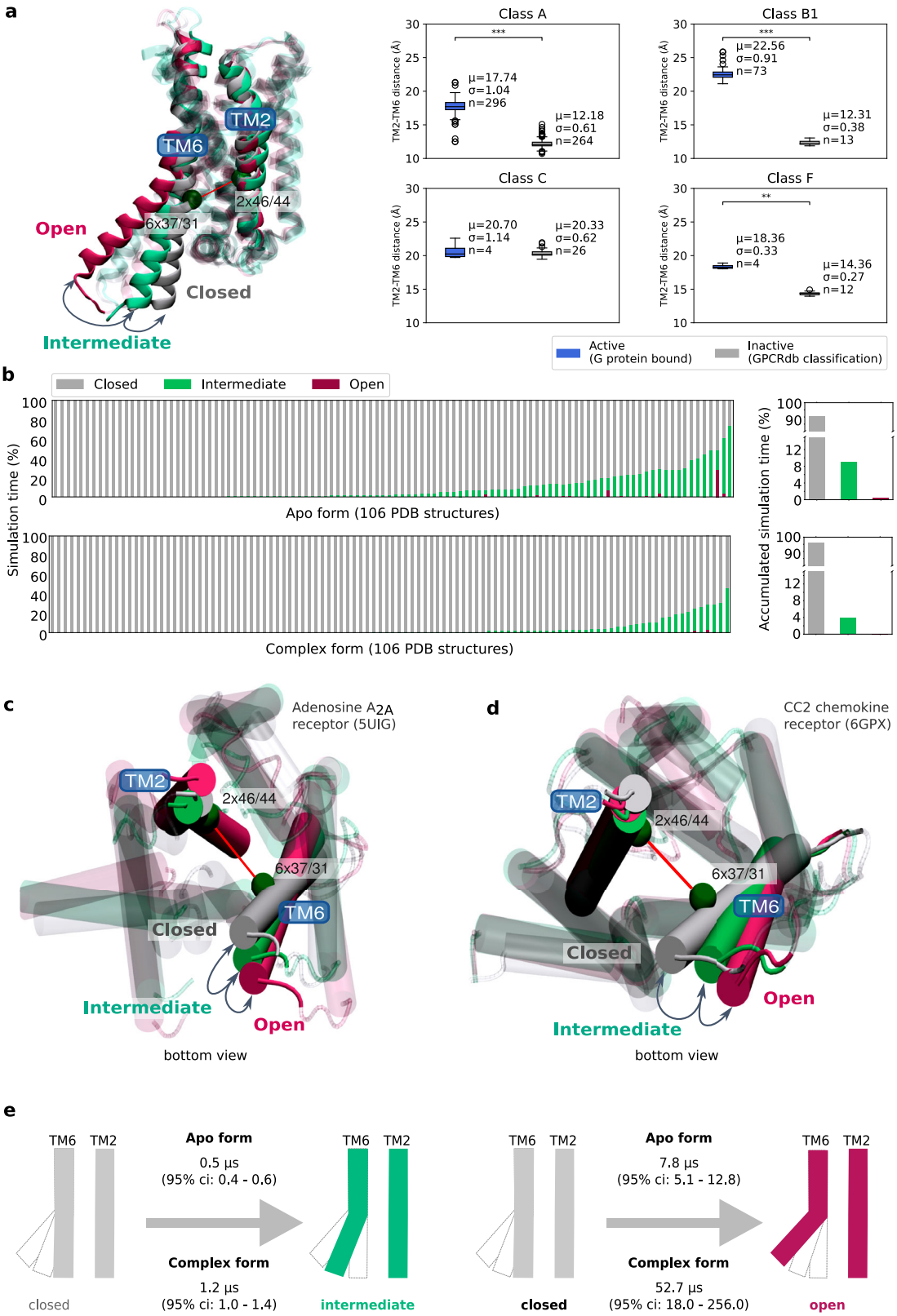


Fig. 2 | Flexibility analysis of the GPCR core. a Structural depiction of the TM6 outward movement and the residues used to measure the TM2-TM6 distance (i.e., class A, B1, C: 2 × 46 and 6 × 37, class F: 2 × 44 and 6 × 31) (left). Distribution plot of the TM2-TM6 distance for class A, B1, C and F for active states (i.e., GPCR structures bound to a G protein) and inactive states according to GPCRdb classification. Boxplots display the average (μ), standard deviation (σ) and sample size (n) for each group. The centre 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. Statistical significance was assessed by a two-sided Mann-Whitney *U* test (class A $P = 3.2 \times 10^{-92}$, class B1 $P = 1.1 \times 10^{-8}$, class C $P = 0.83$, class F $P = 4.3 \times 10^{-3}$). **b** Conformational flexibility in simulated systems of GPCR class A and B1. We monitored closed, intermediate and open states based on the TM2-TM6 distance during 3 × 500 ns. Bar plots show the results for 107 simulated apo GPCRs (top) and the corresponding systems in complex with antagonists, inverse agonists or NAMs, respectively (bottom). Values available in Supplementary Data 2. **c**, **d** Structural depiction of two case studies from our dataset, represented in their open (purple), intermediate (green) and closed (grey) conformations visited during one of its apo simulations replicates. The data for the adenosine A_{2A} receptor (PDB id: 5UIG) are available with GPCRmd id: 773 [https://www.gpcrmd.org/view/773], trajectory id: 15597. The simulation data for the CC2 chemokine receptor (PDB id: 6GPX) are accessible via the GPCRmd id: 799 [https://www.gpcrmd.org/view/799], trajectory 15764. The measured TM2-TM6 distances are marked in red. **e** Estimation of the average breathing kinetics across the 107 GPCR system pairs selected in b. The transition timescales from the closed to intermediate state (left) and from the closed to the open state (right) for apo and complex systems are shown including the 95% confidence interval (ci), respectively. Source data are provided in the Source Data file.

Lipid insertion into the receptor core is a persistent phenomenon in GPCRs

MD simulations offer a detailed view of the dynamic interactions between a GPCR and its membrane environment at high resolution. Here, we investigate the phenomena of shallow and deep lipid insertions (Fig. 3a) across 371 simulation systems (181 complex and their corresponding 181 apo systems, and 9 apo-only systems) by embedding the receptor into a homogenous POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine) membrane bilayer. POPC is a prototypical lipid that is commonly found in biological membranes. Intriguingly, we find that significant lipid insertions occur in all 190 GPCR systems (Supplementary Fig. 2). A detailed analysis (Fig. 3a, right) further shows that most lipid insertions (79%) are ultra-short (<20 ns) and likely represent random interactions. However, we also detect short (20–100 ns), medium (100–300 ns) and long-lasting insertions (>300 ns) with frequencies of 14%, 5% and 2% of occurrence, respectively. Interestingly, the prevalence of shallow insertions (i.e., lipids inserted into cavities at the receptor surface) versus deep insertions (i.e., lipids intercalating between transmembrane helices) correlates with the duration of the insertion. At short durations (20–100 ns), shallow insertions are more common than deep ones, while at longer durations (>300 ns), deep insertions predominate. It is tempting to speculate that the instances of deep and highly stable insertions (>300 ns) may have specific functional significance for GPCRs.

To obtain a more detailed understanding of the observed lipid insertions, we studied their location and insertion magnitude. The location is indicated by the transmembrane helices involved and where they map to on the receptor (i.e., intra (IC)- or extracellular (EC) receptor side). Our data reveals the existence of specific sites that are more prone to lipid insertions (Fig. 3b). The three most frequent insertion sites are TM1-TM7$_{EC}$ (35.8%) on the extracellular side and TM3-TM4$_{IC}$ (33.0%) and TM3-TM5$_{IC}$ (15.2%) on the intracellular side. Interestingly, these three sites coincide with an orphan pocket site at TM1-TM7$_{EC}$ and a known pocket site at TM3-TM4$_{IC}$ and TM3-TM5$_{IC}$ described in a recent in silico study for class A GPCRs[21].

Furthermore, we observe specific insertion patterns for shallow or deep lipid insertions (Fig. 3b). Generally, deep insertions tend to be frequent at the EC side of TM1-TM2-TM7 (i.e., TM1-TM7$_{EC}$, TM1-TM2$_{EC}$, TM2-TM7$_{EC}$). This pattern may result from the highly flexible N-terminus linked to TM1, which introduces plasticity into this region, facilitating lipid traversal through the surrounding pockets. In contrast, shallow insertions are predominantly found on the IC side of TM3-TM4-TM5 (i.e., TM3-TM4$_{IC}$, TM4-TM5$_{IC}$, TM3-TM5$_{IC}$). Intriguingly, this pocket has been previously identified as a binding site for other membrane lipids such as cholesterol in $A_{2A}R$[22], A_1R[22], kappa-opioid[23] and β_1-adrenergic receptors[24] (β_1-adrenoceptor). Furthermore, allosteric modulators have been reported to bind to this site[21], for instance in the GPR40 (PDB ids: 5TZY and 5KW2), the C5a receptor (PDB ids: 6C1Q, 6C1R) and the β_2-adrenoceptor (β_2-AR) (PDB id: 6OBA), underscoring its pharmacological significance[25,26].

Computer simulations recover lipid insertions observed in experimental structure

Static snapshots of lipid insertions in GPCRs are occasionally captured in experimentally solved structures. Among the 190 simulated structures, 42 feature a total of 49 insertions in their experimentally determined structures (Fig. 3c, top) including a large variety of membrane lipids or fatty acids (Supplementary Table 2). We investigated whether all-atomistic MD simulations could replicate these experimentally observed events. Remarkably, we discover that the vast majority of static insertions (46 out of 49) are reproduced as shallow or deep insertions in our simulations (Fig. 3c, bottom, Supplementary Data 5). On the extracellular side, the simulations accurately reproduced insertions for TM1-TM7$_{EC}$ (e.g., $A_{2A}R$, β_2-AR and dopamine 4 receptor (D_4R)) and TM4-TM5$_{EC}$ (e.g., LPA6R, PAFPR). On the intracellular side, we find highly explored insertions between TM3-TM4$_{IC}$ (e.g., $A_{2A}R$, CB_1R, P2YR and prostanoid receptors). Notably, the predominance of TM1-TM7$_{EC}$ and TM3-TM4$_{IC}$ as the most frequently recovered sites (Fig. 3c, bottom) aligns with our observations for the entire dataset (Fig. 3b).

A notable example is the $A_{2A}R$ with insertions at TM1-TM7$_{EC}$, which are found in more than 30% of experimentally solved structures (12 out of 31) (Fig. 3c, top). Remarkably, our simulated $A_{2A}R$ systems reproduce this insertion in 68.17% of the simulation time (Fig. 3c, bottom). A structural view illustrates how the simulated lipid (orange stick) penetrates the receptor occupying the site of the co-solved lipid (cyan stick) (Fig. 3d). This lipid insertion is located near the orthosteric ligand binding pocket, separated only from it by residues L267^{7x31} and Y271^{7x35}. The cavity adjacent to these residues has been described as a potential cryptic allosteric site, since many antagonists (e.g., ZM241385) but not the natural agonist, adenosine, occupy this site[27]. Another frequently recovered site, found in more than 30% of cases, is located between TM3 and TM4 on the intracellular side (Supplementary Fig. 3 and Supplementary Note 1). Thus, the high rate of recovery of experimentally observed lipid insertions underscores MD simulations as a powerful tool for detecting these important events with resolution on the nanosecond timescale.

Lipid insertion sites are hotspots for allosteric modulation of GPCRs

The analysis presented in Fig. 3 demonstrates significant penetration of membrane lipids into the receptor. Based on this observation, we hypothesise that sites of lipid insertion are functionally relevant hotspots for the allosteric modulation of GPCRs. To assess whether such lipid insertions can predict allosteric sites, we screened apo GPCRs in our simulated dataset, which initially had an allosteric modulator co-

solved at the receptor surface (14 systems, Supplementary Fig. 4, Supplementary Table 3).

Our analysis confirms that lipid dynamics in the apo GPCRs can indeed recapitulate a number of those experimentally determined allosteric sites (Fig. 4a). Among these, we observe a higher reproducibility of hydrophobic pockets compared to polar sites, which is expected given the hydrophobic nature of lipid tails. For instance, we find that the hydrophobic TM3-TM4-TM5$_{IC}$ site is consistently identified in our simulations for both the free fatty acid receptor 1 (FFA1R) and the C5a$_1$ receptor (C5a$_1$R) (Supplementary Table 3).

We observe that such lipid insertions can occur very rapidly, within 8 ns after positioning the lipid in front of the pocket, as seen for the FFA1R (PDB id: 5KW2, GPCRmd id: 765 [https://www.gpcrmd.org/view/765/] trajectory id: 15540) (Fig. 4a). Remarkably, once inside the

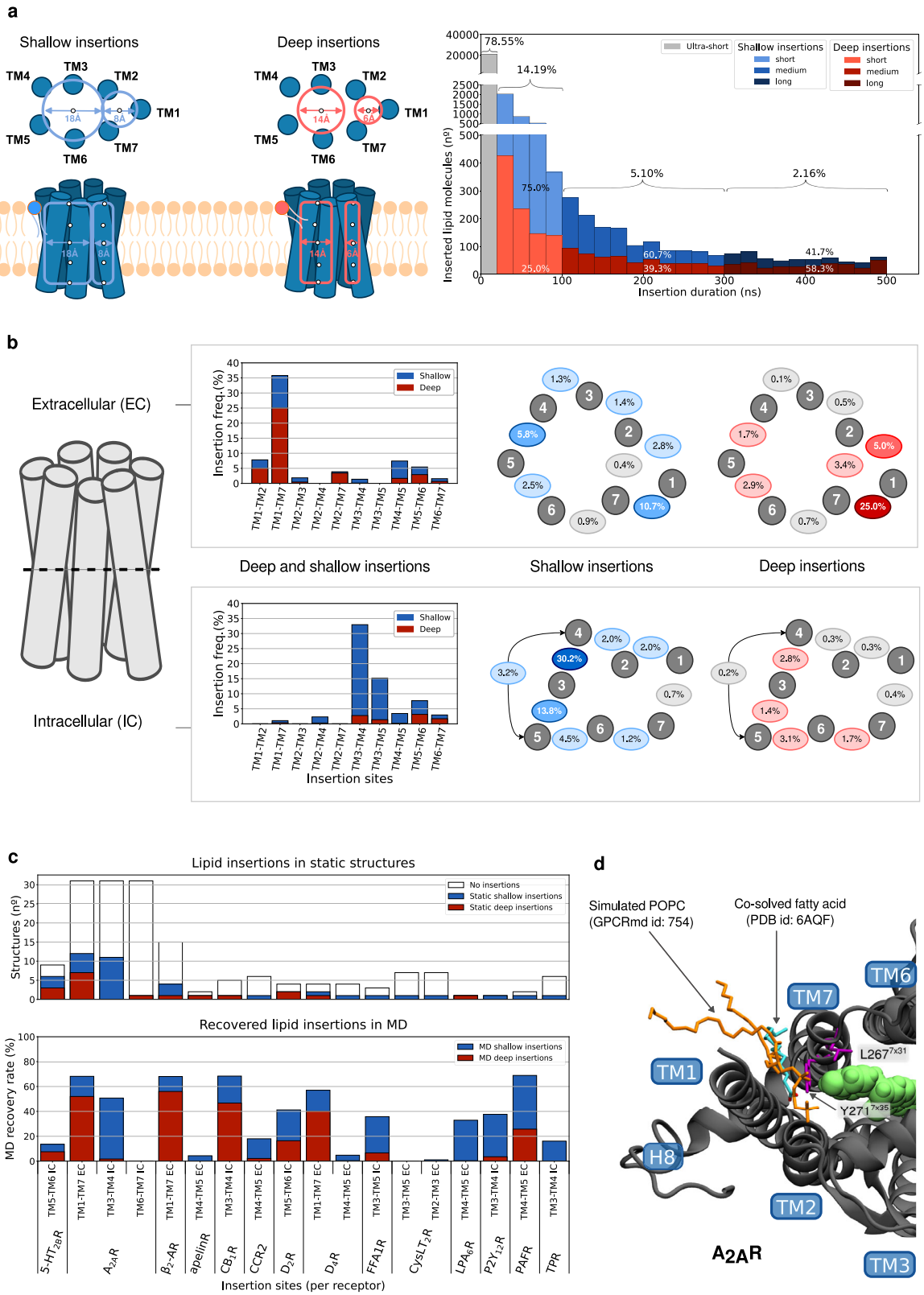


Fig. 3 | Identification and quantification of lipid insertions in GPCR MDs. **a** Visual scheme of the algorithm used to find shallow and deep lipid insertions in GPCR simulations (left) and total count of individual lipid insertions across the dataset classified by duration and depth (right). Lipid molecules found inside any of the inner core receptor selections displayed (left) at least once during a simulation constitute individual lipid insertions. Lipids spending more than 20% of their total insertion time inside the smaller inner core selection (salmon) constitute 'deep' insertions, otherwise being considered 'shallow' insertions. Insertions are classified, according to their duration, as ultra-short (<20 ns), short (20–100 ns), medium (100–300 ns), and long (>300 ns). More information on lipid insertion detection and classification can be found in the 'methods' section. **b** Dataset-wide average lipid insertion frequency by receptor site, classified by surrounding TM helices and GPCR side (either Extracellular (EC) or Intracellular (IC)). Frequency values are displayed in two bar plots (centre) and in four visual schemes (right) where each grey circle represents a TM helix and each coloured oval an insertion frequency value. Frequencies for individual simulated systems are available in Supplementary Data 3 and 4. **c** Barplots representing insertions of lipid-like molecules (Supplementary Table 2) in static PDB structures (top) and the average recovery rate of these insertions in the corresponding simulations, classified by receptor and insertion sites (bottom). Results available in Supplementary Data 5. **d** Recovery of an experimental lipid insertion in an $A_{2A}R$ simulated system (PDB id: 6AQF, GPCRmd id: 754 [https://www.gpcrmd.org/view/754]), where a POPC lipid (licorice, orange) appears inserted in TM1-TM7 EC, the same spot as a previous static insertion produced by an oleic acid co-solved molecule (licorice, cyan). This insertion site is separated from the ligand binding site by residues $L267^{7x31}$ and $Y271^{7x35}$ (licorice, purple). Source data are provided in the Source Data file.

pocket, the aliphatic tail of POPC aligns well with the initially co-solved allosteric ligand. We detected several other insertions in the FFA1R such as the TM3-$TM4_{EC}$ which was not co-solved with a ligand in the 5KW2 structure. Interestingly, this site was experimentally captured with an allosteric ligand in a different FFA1R structure (PDB id: 5TZY, GPCRmd id: 929 [https://www.gpcrmd.org/view/929])[28,29] highlighting the potential of lipid binding regions on the receptor explored via MD simulations on a nanosecond timescale to be predictive of allosteric sites. Unsurprisingly, hydrophobic lipid tails do not detect highly polar sites such as the TM6-TM7 pocket, which contains numerous positively charged residues at the intracellular site for the GLP1R (PDB ids: 5VEX, 5VEW, 6KJV, 6KK7 6KK1) and the GCGR (PDB id: 5EE7) (Supplementary Table 3). In addition, we find a very shallow pocket in TM1-TM2-$TM4_{IC}$ where insertions were not observed, despite its hydrophobic nature (PDB id: 6KQI; GPCRmd id: 1002 [https://www.gpcrmd.org/view/1002]), GPCRmd id: 1136 [https://www.gpcrmd.org/view/1136]).

Lipids dynamics reveal lateral gateways into the receptor core

To investigate whether lipids can reveal lateral entrance gateways into GPCRs, we screened for deep lipid insertions that penetrate the receptor core (Supplementary Data 6 to 7). A notable example is rhodopsin, with important lipid insertions between TM1-$TM7_{EC}$ (3 out of 10 systems) and TM5-$TM6_{EC}$ (8 out of 10 systems) with and without ligands. Interestingly, we observe that the presence of retinal alters the extent of lipid insertion for both gateways: the TM1-$TM7_{EC}$ (apo system: 14.60%, complex system: 7.45%) as well as the TM5-$TM6_{EC}$ gate (apo system: 24.33%, complex system: 10.50%). The overall reduced lipid occupancy in the complex system indicates that the ligand retinal competes with inserted lipids in the receptor core. This competition is illustrated in the rhodopsin apo form (PDB id: 5DYS, GPCRmd id: 872 trajectory id: 16264 [https://www.gpcrmd.org/view/872]) (Fig. 4b) where a lipid tail extends deep into the receptor core through the TM5-$TM6_{EC}$ gate, eventually occupying part of the all-trans-retinal binding pocket (Fig. 4b, inset). At times, the lipid insertion is obstructed by three phenylalanine residues ($F208^{5x43}$, $F273^{6x56}$ and $F276^{6x59}$), which appear to act as gatekeepers. We also detect lipid insertions for the second gate between TM1-$TM7_{EC}$ with residues $Y43^{1x38}$ and $F293^{7x39}$ as effective gatekeepers (Fig. 4b). Our results align with previous findings by Hildebrand et al.[30], who reported both lipid insertion sites as part of a larger channel in rhodopsin facilitating retinal entrance and exit.

Interestingly, the majority of observed lipid insertions into the transmembrane domain of GPCRs are mediated by the hydrophobic tail of membrane lipids. However, we also observe one event in which the polar lipid head opens a lateral channel into the receptor interior, as seen in the cysteinyl leukotriene receptor 1 (PDB id: 6RZ4, GPCRmd id: 738 [https://www.gpcrmd.org/view/738]). Thereby, the positively charged choline group of POPC penetrates the receptor between TM5 and TM6, attracted by two negatively charged residues: the allosteric ion binding site D^{2x50} and D^{7x49} (Supplementary Fig. 5). Notably, such a mechanism has been previously reported for the olfactory receptor OR51E2, which features a dual negatively charged site (D^{2x50}, E^{3x39}) to coordinate divalent ions[31]. In this case, the POPC molecule infiltrates between TM6 and TM7, with its choline group interacting with D^{2x50} and E^{7x49}. Overall, these results suggest that core penetration by positively charged lipid heads is facilitated by a dual unbalanced negative charge within the receptor.

Importantly, our simulation data reveal gateways for lipid insertion in numerous other receptors (Supplementary Data 6 to 7). Among them, some have been previously reported to serve as lateral entrance gates for ligands, such as for the cannabinoid receptor[32] (PDB ids: 5XR8 and 5XRA; GPCRmd ids: 852 [https://www.gpcrmd.org/view/852] and 896 [https://www.gpcrmd.org/view/896], Supplementary Fig. 6 and Supplementary Note 2) and even the β_2-AR (PDB ids: 6OBA, 6PS0 and 5X7D; GPCRmd ids: 1003 [https://www.gpcrmd.org/view/1003], 684 [https://www.gpcrmd.org/view/684] and 769 [https://www.gpcrmd.org/view/769], Supplementary Fig. 7 and Supplementary Note 3), where lipophilic ligands can enter the pocket through the membrane[33]. The widespread occurrence of these observations across several receptor subtypes supports the notion that deep lipid insertions in MD simulations are a reliable predictor of ligand gateways in GPCRs.

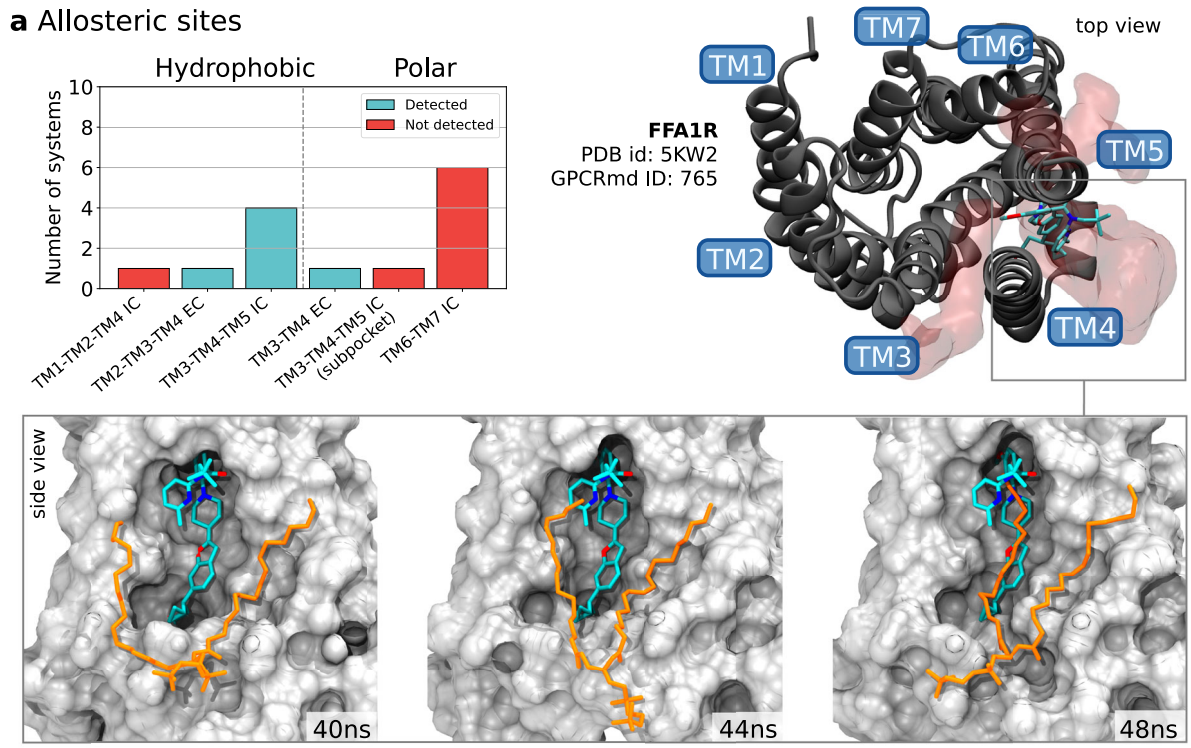
Allosteric sites and lateral channels are often hidden but can be disclosed by lipid dynamics

The identification of allosteric sites and lateral channels is of high relevance for drug discovery programmes. However, the success of virtual screening campaigns into those sites requires that they are captured in an open state. To explore in more depth the stability and dynamics of studied sites, we first monitored the root mean square deviation (RMSD) of pocket residues across the simulations under ligand-bound (complex) and ligand-free (apo) conditions (Fig. 5a). As expected, the RMSD values are significantly lower in ligand-stabilised allosteric sites (complex) (Fig. 5b) whereas ligand-free sites exhibit a notable increase in RMSD due to their higher flexibility. Most importantly, these increased pocket dynamics lead to a pronounced alteration in the site architecture across all investigated cases (Fig. 5b vs Fig. 5c). This manifests in the complete closure of small pockets (i.e., 5NDZ, 5TZR and 5TZY site 1), or partial closure of larger-sized pockets (i.e., 5TZY site 2, 5KW2, 6C1Q and 6C1R) (Fig. 5c). From this data, we conclude that allosteric sites are often hidden in the absence of a molecular modulator. In fact, this can be evidenced in experimental structures such as the FFA1R where an allosteric pocket at TM3-$TM4_{EC}$ is stabilised in an open state when bound to its ligand (PDB id: 5TZY) (Supplementary Fig. 8a) but in a closed state when the ligand is absent (PDB id: 5KW2) (Supplementary Fig. 8b). In this case, the rotation of one phenylalanine (F4.61) is responsible for the closure of this allosteric pocket (Supplementary Fig. 8c).

Importantly, following the evolution of pocket dynamics from a closed state (Fig. 5c), we observe that lipids can facilitate pocket opening at the nanosecond timescale. This can be evidenced for the protease-activated receptor 2 (PDB id: 5NDZ) (Supplementary

Movie 1), the GPR40 (PDB ids: 5TZR, 5TZY), the FFA1R (PDB id: 5KW2) and the complement component 5a receptor 1 (PDB ids: 6C1Q, 6C1R) in our dataset. Not surprisingly, lateral channels into the GPCR are also highly dynamic and undergo closing and opening events at the nanoseconds timescale. Similar to allosteric sites, lipids can facilitate the opening of such channels. This is exemplified in rhodopsin, where the TM5-TM6 gateway is blocked by three phenylalanine residues (F208^{5x44}, F208^{6x56} and F208^{6x59}) (Fig. 4b). Membrane lipid motions push these residues aside, enabling lipid penetration into the receptor core and even occupying part of the orthosteric ligand site (Supplementary Movies 2 and 3).

Overall, our findings indicate that lipid dynamics and their interaction with the receptor surface are able to promote opening of allosteric sites and lateral gateways into GPCRs. Moreover, it suggests



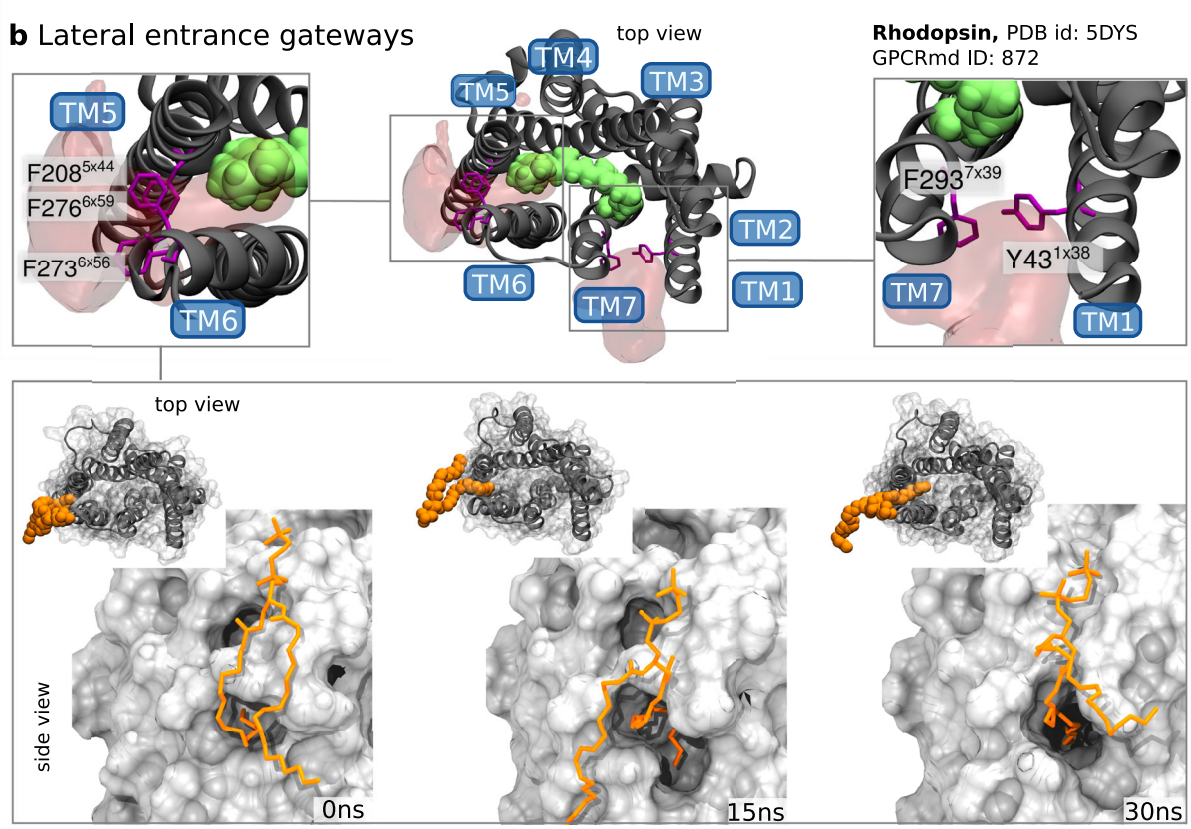


Fig. 4 | Detection of allosteric binding sites and entrance channels through lipid insertions. a Identification of lipid insertions in systems experimentally solved with an allosteric ligand. Allosteric sites are classified according to their location in the receptor and the surrounding residues' polarity. The results are displayed in a barplot (top left) with blue columns representing systems where a lipid insertion is found and red columns for those where no insertion is detected in any simulated replicate. Structural depiction of the free fatty-acid receptor-1 as an example system (top right) (PDB id: 5KW2, GPCRmd id: 765 [https://www.gpcrmd.org/view/765], trajectory id: 15540). A lipid penetration at TM3-TM4-$TM5_{IC}$ is observed where it occupies a known allosteric pocket (bottom). The initially co-solved allosteric modulator (blue colour) is shown as reference (PDB id: 5KW2). **b** Detection of ligand entrance/exit gateways sites through lipid insertions. Gateways identified by a dynamic lipid insertion into a rhodopsin (PDB id: 5DYS, GPCRmd id: 872 [https://www.gpcrmd.org/view/872], trajectory id: 16264). One insertion occurs at site TM5-$TM6_{EC}$ in the apo form system. In the absence of a ligand, the inserted lipid proceeds in this replicate to partially occupy the cavity at the beginning of the simulation. Another remarkable insertion occurs between TM1-$TM7_{EC}$. Gatekeeper residues of the first ($F208^{5x44}$, $F273^{6x56}$ and $F276^{6x59}$) and the second sites ($Y43^{1x38}$ and $F293^{7x39}$) are represented in purple licorice. Density maps in red mark the sites that inserted lipids occupied during the simulations. See Supplementary Movies 2 and 3 for further details. Source data are provided in the Source Data file.

that our compilation of lipid insertion data for individual receptors across our large scale GPCR dataset (available at www.gpcrmd.org) is a valuable resource for identifying unexplored sites and channels. For a better exploitation of this compiled data, GPCRmd provides a tool that maps lipid insertions on the 3D structure for individual structures (Fig. 5e, rhodopsin example available in GPCRmd id: 872 [https://www.gpcrmd.org/view/872]).

Discussion

The exponential growth in computational capabilities and the development of dedicated software[34–36] have made it possible to simulate increasingly large biological systems over longer timescales. In this study, we took advantage of these developments to create the largest MD simulation dataset for the GPCR protein family, ensuring consistency in terms of the force-field and protocols used. All simulated systems are hosted and publicly available on the GPCRmd public server (www.gpcrmd.org), where they can be visualised, analysed and downloaded for research purposes. Leveraging this extensive dataset, we provide insights into GPCR dynamics related to receptor plasticity and lipid regulation.

Our simulations reveal pervasive rapid breathing motions observed across a number of receptors (Fig. 2b–d), which facilitate spontaneous opening and closing of the intracellular receptor core even in apo receptors. Despite having an open intracellular core, receptors typically do not reach the fully active states based on the A100 index (Supplementary Fig. 9)—an index that integrates various metrics of receptor activation[37]. Achieving these states likely requires a longer timescale and/or agonist binding or G protein coupling.

Previous NMR studies have inferred preactivation- and activation-related TM movements (e.g., TM5 or TM6) at the sub-millisecond to sub-microsecond timescale[38,39]. However, the exact timescales and more importantly the modes and amplitudes of these rapid conformational dynamics remain highly challenging to determine by current experimental methods. Our simulations show that transitions from closed to intermediate states in the apo receptor occur within nanoseconds, whereas transitions from closed to open states could take up to a few microseconds (Fig. 2b). These observations suggest a significant energy barrier between the intermediate and the open states. As expected, the binding of antagonists, inverse agonists and NAMs significantly slows down the transition times (i.e., ~2-fold for the transition to the intermediate and 7-fold to the active-like state) between these states. Importantly, the recorded time-resolved state transitions for this large collection of GPCRs at the full atomistic level illuminates a wealth of receptor conformations so far experimentally unexplored. Notably, representative inactive-like, intermediate and active-like GPCR conformations can be directly downloaded from the GPCRmd repository (Fig. 2e) to guide virtual screening and rational drug design efforts.

Moreover, the described intrinsic flexibility can account for GPCRs' ability to pre-couple with heterotrimeric G proteins in the absence of an agonist[40] as well as to promote basal receptor activity, a common feature in the majority of GPCRs[41,42]. Yet, a direct correlation of breathing motions at the intracellular receptor side alone to basal receptor activity is challenging. Other highly flexible regions such as intrinsically disordered intracellular loops (ICL), which are typically not solved in experimental structure and therefore absent in our simulations, might have a significant contribution. For instance, the ICL3 has been reported to play an important role for the regulation of GPCR signalling including G protein specificity[43]. It can be expected that this flexible region at the bottom of the receptor will control the coupling of G proteins also in the absence of an agonist. Importantly, we also find that receptor flexibility is consistently reduced when the receptor is bound to antagonists, inverse agonists, or NAMs (Fig. 2b), thus stabilising the closed (inactive-like) state. The result aligns with the concept that these ligands prevent the receptor from exploring open (active-like) states[44]. Furthermore, our data provide preliminary evidence that the binding of agonists or allosteric agonists destabilises the closed (inactive-like) state, promoting a shift toward intermediate states (Supplementary Fig. 1). However, future studies with a more balanced dataset will be necessary to confirm this observation.

Beyond receptor flexibility, the extensive dataset unveils intriguing aspects of membrane lipids and their consistent ability to insert into the receptor core across all simulated systems. While 75% of lipid insertions occur within a short timeframe (<20 ns) and through rather random contacts, 25% contribute to extensive interactions (20–500 ns) through both shallow and deep penetrations (Fig. 3a). However, to further verify that these insertions are not artefacts of the simulation protocol, we assessed the extent to which experimentally determined lipid insertions are reproduced. The excellent agreement supports the notion that frequent lipid insertions observed in simulation are a biological feature in GPCRs (Fig. 3c, bottom). Interestingly, a receptorome-wide analysis of our simulation data further reveals that there are insertion sites that are common across diverse GPCR types. The most frequent sites are TM1-$TM7_{EC}$ on the extracellular side and TM3-TM4-$TM5_{IC}$ on the intracellular side (Fig. 3b).

Another important question is whether the observed pattern of lipid insertions changes under certain conditions, such as ligand binding. In this respect, our data indicate that the lipid insertion profile in apo receptors closely resembles that of those in complex with ligands, both in terms of insertion sites and the number and duration of lipid insertions (Supplementary Fig. 10a, b). However, differences may arise in specific cases where deep lipid insertions overlap with the ligand binding site. Such lipid insertions are likely to be reduced in the presence of a ligand, as observed in the rhodopsin receptor (see 'Results' section). It is tempting to speculate that the receptor state also impacts the lipid insertion pattern. Unfortunately, we cannot investigate this in our dataset, as we lack data for the same receptor simulated in both inactive and active states. Nevertheless, to address this question, we focused on a specific receptor, the $A_{2A}R$, by including an active-state simulation for direct comparison with the inactive-state simulation (Supplementary Fig. 11). In this active state, we observe several new lipid insertion sites on the intracellular side. This likely results from less tightly packed transmembrane helices which are connected to the core opening upon G protein binding. Our findings indicate that lipid insertion patterns indeed differ between active and

https://doi.org/10.1038/s41467-025-57034-y

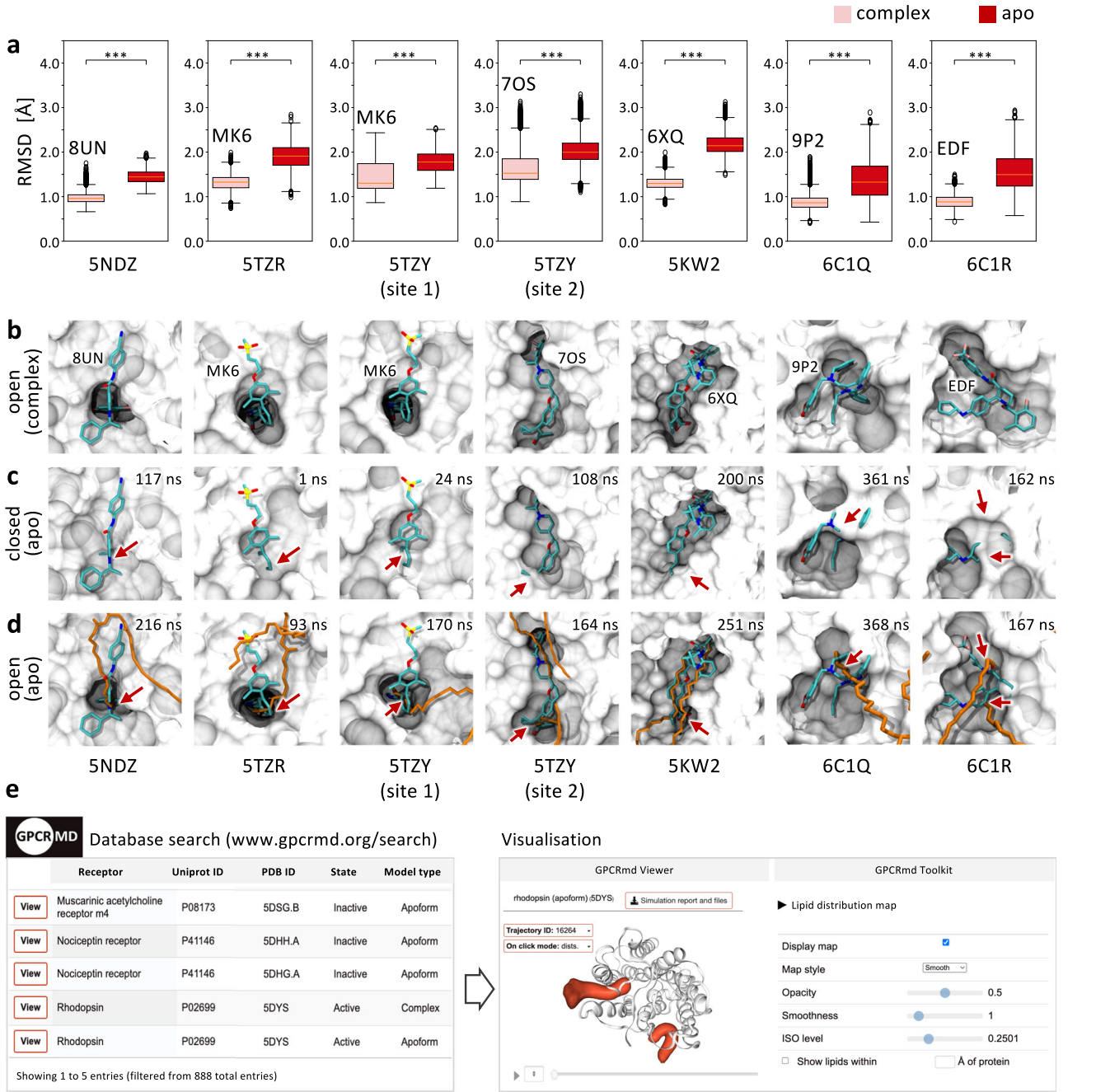


Fig. 5 | Flexibility of druggable allosteric sites. a RMSD boxplots reflect the flexibility of druggable allosteric sites in the presence (complex) and absence (apo) of the ligand computed for a 4 Å region around the ligand (PDB three letter code of the complexed ligand is indicated). For this, the experimentally solved structure is taken as a reference structure. The RMSDs for each box are derived from 7500 simulation frames ($n = 7500$, i.e., 3×2500 frames with one frame corresponding to 0.2 ns). The centre 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 $1.5 \times$ IQR and outliers are represented as circles. Statistical significance was assessed by a two-sided Mann-Whitney U test ($P = 0.0$ for all cases). **b** Surface representation of the experimentally solved allosteric site (white) in complex with the corresponding ligand (cyan). **c** Time evolution of the allosteric sites in the absence of the ligand (apo form) results in their complete (i.e., 5NDZ, 5TZR, 5TZY site 1) or partial closure (i.e., 5TZY site 2, 5KW2, 6C1Q, 6C1R) as indicated by red arrows. The original ligands have been superimposed to the simulated apo system's surface to indicate the location of the initial pocket. The corresponding simulated time of pocket closure is provided in the insets. **d** Lipid penetration facilitates pockets site opening evolving from (**c**) at the indicated simulation time. E.g., the 5NDZ site closes at 117 ns and reopens at 216 ns. Simulation snapshots of (**c**) and (**d**) are taken from 5NDZ (GPCRmd ID: 1018 [https://www.gpcrmd.org/view/1018], trajectory ID: 17253), 5TZR (GPCRmd ID: 793 [https://www.gpcrmd.org/view/793], trajectory ID: 15725), 5TZY site 1 (GPCRmd ID: 928 [https://www.gpcrmd.org/view/928], trajectory ID: 16644), 5TZY site 2 (GPCRmd ID: 928 [https://www.gpcrmd.org/view/928], trajectory ID: 16645), 5KW2 (GPCRmd ID: 765 [https://www.gpcrmd.org/view/765], trajectory ID: 15539), 6C1Q (GPCRmd ID: 1084 [https://www.gpcrmd.org/view/1084], trajectory ID: 18343), 6C1R (GPCRmd ID: 1085 [https://www.gpcrmd.org/view/1085], trajectory ID: 18351). The original ligands have been superimposed to the simulated apo system's surface to indicate the location of the initial pocket. **e** Searchable large scale collection of lipid insertions (left) and their visualisation on GPCRmd by mapping lipid distribution maps (orange volumetric map) on the 3D GPCR structure (right). Persistent lipid insertions are promising hotspots for allosteric sites and entrance channels. Source data are provided in the Source Data file.

inactive receptor states, highlighting an important area for further investigation in future studies.

Of note, several of our detected lipid insertion sites (TM1-TM7$_{EC}$ and TM3-TM4-TM5$_{IC}$) have been frequently associated with allosteric sites[21,45], which prompted us to explore the possibility of using lipid insertions to predict functionally relevant sites. Our subsequent study demonstrates that simulated membrane lipids efficiently identify hydrophobic pockets on the receptor surface previously resolved in experimental structures (Fig. 4a). Most importantly, we find that investigated allosteric sites exhibit remarkable flexibility, often tending to fully or partially close in the absence of an allosteric modulator (Fig. 5b–d). This suggests that experimentally observed allosteric pockets are mostly ligand-induced. Our finding has three significant implications for identifying allosteric drug target sites within GPCRs: First, our data suggests that, in the absence of a modulator, allosteric pockets sample a large conformational space including numerous open but also closed states which can obscure their druggability. This is reflected by the fact that, when allosteric modulators are not present, these pockets disappear from experimentally determined structures. Second, our data points to the existence of multiple unexplored pocket sites within the 3D GPCRome which have not yet been identified in available experimental structures. Third, according to our data, binding sites undergo closing and opening dynamics at the nanosecond timescale. Therefore, there is no efficient experimental method which is capable of thoroughly screening the hidden allosteric pockets with sufficient temporal and spatial resolution. Our study reveals a viable solution to overcome this challenge for GPCRs and any other type of membrane protein: leveraging simulations and lipid-protein interaction as a high-throughput strategy to reveal concealed pockets. Of note, we have compiled an extensive dataset on lipid dynamics and insertion (Supplementary Data 3 and 4) which holds important value for researchers searching for unexplored cavities to allosterically target GPCRs. This data can be interactively streamed for individual receptors on the GPCRmd web platform (Fig. 5e). For instance, the TM3-TM4-TM5$_{IC}$ site, which is a highly conserved lipid insertion site in our simulated GPCR dataset, has not yet been exploited for the apelin receptor (PDB id: 5VBL). Similarly, the TM1-TM7 site has never before hosted a synthetic ligand in an experimentally determined structure[21,25].

In addition to allosteric sites revealed by membrane lipids, we also observe sodium ions allosterically binding in the transmembrane bundle of class A GPCRs—a well-recognised phenomenon[46]. Our data unveil a diverse range of sodium binding profiles, including rapid, delayed, or absent binding at D^{2x50} (Supplementary Fig. 12). Consistent with previous findings[14], the structural characteristics of the extracellular entrance and the intramolecular channel guiding sodium to the allosteric site D^{2x50} appear to modulate the degree of ion binding (Supplementary Fig. 13).

Another important discovery from our study is the identification of entrance gateways from the lateral membrane side into the receptor core, which can serve as potential drug (un)binding pathways. They have been described for a few GPCRs (e.g., rhodopsins, cannabinoid, sphingosine-1-phosphate (S1P), Melatonin and GPR183)[30,32,47–49] but remain largely unexplored for the majority of receptors. Like allosteric pockets, these entrance gateways are highly dynamic and are often closed in experimentally solved structures. Our data demonstrates that membrane lipids are also able to identify such gateways and promote their opening during MD simulations (Supplementary Movies 2 and 3). Importantly, they can even occupy part of the orthosteric site (Fig. 4b). This finding further suggests that membrane lipids can naturally interact and at times compete with orthosteric ligands in GPCRs which aligns with previous observations for the $A_{2A}R$[50]. Altogether, our results highlight the significant role of lipid dynamics in receptor function. We anticipate that modifications of membrane lipid properties[51,52], such as tail length, saturation and polar head will further modulate these effects. This could contribute to diversification of GPCR function across different cell types and tissues with varying membrane compositions[53]. In addition to the reported lateral membrane gateways (e.g., in rhodopsin, CB_1 receptor and β_2-AR), we also uncover previously unknown channels (Supplementary Data 6 to 7). Insights into such potential drug binding and ligand entry pathways are critical for drug development as they determine the chemical space a GPCR recognises. This has been recently demonstrated for the GPR183 - a receptor that signals in response to hydrophobic and polar ligands entering through a lateral membrane gateway or the solvent-exposed extracellular opening, respectively[47].

In our study, we employ a membrane model composed entirely of POPC. This choice is motivated by the prevalence of diacylphosphatidylcholine lipids in animal cell membranes, where they constitute 40–60% of total cellular phospholipids[54] and by its status as a well-characterised lipid in computational research. We recognise that this simplified membrane composition does not fully capture the complexity of biological membranes, notably lacking cholesterol and other lipids that significantly influence GPCR dynamics[50,55–58]. Nonetheless, our results show that a pure POPC membrane setup effectively replicates essential lipid interactions, including lipid insertions and lateral ligand gateways, across a broad range of GPCR types (Fig. 3c). Future research will expand on these findings by systematically incorporating cholesterol and additional lipids to assess the effects of complex membrane environments on GPCR dynamics.

In summary, our resource opens possibilities for incorporating a dynamic view of GPCR plasticity into drug development targeting a large array of diseases (e.g., pain, addiction, inflammation, cancer, etc., Table 1). It provides innovative structural insights into unexplored receptor conformations, allosteric sites and entrance gateways that can be exploited by a large interdisciplinary community including medicinal chemists, pharmacogenomics experts, evolutionary biologists, computational chemists, biomedical data scientists, drug design experts or teachers and students from the natural and biomedical sciences. As a final remark, our study focuses on a selection of highly relevant GPCR features, providing valuable insights into GPCR function. However, the dataset we present extends far beyond these findings, offering significant potential for uncovering yet-to-be-explored features. These could include the discovery of cryptic (sub)pockets within the receptor interior or a systematic GPCR-wide analysis of intramolecular allosteric communication to advance our understanding of GPCR (sub)type-, or class-specific functions among others. For this purpose, all simulations are made freely available on the GPCRmd server for visualisation, analysis and download (www.gpcrmd.org) with the final aim to accelerate GPCR research and drug development.

Methods

Structure selection and curation

The structures simulated in the second edition of GPCRmd were obtained from the GPCRdb's refined structure database (November 2020 update)[59], in which all stabilising mutations have been reverted to wild-type counterparts. We selected all structures that were not co-solved with coupling partners (e.g., G proteins, β-arrestins, nanobodies, among others) and that had not been previously simulated in GPCRmd. During the curation process, structures were automatically protonated using the '*systemPrepare()*' function of the HTMD molecular environment (version 1.22.4)[60], which in turn is based on PROPKA pKa predictions[61]. Internal water molecules were added using the online web server HOMOLWAT[62]. The strongly conserved sodium ion near residue 2x50 was also added using HOMOLWAT in structures deemed inactive (bound to ligands identified as 'antagonist' or 'inverse agonist' by GPCRdb) and in apo form structures, where the sodium was allowed to bind spontaneously. After these automated steps, the structures were manually curated by experts from the GPCRmd

Table 1 | The GPCRmd dataset provides innovation with an added practical value for researchers

Innovation	Description	GPCRmd features[a]	Added practical value for researchers
Time-resolved dynamics of GPCRs implicated in the treatment of a large array of diseases	List of treatment indications and implicated receptors: (1) pain and addiction (opioid, cannabinoid receptors) (2) inflammation or cancer (chemokine, tachykinin, leukotriene, apelin, protease-activated, frizzled and neurotensin receptors) (3) obesity and metabolic disorders (GLP1R, complement peptide, neuropeptide Y, free fatty acid, glucagon receptors) (4) Neurological disorders and neurodegenerative diseases (dopamine, serotonin, adenosine, muscarinic, metabotropic glutamate receptors) (5) insomnia (orexin, melatonin receptors) (6) vision (opsin receptor) (7) cardiovascular disease (angiotensin, adrenergic, endothelin, platelet-activating factor, apelin, neurotensin, P2Y receptors) (8) osteoporosis and bone development disorders (parathyroid hormone receptor) (9) immune disorders (lysophospholipid, P2Y receptors)	GPCRmd viewer and toolkit	**Medicinal chemists**: Study of receptor dynamics, (in)activation, lipid interactions and ligand-receptor interactions for diverse ligand types **Pharmacogenomics experts**: Assessment of the dynamic environment of residues bearing naturally occurring genetic variants to explain their contribution to disease aetiology and differential drug response
Access to a large conformational landscape of GPCRs	Time-resolved opening and closing movements of GPCRs provide access to unexplored GPCR conformations: inactive-like, intermediate and active-like (Fig. 2e)	GPCRmd Breathing Tool	**Evolutionary biologists**: Evolutionary insights into how protein sequence associates with structural dynamics / constitutive activity across closely related receptor types **Computational chemists**: Inactive-like, intermediate and active-like GPCR conformations can be exploited for enriching virtual screening studies **Structural biologists**: Complement to static X-ray/CryoEM data allowing to incorporate a protein dynamics context **Biomedical data scientists**: A massive dataset for machine learning techniques, e.g., linking receptor dynamics data with signalling outcomes, providing a variety of starting structures to train conformation-aware protein modelling algorithms, etc.
Access to unexplored allosteric sites and entrance gateways	Allosteric sites and entrance gateways are highly flexible and often hidden. Lipid dynamics and their insertion into deep cavities can help disclose such functionally relevant sites (Figs. 4 and 5)	GPCRmd Lipid Insertion Tool	**Drug Design Experts**: Unexplored protein sites inspiring original targeting strategies to modulate GPCRs - a highly relevant drug target class **Computational structural biologists**: A protocol that can be applied to any other therapeutically relevant membrane protein (e.g., ion channels, transporters/carriers, tyrosine kinases)
Sharing cutting-edge research with trainees and educators	Exploring protein dynamics visually reveals fascinating processes related to protein physiology, such as receptor (in)activation, the emergence of allosteric sites or lateral channels into receptors	GPCRmd viewer and toolkit	**Natural/biomedical sciences teachers and students**: A tool for science education
Transparency and reproducibility	Input structures, simulation protocols and all generated data are available on GPCRmd (www.gpcrmd.org) GPCRmd simulation submission system	GPCRmd platform	**GPCR research community**: A key resource to accelerate basic and applied GPCR research and to ensure MD simulation data and metadata in scientific publications is citable, reproducible and publicly accessible

[a]Described tools are available on www.gpcrmd.org.

community, who ensured the high quality of the prepared structures to be submitted to the MD simulation process. This includes the remodelling and reconstruction of any missing loops whenever possible, and the revision of the following features:

- Sections and loops modelled by GPCRdb automated refinement algorithm.
- Waters placed by HOMOLWAT.
- Presence or absence of sodium near site 2x50.
- Protonation/tautomeric states of residues.
- Disulphide bridges.
- Chemical structure of non-protein ligand molecules.

Simulation protocol

The entire simulation protocol was performed automatically using a Python pipeline specifically developed for this purpose, based on HTMD (Fig. 6). The proper orientation of receptors within the lipid bilayer was obtained from the OPM database[63], with receptors embedded in a POPC membrane and solvated with TIP3P waters. The resulting systems were built using the '*charmm.build()*' function of HTMD. The system charge was neutralised and ionised using Na^+ and Cl^- ions at a concentration of 0.15 M. System parameters were obtained from CHARMM General Force Field (CGenFF) version c36/July2020[64]. Ligand parameters were assigned using the ParamChem online server[65,66]. Built systems were then subjected to a simulation protocol summarised in Table 2. Briefly, systems were equilibrated using ACEMD3 (version 3.2.3), and then three production runs were performed for each equilibrated system (see further details in Table 2). Running three independent replicates provides the advantage of sampling different regions of conformational space and enhances results reproducibility (i.e., no enhanced sampling is required)[14,67]. In addition, this approach is sufficient to allow the system to comprehensively explore the conformational space around the experimental

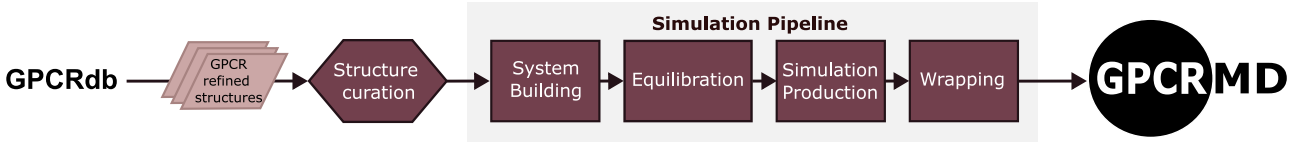


Fig. 6 | Schematic representation of the simulation protocol used for the 2nd simulation round. The parts included in the 'Simulation Pipeline' were performed automatically with a Python-based pipeline specifically designed for this purpose (https://github.com/GPCRmd/simulation_pipeline).

Table 2 | Parameters used for the equilibration and production runs of the 2nd simulation round

	Equilibration	Production
Simulation time	40 ns	500 ns
Timestep	2 fs	4 fs
Trajectory period[a]	0.05 ns (25000 integration steps)	0.2 ns (50000 integration steps)
Minimisation steps	5000	0
Conditions	Constant pressure (NPT)	Constant volume (NVT)
Barostat	Monte Carlo	-
Pressure	1 atm	-
Thermostat	Langevin	
Temperature	310 K	
Thermostat damping	1 ps^{-1}	0.1 ps^{-1}
Restraints applied to	Protein, co-solved ligand molecules, HOMOLWAT, and co-solved waters	-
Restraint strength	0–20 ns: 1 kcal/mol/A^2 20–30 ns: 1 kcal/mol/A^2 linearly tapered to 0 30–40 ns: 0 kcal/mol/A^2	-
Restraints type	Flat-bottom harmonic along the three axes	-
Restraint width	2 Å	-
Force field	CHARMM36, July 2020 version	
Replicates	1	3

[a]Trajectory period: Interval at which frames are saved into the output trajectory.

starting structure which minimises the dependence on the initial configuration. The final simulated systems were wrapped and aligned around the receptor using the MDAnalysis module[68] (version 2.1.0). To ensure data consistency, we created a Python-based pipeline for building, equilibrating, simulating and processing all systems in a fully automated manner, available at https://github.com/GPCRmd/simulation_pipeline. All systems have been uploaded into GPCRmd to ensure their availability to the scientific community. A MD checklist for data reproducibility is provided in Supplementary Table 4.

Conformational sampling

To determine the conformational state of a receptor in a given frame, we measured the distance between two alpha carbons in TM2 and TM6 respectively. The specific residues vary by GPCR class and are listed in Table 3, following the GPCRdb generic numbering scheme. To classify conformations as open, intermediate, or closed, we established a set of distance thresholds. Initially, we tried to determine the distance characteristic of fully 'closed' or 'open' receptor states. Distances indicative of 'closed' conformations were obtained from GPCRdb structures labelled as 'inactive' by their algorithm (https://docs.gpcrdb.org/structures.html). The distances expected for 'open' conformations were obtained from available structures co-solved with a G protein in GPCRdb. Distance value distributions comparing systems labelled as 'closed' and 'open' are displayed in Fig. 2a (right). Statistical significance between the 'closed' and 'open' distance sets was assessed using a two-sided Mann-Whitney *U* test. We then calculated the average distances in both 'open' and 'closed' datasets (Table 3) and the range between them was divided into quartiles. Any frame where the TM2-TM6 distance falls in the first quartile or below is considered 'closed'. Frames falling in the 2nd and 3rd quartiles, above the 'intermediate threshold' and below the 'open threshold', are classified as 'intermediate' (Table 3). Frames in the 4th quartile or above the range are considered 'open'. Using these ranges, we could assign a conformational state to every frame of every simulated replicate in the 2nd simulation round. Classes C and F were excluded from the analysis, as the difference between means was deemed too small to be meaningful.

The applied descriptor is meant to quantify the receptor opening at the intracellular side of the receptor, which is an important feature of receptor activation. However, complete activation can involve subtle structural alterations depending on the receptor class and (sub) type. Many of the algorithms to detect such changes are not scalable to MD simulation data analysis. Therefore, here we refer to the states identified by our descriptor as 'open/close' or 'active-like/inactive-like', which can be correlated to receptor activation but not completely.

Estimation of state transition rates between closed, intermediate, and open GPCR states

Breathing kinetics were estimated for the apo and corresponding complex form of class A and B1 GPCRs bound to an antagonist (90 systems), inverse agonist (9 systems), or NAM (8 systems) (see list of receptors in Supplementary Data 2). In a first step, the time series of the TM2-TM6 distances along the trajectories were smoothed with a cubic spline filter with a parameter $s = 300$ to reduce noise. The smoothed distance was discretized with the class-specific cut-offs for closed, intermediate and open states to obtain time-dependent (instantaneous) state assignments (see methods).

Due to the relatively small amount of transitions observed (rare events on simulated timescales), we modelled the breathing kinetics as two-state irreversible processes. For this, we assume that (1) systems start in the closed state; (2) within the observed time (3×500 ns) they may or may not transition to the intermediate or open state; (3) if the transition occurs, the observation time stops. Irreversible two-state models, which can be considered a special case of discrete-time Markov models, can be estimated quantitatively and reliably even in the presence of few and unobserved events[69] through the class of *survival analysis* statistical methods.

We built models for (1) the closed-to-intermediate and (2) the closed-to-open transitions. For each trajectory, we computed the first time of occurrence of the target state (intermediate or open, respectively), while trajectories for which the target state is never reached count towards the sampling time as 'censored' observations. We used the Kaplan-Meier estimator[70] to evaluate the time-dependent fraction of trajectories still in the closed state. The Kaplan-Meier formula calculates the still-in-closed-state 'survival' probability at time t by multiplying the probabilities of survival at each time point when a transition occurs, calculating these individual probabilities as (number of trajectories surviving - number of events) divided by (number of trajectories at risk) at each time point. Time-to-intermediate and time-to-open were compared between the apo and complex systems by log-rank tests and Cox proportional hazards models, testing the null

Table 3 | Threshold distances used in this analysis to define open, intermediate, and closed conformations in our simulations

Class	Residue pair (GPCR generic numbering)	Mean distance in inactive-labelled systems (Å)	Intermediate minimum threshold (Å)	Open minimum threshold (Å)	Mean distance in G protein-bound systems (Å)
A	2×46–6×37	12.18	13.6	16.36	17.74
B1	2×46–6×37	12.31	14.88	20.01	22.56
C	2×46–6×37	20.33	-	-	20.70
F	2×44–6×31	14.36	-	-	18.36

Fully active and inactive conformations were obtained from static structures stored in GPCRdb, G protein-bound and inactive-labelled systems respectively. Threshold columns represent the minimum distance at which we consider a GPCR in 'intermediate' or 'open' conformations'. Class C and F were not included in the conformational study. The distance value distributions of each class and state are shown in Fig. 2a (right).

hypothesis of no difference between the apo and complex systems. Finally, we estimated transition incidence rates dividing the number of events by the time sampled in the closed state, with 95% confidence intervals provided by Ulm's exact formula[71] (Supplementary Table 1).

Lipid insertions

To study lipid insertions, GPCRs are embedded in a homogenous membrane composed of POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine). Thanks to the use of a homogeneous membrane bilayer, we can exclude lipid interaction artefacts which may arise from a poorly converged multi-component membrane bilayer.

A visual representation of the method used to detect membrane lipid insertions in GPCR simulations is presented in Fig. 3a, of the results. To identify insertions in our simulations, we first established two sets with five mass centres per system: one from atoms in transmembrane helices TM2 to TM7 and another from helices TM1, TM2 and TM7 (the so-called receptor subcore). From these centres of mass, we created several cylindrical selections encompassing most of the receptor's 'inner core'. Any phospholipid found inside these selections in any frame is counted as 'inserted'. The widths of these selections vary: for shallow insertions major cylinders are 18 Å wide, while minor cylinders are 8 Å. To detect deep insertions, we applied a higher level of stringency, with major cylinders being 14 Å wide and minor ones being 6 Å. Only lipids that stay longer than 20% of their 'inserted' frames within this narrower selection are classified as 'deep' insertions, otherwise, they are considered 'shallow'. To quantify the insertions detected by this algorithm in the dataset, we count each lipid molecule that was inserted at some point of the trajectory as a single individual insertion. We assume that lipids mostly insert into a single pocket site during a simulation, as relevant insertions tend to be quite stable and simulations are typically too short for a lipid to experience two different relevant insertion sites. Lipid insertions are further classified, according to the simulation time spent inserted in the receptor, as ultra-short, short, medium, or long (Table 4).

To identify the receptor sites where lipid insertions occur, we analysed the protein residues within 3 Å of the inserted lipid. Subsequently, we classified these protein residues according to their TM and their proximity to either the intra- or the extracellular region of the GPCR. The two most frequent helices and the receptor side with the greatest number of residues were selected as the 'insertion site' for this lipid in this frame (e.g., TM1-TM2$_{IC}$).

Recovery of lipid insertions in static experimental structures was performed using essentially the same method presented above. We first visually inspected all structures in the dataset to identify all adjuvant molecules that could be compared to a phospholipid, or at least to a fatty acid. For instance, OLA, PLM and LPP, among others (see Supplementary Table 2), were the molecules we sought to identify within the cylinder selections to detect 'lipid insertions' in the structures. In assessing recovery, we looked for average, receptor-wise insertion frequencies at the site where the static insertion was found. If insertion frequency values in MD were above 4% (>20 ns on average in each system of the subtype), we considered it as recovered.

Table 4 | Classification of individual lipid insertions (i.e., produced by the same lipid molecule) according to the amount of time they stay inserted in a given simulation

Classification	Min. insertion time (ns)	Max. insertion time (ns)
Ultra-short	0	20
Short	20	100
Medium	100	300
Long	300	-

All occupancy maps shown in the figures were created with the 'VolMap' VMD plugin[72], version 1.1 (see examples in Fig. 4b and Supplementary Figs. 3, 6, 7, and 11).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All simulation files, parameters and protocols are available at GPCRmd https://www.gpcrmd.org/dynadb/publications/1535/. Individual submission codes and links to the simulation data on GPCRmd are listed in Supplementary Data 1. Data for receptor breathing motions are found in Supplementary Data 2. Data for lipid insertions frequencies (% of frames) per receptor site is provided in Supplementary Data 3 (shallow insertions) and Supplementary Data 4 (deep insertions). Data for insertions of lipid-like molecules found in experimental static structures are listed in Supplementary Data 5. Summary of all reported cases of lateral gateways identified through lipid insertions in our simulated systems are found in Supplementary Data 6. Additional information on individual insertions of particular systems is compiled in Supplementary Data 7. Source data are provided with this paper.

Code availability

The code for MD system building and simulation is available at https://github.com/GPCRmd/simulation_pipeline[73].

References

1. Kristiansen, K. Molecular mechanisms of ligand binding, signaling, and regulation within the superfamily of G-protein-coupled receptors: molecular modeling and mutagenesis approaches to receptor structure and function. *Pharmacol. Ther.* **103**, 21–80 (2004).
2. Mederos, Y. & Schnitzler, M. et al. Gq-coupled receptors as mechanosensors mediating myogenic vasoconstriction. *EMBO J.* **27**, 3092–3103 (2008).
3. Hauser, A. S. et al. Pharmacogenomics of GPCR Drug Targets. *Cell* **172**, 41–54.e19 (2018).
4. Lopez-Balastegui, M. et al. Relevance of G protein-coupled receptor (GPCR) dynamics for receptor activation, signalling bias and allosteric modulation. *Br. J. Pharmacol*. https://doi.org/10.1111/bph.16495 (2024).

5. Torrens-Fontanals, M. et al. How do molecular dynamics data complement static structural data of GPCRs. *Int. J. Mol. Sci.* **21**, 5933 (2020).
6. Wu, Y. et al. MD simulations revealing special activation mechanism of cannabinoid receptor 1. *Front. Mol. Biosci.* **9**, 860035 (2022).
7. Jang, J., Kim, S.-K., Guthrie, B. & Goddard, W. A. I. Synergic effects in the activation of the sweet receptor GPCR heterodimer for various sweeteners predicted using molecular metadynamics simulations. *J. Agric. Food Chem.* **69**, 12250–12261 (2021).
8. Bruzzese, A., Dalton, J. A. R. & Giraldo, J. Statistics for the analysis of molecular dynamics simulations: providing P values for agonist-dependent GPCR activation. *Sci. Rep.* **10**, 19942 (2020).
9. Miao, Y. & McCammon, J. A. G-protein coupled receptors: advances in simulation and drug discovery. *Curr. Opin. Struct. Biol.* **41**, 83–89 (2016).
10. Raniolo, S. & Limongelli, V. Ligand binding free-energy calculations with funnel metadynamics. *Nat. Protoc.* **15**, 2837–2866 (2020).
11. Ibrahim, P. & Clark, T. Metadynamics simulations of ligand binding to GPCRs. *Curr. Opin. Struct. Biol.* **55**, 129–137 (2019).
12. Morales-Pastor, A. et al. In silico study of allosteric communication networks in GPCR signaling bias. *Int. J. Mol. Sci.* **23**, 7809 (2022).
13. Vaidehi, N. & Bhattacharya, S. Allosteric communication pipelines in G-protein-coupled receptors. *Curr. Opin. Pharmacol.* **30**, 76–83 (2016).
14. Rodríguez-Espigares, I. et al. GPCRmd uncovers the dynamics of the 3D-GPCRome. *Nat. Methods* **17**, 777–787 (2020).
15. Kooistra, A. J. et al. GPCRdb in 2021: integrating GPCR sequence, structure and function. *Nucleic Acids Res.* **49**, D335–D343 (2021).
16. Pándy-Szekeres, G. et al. GPCRdb in 2023: state-specific structure models using AlphaFold2 and new ligand resources. *Nucleic Acids Res.* **51**, D395–D402 (2023).
17. Herrera, L. P. T. et al. GPCRdb in 2025: adding odorant receptors, data mapper, structure similarity search and models of physiological ligand complexes. *Nucleic Acids Res*. gkae1065. https://doi.org/10.1093/nar/gkae1065 (2024).
18. Kooistra, A. J., Munk, C., Hauser, A. S. & Gloriam, D. E. An online GPCR structure analysis platform. *Nat. Struct. Mol. Biol.* **28**, 875–878 (2021).
19. Shao, Z. et al. Molecular insights into ligand recognition and activation of chemokine receptors CCR2 and CCR3. *Cell Discov.* **8**, 1–11 (2022).
20. Ibrisimovic, E. et al. Constitutive activity of the A2A adenosine receptor and compartmentalised cyclic AMP signalling fine-tune noradrenaline release. *Purinergic Signal.* **8**, 677–692 (2012).
21. Hedderich, J. B. et al. The pocketome of G-protein-coupled receptors reveals previously untargeted allosteric sites. *Nat. Commun.* **13**, 2567 (2022).
22. 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).
23. Che, T. et al. Nanobody-enabled monitoring of kappa opioid receptor states. *Nat. Commun.* **11**, 1145 (2020).
24. Warne, T. et al. The structural basis for agonist and partial agonist action on a β1-adrenergic receptor. *Nature* **469**, 241–244 (2011).
25. Persechino, M., Hedderich, J. B., Kolb, P. & Hilger, D. Allosteric modulation of GPCRs: from structural insights to in silico drug discovery. *Pharmacol. Ther.* **237**, 108242 (2022).
26. Heydenreich, F. M. et al. Molecular determinants of ligand efficacy and potency in GPCR signaling. *Science* **382**, eadh1859 (2023).
27. Sun, B. et al. Crystal structure of the adenosine A2A receptor bound to an antagonist reveals a potential allosteric pocket. *Proc. Natl. Acad. Sci. USA* **114**, 2066–2071 (2017).
28. Lu, J. et al. Structural basis for the cooperative allosteric activation of the free fatty acid receptor GPR40. *Nat. Struct. Mol. Biol.* **24**, 570–577 (2017).
29. Grundmann, M., Bender, E., Schamberger, J. & Eitner, F. Pharmacology of free fatty acid receptors and their allosteric modulators. *Int. J. Mol. Sci.* **22**, 1763 (2021).
30. Hildebrand, P. W. et al. A ligand channel through the G protein coupled receptor opsin. *PLOS ONE* **4**, e4382 (2009).
31. Pirona, L., Ballabio, F., Alfonso-Prieto, M. & Capelli, R. Calcium-driven in silico inactivation of a human olfactory receptor. *J. Chem. Inf. Model.* **64**, 2971–2978 (2024).
32. Jakowiecki, J. & Filipek, S. Hydrophobic ligand entry and exit pathways of the CB1 cannabinoid receptor. *J. Chem. Inf. Model.* **56**, 2457–2466 (2016).
33. 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).
34. Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **9**, 646–652 (2002).
35. Maginn, E. J. & Elliott, J. R. Historical Perspective and current outlook for molecular dynamics as a chemical engineering tool. *Ind. Eng. Chem. Res.* **49**, 3059–3078 (2010).
36. Sinha, S., Tam, B. & Wang, S. M. Applications of molecular dynamics simulation in protein study. *Membranes* **12**, 844 (2022).
37. Ibrahim, P., Wifling, D. & Clark, T. Universal activation index for class A GPCRs. *J. Chem. Inf. Model.* **59**, 3938–3945 (2019).
38. Grahl, A., Abiko, L. A., Isogai, S., Sharpe, T. & Grzesiek, S. A high-resolution description of β1-adrenergic receptor functional dynamics and allosteric coupling from backbone NMR. *Nat. Commun.* **11**, 2216 (2020).
39. Bumbak, F. et al. Ligands selectively tune the local and global motions of neurotensin receptor 1 (NTS1). *Cell Rep.* **42**, 112015 (2023).
40. Nobles, M., Benians, A. & Tinker, A. Heterotrimeric G proteins precouple with G protein-coupled receptors in living cells. *Proc. Natl. Acad. Sci. USA* **102**, 18706–18711 (2005).
41. Zeghal, M., Laroche, G., Freitas, J. D., Wang, R. & Giguère, P. M. Profiling of basal and ligand-dependent GPCR activities by means of a polyvalent cell-based high-throughput platform. *Nat. Commun.* **14**, 3684 (2023).
42. Lu, S., Jang, W., Inoue, A. & Lambert, N. A. Constitutive G protein coupling profiles of understudied orphan GPCRs. *PLoS ONE* **16**, e0247743 (2021).
43. Sadler, F. et al. Autoregulation of GPCR signalling through the third intracellular loop. *Nature* **615**, 734–741 (2023).
44. Hilger, D. The role of structural dynamics in GPCR-mediated signaling. *FEBS J.* **288**, 2461–2489 (2021).
45. Wakefield, A. E., Mason, J. S., Vajda, S. & Keserű, G. M. Analysis of tractable allosteric sites in G protein-coupled receptors. *Sci. Rep.* **9**, 6180 (2019).
46. Katritch, V. et al. Allosteric sodium in class A GPCR signaling. *Trends Biochem. Sci.* **39**, 233–244 (2014).
47. Kjær, V. M. S. et al. Ligand entry pathways control the chemical space recognized by GPR183. *Chem. Sci.* **14**, 10671–10683 (2023).
48. Stanley, N., Pardo, L. & Fabritiis, G. D. The pathway of ligand entry from the membrane bilayer to a lipid G protein-coupled receptor. *Sci. Rep.* **6**, 22639 (2016).
49. Stauch, B. et al. Structural basis of ligand recognition at the human MT1 melatonin receptor. *Nature* **569**, 284–288 (2019).
50. Guixà-González, R. et al. Membrane cholesterol access into a G-protein-coupled receptor. *Nat. Commun.* **8**, 14505 (2017).
51. Mizumura, T. et al. Activation of adenosine A2A receptor by lipids from docosahexaenoic acid revealed by NMR. *Sci. Adv.* **6**, eaay8544 (2020).
52. Thakur, N. et al. Anionic phospholipids control mechanisms of GPCR-G protein recognition. *Nat. Commun.* **14**, 794 (2023).

53. Baccouch, R., Rascol, E., Stoklosa, K. & Alves, I. D. The role of the lipid environment in the activity of G protein coupled receptors. *Biophys. Chem.* **285**, 106794 (2022).
54. van der Veen, J. N. et al. The critical role of phosphatidylcholine and phosphatidylethanolamine metabolism in health and disease. *Biochim. Biophys. Acta Biomembr.* **1859**, 1558–1572 (2017).
55. Duncan, A. L., Song, W. & Sansom, M. S. P. Lipid-dependent regulation of ion channels and g protein–coupled receptors: insights from structures and simulations. *Annu. Rev. Pharmacol. Toxicol.* **60**, 31–50 (2020).
56. 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).
57. Sejdiu, B. I. & Tieleman, D. P. Lipid-protein interactions are a unique property and defining feature of G protein-coupled receptors. *Biophys. J.* **118**, 1887–1900 (2020).
58. Kumar, G. A. et al. A molecular sensor for cholesterol in the human serotonin1A receptor. *Sci. Adv.* **7**, eabh2922 (2021).
59. Pándy-Szekeres, G. et al. GPCRdb in 2018: adding GPCR structure models and ligands. *Nucleic Acids Res.* **46**, D440–D446 (2018).
60. Doerr, S., Harvey, M. J., Noé, F. & De Fabritiis, G. HTMD: high-throughput molecular dynamics for molecular discovery. *J. Chem. Theory Comput.* **12**, 1845–1852 (2016).
61. Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J. Chem. Theory Comput.* **7**, 525–537 (2011).
62. Mayol, E. et al. HomolWat: a web server tool to incorporate 'homologous' water molecules into GPCR structures. *Nucleic Acids Res.* **48**, W54–W59 (2020).
63. Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370–D376 (2012).
64. Vanommeslaeghe, K. et al. CHARMM General Force Field (CGenFF): a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **31**, 671–690 (2010).
65. Vanommeslaeghe, K. & MacKerell, A. D. Jr. Automation of the CHARMM general force field (CGenFF) I: bond perception and atom typing. *J. Chem. Inf. Model.* **52**, 3144–3154 (2012).
66. Vanommeslaeghe, K., Raman, E. P. & MacKerell, A. D. Jr. Automation of the CHARMM general force field (CGenFF) II: assignment of bonded parameters and partial atomic charges. *J. Chem. Inf. Model.* **52**, 3155–3168 (2012).
67. Knapp, B., Ospina, L. & Deane, C. M. Avoiding false positive conclusions in molecular simulation: the importance of replicas. *J. Chem. Theory Comput.* **14**, 6127–6138 (2018).
68. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319–2327 (2011).
69. Giorgino, T., Buch, I. & De Fabritiis, G. Visualizing the induced binding of SH2-phosphopeptide. *J. Chem. Theory Comput.* **8**, 1171–1175 (2012).
70. Kaplan, E. L. & Meier, P. Nonparametric estimation from incomplete observations. *J. Am. Stat. Assoc.* **53**, 457–481 (1958).
71. Ulm, K. A simple method to calculate the confidence interval of a standardized mortality ratio (SMR). *Am. J. Epidemiol.* **131**, 373–375 (1990).
72. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
73. Aranda-García, D. et al. Large scale investigation of GPCR molecular dynamics data uncovers allosteric sites and lateral gateways. Zenodo. https://doi.org/10.5281/zenodo.14678007 (2025).
74. Gulati, S. et al. Photocyclic behavior of rhodopsin induced by an atypical isomerization mechanism. *Proc. Natl. Acad. Sci. USA* **114**, E2608–E2615 (2017).

Acknowledgements

D.A.-G. acknowledges financial support from the Catalan Department of Business and Innovation (2020FISDU/487). A.G.-R. has an Investigo Programme contract funded by the European Union, Next Generation EU. T.M.S. acknowledges support from the Sara Borrell grant CD22/00007 funded by the Institute of Health Carlos III (ISCIII) and resources of grant 2021 SGR 00046 funded by Agència de Gestió d'Ajuts Universitaris i de Recerca Generalitat de Catalunya (AGAUR). B.M. acknowledges financial support from Agència de Gestió d'Ajuts Universitaris i de Recerca Generalitat de Catalunya (AGAUR) (2022 DI 00112) and DIN2021 ayudas para doctorados industriales (DIN2021-011749). A.P.-G. acknowledges support from the Spanish Ministerio de Universidades FPU program (FPU20/03069). T.G. acknowledges financial support from the Spoke 7 of the National Centre for HPC, Big Data and Quantum Computing (Centro Nazionale 01 – CN0000013), funded by the European Union -- NextGenerationEU, Mission 4, Component 2, Investment line 1.4, CUP B93C22000620006; from the PRIN 2022 (BioCat4BioPol) from the Ministero dell'Università e Ricerca, funded by the European Union -- NextGenerationEU, Mission 4 Component C2, CUP B53D23015140006; from the Spoke 5 "Next-Gen Therapeutics" of PNRR M4C2 Investiment 1.3 "HEAL ITALIA" PE_00000019, CUP H43C22000830006 project "PROPHECY-GlycoRARE"; and from the project InvAt-Invecchiamento Attivo e in Salute (FOE 2022) CUP B53C22010140001. We acknowledge CINECA awards aGPCR1 and aGPCR2 under the ISCRA initiative for the availability of high-performance computing resources and support. MMB acknowledges support from ALSAC. JS acknowledges funding from the Horizon Europe Project Obelisk under the grant agreement 101080465, the ERA-NET NEURON & Ministry of Economy, Industry, Competitiveness [AC18/00030] and the MICIU/AEI /10.13039/501100011033 and FEDER, UE (PID2022-137161OB-I00). D.E.G. acknowledges funding from the Novo Nordisk Foundation (NNF18OC0031226) and Lundbeck Foundation (R383-2022-306). H.G.-T. acknowledges funding from Ministerio de Ciencia e Innovación (PID2023-150793OB-I00). I.G.T. is supported by the Biotechnology and Biosciences Research Council responsive mode award BB/R007101/1. M.M-S. is supported by a Royal Society University Research Fellowship (URF\R1\221205). A.D.P. is supported by the Leibniz Programme for Women Professors (grant: P116/2020). D.A.-G., T.M.S, A.G.-R., A.S., A.D.P., D.E.G., G.P-S., R.G.G., T.G., I.G.T, H.G.-T., P.K., M.M.-S., J.S. have been members of COST Action CA18133 'ERNEST'.

Author contributions

D.A.G. scripted the simulation pipeline, performed MD simulations, programmed and performed all presented analyses, prepared figures and wrote the initial version of the manuscript together with J.S. M.T.-F. provided supervision and technical assistance during most of the project. T.M.S., M.T.-F., A.G.-R., M.L.-B., B.M.-L., A.M.-P., A.P.-G., M.D.-E., D.S.-N., T.D., M.D., C.J., J.J., W.J., M.J.-R., V.J.-Y.-L., A.N., U.O., A.S., J.K.T., E.K., D.L., R.G.-G., T.G., H.G.-T., I.T., A.D.P., S.F., P.K., A.C. and M.M.-S. curated GPCR structures for their subsequent simulation. G.P.-S. and D.E.G. provided refined GPCR structures and advice in the curation process. M.T.-F., M.L.-B., B.M.-L., A.M.-P., A.P.-G., F.N.-F., O.C., M.D.-E., D.S.-N., V.L.-M. and S.S.-D. manually revised equilibration and production runs of simulated systems ensuring their quality. P.W.H., M.F. and M.M.B. contributed to the data interpretation. T.G. computed the state transition kinetics during receptor breathing. J.M. assessed receptor flexibility in the light of available experimental data. M.M.-S. supported the general data analysis. All authors contributed to the writing and revision of the manuscript. J.S. supervised the project, conceptualised the studies and wrote together with D.A.G. the final version of the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-57034-y.

Correspondence and requests for materials should be addressed to Jana Selent.

Peer review information *Nature Communications* thanks Riccardo Capelli, Timothy Clark and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

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 Medicine and Life Sciences, Pompeu Fabra University (UPF), Barcelona, Spain. [2]Research Programme on Biomedical Informatics (GRIB), Hospital del Mar Medical Research Institute (IMIM), Barcelona, Spain. [3]InterAx Biotech AG, Villigen, Switzerland. [4]Acellera Labs, Barcelona, Spain. [5]School of Pharmacy, Medical Biology Centre, Queen's University Belfast, Belfast, Northern Ireland, UK. [6]Department of Pharmaceutical Chemistry, University of Marburg, Marburg, 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, Birmingham and Nottingham, Midlands, UK. [12]Sygnature Discovery Ltd., Nottingham, UK. [13]Leibniz Institute for Food Systems Biology at the Technical University of Munich, Freising, Germany. [14]Medizinische Fakultät, Institut für Medizinische Physik und Biophysik, Universität Leipzig, Leipzig, Germany. [15]Department of Drug Design and Pharmacology, University of Copenhagen, Copenhagen, Denmark. [16]Medicinal Chemistry Research Group, HUN-REN Research Center for Natural Sciences, Budapest, 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), Barcelona, Spain. [19]Department of Cell and Molecular Biology, Uppsala University, Biomedical Center, Uppsala, Sweden. [20]Research Center in Nanomaterials and Nanotechnology (CINN/CSIC) and Health Institute of Principado de Asturias (ISPA), 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, USA. [24]Chemoinformatics and Protein Modelling, TUM School of Life Sciences, Technical University of Munich, Freising, Germany. [25]Department of Biochemistry and Molecular Biology, Faculty of Biosciences, Universitat Autònoma de Barcelona (UAB), Cerdanyola del Vallès, Spain. [26]Institute of Biophysics (IBF-CNR), National Research Council of Italy, Milano, Italy. [27]Department of Biosciences, University of Milan, Milan, Italy. [28]Department of Pharmacology, University of Cambridge, Cambridge, UK.
✉e-mail: jana.selent@upf.edu