

This document is confidential and is proprietary to the American Chemical Society and its authors. Do not copy or disclose without written permission. If you have received this item in error, notify the sender and delete all copies.

Probe Confined Dynamic Mapping for GPCR Allosteric Site Prediction

Journal:	<i>ACS Central Science</i>
Manuscript ID	oc-2021-00802q
Manuscript Type:	Research Article
Date Submitted by the Author:	05-Jul-2021
Complete List of Authors:	Ciancetta, Antonella; University of Ferrara Department of Chemical and Pharmaceutical Sciences Gill, Amandeep ; Queen Mary University of London Ding, Tianyi; Queen's University Belfast Karlov, Dmitry; Queen's University Belfast Chalhoub, George; Queen Mary University of London McCormick, Peter ; Queen Mary University of London Tikhonova, Irina; Queen's University Belfast,

SCHOLARONE™
Manuscripts

Probe Confined Dynamic Mapping for GPCR Allosteric Site Prediction

Antonella Ciancetta^{1#§}, Amandeep Kaur Gill^{2‡}, Tianyi Ding¹, Dmitry S. Karlov¹, George Chalhoub², Peter J. McCormick² and Irina G. Tikhonova¹

¹School of Pharmacy, Medical Biology Centre, Queen's University Belfast, Belfast, Northern Ireland, BT9 7BL, UK

²Centre for Endocrinology, William Harvey Research Institute, Bart's and the London School of Medicine and Dentistry, Queen Mary, University of London, London, EC1M 6BQ, UK

[#]Led computational chemistry effort.

[‡]Led molecular pharmacology effort.

* Correspondence, Email: i.tikhonova@qub.ac.uk and p.mccormick@qmul.ac.uk

[§] Present Address: Department of Chemical, Pharmaceutical and Agricultural Sciences, University of Ferrara, 44121, Ferrara, Italy

KEYWORDS: molecular dynamics; fragment-based drug design; enhanced sampling; cosolvent mapping; fragment mapping; probe mapping, mutagenesis

Abstract

Targeting G protein-coupled receptors (GPCRs) through allosteric sites offer advantages over orthosteric sites in identifying drugs with increased selectivity and potentially reduced side effects. In this study, we have developed a probe confined dynamic mapping protocol that allows the prediction of allosteric sites at both the GPCR extracellular and intracellular sides, as well as at the receptor-lipid interface. The applied harmonic wall potential enhanced sampling of probe molecules in a selected area of a GPCR while preventing membrane distortion in molecular dynamics simulations. The specific probes derived from GPCR allosteric ligand structures performed better in allosteric site mapping compared to commonly used cosolvents. The M₂ muscarinic, β₂ adrenergic and P₂Y₁ purinergic receptors were selected for protocol's retrospective validation. The protocol was next validated prospectively to locate the binding site of [5-fluoro-4-(hydroxymethyl)-2-methoxyphenyl]-(4-fluoro-1*H*-indol-1-yl)methanone at the D₂ dopamine receptor and subsequent mutagenesis confirmed the prediction. The protocol provides fast and efficient prediction of key amino acid residues surrounding allosteric sites in membrane proteins and facilitates structure-based design of allosteric modulators.

Introduction

G protein-coupled receptors (GPCRs) are the largest membrane protein family consisting of some 800 members that transduce a signal inside cells from a variety of endogenous ligands including hormones, neurotransmitters, metabolites, pheromones, odorants and light. As a result of their broad influence on human physiology, GPCRs are drug targets in many therapeutic areas such as inflammation, metabolic and neurological disorders, pain, addiction, infertility, viral infections and cancer. 475 GPCR drugs (34% of all drugs) are currently approved by the US Food and Drug Administration (FDA) and ≥ 300 GPCR agents are currently in clinical trials (1). Although GPCR drugs have shown substantial therapeutic success, developing drugs for many GPCR subfamilies has proved challenging. A key challenge is to achieve selectivity when targeting highly conserved orthosteric sites where the endogenous ligands bind.

Most GPCRs can be modulated by small-molecule ligands binding to allosteric sites that are spatially and topologically distinct from the orthosteric sites. Both positive allosteric modulators (PAM), which enhance the binding and signalling of orthosteric agonists, and negative allosteric modulators (NAM), which reduce the activity of orthosteric agonists, have been described. Allosteric drugs have a better potential for receptor subtype selectivity due to greater sequence variability in allosteric sites. Furthermore, allosteric modulators confer agonist dependence and functional selectivity, causing selective receptor activation and thus different tolerance in chronic diseases (2–4). In addition, allosteric modulators have a diverse relationship between duration and intensity of the effect, which can prolong the therapeutic effect without drug overdose (5). Two recent FDA-approved allosteric drug examples are cinacalcet, a PAM of the calcium-sensing receptor against hyperparathyroidism and maraviroc, a NAM of the chemokine CCR5 receptor to prevent the entry of HIV-1 (1). Despite the clear potential benefit of GPCR allosteric modulation, the discovery of allosteric sites and lead compounds has been mostly serendipitous often involving random screening of compound libraries. Recent X-ray crystallography and cryo-electron microscopy (cryo-EM) structures of several GPCRs solved with bound allosteric modulators have revealed remarkably diverse locations of the allosteric binding sites (6,7). Allosteric drugs can reside inside the helical bundle from the extracellular (EC) and intracellular (IC) sides, as well as outside the helical bundle at the lipid interface (LI). For example, in the X-ray ternary complex of the M_2 muscarinic receptor LY2119620, a PAM binds at an allosteric pocket facing the EC medium, including extracellular loop 2 and 3 (ECL2 and ECL3), which lies above the orthosteric site that is occupied by the agonist Iperoxo (8) (**Figure 1A**). In the X-ray complex of the β_2 adrenergic receptor with Cmp-15, this NAM occupies an allosteric site facing the IC side, involving intracellular loop 1 (ICL1) and the tips of helices 1, 2, 7 and 8, which is distant from the orthosteric site (9) (**Figure 1B**). Allosteric modulators sitting at an IC allosteric site have also been found in the X-ray complexes of the chemokine CCR2, CCR7 and CCR9 receptors (10–12).

Of particular interest are the allosteric sites at the protein-lipid interface, identified by X-ray or cryo-EM and confirmed by site-directed mutagenesis. **Figure 1C** shows the X-ray complex of the purinergic P₂Y₁ receptor and BPTU, a NAM occupying an allosteric site facing the membrane environment and in contact with helices 1-3 and ECL1 (13). Allosteric sites at different LI locations were found in the complexes with allosteric modulators of the cannabinoid CB₁ (14), complement C_{5a} (15,16), corticotropin-releasing factor 1 CRF₁ (17), free fatty acid FFA₁ (18), glucagon GCG (19), glucagon-like peptide-1 GLP-1 (20,21) and proteinase-activated PAR₂ (22) and β₂ adrenergic (23) receptors.

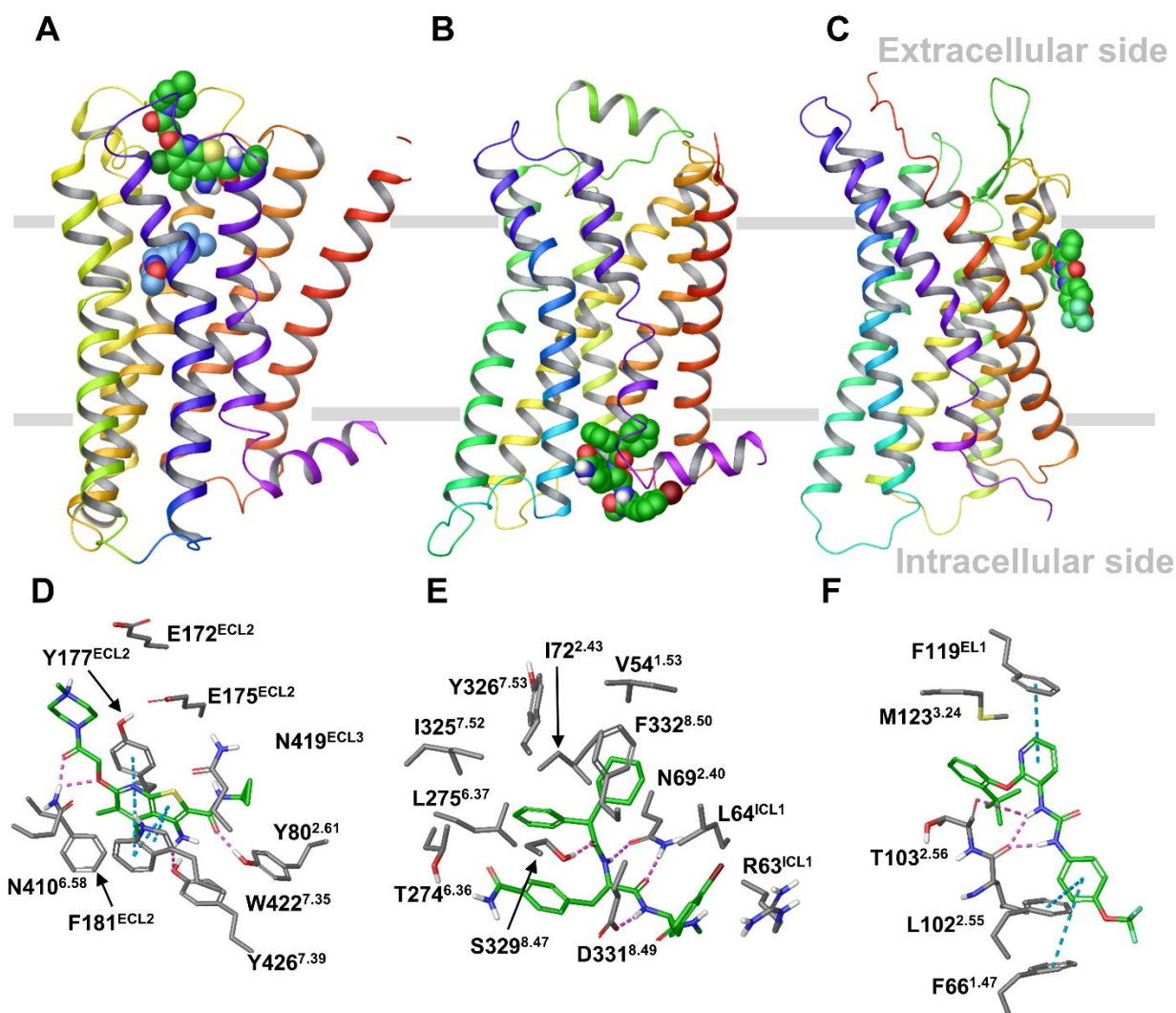


Figure 1. Locations of GPCR allosteric binding sites in the X-ray structures of the M₂, β₂, and P₂Y₁ receptors in complex with allosteric modulators. A-C: The overall view of the receptors with a bound allosteric modulator. **A:** The M₂ receptor bound to LY2119620 (green), a PAM binding at the extracellular side, and the orthosteric ligand, Iperoxo (blue). **B:** The β₂ receptor bound to Cmp-15, a NAM, at the intracellular side. **C:** The P₂Y₁ receptor in complex with BPTU, a NAM bound at the lipid interface. The receptors are in wild type with the rebuilt short intracellular loop 3 (ICL3) fragment. **D-F:** The binding interactions between the allosteric modulator and the receptor obtained from MD simulations of the X-ray receptor-ligand complexes for the M₂, β₂, and P₂Y₁ receptors. The key residues forming strong interactions and the allosteric ligands are in grey and green stick, respectively. Hydrogen bonds and π-π interactions are shown as pink and cyan dashed lines, respectively.

1
2
3 1 These structural data provide the first insights into GPCR allosteric regulation by small molecules and
4 offer opportunities to develop computer-aided methodologies to search for allosteric sites. As the lipid
5 2 bilayer plays a role in the formation of such sites, accounting for the receptor in a realistic environment
6 3 and its dynamics is important for accurate mapping of the allosteric sites.
7 4

8 5 Molecular dynamics (MD) simulations have become an indispensable tool for studying the structure
9 6 and dynamics of drug targets in the cellular environment and predicting ligand binding sites (24).
10 7 Among MD-based computational techniques developed to identify binding sites in proteins, cosolvent
11 8 mapping has recently garnered wide interest (25). In MD-based cosolvent mapping, small organic
12 9 molecules such as isopropanol, acetamide, pyridine and others are used as probes to map the binding
13 10 sites. A molecular probe is a prototype molecule containing polar and/or non-polar groups that can
14 11 quickly diffuse into protein cavities during MD simulations, thus identifying such cavities as accessible
15 12 and therefore as potential binding sites for allosteric modulators. In addition, MD-based cosolvent
16 13 mapping can directly account for protein motion during the site identification process. This approach
17 14 has been developed mainly for soluble proteins to map putative binding sites on the protein surface
18 15 (25). There are only a few examples to date where cosolvent mapping has been used for membrane
19 16 proteins (26–28). Although these studies have successfully mapped the binding sites of membrane-
20 17 bound proteins, the proposed protocols could be challenging to sample all possible locations of
21 18 allosteric sites in GPCRs as identified by X-ray crystallography and cryo-EM while avoiding probe
22 19 non-specific binding and membrane distortion.

23 20 Here, we propose a novel and efficient MD-based probe mapping protocol that is capable of exploring
24 21 all the possible scenarios of allosteric site locations known to date, including the most challenging
25 22 case where the allosteric site resides at the LI. Our methodology overcomes the limitations of
26 23 standard cosolvent mapping protocols through the application of a cylindrical harmonic wall potential
27 24 that enhances probe sampling in a selected area of the receptor. In addition, we use probes derived
28 25 from GPCR allosteric ligands that perform better in mapping allosteric sites compared to organic
29 26 solvents. Our protocol represents a fully automatized pipeline including system setup and simulations
30 27 for different scenarios. We used three exemplar receptors, i.e., the M₂, β₂ and P₂Y₁ receptors (**Figure**
31 28 **1**) for the protocol retrospective validation. We next applied the protocol in a prospective validation
32 29 scenario by predicting the binding site of [5-fluoro-4-(hydroxymethyl)-2-methoxyphenyl]-(4-fluoro-1H-
33 30 indol-1-yl)methanone (the UCB compound), a PAM at the dopamine D₂ receptor and validating the
34 31 prediction by site-directed mutagenesis. The outlined computational approach will facilitate structure-
35 32 based allosteric drug design by predicting receptor binding sites of known allosteric modulators for
36 33 further optimization and/or by identifying binding fragments that could be developed into new allosteric
37 34 modulators.
38 35
39 36
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1 Results

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

In line with previously published work on cosolvent simulations of soluble proteins (25,29–31) and lipid (32–34), while the optimal cosolvent concentration range for soluble proteins is 5-20%, for membrane-embedded proteins the maximum tolerated concentration should be lower, in the 2.8–5.6% range. This is due to membrane distortion caused by higher concentration of organic solvents. Cosolvents are also known to be technically challenging to use in combination with lipid bilayers as they tend to partition from the water layer and be adsorbed and redistribute into the membrane after just a few nanoseconds of simulation (32–34).

To overcome the membrane distortion as the result of cosolvent diffusion into the membrane and to keep a suitable sampling of probe molecules in the receptor while using small concentrations, we applied a cylinder-shaped wall potential to allow the movement of the cosolvent molecules only within the GPCR EC and IC openings or within a defined area at the LI. With such a restrain, as a cosolvent molecule reaches the wall of the cylinder, a repulsive bias is applied to prevent it from visiting regions outside the cylinder. We considered the mapping of different allosteric site locations as separate simulation protocols. Thus, to explore allosteric sites at the EC side, probe molecules were placed at the top water layer (the M₂ receptor case). In the case of an allosteric site at the IC side, the probe molecules were placed at the bottom water layer (the β₂ receptor case). As the LI site in the P₂Y₁ receptor is located close to the EC side, the probe molecules were placed at the top water layer to accelerate probe diffusion to the membrane. This is also supported by the recent MD simulations of BPTU binding, which suggests its entrance to the binding site occurs from the EC side (35). The use of a wall potential allowed us to use a 10% probe concentration.

Probe Confined Dynamic Mapping

Our *ad hoc* protocols (workflow of the M₂ receptor as an example shown in Figure 2A) enable MD co-solvent and fragment system setup, equilibration, and production of membrane-bound receptors. In the system setup, a water-probe (co-solvents or fragments) box is generated and placed at a user-defined distance at either the receptor EC and IC side with (for PAMs) or without (for NAMs) a bound orthosteric ligand. The box height (z dimension) is specified by the user, whereas the box width and depth (x and y sizes) are calculated based on minimum/maximum protein dimensions on the EC/IC sides. After the box is placed, the protein is embedded in the membrane and the system is solvated and neutralized.

During the equilibration, the probe molecules are confined in a closed cylinder and not allowed to interact with the protein-membrane system, (**Figure 2A**). During the production, to map EC/IC

allosteric sites, the cylinder boundary facing the protein is removed to allow the probes to diffuse towards and interact with the protein. During the production, the *distanceZ* collective variable facing the system boundary is maintained, thus defining a semi-closed cylinder, and its boundary lowered by ~ 10 Å (**Figure 2A**). This enables the probes to be confined in the periodic cell and increase their probability to interact with the protein. To sample putative allosteric sites at the LI, during the system production, two additional *distanceZ* collective variables were added to confine the probe movement in the specific area of the cylinder defined by target receptor transmembrane helices to sample. Thus, in the case of the P_2Y_1 receptor the area of the cylinder that are sampled by the probes are defined based on helices 1-3 (**Figure 2B**). In addition, to increase probe membrane penetration, the van der Waals radii of the carbon atoms of the bilayer lipid tails was decreased by 10%. This slight artefact enabled us to preserve membrane integrity during the simulation while allowing the probe to diffuse more easily into the lipid bilayer.

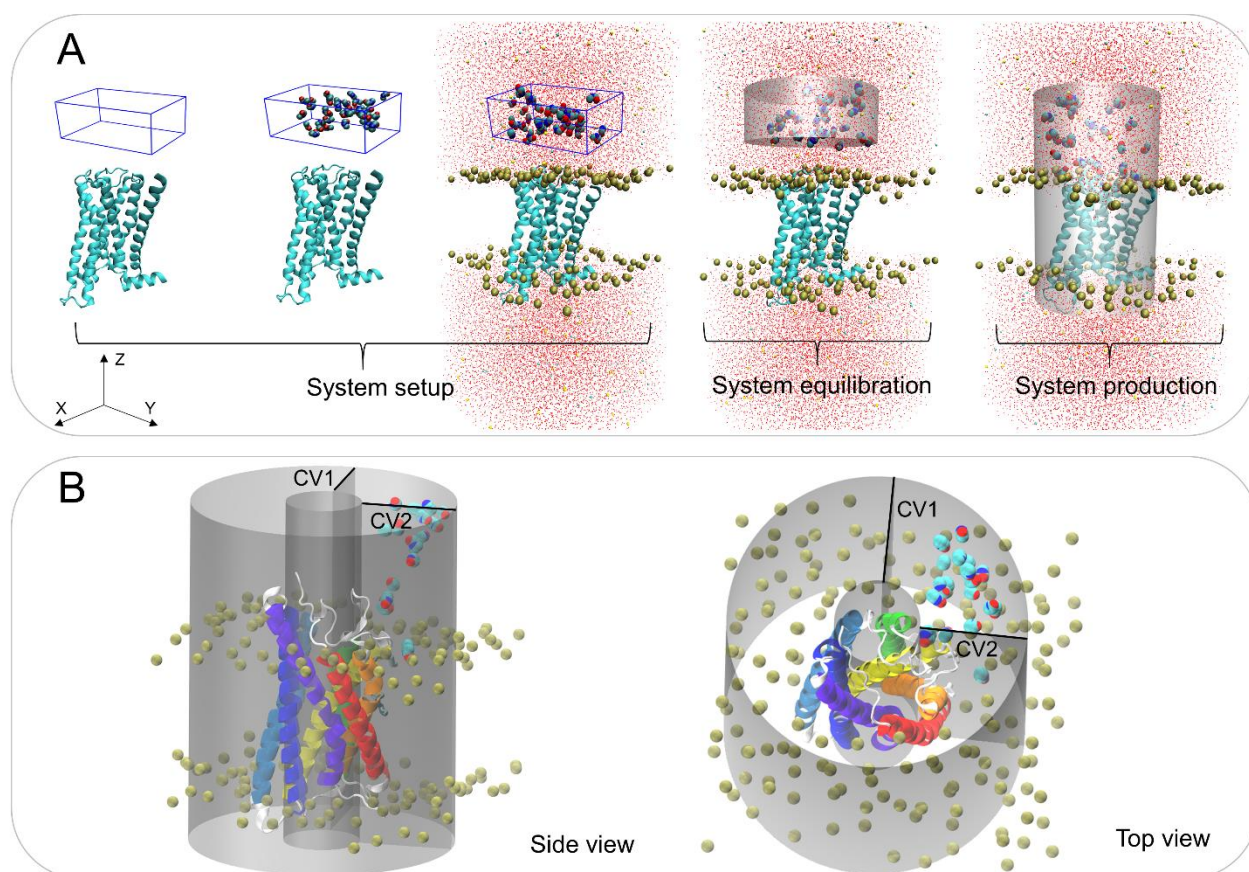


Figure 2. The GPCR probe confined dynamic mapping workflow. **A:** An example of the protocol for the M_2 receptor extracellular allosteric site. **System setup:** A box filled with a mixture of water and probe molecules at a defined concentration is specified using a Z dimension by a user, whereas the box X and Y lengths are calculated based on protein size at the EC side. The protein is then embedded into the POPC membrane and the system is solvated. **System equilibration:** A cylinder-shaped harmonic wall potential to hold probe molecules away from the protein and membrane is applied during the equilibration. **System production:** An extended cylinder-shaped harmonic wall potential is applied to allow the molecules to move towards and interact with the receptor avoiding partitioning to the membrane during the production step. **B:** The cylinder-shaped harmonic wall potential with addition of two collective variables (CV1 and CV2) to confine the movement of the

1 probes at the lipid interface of helices 1-3 (in red, orange and yellow) in the P2Y₁ receptor production
 2 simulations. The collective variables that define a cylinder were selected with lower and upper
 3 boundaries (10 and 35 Å, respectively).

5 Probe Selection from Organic Solvents and Privileged Fragments

6 Organic solvents such as isopropanol, acetamide and pyridine (**Table 1**) are often used as standard
 7 probes to sample donor- and acceptor- hydrogen bonds and hydrophobic interactions in the dynamic
 8 mapping of putative binding sites for soluble proteins (29–31). Allosteric modulators of various drug
 9 targets, however, are characterized by high aromaticity and rigidity in their structures (36,37). Thus,
 10 the probes derived from common/privileged substructures of GPCR allosteric modulators could be
 11 more suitable for mapping GPCR allosteric sites. The muscarinic PAMs including LY2119620, (**Figure**
 12 **1**) and its analogues were subjected to a maximum common substructure search that yielded N-
 13 methylthieno[2,3-b]pyridine-2-carboxamide as a “core fragment” (**Figure S1**). From this structure, two
 14 sub-structures, thieno[2,3-b]pyridine, **THP** and N-methylformamide, **NMF** were identified by ring-chain
 15 fragmentation as probes (**Table 1**). For the β_2 receptor, the fragmentation of **Cmp-15**, (**Figure 1 and**
 16 **S1**) resulted in the selection of benzamide, **BZA** and butyramide, **BTA** as probes. For the P₂Y₁
 17 receptor, the BPTU compound, (**Figure 1 and S1**) was fragmented by functional groups; and phenol,
 18 **PHX** and 2-hydroxypyridine, **P2O** were selected as probes. The choice of these fragments fits to the
 19 recent docking structure-activity analysis of the P2Y₁ allosteric antagonists (38). The three selected
 20 receptors were simulated in the presence of the standard cosolvents and the above-mentioned
 21 fragments (**Table 1**).


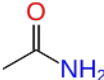

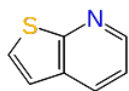
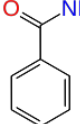
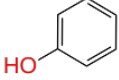
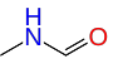
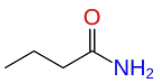
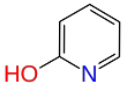
Probe Type	M ₂	β_2	P ₂ Y ₁
Standard co-solvent	 Isopropanol, ISO	 Acetamide, ACE	 Pyridine, PYR
Specific fragments	 Thieno[2,3-b]pyridine, THP	 Benzamide, BZA	 Phenol, PHX
	 N-Methylformamide, NMF	 Butyramide, BTA	 2-Hydroxypyridine, P2O

Table 1. Standard cosolvent and specific probes used for MD mapping of GPCRs allosteric sites.

MD Trajectory Analysis

We have conducted probe confined dynamic mapping of two unbound (apo) receptor structures taken from X-ray complexes of a receptor with and without an allosteric modulator for all three receptors. From the trajectory visual inspection, probe molecules reached and interacted with the amino acid residues of the EC, IC or LI allosteric binding sites in all the receptors (**Videos in SI**). This is particularly true for the specific fragments that are retained in allosteric sites for a long time. The harmonic wall potential prevented the probes from penetrating the lipid bilayer and diffusing to distant water layers. To assess the performance of the protocol, we analysed the trajectories in two different ways. Firstly, we quantified the probe presence in the allosteric binding sites (retrospective analysis). We then evaluated the possibility of allosteric site detection from our probe simulations assuming the location of the allosteric sites is unknown (predictive analysis).

Retrospective Analysis: Probe Allosteric Site Occupancy

To validate the ability of the probes to occupy the allosteric site of the M_2 , β_2 and P_2Y_1 receptors, we calculated the probe occupancy at allosteric interaction spots (**Table 2**), *i.e.* residues contributing most to the ligand-receptor interaction energy as obtained from conventional MD simulations of the X-ray structures of the receptors bound to the allosteric modulators (**Figure 1D-F, Table S1**). The probe occupancy was calculated from the 40 ns production runs and expressed as the percentage of the simulated time in which each spot was occupied by a probe averaged over (at least) three independent trajectories.

Although ISO, PYR and NMF probes occupy two interaction spots for over 70% of the simulated time, in the simulations of the M_2 apo receptor form, the occupancy tends to be higher in the agonist-bound form of the M_2 receptor (**Table 2**). The THP and NMF probes occupy three interaction spots (HB1, HB2 and HYD) for over 85% of the simulation time, whereas the ACE probe occupies the HB1 spot in the agonist-bound form of the M_2 receptor for over 70% of the time. This supports the hypothesis that the orthosteric agonist, Iperoxo stabilizes the allosteric binding site of LY2119620, which is in line with the recent conventional MD simulations of both the M_2 receptor forms (39). The probes derived from the maximum substructure search (THP and NMF) performed better by yielding higher occupancy of several interaction spots than the standard probes. In most of the THP and NMF trajectories we observed one probe molecule occupying the allosteric site and forming persistent interactions with W422^{7.35} (the Ballesteros-Weinstein numbering is given in subscript (40)); or N419^{ECL3} and Y80^{2.61} (**Figure S2, Video 1** for the THP probe). Other probes were less persistent and only occupied the site intermittently. Up to two probe molecules were detected around the residue interaction spots at a 2.5 Å distance (**Figure S2**).

Receptor	Probe, (No)	Occupancy, (%)			
		SB	HB1	HB2	HYD
M₂					
AGO ^{4MQT}	ISO (31)	32±9	67±19	64±16	71±13
APO ^{4MQT}	ISO (31)	36±5	80±8	75±12	76±14
AGO ^{4MQT}	ACE (32)	38±9	81±6	64±18	76±11
APO ^{4MQT}	ACE (32)	41±7	76±15	52±26	73±15
AGO ^{4MQT}	PYR (24)	30±9	69±11	45±19	59±19
APO ^{4MQT}	PYR (24)	38±10	73±12	54±1	79±8
AGO ^{4MQT}	NMF (33)	52±1	94±4	88±11	91±7
APO ^{4MQT}	NMF (33)	48±10	83±13	66±11	81±12
AGO ^{4MQT}	THP (14)	43±13	93±2	85±7	92±2
APO ^{4MQT}	THP (14)	26±9	59±21	54±7	47±15
AGO ^{4MQS}	ISO (31)	2±1	49±9	33±16	28±16
AGO ^{4MQS}	ACE (32)	3±1	29±10	17±7	27±17
AGO ^{4MQS}	PYR (24)	2±1	30±20	27±22	30±22
AGO ^{4MQS}	NMF (33)	3±1	45±5	29±7	39±20
AGO ^{4MQS}	THP (14)	23±3	82±3	63±16	79±5
β₂					
APO ^{5X7D}	ISO (42)	63±7	52±9	46±11	21±25
APO ^{5X7D}	ACE (43)	77±20	55±16	58±21	80±7
APO ^{5X7D}	PYR (32)	87±9	71±21	65±25	51±38
APO ^{5X7D}	BZA (21)	87±10	80±10	77±14	80±12
APO ^{5X7D}	BTA (29)	85±12	42±33	68±8	62±38
APO ^{2RH1}	ISO (42)	65±30	0	30±27	70±19
APO ^{2RH1}	ACE (43)	15±6	0	0	0
APO ^{2RH1}	PYR (32)	41±34	0	53±7	28±20
APO ^{2RH1}	BZA (21)	66±26	0	48±27	75±43
APO ^{2RH1}	BTA (29)	53±22	0	55±22	3±4
P₂Y₁					
APO ^{4XNV}	ISO (30)	4±7	18±11		35
APO ^{4XNV}	ACE (30)	0	11±4		36
APO ^{4XNV}	PYR (23)	12±19	19±12		37
APO ^{4XNV}	PHX (19)	22±19	63±11		38
APO ^{4XNV}	P2O (19)	48±12	79±7		39
APO ^{4XNW}	ISO (30)	0	54±30		40
APO ^{4XNW}	ACE (30)	0	0		41
APO ^{4XNW}	PYR (23)	21±19	12±10		42
APO ^{4XNW}	PHX (19)	40±10	73±13		43
APO ^{4XNW}	P2O (19)	19±11	68±23		44

Table 2. Probe occupancy at allosteric interaction spots. The occupancy is expressed as a percentage of the simulated time averaged over three independent trajectories. Allosteric interaction spots are identified based on ligand-residue interaction energy obtained from the MD simulations of X-ray receptor-ligand complexes (**Table S1**). The allosteric interaction spots include the following residues: **SB**: E172^{ECL2}, **HB1**: N410^{ECL3} and Y177^{ECL2}, **HB2**: Y80^{ECL2} and N419^{ECL3} and **HYD**: W422^{ECL3} and F181^{ECL2} for the M₂ receptor; **HB1**: N69^{ECL3}, D331^{ECL3}, and backbone of R63^{IL3}, **HB2**: T274^{ECL3}, **HYD1**: V54, L64^{ILC1} and F332^{ECL3} and **HYD2**: I72^{ECL3}, L275^{ECL3} and Y326^{ECL3} for the β₂ receptor; and **HB**: L102^{ECL3} backbone and: **HYD**: T103^{ECL3} and M123^{ECL3} for the P₂Y₁ receptor. The number of probes in the box/cylinder used in the simulations is indicated in parenthesis. X-ray structures used in the MD simulations with PDB code: 4MQT, 5X7D and 4XNV contain an allosteric modulator and with PDB code: 4MQS, 2RH1 and 4XNW are without an allosteric modulator.

We next simulated the X-ray structure of the M₂ receptor bound to Iperoxo in the absence of the allosteric modulator (PDB ID:4MQS) to assess whether the probes were able to recognize and occupy the allosteric site. The THP occupancy was high, over 60%, for the HB1, HB2, and HYD interaction

1 spots, whereas the occupancy of other probes was lower with ISO and NMF probes only occupying the HB1 interaction spot by above 45% of the time.

2
3 In the β_2 receptor simulations, three probes (PYR, BZA and BTA) occupied for over 85% of the time the HB1 interaction spot (**Table 2**). The ACE probe was retained in HB1 and HYD2 spots for over 77% of the time, whereas BZA occupies all four interaction spots for over 77% of the time. The occupancy of ISO probe ranged from 21-63%. The probes derived from ligand fragmentation, particularly BZA, had the highest occupancy in all the hotspots. In the trajectories, one or two BZA molecules interacted with N69^{2.40}, S329^{8.47}, D331^{8.49} and R63^{ICL1} (**Figure S3**). In particular, one BTA molecule formed stable hydrogen bonds with N69^{2.40} and T274^{6.36} (**Figure S3, Video 2**). Other probes formed less persistent interactions in the allosteric site. The simulations of the β_2 X-ray structure (PDB ID:2RH1) without an allosteric modulator showed that the BZA and ISO probes occupied the HB1 and HYD2 interaction spots for over 65% of the simulation time. The BTA probe occupied the HB1 and HYD1 spots for above 50% of the time, whereas the PYR molecules occupied the HYD1 spot for over 50% of the time. None of the ACE probe molecules occupied the receptor cavity persistently. The HB2 spot was not sampled in all the simulations.

4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
In the P₂Y₁ receptor simulations, the probe occupancy was generally lower compared to the receptors with the EC and IC allosteric sites (**Table 2**). This is due to the need for a probe to pass through the lipid layer. However, the P2O and PHX probes derived from the allosteric ligand fragmentation yielded a higher occupancy of the HYD spot (79% and 63% for P2O and PHX, respectively) and the HB (48% and 22% for P2O and PHX, respectively) compared to ISO, ACE and PYR probes. One to three molecules of P2O and PHX molecules occupied the allosteric site by forming hydrogen bonds to the backbone of M123^{3.24} and L102^{2.55} (**Figure S4 and Video 3**). The simulations of the P₂Y₁ X-ray structure in complex with an orthosteric antagonist (MRS2500) and in the absence of the allosteric modulator (PDB ID:4XNW) demonstrated that the PHX and P2O yielded better performance than other probes and occupied the HB and HYD binding spots for 40% and 73%; and 19% and 68% of the time, respectively. We also performed probe simulations by sampling the LI area around helices 2, 3 and 4 (1); 3, 4 and 5 (2); 6 and 7 (3); and 1 and 7 (4) at the P₂Y₁ receptor (PDB ID:4XNW) (**Figure S5**). The selection of these lipid-helix interfaces was based on the MDpocket cavity prediction (41) from the P₂Y₁ receptor conventional MD simulations (see the section below) and the available receptor X-ray complexes bound to an allosteric modulator at the LI. For example, allosteric ligands sit at the LI of helices 2, 3 and 4 in the PAR₂ and CB₁ receptors and at the LI of helices 3, 4 and 5 in the C₅a, β_2 and FFA₁ receptors (14–16,18,22,23). Both probes displayed either low occupancy or no occupancy at all in the selected LI areas (**Table S2**), thus demonstrating the specificity of P2O and PHX in binding at the LI of helices 1-3.

Overall, the three receptor examples demonstrate that specific probe molecules perform better in both receptor conformations. The probe occupancy is generally lower for the receptor conformation derived from the X-ray structure obtained in the absence of the allosteric modulator, although most of the

allosteric interactions could still be mapped. The performance of standard probes is particularly low for such a conformation. In addition, the probe occupancy was higher in the presence of an orthosteric agonist when the binding site of the M₂ PAM was mapped, thus indicating the importance of adding an agonist in probe simulations in search for PAM binding sites. Although more probe-MD simulation tests are required, the results obtained on the P₂Y₁ receptor show that the specific probes are capable in detecting a distinct allosteric site.

Predictive Analysis: Probe Density and Cavity Detection

We next analysed the trajectories blind to the nature of the allosteric interaction spots as it would be the case for novel ligands or previously unexplored GPCRs to assess if allosteric sites can be predicted *ab initio*. Therefore, a probe density analysis for the MD simulation trajectories was carried out using the VolMap tool (42). In **Figure 3A** the aggregated view of probe distribution (isovalue 0.5) calculated from the three replicates for each probe obtained for the two X-ray structures considered for each receptor is depicted. From this analysis, we accessed the coordinates of the points with isovalues higher than a threshold (>0.5) and determined the list of residues that interacted with the probe molecules. As shown in **Table 3**, the initial number of residues around a putative allosteric site was high.

To further narrow down the residue selection, we next linked the probe density analysis with cavity detection. The MDpocket tool (41) was used to track putative ligand binding cavities and predict their druggability in conventional MD simulations of the receptor apo forms. Based on the assumption that not all locations where a probe molecule resides in the probe-MD simulations may represent ligand binding cavities, we intersected the residues obtained from the probe density analysis with the residues obtained from cavity detection to find allosteric site residues. **Figures 3B** and **3C** show all detectable and druggable cavities in the receptors from the three simulation replicates of the two receptor X-ray structures. MDpocket identified multiple putative ligand binding sites in different regions of the three receptors. The allosteric sites of the M₂, β₂ and P₂Y₁ receptors were all detected as ligand binding cavities by MDpocket. Among them, the allosteric site at the EC side in the M₂ receptor has also been identified as a druggable cavity in all replicates of the M₂ apo receptor obtained from the X-ray complex bound to an allosteric modulator and in one replicate of the M₂ apo receptor obtained from the X-ray complex not containing an allosteric modulator, according to the Fpocket druggability criteria (43,44). A small region of the allosteric site at the IC side has been only identified as druggable in one simulation of the apo β₂ receptor obtained from the X-ray complex bound to an allosteric modulator. None of the conventional MD simulations of the P₂Y₁ receptor apo forms predicted the allosteric site at the LI side as a druggable cavity. The druggable cavity at the EC side of the M₂ receptor and detectable cavities at the IC site of the β₂ receptor and at the LI side of the

1
2
3 1 P₂Y₁ receptor overlapping with probe density (Figure 3D) were used to determine residues lining
4 these cavities (Table 3).
5 2

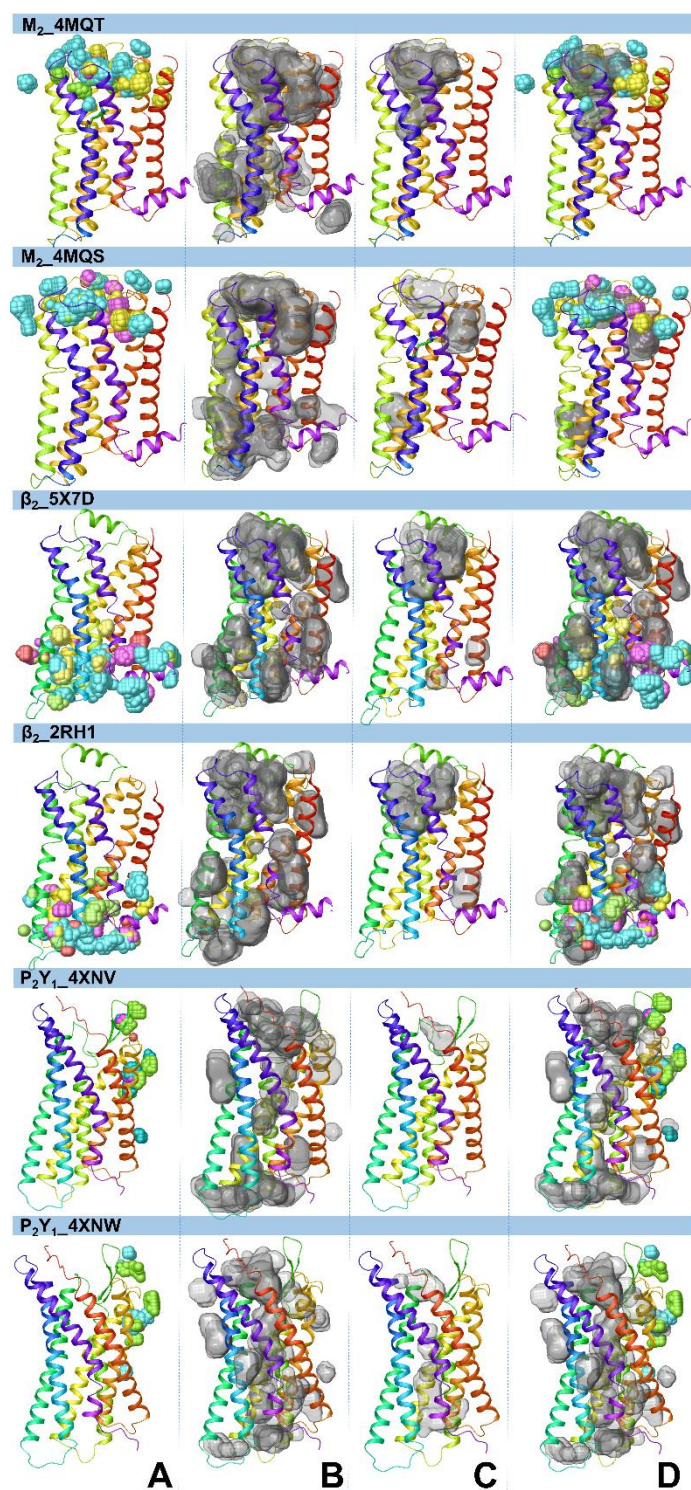


Figure 3. Probe density and cavity detection analyses for MD simulations of the receptor apo forms obtained from X-ray structures of the receptors with (PDB codes: 4MQT, 5X7D and 4XNV) and without an allosteric modulator (PDB codes: 4MQS, 2RH1 and 4XNW). **A:** The aggregated view of probe density from the probe MD simulations. The probe density was calculated using the VMD VolMap tool (isovalue = 0.5). The probe density is in yellow, red and pink for the standard probes and in cyan and green for specific probes. **B:** Binding cavities detected by MDpocket from the three conventional MD simulation runs of the apo receptors. The cavities are shown in transparent surface representation and generated with Maestro 2019-3. **C:** Druggable cavities predicted by Fpocket. **D:** The overlap of probe density and druggable cavities for the M₂ receptor and the overlap of probe density with detectable cavities for the beta₂ and P₂Y₁ receptors.

38
39 We next calculated the interaction energy between the residues selected from the overlap of the probe
40 density and cavity detection results, and the probe molecules to identify the residues forming
41 interactions with the probe below 1 kcal/mol (Table 4). Less than ten residues were identified (Table
42 4) and among them there were residues of the allosteric interaction spots, which were scored at the

top of the list for both receptor conformations of the M₂ and P₂Y₁ receptors (**Table 4**). In the case of the β₂ adrenergic receptor, the allosteric interaction spot residues were more scattered, particularly from the simulations obtained based on the X-ray structure of the receptor without the allosteric ligand. This is because of the partially closed IC cavity in this structure. The specific probes had a major contribution to the interaction energy.

Overall, the binding interactions of probe molecules are matched with the interactions of the allosteric ligands in the X-ray structures. Not all allosteric cavities meet an established druggability rule; this is especially true with cavities at the LI, therefore, all detectable cavities should be potentially considered in a predictive analysis. Our computational protocol was then validated in a prospective study, where the key receptor residues predicted by probe MD mapping were assessed by mutagenesis.

Receptor	Receptor structure used for MD simulations	Number of residues				
		From probe density	From cavity detection	From the intersection between probe density and cavity detection	From probe-residue interaction energy <-1 kcal/mol	Identified allosteric interaction spot residues
M ₂	4MQT	55	35	35	23	8
	4MQS	52	46	40	24	8
β ₂	5X5D	83	45	45	23	9
	2RH1	72	34	33	26	9
P ₂ Y ₁	4XNV	27	18	18	11	3
	4XNW	24	10	10	10	3

Table 3. Prediction of allosteric interaction spot residues from probe mapping, cavity detection and probe-residue interaction energy.

Probe-Residue Interaction Energy (kcal/mol)

M ₂ 4MQT			M ₂ 4MQS			β ₂ 5X7D			β ₂ 2RH1			P ₂ Y ₁ 4XNV			P ₂ Y ₁ 4XNW		
E175	-6.2	NMF	D97	-8.4	NMF	R63	-7.3	BZA	R328	-6.0	ISO	F119	-4.5	P2O	F119	-4.8	PHX
N419	-5.6	THP	E175	-6.5	NMF	D331	-6.6	BZA	R63	-5.7	BZA	M123	-4.3	P2O	L102	-3.9	PHX
W422	-5.1	THP	N410	-4.2	THP	R131	-6.2	BTA	S329	-5.1	ISO	L102	-2.9	PHX	Q127	-3.4	ISO
N410	-4.0	NMF	E172	-4.0	NMF	A271	-3.6	BZA	R131	-5.1	BTA	T103	-2.6	P2O	T103	-3.2	P2O
Y177	-3.9	NMF	Y177	-3.6	ISO	S329	-3.2	BZA	D331	-4.5	BZA	A122	-2.3	PHX	M123	-2.8	PHX
E172	-3.6	NMF	W422	-3.6	THP	N69	-3.2	BZA	K140	-4.3	ACE	L126	-2.1	P2O	W117	-1.9	ISO
F181	-2.6	THP	A414	-3.4	THP	Q142	-3.0	BZA	K243	-3.8	BZA	G120	-1.8	ACE	A106	-1.6	P2O
Y80	-2.4	NMF	N419	-3.3	THP	T274	-2.9	BZA	E244	-3.1	BTA	Q127	-1.5	P2O	L126	-1.1	PHX
Y426	-2.2	NMF	Y80	-2.7	PYR	F264	-2.3	BTA	N69	-2.8	ACE	A106	-1.5	P2O	P105	-1.0	PHX
T170	-2.2	THP	Y83	-2.5	PYR	L275	-2.2	BZA	T274	-2.8	ACE	W117	-1.4	ACE			
S182	-2.1	THP	W162	-2.5	ISO	K267	-2.2	BTA	Q142	-2.5	BZA	P105	-1.0	PHX			
A414	-2.1	THP	F181	-2.2	THP	L145	-1.7	BZA	F332	-2.5	PYR						
T84	-2.0	THP	P415	-2.2	THP	L272	-1.5	BZA	P306	-2.5	BZA						
Y83	-2.0	THP	T423	-2.2	PYR	T68	-1.5	BZA	Y326	-2.4	ISO						
T423	-1.8	THP	T170	-2.1	ACE	F332	-1.4	BZA	Y219	-2.3	ISO						
Y88	-1.6	THP	I417	-2.0	THP	T66	-1.3	BZA	Y141	-2.2	ACE						
Y403	-1.5	NMF	T84	-1.6	PYR	Y326	-1.3	BZA	A247	-2.1	PYR						
T187	-1.5	NMF	T411	-1.6	THP	Y219	-1.2	ISO	T68	-1.4	ACE						
I417	-1.2	ACE	Q179	-1.5	ISO	C327	-1.1	PYR	L275	-1.4	PYR						
N183	-1.1	THP	Y88	-1.5	THP	I72	-1.1	BZA	L248	-1.3	PYR						
G87	-1.1	NMF	S182	-1.3	THP	L143	-1.0	ACE	S143	-1.2	ACE						
T411	-1.0	ACE	V171	-1.2	NMF	T146	-1.0	BTA	I72	-1.2	PYR						
Y104	-1.0	NMF	G174	-1.2	PYR	L64	-1.0	BZA	V67	-1.1	BZA						
			P418	-1.1	THP				T66	-1.1	BZA						
									L64	-1.1	BZA						

Table 4. Probe-residue interaction energy. The residues selected from the overlapping region obtained from the probe density and cavity detection analyses were used to calculate the probe-residues interaction energy. The interaction energies <-1 kcal/mol are shown for the probe that forms the strongest interaction with a residue. Residues that contribute to the allosteric interaction spots (from Table 2) are indicated in bold.

Experimental Validation of Probe Confined Dynamic Mapping

To test our methodology in an experimental setting, we applied our pipeline to the D₂ receptor. As a proof of concept, we explored how the protocol performed in predicting where the previously published UCB PAM might bind (**Figure 4A**) (45). Two probe molecules indole (**IND**) and benzyl alcohol (**BAL**) were generated from the UCB compound (45). Probe confined dynamic mapping on the D₂ crystal structure (46) in a dopamine-bound form was carried out to explore putative binding sites at the EC and IC sides and the LI. From probe density analysis, druggable cavities detection and probe-residue interaction energies, as described above, we have predicted a putative allosteric site for the UCB compound at the EC side of the D₂ receptor consisting of helices 2 and 7, ECL1 and ECL2 (**Figure 4B-C**). From our approach the final list of residues, involving V91^{2.61}, L94^{2.64}, E95^{2.65}, W100^{ECL1},

1
2
3 1 I184^{ECL2} and W413^{7.40} was selected for site-directed mutagenesis by taking into consideration a
4 2 docking pose of the UCB compound in the predicted allosteric site (**Figure 4D**).

5 3 To test if these residues were involved in the function of the UCB compound, we first confirmed the
6 4 compound modulated cAMP production in line with an allosteric modulator of a G_i-coupled receptor
7 5 (**Figure 4E.1**). We then tested each mutant in transfected HEK293 cells to ensure that the mutation
8 6 of each residue did not significantly impact the ability of the receptor to reduce Forskolin-stimulated
9 7 production of cAMP after agonist addition (**Figure 4E, Table S4**) or the expression (**Figure S7**). We
10 8 found either a tryptophan or alanine at position L94^{2.64} was well tolerated (pEC₅₀ = 9.7 and 8.7,
11 9 respectively). At positions 2.61 and 2.65, we were able to mutate these residues to an alanine whilst
12 10 conserving activity with minimal changes in pEC₅₀ (**Figure 4 and Table S3**). Altering W100^{ECL1},
13 11 however, led to a significant decrease in activity, in line with published results (46) (pEC₅₀ = 8.1)
14 12 (**Figure S6**) and showed altered expression (Figure S7). A similar change in activity was seen with
15 13 the I184A^{ECL2} mutation (pEC₅₀ = 8.3) (**Figure S6**). Thus, these two constructs were not explored
16 14 further. For the W413^{7.40} position, we tested mutation to alanine which preserved receptor function
17 15 (pEC₅₀ = 8.8).
18 16
19 17
20 18
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

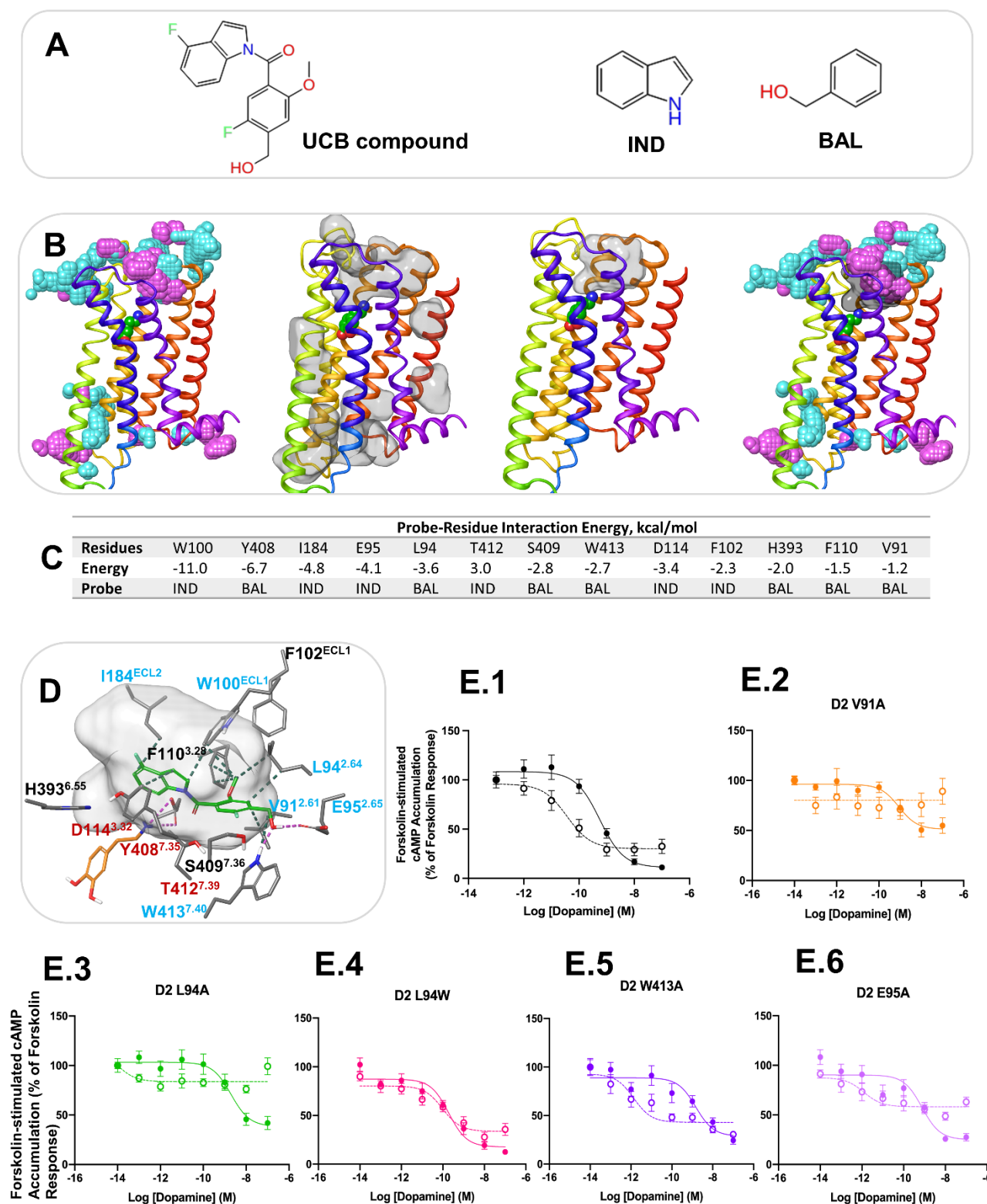


Figure 4. Computational and experimental prediction of the allosteric site for the UCB compound in the D₂ receptor. **A:** The UCB compound and probe molecules used in allosteric site mapping; **B:** The structure of the D₂ receptor in cartoon representation with the results of MD simulation analysis. From left to right: the aggregated view of probe density from the probe simulations, IND and BAL are in cyan and purple, respectively; detectable cavities predicted by MDpocket from the MD trajectories are shown in transparent surface representation; a druggable cavity predicted by Fpocket; the overlap of probe density and a druggable cavity. **C:** The probe-

1
2
3 1 residue interaction energy for the residues selected from probe density and druggable cavity
4 2 detection. The interaction energy < -1 kcal/mol is shown for the probe that forms the strongest
5 3 interaction with a residue. **D**: Docking pose of the UCB compound at the putative allosteric site. The
6 4 UCB compound and dopamine are shown in green and orange sticks, respectively. Only residues
7 5 selected from probe-residue interaction energy calculation are shown. The druggable cavity is shown
8 6 in transparent surface representation. Labels of residues selected for mutagenesis are in blue and
9 7 labels of residues in contact with the orthosteric agonist are in red. **E**: cAMP accumulation assays in
10 8 the absence and presence of the UCB compound at the D₂ receptor WT and mutants. Concentration-
11 9 response curves of the endogenous agonist, dopamine, measuring Forskolin-induced (7.5 μ M) cAMP
12 10 accumulation at the D₂ WT receptor (**1**) and the mutants V91A (**2**), L94A (**3**), L94W (**4**), W413A (**5**)
13 11 and E95A (**6**) in the absence and presence of UCB compound **1** (10 μ M). The absence of the PAM is
14 12 indicated with a solid line whilst the presence of 10 μ M of the PAM is indicated by a dotted line. Each
15 13 data point represents the mean \pm SEM of triplicate wells of three independent experiments. Analysis
16 14 of the pharmacological parameters for these curves can be found in **Supplementary Table S3**.
17 15

18 15
19 16 Next, we tested the effect of these mutations on their ability to influence the effects of the UCB
20 17 compound. In line with the predicted computational analysis, mutations of V91^{2.61}A and L94^{2.64}A did
21 18 appear to reduce the effects of the PAM on dopamine stimulation as measured by forskolin-stimulated
22 19 cAMP accumulation (**Figure 4E.2** and **4E.3**). However, the dose responses were nearly flat in the
23 20 presence of the UCB compound, with a measured ΔE_{\max} \sim 57% and \sim 34%, respectively, which is not
24 21 easy to interpret. Enlarging this pocket by either mutation may provide too much flexibility for the
25 22 compound. This varied to the L94^{2.64}W mutant, where the PAM effect appeared to be completely lost
26 23 (**Figure 4E.4**). The W413^{7.40}A mutation increased the PAM effect by improving dopamine affinity
27 24 ($\Delta pEC_{50} = 3$) with no impact on E_{\max} (**Figure 4E.5**). The E95A^{2.65} mutation too showed a similar but
28 25 milder effect ($\Delta pEC_{50} = 3$) with a measured ΔE_{\max} \sim 35%, suggesting the potential importance of these
29 26 amino acids in maintaining the orientation of the PAM (**Figure 4E.6**). To quantify effects of the UCB
30 27 compound we performed cross titration curves and calculated a K_b value (**Figure S8, A-C**). These
31 28 showed dramatic changes from WT vs L94^{2.64}W and W413^{7.40}A mutants (**Figure S8, D**). Indeed, for
32 29 the L94^{2.64}W no K_b could be obtained as all signs of allosterism were lost, consistent with the data
33 30 obtained above. Together, these data support the predicted binding site of the UCB compound and
34 31 serve to validate our prediction pipeline.
35 32
36 33
37 34
38 35
39 36
40 37
41 38
42 39
43 40
44 41
45 42
46 43
47 44
48 45
49 46
50 47
51 48
52 49
53 50
54 51
55 52
56 53
57 54
58 55
59 56
60 57

50 35 Discussion

51 36 We have developed a probe confined dynamic mapping protocol for fast and efficient detection of
52 37 allosteric sites in GPCRs. The application of the cylinder-shaped harmonic wall potential and the
53 38 specific probes derived from GPCR allosteric ligand structures helped address the limitations of
54 39 currently available cosolvent mapping protocols such as limited probe sampling, membrane distortion,
55 40 probe non-specific binding, and protein denaturation.
56 41
57 42
58 43
59 44
60 45

1
2
3 1 The cylinder-shaped harmonic wall potential allows the probes to explore thoroughly either the EC or
4 2 IC sides of the receptor whilst preventing them from partitioning into the lipid bilayer or distant water
5 3 layers. In the case of the allosteric site mapping at the LI, the probes are only allowed to move from
6 4 a water layer to the lipid bilayer at the interface of the selected helices. To explore the entire protein-
7 5 lipid interface in a blind allosteric site search, we propose to run probe simulations sampling
8 6 separately two to three helices.

9 7 The default simulation length for the production run was set at 40 ns, although the protocol was able
10 8 to identify an allosteric site in simulation times as short as 20 ns. Based on the inspection of recent
11 9 X-ray and cryo-EM structures of GPCR LI allosteric sites located deeper in the membrane, we
12 10 envision that the simulation time for the search of such sites could be extended to enhance the
13 11 sampling at the interface. In blind studies we suggest mapping allosteric sites at the EC, IC and at the
14 12 LI separately. Incorporation of all the scenarios in a single protocol will require too many restraints
15 13 that would result in a reduced simulation speed.

16 14 Here, we extended the dynamic mapping approach with cosolvents to fragments derived from GPCR
17 15 allosteric ligand structures to improve sampling of specific binding in allosteric sites. From the probe
18 16 occupancy and probe-residue interaction energy analysis, the specific probes performed substantially
19 17 better in mapping key allosteric interaction spots compared to the standard organic solvents. In the
20 18 case of the allosteric sites at the LI, the standard probes yielded poor results. In the probe-MD
21 19 simulations of the P₂Y₁ receptor, where we allowed probes to sample the LI for various helices, we
22 20 show that specific probes were also selective in mapping a specific cavity at the LI. In prospective
23 21 search of allosteric sites for a receptor with unknown allosteric ligands, we suggest using a set of
24 22 fragments derived from 'privileged' or common substructures of allosteric modulators of a related
25 23 receptor subtype or receptor subfamilies as probes. Our choice of 10% probe concentration in the
26 24 starting box of the water-probe mixture was enough to sample the allosteric sites. In the case of PAMs,
27 25 the probe mapping simulations should be performed in the presence of an orthosteric agonist.

28 26 We also investigated the performance of receptor X-ray structures without an allosteric ligand.
29 27 Although the probe occupancy was generally low, the specific probes were able to sample the
30 28 allosteric interaction spots in the M₂ and P₂Y₁ receptors. The probes were not able to sample one of
31 29 the interaction spots in the β₂ receptor because the binding cavity was partially occluded. However,
32 30 in the proposed predictive setting of MD trajectory analysis, involving a combination of probe density
33 31 analysis, binding cavity detection and probe-residue interaction energy calculation, these receptor
34 32 conformations allowed identification of the key residues.

35 33 We have applied the developed protocol to identify the binding site of the UCB PAM at the D₂ receptor.
36 34 Currently, the experimental structure of the D₂ receptor bound to an allosteric ligand is not available.
37 35 Recently, the cryo-EM structure of the D₁ receptor has been published bound to LY3154207, a PAM
38 36 at the LI of helices 3-5 (47). The known D₂ NAM, SB269652 (48) and compounds based on thieno[2,3-
39 37 d]pyrimidine scaffold (49,50) are believed to bind to the EC side (49,51). In our study we explored the

1
2
3 1 possible allosteric site of the UCB compound at the EC and IC sides and the LI. We found that the
4 2 probe density was particularly high at the EC side around helices 2 and 7, ECL1 and ECL2. Given
5 3 that this binding cavity is also predicted as druggable, we chose this cavity as a preferable one for
6 4 residue selection. Six residues were selected for site-directed mutagenesis and four residues were
7 5 confirmed to have an impact on the UCB compound of dopamine-induced activation of the receptor,
8 6 thus, validating our computational protocol. Mutating the residues in the EC loops beneath the
9 7 dopamine binding site in direct contact with PAM's F-indole functional group affects receptor function,
10 8 as expected. In particular, I184 is believed to play a role in ligands K_{on} and K_{off} as well as β -arrestin
11 9 signaling in the D_2 receptor and other aminergic GPCRs (52,53). In the X-ray D_2 receptor bound to
12 10 risperidone, an inverse agonist forms a hydrophobic patch involving W100, I194 and L94^{2,64} that
13 11 changes the ECL1/2 orientation with respect to the one previously observed in D_2 -like receptors (46).
14 12 Mutation of these residues reduces the residence time of risperidone. Hence, the effects observed in
15 13 our experiments are in line with acquired knowledge on the D_2 receptor. Mutating the residues
16 14 surrounding the PAM Ph-CH₂OH does not affect receptor function but impacts on the PAM activity.
17 15 This evidence supports the role suggested for the W413^{7,40} and E95^{2,65} side-chains in keeping the
18 16 PAM in an orientation that allows it to occlude the orthosteric pocket and/or interfere with the agonist
19 17 binding.

20 18 FTMap and FTSite tools were used to map allosteric sites in GPCR X-ray structures (54) and MD-
21 19 derived receptor conformations of the M_2 and A_2 adenosine receptors (55,56). These tools were
22 20 successful in mapping EC and IC allosteric sites but failed to map LI sites. The site-identification by
23 21 ligand competitive saturation (SILCS) MD-based approach has been also recently applied to map EC
24 22 allosteric sites of the M_2 and GPR40 receptors (28). Here, we have developed a methodology for
25 23 mapping allosteric sites at the receptor EC and IC sides and, particularly, in the most challenging
26 24 case involving allosteric sites at the interface between and the receptor and membrane. The cylinder-
27 25 shaped harmonic wall potential applied to probe molecules in MD simulations allow efficient mapping
28 26 of allosteric sites at various locations. Our methodology is able to identify allosteric sites in a short
29 27 simulation time and the results of the simulations could be inspected in less than a day, which makes
30 28 it suitable for industry settings. Implementation of a fragment-based drug design approach
31 29 demonstrated that confined dynamic probe mapping can successfully be performed on membrane
32 30 proteins. Our methodology is a computationally feasible solution to initiate rational search of allosteric
33 31 sites and design of allosteric ligands for GPCRs and other membrane-bound drug targets.

32 33 **Materials and Methods**

34 35 ***System Preparation***

36 37 The X-ray structures of the M_2 (4MQT and 4MQS), β_2 (5X7D and 2RH1), P_2Y_1 (4XNV and 4XNW)
38 38 and D_2 (6CM4) receptors were used to revert to the wild type receptors and subsequent MD

1
2
3 1 simulations. The wild types were built based on the sequences with UniProt ID: P08172, P07550,
4 2 P47900 and P14416 with a partial reconstruction of ICL3 using the Prime module of Schrödinger
5 3 software (57).
6
7
8 4

9 5 **Probe Confined Dynamic Mapping Protocols**

10 6

11 7 The automatic procedure of probe confined dynamic mapping are provided as jupyter notebooks
12 8 (MIDAS_EC.ipynb, MIDAS_IC.ipynb and MIDAS_LI.ipynb) available at GitHub
13 9 ([https://github.com/irin12/Probe-Confined-Dynamic-Mapping-Protocols-
14 10 GPCRs_membrane_proteins](https://github.com/irin12/Probe-Confined-Dynamic-Mapping-Protocols-GPCRs_membrane_proteins)). The required input files are: (i) pdb of a receptor (pre-oriented with
15 11 OPM(58)), membrane, and one copy of co-solvent/fragment structures and (ii) an input file containing
16 12 the following information: a receptor, an orthosteric ligand (if applicable), cosolvent and membrane
17 13 file names (1); an orthosteric ligand (if applicable), cosolvent, and lipid residue names (2);
18 14 cosolvent/fragment molecular weight and desired water/probe m/m % (in the closed box) (3); and
19 15 height of the water buffer region between the protein and the water/probe mixture box (4).
20 16

21 16 **System Setup.** A box of water/probe (co-solvents or fragments) mixture of user-defined m/m %
22 17 concentration is generated with Packmol (59) and placed at a user-defined distance (*zoff*) from the
23 18 protein atom with maximum/minimum z coordinate on the EC/IC side. The user can specify the box z
24 19 dimension (*zheight*), whereas the box x and y sizes are automatically calculated based on
25 20 minimum/maximum protein dimensions on the EC/IC sides (based on protein atoms with $z > 0$ and z
26 21 < 0 , respectively). After the box is placed, the protein is embedded in the membrane and the system
27 22 is solvated with water molecules and neutralized with 0.15 NaCl. Final system files are created for
28 23 subsequent MD simulations. All the above-mentioned steps are carried out using HTMD 1.13.6 (60).
29 24 NAMD program (61) input, constraints and collective variable files are automatically generated.
30 25

31 25 **Equilibration.** The cosolvent molecules are confined in a closed cylinder, which is set up by
32 26 *distanceXY* and *distanceZ* collective variables available in the NAMD program and using the
33 27 hypotenuse of the rectangle defined by the Packmol box x and y sizes as radius. A small constraint
34 28 on the protein centre of mass is applied during the third equilibration phase to avoid protein drifting
35 29 and maintain cylinder absolute coordinate reference.
36 30

37 30 **Production.** During the production phase, the cylinder boundary (*distanceZ* collective variable) facing
38 31 the protein is removed and the cylinder radius decreased to the half of the Packmol x or y box size,
39 32 whichever is smaller, to allow the probes to diffuse towards and interact with the protein. During the
40 33 production, the *distanceZ* collective variable facing the system boundary is maintained, thus, defining
41 34 a semi-closed cylinder, and its boundary lowered by ~ 10 Å. To enable probes adaptation to the new
42 35 cylinder size, the system is subjected to 240 steps of minimization, before 40 ns of NPAT (constant
43 36 normal pressure and lateral surface area of membranes and constant temperature) production is ran
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 1 (at least in triplicate). Protein C α atoms with z coordinates $\pm 5\text{\AA}$ from the cell origin are restrained to
4 2 avoid protein drifting.

5 3 To sample allosteric sites at the membrane/protein interface two additional *distanceZ* collective
6 4 variables (one on the x and one on the y axis) were added to confine the probes in the specific area
7 5 of the cylinder defined by the user based on target receptor transmembrane helices to sample. To
8 6 enable probes adaptation to the new area size, at least 1800 minimization steps were required before
9 7 running the actual production (40 ns, NPAT condition, at least 5 replicas). The collective variables
10 8 that define a cylinder were selected with lower and upper boundaries (10 and 35 \AA). The 10 \AA was
11 9 set to prevent the probes from sampling MRS2500 orthosteric site, which is readily accessible from
12 10 the EC side. The van der Waal radii of POPC CLT2 and CTL3 atom types was decreased by 10%,
13 11 while the parameters of all the other POPC atom types remained unaltered.

24 15 **Molecular Dynamics Simulations**

25 16
26 17 CHARMM36 force field was used for proteins, lipids and water(62,63). The parameters for ligands
27 18 and probes were derived from the CHARMM General Force Field (CgenFF), v 1.0.0 (64). The
28 19 receptors were placed in a 90x90 \AA^2 -palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC)
29 20 membrane patch. The receptor-lipid system was solvated with a 30 \AA buffer from both sides of the
30 21 lipid bilayer. The systems were neutralized by 0.15M Na⁺ and Cl⁻ ions.

31 22 All MD simulations were performed using NAMD Git-2017-12-19, Linux-x86_64-multicore-CUDA (65).
32 23 The first equilibration step included 1000 step of minimization followed by 0.5 ns of NVT simulations
33 24 with the protein, lipid headgroups, ions, cosolvent and water atoms fixed. The second equilibration
34 25 step comprised of 500 minimization steps followed by 2.0 ns of NPT simulations, where harmonic
35 26 restraints on all protein atoms were applied; and a small force was applied to water molecules to
36 27 prevent them from entering the membrane. In the case of probe simulations, the probes were confined
37 28 in a closed cylinder. The third equilibration step involved 10 ns of NPT ensemble with the receptor
38 29 free to relax with translation on the centre of mass removed. Probe molecules were kept in a closed
39 30 cylinder. The production step included 240 steps of minimization and 40 ns of simulations, where the
40 31 whole system was free to relax. In the case of probe simulations, harmonic restraints on the protein
41 32 C α atoms with z $\pm 5\text{\AA}$ from the origin (0,0,0) were applied and translation on the protein centre of
42 33 mass was removed. The probe molecules were confined in a smaller semi-closed cylinder, with the
43 34 boundary facing the receptor opened. The temperature of all simulations was 310 K.

56 36 **Cheminformatics Analysis**

57 37
58
59
60

1
2
3 1 Maximum common substructure search and ligand fragmentation based on a ring-chain method or
4 functional groups were performed using the cheminformatics toolkit (frags2img.py, getcore.py and
5 2 enumfras2pdf.py of OpenEye (OEChem TK 2.2.0) (66).
6 3
7 4
8 5

10 6 **Trajectory Analysis**

11 7

12 8 The probe occupancy at the distance of 4 Å from the allosteric spots was calculated using an in-house
13 9 tcl-script using VMD 1.9.3 (42). The probe density was calculated using the Volmap tool of VMD 1.9.3
14 10 with a cell side of 1 Å and the density was averaged over all frames of the top molecule. The Volmap
15 11 probe density was analysed at isovalues of 0.3, 0.5 and 0.8. Probe density at isovalue of 0.5 was
16 12 selected for the selection of residues in the interaction with probe molecules at a 5 Å distance.
17 13 MDpocket (41) with Fpocket 3.0 (43,44) was used to predict detectable and druggable pockets in MD
18 14 trajectories. The residue–probe or residue–ligand interaction energy was calculated using the
19 15 ‘namdenergy.tcl’ script v 1.6 of NAMD (61). The residues at 5 Å distance from a ligand/probe were
20 16 selected for the interaction energy analysis. Modelling pictures were created with Maestro 2019-3
21 17 (57) and MD videos were generated with VMD 1.9.3.
22 18
23 19

30 19 **Molecular Docking**

31 20

32 21 The induced fit docking program of Schrodinger software 2019-3 (67,68) was used for docking of the
33 22 UCB compound and dopamine. Prior to docking, ligands were prepared using the ‘Ligand Prep’
34 23 module and the D₂ receptor was pre-processed according to the protein preparation procedure of the
35 24 Schrodinger software. All docking calculations were run in the ‘Standard Precision’ (SP) mode (69)
36 25 with default values for all parameters. The docking box was set based on the residues predicted for
37 26 the putative allosteric site from the probe-MD simulations. The best-docked structure was chosen
38 27 using the Glide Score (70). Dopamine was maintained in the orthosteric site during the docking of the
39 28 UCB compound.
40 29
41 30

47 30 **Site-Directed Mutagenesis**

48 31

49 32 All cDNA templates and primer sequences are listed in **Table S4**. Per mutation, a reaction mix (25
50 33 μL) was prepared containing 1X Phusion™ HF Buffer, 200 μM of dNTP mix, 0.5 μM of forward primer,
51 34 0.5 μM of reverse primer, 25 ng of template DNA and 0.02 U/μL of Phusion™ High-Fidelity DNA
52 35 Polymerase. PCR cycles were carried out using Veriti™ thermocycler (Applied Biosystems). A three-
53 36 step protocol was set such that initial denaturation occurred at 98 °C for 30 seconds, followed by 35
54 37 amplification cycles. Each amplification cycle consisted of denaturation for 10 minutes, an annealing
55 38 gradient of 2-3 temperatures depending on T_m of primers for 60 seconds followed by an extension at
56 39 72 °C (22.5 s/kb). To terminate the PCR amplification cycles, a final extension step at 72 °C for 10

1
2
3 1 minutes was applied. For the mutation W413A, a two-step protocol was used. Agarose gel
4 2 electrophoresis (1% *w:v*) was then used to analyse PCR products. 5U of DpnI restriction enzyme (2%
5 3 of total volume in PCR tube) was added and incubated for 37 °C overnight. PCR products were
6 4 transformed using *E. Coli* DH5 α competent cells, amplified and purified. DNA Sanger sequencing
7 5 (Eurofins) was then used to confirm if PCR had been successful.
8 6
9 7

10 8 **Transient Transfection**

11 9 Reverse transfection was used to transiently transfect HEK293 cells using Lipofectamine™ 3000
12 10 (Thermo Fisher), using the method provided by the manufacturer. Transfections were performed such
13 11 that each well contained 150 ng of the D2 WT receptor and 50 ng of the pGlo-Sensor™-22F cAMP
14 12 protein sensor with Lipofectamine added in a 1:1.5 *w:v* ratio respectively. 50 μ L of this mix was added
15 13 to a poly-D-lysine (Sigma-Aldrich) coated F white clear-bottom plate (Greiner Bio-One). To this, 100
16 14 μ L of HEK293 cells at a viable cell density of 75,000 cells were then added. Plates were then
17 15 incubated in a 5% CO₂ atmosphere at 37 °C for 24 hours prior to performing intracellular cAMP
18 16 Accumulation Assays.
19 17

20 18 **Intracellular cAMP Accumulation Assays**

21 19
22 20 24 hours post-transfection, the cell culture media was removed slowly, minimising disruption to
23 21 attached cells adhered to bottom of the wells. Cells were initially washed using HBSS-based cAMP
24 22 assay buffer (pH 7.4). Thereafter, cAMP buffer supplemented with Firefly D-Luciferin (0.45 mg/mL;
25 23 NanoLight Technologies) was added (90 μ L for functionality assays or 80 μ L for testing the PAM
26 24 assays). The plate was then left to pre-equilibrate in the dark at 28 °C for 1 hour. During this time, the
27 25 CLARIOstar PLUS (BMG Labtech) was set to 28 °C. Remaining cAMP buffer was used to prepare
28 26 ascorbic acid (0.01% *w:v*) and then further supplemented with Forskolin (7.5 μ M). This was used to
29 27 prepare dopamine dilutions. The allosteric modulator, **(1)** was custom synthesized by Enamine and
30 28 resuspended in DMSO to a stock concentration of 10 mM and then aliquoted. Prior to assays, dilutions
31 29 were prepared in filtered cAMP buffer absent of ascorbic acid or forskolin. Post-equilibration, 10 μ L
32 30 of the allosteric modulator was added to the wells and left to equilibrate for a further 15 minutes.
33 31 Bioluminescence readings were then conducted to measure basal luminescence signal (~6-10 cycles)
34 32 prior to agonist addition. Upon agonist addition (10 μ L), luminescence readings were taken for ~ 1
35 33 hour. For functionality assays, post-equilibration, basal luminescence reads were performed, followed
36 34 by agonist addition (10 μ L) only. Variability in luciferase signal was taken into consideration by using
37 35 the average of the last three stable basal luminescence reads to normalise the response of each well.
38 36 GraphPad Prism 9.0 was used to plot data.
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

1
2
3 1 Cross-titration curves of the UCB compound (30 μ M, 10 μ M, 3 μ M, 1 μ M, 0.3 μ M and 0 μ M) were
4 2 performed to calculate the K_b value at the WT, L94W and W413A. Luminescence readings were
5 3 performed for ~ 30 minutes. A nonlinear regression analysis fit GraphPad Prism 9.0 was used to
6 4 calculate the K_b value by using a nonlinear regression analysis fit.
7
8
9
10
11
12

13 7 **Immunofluorescence Assays**

14 8

15
16 9 HEK293 cells were grown on coverslips and transiently transfected using polyethylenimine (PEI).
17 10 Prior to fixation they were rinsed 3X with PBS and fixed at room temperature for 10 min with 4% v/v
18 11 paraformaldehyde. The cells were then permeabilised using Triton X-100 (0.2%) in PBS buffer 1min
19 12 followed by 1X wash with PBS. then a quenching of Aldehyde step was performed to reduce
20 13 background with NH_4Cl (50 mM, 15 min). The coverslips were then washed 1X for 5 min in PBS and
21 14 blocked for 1 hr at room temperature in 5% w/v BSA. The coverslips were subsequently washed in
22 15 PBS 1X and primary (Rabbit D2, Dopamine Receptor 2 Antibody) diluted in 5% w/v BSA added at
23 16 1:200 overnight at 4°C. The next morning the coverslips were washed again in PBS 3X and the
24 17 secondary (Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 568) added
25 18 at 1:1000 in 5% w/v BSA for 1 hour at room temperature. The cells were washed again 3X in PBS
26 19 and then mounted using ProLong Gold antifade reagent with DAPI and imaged using a Zeiss LSM
27 20 880 Laser Scanning Confocal Microscope.
28
29
30
31
32
33
34
35
36
37

38 23 **Acknowledgments**

39 24

40
41 25 This work was supported by the European Union's Horizon 2020 research and innovation programme
42 26 under the Marie-Skłodowska-Curie grant agreement No. 748830 (awarded to A.C.) and the
43 27 Biotechnology and Biosciences Research Council (BBSRC) responsive mode award BB/R007101/1
44 28 to I.G.T.; and BB/R006946/1, Medical Research Council (MRC) response mode award
45 29 MR/S008608/1, and Barts Charity award MRC0227 to P.J.M. This project made use of computational
46 30 time on Kelvin-2 (grant no. EP/T022175/1) and ARCHER2 granted via the UK High-End Computing
47 31 Consortium for Biomolecular Simulation, HECBioSim (hecbiosim.ac.uk), supported by EPSRC (grant
48 32 no. EP/R029407/1). I.G.T and P.J.M. participate in the European COST Action CA18133 (ERNEST).
49 33 We thank Prof Graeme Milligan for several helpful discussions and suggestions.
50
51
52
53
54
55
56
57
58
59
60

1 Author contribution

2 AC: Data curation, Software, Formal analysis, Validation, Visualization, Methodology,
3 Writing - review and editing.

4 AKG: Data curation, Formal analysis and Visualization.

5 TD: Data curation, Formal analysis and Visualization.

6 DSK: Validation and Formal analysis.

7 GC: Data curation and Formal analysis.

8 PJM: Resources, Supervision, Funding acquisition, Writing - review and editing.

9 IGT: Conceptualization, Resources, Data curation, Software, Supervision, Funding
10 acquisition, Validation, Methodology, Writing - original draft, Project administration, Writing -
11 review and editing.

17 References

- 19 1. Hauser AS, Attwood MM, Rask-Andersen M, Schiöth HB, Gloriam DE. Trends in GPCR drug
20 discovery: new agents, targets and indications. *Nature reviews Drug discovery*.
21 2017;16(12):829–42.
- 22 2. Wootten D, Christopoulos A, Marti-Solano M, Babu MM, Sexton PM. Mechanisms of
23 signalling and biased agonism in G protein-coupled receptors. *Nat Rev Mol Cell Biol*.
24 2018;19(10):638–53.
- 25 3. Kenakin T. G protein coupled receptors as allosteric proteins and the role of allosteric
26 modulators. *J Recept Signal Transduct Res*. 2010 Oct;30(5):313–21.
- 27 4. Gentry PR, Sexton PM, Christopoulos A. Novel Allosteric Modulators of G Protein-coupled
28 Receptors. *J Biol Chem*. 2015 Aug 7;290(32):19478–88.
- 29 5. Keov P, Sexton PM, Christopoulos A. Allosteric modulation of G protein-coupled receptors: a
30 pharmacological perspective. *Neuropharmacology*. 2011;60(1):24–35.
- 31 6. Thal DM, Glukhova A, Sexton PM, Christopoulos A. Structural insights into G-protein-coupled
32 receptor allostery. *Nature*. 2018;559(7712):45–53.
- 33 7. Congreve M, Oswald C, Marshall FH. Applying structure-based drug design approaches to
34 allosteric modulators of GPCRs. *Trends in pharmacological sciences*. 2017;38(9):837–47.

- 1
2
3 1 8. Kruse AC, Ring AM, Manglik A, Hu J, Hu K, Eitel K, et al. Activation and allosteric
4 2 modulation of a muscarinic acetylcholine receptor. *Nature*. 2013;504(7478):101–6.
5
6 3 9. Liu X, Masoudi A, Kahsai AW, Huang L-Y, Pani B, Staus DP, et al. Mechanism of β 2AR
7 4 regulation by an intracellular positive allosteric modulator. *Science*. 2019;364(6447):1283–7.
8
9
10 5 10. Zheng Y, Qin L, Zacarías NVO, de Vries H, Han GW, Gustavsson M, et al. Structure of CC
11 6 chemokine receptor 2 with orthosteric and allosteric antagonists. *Nature*. 2016;540(7633):458–
12 7 61.
13
14 8 11. Jaeger K, Bruenle S, Weinert T, Guba W, Muehle J, Miyazaki T, et al. Structural basis for
15 9 allosteric ligand recognition in the human CC chemokine receptor 7. *Cell*. 2019;178(5):1222-
16 10 1230. e10.
17
18
19 11 12. Oswald C, Rappas M, Kean J, Doré AS, Errey JC, Bennett K, et al. Intracellular allosteric
20 12 antagonism of the CCR9 receptor. *Nature*. 2016;540(7633):462–5.
21
22 13 13. Zhang D, Gao Z-G, Zhang K, Kiselev E, Crane S, Wang J, et al. Two disparate ligand-binding
23 14 sites in the human P2Y 1 receptor. *Nature*. 2015;520(7547):317–21.
24
25 15 14. Shao Z, Yan W, Chapman K, Ramesh K, Ferrell AJ, Yin J, et al. Structure of an allosteric
26 16 modulator bound to the CB1 cannabinoid receptor. *Nature chemical biology*.
27 17 2019;15(12):1199–205.
28
29
30 18 15. Robertson N, Rappas M, Doré AS, Brown J, Bottegoni G, Koglin M, et al. Structure of the
31 19 complement C5a receptor bound to the extra-helical antagonist NDT9513727. *Nature*. 2018
32 20 03;553(7686):111–4.
33
34 21 16. Liu H, Kim HR, Deepak RK, Wang L, Chung KY, Fan H, et al. Orthosteric and allosteric
35 22 action of the C5a receptor antagonists. *Nature structural & molecular biology*. 2018;25(6):472–
36 23 81.
37
38
39 24 17. Hollenstein K, Kean J, Bortolato A, Cheng RK, Doré AS, Jazayeri A, et al. Structure of class B
40 25 GPCR corticotropin-releasing factor receptor 1. *Nature*. 2013;499(7459):438–43.
41
42 26 18. Lu J, Byrne N, Wang J, Bricogne G, Brown FK, Chobanian HR, et al. Structural basis for the
43 27 cooperative allosteric activation of the free fatty acid receptor GPR40. *Nature structural &*
44 28 *molecular biology*. 2017;24(7):570–7.
45
46 29 19. Jazayeri A, Doré AS, Lamb D, Krishnamurthy H, Southall SM, Baig AH, et al. Extra-helical
47 30 binding site of a glucagon receptor antagonist. *Nature*. 2016;533(7602):274–7.
48
49
50 31 20. Song G, Yang D, Wang Y, de Graaf C, Zhou Q, Jiang S, et al. Human GLP-1 receptor
51 32 transmembrane domain structure in complex with allosteric modulators. *Nature*.
52 33 2017;546(7657):312–5.
53
54 34 21. Bueno AB, Sun B, Willard FS, Feng D, Ho JD, Wainscott DB, et al. Structural insights into
55 35 probe-dependent positive allosterism of the GLP-1 receptor. *Nature Chemical Biology*. 2020
56 36 Oct;16(10):1105–10.
57
58
59
60

- 1
2
3 1 22. Cheng RK, Fiez-Vandal C, Schlenker O, Edman K, Aggeler B, Brown DG, et al. Structural
4 2 insight into allosteric modulation of protease-activated receptor 2. *Nature*.
5 3 2017;545(7652):112–5.
6
7
8 4 23. Liu X, Kaindl J, Korczynska M, Stöbel A, Dengler D, Stanek M, et al. An allosteric modulator
9 5 binds to a conformational hub in the β 2 adrenergic receptor. *Nature Chemical Biology*.
10 6 2020;1–7.
11
12 7 24. Wagner JR, Lee CT, Durrant JD, Malmstrom RD, Feher VA, Amaro RE. Emerging
13 8 Computational Methods for the Rational Discovery of Allosteric Drugs. *Chem Rev*. 2016 Jun
14 9 8;116(11):6370–90.
15
16
17 10 25. Ghanakota P, Carlson HA. Driving Structure-Based Drug Discovery through Cosolvent
18 11 Molecular Dynamics: Miniperspective. *Journal of medicinal chemistry*. 2016;59(23):10383–99.
19
20 12 26. Prakash P, Hancock JF, Gorfe AA. Binding hotspots on K-ras: Consensus ligand binding sites
21 13 and other reactive regions from probe-based molecular dynamics analysis. *Proteins: Structure,*
22 14 *Function, and Bioinformatics*. 2015;83(5):898–909.
23
24 15 27. Lakkaraju SK, Yu W, Raman EP, Hershfeld AV, Fang L, Deshpande DA, et al. Mapping
25 16 functional group free energy patterns at protein occluded sites: nuclear receptors and G-protein
26 17 coupled receptors. *Journal of chemical information and modeling*. 2015;55(3):700–8.
28
29 18 28. MacKerell AD, Jo S, Lakkaraju SK, Lind C, Yu W. Identification and characterization of
30 19 fragment binding sites for allosteric ligand design using the site identification by ligand
31 20 competitive saturation hotspots approach (SILCS-Hotspots). *Biochim Biophys Acta Gen Subj*.
32 21 2020;1864(4):129519.
33
34 22 29. Alvarez-Garcia D, Barril X. Molecular simulations with solvent competition quantify water
35 23 displaceability and provide accurate interaction maps of protein binding sites. *Journal of*
36 24 *medicinal chemistry*. 2014;57(20):8530–9.
38
39 25 30. Bakan A, Nevins N, Lakdawala AS, Bahar I. Druggability Assessment of Allosteric Proteins by
40 26 Dynamics Simulations in the Presence of Probe Molecules. *J Chem Theory Comput*. 2012 Jul
41 27 10;8(7):2435–47.
42
43 28 31. Ung PMU, Ghanakota P, Graham SE, Lexa KW, Carlson HA. Identifying binding hot spots on
44 29 protein surfaces by mixed-solvent molecular dynamics: HIV-1 protease as a test case.
45 30 *Biopolymers*. 2016;105(1):21–34.
47
48 31 32. Posokhov YO, Kyrychenko A. Effect of acetone accumulation on structure and dynamics of
49 32 lipid membranes studied by molecular dynamics simulations. *Computational Biology and*
50 33 *Chemistry*. 2013;46:23–31.
51
52 34 33. Odinkov A, Ostroumov D. Structural degradation and swelling of lipid bilayer under the
53 35 action of benzene. *The Journal of Physical Chemistry B*. 2015;119(48):15006–13.
54
55 36 34. Reigada R. Atomistic study of lipid membranes containing chloroform: looking for a lipid-
56 37 mediated mechanism of anesthesia. *PLoS One*. 2013;8(1):e52631.
58
59
60

- 1
2
3 1 35. Yuan X, Raniolo S, Limongelli V, Xu Y. The Molecular Mechanism Underlying Ligand
4 2 Binding to the Membrane-Embedded Site of a G-Protein-Coupled Receptor. *J Chem Theory*
5 3 *Comput.* 2018 May 8;14(5):2761–70.
- 6
7 4 36. Smith RD, Lu J, Carlson HA. Are there physicochemical differences between allosteric and
8 5 competitive ligands? *PLoS computational biology.* 2017;13(11):e1005813.
- 9
10 6 37. Wang Q, Zheng M, Huang Z, Liu X, Zhou H, Chen Y, et al. Toward understanding the
11 7 molecular basis for chemical allosteric modulator design. *Journal of Molecular Graphics and*
12 8 *Modelling.* 2012;38:324–33.
- 13
14 9 38. Ciancetta A, O'Connor RD, Paoletta S, Jacobson KA. Demystifying P2Y1 Receptor Ligand
15 10 Recognition through Docking and Molecular Dynamics Analyses. *J Chem Inf Model.* 2017 Dec
16 11 26;57(12):3104–23.
- 17
18 12 39. Jiménez-Rosés M, Matsoukas M-T, Caltabiano G, Cordoní A. Ligand-Triggered Structural
19 13 Changes in the M2 Muscarinic Acetylcholine Receptor. *J Chem Inf Model.* 2018 May
20 14 29;58(5):1074–82.
- 21
22 15 40. Ballesteros JA, Weinstein H. [19] Integrated methods for the construction of three-dimensional
23 16 models and computational probing of structure-function relations in G protein-coupled
24 17 receptors. In: Sealfon SC, editor. *Methods in Neurosciences* [Internet]. Academic Press; 1995
25 18 [cited 2021 Mar 29]. p. 366–428. (Receptor Molecular Biology; vol. 25). Available from:
26 19 <https://www.sciencedirect.com/science/article/pii/S1043947105800497>
- 27
28 20 41. Schmidtke P, Bidon-Chanal A, Luque FJ, Barril X. MDpocket: open-source cavity detection
29 21 and characterization on molecular dynamics trajectories. *Bioinformatics.* 2011;27(23):3276–85.
- 30
31 22 42. Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *Journal of molecular*
32 23 *graphics.* 1996;14(1):33–8.
- 33
34 24 43. Le Guilloux V, Schmidtke P, Tuffery P. Fpocket: An open source platform for ligand pocket
35 25 detection. *BMC Bioinformatics.* 2009 Jun 2;10:168.
- 36
37 26 44. Schmidtke P, Le Guilloux V, Maupetit J, Tufféry P. fpocket: online tools for protein ensemble
38 27 pocket detection and tracking. *Nucleic Acids Res.* 2010 Jul 1;38(Web Server issue):W582–9.
- 39
40 28 45. Wood M, Ates A, Andre VM, Michel A, Barnaby R, Gillard M. In Vitro and In Vivo
41 29 Identification of Novel Positive Allosteric Modulators of the Human Dopamine D2 and D3
42 30 Receptor. *Mol Pharmacol.* 2016 Feb;89(2):303–12.
- 43
44 31 46. Wang S, Che T, Levit A, Shoichet BK, Wacker D, Roth BL. STRUCTURE OF THE D2
45 32 DOPAMINE RECEPTOR BOUND TO THE ATYPICAL ANTIPSYCHOTIC DRUG
46 33 RISPERIDONE. *Nature.* 2018 Mar 8;555(7695):269–73.
- 47
48 34 47. Zhuang Y, Krumm B, Zhang H, Zhou XE, Wang Y, Huang X-P, et al. Mechanism of dopamine
49 35 binding and allosteric modulation of the human D1 dopamine receptor. *Cell Res.* 2021 Mar 9;
- 50
51 36 48. Rossi M, Fasciani I, Marampon F, Maggio R, Scarselli M. The First Negative Allosteric
52 37 Modulator for Dopamine D2 and D3 Receptors, SB269652 May Lead to a New Generation of
53 38 Antipsychotic Drugs. *Mol Pharmacol.* 2017 Jun;91(6):586–94.
- 54
55
56
57
58
59
60

- 1
2
3 1 49. Lane JR, Chubukov P, Liu W, Canals M, Cherezov V, Abagyan R, et al. Structure-Based
4 2 Ligand Discovery Targeting Orthosteric and Allosteric Pockets of Dopamine Receptors. *Mol*
5 3 *Pharmacol.* 2013 Dec;84(6):794–807.
- 7 4 50. Fyfe TJ, Zarzycka B, Lim HD, Kellam B, Mistry SN, Katrich V, et al. A Thieno[2,3-
8 5 d]pyrimidine Scaffold Is a Novel Negative Allosteric Modulator of the Dopamine D2 Receptor.
9 6 *J Med Chem.* 2019 Jan 10;62(1):174–206.
- 12 7 51. Draper-Joyce CJ, Michino M, Verma RK, Herenbrink CK, Shonberg J, Kopinathan A, et al.
13 8 The structural determinants of the bitopic binding mode of a negative allosteric modulator of
14 9 the dopamine D2 receptor. *Biochem Pharmacol.* 2018 Feb;148:315–28.
- 16 10 52. Free RB, Chun LS, Moritz AE, Miller BN, Doyle TB, Conroy JL, et al. Discovery and
17 11 characterization of a G protein-biased agonist that inhibits β -arrestin recruitment to the D2
18 12 dopamine receptor. *Mol Pharmacol.* 2014 Jul;86(1):96–105.
- 21 13 53. McCorvy JD, Butler KV, Kelly B, Rechsteiner K, Karpiak J, Betz RM, et al. Structure-inspired
22 14 design of β -arrestin-biased ligands for aminergic GPCRs. *Nat Chem Biol.* 2018 Feb;14(2):126–
23 15 34.
- 25 16 54. Wakefield AE, Mason JS, Vajda S, Keserú GM. Analysis of tractable allosteric sites in G
26 17 protein-coupled receptors. *Scientific reports.* 2019;9(1):1–14.
- 29 18 55. Miao Y, Nichols SE, McCammon JA. Mapping of Allosteric Druggable Sites in Activation-
30 19 Associated Conformers of the M2 Muscarinic Receptor. *Chem Biol Drug Des.* 2014
31 20 Feb;83(2):237–46.
- 33 21 56. Caliman AD, Miao Y, McCammon JA. Mapping the Allosteric Sites of the A2A Adenosine
34 22 Receptor. *Chem Biol Drug Des.* 2018 Jan;91(1):5–16.
- 36 23 57. Schrödinger Release 2018-4: Maestro, Schrödinger, LLC, New York, NY, 2018.
- 38 24 58. Lomize MA, Lomize AL, Pogozheva ID, Mosberg HI. OPM: orientations of proteins in
39 25 membranes database. *Bioinformatics.* 2006;22(5):623–5.
- 41 26 59. Martínez L, Andrade R, Birgin EG, Martínez JM. PACKMOL: A package for building initial
42 27 configurations for molecular dynamics simulations. *Journal of Computational Chemistry.*
43 28 2009;30(13):2157–64.
- 46 29 60. Doerr S, Harvey MJ, Noé F, De Fabritiis G. HTMD: High-Throughput Molecular Dynamics for
47 30 Molecular Discovery. *J Chem Theory Comput.* 2016 Apr 12;12(4):1845–52.
- 49 31 61. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular
50 32 dynamics with NAMD. *Journal of computational chemistry.* 2005;26(16):1781–802.
- 52 33 62. Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, et al. Optimization of the additive
53 34 CHARMM all-atom protein force field targeting improved sampling of the backbone ϕ , ψ and
54 35 side-chain χ_1 and χ_2 dihedral angles. *J Chem Theory Comput.* 2012 Sep 11;8(9):3257–73.
- 57 36 63. Klauda JB, Venable RM, Freites JA, O'Connor JW, Tobias DJ, Mondragon-Ramirez C, et al.
58 37 Update of the CHARMM all-atom additive force field for lipids: validation on six lipid types. *J*
59 38 *Phys Chem B.* 2010 Jun 17;114(23):7830–43.

- 1
2
3 1 64. Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, et al. CHARMM
4 2 general force field: A force field for drug-like molecules compatible with the CHARMM all-
5 3 atom additive biological force fields. *J Comput Chem*. 2010 Mar;31(4):671–90.
6
7
8 4 65. Phillips JC, Hardy DJ, Maia JD, Stone JE, Ribeiro JV, Bernardi RC, et al. Scalable molecular
9 5 dynamics on CPU and GPU architectures with NAMD. *The Journal of Chemical Physics*.
10 6 2020;153(4):044130.
11
12 7 66. OpenEye Toolkits 2020.1.0 OpenEye Scientific Software, Santa Fe, NM.
13 8 <http://www.eyesopen.com>.
14
15 9 67. Farid R, Day T, Friesner RA, Pearlstein RA. New insights about HERG blockade obtained from
16 10 protein modeling, potential energy mapping, and docking studies. *Bioorg Med Chem*. 2006
17 11 May 1;14(9):3160–73.
18
19
20 12 68. Sherman W, Day T, Jacobson MP, Friesner RA, Farid R. Novel procedure for modeling
21 13 ligand/receptor induced fit effects. *J Med Chem*. 2006 Jan 26;49(2):534–53.
22
23 14 69. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: a new
24 15 approach for rapid, accurate docking and scoring. 1. Method and assessment of docking
25 16 accuracy. *J Med Chem*. 2004 Mar 25;47(7):1739–49.
26
27
28 17 70. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and testing of the OPLS all-atom
29 18 force field on conformational energetics and properties of organic liquids. *Journal of the*
30 19 *American Chemical Society*. 1996;118(45):11225–36.
31
32 20
33
34 21
35 22
36 23
37 24
38 25
39 26
40 27
41 28
42 29
43 30
44 31
45 32
46 33
47 34
48 35
49 36
50 37
51 38
52 39
53 40
54 41
55 42
56 43
57 44
58
59
60

1
2
3 1
4 2
5 3
6 4
7 5
8 6
9 7
10 8
11 9
12 10
13 11
14 12
15 13
16 14
17 15
18 16
19 17
20 18
21 19
22 20
23 21
24 22
25 23
26 24
27 25
28 26
29 27
30 28
31 29
32 30
33 31
34 32
35 33
36 34
37 35
38 36
39 37
40 38
41 39
42 40
43 41
44 42
45 43
46 44
47 45
48 46
49 47
50 48
51 49
52 50
53 51
54 52
55 53
56 54

Video 1. A video of the probe confined dynamic mapping trajectory for the THP probe molecules binding to the EC side of the M₂ receptor. The receptor is shown in white cartoon representation with residues forming the allosteric site in stick representation. THP probe molecules are in space-filling representation and the orthosteric ligand, Iperoxo is in orange stick.

Video 2. A video of the probe confined dynamic mapping trajectory for the BTA probe molecules binding to the IC side of the β₂ receptor. The receptor is shown in white cartoon representation with residues forming the allosteric site in stick representation and the BTA probe molecules are in space-filling representation.

Video 3. A video of the probe confined dynamic mapping trajectory for the P2Y₁ probe molecules binding to the LI of helices 1-3 of the P2Y₁ receptor. The receptor is shown in grey cartoon representation with residues forming the allosteric site in stick representation, the P2O probe molecules are in space-filling representation and the POPC bilayer in stick representation.

Table S1. The strongest ligand-residue interaction energy (IE) from the triplicate conventional MD simulations trajectories of the X-ray receptor-ligand complexes. Residues selected for allosteric interaction spots are in bold.

Top residues selected from the lowest ligand-residue interaction energy, (kcal/mol)					
^a M ₂ -LY211960		^b β ₂ -Cmp-15		^c P ₂ Y ₁ -BPTU	
E175	-18	D331	-30	L102	-9.9
Y177	-13.1	S329	-12.7	T103	-5.9
W422	-11.4	N69	-11.3	M123	-4.9
N410	-8.7	R63	-9.7	P105	-3.4
E172	-7.4	L64	-6.7	F119	-2.5
Y80	-7	F332	-5.6	Q127	-1.8
T170	-4.7	K270	-4.8	L126	-1.7
Y426	-3.7	T68	-4.6	F66	-1.4
Tyr83	-3.1	I72	-3.1	V101	-1.4
N419	-3	T274	-2.9	F62	-1.3
F181	-2	A271	-2.4	A106	-1.3
S182	-1.5	Y326	-1.8	L99	-1
T84	-1.3	A335	-1.8	C124	-0.8
T423	-1.3	P330	-1.7	W117	-0.7
V407	-0.3	V54	-1.6	I130	-0.5
Y403	-0.2	F61	-1.5		
		I58	-1.3		
		R328	-1.3		
		T66	-1.2		
		I334	-1.2		
		I325	-1		
		T73	-0.8		
		L275	-0.6		

^a For the M₂ receptor ternary complex with LY211960, we identified the residue establishing the strongest and most persistent electrostatic interactions - E175^{ECL2}. This residue initially lied far away from the allosteric site, however, the PAM piperazine ring underwent conformational changes by adopting alternative orientations during the MD simulations. As for the van der Waals contribution to the total IE, the residues most contributing were Y177^{ECL2} and W422^{7.35}.

^bThe ligand consistently shifted toward helices 6-7 interface by ~1.5 Å in all three trajectories. D331^{8.49}, S329^{8.47} and N69^{2.40} are the residues mostly contributing via electrostatic interactions to the ligand binding and L64^{ICL1} is the residue establishing the strongest and most persistent van der Waals interaction.

^c Trajectory visual inspection identified that the ligand *p*-trifluoromethyl phenyl ring and, to a lesser extent, the pyridine ring slightly moved away from the transmembrane bundle towards the membrane in all the trajectories. These fluctuations are reflected in the IE analysis, as the residue establishing the strongest and most persistent electrostatic interactions with the ligand is L102^{2.55}, whereas the strongest and most persistent van der Waals interactions engaged L102^{2.55}, T103^{2.56}, P105^{2.58}, and M123^{3.24} sidechains surrounding the ligand *t*-butyl phenyl and pyridine rings.

Table S2. P2O and PHX probe occupancy at the P₂Y₁ receptor from probe confined dynamic mapping of various helix-lipid interfaces (calculated over five independent trajectories). The amino acid residues forming a binding cavity at the lipid interface and used to define interaction spots are also reported in the table.

Receptor Area at the lipid interface	Binding cavity detected by MDpocket	P2O occupancy, %	PHX occupancy, %
1. Interface of helices 2, 3 and 4 (L95 ^{2.48} , A96 ^{2.49} , I130 ^{3.31} , V133 ^{3.34} , N134 ^{3.35} , W176 ^{4.50} , and V180 ^{4.54})	Yes	18±18	9±9
2. Interface of helices 3, 4 and 5	No	0	0
3. Interface of helices 6 and 7 (T267 ^{6.42} , V268 ^{6.43} , V271 ^{6.46} , S272 ^{6.47} , L315 ^{7.44} , N316 ^{7.45} , and V319 ^{7.48})	Yes	25±30	10±17
4. Interface of helices 7 and 1 (F49 ^{1.30} , L54 ^{1.35} , V57 ^{1.38} , V308 ^{7.37} , G311 ^{7.40} , L312 ^{7.41} , and L315 ^{7.44})	Yes	0	0

Table S3. Pharmacological parameter analysis of Forskolin-Induced cAMP accumulation assays of D₂ receptor wild type and mutants in the absence and presence of UCB compound. Data values correspond to **Figure 4** in the absence and presence of the allosteric modulator. In the absence of the PAM, ΔpEC_{50} is calculated relative to the pEC_{50} of the D₂ WT with dopamine only. In the presence of the PAM, ΔpEC_{50} is calculated relative to the calculated pEC_{50} of the respective construct in the absence of the PAM. Each data value represents the mean \pm SEM from three independent experiments, each condition being in triplicate.

	Dopamine			Dopamine + PAM (10 μ M)			ΔE_{max} (%)
	pEC_{50}	ΔpEC_{50}	E_{max} (%)	pEC_{50}	ΔpEC_{50}	E_{max} (%)	
D2 WT	9.25 \pm 0.12	0.00	11.22 \pm 1.87	10.43 \pm 0.24	-1.18	32.60 \pm 7.23	21.40
V91A	9.06 \pm 0.95	-0.19	54.90 \pm 7.74	/	/	89.17 \pm 12.9	34.27
L94A	8.71 \pm 0.54	-0.54	41.97 \pm 6.63	/	/	99.14 \pm 8.59	57.17
L94W	9.72 \pm 0.31	0.47	12.56 \pm 2.72	10.02 \pm 0.66	0.30	35.64 \pm 6.21	23.08
E95A	9.12 \pm 0.97	-0.13	27.44 \pm 3.73	11.91 \pm 0.97	2.79	63.10 \pm 4.80	35.66
W100A	8.11 \pm 0.75	-1.14	49.11 \pm 13.8	-	-	-	-
I184A	8.29 \pm 0.11	-0.96	28.67 \pm 2.58	-	-	-	-
W413A	8.83 \pm 0.59	-0.42	24.39 \pm 3.99	11.87 \pm 0.87	3.04	30.54 \pm 1.33	6.15

Table S4 Primer Sequences for Site-Directed Mutagenesis of the D₂ Mutants. Both the forward (*FW*) and reverse (*RV*) primer sequences used for site-directed mutagenesis of the D₂ mutants, using the D_{2L} plasmid as the template, are denoted in the 5' to 3' direction. All primers were designed using Benchling. Mutated residues are indicated in red.

Mutant	Template	Primer Sequence (5' → 3')	
pcDNA3.1 - D _{2L} L94W	pcDNA3.1 - D ₂	<i>FW</i>	CTACTGGGAGGTGGTAGGTGAGTGGAAATTC
		<i>RV</i>	CACCTCCAGTAGACAACCCAGGGCAT
pcDNA3.1 - D _{2L} E95A	pcDNA3.1 - D ₂	<i>FW</i>	CCTGGCGGTGGTAGGTGAGTGG
		<i>RV</i>	ACCGCCAGGTAGACAACCCAGG
pcDNA3.1 - D _{2L} W100A	pcDNA3.1 - D ₂	<i>FW</i>	GTGAGGCGAAATTCAGCAGGATTCCT
		<i>RV</i>	ATTTCCGCTCACCTACCACCTCCAGG
pcDNA3.1 - D _{2L} L94A	pcDNA3.1 - D ₂	<i>FW</i>	CTACGCGGAGGTGGTAGGTGAGTGGAAATTC
		<i>RV</i>	CCTCCGCGTAGACAACCCAGGGC
pcDNA3.1 - D _{2L} V91A	pcDNA3.1 - D ₂	<i>FW</i>	CTGGGCTGTCTACCTGGAGGTGG
		<i>RV</i>	GACAGCCCAGGGCATGACCAGT
pcDNA3.1 - D _{2L} I184A	pcDNA3.1 - D ₂	<i>FW</i>	GCATCGCTGCCAACCCGGCCTT
		<i>RV</i>	GCA GCGATGCACTCGTTCTGGTCTG
pcDNA3.1 - D _{2L} W413A	pcDNA3.1 - D ₂	<i>FW</i>	CACGCGCTGGGCTATGTCAACAGCG
		<i>RV</i>	CAGCGCGTGAAGGCGCTGTACAG

Figure S1. Search for privileged fragments from known allosteric modulators. Maximum common substructure search for the muscarinic PAMs led to Structure 1. Fragmentation of structures of the known NAM for the β_2 and P_2Y_1 receptors. The ligand structure was fragmented by functional groups for both the β_2 and P_2Y_1 receptors.

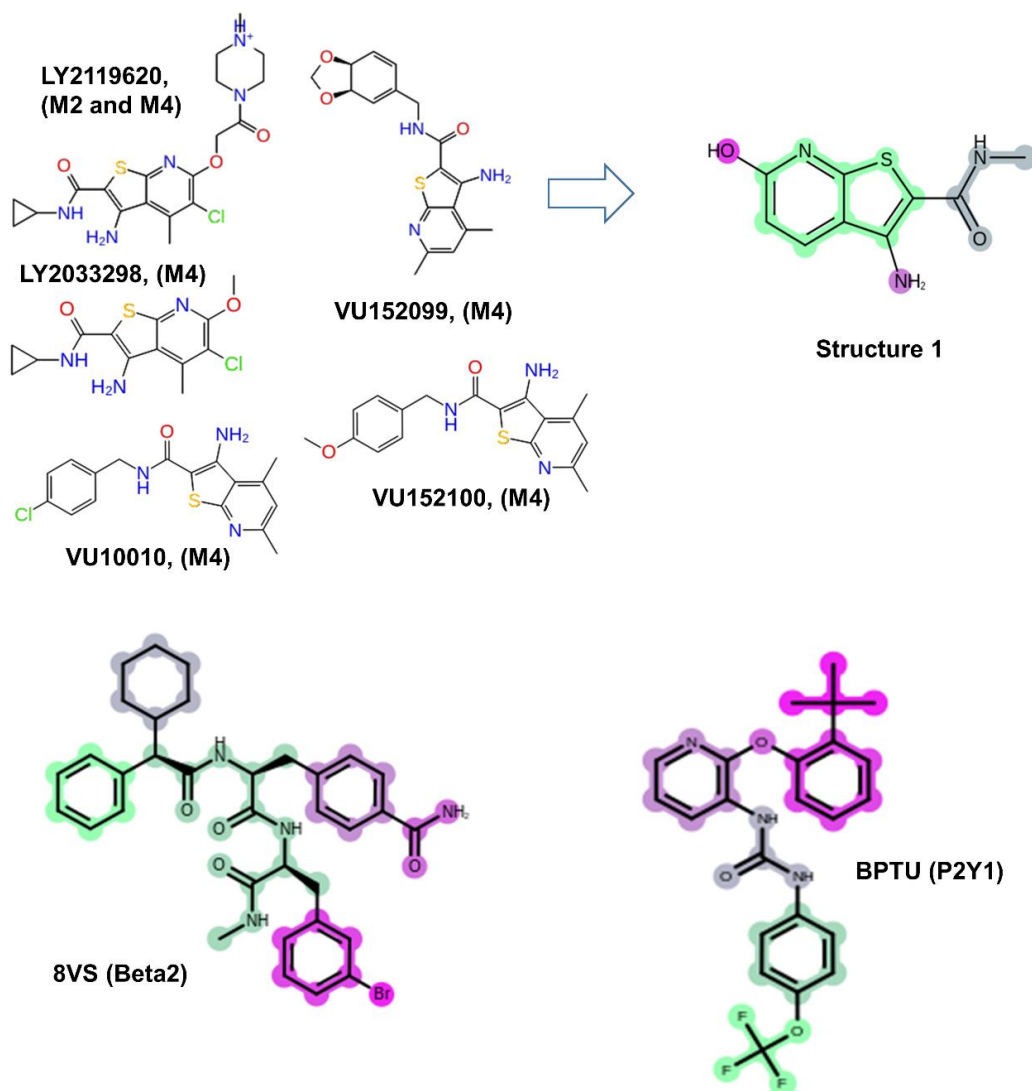


Figure S2. Probe positions and interactions in the allosteric site of the M₂ receptor. The probe molecules and the orthosteric ligand are shown with green and orange carbon atoms, respectively. Only the residues that are used for probe interaction spots are shown in stick-like representation. Hydrogen bonds and π - π interactions are shown in black and cyan dashed lines, respectively. A MD snapshot with the probe occupying the allosteric site was selected to generate the images.

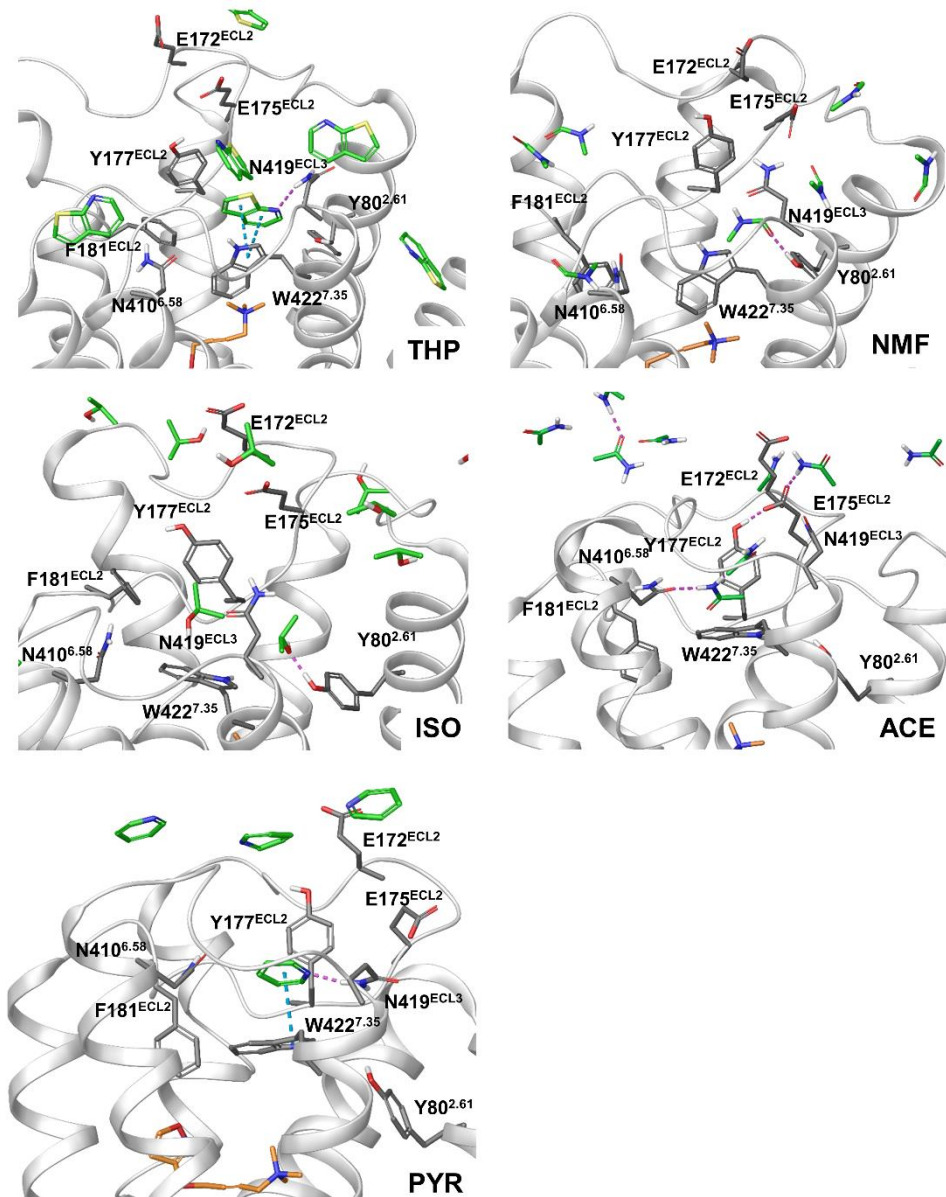


Figure S3. Probe positions and interactions in the allosteric site of the β_2 receptor. Probe molecules are shown with green carbon atoms. Only the residues used for probe interaction spots are shown in stick-like representation. Hydrogen bonds and π - π interactions are shown in black and cyan dashed lines, respectively. A MD snapshot with probe occupying the allosteric site was selected to generate the images.

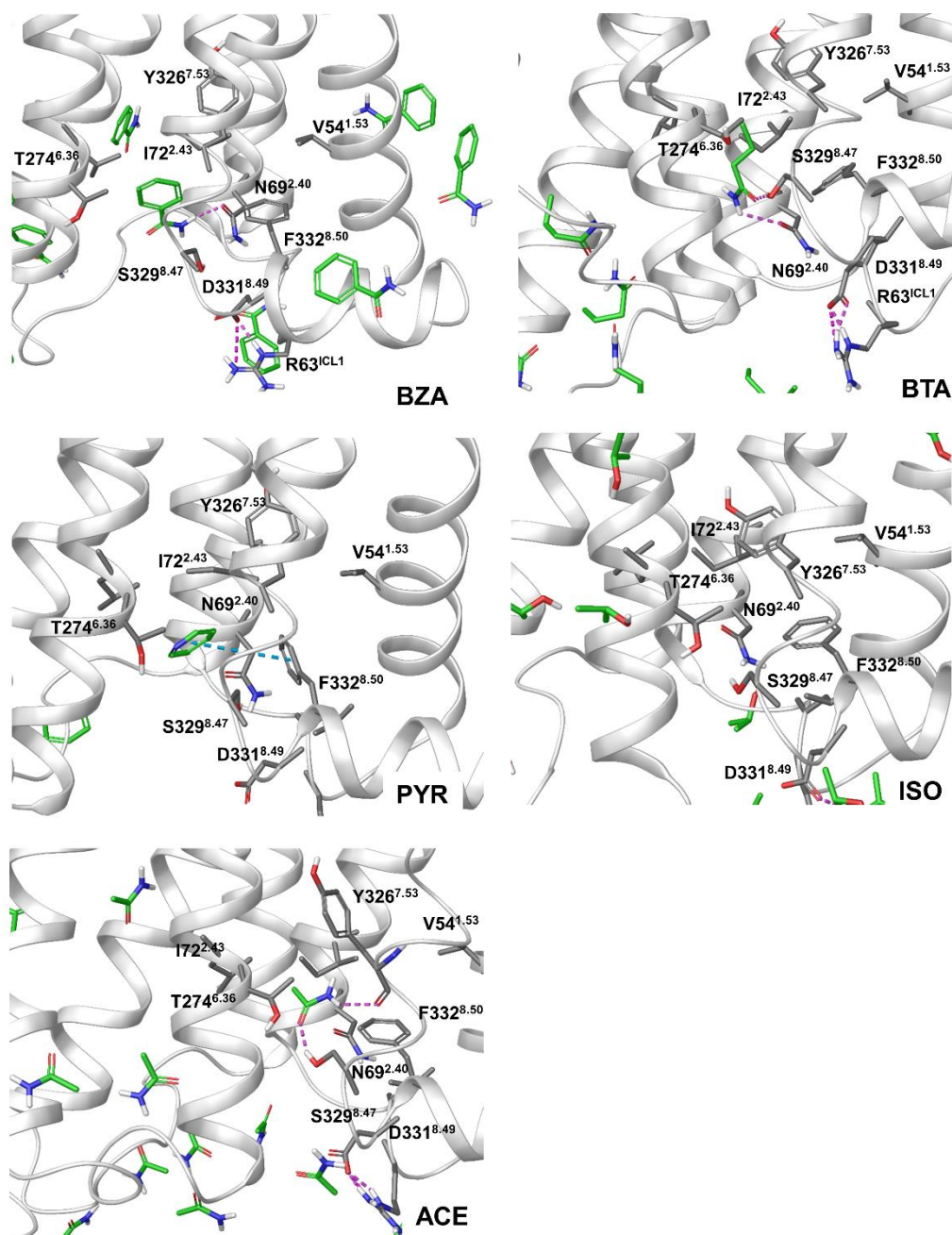
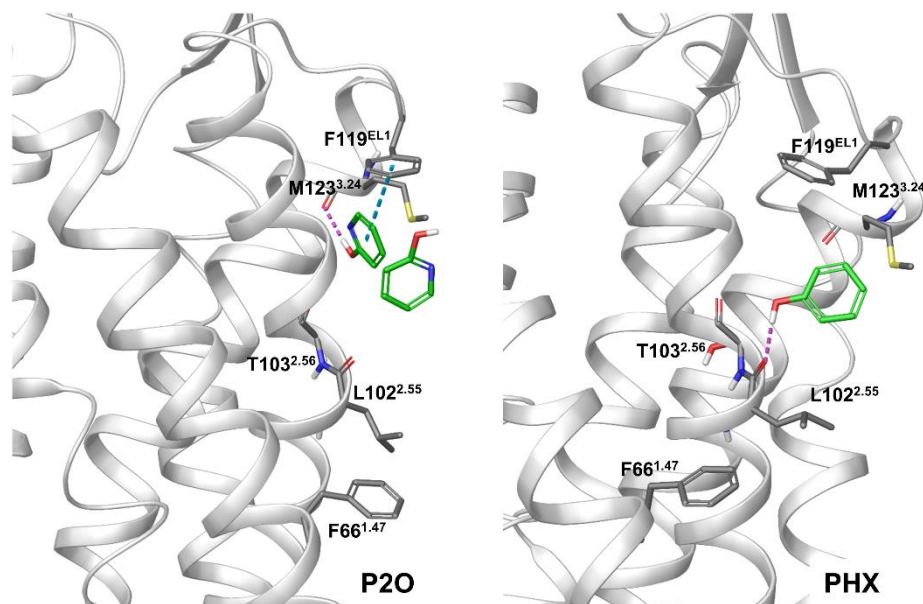
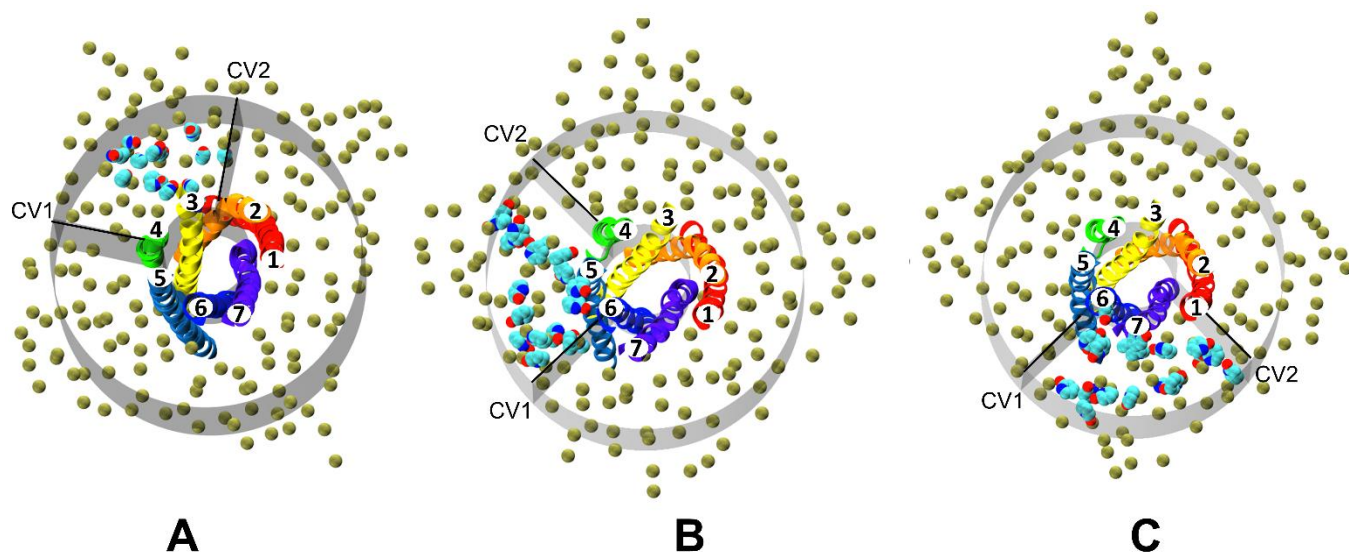


Figure S4. Probe positions and interactions in the allosteric site of the P₂Y₁ receptor. Probe molecules are shown with green carbon atoms. Only the residues that are used for probe interaction spots are shown in stick-like representation, respectively. Hydrogen bonds and π - π interactions are shown in black and cyan dashed lines, respectively. A MD snapshot with the probes occupying the allosteric site was selected to generate the images.



1
2
3 1 **Figure S5. The cylinder-shaped harmonic wall potential with addition of two collective**
4 2 **variables (CV1 and CV2) to confine the movement of the probes at the lipid interface of**
5 3 **helices 2-4 (A), helices 3-5 (B) and helices 1-6-7 (C) in the P2Y1 production simulations.** The
6 4 collective variables that define a cylinder were selected with lower and upper boundaries (10 and 35
7 5 Å). The wall potential in **C** shows the sampling of cavities at the interface of 6 and 7 and at the
8 6 interface of 1 and 7 (Table S2). Only transmembrane helices are shown in rainbow cartoon.



1
2
3 1 Figure S6. cAMP Accumulation assays in the absence and presence of the allosteric
4 2 modulator at the D₂ WT, I184A and W413A mutants
5
6 3
7 4
8
9

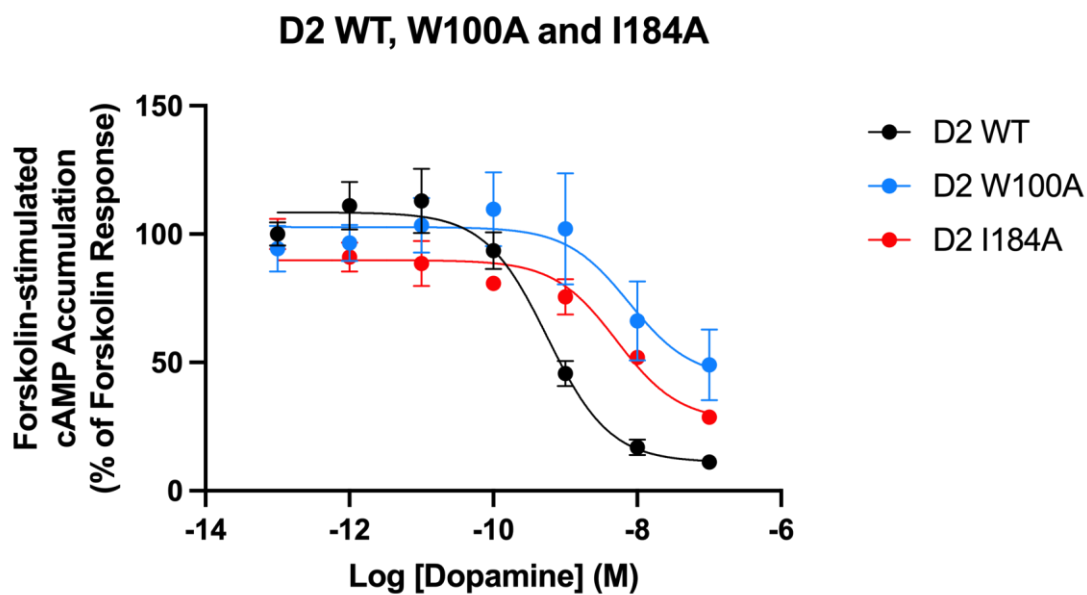


Figure S7. Expression of D₂ WT and mutants as measured by immunofluorescence. HEK293 cells were transiently transfected as reported in materials and methods. D2 receptor construct staining is shown in red and DAPI staining in blue.

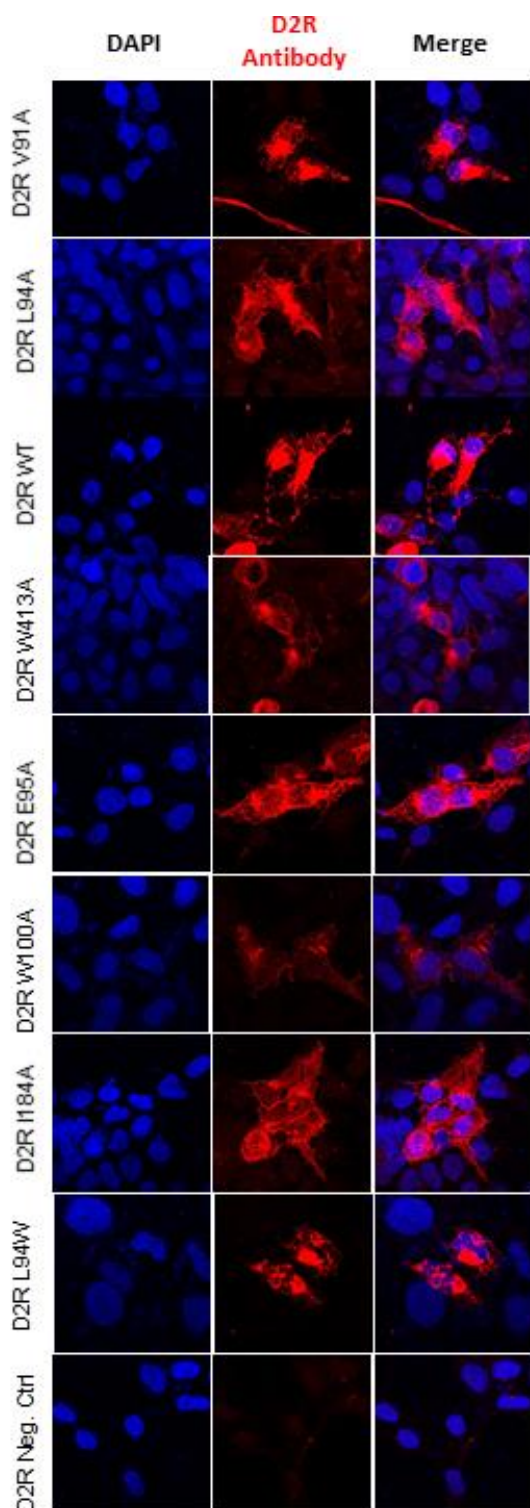


Figure S8. Cross-titration curves to calculate K_b . Concentration-response curves measuring cAMP accumulation using the endogenous agonist, dopamine, were performed with A) D2 WT, B) D2 W413A and C) D2 L94W. The UCB compound was added at 30 μ M, 10 μ M, 3 μ M, 1 μ M, 0.3 μ M and 0 μ M. The Allosteric EC_{50} shift was used to plot the curves and calculate the K_b value (D). Each data point represents the mean \pm SEM of duplicate wells of five independent experiments.

