Dynamic Mechanism for Subtype Selectivity of Endocannabinoids

Soumajit Dutta,[†] Lawrence Zhao,[‡] and Diwakar Shukla[*,†,¶,§,‖]

[†]Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, 61801

[‡]Department of Computer Science, Yale University, New Haven, Connecticut, 06520

[¶]Center for Biophysics and Quantitative Biology, University of Illinois at Urbana-Champaign, Urbana, IL, 61801

[§]Department of Bioengineering, University of Illinois at Urbana-Champaign, Urbana, IL, 61801

[‖]Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL, 61801

E-mail: diwakar@illinois.edu

Abstract

Endocannabinoids are naturally occurring lipid-like molecules that bind to cannabinoid receptors (CB_1 and CB_2) and regulate many of human bodily functions via the endocannabinoid system. There is a tremendous interest in developing selective drugs that target the CB receptors. However, the biophysical mechanisms responsible for the subtype selectivity for endocannbinoids have not been established. Recent experimental structures of CB receptors show that endocannbinoids potentially bind via membrane using the lipid access channel in the transmembrane region of the receptors. Furthermore, the N-terminus of the receptor

1

could move in and out of the binding pocket thereby modulating both the pocket volume and its residue composition. On the basis of these observations, we propose two hypothesis to explain the selectivity of the endocannabinoid, anandamide for CB_1 receptor. First, the selectivity arises from distinct enthalpic ligand-protein interactions along the ligand binding pathway formed due to the movement of N-terminus and subsequent shifts in the binding pocket composition. Second, selectivity arises from the volumetric differences in the binding pocket allowing for differences in ligand conformational entropy. To quantitatively test these hypothesis, we perform extensive molecular dynamics simulations ($\sim$0.9 milliseconds) along with Markov state modeling and deep learning-based VAMPnets to provide an interpretable characterization of the anandamide binding process to cannabinoid receptors and explain its selectivity for CB_1. Our findings reveal that the distinct N-terminus positions along lipid access channels between TM1 and TM7 lead to different binding mechanisms and interactions between anandamide and the binding pocket residues. To validate the critical stabilizing interactions along the binding pathway, relative free energy calculations of anandamide analogs are used. Moreover, the larger CB_2 pocket volume increases the entropic effects of ligand binding by allowing higher ligand fluctuations but reduced stable interactions. Therefore, the opposing enthalpy and entropy effects between the receptors shape the endocannabinoid selectivity. Overall, the CB_1 selectivity of anandamide is explained by the dominant enthalpy contributions due to ligand-protein interactions in stable binding poses. This study shed lights on potential selectivity mechanisms for endocannabinoids that would aid in the discovery of CB selective drugs.

keywords

G-protein-coupled receptor, cannabinoid receptors, allosteric binding site, molecular dynamics, Markov state model, VAMPnets subtype selectivity, Relative binding free energy calculation, anandamide

Introduction

The discovery of cannabinoid receptors (CBRs) busted the myth of "membrane-mediated signaling" of lipid-like cannabinoid molecules.[1] Two types of cannabinoid receptors have been confirmed: Cannabinoid receptor 1 (CB_1) and 2 (CB_2).[2] CB_1 is majorly expressed in central nervous systems, whereas CB_2 is primarily expressed in the immune system.[3,4] The discoveries of CBRs established a complex signaling system named the endocannabinoid system (ECS).[5,6] The ECS consists of neurotransmitter endocannabinoids, enzymes that help produce or degrade endocannabinoids, and receptors where the endocannabinoids bind to modulate downstream signaling.[7] This complex system maintains homeostasis in neuronal signaling and has been targeted for different physiological and psychological disorders such as obesity, pain, and inflammation.[8]

Unlike other neurotransmitters, which are stored for future usage, endocannabinoids are produced in demand from the membranes of postsynaptic neurons due to an increase in calcium ion concentration.[9] These molecules travel to the presynaptic neuron via a retrograde signaling system to bind CBRs.[10] Binding of endocannabinoid leads to downstream signaling by G-protein and β-arrestin.[11–13] Enzymes that are part of ECS finally degrade these molecules in presynaptic neurons.[14]

Two well-known endocannabinoids are anandamide (N-Arachidonoylethanolamine) and 2-AG (2-arachidonoylglycerol) (Figure 1A).[17,18] These molecules have four pharmacophore groups: a polar head group, a propyl linker, a polyene linker, and an alkyl tail. Although both molecules are precursors of arachidonic acid, they are produced via distinct enzymatic mechanisms.[19,20] Anandamide is a partial agonist for both CB_1 and CB_2, whereas 2-AG is a full agonist for both receptors.[21] As full agonists of CBRs are associated with higher side effects, anandamide has gained more attention in drug design due to its partial agonistic properties like Δ^9-THC.[22–24]

GPCRs belonging to the same subfamily share high sequence and structural similarities.[25] Consequently, most designed drugs bind to all receptors from the same subfamily, causing
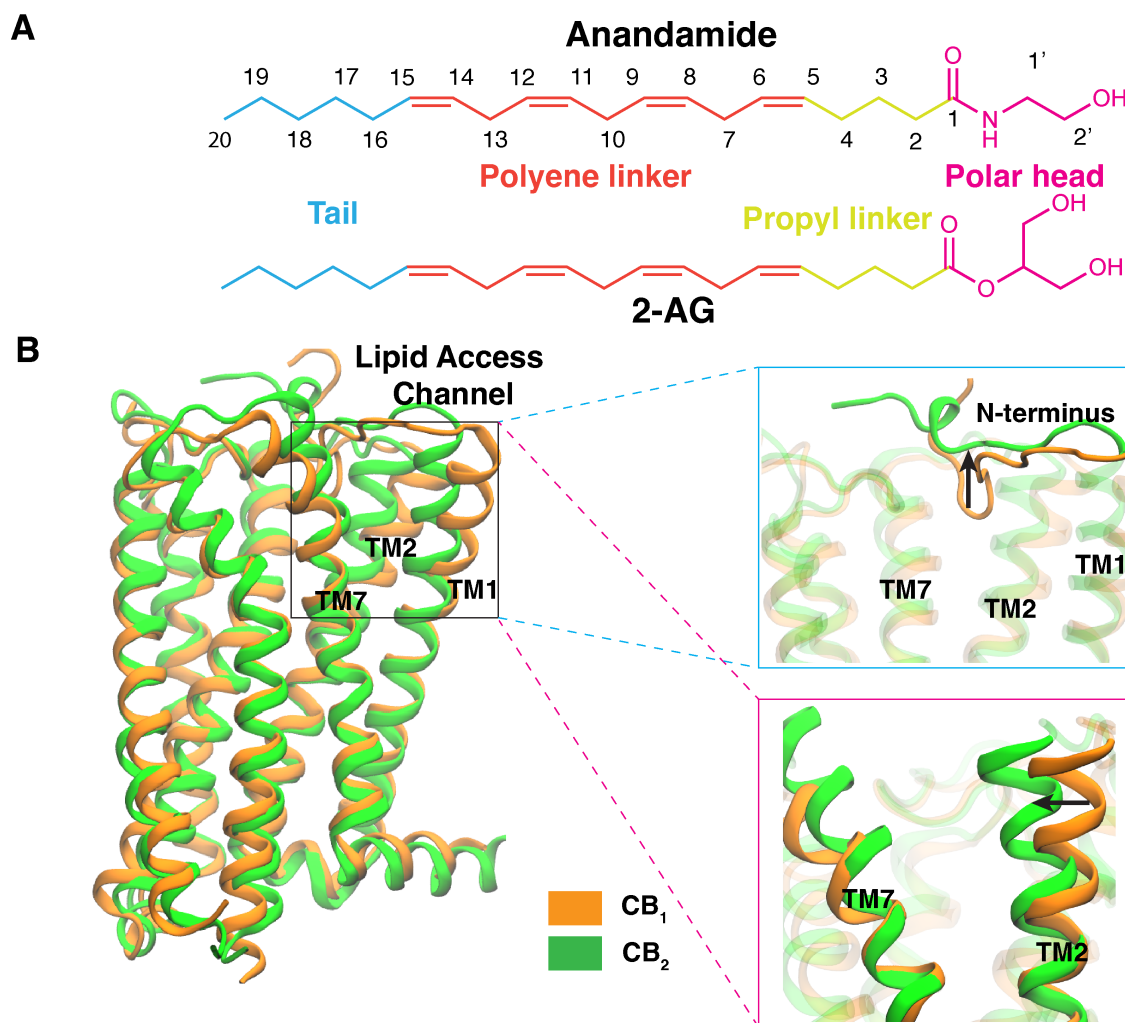
Figure 1: (A) Structures of the two well-known endocannabinoids: anandamide and 2-AG. Four pharmacophore groups of each endocannabinoid are represented in different colors. The carbon atom numbering of anandamide is shown. (B) Superposition of inactive CB_1 (PDB: 5TGZ;[15] color: orange) and CB_2 (PDB: 5ZTY;[16] color: green) are shown as cartoon representations with structural differences along the lipid access channel highlighted in separate panels.

off-target side effects. CB_1 and CB_2 share a 44% total sequence identity and 68% sequence homology in the transmembrane regions.[26] Previous experiments have shown anandamide to be selective towards CB_1 (K_i: 61-543 nM) as compared to CB_2 (K_i: 279-1940 nM).[18,21] Several anandamide analogs designed to study the structure-activity relationship have been shown to retain or increase ligand selectivity for CB_1.[27–29] Therefore, anandamide and its analogs provide an avenue to design subtype-selective ligands for CB_1, which is an important issue not just for CBRs but for all Class A GPCRs.[30–32]

This study proposes two hypotheses to provide mechanistic explanation of endocannabinoid subtype selectivity. The first hypothesis focuses on the influence of structural changes along the endocannabinoid binding pathway within the receptors. The second hypothesis explores the role of entropy in the binding process.

Previous studies suggested that endocannabinoids, due to their lipidic nature, bind to the CBRs from the membrane-accessible transmembrane regions via lateral diffusion.[33,34] This proposition was supported by the recent discoveries of experimentally determined structures that show a lipid accession channel between TM1 and TM7 for CB_1.[22,35] Additionally, mutagenesis and molecular dynamics (MD) studies on cannabinoid receptors provide further evidence for this ligand binding pathway.[33,36,37] However, comparison of the lipid access channel of CB_1 and CB_2 inactive structures shows major structural differences (Figure 1B). Membrane-proximal N-terminus remains inside the orthosteric pocket for CB_1 inactive structure.[15,38] Agonist binding moves the N-terminus out of the pocket in order to bind in the agonist-like pose.[23,35] Conversely, the downward N-terminus creates a large space between TM2 and TM7 for CB_1 compared to CB_2, which might facilitate the entrance of the agonists in the pocket (Figure 1B). These structural differences might influence endocannabinoid binding mechanism and ligand interactions in bound pose within each receptor. Thus, in our first hypothesis, we propose the distinct interactions formed by the ligand and receptors lead to differential enthalpy contributions to binding affinity making anandamide subtype selective.

The second hypothesis is based on the volumetric differences in cannabinoid receptors' binding pocket.[38] CB_1 and CB_2 binding pockets have been shown to accommodate much bulkier ligands compared to the long-chain endocannabinoids with an abundance of sp^3 hybridized carbon atoms.[15,35] Flexible endocannabinoids may obtain multiple reversible stable poses inside large pocket of CBRs, which increases the importance of entropy contribution in ligand binding. Recent experimental studies have also revealed that entropy plays a major role in determining subtype selectivity of CBRs.[39] Therefore, we propose that distinct pocket

5

volumes of CBRs leads to dissimilar entropy contribution in binding free energy, causing subtype selective behavior of anandamide. Hence, to investigate whether distinct ligand-protein interactions or ligand flexibility effects determine ligand selectivity, we compared the binding of the endocannabinoid anandamide to cannabinoids using atomistic MD simulations.

Extensive $\sim 900\mu s$ of unbiased molecular dynamics were performed for the exploration of binding pathways and interaction sites of anandamide for CB_1 and CB_2. The thermodynamics and kinetics of these processes were estimated using Markov state models (MSM).[40] Deep learning based VAMPnets was used to identify intermediate states along the pathway.[41] Binding simulations reveal that anandamide can diffuse through the membrane to bind to CB_1 and CB_2 from the lipid access channel of TM1 and TM7 confirming previous observations. A significant conformational change in membrane-proximal N-terminus of CB_1 during ligand binding suggests an induced fit binding mechanism. Conversely, anandamide binding to CB_2 does not lead to a major conformational change in the binding pocket, aligning with a conformational selection mechanism. Unbiased MD simulations also detect a non-canonical transmembrane pathway only for CB_2 (between TM5 and TM6), which converges to the orthosteric binding pocket similar to the canonical pathway but appears less favorable for ligand binding. This difference in binding mechanisms contributes to varying interactions within the binding pockets of CB_1 and CB_2. These interaction differences were validated by relative binding free energy (RBFE) calculations for anandamide analogs, which show consistent trends with experimental binding affinity. Further analyses reveal that anandamide interacts strongly with the displaced N-terminus in the orthosteric pocket of CB_1, increasing enthalpy contribution to binding free energy compared to CB_2, where the N-terminus remains outside the pocket. Additionally, larger volume of orthosteric binding pocket of CB_2 allow more flexibility for the anandamide compared to CB_1 in the ligand bound state. This increased flexibility leads to a higher entropy contribution towards ligand binding free energy for CB_2. Comparing these opposing entropy and enthalpy contributions reveals that the enthalpic contributions are dominant and facilitate more selective ligand binding in CB_1.

Overall, this study provides a detailed quantitative evaluation of mechanistic hypothesis for understanding of endocannabinoid subtype selectivity, which will aid in the discovery of selective drugs targeting cannabinoid receptors.

Results and Discussion

Anandamide binds to cannabinoid receptors via the lipid access channels in between transmembrane helices. Approximately ~ 423 and 483 μs of atomistic molecular dynamics simulations were performed for CB_1 and CB_2 to capture the anandamide binding. Realistic membrane compositions were used to accurately estimate thermodynamics and kinetics as the ligand binding happens via the membrane bilayer.[42,43] Average brain membrane composition was used for CB_1 as these receptors are majorly expressed in the central nervous system (Table S1). In contrast, average eukaryotic membrane composition was used for CB_2 as they are expressed in immune cells (Table S2). Anandamide was initially placed in the extracellular region and was observed to diffuse to the membrane bilayer for both CB_1 and CB_2 during the simulation. The higher probability density of the ligand in the membrane bilayer compared to the salt solution is consistent with the lipidic nature of anandamide (Figure S1). The frames where ligand stays in the membrane bilayer were selected to characterize its transmembrane diffusion and binding pathway to the receptors. Markov State Models were used to obtain the unbiased free energy estimates of the ligand binding process. MSM-weighted free energy landscape between x and y components of the center of mass of anandamide reveal a binding pathway for cannabinoid receptors (CB_1 and CB_2) from the lipid access channel between TM1 and TM7 (Figure 2A and 2B). A separate binding pathway was also observed in between TM5 and TM6 for CB_2, which converges to the similar bound pose inside the pocket (Figure 2B).

Characterization of the binding process was done using VAMPnets, which cluster the simulation ensemble into different macrostates. VAMPnets use deep learning architecture
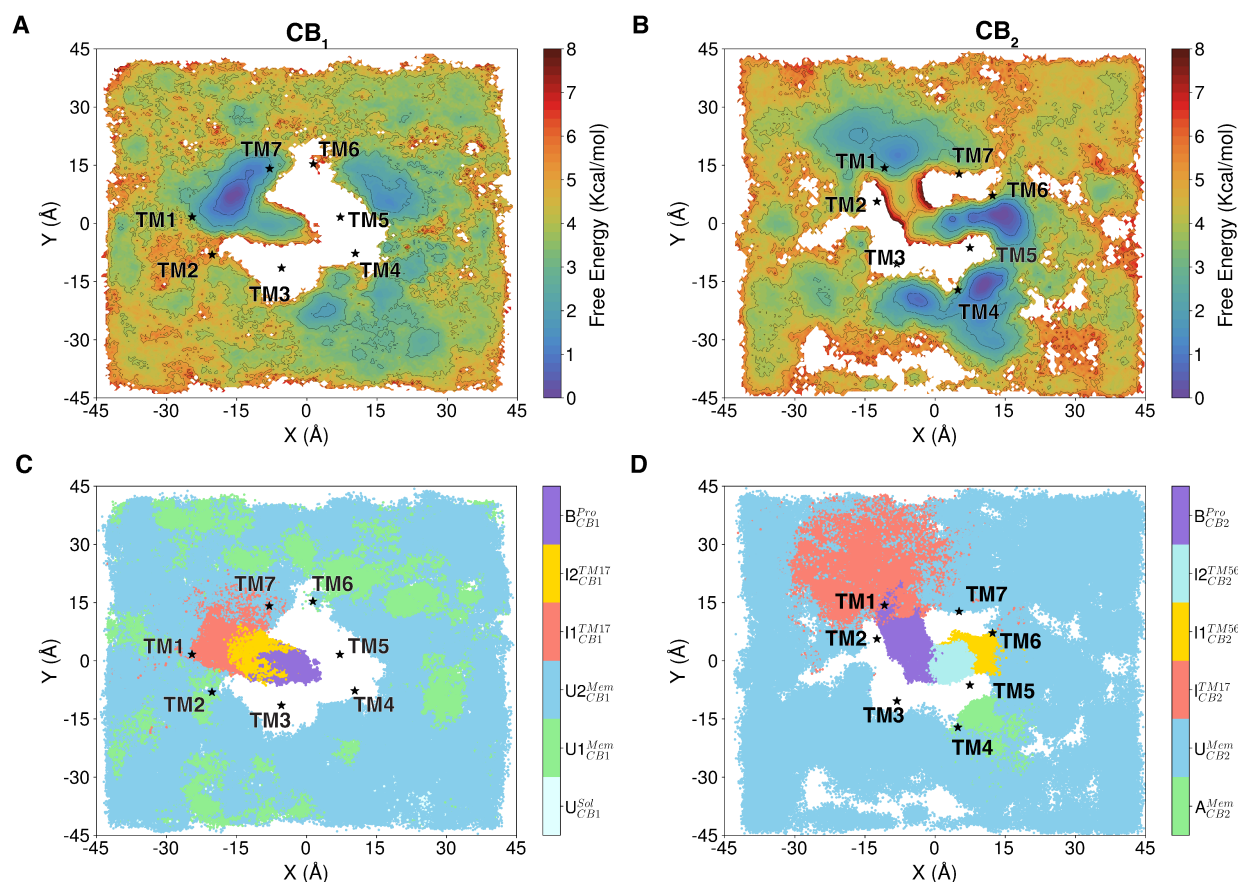
Figure 2: MSM weighted free energy landscapes are projected between x and y coordinates of center of mass of anandamide for CB_1 (A) and CB_2 (B). Macrostates obtained by training the VAMPnets are plotted as scatter plots between x and y coordinates of the center of mass of anandamide for CB_1 (C) and CB_2 (D). Frames from the transmembrane region are only shown in the above plots. The extracellular residue of each helix is shown as a star marker to show its position. Macrostate names A, U, I, and B represent anandmide in allosterically bound, unbound, intermediate and orthosterically bound state in the receptors. Superscript in the intermediate macrostate stands for the ligand binding pathway between transmembrane helices.

to divide the ensemble into macrostates, where intrastate conformational changes are faster compared to interstate transitions.[41] The optimal number of macrostates were selected for CB_1 (six) and CB_2 (six) with the criteria of minimizing the error in implied timescale and state populations (Figures S2, S3, S4, and S5). Relative positions of the macrostates are shown as scatter plots based on the x and y components of the ligand center of mass of frames where anandamide stay inside membrane-bilayer or protein (Figures 2C and 2D). The agonist bound pose in CB_1 and CB_2 are characterized by the macrostate B_{CB1}^{Pro} and B_{CB2}^{Pro}, respectively. Two intermediate states ($I1_{CB1}^{TM17}$, $I2_{CB1}^{TM17}$) mark the anandamide binding

pathway in between TM1 and TM7 for CB_1 before the ligand reach bound state from the bulk membrane. In contrast, one intermediate state (I_{CB2}^{TM17}) was observed along the similar pathway for CB_2. An alternate pathway was also observed for CB_2 in between TM5 and TM6 was characterized by two intermediate states ($I1_{CB2}^{TM56}$, $I2_{CB2}^{TM56}$). Detailed description of each pathway and its role in determining selectivity are explained in the subsequent sections.

Induced-fit binding of anandamide to CB_1 via the lipid access channel of TM1 and TM7. Ligand diffuse from the bulk membrane to the metastable minima at the membrane exposed interface of TM1 and TM7 before anandamide binds inside the pocket (Figure 2A). This stable region of the landscape is represented as Macrostate $I1_{CB1}^{TM17}$ (Figure 2C). The timescale for ligand diffusion from the bulk of the membrane was determined using mean free passage time (MFPT) calculation. MFPT was calculated from transition path theory, which estimates the timescale from the transition probability matrix as determined by MSM. MFPT calculation shows that the ligand can diffuse from bulk to macrostate $I1_{CB1}^{TM17}$ within a sub-microsecond timescale ($0.6 \pm 0.1\mu s$).

In this metastable macrostate $I1_{CB1}^{TM17}$, the ligand is stabilized in the protein-membrane surface while maintaining contact with the lipid molecules (Figures 3A and 3B). Protein and lipid exposed surface area calculation of anandamide in macrostate $I1_{CB1}^{TM17}$ shows that more than 50% of average area maintains contact with protein, depicting a stable interaction between the ligand and protein (Figure S6). The protein-ligand contact probabilities and interaction energies were calculated to identify the residues important for anandamide stabilization at the receptor surface (Figure 4A). These calculations reveal that the ligand form major interactions with N-terminus ($F108^{N-term}$, $M109^{N-term}$) and TM7 ($F381^{7.37}$, $M384^{7.40}$) residues (Figure 3C). The most frequent and stable interaction is the polar hydrogen bond interaction between the nitrogen group of anandamide and backbone oxygen of the N-terminus $F108^{N-term}$ residue. Contact probability calculations were also performed to capture major interactions of anandamide pharmacophore groups with different lipids (Figure 3B). The lipids in the membrane were classified based on the eight different head

groups. The analysis results reveal that major interactions with the ligand are observed with phosphatidylcholine (PC), sphingolipids (SM), and cholesterol (CHL). Furthermore, stable lipid contacts are majorly formed with the polyene linker and polar headgroups, while the propyl chain and hydrophobic tail form weaker interactions (Figure 3B). Due to the stable interactions, the probability density plot for ligand RMSD reveals only one major conformation of anandamide, where the ligand is parallel to the protein helices (Figures 3C and S7).

From macrostate $I1_{CB1}^{TM17}$, anandamide moves inside the binding pocket between the space formed by TM1, TM2, and TM7. Conformational ensemble of anandamide in this position was represented by macrostate $I2_{CB1}^{TM17}$. The binding position of anandamide in this macrostate is similar to the experimentally determined antagonist-bound pose (Figure S8). Contact probability and interaction energy calculations reveal that major ligand-protein interactions are formed between the polar group of the ligand with TM2 (F174$^{2.61}$, F177$^{2.64}$, and H178$^{2.65}$) and N-terminus (M109^{N-term}) of the protein (Figures 3A and 3C). Previous experiments have shown that these interacting residues in TM2 have major effects on the ligand binding and potency for CB$_1$.[15,44] Few lipid interactions are also present in the macrostate $I2_{CB1}^{TM17}$ (Figure 3B). The polar part of the ligand still maintains contact with cholesterol (CHL) and sphingolipids (SM). Although macrostate $I2_{CB1}^{TM17}$ has more protein interactions than macrostate $I1_{CB1}^{TM17}$, it was observed from the free energy landscape that macrostate $I2_{CB1}^{TM17}$ is comparatively less stable than macrostate $I1_{CB1}^{TM17}$ (Figure 2A). This observation shows the importance of lipid interactions in stabilizing anandamide in the protein interface. As the ligand moves from a low to high energy macrostate, kinetically, this transition is slower ($29.0 \pm 3.8 \mu s$) compared to the reverse transition ($0.5 \pm 0.1 \mu s$).

The binding of the ligand to the protein leads to the conformational change. K-L divergence analysis was performed between the different macrostates to highlight the regions of the protein conformational differences. Comparing the K-L divergence between these macrostates, a major conformational change was observed in the extracellular TM1, TM2,

Figure 3: (A) Contact probability (color: Blue) and interaction energy (color: Orange) calculations between anandamide and CB_1 in different macrostates in the binding pathway. Residues that maintain at least 50% contact probabilities in one of the macrostates are shown in the figure. (B) Lipid contact probabilities in different macrostates of CB_1 with anandamide are shown as a heatmap. (C) A representative frame from each of the macrostates is shown. CB_1 (color: Orange) is shown as a cartoon representation with transparent TM6 and TM7. Anandamide (color: Purple) and the four highest interacting residues (color: Cyan) are represented as sticks. (D) K-L divergences between two macrostates are shown in color and thickness gradients. Thickness gradients are shown as moving averages.

11

and N-terminus, confirming that the ligand binding is leading local conformational changes in the binding pocket (Figure 3D). Within the local conformational changes, specific movements are found to facilitate the ligand binding: upward movement of N-terminus and F174$^{2.61}$. To capture these changes, the N-terminus (M103^{N-term}) distance from D163$^{2.50}$ was plotted against the sidechain dihedral (χ_1) of F174$^{2.61}$ as MSM weighted free energy landscape (Figures S9A and S9B). We observed that macrostate $I1_{CB1}^{TM17}$ to macrostate $I2_{CB1}^{TM17}$ transition leads to the shift in the population distribution of the χ_1 angle of F174$^{2.61}$. The binding of the macrostate $I2_{CB1}^{TM17}$ also partially shifts the N-terminus in the extracellular direction to accommodate the ligand (Figure S9A).

From macrostate $I2_{CB1}^{TM17}$, anandamide moves further deep in the pocket to macrostate B_{CB1}^{Pro}, which represents the agonist binding pocket (Figure 3C). The major interactions are still maintained between the polar group of anandamide and TM2 in this macrostate (Figure 3A). K-L divergence calculation between these macrostates also shows the major conformational change associated with the N-terminus (Figure 3D). MSM weighted free energy landscape also points out the population shift of N-terminus away from the binding pocket (Figures S9A and S9B). This required large conformational change of the N-terminus in the binding pocket makes transition from $I2_{CB1}^{TM17}$ to B_{CB1}^{Pro} order of magnitude slower compared to previous transitions along this pathway.

These significant conformational changes in the N-terminus alter the shape of the binding pocket. Further analyses reveal changes in the pocket volume during ligand binding (Figure S10). These observations suggest that anandamide binds to CB$_1$ via an induced fit mechanism, where the binding of anandamide induces a conformational change in the N-terminus.

Conformational selection mechanism dominates anandamide binding to CB$_2$ via TM1-TM7 lipid access channel. Mutagenesis experiment and MD simulation have also shown similar ligand binding pathway in between TM1 and TM7 for CB$_2$. However, the N-terminus of CB$_2$ remain above the orthosteric pocket and far from this pathway.
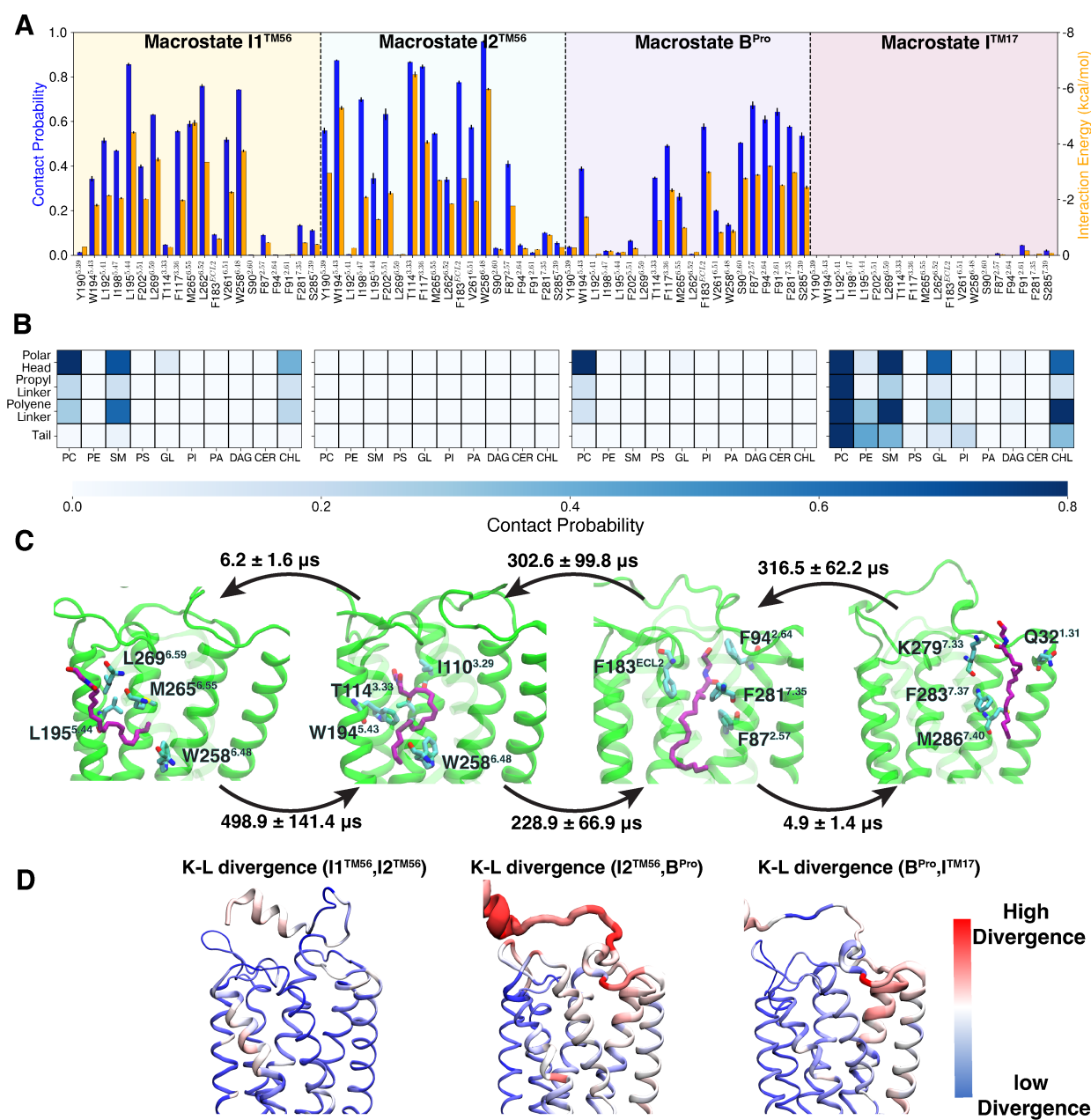
Figure 4: (A) Contact probability (color: Blue) and interaction energy (color: Orange) calculations between anandamide and CB_1 in different macrostates in the binding pathway. Residues which maintain atleast 50% contact probabilities in one of the macrostates are shown in the figure. (B) Lipid contact probabilities in different macrostates of CB_1 with anandamide are shown as a heatmap. (C) Representative frame from the each of the macrostate is shown. CB_1 (color: Green) are shown as a cartoon representation with transparent TM6 and TM7. Anandamide (color: Purple) and four highest interacting residues (color: Cyan) are represented as sticks. (D) K-L divergences between two macrostates are shown color and thickness gradients. Thickness gradients are shown as moving average.

Simulations show this structural distinction causes major difference in the pathway. VAMP-nets analysis captures one intermediate macrostate I^{TM17}_{CB2} near the protein surface between TM1 and TM7 before anandamide binds to agonist-like poses in the orthosteric pocket. Spontaneous transition to this macrostate from membrane-bilayer was observed within the microsecond timescale ($8.7 \pm 1.2\mu s$). Contrasting to CB$_1$, this macrostate does not represent a metastable minima in the protein surface between TM1 and TM7, which depicts less stable interactions with the protein surface (Figure 2B). The less stability of the ligand in this receptor surface can be explained by the lack of stable contact in the surface (Figures 4A). A lesser ratio of protein and lipid accessible surface area of anandamide compared to the Macrostate $I1^{TM17}_{CB1}$ further validates this phenomena (Figure S6). Major interaction differences between CB$_1$ and CB$_2$ for anandamide was observed at the N-terminus. Due to lack of N-terminus interaction, highest protein-ligand interactions in CB$_2$ macrostate I^{TM17}_{CB2} are observed in TM1 (Q32$^{1.31}$) and TM7 (K279$^{7.33}$, F283$^{7.37}$, M286$^{7.40}$). The anandamide also forms interactions with transmembrane lipids (phosphatidylcholine (PC), sphingolipids (SM), and cholesterol (CHL)) similar to the CB$_1$ (Figures 4B).

Anandamide moves to the bound pose B^{Pro}_{CB2} from macrostate I^{TM17}_{CB2} (Figure 4C). In the macrostate, the major interactions with anandamide were observed with aromatic residues in TM2 (F91$^{2.61}$, F94$^{2.64}$), TM7 (F281$^{7.35}$) and ECL2 (F183^{ECL2}) (Figure 4C). K-L divergence analysis shows that this binding from macrostate I^{TM17}_{CB2} to B^{Pro}_{CB2} leads to the higher residue fluctuation in the N-terminus, extracellular TM1, and TM2 as ligand binds from the region (Figure 4D). Comparison of inactive structure of CB$_1$ and CB$_2$ points out less space between the TM2 and TM7 for CB$_2$ along this pathway. To characterize major conformational changes in anandamide binding, F91$^{2.61}$ χ_1 dihedral angle and TM2 (F91$^{2.61}$) - TM7 (S285$^{7.39}$) distance were plotted against each other (Figure S11A). MSM weighted free energy landscape reveal that F91$^{2.61}$ χ_1 angle ensemble average does not change as the ligand moves from macrostate I^{TM17}_{CB2} to B^{Pro}_{CB2}. However, the TM2 and TM7 distance increases to facilitate the anandamide binding (Figures S11A and S11B). Comparison of similar distances for CB$_1$

shows that TM2 and TM7 remain far away to be affected by ligand binding (Figure S12). Angle rotation of $F174^{2.61}$ is enough for anandamide to move inside the pocket (Figure S12). As the helical movement is needed for CB_2, the kinetic transition from macrostate I_{CB2}^{TM17} to B_{CB2}^{Pro} is much slower ($300.2 \pm 99.8 \mu s$) compared to the equivalent transition in CB_1 (macrostate $I1_{CB1}^{TM17}$ to $I2_{CB1}^{TM17}$: $29.0 \pm 3.8 \mu s$) (Figures 3C and 4C).

These analyses highlight the differences in the binding mechanisms of anandamide to CB_1 and CB_2 despite of binding from same the lipid access channel between TM1 and TM7. In CB_1, the binding pocket shape changes during ligand binding due to the repositioning of the N-terminus, indicating an induced fit mechanism. Conversely, in CB_2, the shape of binding pocket remains largely unchanged, suggesting a conformation selection mechanism. This demonstrates that endocannabinoids can utilize different binding mechanisms even when interacting with receptors from the same subfamily.

Alternative anandamide binding pathway for CB_2 in between TM5 and TM6. Another membrane-embedded ligand-binding pathway was detected for CB_2 between TM5 and TM6. VAMPnets analysis shows two macrostates along this pathway before it converges back to the bound state (Figure 4C). In the macrostate closer to the membrane bilayer, the ligand is stabilized between the surface of TM5 and TM6. Diffusion from the membrane bilayer to $I1_{CB2}^{TM56}$ is also relatively fast ($3.4 \pm 0.4 \mu s$). This macrostate also represents a metastable region as the ligand is stabilized by the residues in TM5 ($L195^{5.44}$) and TM6 ($W258^{6.48}$, $M265^{6.55}$, $L259^{6.59}$) surface (Figures 4A and 4C). In previous experimentally determined structures, long chain lipid molecules were found to be bound to TM5 and TM6 surface, supporting our observation[16,45] (Figure S13). The ratio of the protein and lipid surface area calculations indicates stronger interactions with CB_2 surface as more than 80% of the ensemble averaged surface area embedded to the protein (Figure S6). Lipid interactions in this macrostate are limited to the polar head group of the anandamide as the hydrophobic tail remains inside the receptor (Figure 4B). K-L divergence analysis compared to macrostate $I1_{CB2}^{TM56}$ shows a higher divergence in TM5 and TM6 surface to accommodate

15

the tail of anandamide (Figure S14). Further analysis reveal that the pocket volume between the TM5 and TM6 interface of the CB_2 increases as the ligand diffuses from the membrane to this intersurface (Figure S15).

Anandamide moves inside the receptor to macrostate $I2_{CB2}^{TM56}$ from macrostate $I1_{CB2}^{TM56}$. Lack of lipid interactions confirms the ligand position inside the protein in this macrostate (Figure 4B). However, the ligand pose is different than compared to macrostate B_{CB2}^{Pro}. The major protein-ligand interactions were observed between the TM3 ($I110^{3.29}$, $T114^{3.33}$), TM5 ($W194^{5.43}$), TM6 ($W258^{6.48}$) (Figures 4A and 4C). Due to the stable interactions, RMSD calculations show that the ligand stabilizes in one stable pose in this macrostate (Figures 4C and S16). K-L divergence analysis shows that the major fluctuation happens in the TM6 due to the transition between macrostate $I1_{CB2}^{TM56}$ to $I2_{CB2}^{TM56}$ (Figure 4D). Major dynamic change was observed in the angle distribution of $F202^{5.51}$ rotational movement to create space between TM3, TM5, and TM6 for the ligand. Anandamide bound in this macrostate increases the population of $F202^{5.51}$ in the outward facing direction (Figure S17). Interstate transition further in the pocket between macrostate $I2_{CB2}^{TM56}$ and B_{CB2}^{Pro} was also observed. Although anandamide binding is possible from both the lipid access channels for CB_2, kinetic comparison between the pathways shows that pathway between TM1 and TM7 is kinetically more favorable (Figure 4C).

MD simulations were able to capture anandamide TM5 and TM6 for CB_1, however, no metastable minima or ligand binding pathway were observed from this lipid access channel (Figure 2A). We explained this difference between CB_1 and CB_2 by calculating the lipid accessible surface area for the membrane-facing residues in the TM5 and TM6 surfaces (Figures S18A). Higher accessible surface area would increase the contact probability between anandamide and receptor to stabilize the ligand on the surface. For CB_2, area per residue is much higher compared to the CB_1 (Figure S18B). Less per residue surface area for CB_1 can be attributed to bulky $F368^{6.60}$ residue in the protein surface instead of the $A270^{6.60}$ in CB_2 (Figure S18A). This decreasing surface area implies that the anandamide cannot reach the

groove between TM5 and TM6, decreasing stable interactions between the ligand and CB_1.

Distinct binding pocket interactions contributes to anandamide selectivity.
Difference in the binding mechanism leads to dissimilar binding poses and interactions of anandamide inside the pocket. Recently, cryo-EM structure of anandamide analog (AMG315) bound CB_1 was determined in the active state with downstream G-protein.[46] AMG315 is functionally a full agonist and is observed in the canonical agonist-bound pose in the cyro-EM structure. Comparison of AMG315 bound structure with MD sampled structures of the bound poses shows that MD simulations are able to capture the experimentally determined ligand pose for both CB_1 and CB_2 (Figures S19A and S19B). However, distance RMSD calculations of anandamide in bound poses show that the canonical agonist bound conformation is not the most stable conformation for anandamide in the pocket (Figures S19C and S19D). We performed relative binding free energy (RBFE) calculation with respect to anandamide analogs with known experimental binding affinity to computationally validate these obtained distinct poses. One major drawback of this method of validation is that RBFE calculation depends on the reference ligand's pose. However, anandamide shows significant flexibility inside the pocket for both receptors. There is high enough probability that ligand can obtain other poses in the receptor, which can affect the calculated $\Delta\Delta G$. Hence, instead of the quantitative agreement, we focus on the qualitative trend of $\Delta\Delta G$ with respect to most stable pose of anandamide.

The analogs that we are studying are the ones where scaffold changes have been focused on the polar group of the anandamide, as polar groups form the most stable interactions in the pocket (Figure S20A). The selected analogs were classified into three categories to better understand the effect of scaffold change: (1) addition of alkyl groups in C1' and C2'. It has been shown that adding an alkyl group in this position makes the ligand more resistant to hydrolysis by degrading enzymes, increasing its apparent binding affinity.[47] (2) removal of terminal hydroxyl group. Previous studies have shown that the hydroxyl group is not essential for binding.[27] (3) substitution of hydroxyl group with fluorine group. These

17

analogs increase the CB_1 and CB_2 affinity.[27] Estimated $\Delta\Delta G$ reveals that binding affinity increases for CB_1 and CB_2 for all three categories of the analogs and magnitudes of the $\Delta\Delta G$ also remain within the realistic limits (Figure S20B). Considering the assumption of RBFE calculations, these calculations validate the binding pose obtained from the simulations.

These distinct binding poses and differences in the receptors' conformation due to ligand binding lead to difference of ligand interaction with receptors. To observe these differences, we compared the interactions of anandamide with equivalent loops and helices surrounding orthosteric pocket of receptors. Interaction energies were calculated using LIE method for 1000 representative structures from bound state of each receptor, which were selected based on their MSM probabilities. Ligand interaction energy calculations in these structures show major differences in N-terminus region of the two receptors (Figure 5A). In spite of extracellular motion, stronger interactions with N-terminus hydrophobic residues were observed for CB_1 compared to CB_2. Furthermore, overall comparison of enthalpy contribution due to protein and ligand interactions indicate higher enthalpy contribution for CB_1. This makes the anandamide binding more enthalpically favorable to CB_1.

Explanation of CB_1 selectivity of anandamide. Although there are conflicting reports on the binding affinities of anandamide, the general consensus is that anandamide is more selective towards CB_1 compared to CB_2.[18] We showed that the ligand binding is enthalpically favorable to CB_1. However, significant ligand dynamics inside the pocket due to large pocket volume of receptors and plethora of sp^3 hybridized carbon atoms may lead to a significant entropy effect in the binding free energy. (Figures S19C and S19D).

To compare the anandamide dynamics inside CB_1 and CB_2 pocket, we calculated angle between carbon atoms at three different parts of ligand: tail (C16), polyene linker (C11), and propyl linker (C3) group (Figure 5B). This analysis shows that anandamide is stabilized in one conformation for majority of the population in B_{CB1}^{Pro}. In contrast, the ligand has much higher divergence in angle distribution for CB_2 with multiple equally probable angular conformations (Figure 5B). This higher conformational divergence in macrostate B_{CB2}^{Pro} can
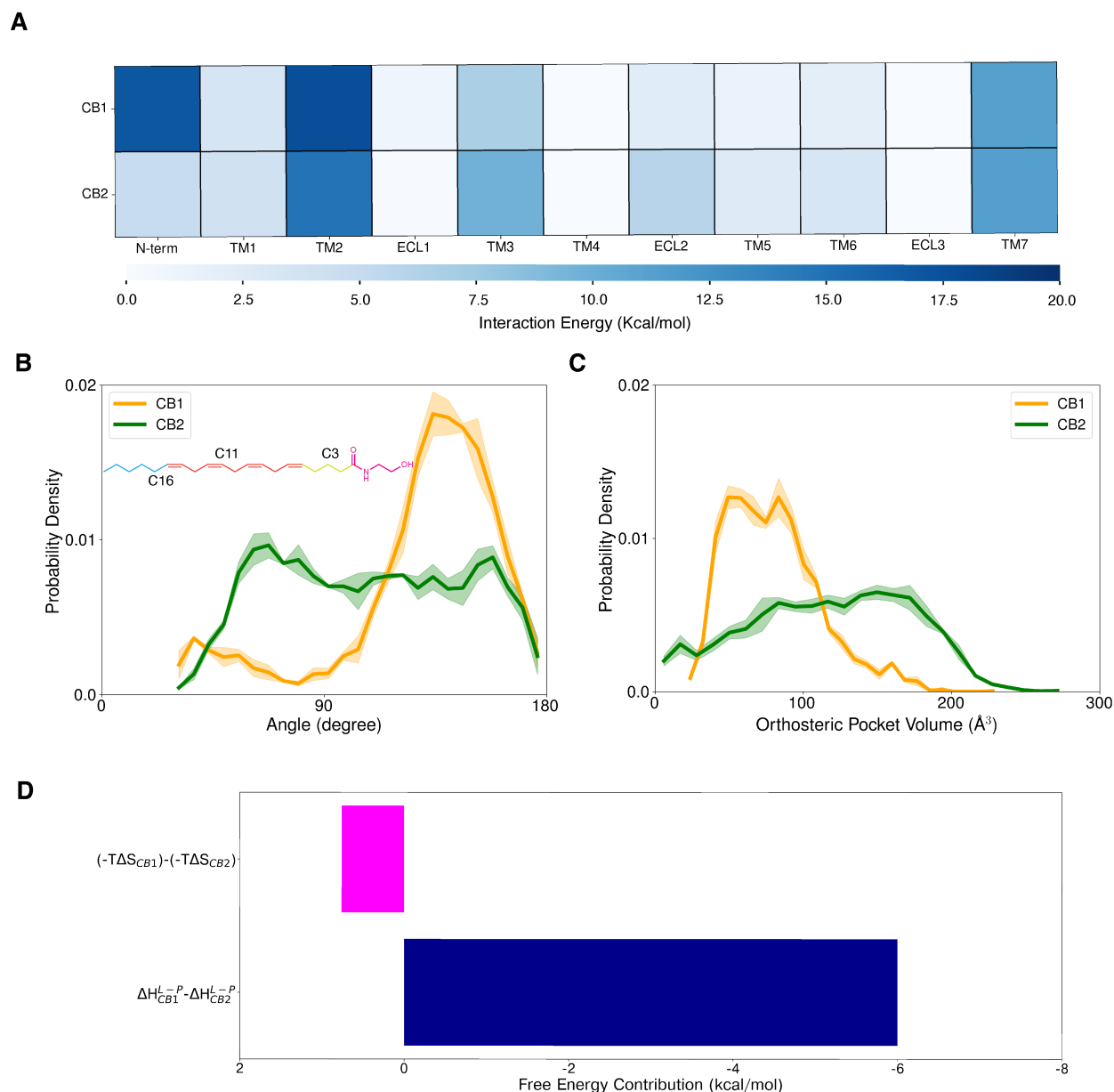
Figure 5: (A) Anandamide interaction energies of different structural elements of cannabinoid receptors in the bound states. (B) Probability density plot of three sp^3 carbons of anandamide representing hydrophobic tail (C16), polyene linker (C11), and propyl linker (C3) for B^{Pro}_{CB1} (color: orange) and B^{Pro}_{CB2} (color: green). (C) Probability density plot of binding volume calculation of orthosteric binding pocket for B^{Pro}_{CB1} (color: orange) and B^{Pro}_{CB2} (color: green). Error in angle and pocket volume distributions were calculated from on 3 bootstrap samples where each sample contains 1000 frames obtained from macrostate based on MSM weighted probability. (D) Bar plot represents the differences of enthalpy (color: blue) and entropy (color: magenta) contribution between the B^{Pro}_{CB1} and B^{Pro}_{CB2}.

be explained by the binding pocket volume calculations (Figure 5C). These calculations show that anandamide has larger accessible volume to move in CB_2 compared to the CB_1, which

19

matches with our previous observations.[38] Therefore, in CB_1, degrees of freedom for the movement of sp^3 hybridized carbons are restricted compared to CB_2. The higher fluctuation in ligand due to larger volume in CB_2 leads to higher entropic contribution in free energy (Figure 5D). Estimation of Gibbs entropy ($\Delta S = -\sum_i k_B p_i ln(p_i)$) show higher entropic contribution for anandamide binding in CB_2, which contradicts with the higher enthalpic effects in CB_1. Comparison of these contrasting effects of enthalpy and entropy between two receptors, enthalpy contribution to free energy dominates in the ligand binding process (Figure 5D). Hence, anandamide binds favorably to CB_1 than CB_2.

Conclusions

Endocannabinoids are lipid-like long-chain molecules that bind to cannabinoid receptors to transduce intracellular signaling. One of the endocannabinoids, anandamide, has garnered significant attention due to its partial agonistic nature and CB_1 selectivity. Designing analogs of anandamide as CB_1 selective drugs necessitates a deep understanding of its binding mechanisms. In this study, we developed two hypotheses to elucidate the CB_1 selectivity of anandamide.

First, we proposed that structural differences along the ligand-binding pathway lead to distinct bound poses and the resulting enthalpy contributions to the ligand-binding free energy. Second, we suggested that the distinct pocket volumes of CB_1 and CB_2 lead to differences in ligand dynamics, influencing the entropy contributions to free energy. To test these hypotheses, we conducted unbiased extensive MD simulations of anandamide binding to CB_1 and CB_2.

Our MD simulations reveal mechanistic differences in anandamide binding between CB_1 and CB_2 via the pathway between TM1 and TM7. Binding to CB_1 is characterized by an induced fit mechanism, where the ligand displaces the N-terminus from the orthosteric pocket. In contrast, CB_2 binding suggests a conformation selection mechanism, with no

major changes observed. For CB_2, a separate non-canonical binding pathway was observed in between TM5 and TM6. Although this pathway finally converges to the same orthosteric binding pocket, it is kinetically unfavorable compared to the canonical pathway between TM1 and TM7. These distinct binding mechanisms result in different ligand binding poses and interactions with the receptors. Major differences in ligand interactions are observed between the N-terminus of the receptors. Higher interaction with N-terminus for CB_1 makes the ligand enthalpically more stable in CB_1 binding pocket, which supports our first hypothesis. Additionally, we observed that the larger pocket volume of CB_2 allows greater ligand dynamics, leading to a more favorable entropy contribution to free energy for CB_2. Comparison of this relatively contradicting enthalpy and entropy effect on binding free energy reveal that the dominant enthalpy effect causes the ligand to be more selective towards CB_1.

In summary, our work elucidates the orthosteric selectivity of anandamide for cannabinoid receptors using extensive molecular simulations and post-processing Markovian techniques. These insights provide crucial information for designing selective drugs.

Methods

System Preparation

Inactive structures of CB_1 (PDB ID: 5TGZ[15]) and CB_2 (PDB ID: 5ZTY[16]) were selected to perform anandamide binding simulations. Non-protein residues (ligands and water molecules) and fused proteins (Flavodoxin and T4-lysozyme) were removed for the structure file. Thermostabilized mutations were modified back to the original residues. Missing residues in ICL3 were modeled using Rosetta loop modeling for CB_1 and CB_2.[48,49] For CB_2, missing 21 residues in the N-terminus were also modeled using Robetta webserver.[50] Ninety-eight residues were missing from the CB_1 N-terminus, and no template is available to model from homologous protein.[51] Furthermore, it is challenging to model 98 residues accurately with template-free modeling because of the combinatorial expansions of conformational space.

However, the membrane-proximal regions of the N-terminus were shown to be important for CB_1 signaling by allosterically modulating ligand affinity.[52] Hence, the closest ten residues (89 to 98) were modeled using Rosetta with a disulfide bond constraint between $C98^{N-term}$ and $C107^{N-term}$ residues.[52,53] The remodeled structure with the least energy was further refined in each step using kinematic closure protocol.[54]

Proteins with the modeled loops were embedded in the membrane bilayer and salt solution (extracellular and intracellular region) using CHARMM-GUI.[55] Asymmetric complex membrane compositions were selected for CB_1 and CB_2. Average brain membrane composition was used for CB_1 as it is majorly expressed in the central nervous system. Average mammalian cell composition was used for CB_2 as it is majorly expressed in immune cells. The membrane compositions were obtained from Ingólfsson et al. and proportionally downsized according to our system.[56] The final membrane composition is shown in Tables S1 and S2. 150 mM NaCl salt solution with TIP3P water model is used in the extracellular and intracellular regions.[57] CHARMM36m forcefield was used to parameterize the system.[58]

Additional steps were implemented to include anandamide in MD systems. Anandamide was parameterized with CGenFF using the "Ligand Reader & Modeler" of CHARMM-GUI.[59] PACKMOL v18.169 software package randomly placed the anandamide in the extracellular solution.[60] Ligand added systems were rebuilt using the psfgen v2.0 module of VMD.[61] ParmEd v3.2.0 software package was used to perform hydrogen mass repartitioning on the systems and convert the systems into AMBER-recognized format.[62–64]

Simulation Details

Minimization, equilibration, and initial round of production runs were performed in AMBER v18 simulation package.[65] Later rounds of production runs were run on openMM v7.7 using the distributed computing platform folding@home.[66,67] Systems were minimized for a total of 14000 steps with an initial 5000 steps using the steepest descent algorithm and the remaining 9000 steps with conjugate gradient algorithm.[68] Temperature was increased

to 300K in NVT ensemble using Langevin dynamics as thermostat.[69] Collision frequency for Langevin dynamics was used as two ps^{-1}. NVT ensemble simulations were performed for one ns. In the next step, the pressure was set to 1 bar in the NPT ensemble using a Monte Carlo barostat with anisotropic pressure control.[70] This step was performed for two ns. In NVT and the initial stage of NPT ensemble simulation, the protein backbone atoms were restrained using a harmonic force with spring constant 10 kcal/mol-Å^2. These steps were run using the AMBER v18 module sander. The restraints were removed in the final equilibration step, and systems were equilibrated for 50 ns in the NPT ensemble before the production run. Production runs were also continued in the NPT ensemble. During the simulation, a cut-off distance of 12 Å was used for non-bonded force calculations. Particle Mesh Ewald method was used for long-range electrostatic calculations.[71] SHAKE algorithm was used to constraint bond lengths for bonds with hydrogen atoms.[72] As HMR was used to perform simulations, four fs simulation step was chosen.[63,64] Periodic boundary conditions were maintained throughout the simulation. All the simulation parameters were kept the same during AMBER v18 and openMM v7.7 production runs.

MD sampling

As GPCR orthosteric pockets are buried inside the transmembrane domain, ligand binding often involves stable intermediate states.[23,73–76] Therefore, long timescale MD simulations are not efficient for capturing ligand binding as it requires significant time to sample the entire process. Here, we employed adaptive sampling,[77,78,78–86] a exploitative and explorative iterative approach for sampling the ligand binding by running parallel short simulations. Adaptive sampling has been successfully employed to investigate the dynamics of mmebrane proteins including GPCRs.[23,38,75,87–92] Initial rounds of sampling were exploitative, where we ran simulations from the frames with ligand positions closest to the binding pocket. The distance of the ligand center of mass from the protein binding pocket center of mass was estimated along the y and z projection, and the closest 15 to 20 frames were selected to

run the next round of sampling. Residues to calculate the center of mass of protein binding pocket for CB_1 and CB_2 are shown in Figure S21. Each trajectory was run for 80 ns. 28 and 25 rounds of iterative simulations were run for CB_1 and CB_2 to observe the complete binding of the ligands.

As MSM building requires sampling of reversible local transitions during the binding procedure, further sampling was done from the entire simulation space using folding@home distributive computing.[66] The entire simulation space was clustered into 1000 clusters based on x, y, and z projection of the center of mass (COM) distance from the ligand. One frame was picked randomly from each cluster, and four independent 100 ns atomistic unbiased simulations were run using folding@home distributive computing.

Finally, explorative sampling was performed to sample the ligand diffusion in the membrane bilayer. Simulations were started from the frames where ligand exists in the membrane. For each round of simulations, these frames are clustered into 20 clusters based on x, y, and z projection of the center of mass (COM) distance from the ligand, and one frame is selected randomly from each cluster. From the selected frame, four independent simulations were started using folding@home. Simulations were continued until the membrane adjacent area of the protein was appropriately sampled.

Featurization of Protein Conformational Ensemble

To build Markov state model (Discussed in the next subsection), each conformation of the biomolecular simulations needs to be featurized to capture the important dynamics. Here, ligand binding and resulting protein conformational changes are the process of interest. To featurize the ligand binding dynamics, the distance from the hydrophobic tail's terminal carbon (C20) and anandamide's carboxyl oxygen to every binding pocket residue (Cα) was calculated. For capturing protein movements, we considered the pairwise residue distances of residues that undergo significant conformational change during activation as measured by the residue-residue contact scores (RRCS).[93] The entire process of residue pair selection

using RRCS is discussed in Dutta and Shukla.[38]

Markov State Models

As short trajectories were run to observe ligand binding, each trajectory can only sample the local transitions. Markov state model (MSM) is a technique that connects the local transitions observed in short trajectories to provide global information about the dynamics such as free energy barriers and kinetics of transitions.[94–98]

In MSM, the MD simulations are discretized into microstates. It is assumed these states would follow Markovian dynamics where the jump between states would only depend on the current state, not the past state.[99] A concept of lag time was introduced to maintain this Markovian property between interstate transitions. Lag time is the minimum time when the interstate transitions become memoryless. Lag time (τ) for CB_1 and CB_2 ligand binding systems were considered 35 ns (Figures S22A and S22B).

According to the assumption of Markovian property, all conformations in a macrostate have to be kinetically similar. Therefore, before discretizing the simulation spaces, featurized MD frames were converted into time-independent components (tiCs) using the variational approach of MSM.[100,101] tiCs are the linear combinations of slow features. Here, tICs were built with a lag time of 25 ns. tiCs were further clustered, and transitional probabilities (p_{ij}) between the clusters were calculated using maximum likelihood estimation.

To build an optimized model, hyperparameters such as the number of tIC components and microstates were optimized by maximizing the VAMP-2 score.[102] VAMP-2 scores are a measure to represent the quality of the slowest processes captured by Markov state models (Figures S22C and S22D). In this case, the VAMP-2 score was calculated using a 10-fold cross-validation of the MD data. The entire MSM building process was implemented using pyEMMA 2.5.6 python package.[103] Optimized MSM redistributes the population of Markovian states, removing the bias created by running short simulations (Figures S23A and S23B).

25

Trajectory Analysis and Visualization

Feature calculations (e.g., distances, angles) for MSM building and other analyses were performed using MDTraj 1.9.8 python package.[104] Trajectory visualization was performed VMD v1.9.3.[61] Frame selection and other modulations in trajectories were performed using CPPTraj v18.01. All analysis codes were written using written in python programming language. Pocket volume calculations were performed using POVME3.0.[105] Getcontacts software was used for contact probability calculations.[106] Linear interaction energy (LIE) analysis was performed using CPPTraj implementation to calculate the interaction energies between the ligand and protein residues.[107,108]

Metastable State Estimation Using VAMPnets

VAMPnets is a deep learning based technique to capture the slow processes from MD ensemble.[41] Using neural network architecture, VAMPnets tries to find the non-linear transformation of features that can represent slow processes. VAMPnets is trained by maximizing the VAMP-2 score. In this work, VAMPnets were used to discretize the space into macrostates. Macrostate represents a collection of Markovian states that encompasses kinetically close regions of the MD ensemble.[38] To achieve the discretization, we used the softmax activation function at the last layer of the neural network. Softmax function returns a discrete probability distribution, where each value represents a probability of an MD frame belonging to a particular macrostate.

We used features that have been used to build MSM to train VAMPnets. A forward neural network with eight hidden layers was used for training, where number of nodes per layer is 30, 100, 100, 100, 100, 100, 100, 30, respectively. ELU activation was used on the input of every hidden layer. The network was trained for 100 epochs on 80% of total trajectories.

The number of the macrostates for every system was selected based on the least errors in the implied time scale and macrostate population[109] (Figures S2, S3, S4, and S5). Implied

time scale is obtained from the VAMP transformation of the last layer. The VAMPnets analysis was performed with Deeptime 0.4.3 python package.[110]

Transition Path Theory

Transition Path Theory (TPT) is a probabilistic approach to calculate the kinetics from the Markov state model.[111,112] This approach introduces a variable named committer (q_i^+). In a probabilistic transition between an initial and final macrostate, committer (q_i^+) is the probability of an intermediate Markovian state (i) reaching the final state before the initial state. Mathematically, the committer can be calculated from the system of linear equations as shown in Equation 1. From the committer, the mean free passage time between the two macrostates is calculated using Equation 2, where π_i is the stationary density of ith Markovian state and τ is the lagtime of MSM. TPT calculations were performed using PyEMMA v2.5.6.[103] Error estimation in TPT calculations was performed using the bootstrapping approach, where MSM was built using randomly selected 80% of all trajectories.[113]

$$-q_i^+ + \sum_{k \notin B} p_{ik} q_i^+ = \sum_{k \in B} p_{ik} q_k^+ \tag{1}$$

$$MFPT_{AB} = \frac{\tau \sum_{i=1}^{m} (\pi_i (1 - q_i^+))}{\sum_{i \in A} \sum_{k \notin A} \pi_i p_{ik} q_k^+} \tag{2}$$

Kullback-Leibler Divergence

Kullback-Leibler divergence (K-L divergence) is an asymmetric statistical distance measurement between two probability distributions. In this context, K-L divergence was performed between the distance distributions of difference macrostates to compare protein conformational changes observed during the ligand binding process.[23,38,114] For this analysis, a thousand representative conformations from each macrostate were selected based on the probability of the MSM states belonging to the macrostate. All possible pairwise closest-heavy atom

distances (except for the two next neighbors per residue) were calculated for these conformations of each macrostate. K-L divergence was estimated between the inverse of each pair of these distances. Per residue pair, we averaged two measurements where each macrostate is considered a reference to make it symmetric. To estimate per-residue contribution, we averaged K-L divergences for all residue pairs containing that certain residue.

Relative Free Energy Calculation

We selected 11 synthetic analogs of anandamide to perform relative binding free energy calculations on the stable bound pose of anandamide observed in the simulation. These calculations were performed on the GROMACS v2020.7 simulations package with a single topology alchemical free energy approach.[115] System building was performed with the following steps.

- As anandamide and its analogs are flexible, a constrained optimization module of the RDKit was used to generate the coordinates of analogs from isomeric smiles stings similar to the bound pose of anandamide in CB_1 and CB_2. The isomeric smiles were obtained from the PubChem database.[116]

- OpenBabel v2.4.1 software was used to convert the obtained mol file from RDKit to mol2 format and to add hydrogen.[117] The mol2 file was parameterized using the CHARMM-GUI ligand generator module. The ParmEd package was used to create the top and gro files.

- alchemical-setup module was used to create single topology top and gro files of the analogs by mapping the similar atoms against the anandamide.

- The new ligand system was solvated using a gmx module of GROMACS, and a 150 mM concentration of NaCl was maintained.

- To create the complex system, the anandamide-bound receptor system was converted into the top and gro format using ParmEd software. Further, anandamide topology parameters and coordinates were replaced with the ligand system.

To perform Alchemical free energy calculations, the system is divided along a hyperparameter λ, where λ equal to 0 represents the reference ligand and λ value 1 represents query ligand.[118] Here, twenty-one λ values were considered for both ligand and complex systems. In Intermediate λ values, force-field parameters were obtained by the linear combination of both ligand parameters.

Ligand systems and complex systems were minimized and equilibrated using NVT and NPT ensemble. The steepest descent algorithm was used for energy minimization for 15000 steps. Langevin dynamics was used as a thermostat to maintain the temperature at 300K.[69] Parrinello-Rahman barostat was used to control the pressure at 1 atm.[119] Cutoff distances for electrostatics and other non-bonded interactions were 12 Å. The Particle Mesh Ewald method was used for long-range electrostatic calculations. To take into account the particle disappearance, the soft-core version of the potential was used to with hyperparameter σ, α and soft-core λ power set to 0.3, 0.5, and 1 respectively.[120] Each NVT and NPT step was run for 100 ps. For a complex system, the backbone atoms of the protein are restrained with 1000 kJ/mol. Production runs for each λ value of each complex and ligand system were performed for five ns. Free energy calculations from these simulations were performed with a thermodynamics integration approach using the Python module alchemical-analysis.[121]

Acknowledgements

Author contributions

D.S. obtained funding for the project. D.S. supervised the research. D.S. and S.D. designed the research. S.D. and L.Z built the MD systems and performed simulations. D.S. participated in discussing results. S.D. analyzed data, and wrote the article with inputs from D.S. All authors reviewed the manuscript.

Conflict of interest

The authors declare no conflict of interest.

Code availability

Python scripts to reproduce the results are publicly available at GitHub. `https://github.com/ShuklaGroup/Dutta_Shukla_Cannabinoid_2023b.git`

Data availability

Simulation topology, trajectory files, MSM objects and other necessary files to reproduce the analyses shown in the manuscript can be obtained from `https://uofi.box.com/s/q9w1f99mybxo6hrgo8sfr3sbm3l8xrmu`.

References

(1) Martin, B.; Mechoulam, R.; Razdan, R. Discovery and characterization of endogenous cannabinoids. *Life Sciences* **1999**, *65*, 573–595.

(2) Zou, S.; Kumar, U. Cannabinoid Receptors and the Endocannabinoid System: Signal-

ing and Function in the Central Nervous System. *International Journal of Molecular Sciences* **2018**, *19*, 833.

(3) Kendall, D. A.; Yudowski, G. A. Cannabinoid Receptors in the Central Nervous System: Their Signaling and Roles in Disease. *Frontiers in Cellular Neuroscience* **2017**, *10*, 294.

(4) Turcotte, C.; Blanchet, M.-R.; Laviolette, M.; Flamand, N. The CB2 receptor and its role as a regulator of inflammation. *Cellular and Molecular Life Sciences* **2016**, *73*, 4449–4470.

(5) Kaur, I.; Behl, T.; Bungau, S.; Zengin, G.; Kumar, A.; El-Esawi, M. A.; Khullar, G.; Venkatachalam, T.; Arora, S. The endocannabinoid signaling pathway as an emerging target in pharmacotherapy, earmarking mitigation of destructive events in rheumatoid arthritis. *Life Sciences* **2020**, *257*, 118109.

(6) Piomelli, D. The molecular logic of endocannabinoid signalling. *Nature Reviews Neuroscience* **2003**, *4*, 873–884.

(7) Castillo, P. E.; Younts, T. J.; Chávez, A. E.; Hashimotodani, Y. Endocannabinoid Signaling and Synaptic Function. *Neuron* **2012**, *76*, 70–81.

(8) Grotsch, K.; Fokin, V. V. Between Science and Big Business: Tapping Mary Jane's Uncharted Potential. *ACS Central Science* **2022**, *8*, 156–168.

(9) Diana, M. A.; Bregestovski, P. Calcium and endocannabinoids in the modulation of inhibitory synaptic transmission. *Cell Calcium* **2005**, *37*, 497–505.

(10) Vaughan, C. W.; Christie, M. J. *Handbook of Experimental Pharmacology*; Springer-Verlag, pp 367–383.

(11) Rosenbaum, D. M.; Rasmussen, S. G. F.; Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **2009**, *459*, 356–363.

(12) Latorraca, N. R.; Venkatakrishnan, A. J.; Dror, R. O. GPCR Dynamics: Structures in Motion. *Chemical Reviews* **2016**, *117*, 139–155.

(13) Weis, W. I.; Kobilka, B. K. The Molecular Basis of G Protein–Coupled Receptor Activation. *Annual Review of Biochemistry* **2018**, *87*, 897–919.

(14) Basavarajappa, B. Critical Enzymes Involved in Endocannabinoid Metabolism. *Protein & Peptide Letters* **2007**, *14*, 237–246.

(15) Hua, T.; Vemuri, K.; Pu, M.; Qu, L.; Han, G. W.; Wu, Y.; Zhao, S.; Shui, W.; Li, S.; Korde, A.; Laprairie, R. B.; Stahl, E. L.; Ho, J.-H.; Zvonok, N.; Zhou, H.; Kufareva, I.; Wu, B.; Zhao, Q.; Hanson, M. A.; Bohn, L. M.; Makriyannis, A.; Stevens, R. C.; Liu, Z.-J. Crystal Structure of the Human Cannabinoid Receptor CB1. *Cell* **2016**, *167*, 750–762.e14.

(16) Li, X.; Hua, T.; Vemuri, K.; Ho, J.-H.; Wu, Y.; Wu, L.; Popov, P.; Benchama, O.; Zvonok, N.; Locke, K.; Qu, L.; Han, G. W.; Iyer, M. R.; Cinar, R.; Coffey, N. J.; Wang, J.; Wu, M.; Katritch, V.; Zhao, S.; Kunos, G.; Bohn, L. M.; Makriyannis, A.; Stevens, R. C.; Liu, Z.-J. Crystal Structure of the Human Cannabinoid Receptor CB2. *Cell* **2019**, *176*, 459–467.e13.

(17) Luchicchi, A.; Pistis, M. Anandamide and 2-arachidonoylglycerol: Pharmacological Properties, Functional Features, and Emerging Specificities of the Two Major Endocannabinoids. *Molecular Neurobiology* **2012**, *46*, 374–392.

(18) Howlett, A. C. International Union of Pharmacology. XXVII. Classification of Cannabinoid Receptors. *Pharmacological Reviews* **2002**, *54*, 161–202.

(19) Guindon, J.; Hohmann, A. The Endocannabinoid System and Pain. *CNS & Neurological Disorders - Drug Targets* **2009**, *8*, 403–421.

(20) Scherma, M.; Masia, P.; Satta, V.; Fratta, W.; Fadda, P.; Tanda, G. Brain activity of anandamide: a rewarding bliss? *Acta Pharmacologica Sinica* **2018**, *40*, 309–323.

(21) Reggio, P. Endocannabinoid Binding to the Cannabinoid Receptors: What Is Known and What Remains Unknown. *Current Medicinal Chemistry* **2010**, *17*, 1468–1486.

(22) Kumar, K. K.; Shalev-Benami, M.; Robertson, M. J.; Hu, H.; Banister, S. D.; Hollingsworth, S. A.; Latorraca, N. R.; Kato, H. E.; Hilger, D.; Maeda, S.; Weis, W. I.; Farrens, D. L.; Dror, R. O.; Malhotra, S. V.; Kobilka, B. K.; Skiniotis, G. Structure of a Signaling Cannabinoid Receptor 1-G Protein Complex. *Cell* **2019**, *176*, 448–458.e12.

(23) Dutta, S.; Selvam, B.; Das, A.; Shukla, D. Mechanistic origin of partial agonism of tetrahydrocannabinol for cannabinoid receptors. *Journal of Biological Chemistry* **2022**, *298*, 101764.

(24) Dutta, S.; Shukla, D. Transition-based reweighting and neural rational inference analysis of effects of new psychoactive substances on cannabinoid receptors. *eLife* **2024**, doi: 10.7554/eLife.98798.1.

(25) Pándy-Szekeres, G.; Munk, C.; Tsonkov, T. M.; Mordalski, S.; Harpsøe, K.; Hauser, A. S.; Bojarski, A. J.; Gloriam, D. E. GPCRdb in 2018: adding GPCR structure models and ligands. *Nucleic Acids Research* **2017**, *46*, D440–D446.

(26) Abood, M. E. *Handbook of Experimental Pharmacology*; Springer-Verlag, pp 81–115.

(27) Khanolkar, A. D.; Makriyannis, A. Structure-activity relationships of anandamide, an endogenous cannabinoid ligand. *Life Sciences* **1999**, *65*, 607–616.

(28) Lin, S.; Khanolkar, A. D.; Fan, P.; Goutopoulos, A.; Qin, C.; Papahadjis, D.; Makriyannis, A. Novel Analogues of Arachidonylethanolamide (Anandamide): Affinities for the CB1 and CB2 Cannabinoid Receptors and Metabolic Stability. *Journal of Medicinal Chemistry* **1998**, *41*, 5353–5361.

33

(29) Marzo, V. D.; Bisogno, T.; Petrocellis, L. D.; Brandi, I.; Jefferson, R.; Winckler, R.; Davis, J.; Dasse, O.; Mahadevan, A.; Razdan, R.; Martin, B. Highly Selective CB1 Cannabinoid Receptor Ligands and Novel CB1/VR1 Vanilloid Receptor "Hybrid" Ligands. *Biochemical and Biophysical Research Communications* **2001**, *281*, 444–451.

(30) Peng, Y.; McCorvy, J. D.; Harpsøe, K.; Lansu, K.; Yuan, S.; Popov, P.; Qu, L.; Pu, M.; Che, T.; Nikolajsen, L. F.; Huang, X.-P.; Wu, Y.; Shen, L.; Bjørn-Yoshimoto, W. E.; Ding, K.; Wacker, D.; Han, G. W.; Cheng, J.; Katritch, V.; Jensen, A. A.; Hanson, M. A.; Zhao, S.; Gloriam, D. E.; Roth, B. L.; Stevens, R. C.; Liu, Z.-J. 5-HT2C Receptor Structures Reveal the Structural Basis of GPCR Polypharmacology. *Cell* **2018**, *172*, 719–730.e14.

(31) Lee, S.-M.; Booe, J. M.; Pioszak, A. A. Structural insights into ligand recognition and selectivity for classes A, B, and C GPCRs. *European Journal of Pharmacology* **2015**, *763*, 196–205.

(32) Bonner, G.; Meng, F.; Akil, H. Selectivity of μ-opioid receptor determined by interfacial residues near third extracellular loop. *European Journal of Pharmacology* **2000**, *403*, 37–44.

(33) Jakowiecki, J.; Filipek, S. Hydrophobic Ligand Entry and Exit Pathways of the CB1 Cannabinoid Receptor. *Journal of Chemical Information and Modeling* **2016**, *56*, 2457–2466.

(34) Szlenk, C. T.; GC, J. B.; Natesan, S. Does the Lipid Bilayer Orchestrate Access and Binding of Ligands to Transmembrane Orthosteric/Allosteric Sites of G Protein-Coupled Receptors? *Molecular Pharmacology* **2019**, *96*, 527–541.

(35) Hua, T.; Vemuri, K.; Nikas, S. P.; Laprairie, R. B.; Wu, Y.; Qu, L.; Pu, M.; Korde, A.; Jiang, S.; Ho, J.-H.; Han, G. W.; Ding, K.; Li, X.; Liu, H.; Hanson, M. A.; Zhao, S.;

Bohn, L. M.; Makriyannis, A.; Stevens, R. C.; Liu, Z.-J. Crystal structures of agonist-bound human cannabinoid receptor CB1. *Nature* **2017**, *547*, 468–471.

(36) Li, X.; Chang, H.; Bouma, J.; de Paus, L. V.; Mukhopadhyay, P.; Paloczi, J.; Mustafa, M.; van der Horst, C.; Kumar, S. S.; Wu, L.; Yu, Y.; van den Berg, R. J. B. H. N.; Janssen, A. P. A.; Lichtman, A.; Liu, Z.-J.; Pacher, P.; van der Stelt, M.; Heitman, L. H.; Hua, T. Structural basis of selective cannabinoid CB2 receptor activation. *Nature Communications* **2023**, *14*, 1447.

(37) Casajuana-Martin, N.; Navarro, G.; Gonzalez, A.; del Torrent, C. L.; Gómez-Autet, M.; García, A. Q.; Franco, R.; Pardo, L. A Single Point Mutation Blocks the Entrance of Ligands to the Cannabinoid CB_2 Receptor via the Lipid Bilayer. *Journal of Chemical Information and Modeling* **2022**, *62*, 5771–5779.

(38) Dutta, S.; Shukla, D. Distinct activation mechanisms regulate subtype selectivity of Cannabinoid receptors. *Communications Biology* **2023**, *6*, 485.

(39) Yang, X.; Zhou, P.; Shen, S.; Hu, Q.; Tian, C.; Xia, A.; Wang, Y.; Yang, Z.; Nan, J.; Zhou, Y.; Chen, S.; Tian, X.; Wu, C.; Lin, G.; Zhang, L.; Wang, K.; Zheng, T.; Zou, J.; Yan, W.; Shao, Z.; Yang, S. Entropy drives the ligand recognition in G-protein-coupled receptor subtypes. *Proceedings of the National Academy of Sciences* **2024**, *121*, e2401091121.

(40) Pande, V. S.; Beauchamp, K.; Bowman, G. R. Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* **2010**, *52*, 99–105.

(41) Mardt, A.; Pasquali, L.; Wu, H.; Noé, F. VAMPnets for deep learning of molecular kinetics. *Nature Communications* **2018**, *9*, 4443.

(42) Weigle, A. T.; Carr, M.; Shukla, D. Impact of Increased Membrane Realism on Conformational Sampling of Proteins. *Journal of Chemical Theory and Computation* **2021**, *17*, 5342–5357.

(43) Pogozheva, I. D.; Armstrong, G. A.; Kong, L.; Hartnagel, T. J.; Carpino, C. A.; Gee, S. E.; Picarello, D. M.; Rubin, A. S.; Lee, J.; Park, S.; Lomize, A. L.; Im, W. Comparative Molecular Dynamics Simulation Studies of Realistic Eukaryotic, Prokaryotic, and Archaeal Membranes. *Journal of Chemical Information and Modeling* **2022**, *62*, 1036–1051.

(44) Shim, J.-Y.; Bertalovitz, A. C.; Kendall, D. A. Probing the Interaction of SR141716A with the CB1 Receptor. *Journal of Biological Chemistry* **2012**, *287*, 38741–38754.

(45) Xing, C.; Zhuang, Y.; Xu, T.-H.; Feng, Z.; Zhou, X. E.; Chen, M.; Wang, L.; Meng, X.; Xue, Y.; Wang, J.; Liu, H.; McGuire, T. F.; Zhao, G.; Melcher, K.; Zhang, C.; Xu, H. E.; Xie, X.-Q. Cryo-EM Structure of the Human Cannabinoid Receptor CB2-Gi Signaling Complex. *Cell* **2020**, *180*, 645–654.e13.

(46) Kumar, K. K.; Robertson, M. J.; Thadhani, E.; Wang, H.; Suomivuori, C.-M.; Powers, A. S.; Ji, L.; Nikas, S. P.; Dror, R. O.; Inoue, A.; Makriyannis, A.; Skiniotis, G.; Kobilka, B. Structural basis for activation of CB1 by an endocannabinoid analog. *Nature Communications* **2023**, *14*, 2672.

(47) Adams, I. B.; Ryan, W.; Singer, M.; Razdan, R. K.; Compton, D. R.; Martin, B. R. Pharmacological and behavioral evaluation of alkylated anandamide analogs. *Life Sciences* **1995**, *56*, 2041–2048.

(48) Stein, A.; Kortemme, T. Improvements to Robotics-Inspired Conformational Sampling in Rosetta. *PLoS ONE* **2013**, *8*, e63090.

(49) Alford, R. F.; Leaver-Fay, A.; Jeliazkov, J. R.; O'Meara, M. J.; DiMaio, F. P.; Park, H.; Shapovalov, M. V.; Renfrew, P. D.; Mulligan, V. K.; Kappel, K.; Labonte, J. W.; Pacella, M. S.; Bonneau, R.; Bradley, P.; Dunbrack, R. L.; Das, R.; Baker, D.; Kuhlman, B.; Kortemme, T.; Gray, J. J. The Rosetta All-Atom Energy Function for

Macromolecular Modeling and Design. *Journal of Chemical Theory and Computation* **2017**, *13*, 3031–3048.

(50) Raman, S.; Vernon, R.; Thompson, J.; Tyka, M.; Sadreyev, R.; Pei, J.; Kim, D.; Kellogg, E.; DiMaio, F.; Lange, O.; Kinch, L.; Sheffler, W.; Kim, B.-H.; Das, R.; Grishin, N. V.; Baker, D. Structure prediction for CASP8 with all-atom refinement using Rosetta. *Proteins: Structure, Function, and Bioinformatics* **2009**, *77*, 89–99.

(51) Kooistra, A. J.; Mordalski, S.; Pándy-Szekeres, G.; Esguerra, M.; Mamyrbekov, A.; Munk, C.; Keserű, G. M.; Gloriam, D. E. GPCRdb in 2021: integrating GPCR sequence, structure and function. *Nucleic Acids Research* **2020**, *49*, D335–D343.

(52) Fay, J. F.; Farrens, D. L. The Membrane Proximal Region of the Cannabinoid Receptor CB1 N-Terminus Can Allosterically Modulate Ligand Affinity. *Biochemistry* **2013**, *52*, 8286–8294.

(53) Yang, X.; Wang, X.; Xu, Z.; Wu, C.; Zhou, Y.; Wang, Y.; Lin, G.; Li, K.; Wu, M.; Xia, A.; Liu, J.; Cheng, L.; Zou, J.; Yan, W.; Shao, Z.; Yang, S. Molecular mechanism of allosteric modulation for the cannabinoid receptor CB1. *Nature Chemical Biology* **2022**, *18*, 831–840.

(54) Mandell, D. J.; Coutsias, E. A.; Kortemme, T. Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. *Nature Methods* **2009**, *6*, 551–552.

(55) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* **2008**, *29*, 1859–1865.

(56) Ingólfsson, H. I.; Carpenter, T. S.; Bhatia, H.; Bremer, P.-T.; Marrink, S. J.; Lightstone, F. C. Computational Lipidomics of the Neuronal Plasma Membrane. *Biophysical Journal* **2017**, *113*, 2271–2280.

(57) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *The Journal of Physical Chemistry A* **2001**, *105*, 9954–9960.

(58) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* **2016**, *14*, 71–73.

(59) Vanommeslaeghe, K.; MacKerell, A. D. Automation of the CHARMM General Force Field (CGenFF) I: Bond Perception and Atom Typing. *Journal of Chemical Information and Modeling* **2012**, *52*, 3144–3154.

(60) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry* **2009**, *30*, 2157–2164.

(61) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *Journal of Molecular Graphics* **1996**, *14*, 33–38.

(62) Shirts, M. R.; Klein, C.; Swails, J. M.; Yin, J.; Gilson, M. K.; Mobley, D. L.; Case, D. A.; Zhong, E. D. Lessons learned from comparing molecular dynamics engines on the SAMPL5 dataset. *Journal of Computer-Aided Molecular Design* **2016**, *31*, 147–161.

(63) Hopkins, C. W.; Grand, S. L.; Walker, R. C.; Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *Journal of Chemical Theory and Computation* **2015**, *11*, 1864–1874.

(64) Balusek, C.; Hwang, H.; Lau, C. H.; Lundquist, K.; Hazel, A.; Pavlova, A.; Lynch, D. L.; Reggio, P. H.; Wang, Y.; Gumbart, J. C. Accelerating Membrane Simulations with Hydrogen Mass Repartitioning. *Journal of Chemical Theory and Computation* **2019**, *15*, 4673–4686.

(65) Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Grand, S. L.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *Journal of Chemical Theory and Computation* **2013**, *9*, 3878–3888.

(66) Voelz, V. A.; Pande, V. S.; Bowman, G. R. Folding@home: Achievements from over 20 years of citizen science herald the exascale era. *Biophysical Journal* **2023**, *122*, 2852–2863.

(67) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Computational Biology* **2017**, *13*, e1005659.

(68) Meza, J. C. Steepest descent. *WIREs Computational Statistics* **2010**, *2*, 719–722.

(69) Schneider, T.; Stoll, E. Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Physical Review B* **1978**, *17*, 1302–1322.

(70) Braun, E.; Gilmer, J.; Mayes, H. B.; Mobley, D. L.; Monroe, J. I.; Prasad, S.; Zuckerman, D. M. Best Practices for Foundations in Molecular Simulations [Article v1.0]. *Living Journal of Computational Molecular Science* **2019**, *1*, 5957.

(71) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **1995**, *103*, 8577–8593.

(72) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* **1977**, *23*, 327–341.

(73) Saleh, N.; Ibrahim, P.; Saladino, G.; Gervasio, F. L.; Clark, T. An Efficient Metadynamics-Based Protocol To Model the Binding Affinity and the Transition State Ensemble of G-Protein-Coupled Receptor Ligands. *Journal of Chemical Information and Modeling* **2017**, *57*, 1210–1217.

(74) Mattedi, G.; Deflorian, F.; Mason, J. S.; de Graaf, C.; Gervasio, F. L. Understanding Ligand Binding Selectivity in a Prototypical GPCR Family. *Journal of Chemical Information and Modeling* **2019**, *59*, 2830–2836.

(75) Dutta, S.; Selvam, B.; Shukla, D. Distinct Binding Mechanisms for Allosteric Sodium Ion in Cannabinoid Receptors. *ACS Chemical Neuroscience* **2022**, *13*, 379–389.

(76) Saleh, N.; Hucke, O.; Kramer, G.; Schmidt, E.; Montel, F.; Lipinski, R.; Ferger, B.; Clark, T.; Hildebrand, P. W.; Tautermann, C. S. Multiple Binding Sites Contribute to the Mechanism of Mixed Agonistic and Positive Allosteric Modulators of the Cannabinoid CB1 Receptor. *Angewandte Chemie* **2018**, *130*, 2610–2615.

(77) Hinrichs, N. S.; Pande, V. S. Calculation of the distribution of eigenvalues and eigenvectors in Markovian state models for molecular dynamics. *The Journal of Chemical Physics* **2007**, *126*, 244101.

(78) Weber, J. K.; Pande, V. S. Characterization and Rapid Sampling of Protein Folding Markov State Model Topologies. *Journal of Chemical Theory and Computation* **2011**, *7*, 3405–3411.

(79) Zimmerman, M. I.; Bowman, G. R. FAST Conformational Searches by Balancing Exploration/Exploitation Trade-Offs. *Journal of Chemical Theory and Computation* **2015**, *11*, 5747–5757.

(80) Shamsi, Z.; Moffett, A. S.; Shukla, D. Enhanced unbiased sampling of protein dynamics using evolutionary coupling information. *Scientific Reports* **2017**, *7*, 12700.

(81) Zimmerman, M. I.; Porter, J. R.; Sun, X.; Silva, R. R.; Bowman, G. R. Choice of Adaptive Sampling Strategy Impacts State Discovery, Transition Probabilities, and the Apparent Mechanism of Conformational Changes. *Journal of Chemical Theory and Computation* **2018**, *14*, 5459–5475.

(82) Shamsi, Z.; Cheng, K. J.; Shukla, D. Reinforcement Learning Based Adaptive Sampling: REAPing Rewards by Exploring Protein Conformational Landscapes. *The Journal of Physical Chemistry B* **2018**, *122*, 8386–8395.

(83) Buenfil, J.; Koelle, S. J.; Meila, M. Tangent Space Least Adaptive Clustering. *Proceedings of the 38th International Conference on Machine Learning, Proceedings of Machine Learning Research* **2021**, *139*.

(84) Kleiman, D. E.; Shukla, D. Active Learning of the Conformational Ensemble of Proteins Using Maximum Entropy VAMPNets. *Journal of Chemical Theory and Computation* **2023**, *19*, 4377–4388.

(85) Kleiman, D. E.; Nadeem, H.; Shukla, D. Adaptive Sampling Methods for Molecular Dynamics in the Era of Machine Learning. *The Journal of Physical Chemistry B* **2023**, *127*, 10669–10681.

(86) Hénin, J.; Lelièvre, T.; Shirts, M. R.; Valsson, O.; Delemotte, L. Enhanced Sampling Methods for Molecular Dynamics Simulations [Article v1.0]. *Living Journal of Computational Molecular Science* **2022**, *4*, 1583.

(87) Kim, K.; Bansal, P. D.; Shukla, D. Cyclopamine modulates smoothened receptor activity in a binding position dependent manner. *Communications Biology* **2024**, *7*, 1207.

(88) Deganutti, G.; Pipitò, L.; Rujan, R. M.; Weizmann, T.; Griffin, P.; Ciancetta, A.; Moro, S.; Reynolds, C. A. Hidden GPCR structural transitions addressed by

multiple walker supervised molecular dynamics (mwSuMD). *elife* **2024**, doi: 10.7554/elife.96513.1.

(89) Bansal, P. D.; Dutta, S.; Shukla, D. Activation mechanism of the human Smoothened receptor. *Biophysical Journal* **2023**, *122*, 1400–1413.

(90) Song, W.; Duncan, A. L.; Sansom, M. S. Modulation of adenosine A2a receptor oligomerization by receptor activation and PIP2 interactions. *Structure* **2021**, *29*, 1312–1325.e3.

(91) Betz, R. M.; Dror, R. O. How Effectively Can Adaptive Sampling Methods Capture Spontaneous Ligand Binding? *Journal of Chemical Theory and Computation* **2019**, *15*, 2053–2063.

(92) Kohlhoff, K. J.; Shukla, D.; Lawrenz, M.; Bowman, G. R.; Konerding, D. E.; Belov, D.; Altman, R. B.; Pande, V. S. Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways. *Nature Chemistry* **2013**, *6*, 15–21.

(93) Zhou, Q.; Yang, D.; Wu, M.; Guo, Y.; Guo, W.; Zhong, L.; Cai, X.; Dai, A.; Jang, W.; Shakhnovich, E. I.; Liu, Z.-J.; Stevens, R. C.; Lambert, N. A.; Babu, M. M.; Wang, M.-W.; Zhao, S. Common activation mechanism of class A GPCRs. *eLife* **2019**, *8*, e50279.

(94) Noé, F.; Rosta, E. Markov Models of Molecular Kinetics. *The Journal of Chemical Physics* **2019**, *151*, 190401.

(95) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov models of molecular kinetics: Generation and validation. *The Journal of Chemical Physics* **2011**, *134*, 174105.

(96) Husic, B. E.; Pande, V. S. Markov State Models: From an Art to a Science. *Journal of the American Chemical Society* **2018**, *140*, 2386–2396.

(97) Wang, W.; Cao, S.; Zhu, L.; Huang, X. Constructing Markov State Models to elucidate the functional conformational changes of complex biomolecules. *WIREs Computational Molecular Science* **2017**, *8*, e1343.

(98) Shukla, D.; Hernández, C. X.; Weber, J. K.; Pande, V. S. Markov State Models Provide Insights into Dynamic Modulation of Protein Function. *Accounts of Chemical Research* **2015**, *48*, 414–422.

(99) Schütte, C.; Fischer, A.; Huisinga, W.; Deuflhard, P. A direct approach to conformational dynamics based on hybrid Monte Carlo. *Journal of Computational Physics* **1999**, *151*, 146–168.

(100) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; Fabritiis, G. D.; Noé, F. Identification of slow molecular order parameters for Markov model construction. *The Journal of Chemical Physics* **2013**, *139*, 015102.

(101) Noé, F.; Clementi, C. Kinetic Distance and Kinetic Maps from Molecular Dynamics Simulation. *Journal of Chemical Theory and Computation* **2015**, *11*, 5002–5011.

(102) Wu, H.; Noé, F. Variational Approach for Learning Markov Processes from Time Series Data. *Journal of Nonlinear Science* **2019**, *30*, 23–66.

(103) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *Journal of Chemical Theory and Computation* **2015**, *11*, 5525–5542.

(104) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* **2015**, *109*, 1528–1532.

(105) Wagner, J. R.; Sørensen, J.; Hensley, N.; Wong, C.; Zhu, C.; Perison, T.; Amaro, R. E. POVME 3.0: Software for Mapping Binding Pocket Flexibility. *Journal of Chemical Theory and Computation* **2017**, *13*, 4584–4592.

(106) *GetContacts: Interactive analysis for atomic interaction in protein structures*, Stanford University. https://getcontacts.github.io/ (Accessed 2023-19-08).

(107) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *Journal of Chemical Theory and Computation* **2013**, *9*, 3084–3095.

(108) de Amorim, H. L.; Caceres, R.; Netz, P. Linear Interaction Energy (LIE) Method in Lead Discovery and Optimization. *Current Drug Targets* **2008**, *9*, 1100–1105.

(109) Löhr, T.; Kohlhoff, K.; Heller, G. T.; Camilloni, C.; Vendruscolo, M. A kinetic ensemble of the Alzheimer's Aβ peptide. *Nature Computational Science* **2021**, *1*, 71–78.

(110) Hoffmann, M.; Scherer, M.; Hempel, T.; Mardt, A.; de Silva, B.; Husic, B. E.; Klus, S.; Wu, H.; Kutz, N.; Brunton, S. L.; Noé, F. Deeptime: a Python library for machine learning dynamical models from time series data. *Machine Learning: Science and Technology* **2021**, *3*, 015009.

(111) E, W.; Vanden-Eijnden, E. Transition-Path Theory and Path-Finding Algorithms for the Study of Rare Events. *Annual Review of Physical Chemistry* **2010**, *61*, 391–420.

(112) Noé, F.; Schütte, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proceedings of the National Academy of Sciences* **2009**, *106*, 19011–19016.

(113) Sultan, M. M.; Kiss, G.; Pande, V. S. Towards simple kinetic models of functional dynamics for a kinase subfamily. *Nature Chemistry* **2018**, *10*, 903–909.

(114) Fleetwood, O.; Carlsson, J.; Delemotte, L. Identification of ligand-specific G protein-coupled receptor states and prediction of downstream efficacy via data-driven modeling. *eLife* **2021**, *10*, e60715.

(115) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.

(116) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem 2023 update. *Nucleic Acids Research* **2022**, *51*, D1373–D1380.

(117) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An open chemical toolbox. *Journal of Cheminformatics* **2011**, *3*, 33.

(118) Mey, A. S.; Allen, B. K.; Macdonald, H. E. B.; Chodera, J. D.; Hahn, D. F.; Kuhn, M.; Michel, J.; Mobley, D. L.; Naden, L. N.; Prasad, S.; Rizzi, A.; Scheen, J.; Shirts, M. R.; Tresadern, G.; Xu, H. Best Practices for Alchemical Free Energy Calculations [Article v1.0]. *Living Journal of Computational Molecular Science* **2020**, *2*, 18378.

(119) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **1981**, *52*, 7182–7190.

(120) Steinbrecher, T.; Joung, I.; Case, D. A. Soft-core potentials in thermodynamic integration: Comparing one- and two-step transformations. *Journal of Computational Chemistry* **2011**, *32*, 3253–3263.

(121) Klimovich, P. V.; Shirts, M. R.; Mobley, D. L. Guidelines for the analysis of free energy calculations. *Journal of Computer-Aided Molecular Design* **2015**, *29*, 397–411.