Energetics of cholesterol perception and translocation in the human Smoothened receptor

Prateek D. Bansal,[†] Maia Kinnebrew,[‡] Rajat Rohatgi,[‡,¶] and Diwakar Shukla[*,†,§,∥,⊥]

†*Department of Chemical and Biomolecular Engineering, University of Illinois Urbana-Champaign, Urbana, IL 61801, United States*

‡*Department of Biochemistry, Stanford University School of Medicine, Stanford, CA 94305, United States.*

¶*Department of Medicine, Stanford University School of Medicine, Stanford, CA 94305, United States.*

§*Department of Chemistry, University of Illinois at Urbana-Champaign, Urbana, IL, 61801*

∥*Department of Bioengineering, University of Illinois Urbana-Champaign, Urbana, IL 61801, United States*

⊥*Cancer Center at Illinois, University of Illinois at Urbana-Champaign, Urbana, IL, 61801*

E-mail: diwakar@illinois.edu

Abstract

Smoothened (SMO), a member of the G Protein-Coupled Receptor superfamily, mediates Hedgehog signaling and is linked to cancer and birth defects. SMO responds to accessible cholesterol in the ciliary membrane, translocating it via a longitudinal tunnel to its extracellular domain. Reaching a complete mechanistic understanding of the cholesterol translocation process would help in the development of cancer therapies. Competing hypotheses based on available structures support entry of cholesterol from

outer and inner membrane leaflets, but the exact mechanism of translocation remains unclear. Using atomistic molecular dynamics simulations (∼2 millisecond simulations) and biochemical assays of SMO mutants, we assess the energetic feasibilities of proposed hypotheses. We show that the energetic barriers for cholesterol translocation from either leaflets are comparable. Mutagenesis experiments and complementary simulations of SMO mutants validate the role of critical amino acid residues along the translocation pathways. Our data suggests that cholesterol can take either pathway to enter SMO, thus explaining contradictory experimental observations in literature. Thus, our results illuminate the energetics and provide a first molecular description of cholesterol translocation in SMO.

keywords

G Protein-Coupled Receptors, Smoothened, cholesterol, Hedgehog signaling, Markov State Models, mutagenesis, translocation

Introduction

Smoothened (SMO) is a member of the G Protein-Coupled Receptor (GPCR) superfamily with a typical heptahelical transmembrane fold.[1,2] SMO has been identified as an oncoprotein, and mutations that cause overactivity of SMO drive tumorigenesis in basal cell carcinoma and medulloblastoma.[3] SMO antagonists are validated anti-cancer drugs,[4,5] but are limited by drug resistance and side effects.[6] SMO functions as a transmembrane signal transducer in the Hedgehog (HH) signaling pathway which is known to mediate cell differentiation during embryonic development.[7,8] It transduces signals across the cell membrane, particularly in the primary cilia.[9,10] However, the mechanism of endogenous activation of SMO is still a debated question in the field. SMO activation has been linked to membrane sterols in numerous studies.[11–20] Cholesterol, a steroidal molecule found abundantly

in the plasma membranes of vertebrates, has been identified as the endogenous agonist for SMO.[16–18]Membrane cholesterol has been shown to interact with and modulate GPCR activity,[21–23] but in the case of SMO, cholesterol has been uniquely shown to be necessary and sufficient for SMO activation.[16,17] Recent work has shown that SMO is activated by a minor pool of membrane cholesterol, termed accessible cholesterol, at the primary cilium.[20,24,25] Patched-1 (PTCH1), a twelve-pass transmembrane protein, sits directly upstream of SMO in the HH pathway, and inhibits SMO.[26–29] PTCH1 is proposed to modulate the HH signal by decreasing the ability of membrane cholesterol[20,25,30–34] to access SMO's extracellular Cysteine-Rich Domain (CRD).[35]

The mechanism by which SMO is activated has been the subject of much speculation over the recent years, with multiple studies theorizing the mechanism of PTCH1's inhibition on SMO. Previous studies have suggested that PTCH1 could function as a sterol transporter.[25,29,31,36,37] A recent study using coarse-grained Molecular Dynamics simulations investigated the possibility of this process, concluding that the overall process might occur at a energetic cost of $\sim$ 3-5 kcal/mol.[38] Recently resolved structures of PTCH1[30,33,34,39] have reported the presence of a Sterol Binding Domain (SBD) and a hydrophobic conduit that extends from the inner leaflet to the extracellular space. These observations conclude that PTCH1 controls SMO's access to membrane cholesterol.

The CRD of SMO contains a binding site for steroidal molecules.[13,14,40,41] The orthosteric agonist cholesterol,[18] as well as the naturally occurring alkaloidal agonist cyclopamine,[17,40] bind in the CRD. In addition, SMO has a pocket in the Transmembrane Domain (TMD), which is known to bind to multiple antagonists such as LY2940680,[1] SANT1 and AntaXV,[2] cyclopamine,[42] TC114,[43] Vismodegib,[18] the synthetic agonist SAG (and variants SAG1.5, SAG21k),[2,44,45] the steroidal agonists 24S,25-epoxy cholesterol,[46] and cholesterol.[44] In recent years, there has been a vigorous debate on the orthosteric site of SMO - whether the endogenous activity of SMO is controlled by its CRD site or the TMD site. Recently, a study by Kinnebrew et al.[35] reported that SMO's CRD is the primary domain controlled by PTCH1

activity, which suggests that the CRD is the orthosteric site. Frizzleds (FZDs), members of the same Class F family of GPCRs as SMO, also bind to their orthosteric ligands, specifically the fatty acid moiety of Wnt ligands,[47–49] at the CRD. Therefore, if cholesterol mediating the activity of SMO originates in the membrane,[18,20] then cholesterol must travel through the SMO TMD to reach the orthosteric site in the CRD.[50] This is further supported by multiple structures of SMO bound to sterols in the TMD.[44–46] (Fig. 1a)

Huang et al. (2018) theorized that SMO's activation involved the translocation of cholesterol from the membrane via a hydrophobic conduit to the binding site in the CRD.[50] This study resolved a structure of active *Xenopus laevis* SMO; which showed outward movements of transmembrane helices 5 and 6 (TM5-TM6) on the intracellular end, which opened a cavity in SMO extending to the inner leaflet laterally, between TM5 and TM6. This led to the hypothesis that the entry of cholesterol into SMO could happen from the inner leaflet of the membrane, between TMs 5 and 6. There exists, however, contradictory evidence showing that the activity of SMO is regulated by cholesterol in the outer leaflet, entering SMO between TM2 and TM3. In 2019, Hedger et al.[51] reported a cholesterol binding site present at the outer leaflet, between the TM2 and TM3 helices. Using coarse-grained simulations, the authors tested a variety of membrane compositions around SMO and found that a cholesterol binding site existed in hSMO (human SMO). In 2021, Kinnebrew et al.[25] used total internal reflection fluorescence microscopy (TIRFM) to assess the effect of PTCH1's activity on membrane cholesterol. They concluded that PTCH1 activity caused a decrease in the accessibility of cholesterol in the outer leaflet, suggesting that outer leaflet cholesterol is sensed by SMO. Furthermore, in our previous work,[52] we observed that cholesterol accumulated outside TM2-TM3 in the outer leaflet of inactive SMO. This was supported by the observation that when SMO was bound to its agonist SAG, we observed a tunnel opening between TM2-TM3 in the outer leaflet, which may facilitate the translocation of cholesterol into the core of SMO's TMD. We therefore hypothesize that cholesterol may enter the TMD from the membrane through different pathways - (1) starting at the outer leaflet,

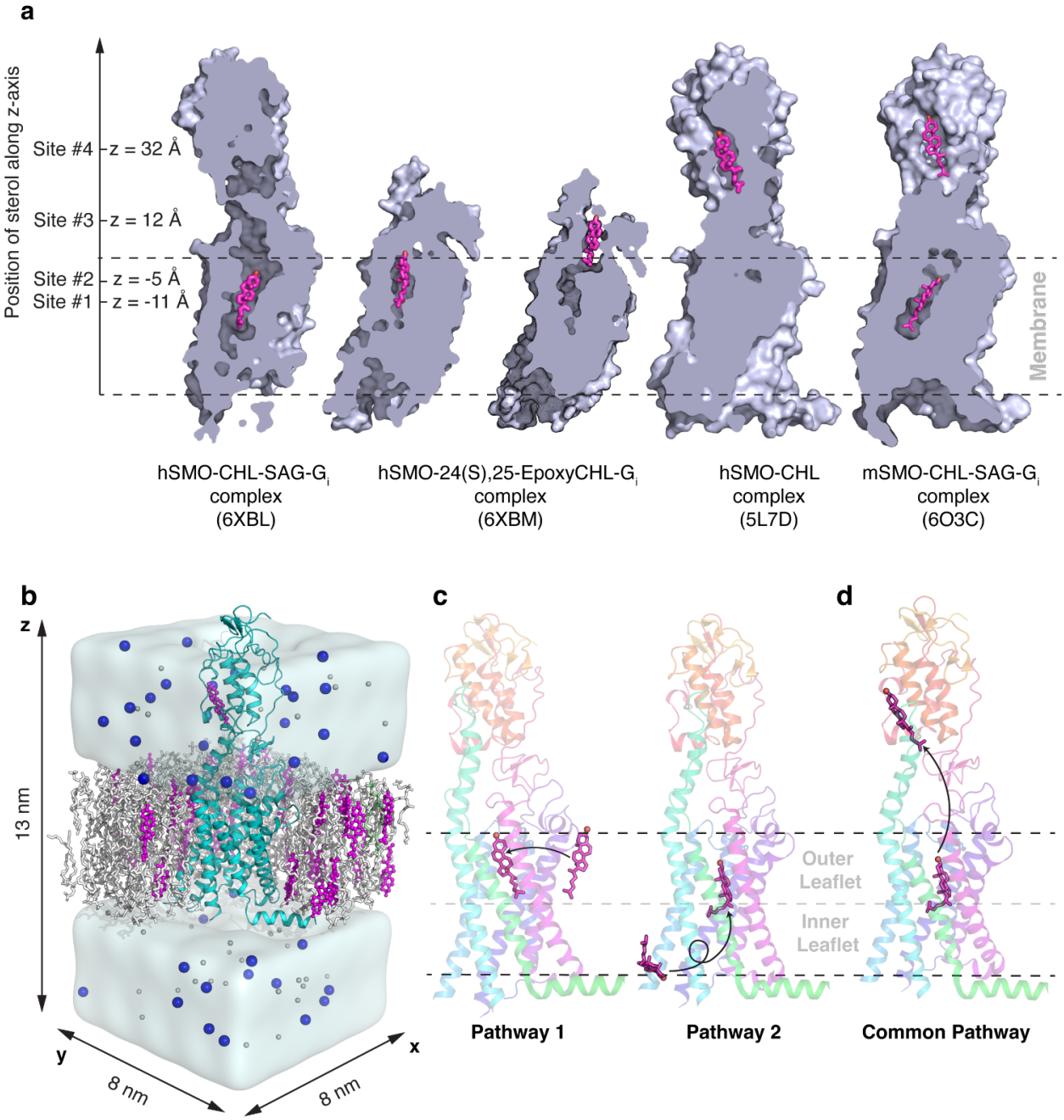

Figure 1: (a) The binding sites of the sterols along the hypothesized tunnel in SMO. Sterol binding sites have been identified deep in the TMD (6XBL),[44] in the extracellular TMD, at the interface of CRD and TMD (6XBM),[44] at the orthosteric binding site in the CRD (5L7D)[18] and a dual-bound mode where cholesterol is bound to TMD and CRD (6O3C).[45] (b) Example simulation system showing SMO (5L7D, cyan) embedded in a membrane (white/magenta). Water is shown as a surface, while sodium (blue) and chloride (gray) ions are shown as spheres. (c) The two pathways explored in this study for the translocation of cholesterol from membrane to SMO's TMD. (d) The common pathway followed by cholesterol once cholesterol enters the protein to reach SMO's CRD.

between TM2-TM3, or (2) starting at the inner leaflet, between TM5-TM6. Alternatively, (3) cholesterol could use both pathways if they show similar energetic behaviors.

How cholesterol moves from the membrane into the core of the TMD of SMO is still widely debated. Cholesterol traverses SMO to ultimately reach the CRD binding site but the mechanism of cholesterol perception has not been elucidated yet, which gives us an opportunity to explore the mechanistic aspects of this process from both computational and experimental viewpoints. In this study, we simulate SMO by embedding it in a membrane (Fig. 1b). We report the entire translocation path of cholesterol from the membrane to SMO's CRD for both of the hypothesized pathways (Fig. 1c) - between TM2-TM3 in the outer leaflet of the membrane (hereafter referred to as "Pathway 1"), and between TM5-TM6 in the inner leaflet of the membrane (hereafter referred to as "Pathway 2"). We observe that cholesterol can translocate via both pathways, and the free energy barriers associated with both pathways are comparable. We test mutations in SMO that can disrupt the movement or either pathway, and show that the experimental results are further supported by simulations of cholesterol translocation in SMO mutants.

Following the entry of cholesterol in our simulations, we observe cholesterol moving along TM6 to the TMD-CRD interface (common pathway, Fig. 1d) to access the orthosteric binding site in the CRD.[35,50] One of the unique features of SMO is the presence of a long helix 6 (TM6),[1] (Fig. **??**)) which acts as an connector between the CRD and the TMD. We test mutations in SMO that can disrupt cholesterol movement along this common pathway, and show that these mutants can halt the translocation process by loss of hydrophobic contacts. These results for SMO mutants are further validated by additional MD simulations. Therefore, in this study, the entire process of cholesterol translocation was observed using aggregate 2 milliseconds of unbiased all-atom molecular dynamics simulations. Exploring the mechanism by which SMO translocates cholesterol would provide insights into the endogenous regulation of HH activity and suggest strategies for the next generation of drugs targeting SMO.

Results and Discussion

The entry of cholesterol from the outer leaflet into the TMD exhibits the highest energetic barrier

We first tested the hypothesis that cholesterol enters the TMD of SMO from the outer leaflet (Pathway 1). To model this process, cholesterol was placed outside of the TMD in the membrane outer leaflet, between the TM2-TM3 interface. The entry of cholesterol from the membrane into the protein was steered towards the TMD to sample the entry pathway. The frames generated were then used to seed unbiased simulations for adaptive sampling (see Methods section for details). This was done to estimate the energetic barriers involved in the cholesterol entry from the outer leaflet.

To enter the TMD from the membrane, cholesterol must overcome the entropic barrier of being restrained inside the protein. Additionally, cholesterol faces an enthalpic barrier caused by the change in environment which is mostly hydrophobic in the membrane and both hydrophobic and hydrophilic inside the protein. This can make the entry process energetically expensive. To identify the barrier associated with entry, we projected the entire dataset on different translocation related metrics and identifed the mechanism of cholesterol translocation. The z-coordinate of cholesterol's center of mass (z-axis is perpendicular to the plane of the membrane) was used as a proxy for the progress of the overall translocation from the membrane to the CRD. The angle of cholesterol with the x-y plane (the plane of the membrane) was calculated to identify cholesterol pose. We observe that this angle (Fig. 2a and **??**a) shows multiple minima ($\boldsymbol{\alpha}/\boldsymbol{\beta}$ along the Pathway 1). This provides evidence for multiple stable poses along the pathway as observed in the multiple stable poses of cholesterol in Cryo-EM structures of SMO bound to sterols.[44–46] Overall, the highest barrier along Pathway 1 is $\sim 5.8 \pm 0.7$ kcal/mol and it is associated with the entry of cholesterol into the TMD (Fig **??**)). Several factors contribute to this high barrier; first, cholesterol needs to be at the correct position - just outside the protein. In addition, cholesterol needs

to be in the correct orientation - angled, with the isooctyl tail pointing towards the protein core. Additionally, the steric hinderances from the hydrophobic residues at the interface provide further obstacles for entry. Therefore, the TM2/3 residues at the entry point need to undergo conformational change to facilitate cholesterol entry – making it a rare event.

In the first stage of translocation, cholesterol is in the membrane ($\boldsymbol{\alpha}$, Fig. 2), just outside the TM2/3 helices of SMO as identified in literature.[51] At this cholesterol recognition site, SMO primarily contains hydrophobic residues (Fig. 2 c-f), which preferentially interact with the hydrophobic isooctyl tail of cholesterol. Therefore, the cholesterol entry involves the insertion of the tail into the TMD first, which is then followed by the hydrophilic androsterolic moiety. For the sake of clarity, we divide this entry pocket at the TM2-TM3-membrane interface into two parts - 'lower' and 'upper' pocket. The 'upper' pocket corresponds to the residues that coordinate with the androsterolic moiety of cholesterol (Fig. 2c, d), and the 'lower' pocket corresponds to residues that lie closer to the isooctyl tail of cholesterol (Fig. 2e, f). The upper (M286^{ECL1}, A289^{ECL1} and I389^{ECL2}) and the lower (W281$^{2.58f}$, F391^{ECL2} and M525$^{7.45f}$) pocket residues undergo conformational changes to open the space for the entry of cholesterol (Fig. 2c-f). Here, the superscript refers to the Wang numbering scheme.[2] This flexible motion is facilitated by G280$^{2.57f}$ in the lower pocket and G288^{ECL1} in the upper pocket. Upon the entry of the cholesterol tail, an intermediate state, $\boldsymbol{\beta}$ (Fig. 2d, f) is observed where cholesterol lies flat with respect to the membrane plane and forms extensive hydrophobic contacts with A283$^{2.60f}$, I317$^{3.28f}$, F318$^{3.29f}$ and the disulfide bridge forming residues C314$^{3.25f}$ - C390^{ECL2}.

To reach the TMD binding site, cholesterol must first "rock"back to its upright pose ($\boldsymbol{\alpha}^*$) from its flat conformation in the state $\boldsymbol{\beta}$. This rocking motion is facilitated by the polar interactions between S313$^{3.24f}$, Q284$^{2.63f}$ and the alcoholic oxygen in cholesterol. The entry of cholesterol is thus captured by a $\boldsymbol{\alpha} \rightarrow \boldsymbol{\beta} \rightarrow \boldsymbol{\alpha}^*$ transition. To further elucidate the position of cholesterol as it enters the protein, we projected the x and y coordinates of cholesterol on a free energy landscape (Fig. 2b, Fig. 4b and Fig. **??**bb, d). In this figure, the

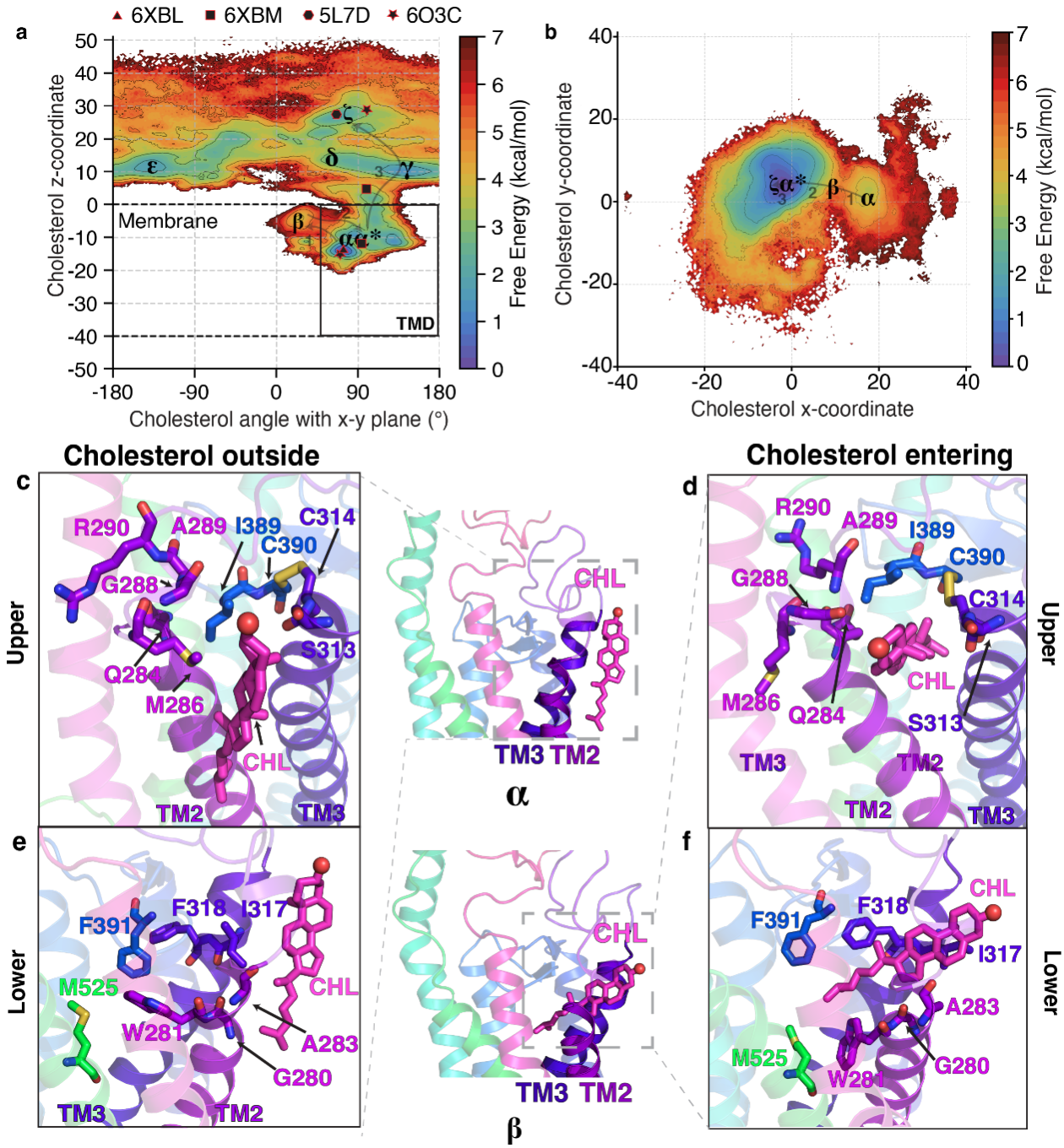

Figure 2: The molecular events as cholesterol enters the core of SMO's TMD from the outer leaflet of the membrane. (a) Free energy plot showing the angle of cholesterol with the x-y plane, the plane of the membrane, v/s the z-coordinate of cholesterol for Pathway 1. The pathway followed by cholesterol is $\boldsymbol{\alpha} \rightarrow \boldsymbol{\beta} \rightarrow \boldsymbol{\alpha}^*$. The experimental structures of SMO are shown as black polygons. (b) Free energy landscape of cholesterol's y-coordinate plotted v/s cholesterol's x-coordinate. Cholesterol interacts with residues in TM2-TM3 while entering core TMD of SMO. (c-f) Insets show cholesterol's interactions with residues at the membrane-protein interface for Pathway 1. (c,e) show cholesterol outside the protein ($\boldsymbol{\alpha}$), while (d,f) show cholesterol entering the protein ($\boldsymbol{\beta}$).

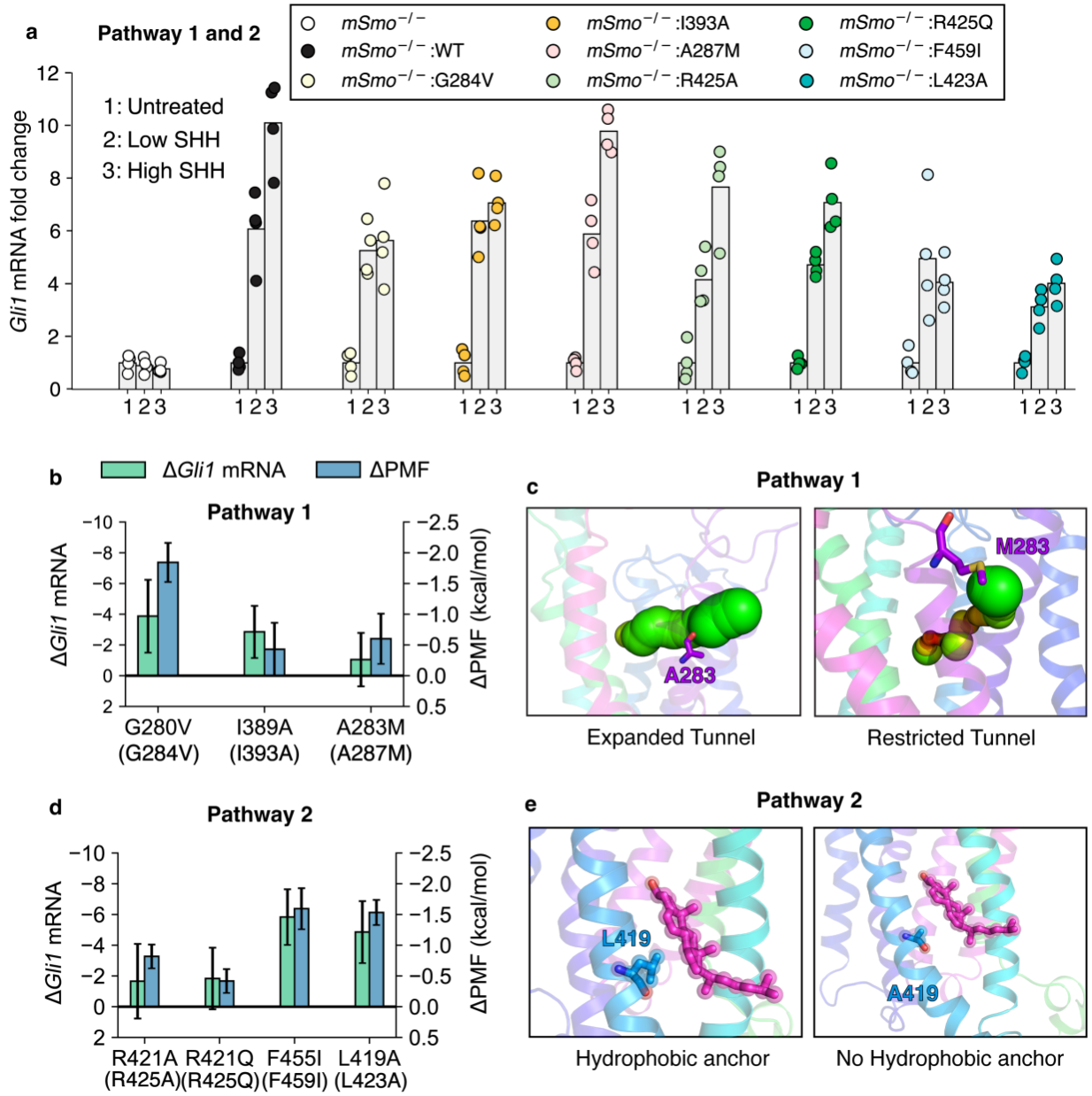

Figure 3: Effects of mutations along Pathways 1 and 2 on the activation of SMO. (a) *Gli1* mRNA fold change (indicator of SMO activation) plotted for SMO mutants, showing fold change when the mutants are untreated, treated with a low concentration of SHH and treated with a saturating concentration of SHH. (b) Δ*Gli1* mRNA fold change (high SHH vs untreated) and Δ PMF (difference of peak PMF) plotted for the mutants in Pathway 1. *hSMO* numbering is used as default, with *mSMO* numbering given in parentheses. (c) Example mutant A283M shows that cholesterol is able to enter SMO through Pathway 1 even on a bulky mutation. (d) Same as (b) but for Pathway 2. *hSMO* numbering is used as default, with *mSMO* numbering given in parentheses. (e) Example mutant L419A shows that cholesterol is able to enter SMO through Pathway 2 due to lesser steric hinderance.

state $\boldsymbol{\beta}$ clearly marks the transition state between cholesterol outside and inside the TMD. Overall, the angle between cholesterol and the x-y plane transitions from $90^\circ \rightarrow 0^\circ \rightarrow 90^\circ$ in order for cholesterol to enter the protein.

To further dive into the details of cholesterol translocation, we designed mutations along Pathway 1, and measured the change in activity with respect to Wild-Type (WT) *mSMO*. Activity was measured using *Gli1* mRNA fold change in the presence of SHH (Fig. 3a). To further validate the mutations, simulations were performed on hSMO to compute the Potential of Mean Force (PMF) of cholesterol entry[53,54] (additional details presented in Methods). Here, the PMF is able to characterize the barriers associated with the translocation of cholesterol, and a difference in the peak value of PMF is presented for each mutant (Fig. 3b). We mutated G280$^{2.57f}$ to valine - G280V to test whether reducing the flexibility of TM2 prevents cholesterol entry into the TMD. Consequently, the activity of *mSMO* showed a decrease. We designed I389A to check the importance of hydrophobic contacts during translocation along Pathway 1. However this mutant did not affect the activity nor the barrier for translocation significantly (Fig. 3b, **??**)). Finally, we mutated A283$^{2.60f}$ to methionine - A283M to test whether the presence of a bulkier residue would block translocation. Surprisingly, the effect on activity was not significant. When we calculated the PMF for cholesterol entry, A283M mutant showed restricted tunnel but it did not fully block the tunnel. Therefore, the change in the PMF and experimentally measured activity were not significant (Fig. 3b, c, **??**)).

Cholesterol 'flipping' corresponds to the highest barrier in its translocation from the inner leaflet

To quantitatively assess the validity of cholesterol translocation from inner leaflet, we performed adaptive sampling simulations to obtain the associated free energy barriers. For Pathway 2, cholesterol first binds at the interface between TM5 and TM6 in the inner leaflet.[45,50] However, cholesterol in the inner leaflet has a downward orientation, with the polar hydroxyl group pointing intracellularly ($\boldsymbol{\eta}$, Fig. 4). Thus, if cholesterol has to translo-

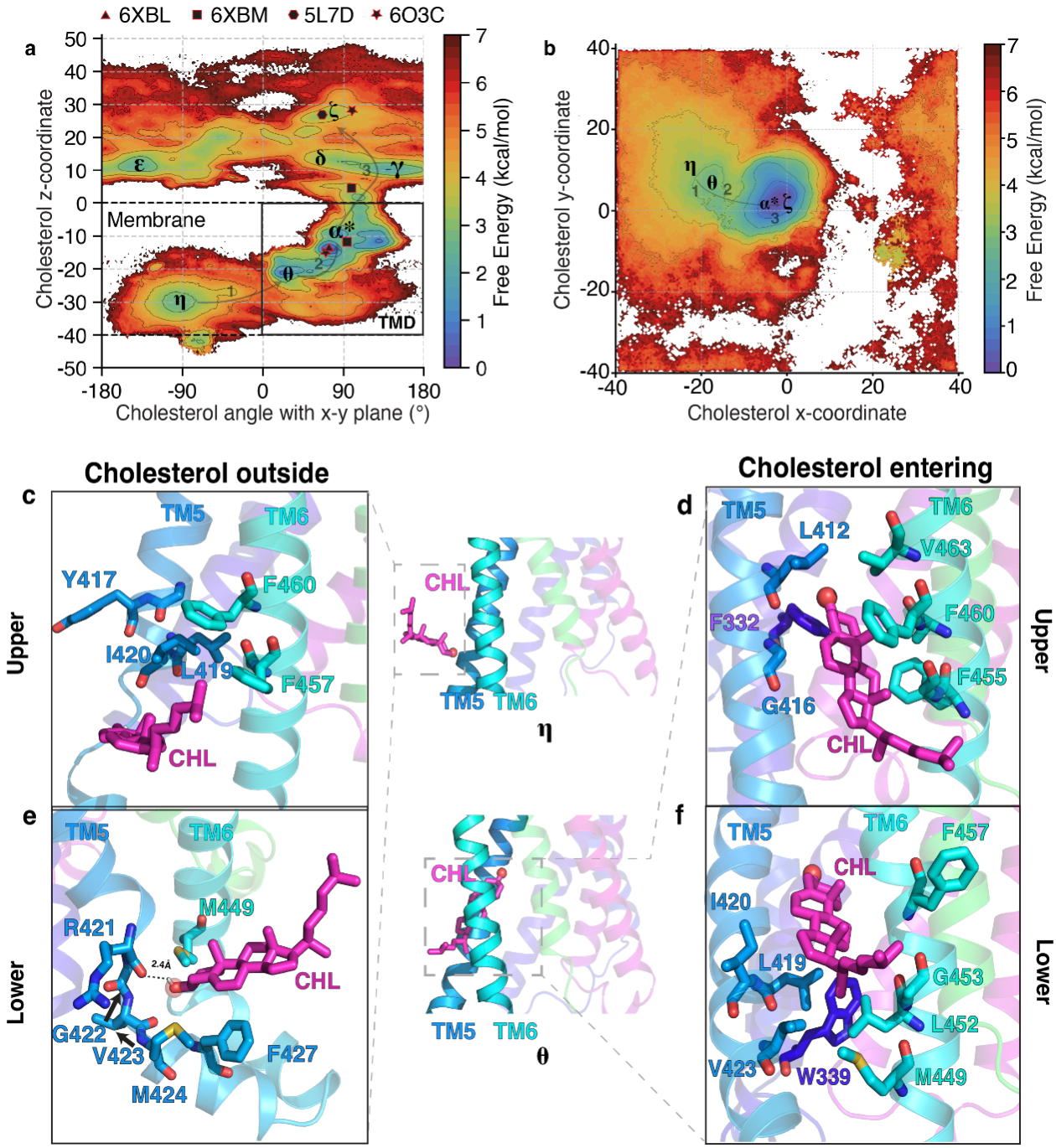

Figure 4: Molecular events at the entry of cholesterol from the membrane into the core SMO TMD for Pathway 2. (a) Free energy plot showing the angle of cholesterol with the x-y plane, the plane of the membrane, v/s the z-coordinate of cholesterol for Pathway 2. The pathway followed by cholesterol is $\boldsymbol{\eta} \rightarrow \boldsymbol{\theta} \rightarrow \boldsymbol{\alpha}^*$. The experimental structures of SMO are shown as black polygons. (b) Free energy landscape of cholesterol's y-coordinate plotted v/s cholesterol's x-coordinate for Pathway 2. Cholesterol interacts with residues in TM5-TM6 for Pathway 2 while entering SMO core TMD. (e-h) Insets show cholesterol's interactions with residues at the membrane-protein interface for Pathway 2. (c,e) show cholesterol outside the protein ($\boldsymbol{\eta}$), while (d,f) show cholesterol entering the protein ($\boldsymbol{\theta}$) for Pathway 2.

cate from the inner leaflet, it has to undergo "flipping" motion, as the resolved structure with the lowermost position of a sterol inside the tunnel[44] shows the alcoholic moiety pointing towards the CRD[45] (Fig. 1a). The energetic barrier associated with flipping the motion of cholesterol can be observed by estimating the angle cholesterol makes with the x-y plane and the barrier associated with translocation towards the TMD binding site can be estimated by projecting the data on the z-coordinate of cholesterol.

In Fig. 4a, multiple free energy minima are observed. The state $\boldsymbol{\eta}$ corresponds to cholesterol outside of the TMD and pointing downwards, forming a -90° angle with the x-y plane. The state $\boldsymbol{\epsilon}$ corresponds to cholesterol at the membrane-TMD interface and forming a 45° angle with the x-y plane. Finally, the state $\boldsymbol{\alpha}^*$ corresponds to cholesterol inside the TMD and forming a +90° angle with the x-y plane. The state $\boldsymbol{\alpha}^*$ which represents the TMD binding site is the most stable pose of cholesterol inside SMO. According to the free-energy landscapes (Fig. 2a, Fig. 4a and 2c,d and Fig. 4c,d), the entry of cholesterol from the inner leaflet is associated with the highest barrier for the entire translocation process, $\sim$6.5 $\pm$ 0.8 kcal/mol and it corresponds to the transition between the states $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$. These values are comparable to ATP-Binding Cassette (ABC) transporters of membrane lipids which use ATP hydrolysis (-7.54 $\pm$ 0.3 kcal/mol)[55] to drive lipid transport from the membrane to an extracellular acceptor. Some of these transporters share the same mechanism as SMO where the lipid from the inner leaflet is flipped and transported to the extracellular acceptor protein.[56]

The computed values of free energy barriers are dependent on the projections of the data. Upon projecting the data using the z-coordinate of cholesterol versus the angle between the cholesterol and the x-y plane (Fig. 4a) and the y-coordinate of cholesterol versus the x-coordinate (Fig. 4b), the barrier between $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$ is different. A reliable estimate of the barriers in such a case comes from using the time-lagged Independent Components (tICs), which project the entire dataset along the slowest kinetic degrees of freedom. On plotting the first two components of tICs, (Fig. **??**a,c), we observe that energetic barrier between $\boldsymbol{\eta}$

and $\boldsymbol{\theta}$ is $\sim$6.5 $\pm$ 0.8 kcal/mol. We further validate that the slowest degrees of freedom in the model correspond to the entry of cholesterol from the inner leaflet of the membrane into SMO TMD (Fig. **??**b,d).

Contrary to Pathway 1, the entry of cholesterol into the TMD happens via the alcoholic moiety first, which forms a hydrogen bond with the backbone oxygen of R421$^{5.64f}$ (Fig. 4e). Furthermore, the TM5 loses some of its helicity, between residues R421$^{5.64f}$-M424$^{5.67f}$ due to the flexibility provided by G422$^{5.65f}$, which is a part of the conserved WGM motif implicated in SMO activation.[52] Here, we also designate the "upper"and "lower"pocket, which coordinate with the isooctyl and androsterolic moieties of cholesterol, respectively (Fig. 4c-f). In the upper pocket, the isooctyl tail entry is blocked by the strong hydrophobic patch formed by Y417$^{5.60f}$, L419$^{5.62f}$, I420$^{5.63f}$, F457$^{6.38f}$ and F460$^{6.41f}$ (Fig. 4c,e). Flexible residues such as glycine are also not present that would allow for the fluctuations leading to opening of the hydrophobic patch and enable entry of the cholesterol tail. Therefore, the androsterolic moiety of cholesterol enters first followed by the isooctyl tail of cholesterol to reach the state $\boldsymbol{\epsilon}$. To stabilize and facilitate the androsterolic moiety's entry, F460$^{6.41f}$ and F455$^{6.36f}$ form $\pi-\pi$ contacts, and L419$^{5.62f}$, L452$^{6.33f}$ form hydrophobic contacts (Fig. 4d,f) in state $\boldsymbol{\theta}$.

Once cholesterol has flipped in $\boldsymbol{\theta}$, it allows for further translocation towards the TMD binding site ($\boldsymbol{\alpha}^*$) and finally arrival at the CRD site ($\boldsymbol{\zeta}$) with the hydroxyl group pointing extracellularly. The translocation of cholesterol from membrane to the CRD binding site via Pathway 2 is captured by a $\boldsymbol{\eta} \rightarrow \boldsymbol{\theta} \rightarrow \boldsymbol{\alpha}^* \rightarrow \boldsymbol{\zeta}$ transition. Overall, the angle between cholesterol and the x-y plane involves an entire cycle of -90°$\rightarrow$ 0°$\rightarrow$ 90° and cholesterol must flip, in order for cholesterol to enter the protein.

Interestingly, mutants along Pathway 2 showed a significant decrease in activity compared to Pathway 1 (Fig. 3a, 3d, **??**)), along with an increased thermodynamic barrier for translocation (Fig. 3d, **??**)). Mutating R421$^{5.64f}$ to alanine or glutamine did not decrease SMO activity significantly (Fig. 3d), because the interaction with cholesterol is mediated

by the protein backbone, and not the side-chain (Fig. 4e). However, mutations like F455I and L419A reduce SMO activity. Their expression levels of these mutants are comparable to the wild-type *mSMO* (Fig. **??**)). The mutants compared to WT SMO showed a significant increase in PMF, due to the lack of the hydrophobic π-stacking provided by F455$^{6.36f}$ and hydrophobic contacts provided by L419$^{5.62f}$ during cholesterol translocation. Overall, we report that the mutants for Pathway 2 show a decrease in the activity of SMO and show a strong correlation between the reduction in activity and the barrier for cholesterol translocation (Fig. 3b,d). These results validate the role of critical residues involved in cholesterol translocation from the inner leaflet as observed in the simulations.

On the basis of our experimental and computational data, we conclude that cholesterol translocation can happen via either pathway. This is supported on the basis of the following observations: mutations along pathway 2 affect SMO activity more significantly, and the presence of a direct conduit that connects the inner leaflet to the TMD binding site. However, since pathway 1 shows a lower thermodynamic barrier, PTCH1 controls cholesterol accessibility in the outer leaflet,[25] and there is no experimental structure with cholesterol in the inner leaflet region of SMO TMD. This shows that there is a possibility for transport from both leaflets.

The pathway connecting the TMD to the CRD binding sites shows off-pathway intermediates

Once cholesterol has reached the TMD binding site ($\boldsymbol{\alpha}^*$), it must translocate along the TMD-CRD interface to reach the orthosteric binding site in the CRD. This common pathway is shared by cholesterol molecules translocating from the inner and outer leaflets. The translocation of cholesterol from the TMD to the CRD binding site involves a linear movement of the androsterolic moeity through the extracellular end of the TMD, where cholesterol maintains a primarily upright position, with the polar androsterolic moeity pointing towards the CRD site. The energetic barrier associated with the transition ($\boldsymbol{\alpha}^*$(TMD site) $\rightarrow$ $\boldsymbol{\zeta}$(CRD site))

can be visualized by plotting the z-coordinate of cholesterol versus the angle it forms with the x-y plane (Fig. 2a, 4a). We observe that the highest barrier along the common pathway is $\sim 4.1 \pm 0.3$ kcal/mol, which is lower than the highest energetic barrier for cholesterol entry from the inner and outer leaflet. Another interesting observation is that the stability of the cholesterol in the TMD binding site ($\boldsymbol{\alpha}^*$) is higher than the CRD binding site ($\boldsymbol{\zeta}$). This is counter-intuitive, as the CRD binding site is considered the orthosteric binding site of SMO.[35] This observation can be explained on the basis of the thermodynamic driving forces in the two pockets. The TMD binding site is composed mainly of hydrophobic residues, which forms strong interactions with cholesterol (Fig. **??**)) in contrast to the CRD binding site, where the cholesterol is exposed to the solvent. The CRD binding site being exposed to the solvent increases the conformational entropy associated with the cholesterol compared to the restrained TMD binding site. In addition, the CRD binding site is more flexible than the TMD binding site in the absence of cholesterol as reported previously.[35] This increased flexibility of the CRD binding site further leads to formation of multiple conformational states between the CRD and cholesterol.

Once cholesterol reaches the TMD-CRD interface, it can adopt multiple poses before reaching the CRD binding site (Fig. 2a, 4a). Cholesterol at this position can be upright ($\boldsymbol{\delta}$, Fig. 2a, 4a), where it interacts with F484$^{6.65f}$, W480$^{6.61f}$, V488$^{6.69f}$ and L221^{LD}, forming hydrophobic contacts (Fig. 5a). However, there exists a thermodynamic barrier to take a completely upright path to the CRD binding site (angle $\sim +90°$). This is due to the presence of the long beta sheet of the linker domain (residues L197^{LD}-I215^{LD}) of SMO at the TMD-CRD interface, blocking the direct upright translocation of cholesterol (Fig. **??**)). Hence, the major conformation of cholesterol at this position is slightly tilted, away from the plane of the membrane ($\boldsymbol{\gamma}$, Fig. 2a, 4a). Additionally, in $\boldsymbol{\gamma}$, Y207^{LD} creates a hydrophobic interaction with cholesterol, stabilizing the bent pose (Fig. 5b). $\boldsymbol{\gamma}$ has been identified as the binding site of the synthetic agonist SAG in SMO.[44] This provides further validation for $\boldsymbol{\gamma}$ being the major intermediate state in the common pathway.

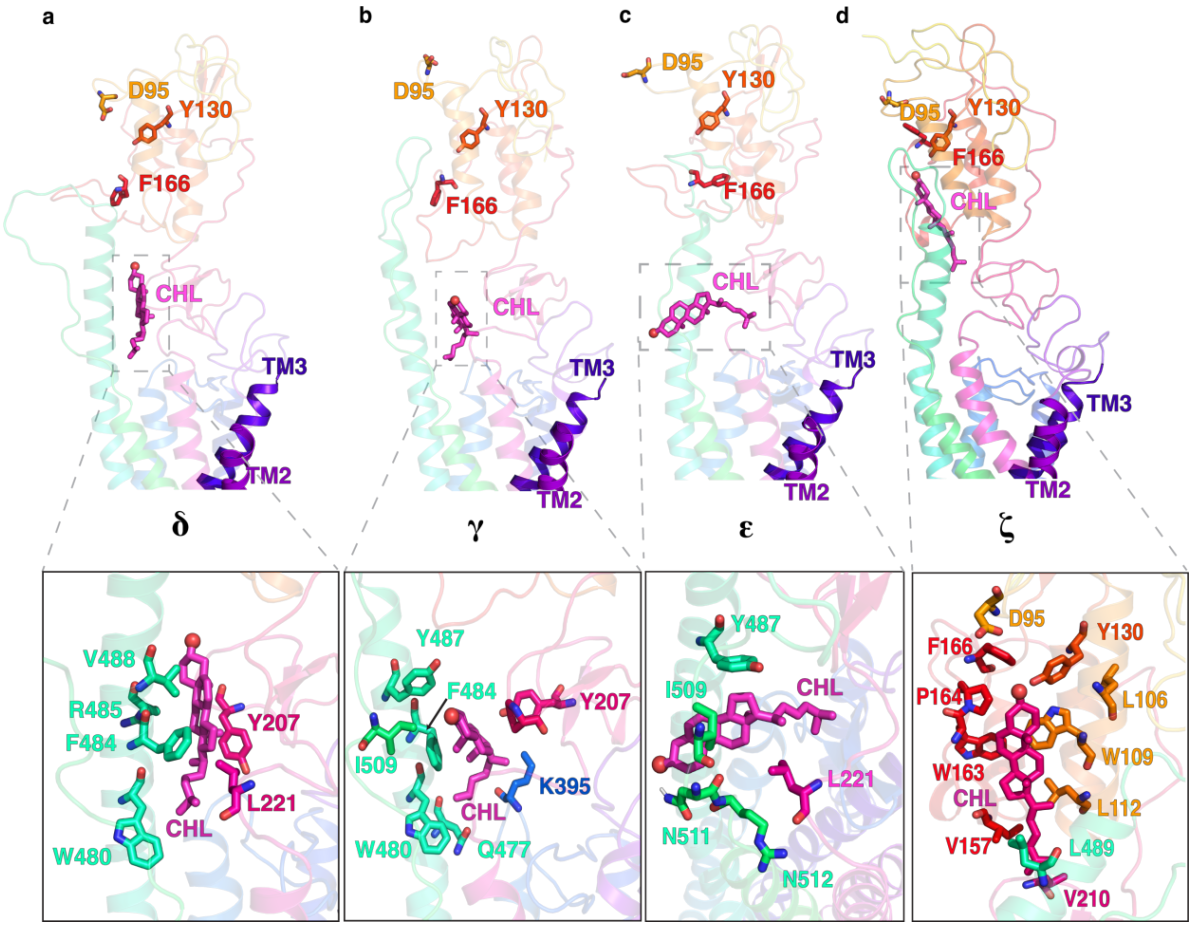

Figure 5: Multiple positions of cholesterol as it translocates through the common pathway, including the off-pathway intermediate. (a) upright ($\boldsymbol{\delta}$) (b) tilted ($\boldsymbol{\gamma}$) (c) overtilted ($\boldsymbol{\epsilon}$), the off pathway intermediate and (d) Cholesterol at the CRD binding site ($\boldsymbol{\zeta}$).

Since the degrees of freedom accessible to cholesterol at this point in the pathway are higher than at the TMD, cholesterol can can undergo "overtilting "as it approaches the CRD ($\boldsymbol{\epsilon}$, Fig. 5c). This state ($\boldsymbol{\epsilon}$) attributes to an off-pathway intermediate state in the cholesterol translocation process, raising the timescales required for cholesterol to translocate to the CRD binding site. This overtilted pose is stabilized by hydrophobic interactions between the sterol and Y487$^{6.78f}$, L221^{LD} and I509^{ECL3}, and a hydrogen bond between the side-chain of N511^{ECL3} and the alcoholic oxygen (Fig. 5c). Once the cholesterol has crossed the TMD-CRD interface, it can reach the orthosteric binding site ($\boldsymbol{\zeta}$, Fig. 5d) in the CRD. In summary, we identify numerous conformational states of cholesterol bound SMO which are distinct from the available structures of sterol bound SMO.

We performed experimental mutagenesis to validate the critical residues along this pathway, from the TMD to the CRD binding site. These mutants - Y207A, F484A and I509A, all showed a significant decrease in activity compared to WT SMO (Fig. 6a,b, **??**)). When compared to the difference in PMFs induced by the mutants, we observe that the mutants consistently show a higher peak PMF, suggesting that the force required to translocate cholesterol is higher than the wild-type residue in this position (Fig. 6b, **??**)). This further implies that the mutants reduce the activity of SMO by increasing the energetic barriers for cholesterol translocation. In particular, the effect is pronounced for Y207A, which forms hydrophobic contacts along the pathway (Fig. 6c). F484A showed a significant decrease in SMO activity, which can be attributed to reduction in hydrophobic stabilizing contacts that enable cholesterol's entry to the CRD (Fig. 6d).

Overall, the entire cholesterol translocation process can be divided into two parts - entry from the membrane to the TMD binding site via Pathway 1 or 2, and translocation from the SMO TMD to CRD binding site via common pathway. The entire process is mapped out in Fig. **??**.. The experimental and computational analysis shows that both Pathway 1 and Pathway 2 are thermodynamically feasible pathways and exhibit similar energetic barriers for cholesterol to take from the membrane to the CRD binding site.

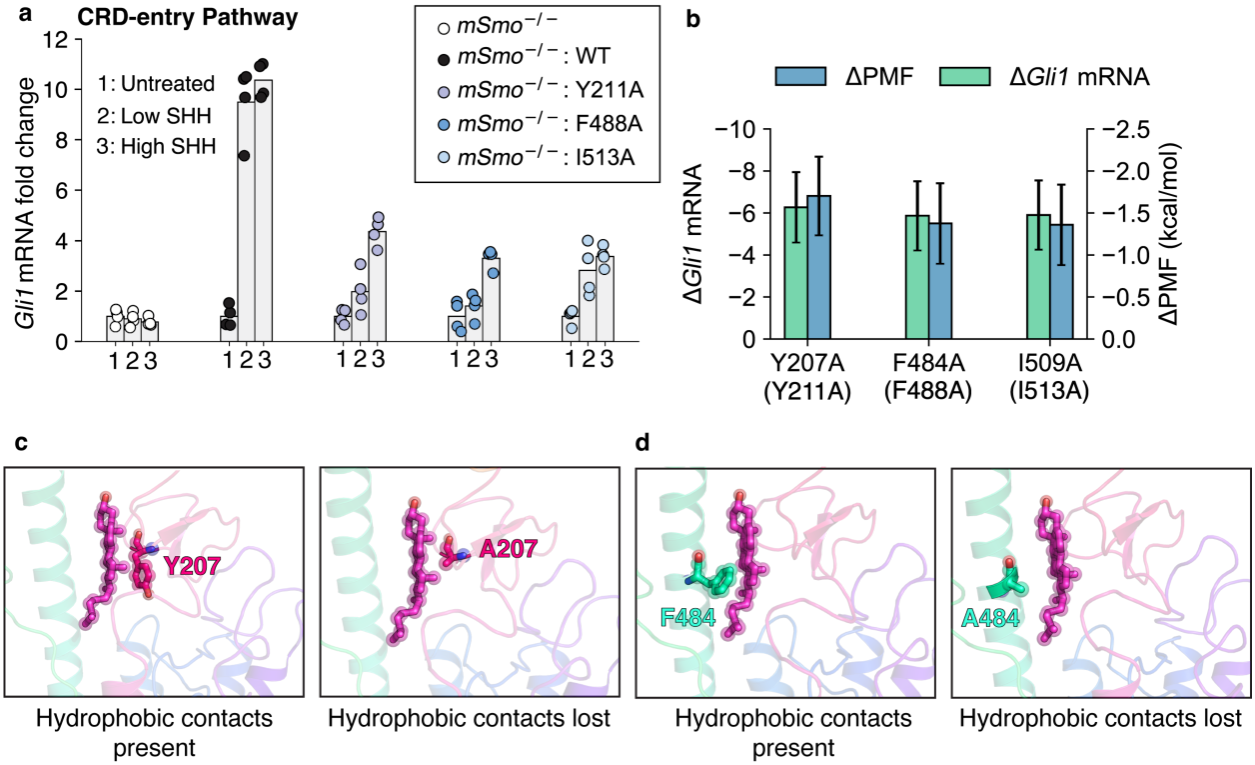

Figure 6: Effects of mutations along the translocation pathway connecting the TMD and CRD binding sites on activation of SMO. (a) *Gli1* mRNA fold change (indicator of SMO activation) plotted for SMO mutants, showing fold change when the mutants are untreated, treated with a low concentration of SHH and treated with a saturating concentration of SHH. (b) Δ*Gli1* mRNA fold change (high SHH vs untreated) and Δ PMF (difference of peak PMF) plotted for mutants along the TMD-CRD pathway. *hSMO* numbering is used as default, with *mSMO* numbering given in parentheses. (c, d) Example mutants Y207A and F484A show that cholesterol is unable to translocate through this pathway because of the loss of crucial hydrophobic contacts provided by Y207 and F484 and along the solvent exposed pathway.

A squeezing mechanism for translocation of cholesterol in SMO

To further elucidate the structural changes that happen in SMO during cholesterol translocation, we sought to characterize the conformation of the hydrophobic tunnel inside SMO. We calculated the tunnel diameter along the channel as cholesterol traverses through the protein (Fig. 7 and Fig. **??**)). The tunnel calculations were done with cholesterol at different points along the pathway as indicated by the metastable states in the Markov state models - with cholesterol in the TMD binding site, cholesterol at the CRD-TMD interface,

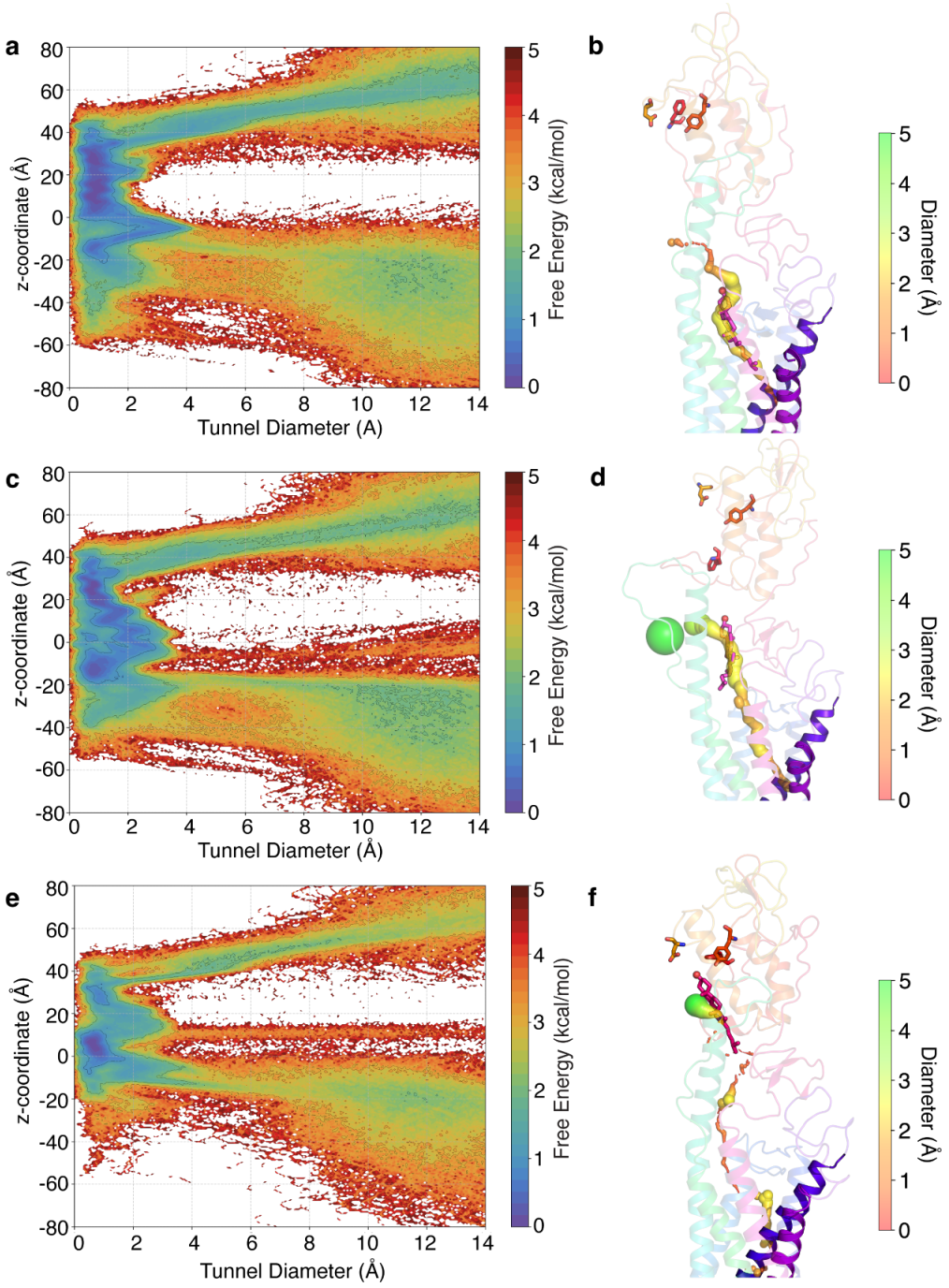

Figure 7: The tunnel profile during cholesterol translocation in SMO. (a) Free energy plot of the z-coordinate v/s the tunnel diameter when cholesterol is present in the core TMD. The tunnel shows a spike in the radius in the TMD domain, indicating the presence of a cholesterol accommodating cavity. (b) Representative figure for the tunnel when a cholesterol is in the TMD. (c) Same as (a), when cholesterol is at the TMD-CRD interface. (e) same as (b), when cholesterol is at the TMD-CRD interface. (e) same as (a), when cholesterol is at the CRD binding site. (f) same as (b), when cholesterol is at the CRD binding site. Tunnel diameters shown as spheres.

and cholesterol present at the binding site in the CRD. When we plot cholesterol's position at different points along its transit through the TMD, we observed that the tunnel diameter varies along the z-coordinate as cholesterol moves through the channel. This corroborates that the tunnel radius is a function of the position of cholesterol in SMO. Furthermore, we observe that the tunnel radius shows a peak at z $\sim$ -7 when cholesterol is present in the core of the TMD (Fig. 7 a,b). In the rest of the tunnel, the average radius remains relatively small (Fig. **??**)). This peak in diameter is seen to move along with the cholesterol position (Fig. 7 c,d). This provides evidence that SMO uses a squeezing mechanism to translocate cholesterol. The term 'squeezing' here implies that the tunnel remains open only around cholesterol and it closes as cholesterol moves away from its current position in the tunnel. Once cholesterol has reached the binding site in the CRD, the tunnel in the TMD region is closed (Fig. 7 e,f, Fig. **??**)).

The cholesterol translocation mechanism of SMO shows similarities to the alternating access model proposed for substrate transport in membrane transporter proteins - where the transport tunnel closes behind the substrate to facilitate substrate transport across membrane. Membrane transporters including lipid transporters use either an ion-based gradient or ATP hydrolysis to facilitate the substrate transport. An ion binding site for SMO is currently unknown in contrast to Class A receptors, which have a sodium binding site.[57–59] There is no experimental evidence of ion-coupling with the cholesterol export via SMO. Therefore, we posit that cholesterol translocation through SMO involves passive or concentration-dependent diffusion driven by a shift in the pool of accessible cholesterol, which rises once PTCH1 is inhibited. Thus, we provide a mechanistic overview of the dynamics of the cavity during cholesterol translocation in SMO.

The translocation of cholesterol occurs on a millisecond timescale

To give a perspective on the overall translocation process from a kinetic standpoint, we sought to calculate the timescales associated with the translocation of cholesterol from the

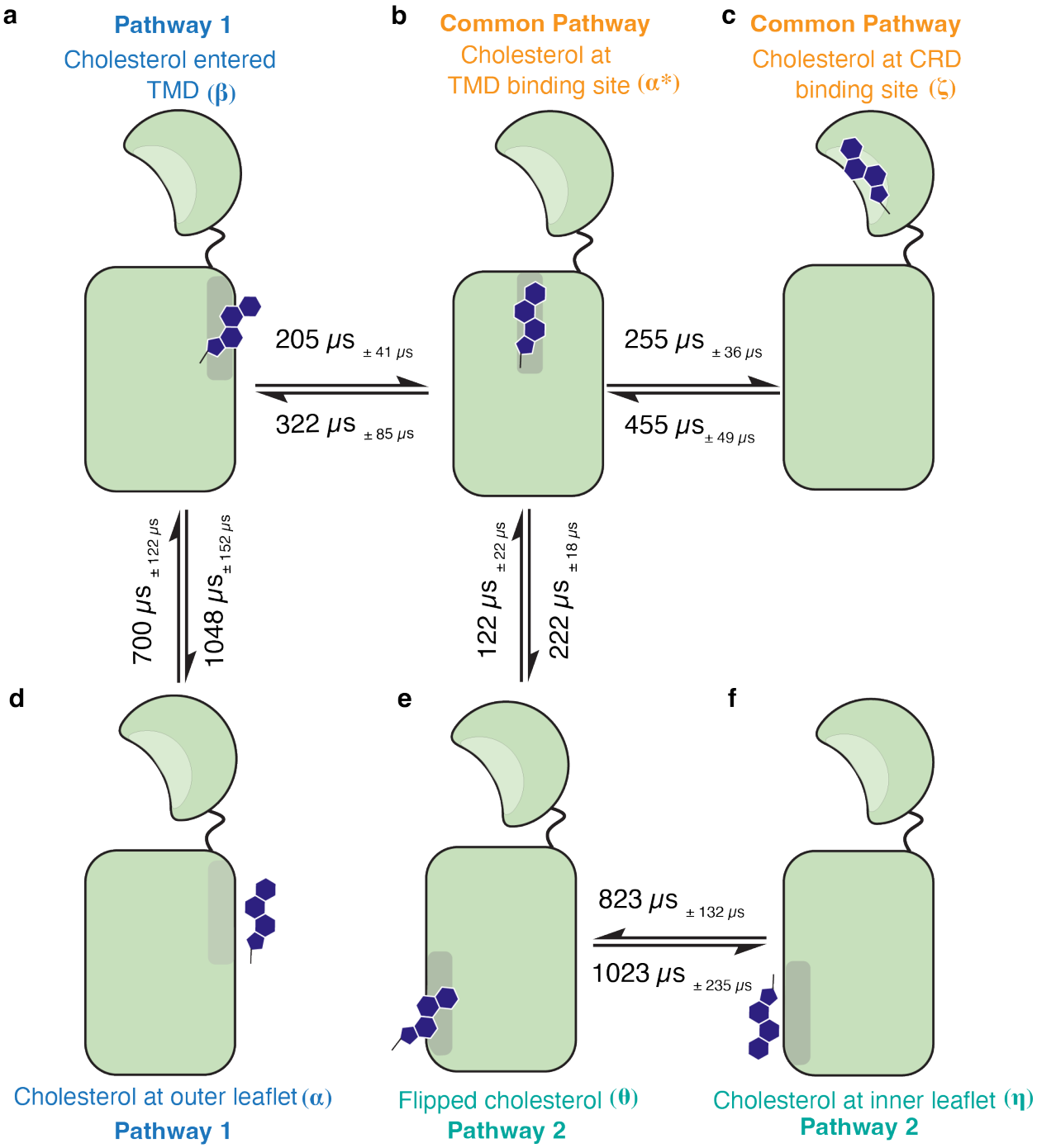

Figure 8: The timescales associated with the translocation of cholesterol through SMO. Each major intermediate state has been marked (a-f). Timescales were obtained by calculating the mean first passage time (MFPT) using the markov state model. Errors in timescales are shown as subscripts. The arrows represent the relative flux for the translocation between subsequent steps. The overall process occurs at a timescale of $\sim$ 1ms.

membrane to the binding site in the CRD. Using a combination of Transition Path Theory and Markov state models, the reactive fluxes associated with each stage of the translocation cycle can be calculated. This enables us to calculate the mean first passage time (MFPT), which gives us an estimate of the timescales associated with the process (Fig. 8). The Pathway 1 was divided into two stages - the first being the translocation of cholesterol from the outer leaflet (Fig. 8d) to the TMD - yielding a mean first passage time of 700 $\pm$ 122 μs. Here, the cholesterol tail has entered in the TMD (Fig. 8a). This was followed by cholesterol reaching the TMD binding site, with a timescale of 205 $\pm$ 41 μs (Fig. 8b). For Pathway 2, cholesterol starts in the inner leaflet (Fig. 8f). In the first stage, cholesterol undergoes flipping which leads to a higher timescale than stage 1 of Pathway 1 - 823 $\pm$ 132 μs. This is followed by 122 $\pm$ 22 μs for cholesterol to reach the TMD binding site (Fig. 8b).

Once cholesterol has reached the TMD binding site, it is followed by translocation of cholesterol from the TMD to the CRD binding site with a timescale of 255 $\pm$ 36 μs. Overall, the calculated timescales for cholesterol to reach the CRD site from the inner/outer leaflet of the membrane are 1023 $\pm$ 223 μs (Pathway 1) and 1134 $\pm$ 188 μs (Pathway 2). These timescales are comparable to the substrate transport timescales of Major Facilitator Superfamily (MFS) transporters.[60] Interestingly, the timescales for the reverse process of translocating cholesterol from the CRD binding site to the membrane are higher for each step (Fig. 8) indicating that the reverse process has higher energetic barrier associated with it. This provides kinetic evidence for the overall translocation process from membrane to the CRD binding site is thermodynamically favorable. Thus, it can be concluded that SMO facilitates the translocation of cholesterol.

Conclusions

In this study, we have used a combination of millisecond-scale atomistic molecular dynamics simulations, Markov state modeling and experimental mutagenesis to describe the step-

by-step process of cholesterol translocation through SMO. Previous structural studies have delineated multiple translocation pathways for cholesterol transport via SMO. In this study, we have examined the mechanism of cholesterol translocation from the membrane to the orthosteric binding site in the CRD of SMO via two pathways - the outer leaflet pathway, between TM2 and TM3, and the inner leaflet pathway, between TM5 and TM6. We quantitatively assess the thermodynamic viability of the translocation pathways and estimate the timescales associated with the process. The key intermediate cholesterol bound conformations of the SMO and the role of specific residues in the translocation process were identified computationally and validated using both experimental and *in silico* mutagenesis.

We observe that cholesterol moves through a conduit, starting from the outer leaflet in the membrane, up to the cholesterol binding site in the CRD. In the first pathway, cholesterol enters the protein between TM2 and TM3 in the outer leaflet and then it is translocated along the extended TM6 to reach the CRD binding site. The highest barrier along the first pathway is associated with the cholesterol entry from the outer leaflet of the membrane to the TMD binding site. Similarly, the highest barrier for the second pathway also involves cholesterol translocation from the inner leaflet membrane to the TMD of SMO. We show that the barriers associated with these two pathways are comparable, and that cholesterol can enter SMO via both leaflets. The second highest barrier in cholesterol translocation is at the TMD-CRD interface but it is $\sim$2 kcal/mol lower than the barrier for cholesterol entry. We also show that the cholesterol translocation process occurs via a squeezing mechanism that maintains the forward flux of cholesterol from membrane to the CRD binding site. Overall, the translocation process takes place on the millisecond timescale with multiple intermediate states identified using simulations and validated using the sterol bound structures of SMO.

Despite the extensive MD simulations reported in this study, there is still a need to further probe the endogenous "activation" of SMO by cholesterol in a position dependent manner. GPCRs exist in a conformational equilibrium of the active and inactive states and the job of the agonist (cholesterol) is to lower the barrier to activation and shift the

equilibrium towards the active state. Our recent work[61] discusses how the binding position of cyclopamine modulates SMO activity. Cyclopamine acts as an antagonist when bound to SMO TMD and acts as an agonist when bound to SMO CRD. Cholesterol binding at different positions along the translocation pathway leads to a position dependent modulation of SMO activity.[35,62] Kinnebrew et al. have shown that binding of cholesterol to the TMD leads to basal SMO activity, whereas binding to the CRD site leads to 'medium' activity. They have also propose that cholesterol binding to both the sites leads to 'high' activity. In this study, we only focused on the cholesterol movement from the membrane to CRD binding site. Therefore, a future investigation is needed to fully sample the activation process of SMO with cholesterol bound at different positions.

Our study, however, provides computational and experimental evidence for the translocation of cholesterol via SMO. We outline the entire translocation mechanism from a kinetic and thermodynamic perspective. These findings underscore SMO's role as a unique and nuanced regulator that undergoes activation by translocating cholesterol. We provide an enhanced understanding of the endogenous regulation of HH signaling via SMO, and provide a framework for developing drugs that target SMO by outlining the intermediate states along the translocation process, to counter aberrant SMO activity.

Methods

Molecular Dynamics (MD) Simulations

Simulation setup

Structures of SMO bound to sterols - 6XBL[44] (SMO-CHL-1), 6XBM[44] (SMO-CHL-2, SMO-CHL-3) and 5L7D[18] (SMO-CHL-4) were used as the starting structures for simulations. For SMO-CHL-1, the bound agonist SAG was removed. The two sterols occupying different positions in the tunnel in 6XBM were used to build 2 separate systems with cholesterol at

different sites in the pathway (SMO-CHL-2, SMO-CHL-3). For SMO-CHL-2 and SMO-CHL-3, to account for the lack of CRD in the structure 6XBM, the sterol positions were aligned to the full-length SMO (6XBL, 0.8 Å RMSD from 6XBM) and the 24(S),25-epoxycholesterol for each system was replaced by cholesterol. For SMO-CHL-4, The inactivating mutation V329F was mutated back to Wild Type. For all systems, any stabilizing antibodies and bound G proteins were removed. The missing residues in the intra/extracellular loop for every protein were modeled using MODELLER[63] (Table **??**)). Termini for all proteins were capped using acetyl (ACE) and N-methylamino (NME) at the N- and C-termini to ensure neutrality. All four protein systems with cholesterols at different points along the translocation path were embedded in a lipid bilayer. The composition of the bilayer was set similar to mice cerebellum[64] (Table **??**)), to mimic physiological conditions, using CHARMM-GUI.[65,66] Interactions between the atoms - bonded and non-bonded, were modeled using the CHARMM36 Force Field.[67,68] TIP3P water[69] and 0.15 M NaCl was used to solvate the system, to mimic physiological conditions. Non-protein hydrogen masses were repartitioned to 3.024 Da, to enable use of a longer timestep (4fs).[70] Starting points for Pathways 1 and 2 were chosen from already simulated data, according to the closest cholesterol distance from the respective helices (outer leaflet, TM2-TM3 for Pathway 1, lower leaflet, TM5-TM6 for Pathway 2). This was done once the rest of the pathway was completely explored.

Pre-Production MD

All systems were subject to 50,000 steps of initial minimization. Further, minimization was performed for another 10000 steps by constraining the hydrogens using SHAKE.[71] Systems were then heated to 310K to mimic physiological conditions, at NVT for 10 ns. This was followed by equilibration at NPT and 1 bar for 5 ns. Backbone constraints of 10 kcal/mol/$Å^2$ were applied during NVT and NPT. Next, systems were equilibrated at NPT for 40 ns without constraints to ensure system stability. All pre-production steps were performed using AMBER18.[72–76]

Production MD

Systems were then subjected to some initial sampling of 100ns each. This was followed by clustering and performing adaptive sampling, to enable divide-and-conquer approach to sample the conformational landscape. 3 rounds of sampling was performed on each system. This was followed by similar rounds of adaptive sampling on the distributed computing project Folding@Home (http://foldingathome.org). OpenMM 7.5.1[77] was used for running simulations on Folding@Home.

In all simulations, 4fs was the chosen integration timestep. Particle Mesh Ewald (PME)[78] method was used to account for long-range electrostatics. The cutoff for considering non-bonded interactions was set to 10 Å. Temperature was maintained using the Langevin Thermostat.[79] Pressure was maintained using the Monte-Carlo Barostat.[80] All hydrogen bonds were constrained using SHAKE.[71]

Steered MD

Steered MD was performed to steer cholesterol from the membrane into the protein, for generating the starting frames for Pathway 1 and Pathway 2. This was done by first finding frames from the existing data, where cholesterols were closest to the respective starting points. Then, cholesterols were steered towards the center of TMD with the end point being cholesterol binding site deep in the TMD, as resolved in the structure 6XBL. This was done using a steering force of 20 kcal/(mol $Å^2$), over a course of 500ns. The entire protein except the helices involved at the entry (TM2-TM3 : Pathway 1; TM5-TM6 : Pathway 2), were constrained using two RMSD restraints during simulations to prevent any unphysical effects. The force constants used for RMSD restraints are as follows: 10 kcal/(mol $Å^2$) for restraining residues from CRD to ICL1, and 35 kcal/(mol $Å^2$) for restraining residues from ICL2 to helix8 for Pathway 1. For Pathway 2, an RMSD restraint of 35 kcal/(mol $Å^2$) was used to restrain residues from CRD to ECL2, and 10 kcal/(mol $Å^2$) to restrain residues from ECL3 to helix8. Three replicates of steered MD were performed to ensure that the pathway

explored converged. The frames generated from these runs were then used as seed frames to start simulations from for exploring the entry of cholesterol into SMO. Steered MD was performed using NAMD.[81,82] Each frame generated by Steered MD was minimized for 50000 steps, and then equilibrated for 40 ns, using AMBER using the same methodology for these seed frames as the described in section Pre-Production MD.

Adaptive Biasing Force-Based Sampling for PMF generation

To elucidate the effect of mutations on the cholesterol translocation barriers, we used an adaptive-biasing force (ABF) based sampling for generating the potential of mean force (PMF) profiles for each case, and compared it to the Wild Type (WT) translocation barriers. For generating the starting files for every mutant system, psfgen[83,84] was used. SMO WT PMF profiles were generated for each pathway, and mutant profiles were generated for their respective pathways. A biasing potential of 45 kcal/(mol Å^2) was used for both lower wall and upper wall of the ABF potential. Each mutant was run for 3 replicates, and 10^7 samples were generated for each pathway to compute the PMF. NAMD was used for this purpose.[81,82]

Adaptive sampling, feature selection and clustering

The cholesterol translocation process in SMO was simulated in a stage-wise process - with the four starting points having cholesterol present at different points along the translocation pathway (Fig. 1). To accelerate the sampling of the entire translocation process, simulations were performed using an approach that parallelizes the exploration of conformational space. Adaptive sampling[85–87] was utilized to achieve this acceleration - which uses an iterative sampling approach involving picking the next round of simulation starting points based on the current data. Several different types of machine learning and heuristic based adaptive sampling approaches have been proposed in the literature.[87–93] However, least count based sampling where the least visited states are chosen as the starting points for the next set of simulations has been shown to be among the best sampling strategies for exploration of the

conformational free energy landscapes.[93,94] The following approach was used to collect the data:

1. Initially, sampling was started from the 4 starting points with cholesterol at different points along the tunnel (Fig. 1). Each starting point was simulated as a separate system.

2. All systems were subject to Pre-Production MD (Refer Pre-Production MD).

3. Following Pre-Production, all systems were subject to 200 ns of Production MD (Refer Production MD).

4. The Production data until this point was combined and clustered on the basis of Adaptive Sampling metrics (Table **??**)) using k-means clustering (for the first 4 rounds) and then mini-batch k-means later on. pyEMMA[95] was used for this purpose.

5. Frames were chosen from the clusters with the least populations, and were used as seeds for the next round of simulations. This enabled a parallel iterative approach for sampling the conformational landscape (Table **??**,, Fig. **??**))

6. Steps 4-5 were repeated until the entire landscape was explored.

The entire process of cholesterol translocation - from cholesterol present in the membrane to cholesterol bound to orthosteric binding site in the extracellular domain - was simulated. Total 1.5 milliseconds of unbiased simulation data was collected using this approach.

Dimensionality reduction using tICA

Dimensionality reduction was performed on the high-dimensional data set using Time-Lagged Independent Component Analysis (tICA).[96] tICA uses a linear transformation to project the input data into a lower-dimensional basis set, the components of which approximate the slowest processes observed in the simulations. In our case, the slowest process simulated was

identified by tICA as the translocation of cholesterol (Fig. **??**)). The dataset was divided into 2 different groups - separately for Pathways 1 and 2. To further gain insights from the simulations, we constructed a Markov state model from the tICA-projected simulation data.

Markov state model construction

Markov State Model (MSM) is a kinetic modeling technique that uses short trajectory data that sample local transitions to provide a global estimate of the thermodynamics and kinetics of the physical process.[97–102] A MSM discretizes the dataset into kinetically distinct microstates, and calculates the rates of transitions among such microstates. This methodology has been used extensively to investigate the conformational dynamics of membrane proteins including G-protein coupled receptors.[52,59–61,103–106] To construct the Markov state model, the data collected from simulations was first featurized (Table **??**)). Pathways 1 and 2 were treated independently of each other. Time-lagged independent component analysis (tICA) was performed on the data to reduce the dimensionality of the input data and identify the slowest processes observed in the simulations. A Markov state model was then constructed on the 42 and 28 components from the tIC space for Pathways 1 and 2 respectively. To construct the models for both the pathways, the following approach was used:

1. The data was clustered into different number of clusters, ranging from 100-1000 and the implied timescales were calculated as a function of the MSM lag time. The lag time after which the implied timescales converged (30ns) in both cases, was chosen as the MSM lag time (Fig. **??**aa, **??**aa).

2. Once the MSM lag times were chosen, a grid-search based approach was chosen to compute the optimal number of clusters. Each constructed MSM was evaluated using a VAMP2 score,[107] where the sum of squares of top five eigenvalues was computed. Each pathway's dataset was clustered into number of clusters (200 - Pathway 1, 400 - Pathway 2) that gave the highest VAMP2 score (Fig. **??**bb, **??**bb).

3. Once the MSMs were constructed, bootstrapping was performed, with 200 rounds and 80% of the data in each round, to compute the error associated with the probabilities.

4. To validate the final MSM, a Chapman-Kolmogorov test was performed (Fig. **??**,, **??**)) to show the long-timescale validity of the Markovian property followed by the constructed models.

This information can be used to calculate the probability of each state, which can be used to recover the thermodynamic and kinetic properties of the entire ensemble. One of the use cases of the probabilities is that they can be used to reweigh the projected free energy of each datapoint along a said reaction co-ordinate, which has been used in Fig. 2, 4, 5, 7, **??**.. The errors in free energies were computed using the errors in the projected probabilities from the bootstrapped MSMs (Fig. **??**,, **??**,, **??**,, **??**))

Trajectory Analysis and Visualization

Trajectories were stripped of waters and imaged before analysis to allow faster computation. cpptraj[108] was used for this purpose. For constructing the figures, VMD[83,84] and open-source PyMOL[109] (rendering) MDTraj[110] (computing observables from trajectories), matplotlib[111] and seaborn[112] (Plot rendering), Numpy[113] (numerical computations), HOLE[114] (tunnel diameter calculations) were used.

Cell culture and cell line generation

Mouse embryonic fibroblasts (MEFs) lacking Smoothened ($Smo^{-/-}$) were tested to ensure lack of endogenous SMO protein using immunoblotting, as described previously.[9] These $Smo^{-/-}$ MEF cells were used to generate stable cell lines expressing SMO mutants, which were then authenticated by immunoblotting to ensure stable expression of the transgene.[40] Cell lines were confirmed to be negative for mycoplasma infection.

MEF cells were grown in high-glucose DMEM (Thermo Fisher Scientific, catalog no. SH30081FS) containing 10% FBS (Sigma-Aldrich, catalog no. S11150) and the following supplements: 1 mM sodium pyruvate (Gibco, catalog no. 11-360-070), 2 mM L-glutamine (GeminiBio, catalog no. 400106), 1× minimum essential medium NEAA solution (Gibco, catalog no. 11140076), penicillin (40 U/ml), and streptomycin (40 μg/ml) (GeminiBio, catalog no. 400109). This media, hereafter referred to as supplemented DMEM, was sterilized through a 0.2-μm filter and stored at 4°C.

To measure Hedgehog responsiveness by quantitative polymerase chain reaction (PCR) or Western blotting, cells were seeded in 10% FBS-supplemented DMEM and grown to confluence. To induce ciliation, a requirement for Hedgehog signaling, cells were serum-starved in 0.5% FBS supplemented DMEM and simultaneously treated with Sonic Hedgehog (SHH) for 24 hours before analysis.

Western blotting was carried out to assess SMO protein expression for all mutants. Briefly, whole-cell extracts were prepared in lysis buffer containing 150 mM NaCl, 50 mM Tris-HCl (pH 8), 1% NP-40, 1× protease inhibitor (SigmaFast Protease inhibitor cocktail, EDTA-free; Sigma-Aldrich, catalog no. S8830), 1 mM $MgCl_2$, and 10% glycerol. After lysate clarification by centrifugation at 20,000g, samples were resuspended in 50 mM tris(2-carboxyethyl)phosphine and 1× Laemmli buffer for 30 min at 37°C. Samples were then subjected to SDS-polyacrylamide gel electrophoresis, followed by immunoblotting with antibodies against GLI1 [anti-GLI1 mouse monoclonal (clone L42B10); Cell Signaling Technology, catalog no. 2643, RRID: AB_2294746], SMO (rabbit polyclonal),[9] or GAPDH [anti-GAPDH mouse monoclonal (clone 1E6D9); Protein tech, catalog no. 60004-1-Ig, RRID: AB_2107436].

Measuring Hedgehog signaling with quantitative PCR

Gli1 mRNA transcript levels were measured using the Power SYBR Green Cells-to-CT kit (Thermo Fisher Scientific). *Gli1* levels relative to *Gapdh* were calculated using the Delta-Ct method (CT(*Gli1*) - CT(*Gapdh*)). The RT-PCR was carried out using custom primers for

Gli1 (forward primer: 5′-ccaagccaactttatgtcaggg-3′ and reverse primer: 5′-agcccgcttctttgttaatttga-3′), and *Gapdh* (forward primer: 5′-agtggcaaagtggagatt-3′ and reverse primer: 5′-gtggagtcatactggaaca-3′).

Acknowledgements

The authors thank The Blue Waters Petascale Computing Facility and National Center for Supercomputing Applications, which is supported by the National Science Foundation (awards OCI-0725070 and ACI-1238993) and the state of Illinois. Blue Waters is a joint effort of the University of Illinois at Urbana-Champaign and its National Center for Supercomputing Applications. D.S. acknowledges support from NIH grant R35GM142745 and Cancer Center at Illinois. R.R. acknowledges support from NIH grant GM118082. The authors also acknowledge support from citizen scientists providing compute hours for simulations performed Folding@Home, which have enabled us to collect data on a large scale for this project.

Author Contributions

D.S. and R.R. acquired funding for the project. D.S. supervised the project. D.S. and P.B. designed the research. P.B. performed simulations and analyzed the computational data. M.K. designed and performed experiments. P.B. wrote the manuscript with inputs from M.K., R.R. and D.S..

Competing interests

The authors declare no competing interests.

Data Availability

Relevant trajectory files have been shared via Box (https://uofi.box.com/s/zldrqparb39ax72c637amcxi43hv2yaq) and are available publicly. Python scripts uses for analysis are available on Github (https://github.com/ShuklaGroup/Bansal_et_al_Cholesterol_Smoothened_2024).

References

(1) Wang, C.; Wu, H.; Katritch, V.; Han, G. W.; Huang, X.-P.; Liu, W.; Siu, F. Y.; Roth, B. L.; Cherezov, V.; Stevens, R. C. Structure of the human smoothened receptor bound to an antitumour agent. *Nature* **2013**, *497*, 338–343.

(2) Wang, C.; Wu, H.; Evron, T.; Vardy, E.; Han, G. W.; Huang, X.-P.; Hufeisen, S. J.; Mangano, T. J.; Urban, D. J.; Katritch, V.; Cherezov, V.; Caron, M. G.; Roth, B. L.; Stevens, R. C. Structural basis for Smoothened receptor modulation and chemoresistance to anticancer drugs. *Nature Communications* **2014**, *5*, 4355.

(3) Raleigh, D. R.; Reiter, J. F. Misactivation of Hedgehog signaling causes inherited and sporadic cancers. *Journal of Clinical Investigation* **2019**, *129*, 465–475.

(4) Axelson, M.; Liu, K.; Jiang, X.; He, K.; Wang, J.; Zhao, H.; Kufrin, D.; Palmby, T.; Dong, Z.; Russell, A. M.; Miksinski, S.; Keegan, P.; Pazdur, R. US Food and Drug Administration approval: vismodegib for recurrent, locally advanced, or metastatic basal cell carcinoma. *Clinical Cancer Research* **2013**, *19*, 2289–2293.

(5) Jain, S.; Song, R.; Xie, J. Sonidegib: mechanism of action, pharmacology, and clinical utility for advanced basal cell carcinomas. *OncoTargets and Therapy* **2017**, *Volume 10*, 1645–1653.

(6) Meani, R. E.; Lim, S.-W.; Chang, A. L. S.; Kelly, J. W. Emergence of chemoresistance

in a metastatic basal cell carcinoma patient after complete response to hedgehog pathway inhibitor vismodegib (GDC-0449). *Australasian Journal of Dermatology* **2014**, *55*, 218–221.

(7) Nüsslein-Volhard, C.; Wieschaus, E. Mutations affecting segment number and polarity in Drosophila. *Nature* **1980**, *287*, 795–801.

(8) Ingham, P. W. *Cell Signaling Pathways in Development*; Elsevier, 2022; pp 1–58.

(9) Rohatgi, R.; Milenkovic, L.; Scott, M. P. Patched1 Regulates Hedgehog Signaling at the Primary Cilium. *Science* **2007**, *317*, 372–376.

(10) Corbit, K. C.; Aanstad, P.; Singla, V.; Norman, A. R.; Stainier, D. Y. R.; Reiter, J. F. Vertebrate Smoothened functions at the primary cilium. *Nature* **2005**, *437*, 1018–1021.

(11) Cooper, M. K.; Wassif, C. A.; Krakowiak, P. A.; Taipale, J.; Gong, R.; Kelley, R. I.; Porter, F. D.; Beachy, P. A. A defective response to Hedgehog signaling in disorders of cholesterol biosynthesis. *Nature Genetics* **2003**, *33*, 508–513.

(12) Stottmann, R. W.; Turbe-Doan, A.; Tran, P.; Kratz, L. E.; Moran, J. L.; Kelley, R. I.; Beier, D. R. Cholesterol Metabolism Is Required for Intracellular Hedgehog Signal Transduction In Vivo. *PLoS Genetics* **2011**, *7*, e1002224.

(13) Nachtergaele, S.; Mydock, L. K.; Krishnan, K.; Rammohan, J.; Schlesinger, P. H.; Covey, D. F.; Rohatgi, R. Oxysterols are allosteric activators of the oncoprotein Smoothened. *Nature Chemical Biology* **2012**, *8*, 211–220.

(14) Myers, B. R.; Sever, N.; Chong, Y. C.; Kim, J.; Belani, J. D.; Rychnovsky, S.; Bazan, J. F.; Beachy, P. A. Hedgehog Pathway Modulation by Multiple Lipid Binding Sites on the Smoothened Effector of Signal Response. *Developmental Cell* **2013**, *26*, 346–357.

(15) Blassberg, R.; Macrae, J. I.; Briscoe, J.; Jacob, J. Reduced cholesterol levels impair Smoothened activation in Smith–Lemli–Opitz syndrome. *Human Molecular Genetics* **2015**, *25*, 693–705.

(16) Luchetti, G.; Sircar, R.; Kong, J. H.; Nachtergaele, S.; Sagner, A.; Byrne, E. F.; Covey, D. F.; Siebold, C.; Rohatgi, R. Cholesterol activates the G-protein coupled receptor Smoothened to promote Hedgehog signaling. *eLife* **2016**, *5*, e20304.

(17) Huang, P.; Nedelcu, D.; Watanabe, M.; Jao, C.; Kim, Y.; Liu, J.; Salic, A. Cellular Cholesterol Directly Activates Smoothened in Hedgehog Signaling. *Cell* **2016**, *166*, 1176–1187.e14.

(18) Byrne, E. F. X.; Sircar, R.; Miller, P. S.; Hedger, G.; Luchetti, G.; Nachtergaele, S.; Tully, M. D.; Mydock-McGrane, L.; Covey, D. F.; Rambo, R. P.; Sansom, M. S. P.; Newstead, S.; Rohatgi, R.; Siebold, C. Structural basis of Smoothened regulation by its extracellular domains. *Nature* **2016**, *535*, 517–522.

(19) Myers, B. R.; Neahring, L.; Zhang, Y.; Roberts, K. J.; Beachy, P. A. Rapid, direct activity assays for Smoothened reveal Hedgehog pathway regulation by membrane cholesterol and extracellular sodium. *Proceedings of the National Academy of Sciences U.S.A.* **2017**, *114*.

(20) Kinnebrew, M.; Iverson, E. J.; Patel, B. B.; Pusapati, G. V.; Kong, J. H.; Johnson, K. A.; Luchetti, G.; Eckert, K. M.; McDonald, J. G.; Covey, D. F.; Siebold, C.; Radhakrishnan, A.; Rohatgi, R. Cholesterol accessibility at the ciliary membrane controls hedgehog signaling. *eLife* **2019**, *8*, e50051.

(21) Kumar, G. A.; Chattopadhyay, A. Statin-Induced Chronic Cholesterol Depletion Switches GPCR Endocytosis and Trafficking: Insights from the Serotonin Receptor. *ACS Chemical Neuroscience* **2019**, *11*, 453–465.

(22) Kumar, G. A.; Sarkar, P.; Stepniewski, T. M.; Jafurulla, M.; Singh, S. P.; Selent, J.; Chattopadhyay, A. A molecular sensor for cholesterol in the human serotonin receptor. *Science Advances* **2021**, *7*, 30.

(23) Serdiuk, T.; Manna, M.; Zhang, C.; Mari, S. A.; Kulig, W.; Pluhackova, K.; Kobilka, B. K.; Vattulainen, I.; Müller, D. J. A cholesterol analog stabilizes the human β_2-adrenergic receptor nonlinearly with temperature. *Science Signaling* **2022**, *15*, 737.

(24) Radhakrishnan, A.; Rohatgi, R.; Siebold, C. Cholesterol access in cellular membranes controls Hedgehog signaling. *Nature Chemical Biology* **2020**, *16*, 1303–1313.

(25) Kinnebrew, M.; Luchetti, G.; Sircar, R.; Frigui, S.; Viti, L. V.; Naito, T.; Beckert, F.; Saheki, Y.; Siebold, C.; Radhakrishnan, A.; Rohatgi, R. Patched 1 reduces the accessibility of cholesterol in the outer leaflet of membranes. *eLife* **2021**, *10*, e70504.

(26) Ingham, P.; Taylor, A.; Nakano, Y. Role of the Drosophila patched gene in positional signalling. *Nature* **1991**, *353*, 184–187.

(27) Denef, N.; Neubüser, D.; Perez, L.; Cohen, S. M. Hedgehog Induces Opposite Changes in Turnover and Subcellular Localization of Patched and Smoothened. *Cell* **2000**, *102*, 521–531.

(28) Ingham, P.; Nystedt, S.; Nakano, Y.; Brown, W.; Stark, D.; van den Heuvel, M.; Taylor, A. Patched represses the Hedgehog signalling pathway by promoting modification of the Smoothened protein. *Current Biology* **2000**, *10*, 1315–1318.

(29) Taipale, J.; Cooper, M. K.; Maiti, T.; Beachy, P. A. Patched acts catalytically to suppress the activity of Smoothened. *Nature* **2002**, *418*, 892–896.

(30) Gong, X.; Qian, H.; Cao, P.; Zhao, X.; Zhou, Q.; Lei, J.; Yan, N. Structural basis for the recognition of Sonic Hedgehog by human Patched1. *Science* **2018**, *361*.

(31) Zhang, Y.; Bulkley, D. P.; Xin, Y.; Roberts, K. J.; Asarnow, D. E.; Sharma, A.; Myers, B. R.; Cho, W.; Cheng, Y.; Beachy, P. A. Structural Basis for Cholesterol Transport-like Activity of the Hedgehog Receptor Patched. *Cell* **2018**, *175*, 1352–1364.e14.

(32) Rudolf, A. F.; Kinnebrew, M.; Kowatsch, C.; Ansell, T. B.; El Omari, K.; Bishop, B.; Pardon, E.; Schwab, R. A.; Malinauskas, T.; Qian, M.; Duman, R.; Covey, D. F.; Steyaert, J.; Wagner, A.; Sansom, M. S. P.; Rohatgi, R.; Siebold, C. The morphogen Sonic hedgehog inhibits its receptor Patched by a pincer grasp mechanism. *Nature Chemical Biology* **2019**, *15*, 975–982.

(33) Qi, C.; Di Minin, G.; Vercellino, I.; Wutz, A.; Korkhov, V. M. Structural basis of sterol recognition by human hedgehog receptor PTCH1. *Science Advances* **2019**, *5*.

(34) Zhang, Y.; Lu, W.-J.; Bulkley, D. P.; Liang, J.; Ralko, A.; Han, S.; Roberts, K. J.; Li, A.; Cho, W.; Cheng, Y.; Manglik, A.; Beachy, P. A. Hedgehog pathway activation through nanobody-mediated conformational blockade of the Patched sterol conduit. *Proceedings of the National Academy of Sciences U.S.A.* **2020**, *117*, 28838–28846.

(35) Kinnebrew, M.; Woolley, R. E.; Ansell, T. B.; Byrne, E. F. X.; Frigui, S.; Luchetti, G.; Sircar, R.; Nachtergaele, S.; Mydock-McGrane, L.; Krishnan, K.; Newstead, S.; Sansom, M. S. P.; Covey, D. F.; Siebold, C.; Rohatgi, R. Patched 1 regulates Smoothened by controlling sterol binding to its extracellular cysteine-rich domain. *Science Advances* **2022**, *8*, 22.

(36) Davies, J. P.; Chen, F. W.; Ioannou, Y. A. Transmembrane Molecular Pump Activity of Niemann-Pick C1 Protein. *Science* **2000**, *290*, 2295–2298.

(37) Bidet, M.; Joubert, O.; Lacombe, B.; Ciantar, M.; NehmÃ©, R.; Mollat, P.; BrÃ©tillon, L.; Faure, H.; Bittman, R.; Ruat, M.; Mus-Veteau, I. The Hedgehog Receptor Patched Is Involved in Cholesterol Transport. *PLoS ONE* **2011**, *6*, e23834.

(38) Ansell, T. B.; Corey, R. A.; Viti, L. V.; Kinnebrew, M.; Rohatgi, R.; Siebold, C.; Sansom, M. S. P. The energetics and ion coupling of cholesterol transport through Patched1. *Science Advances* **2023**, *9*, 34.

(39) Qi, X.; Schmiege, P.; Coutavas, E.; Li, X. Two Patched molecules engage distinct sites on Hedgehog yielding a signaling-competent complex. *Science* **2018**, *362*, 6410.

(40) Nachtergaele, S.; Whalen, D. M.; Mydock, L. K.; Zhao, Z.; Malinauskas, T.; Krishnan, K.; Ingham, P. W.; Covey, D. F.; Siebold, C.; Rohatgi, R. Structure and function of the Smoothened extracellular domain in vertebrate Hedgehog signaling. *eLife* **2013**, *2*.

(41) Nedelcu, D.; Liu, J.; Xu, Y.; Jao, C.; Salic, A. Oxysterol binding to the extracellular domain of Smoothened in Hedgehog signaling. *Nature Chemical Biology* **2013**, *9*, 557–564.

(42) Weierstall, U.; James, D.; Wang, C.; White, T. A.; Wang, D.; Liu, W.; Spence, J. C. H.; Doak, R. B.; Nelson, G.; Fromme, P.; Fromme, R.; Grotjohann, I.; Kupitz, C.; Zatsepin, N. A.; Liu, H.; Basu, S.; Wacker, D.; Han, G. W.; Katritch, V.; Boutet, S.; Messerschmidt, M.; Williams, G. J.; Koglin, J. E.; Seibert, M. M.; Klinker, M.; Gati, C.; Shoeman, R. L.; Barty, A.; Chapman, H. N.; Kirian, R. A.; Beyerlein, K. R.; Stevens, R. C.; Li, D.; Shah, S. T. A.; Howe, N.; Caffrey, M.; Cherezov, V. Lipidic cubic phase injector facilitates membrane protein serial femtosecond crystallography. *Nature Communications* **2014**, *5*, 3309.

(43) Zhang, X.; Zhao, F.; Wu, Y.; Yang, J.; Han, G. W.; Zhao, S.; Ishchenko, A.; Ye, L.; Lin, X.; Ding, K.; Dharmarajan, V.; Griffin, P. R.; Gati, C.; Nelson, G.; Hunter, M. S.; Hanson, M. A.; Cherezov, V.; Stevens, R. C.; Tan, W.; Tao, H.; Xu, F. Crystal structure of a multi-domain human smoothened receptor in complex with a super stabilizing ligand. *Nature Communications* **2017**, *8*, 15383.

(44) Qi, X.; Friedberg, L.; Bose-Boyd, R. D.; Long, T.; Li, X. Sterols in an intramolecular channel of Smoothened mediate Hedgehog signaling. *Nature Chemical Biology* **2020**, *16*, 1368–1375.

(45) Deshpande, I.; Liang, J.; Hedeen, D.; Roberts, K. J.; Zhang, Y.; Ha, B.; Latorraca, N. R.; Faust, B.; Dror, R. O.; Beachy, P. A.; Myers, B. R.; Manglik, A. Smoothened stimulation by membrane sterols drives Hedgehog pathway activity. *Nature* **2019**, *571*, 284–288.

(46) Qi, X.; Liu, H.; Thompson, B.; McDonald, J.; Zhang, C.; Li, X. Cryo-EM structure of oxysterol-bound human Smoothened coupled to a heterotrimeric Gi. *Nature* **2019**, *571*, 279–283.

(47) Janda, C. Y.; Waghray, D.; Levin, A. M.; Thomas, C.; Garcia, K. C. Structural Basis of Wnt Recognition by Frizzled. *Science* **2012**, *337*, 59–64.

(48) Nile, A. H.; Mukund, S.; Stanger, K.; Wang, W.; Hannoush, R. N. Unsaturated fatty acyl recognition by Frizzled receptors mediates dimerization upon Wnt ligand binding. *Proceedings of the National Academy of Sciences U.S.A.* **2017**, *114*, 4147–4152.

(49) Hirai, H.; Matoba, K.; Mihara, E.; Arimori, T.; Takagi, J. Crystal structure of a mammalian Wnt–frizzled complex. *Nature Structural & Molecular Biology* **2019**, *26*, 372–379.

(50) Huang, P.; Zheng, S.; Wierbowski, B. M.; Kim, Y.; Nedelcu, D.; Aravena, L.; Liu, J.; Kruse, A. C.; Salic, A. Structural Basis of Smoothened Activation in Hedgehog Signaling. *Cell* **2018**, *174*, 312–324.e16.

(51) Hedger, G.; Koldsø, H.; Chavent, M.; Siebold, C.; Rohatgi, R.; Sansom, M. S. Cholesterol Interaction Sites on the Transmembrane Domain of the Hedgehog Signal Transducer and Class F G Protein-Coupled Receptor Smoothened. *Structure* **2019**, *27*, 549–559.e2.

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

(53) Darve, E.; Rodríguez-Gómez, D.; Pohorille, A. Adaptive biasing force method for scalar and vector free energy calculations. *The Journal of Chemical Physics* **2008**, *128*.

(54) HÃ©nin, J.; Fiorin, G.; Chipot, C.; Klein, M. L. Exploring Multidimensional Free Energy Landscapes Using Time-Dependent Biases on Collective Variables. *Journal of Chemical Theory and Computation* **2009**, *6*, 35–47.

(55) Meurer, F.; Do, H. T.; Sadowski, G.; Held, C. Standard Gibbs energy of metabolic reactions: II. Glucose-6-phosphatase reaction and ATP hydrolysis. *Biophysical Chemistry* **2017**, *223*, 30–38.

(56) Tarling, E. J.; Vallim, T. Q. d. A.; Edwards, P. A. Role of ABC transporters in lipid transport and human disease. *Trends in Endocrinology & Metabolism* **2013**, *24*, 342–350.

(57) Katritch, V.; Fenalti, G.; Abola, E. E.; Roth, B. L.; Cherezov, V.; Stevens, R. C. Allosteric sodium in class A GPCR signaling. *Trends in Biochemical Sciences* **2014**, *39*, 233–244.

(58) Selvam, B.; Shamsi, Z.; Shukla, D. Universality of the Sodium Ion Binding Mechanism in Class A G-Protein-Coupled Receptors. *Angewandte Chemie* **2018**, *57*, 3048–3053.

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

(60) Chan, M. C.; Selvam, B.; Young, H. J.; Procko, E.; Shukla, D. The substrate import mechanism of the human serotonin transporter. *Biophysical Journal* **2022**, *121*, 715–730.

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

(62) Kumari, S.; Mitra, A.; Bulusu, G. Putative Role of Cholesterol in Shaping the Structural and Functional Dynamics of Smoothened (SMO). *The Journal of Physical Chemistry B* **2023**, *127*, 9476–9495.

(63) Webb, B.; Sali, A. Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics* **2016**, *54*, 5–6.

(64) Scandroglio, F.; Venkata, J. K.; Loberto, N.; Prioni, S.; Schuchman, E. H.; Chigorno, V.; Prinetti, A.; Sonnino, S. Lipid content of brain, brain membrane lipid domains, and neurons from acid sphingomyelinase deficient mice. *Journal of Neurochemistry* **2008**, *107*, 329–338.

(65) 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.

(66) Lee, J.; Patel, D. S.; Ståhle, J.; Park, S.-J.; Kern, N. R.; Kim, S.; Lee, J.; Cheng, X.; Valvano, M. A.; Holst, O.; Knirel, Y. A.; Qi, Y.; Jo, S.; Klauda, J. B.; Widmalm, G.; Im, W. CHARMM-GUI Membrane Builder for Complex Biological Membrane Simulations with Glycolipids and Lipoglycans. *Journal of Chemical Theory and Computation* **2018**, *15*, 775–786.

(67) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *The Journal of Physical Chemistry B* **2010**, *114*, 7830–7843.

(68) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the additive CHARMM all-atom protein force field targeting improved

sampling of the backbone ϕ, ψ and side-chain $\chi1$ and $\chi2$ dihedral angles. *Journal of Chemical Theory and Computation* **2012**, *8*, 3257–3273.

(69) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics* **1983**, *79*, 926–935.

(70) 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.

(71) Andersen, H. C. Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *Journal of Computational Physics* **1983**, *52*, 24–34.

(72) Case, D.; Belfon, K.; Ben-Shalom, I.; Brozell, S.; Cerutti, D.; Cheatham III, T.; Cruzeiro, V.; Darden, T.; Duke, R. AMBER 2020.

(73) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber biomolecular simulation package. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2012**, *3*, 198–210.

(74) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *Journal of Computational Chemistry* **2005**, *26*, 1668–1688.

(75) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Grand, S. L.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *Journal of Chemical Theory and Computation* **2012**, *8*, 1542–1555.

(76) 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.

(77) 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.

(78) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. *Journal of Chemical Physics* **1993**, *98*, 10089–10092.

(79) Davidchack, R. L.; Handel, R.; Tretyakov, M. V. Langevin thermostat for rigid body dynamics. *Journal of Chemical Physics* **2009**, *130*, 234101.

(80) Åqvist, J.; Wennerström, P.; Nervall, M.; Bjelic, S.; Brandsdal, B. O. Molecular dynamics simulations of water and biomolecules with a Monte Carlo constant pressure algorithm. *Chemical Physics Letters* **2004**, *384*, 288–294.

(81) Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B. K.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kalé, L. V.; Schulten, K.; Chipot, C.; Tajkhorshid, E. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *Journal of Chemical Physics* **2020**, *153*, 044130.

(82) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **2005**, *26*, 1781–1802.

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

(84) Stone, J. *An Efficient Library for Parallel Ray Tracing and Animation*. M.Sc. thesis, Computer Science Department, University of Missouri-Rolla, 1998.

(85) 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.

(86) Bowman, G. R.; Ensign, D. L.; Pande, V. S. Enhanced Modeling via Network Theory: Adaptive Sampling of Markov State Models. *Journal of Chemical Theory and Computation* **2010**, *6*, 787–794.

(87) 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.

(88) 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.

(89) Kleiman, D. E.; Shukla, D. Multiagent Reinforcement Learning-Based Adaptive Sampling for Conformational Dynamics of Proteins. *Journal of Chemical Theory and Computation* **2022**, *18*, 5422–5434.

(90) 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.

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

(92) 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.

(93) 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.

(94) Nadeem, H.; Shukla, D. Optimizing adaptive sampling via Policy Ranking. *arXiv* **2024**, doi: 10.48550/ARXIV.2410.15259.

(95) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; PÃ©rez-HernÃ¡ndez, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; NoÃ©, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *Journal of Chemical Theory and Computation* **2015**, *11*, 5525–5542.

(96) Schwantes, C. R.; Pande, V. S. Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *Journal of Chemical Theory and Computation* **2013**, *9*, 2000–2009.

(97) 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.

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

(99) 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.

(100) 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.

(101) 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.

(102) 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.

(103) 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.

(104) 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.

(105) 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.

(106) 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.

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

(108) 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.

(109) Schroedinger, LLC, The PyMOL Molecular Graphics System. `https://pymol.org/2/`.

(110) 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.

(111) Hunter, J. D. Matplotlib: A 2D graphics environment. *Computing in Science & Engineering* **2007**, *9*, 90–95.

(112) Waskom, M. L. seaborn: statistical data visualization. *Journal of Open Source Software* **2021**, *6*, 3021.

(113) Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; van Kerkwijk, M. H.; Brett, M.; Haldane, A.; del Río, J. F.; Wiebe, M.; Peterson, P.; Gérard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; Oliphant, T. E. Array programming with NumPy. *Nature* **2020**, *585*, 357–362.

(114) Smart, O.; Goodfellow, J.; Wallace, B. The pore dimensions of gramicidin A. *Biophysical Journal* **1993**, *65*, 2455–2460.