Network analysis reveals how lipids and other cofactors influence membrane protein allostery

Annie M. Westerlund, Oliver Fleetwood‡, Sergio Perez-Conesa‡, and Lucie Delemotte*

KTH Royal Institute of Technology, Science for Life Laboratory, Stockholm

Abstract

Many membrane proteins are modulated by external stimuli, such as small molecule binding or change in pH, transmembrane voltage or temperature. This modulation typically occurs at sites that are structurally distant from the functional site. Revealing the communication, known as allostery, between these two sites is key to understanding the mechanistic details of these proteins. Residue interaction networks of isolated proteins are commonly used to this end. Membrane proteins, however, are embedded in a lipid bilayer which may contribute to allosteric communication. The fast diffusion of lipids hinders direct use of standard residue interaction networks. Here, we present an extension which includes cofactors such as lipids and small molecules in the network. The novel framework is applied to three membrane proteins: a voltage-gated ion channel (KCNQ1), a G-protein coupled receptor (GPCR - β2 adrenergic receptor) and a pH-gated ion channel (KcsA). Through systematic analysis of the obtained networks and their components, we demonstrate the importance of lipids for membrane protein allostery. Finally, we reveal how small molecules may stabilize different protein states by allosterically coupling and decoupling the protein from the membrane.

Keywords: GPCR, KcsA, KCNQ1, Kv7.1, beta2 adrenergic receptor, allostery, network analysis, lipid-protein interactions, molecular dynamics

‡ Contributed equally to the work.

* lucied@kth.se; www.github.com/delemottelab

I. INTRODUCTION

Proteins are molecular machines which function via complex networks of interacting residues. A perturbation at a specific site in a protein may travel across the protein to modulate a distal site. The specific chain of residues that propagate this information to the distal site constitutes an *allosteric pathway* [1]. Allostery is the major functional contributor in membrane proteins such as receptors and ion channels [2–5]. G-protein coupled receptors (GPCRs), for example, bind specific ligands at an extracellular binding site. These binding events yield conformational changes at an intracellular binding site, altering the receptor's active/inactive equilibrium [6]. Voltage-gated ion channels, on the other hand, sense and respond to membrane potentials in domains which are located far away from the pore gate. Activation of these domains are allosterically communicated to the pore to trigger channel opening [7–11]. Understanding protein allostery and its governing components may therefore reveal the mechanistic details of these proteins.

All-atom molecular dynamics (MD) simulations provide atomistic details of how protein residues interact with each other, as well as with their surroundings. MD simulations thus contain explicit information about protein allostery. However, the large number of atoms in physiologically relevant systems prevents direct visual identification of residues that are especially important for communicating allostery. This requires computational methods able to synthesize and interpret the raw data. Residue interaction networks of isolated proteins have previously been used to reveal protein allostery [11–15]. Membrane proteins, however, are surrounded by lipids and may be modulated by small molecules, both of which play an integral role in communicating allostery between different protein domains [2, 16]. The original framework therefore has to be extended to reveal the allosteric role of cofactors, such as lipids, ligands or highly retained water molecules.

The difficulty of including cofactors in these networks relates to their redundancy, and the network topology. A network consists of two components: nodes and connection between nodes (edges), which are static. Cofactors, however, constantly move and may exchange positions. To understand the role of lipids, for example, the main goal is to reveal whether a lipid bound in a specific location relative to the protein is allosterically important, rather than a specific lipid molecule which may diffuse away. The redundancy of cofactors therefore has to be translated to the nodes of the static network.

Here, we extend the framework of standard residue interaction networks to include cofactors

such as lipids and ligands. We assign cofactor residues to network nodes by solving a combinatorial optimization problem in each trajectory frame, i.e. minimizing the distances between cofactor residues and network nodes. This new framework is applied to MD simulations of three membrane proteins with different functions and topologies: a voltage-gated ion channel, a GPCR, and a pH-gated ion channel. By applying current flow analysis [17, 18] to these networks, we reveal their allosteric fingerprints. Finally, we access the allosteric role of lipids, and show how this role may be modulated by ligands. The method can easily be applied to other systems. A jupyter notebook tutorial is publicly available on our lab github [19].

II. THEORY

A. Constructing a residue interaction network

An ordinary residue interaction network is built by grouping the atoms of each protein residue to a node and forming edges between nodes of interacting residues. The weight of an edge represents the extent of residue interaction. Practically, the network is here obtained from the elementwise product of a contact map, C, and a matrix representing the correlation of node movement, M,

$$A_{ij} = C_{ij} M_{ij} \quad \forall i, j, \tag{1}$$

where A is the network adjacency matrix. Fig. 1 shows a flowchart of how the network is built using MD simulation data. We use a continuous contact map and positional mutual information to model spatial proximity and correlation of residue movements, respectively (see details below) [20]. The analysis is performed on non-hydrogen atoms of the protein residues, ligand and lipids. MDtraj [21] is used as backbone to perform analysis.

B. Assigning lipids and ligands to nodes

The network can be extended to represent cofactors as additional nodes. Similar to the protein, we define *interactors* by grouping cofactor atoms into domains. These interactors will correspond to interactor-nodes in the network. Here, we describe a ligand with one interactor while lipids are divided into three interactors; one corresponding to the head group and two to the hydrophobic tails, Fig. 1, S1.

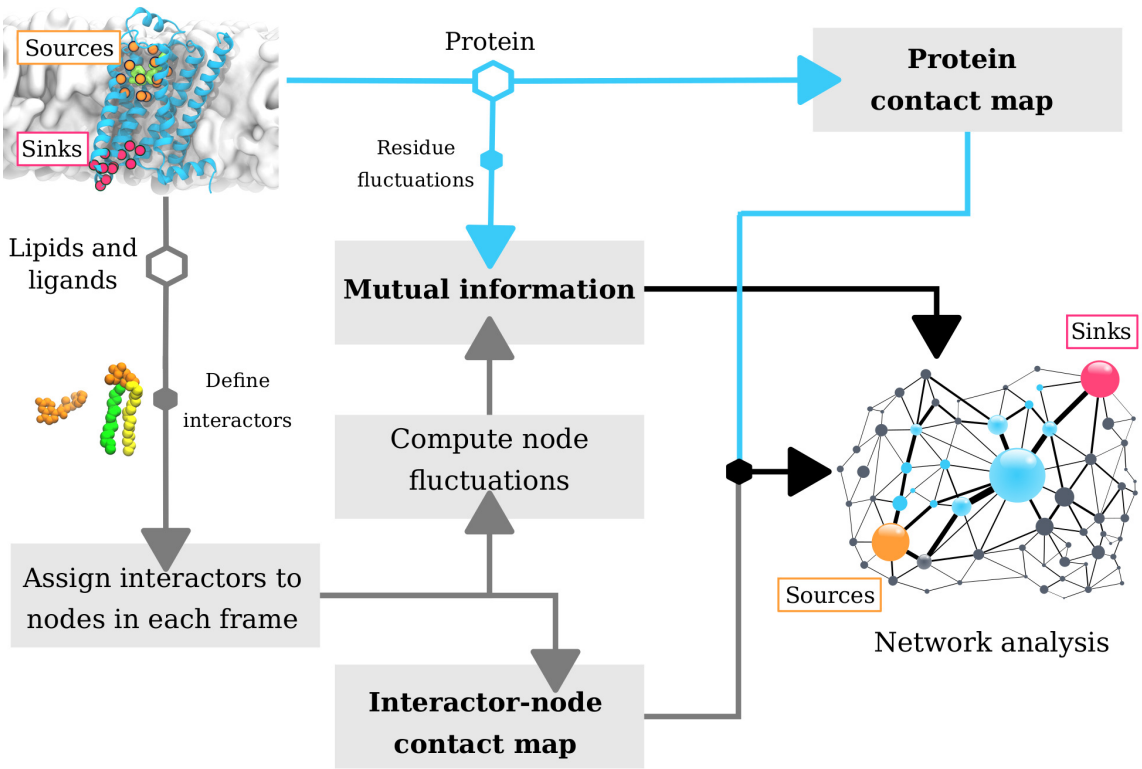

FIG. 1. Steps carried out to build a residue interaction network which includes lipids and ligands as well as the protein. Starting from a system with a protein embedded in a lipid membrane and bound ligand, the protein is first isolated. Each protein residue is assigned a node in the isolated protein network and the contact map and mutual information for all protein residue pairs are calculated (blue arrows) The lipids and ligands are divided into "interactors". These interactors are assigned network nodes depending on their positions relative to the initial frame. The nodes are used to compute the interactor-protein contact map (gray arrows), and the fluctuations are used to compute mutual information between node movements. The contact maps are merged. The full contact map and mutual information make up the final network (black arrows). Once the network is built, we can extract allosteric pathways between source (orange) and sink (red) nodes employing network analysis.

A first solvation shell lipid may exchange position with a second shell lipid. In such case, the protein environment is essentially unaltered, while the individual lipid environments change abruptly. We therefore keep the network nodes static by permuting the lipid interactors and reassigning interactors to nodes in each frame. To do this, the trajectory is first aligned on the protein C_α atoms. Nodes are assigned to interactors in the first frame of the trajectory. Next, we permute the interactor-nodes such that the network remains similar to the initial frame. Interactors

are assigned to nodes in each frame by minimizing the distances between interactor centroids of the current and initial frames [22]. The distances are computed with the minimum image convention using an orthorhombic box periodic boundary condition, and the optimization is performed with a linear assignment problem solver (LAPJV) [23–25].

C. Continuous contact map

Continuous contact maps yield more stable networks than binary ones [14]. A switching-function, $K(d)$ introduced in [20] is used. Given the minimum distance $d_{ij}(n)$ between atoms of nodes i and j in frame n,

$$K(d_{ij}(n)) = \begin{cases} 1 & \text{if } d_{ij}(n) \leq c \\ \exp\left[-\frac{d_{ij}(n)^2}{2\sigma^2}\right] / \exp\left[-\frac{c^2}{2\sigma^2}\right] & \text{otherwise} \end{cases} \tag{2}$$

Distances below c are given a weight of 1, while weights of larger distances are continuously decreased. Similar to [20], we used $c = 0.45$ nm, and $K(d_{cut}) = 10^{-5}$, where $d_{cut} = 0.8$ nm, leading to $\sigma = \sqrt{\frac{c^2 - d_{cut}^2}{2 \ln K(d_{cut})}} \approx 0.138$. The final contact map is averaged over frames,

$$C_{ij} = \frac{1}{N_{\text{points}}} \sum_{n=1}^{N_{\text{points}}} K(d_{ij}(n)). \tag{3}$$

The minimum image convention is used to compute distances involving interactor nodes [21]. This makes the network periodically connected, allowing allosteric pathways to cross simulation box boundaries.

D. Mutual information

Correlation of residue and interactor node movements are modeled with positional mutual information (MI) of node centroids [20, 26]. This measures correlation of fluctuation distances around an equilibrium position by characterizing the amount of information that can be inferred about a specific node position knowing the position of another node. Given the fluctuation density, $\rho_i(x)$, where x is the distance to the equilibrium position, the entropy of residue i is calculated as

$$H_i = -\int_X \rho_i(x) \ln \rho_i(x) dx \approx \frac{1}{N_{\text{points}}} \sum_{n=1}^{N_{\text{points}}} [-\ln \rho_i(x_n)], \tag{4}$$

and the MI between nodes i and j is obtained with

$$\hat{M}_{ij} = H_i + H_j - H_{ij}.$$

To improve accuracy of entropy estimation, we sample N_{points} new data points from the density, $\rho_{ij}(x)$, 10 times, followed by repeating entropy and MI calculations and averaging over datasets.

Each density is represented by a Gaussian mixture model (GMM) [27]; a linear combination of Gaussian basis functions. Because Gaussians are continuous, the density estimation is more stable where the data is sparse and does not introduce unnecessary approximations compared to discrete methods such as histogramming [28, 29]. The likelihood of the GMM will increase arbitrarily when increasing the number of Gaussians. To choose a non-overfitting but detailed GMM, we use the Bayesian information criterion (BIC) [30]. This adds a penalty to the likelihood, which increases with model complexity. For each density, the number of Gaussians is varied between 1 and 4. The GMM with the smallest BIC is ultimately chosen.

Because non-protein interactors may cross the simulation box boundaries, their node equilibrium positions are calculated by first replicating the node frame coordinates to all periodic images. The replication creates clusters of points. One cluster corresponds to the fluctuations around one image equilibrium, which is given by the centroid of this cluster. We then identify a full cluster using a cutoff around the node's first frame coordinate in the original box. The cutoff to identify the full cluster is taken as half the smallest box side. Although this cutoff may seem conservative for connecting all points within the same cluster, it effectively avoids connecting points that are part of different clusters. The equilibrium position is then defined as the cluster centroid.

III. INFERRING ALLOSTERY WITH CURRENT FLOW ANALYSIS

Current flow analysis [17, 18], named after the analogy to the flow of current through an electrical network, measures the net flow of information between a set of source (S_0) and sink (S_1) nodes. Allostery within a protein is communicated via multiple pathways [1]. Current flow betweenness takes all effective pathways between the source and sink nodes into account. It is thus well-suited to study allostery from molecular dynamics simulations of proteins [14, 20]. The result highlights residues and interactors that are important for defining the allosteric pathways.

To understand the role of lipids in defining the allosteric pathways, a measure of signaling efficiency is needed. Current flow closeness centrality [17] estimates efficiency using an inverse

”current flow distance”. The details of current flow analysis are described in Supplementary Materials (Section I A), as well as in the original papers [17, 18]. Here, we used current flow betweenness and closeness centrality to extract allosterically important residues in the proteins, and to reveal the allosteric role of lipids and ligands. The current flow betweenness and closeness are averaged over subunits in homomeric proteins. This is done by first replicating and summing the current flow from each subunit over the structure.

We studied the allosteric role of lipids and ligands in three well-known systems; the voltage-gated ion channel KCNQ1, the GPCR $\beta 2$ adrenergic receptor, and the pH-gated ion channel KcsA. Details about simulation setup and protocols are described in Supplementary Materials (Section I B). The structures were visualized with VMD [31].

IV. RESULTS AND DISCUSSION

A. KCNQ1 exploits the membrane for communicating voltage sensitivity to the pore

KCNQ1 is a voltage-gated potassium channel critical for action potential repolarization in the heart. KCNQ1 contains the common elements of a homotetrameric voltage-gated ion channel; each subunit consists of six transmembrane helices [32]. The first four helices form the voltage-sensing domain (VSD), while the last two make up the pore domain, Fig. 2a. The Ca^{2+}-sensor calmodulin (CaM) is required for KCNQ1 channel assembly [33, 34], and binds to the KCNQ1 C-terminal domain (CTD) [32, 35]. Specific to KCNQ1 is that PIP_2 is required to couple VSD activation to pore opening [36, 37]. The two cryo-EM structures of human KCNQ1 [35] revealed ”bent” CTD helices in the absence of PIP_2, and ”straight” CTD helices in the presence of PIP_2, Fig. S2. We call these states *CTD-bent* and *CTD-straight*, respectively. The CTD-bent state allows interactions between CaM and the VSD, which are disrupted in the CTD-straight channel. Recent studies showed that both states are present in the channel activation pathway [20].

We studied the allosteric roles of PIP_2 and the membrane to allow communication between the VSD and pore. Efficiency of communication was measured in terms of current flow closeness using the KCNQ1 gating charges (R1, R2, Q3. R4) as sources and the pore residue S349 as sink. The allosteric contribution of each system component was inferred by performing the efficiency analysis on the networks including: 1) the protein, 2) the protein and PIP_2, and 3) the protein and the full membrane.

Fig. 2b shows the allosteric efficiency when including PIP_2 and the lipid membrane in the network. The upper row displays the results obtained from the CTD-bent simulations, while the lower row shows the corresponding results from the CTD-straight simulations. The left and right columns correspond to PIP_2 being absent or present in the simulations, respectively. Regardless of PIP_2-bound state, the membrane increases the allosteric efficiency, indicating that the membrane is involved in communicating voltage sensitivity to the pore, Fig. 2b. Fig. 2c depicts a top-view of the current flow through each residue projected onto the CTD-bent (upper) and CTD-straight (lower) structures. The lipid interactors with the highest current flow are shown in surface representation. KCNQ1 exploits the membrane lipids tucked in the cavities between VSDs and the pore to communicate allostery, Fig. 2c. Fig. 2d shows the difference in efficiency between networks which include the additional interactors compared to the isolated protein. PIP_2 appears to not significantly increase signaling efficiency alone. The effect of tightly bound ligands may thus already be encoded in the isolated protein network. Moreover, the CTD-straight channel uses the membrane to a larger extent than the CTD-bent, while PIP_2 appears to not alter the involvement of the membrane. The role of the bound PIP_2 may thus not be to explicitly couple the VSD to the pore via the membrane, but to induce the large rearrangements in the CTD related to pore opening.

B. Agonist ligand-binding decreases the membrane allosteric role in the inactive β2 adrenergic receptor

The β2 adrenergic receptor (β2AR) is a G-protein coupled receptor (GPCR) and a common target for asthma medication. It sustains communication through the cell membrane by recognizing and binding ligands. The binding of agonist ligands in the receptor extracellular domain induces structural rearrangements throughout the protein. The induced state, known as the active state, allows for intracellular G-protein association, which triggers cellular downstream responses [6, 38, 39]. As such, understanding the molecular mechanisms of agonist-imposed allosteric signaling between the two binding sites may be an essential element to facilitate development of novel drugs.

Fig. 3a shows the structural elements of the β2AR with the bound agonist ligand BI-167107 and G-protein mimetic nanobody (Nb) [40]. We applied the current flow analysis to the agonist bound as well as ligand-free (apo) β2AR in the active/Nb-bound and inactive/Nb-free states, respectively. To investigate the allosteric roles of the agonist and membrane, we performed the analysis on

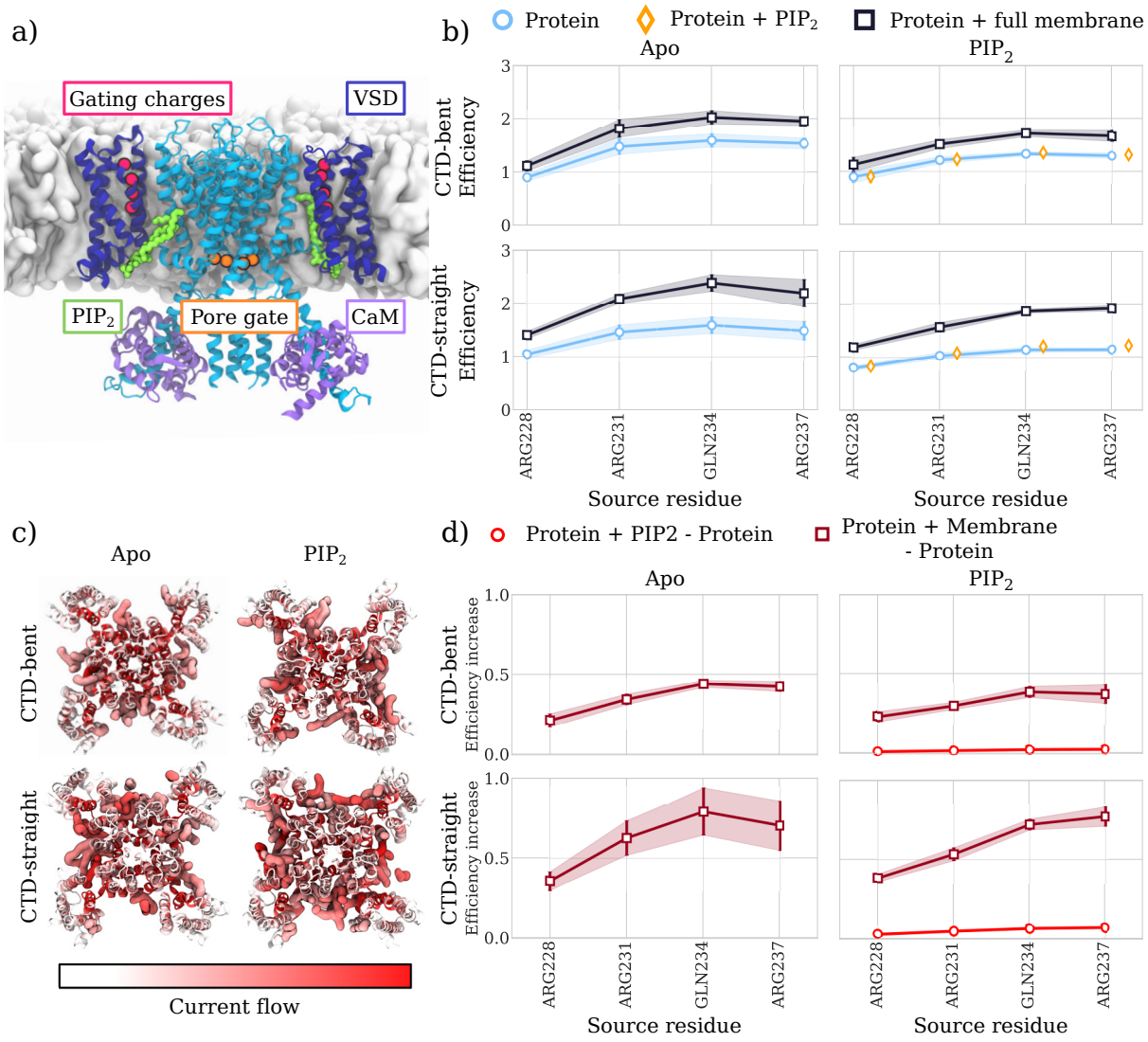

FIG. 2. a) Structure of CTD-straight KCNQ1, showing important structural features, and sources and sinks. b) State- and PIP_2-dependent allosteric efficiency obtained with current flow closeness. Efficiency is measured from the source residues listed on the x-axis to the sink residue S349 depicted in a). Blue: protein, yellow: protein+bound PIP_2, and black: protein+full membrane. Left column: apo channels, and right: PIP_2-bound channels. Upper row: CTD-bent state, and bottom row: CTD-straight. c) Average current flow through each protein residue and lipid interactor projected onto the first trajectory frame (top view). Red signifies larger current flow values. Lipid interactors (current flow > 0.008) are shown in surface representation. d) Difference in allosteric efficiency of networks built by including different components of the systems. Dark red: comparing networks that include both protein and membrane to networks that only include the protein, and red: comparing networks that include protein and PIP_2 to networks that only include the protein. Averages and standard errors (SEM) are computed across subunits.

networks including different system components: the protein, the protein and agonist, as well as the protein, agonist and membrane. Source nodes were defined as residues in the orthosteric site which are known to interact with agonists, Fig. 3a-b [41]. The sink nodes were taken as residues in contact with the G protein in the active state structure 3SN6 [42], Fig. 3a and Table S1.

The apo inactive β2AR exploits the membrane to transfer allostery between the ligand and G-protein binding sites, as shown by an increased signaling efficiency after including the membrane in the network Fig. 3b. Specifically, it uses the closest layer of the lipid membrane, Fig. 3c. The agonist-bound inactive receptor, on the other hand, demonstrates a significantly decreased allosteric involvement of the membrane relative to the apo receptor, Fig. 3b-d. This is shown by the smaller increase of signaling efficiency when including the membrane, Fig. 3b,d, as well as a lower amount of current flow through lipids, Fig. 3c. This suggests that the bound ligand decouples the allosteric pathways from the membrane in the inactive state. Conversely, the membrane increases allosteric efficiency of the active/Nb-bound receptor, in both apo and agonist-bound systems, Fig. 3b-d. Thus, an intracellular binding partner may again allosterically couple β2AR to the surrounding membrane in its active state.

To understand the allosteric details related to exploiting the membrane, we plotted the current flow along protein residue sequence and contrasted the profiles obtained from the apo (black) and agonist-bound (blue) simulations, Fig. 4a. The profiles show a large allosteric involvement of TM3 and TM5-7. These helices are important for activation and interact with both binding partners [42–45]. An interesting feature related to ligand binding emerged; namely an increased involvement of TM1, in particular of the residues I43, V44, I47 and N51, Fig. 4b. N51 is highly conserved across all class A GPCRs [46] and supports a stabilizing water network in the receptor's inactive state [47]. Notably, the inactive agonist-bound receptor also exhibited more current flow at TM3, compared to the apo inactive receptor. The difference between agonist-bound and apo current flow profiles is smaller in the active/Nb state than in the inactive state. Together with the results of Fig. 3d, this indicates that an intracellular binding partner overshadows the ligand-induced redirection of allosteric pathways within the protein.

C. Binding of DOPG lipid changes the KcsA allosteric signaling at the open state

KcsA is a bacterial pH-gated and homotetrameric potassium channel [48]. The simplicity and high conservation of structural features make the channel a well-suited model for complex

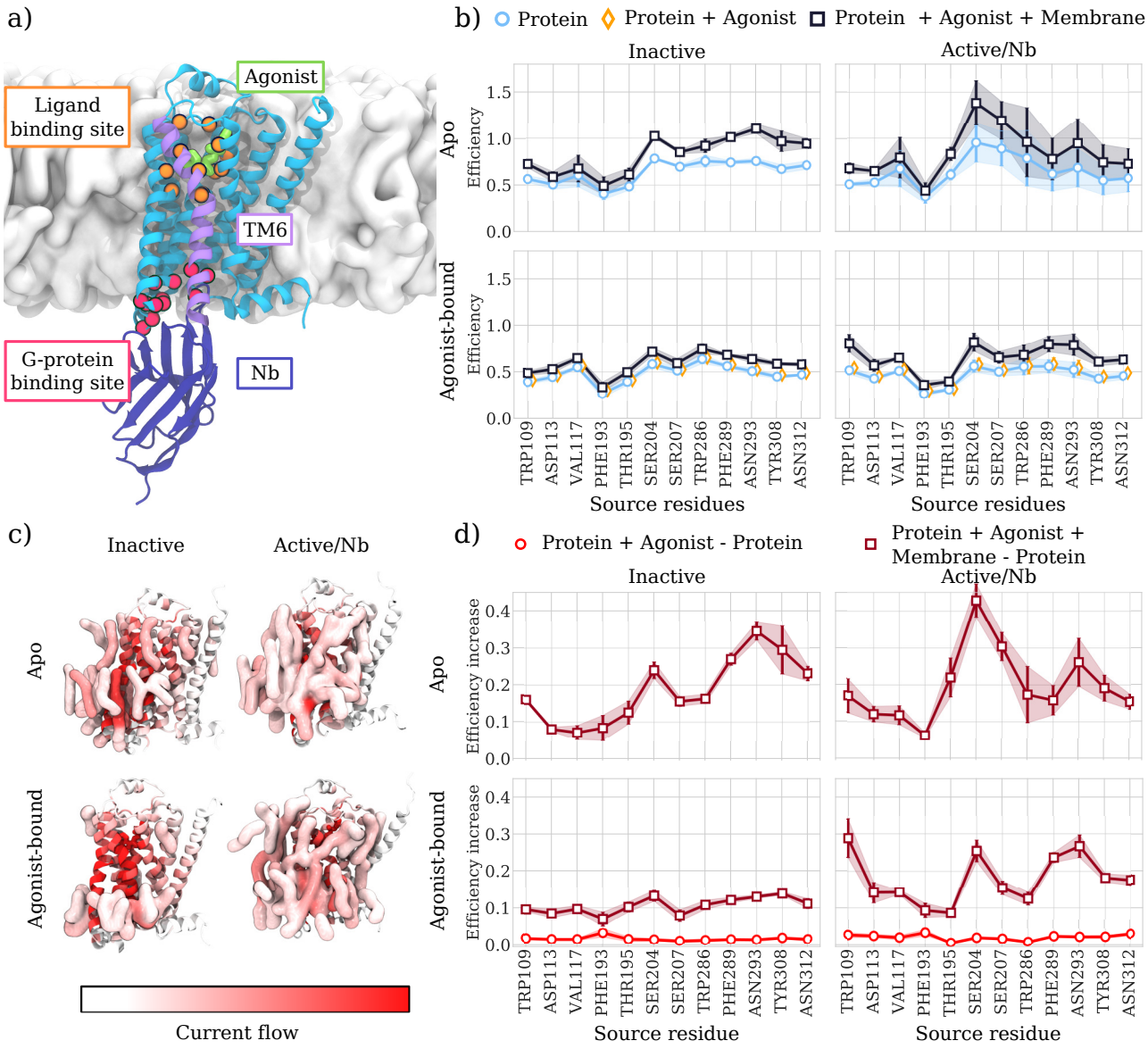

FIG. 3. a) Structure of active/Nb-bound β2AR, showing important structural features, and sources and sinks. b) State- and ligand-dependent allosteric efficiency (blue: protein, yellow: protein + agonist, black: protein + agonist + full membrane). Left column: inactive state, middle: active, and right: active/Nb. Upper row: apo β2AR, bottom row: agonist-bound β2AR. c) Average current flow through each protein residue and lipid interactors (current flow > 0.012) projected onto the first trajectory frame. Red signifies larger current flow values. d) Difference in allosteric efficiency of networks built by including different components of the β2AR systems. Dark red: comparing networks that include both protein and membrane to networks that only include the protein, and red: comparing protein/agonist networks to protein networks. Averages and SEM are obtained by splitting the simulations into three parts.

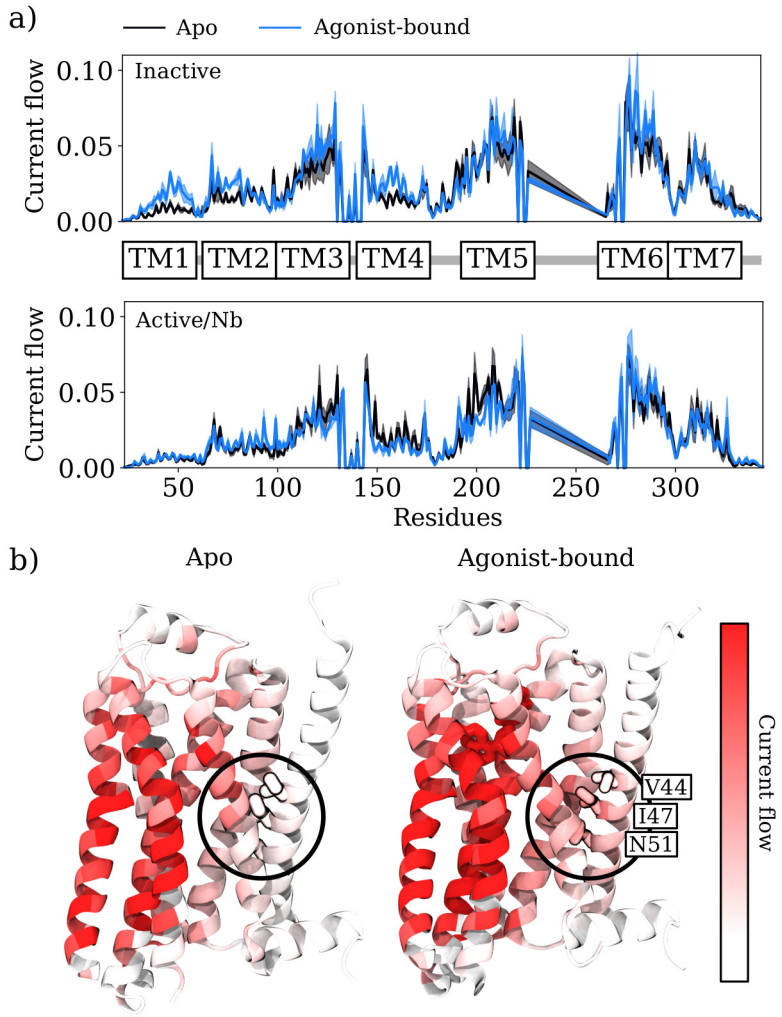

FIG. 4. a) Average and standard deviation of protein current flow plotted along residue sequence of apo (black) and agonist-bound (blue) β2AR in inactive (upper) and active/Nb (lower) states. b) Current flow projected onto apo (left) and agonist-bound (right) inactive structures. The black circle shows a region (encompassing TM1, TM2 and TM7) which becomes more involved upon agonist-binding. Averages and standard deviations are obtained by splitting the simulations into three parts.

eukaryotic potassium channels [49]. Mechanisms related to its gating and lipid regulation are therefore widely studied [49–52]. Two structural components are important for conduction of potassium ions in KcsA: the *selectivity filter* (SF) and the *inner gate*, Fig. 5a. The SF, which selectively coordinates potassium ions, is located at the extracellular side of the protein. The sequence of SF residues is highly conserved across potassium channels [53]. The inner gate, on the other hand, is located on the intracellular side of the protein. It is formed by the crossing of

the second transmembrane helix (TM2) from each subunit [49]. A pH drop in the cell causes the inner gate to open, yielding the KcsA open state channel. Within milliseconds after opening, the selectivity filter collapses and hinders ion permeation. This process, known as *inactivation*, is crucial to control electrical signals and membrane potentials in excitable cells [53]. The constriction of the SF, which is responsible for inactivation, is thus allosterically triggered by a structural change at the inner gate. [52] In addition to the inner gate transition, inactivation is facilitated by the binding of a DOPG lipid to each intersubunit surface, Fig. 5a. [51, 54]

We investigated the allosteric role of a bound DOPG molecule and of the lipid membrane in the open, inactivated and closed states using the inner gate residues A108, T112, G116 and E120 as sources and SF residues T75 to G79 as sinks, Fig. 5a. Fig. 5b shows the state-dependent allosteric efficiency of each system component; the protein, the protein and the bound DOPG, and the protein and the full membrane. The DOPG-unbound open state channel appears to allosterically rely on the membrane slightly less than the DOPG-unbound inactivated and closed states, Fig. 5b. The current flow projected onto structures and the most important lipids shows that the DOPG-bound open channel uses lipids located at the intracellular side of the protein to transmit allosteric signals, Fig. 5c. The inactivated and closed channels, on the other hand, exploit lipids from the intersubunit surface at the upper leaflet all the way down to the lower leaflet and the inner gate, Fig. 5c. Interestingly, DOPG binding appears to specifically increase the allosteric role of the lipid membrane in the open state channel, such that the bound lipid connects the upper and lower leaflet lipids, Fig. 5b,c.

To further investigate KcsA allostery related to the open state, we compared the current flow profiles between different states of DOPG-unbound and DOPG-bound channels, Fig. 6a, S3. This revealed differences at the central residues of TM2 helix, and especially at the TM1 helix, Fig. 6a,b, S3. Specifically, the allosteric importance of TM1 is increased in the DOPG-unbound open state channel. This may be a consequence of the decreased allosteric involvement of the membrane. We therefore hypothesize that the bound DOPG promotes inactivation by coupling the protein to the membrane, thereby altering the allosteric pathways within the protein.

V. CONCLUSIONS

We presented a framework which extends ordinary residue interaction networks to include cofactors such as lipids and ligands. By solving a linear assignment problem, and thus optimally

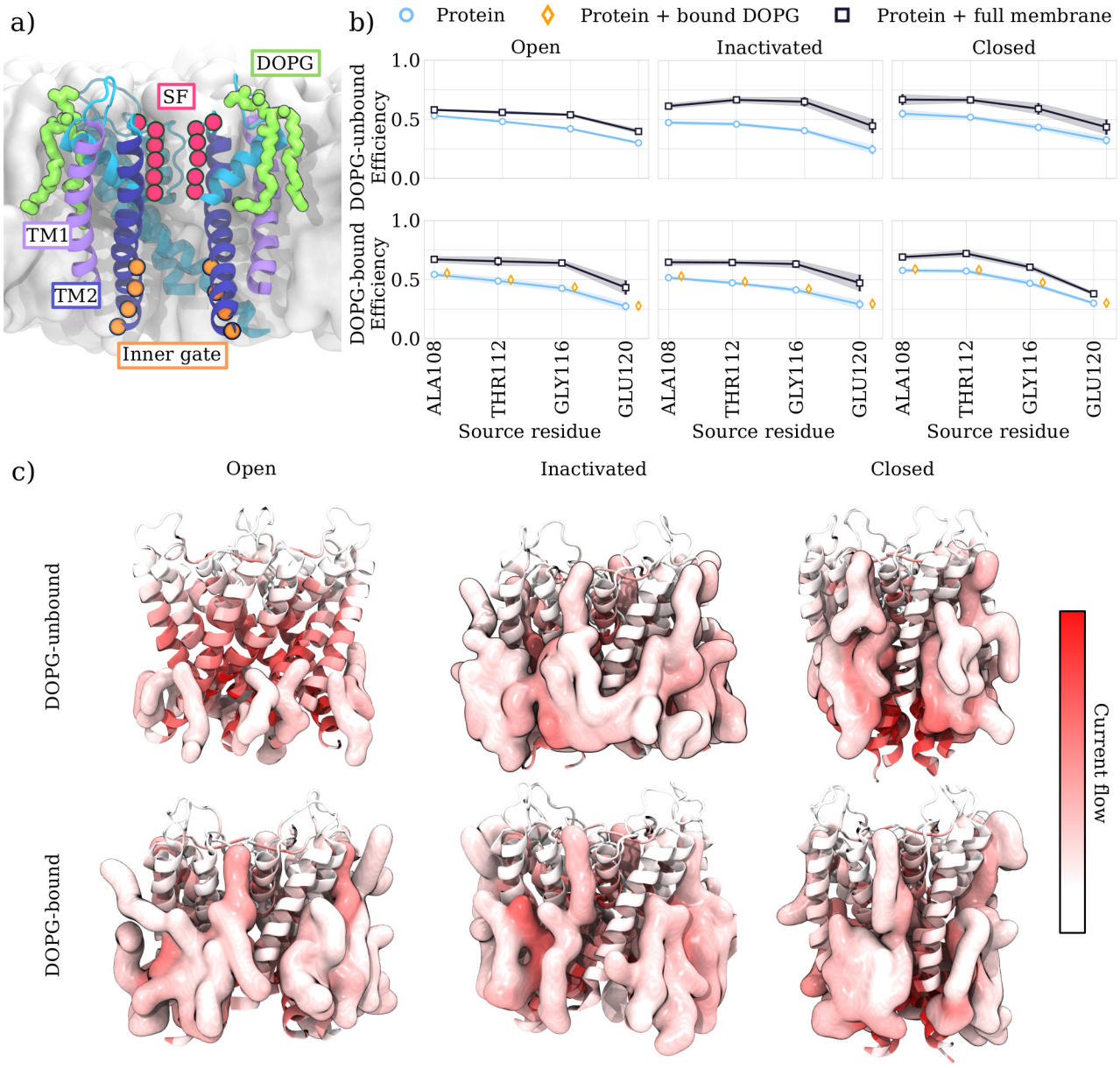

FIG. 5. a) Structure of open/DOPG-bound KcsA, showing important structural features, and sources and sinks. The subunit in front of the channel pore is omitted for clarity. b) State- and DOPG-dependent allosteric efficiency. Blue: protein, yellow: protein + bound DOPG, and black: protein + full membrane. Left column: open state, middle: inactivated state, and right: closed state. Upper row: DOPG-unbound channels, and bottom row: DOPG-bound channels. c) Average current flow through each protein residue and lipid interactor (current flow > 0.012) projected onto the first trajectory frame. Red signifies larger current flow values. Averages and SEMs are computed across subunits.

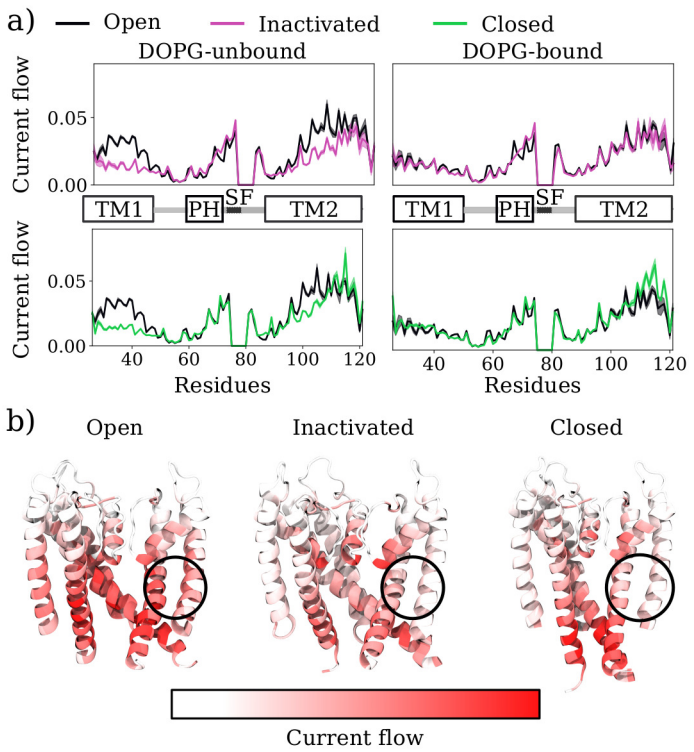

FIG. 6. Average and standard deviation of protein current flow plotted along residue sequence of open (black), inactivated (purple) and closed (green) states with (right) and without (left) bound DOPG. KcsA structural features are labeled along sequence: TM1, pore helix (PH), selectivity filter (SF), TM2. b) Current flow projected onto the DOPG-unbound protein structures. The circle shows the region on TM1 and TM2 which is more allosterically involved in the open/DOPG-unbound channel. The subunit in front of the channel pore is omitted for clarity. Averages and standard deviations are computed across subunits.

assigning interactors to nodes in each frame, the problem of cofactor redundancy is avoided. We applied this framework to obtain networks of three different types of membrane proteins and their cofactors. The network analysis revealed interesting features of protein allostery modulation by lipids and ligands. Importantly, we showed that the lipid membrane plays an integral role in the state-dependent allosteric communication of these proteins. The strong involvement of the membrane in KCNQ1 allostery may be attributed to the nonconvex structural topology of KCNQ1. The β2AR and KcsA datasets specifically demonstrated that subtle redirections of allosteric pathways induced by a modulator may change the membrane involvement. Network analysis including cofactors may thus serve to reveal intricate allosteric mechanisms of membrane proteins.

AUTHOR CONTRIBUTIONS

L.D. initiated and coordinated the project. L.D. and A.M.W. defined the project scope. A.M.W. developed and implemented the framework, and analyzed the datasets. O.F., S.P.C. and A.M.W. performed simulations of the $\beta 2$ adrenergic receptor, KcsA and KCNQ1, respectively. A.M.W. prepared the figures and the first manuscript draft. All authors participated in interpreting the results and writing the paper.

SUPPLEMENTARY MATERIAL

Supplementary material includes additional Methods section and Figs. S1 and S2, and Table S1. The SI Methods section describes MD simulation parameters and current flow calculations.

AIP PUBLISHING DATA SHARING POLICY

The data that support the findings of this study are openly available in the open science framework repository https://osf.io/2c6zd/, [55]. Code and tutorial for building and analyzing networks are available at https://www.github.com/delemottelab/allosteric-pathways [19].

ACKNOWLEDGMENTS

This work was supported by grants from the Science for Life Laboratory, the Gustafsson Foundation and the swedish research council (VR-2018-04905). The simulations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at PDC Centre for High Performance Computing (PDC-HPC). The authors thank Christian Blau for insightful discussions during method development.

---

[1] A. del Sol, C.-J. Tsai, B. Ma, and R. Nussinov, Structure **17**, 1042 (2009).

[2] Z. Cournia and A. Chatzigoulas, Current Opinion in Structural Biology **62**, 197 (2020).

[3] J.-P. Changeux and A. Christopoulos, Cell **166**, 1084 (2016).

[4] J.-P. Changeux, Drug Discovery Today: Technologies Allosteric modulation, **10**, e223 (2013).

[5] M. Canals, P. M. Sexton, and A. Christopoulos, Trends in Biochemical Sciences **36**, 663 (2011).

[6] N. R. Latorraca, A. J. Venkatakrishnan, and R. O. Dror, Chemical Reviews **117**, 139 (2017).

[7] S. B. Long, E. B. Campbell, and R. MacKinnon, Science **309**, 903 (2005).

[8] M. Jensen, V. Jogini, D. W. Borhani, A. E. Leffler, R. O. Dror, and D. E. Shaw, Science **336**, 229 (2012).

[9] J. Cui, Biophysical Journal **110**, 14 (2016).

[10] C. A. Ahern, J. Payandeh, F. Bosmans, and B. Chanda, Journal of General Physiology **147**, 1 (2016).

[11] A. I. Fernández-Mariño, T. J. Harpole, K. Oelstrom, L. Delemotte, and B. Chanda, Nature Structural & Molecular Biology **25**, 320 (2018).

[12] W. M. Botello-Smith, A. Alsamarah, P. Chatterjee, C. Xie, J. J. Lacroix, J. Hao, and Y. Luo, PLOS Computational Biology **13**, e1005711 (2017).

[13] N. Duclert-Savatier, G. Bouvier, M. Nilges, and T. E. Malliavin, Journal of Chemical Information and Modeling **56**, 1762 (2016).

[14] W. M. Botello-Smith and Y. Luo, Journal of Chemical Theory and Computation **15**, 2116 (2019).

[15] K. Park and D. Kim, BMC Bioinformatics **12**, S23 (2011).

[16] M. C. R. Melo, R. C. Bernardi, C. d. l. Fuente-Nunez, and Z. Luthey-Shulten, bioRxiv , 2020.06.18.160572 (2020).

[17] U. Brandes and D. Fleischer, in *STACS 2005*, Lecture Notes in Computer Science, edited by V. Diekert and B. Durand (Springer, Berlin, Heidelberg, 2005) pp. 533–544.

[18] M. E. J. Newman, Social Networks **27**, 39 (2005).

[19] A. M. Westerlund, "delemottelab/allosteric-pathways," (2020), library catalog: www.github.com.

[20] P. W. Kang, A. M. Westerlund, J. Shi, K. M. White, A. K. Dou, A. H. Cui, J. R. Silva, L. Delemotte, and J. Cui, bioRxiv , 2020.07.04.187161 (2020).

[21] R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L.-P. Wang, T. J. Lane, and V. S. Pande, Biophysical Journal **109**, 1528 (2015).

[22] F. Reinhard, O. F. Lange, J. S. Hub, J. Haas, and H. Grubmüller, Computer Physics Communications **180**, 455 (2009).

[23] R. Jonker and A. Volgenant, Computing **38**, 325 (1987).

[24] A. Volgenant, Computers & Operations Research **23**, 917 (1996).

[25] T. Kazmar, "gatagat/lap," (2020), library catalog: www.github.com.

[26] G. A. Cortina and P. M. Kasson, Bioinformatics **32**, 3420 (2016).

[27] A. P. Dempster, N. M. Laird, and D. B. Rubin, Journal of the Royal Statistical Society. Series B (Methodological) **39**, 1 (1977).

[28] A. M. Westerlund, T. J. Harpole, C. Blau, and L. Delemotte, Journal of Chemical Theory and Computation **14**, 63 (2018).

[29] M. R. Shirts and A. L. Ferguson, Journal of Chemical Theory and Computation **0** (2020), 10.1021/acs.jctc.0c00077.

[30] G. Schwarz, Annals of Statistics **6**, 461 (1978).

[31] W. Humphrey, A. Dalke, and K. Schulten, Journal of Molecular Graphics **14**, 33 (1996).

[32] J. Sun and R. MacKinnon, Cell **169**, 1042 (2017).

[33] S. Ghosh, D. A. Nunziato, and G. S. Pitt, Circulation Research **98**, 1048 (2006).

[34] L. Shamgar, L. Ma, N. Scmitt, Y. Haitin, A. Peretz, R. Wiener, J. Hirsch, O. Pongs, and B. Attali, Circulation Research **98**, 1055 (2006).

[35] J. Sun and R. MacKinnon, Cell **180**, 340 (2020).

[36] M. A. Zaydman, J. R. Silva, K. Delaloye, Y. Li, H. Liang, H. P. Larsson, J. Shi, and J. Cui, Proceedings of the National Academy of Sciences **110**, 13180 (2013).

[37] M. A. Kasimova, M. A. Zaydman, J. Cui, and M. Tarek, Scientific Reports **5**, 7474 (2015).

[38] W. I. Weis and B. K. Kobilka, Annual Review of Biochemistry **87**, 897 (2014).

[39] A. Manglik and A. C. Kruse, Biochemistry **56**, 5628 (2017).

[40] S. G. F. Rasmussen, H.-J. Choi, J. J. Fung, E. Pardon, P. Casarosa, P. S. Chae, B. T. DeVree, D. M. Rosenbaum, F. S. Thian, T. S. Kobilka, A. Schnapp, I. Konetzki, R. K. Sunahara, S. H. Gellman, A. Pautsch, J. Steyaert, W. I. Weis, and B. K. Kobilka, Nature **469**, 175 (2011).

[41] H. C. S. Chan, S. Filipek, and S. Yuan, Scientific Reports **6**, 34736 (2016).

[42] S. G. F. Rasmussen, B. T. DeVree, Y. Zou, A. C. Kruse, K. Y. Chung, T. S. Kobilka, F. S. Thian, P. S. Chae, E. Pardon, D. Calinski, J. M. Mathiesen, S. T. A. Shah, J. A. Lyons, M. Caffrey, S. H. Gellman, J. Steyaert, G. Skiniotis, W. I. Weis, R. K. Sunahara, and B. K. Kobilka, Nature **477**, 549 (2011).

[43] A. J. Venkatakrishnan, X. Deupi, G. Lebon, C. G. Tate, G. F. Schertler, and M. M. Babu, Nature **494**, 185 (2013).

[44] R. O. Dror, D. H. Arlow, P. Maragakis, T. J. Mildorf, A. C. Pan, H. Xu, D. W. Borhani, and D. E. Shaw, Proceedings of the National Academy of Sciences **108**, 18684 (2011).

[45] O. Fleetwood, P. Matricon, J. Carlsson, and L. Delemotte, Biochemistry **59**, 880 (2020).

[46] V. Isberg, B. Vroling, R. van der Kant, K. Li, G. Vriend, and D. Gloriam, Nucleic Acids Research **42**, D422 (2014).

[47] A. J. Venkatakrishnan, A. K. Ma, R. Fonseca, N. R. Latorraca, B. Kelly, R. M. Betz, C. Asawa, B. K. Kobilka, and R. O. Dror, Proceedings of the National Academy of Sciences **116**, 3288 (2019).

[48] D. A. Doyle, J. M. Cabral, R. A. Pfuetzner, A. Kuo, J. M. Gulbis, S. L. Cohen, B. T. Chait, and R. MacKinnon, Science **280**, 69 (1998).

[49] M. L. Renart, A. M. Giudici, C. Díaz-García, M. L. Molina, A. Morales, J. M. González-Ros, and J. A. Poveda, International Journal of Molecular Sciences **21**, 2554 (2020).

[50] W. Kopec, B. S. Rothberg, and B. L. de Groot, Nature Communications **10**, 5366 (2019).

[51] J. A. Poveda, A. M. Giudici, M. L. Renart, O. Millet, A. Morales, J. M. González-Ros, V. Oakes, S. Furini, and C. Domene, Biochimica et Biophysica Acta (BBA) - Biomembranes **1861**, 183029 (2019).

[52] L. Delemotte, Journal of General Physiology **150**, 1356 (2018).

[53] B. Hille, *Ion Channels of Excitable Membranes*, 3rd ed. (Sinauer Associates, Oxford University Press, Sunderland, Mass., 2001).

[54] J. Li, J. Ostmeyer, L. G. Cuello, E. Perozo, and B. Roux, Journal of General Physiology **150**, 1408 (2018).

[55] A. M. Westerlund, O. Fleetwood, S. Perez-Conesa, and L. Delemotte, (2020), https://osf.io/2c6zd/.