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, b2AR, 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, 17]. 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. In the Theory section, we explain how to 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 framework is applied in the Results section 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 [18, 19] 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 [20].

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) [21]. The analysis is performed on non-hydrogen atoms of the protein residues, ligand and lipids. MDtraj [22] 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

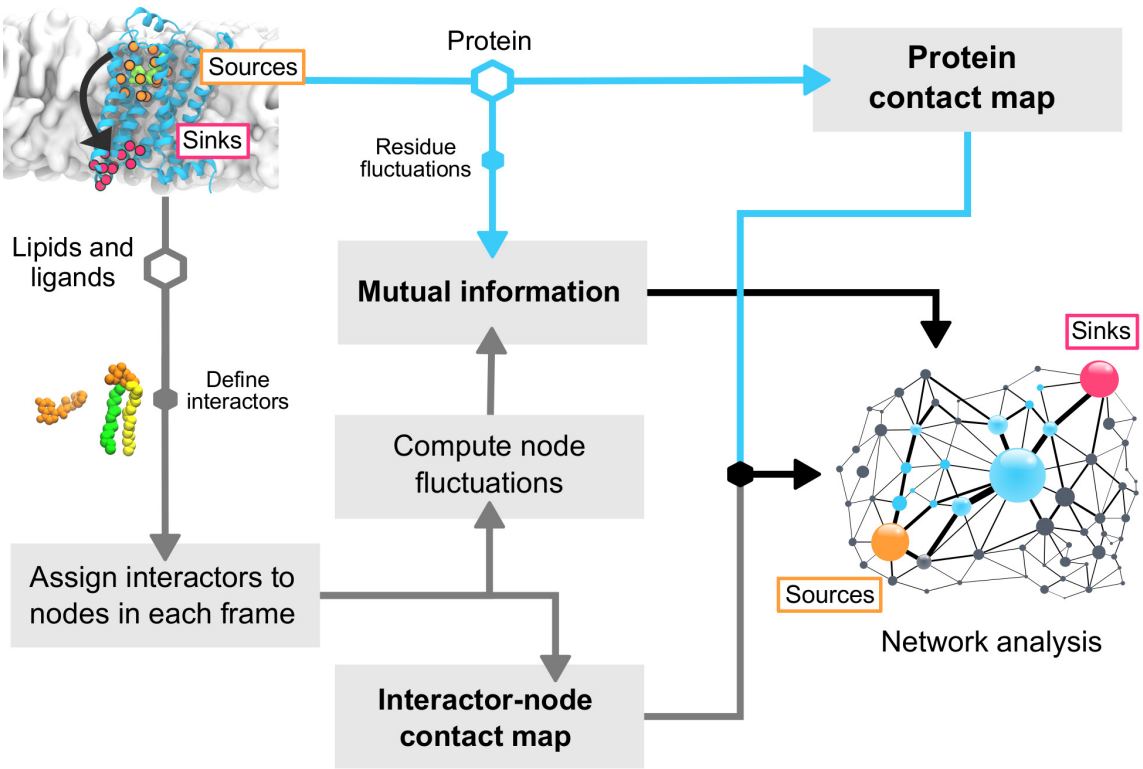

FIG. 1. Steps carried out to build a residue interaction network which includes lipids and ligands as well as the protein. Starting from a system with a protein embedded in a lipid membrane and bound ligand, the protein is first isolated. Each protein residue is assigned a node in the isolated protein 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.

divided into three interactors; one corresponding to the head group and two to the hydrophobic tails, Fig. 1, S1.

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 in each trajectory frame such that the network remains similar to the initial frame. In practice, for a specific frame, we wish to assign each interactor node to a node in the initial frame network. This is done by minimizing all centroid distances between pairs of the specific and initial frame nodes [23]. We solve this optimization problem with a linear assignment problem solver (the python package LAPJV) [24–26]. The centroid distances are computed with the minimum image convention using an orthorhombic box periodic boundary condition.

C. Continuous contact map

Continuous contact maps yield more stable networks than binary ones [14]. A switching-function, $K(d)$ introduced in [21] 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. The cutoff $c = 4.5$Å is suitable to define contacts between heavy atoms [27] and is commonly used in binary contact maps. We therefore 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 chosen d_{cut} allows for small deviations from the hard cutoff, c, to be included as weaker interactions. 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 [22]. 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 [21, 28]. 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) [29]; 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 [30, 31]. Given the number of Gaussians, GMM parameters (amplitudes, means and covariances) are estimated with expectation-maximization [29]. Increasing the number of Gaussians arbitrarily increases the likelihood. To choose a non-overfitting but detailed model, we use the Bayesian information criterion (BIC) [32], a criterion often used for mixture models [33]. BIC adds a penalty to the likelihood, which increases with model complexity. The GMM with the smallest BIC is selected. Ultimately, BIC is prior-free with respect to GMM parameters, and assumes a flat prior over models such that the most likely model is chosen [33]. For each density, we vary the number of Gaussians between 1 and 4.

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 INFORMATION FLOW ANALYSIS

Information flow, often called *current flow* after the original analogy to electrical networks [18, 19], 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]. As opposed to shortest path betweenness [34], current flow betweenness takes all pathways between source and sink nodes into account (details in Supplementary Material) [19]. The net flow across edges naturally removes uninteresting regions in which ineffective pathways might move back and forth before reaching the sink. Because of this, explicit pathways may sometimes be difficult to infer directly. Nonetheless, the result highlights nodes that carry effective pathways, making current flow betweenness well-suited to extract allosterically important residues and interactors.

Current flow betweenness was first introduced to biomolecular simulations in the context of a soluble enzyme [14]. A homomeric-averaged variant was recently presented and experimentally validated on KCNQ1 [21]. The current flow betweenness from each subunit was replicated and summed over the structure prior to averaging. To understand the role of cofactors for communicating allostery, we extended this analysis with a measure of signaling efficiency. Here, efficiency is characterized by *current flow closeness centrality* [18] (see Supplementary Material). We thus used current flow betweenness and closeness centrality to extract allosterically important protein residues, and to reveal the allosteric role of lipids and ligands. To avoid confusion with ion channel currents, we will use *information flow* to denote current flow betweenness and *information flow closeness* to denote current flow closeness.

We studied three well-known systems; the voltage-gated ion channel KCNQ1, the GPCR $\beta 2$ adrenergic receptor, and the pH-gated ion channel KcsA. Technical details about simulation setup and protocols are described in Supplementary Material. The structures were visualized with VMD [35].

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 [36]. 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 [37, 38], and binds to the KCNQ1 C-terminal domain (CTD) [36, 39]. Specific to KCNQ1 is that PIP_2 is required to couple VSD activation to pore opening [40, 41]. The two cryo-EM structures of human KCNQ1 [39] 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 [21].

A thorough investigation of the KCNQ1 protein information flow from the voltage-sensor domain to the pore, validated by electrophysiology, is presented elsewhere [21]. Here, on the other hand, 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 information 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 information flow through each residue projected onto the CTD-bent (upper) and CTD-straight (lower) structures. The lipid interactors with the highest information flow are shown in surface representation, indicating their location with respect to the protein. 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. Previous studies suggested that state-dependent PIP_2-binding is important for stabilizing different channel conformations [41]. Accordingly, our results suggest that the role of

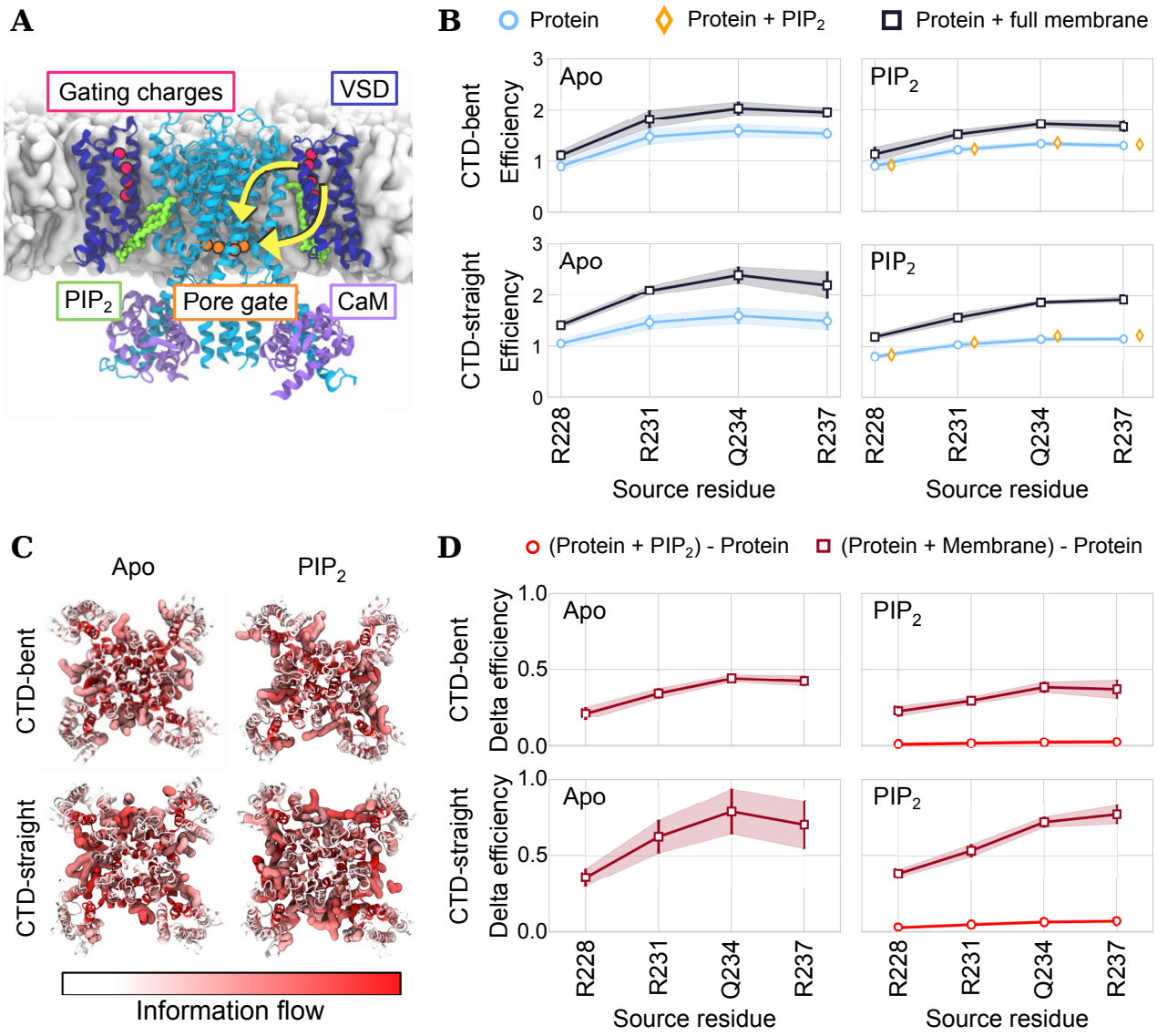

FIG. 2. **(A)** Structure of CTD-straight KCNQ1 with important structural features, and sources and sinks. The yellow arrows show cartoon pathways, highlighting the direction between the source and sink residues. **(B)** State- and PIP_2-dependent allosteric efficiency (information flow closeness) measured from the source residues listed on the x-axis to the sink residue S349. Blue: protein, yellow: protein+bound PIP_2, and black: protein+full membrane. Left column: apo, and right: PIP_2-bound. Upper row: CTD-bent state, and bottom row: CTD-straight. **(C)** Average information flow through each protein residue and lipid interactor projected onto the first trajectory frame (top view). Red signifies larger information flow values. Lipid interactors (information flow > 0.008) are shown in surface representation. **(D)** Difference in allosteric efficiency of networks including different components of the systems. Dark red: comparing protein/membrane networks to protein networks, and red: comparing protein/PIP_2 networks to protein networks. Averages and standard errors (SEM) are computed across subunits.

the bound PIP_2 is not 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, 42, 43]. 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) [44]. We applied the information 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 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 [45]. The sink nodes were taken as residues in contact with the G protein in the active state structure 3SN6 [46], 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 information 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

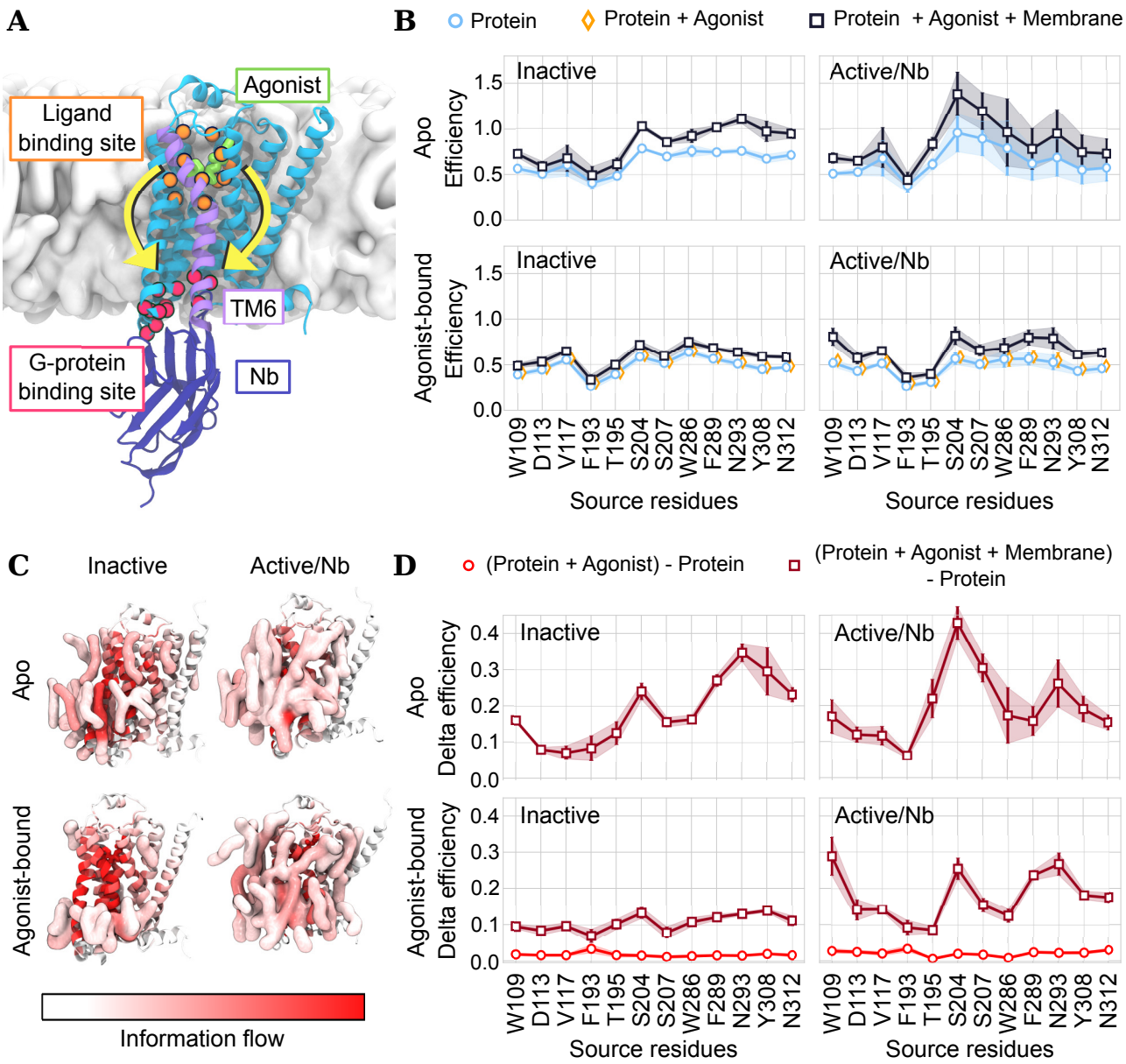

FIG. 3. **(A)** Structure of active/Nb-bound β2AR showing important structural features, and sources and sinks. The yellow arrows show cartoon pathways, highlighting the direction between the source and sink residues. **(B)** State- and ligand-dependent allosteric efficiency (blue: protein, yellow: protein + agonist, black: protein + agonist + full membrane). Left column: inactive state, and right: active/Nb-bound state. Upper row: apo β2AR, bottom row: agonist-bound β2AR. **(C)** Average information flow through each protein residue and lipid interactors (information flow > 0.012) projected onto the first trajectory frame. Red signifies larger information flow values. **(D)** Difference in allosteric efficiency of networks built by including different components of the β2AR systems. Dark red: comparing protein/agonist/membrane networks to protein networks, and red: comparing protein/agonist networks to protein networks. Averages and SEM are obtained by splitting the simulations into three parts.

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 information 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 [46–49]. 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 [50] and supports a stabilizing water network in the receptor's inactive state [51]. Notably, the inactive agonist-bound receptor also exhibited more information flow at TM3, compared to the apo inactive receptor. The difference between agonist-bound and apo information 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 [52]. The simplicity and high conservation of structural features make the channel a well-suited model for complex eukaryotic potassium channels [53]. Mechanisms related to its gating and lipid regulation are therefore widely studied [53–56]. 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 [57]. 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 [53]. 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, which hinders ion permeation. This process, known as *inactivation*, is crucial to control electrical signals and membrane potentials in excitable cells [57]. The constriction of the SF, which is responsible for inactivation, is thus allosterically triggered by a structural change at the inner gate [56].

Lipids are important for KcsA modulation. The membrane thickness, for example, affects

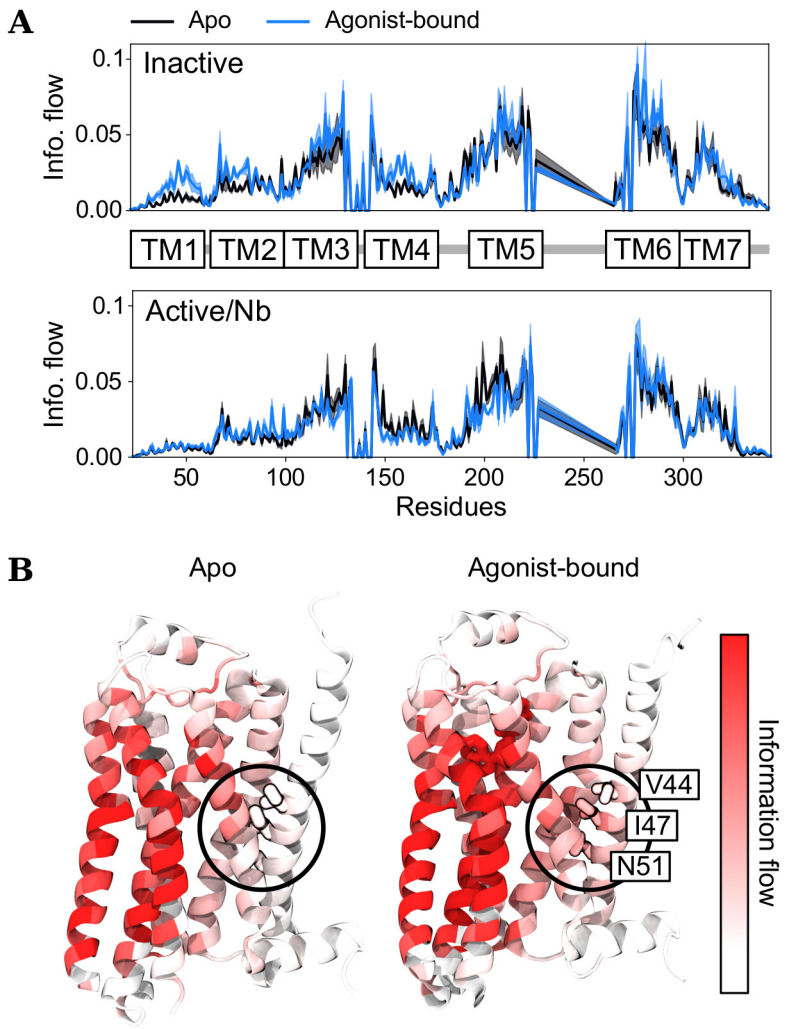

FIG. 4. **(A)** Average and standard deviation of protein information flow plotted along residue sequence of apo (black) and agonist-bound (blue) β2AR in inactive (upper) and active/Nb (lower) states. **(B)** information flow projected onto apo (left) and agonist-bound (right) inactive structures. The 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.

the open state probability of the E71A mutant [58]. In addition to this, experiments indicate that inactivation is facilitated by the binding of an anionic lipid to each intersubunit cavity, Fig. 5A [55]. These results are obtained by mutating residues at the lipid binding sites. Such mutations may, however, yield unknown structural effects. In contrast, the approach suggested here allows for accessing the molecular details in the wild-type channel, given the general limitations of molecular modeling.

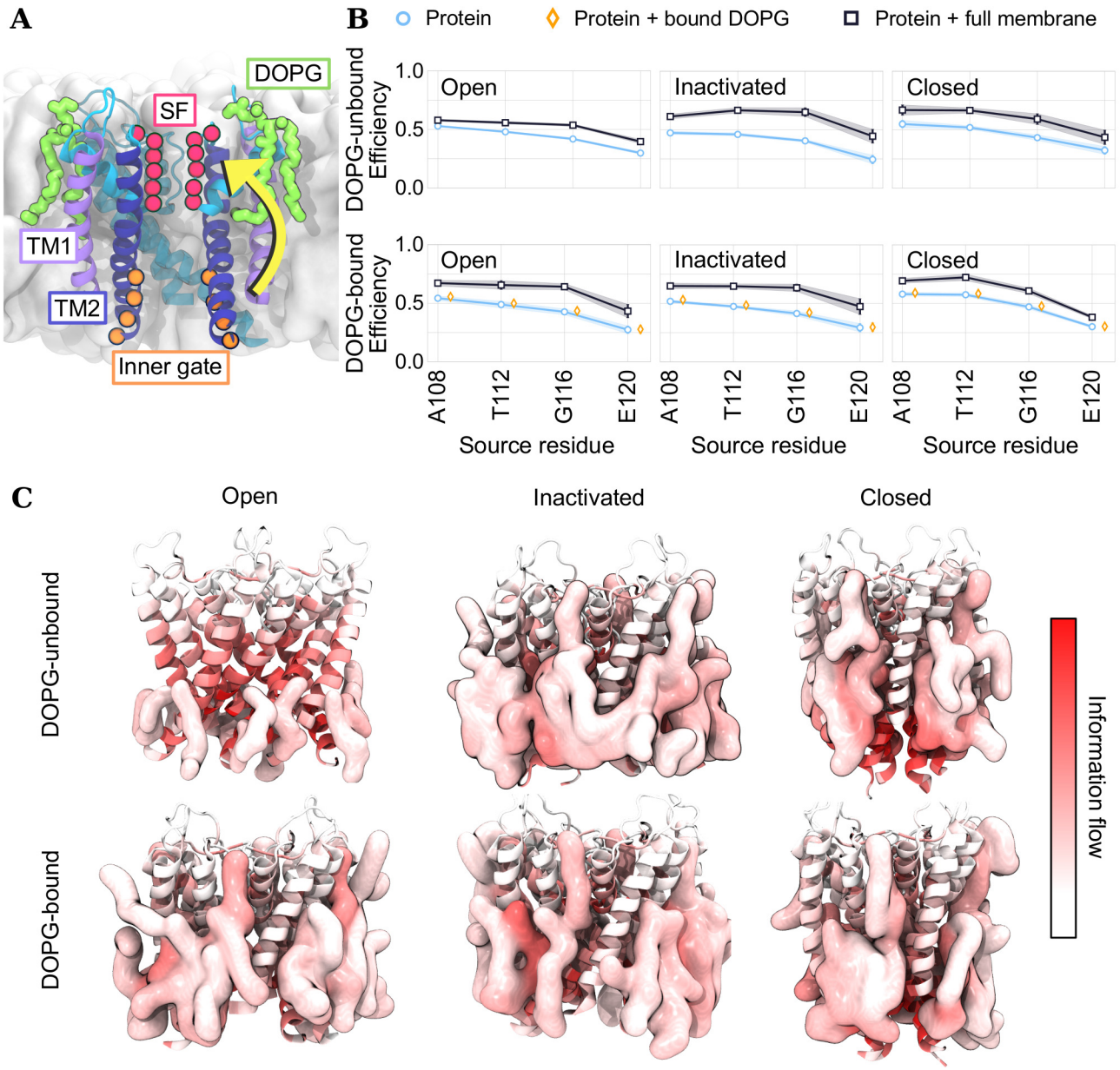

FIG. 5. **(A)** Structure of open/DOPG-bound KcsA showing important structural features, and sources and sinks. The subunit in front of the channel pore is omitted for clarity. The yellow arrow shows a cartoon pathway, highlighting the direction between the source and sink residues. **(B)** State- and DOPG-dependent allosteric efficiency (blue: protein, yellow: protein + bound DOPG, 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 information flow through each protein residue and lipid interactor (information flow > 0.012) projected onto the first trajectory frame. Red signifies larger information flow values. Averages and SEMs are computed across subunits.

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 channel appears to allosterically rely on the membrane slightly less than the DOPG-unbound inactivated and closed channels, Fig. 5B. 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 information 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 particularly at the TM1 helix, Fig. 6A,B, S3. The allosteric importance of TM1 (W26-L40) is increased in the DOPG-unbound open state channel. L40 is specifically important for the coupling between inner gate and SF at the open state [59]. The increased importance of this region in the DOPG-unbound state may be a consequence of the decreased allosteric involvement of the membrane. We 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 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

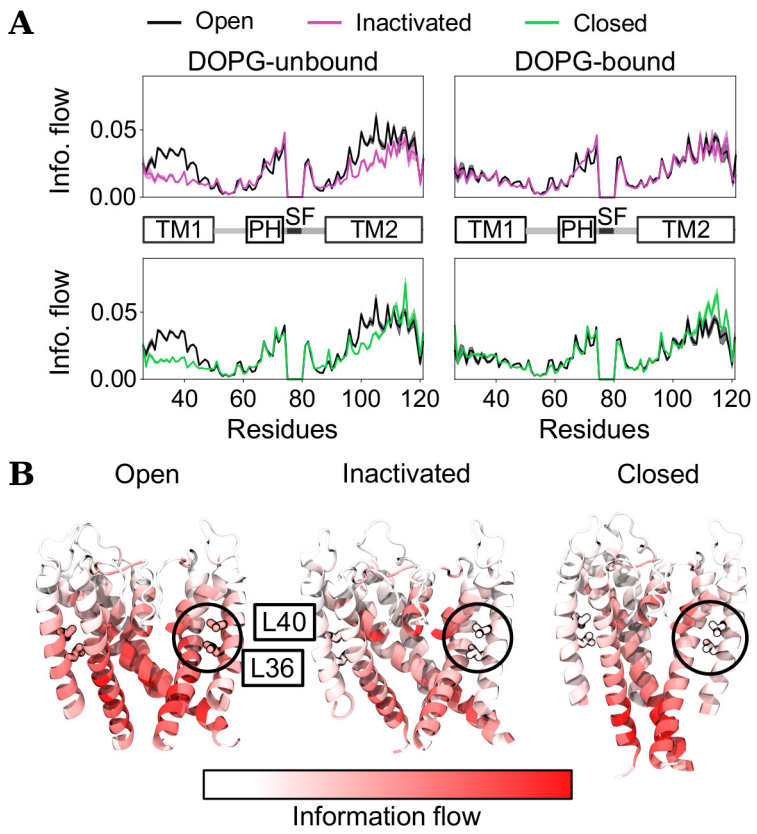

FIG. 6. **(A)** Average and standard deviation of protein information flow plotted along residue sequence of open (black), inactivated (purple) and closed (green) states with (right) and without (left) bound DOPG. KcsA structural features are labeled along sequence: TM1, pore helix (PH), selectivity filter (SF), TM2. **(B)** information 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.

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, Figs. S1-S3, and Table S1. The Methods section describes MD simulation parameters and information flow (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/, [60]. Code and tutorial for building and analyzing networks are available at https://www.github.com/delemottelab/allosteric-pathways [20].

ACKNOWLEDGMENTS

This work was supported by grants from the Science for Life Laboratory, the Göran 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] V. Corradi, E. Mendez-Villuendas, H. I. Ingólfsson, R.-X. Gu, I. Siuda, M. N. Melo, A. Moussatova, L. J. DeGagné, B. I. Sejdiu, G. Singh, T. A. Wassenaar, K. Delgado Magnero, S. J. Marrink, and D. P. Tieleman, ACS Central Science **4**, 709 (2018), publisher: American Chemical Society.

[18] 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.

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

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

[21] 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).

[22] 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).

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

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

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

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

[27] C. Yuan, H. Chen, and D. Kihara, BMC Bioinformatics **13**, 292 (2012).

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

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

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

[31] M. R. Shirts and A. L. Ferguson, Journal of Chemical Theory and Computation **16**, 4107 (2020).

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

[33] S. Konishi and G. Kitagawa, *Information Criteria and Statistical Modeling*, Springer Series in Statistics (Springer, New York, 2008).

[34] L. C. Freeman, Sociometry **40**, 35 (1977).

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

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

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

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

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

[40] 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).

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

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

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

[44] 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).

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

[46] 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).

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

[48] 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).

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

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

[51] 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).

[52] 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).

[53] 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).

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

[55] 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).

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

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

[58] K. M. Callahan, B. Mondou, L. Sasseville, J.-L. Schwartz, and N. D'Avanzo, Channels **13**, 424 (2019).

[59] F. T. Heer, D. J. Posson, W. Wojtas-Niziurski, C. M. Nimigean, and S. Bernèche, eLife **6**, e25844 (2017).

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