Dynamics of activation in the voltage-sensing domain of Ci-VSP

Spencer C. Guo,[1] Rong Shen,[2] Benoît Roux,[1,2,*)] and Aaron R. Dinner[1,3,†)]

[1)] *Department of Chemistry, The University of Chicago, Chicago, Illinois 60637*

[2)] *Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, Illinois 60637*

[3)] *Institute for Biophysical Dynamics, The University of Chicago, Chicago, Illinois 60637*

The *Ciona intestinalis* voltage-sensing phosphatase (Ci-VSP) is a membrane protein containing a voltage-sensing domain (VSD) that is homologous to VSDs from voltage-activated ion channels responsible for cellular excitability. Two crystal structures of Ci-VSD in putative resting and active conformations suggest a stepwise voltage-sensing mechanism involving translocation and rotation of the S4 helix. Relying on a theoretical framework based upon dynamical operators, we use mechanistic statistics estimated from an ensemble of unbiased molecular dynamics trajectories to elucidate the molecular determinants of the resting-active transition. Sparse regression with a small set of coordinates reveals that activation is primarily governed by the movement of arginine side chains through a central hydrophobic constriction, independent of global S4 movements. While translocation and rotation do provide a meaningful representation of the overall motion of the S4 helix, the analysis indicates that the activation mechanism cannot be accurately described without incorporating more fine-grained details such as the sequential exchange of salt bridges between arginines along S4 and negative countercharges along the S1–S3 helices. The importance of these salt bridges has been previously suggested by structural and functional studies. These results highlight how molecular dynamics simulations can now quantitatively characterize complex and functionally important conformational changes of proteins.

*)Electronic mail: roux@uchicago.edu

†)Electronic mail: dinner@uchicago.edu

I. INTRODUCTION

Voltage sensing in membrane proteins is an essential biophysical phenomenon that underlies many physiological processes. Experiments and models suggest that voltage-sensing proteins respond to changes in the transmembrane potential through movement of positively-charged amino acids (primarily arginines)[1–3], but a detailed understanding of the mechanism at the atomic level remains lacking. Despite the availability of several structures of voltage-gated ion channels (VGICs) in their active states[4–6], there are comparatively few structures of VGICs in their resting states without the perturbing effects of engineered crosslinks (metal binding sites and disulfide bridges) or toxins[7–11].

Like VGICs, the voltage-sensitive phosphatase from *Ciona intestinalis* (Ci-VSP) exhibits sensing currents in response to changes in transmembrane potential, as well as voltage-regulated enzymatic activity[12–15]. Furthermore, crystal structures of the voltage-sensing domain (VSD) from Ci-VSP in conformations thought to represent the resting (down) and active (up) states are available[16]. The down state is occupied by the wild-type protein at 0 mV applied voltage, whereas the up state is obtained by mutating a basic residue to an acidic one (R217E), which shifts the voltage dependence. These structures present a unique opportunity for elucidating a mechanism of voltage-sensing in atomic detail using molecular dynamics (MD) simulations.

Consistent with structures of a number of VGICs[4,6,15], Ci-VSD consists of four transmembrane helices, labeled S1–S4 (Fig. 1). The voltage-sensing response is mediated by five arginines (R217, R223, R226, R229, and R232) spaced by three residues along S4[1,3,12,15–17]; these are the analogs of the gating arginines in ion channels. These arginines form salt bridges with acidic side chains acting as countercharges, which are located on helices S1–S3: D129, E183, and D186. Studies that manipulate these interactions and corresponding ones in other VSDs indicate that these interactions are essential for the voltage-sensing response[7,8,18–20]. Finally, there is a group of three highly-conserved, bulky hydrophobic residues, I126, F161, I190, frequently referred to as the "hydrophobic plug" or "gasket", which serve as a charge transfer center for the sensing arginines by focusing the transmembrane electric field into a small region[21–24]. The structures of the down and up states and simulations suggest that the S4 helix undergoes a helical-screw mechanism in discrete steps ("clicks") that couple a

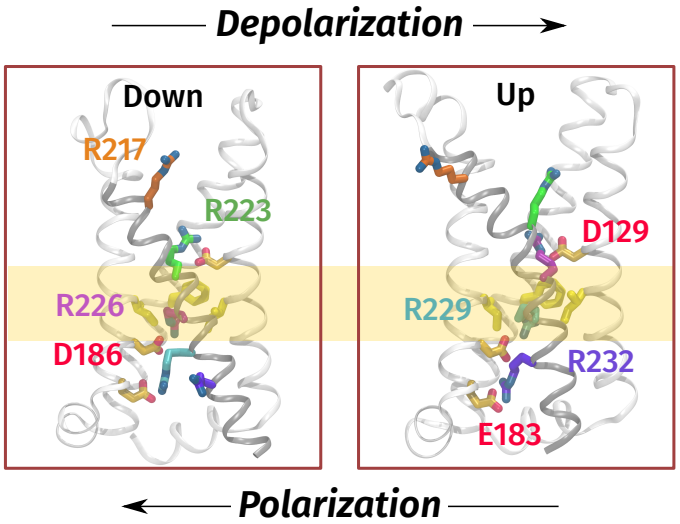

FIG. 1. Structure of Ci-VSD. Left, down (PDB: 4G80) and right, up (PDB: 4G7V) states. The side chains of sensing arginines on the S4 helix (gray) which participate in voltage-sensing, R217 (orange), R223 (lime), R226 (magenta), R229 (cyan), and R232 (purple), as well as of residues that form salt bridges with them (D129, E183, D186) are rendered as sticks. Nitrogen and oxygen atoms on these side chains are colored blue and red, respectively. Hydrophobic plug residues (I126, F161, and I190) are highlighted in yellow.

translocation of about 5 Å to a rotation of 60° around the helix axis[16,20]. This enables exchange of salt bridge partners between neighboring arginines. At the same time, the S1–S3 helices differ little in between the down and up states. There is evidence for states that are one click in each direction beyond the down and up states (down– and up+, respectively)[20].

Although there have been previous MD studies of VSDs[20,25–31], direct simulation of the transition between activated and inactivated states at physiological voltages remains computationally intractable. In ref. 32, the authors were able to recapitulate the gating cycle of the Kv1.2 channel by applying hyperpolarizing or depolarizing potentials of over 300 mV. In these simulations the S4 helices translocated 15 Å and rotated about 120° while the sensing arginines moved past a central hydrophobic residue in discrete steps. However, such strongly depolarizing voltages can impact the mechanism, and only a few transitions were observed during hundreds of microseconds of simulation time, so the statistical significance of these observations remains unclear. Enhanced sampling methods have been used to compute free energy landscapes for Ci-VSD and Kv1.2[20,31], but such calculations do not provide information about kinetics. Furthermore, they rely on *a priori* assumptions about the coordinates

that are important for transition.

Here, we use recently introduced methods to compute kinetic statistics from an ensemble of short MD trajectories[33,34] to characterize pathways and intermediate states for the down-up transition of Ci-VSD. We interpret the simulations through sparse regression on the committor, a measure of the progress of a stochastic reaction[35]. By combining regression on the committor with statistics that reveal the pathways between states and their relative free energies, we find that voltage activation in Ci-VSD is described by stepwise rearrangements of salt bridges involving sensing arginines. These rearrangements are coupled more tightly to S4 helix rotation than translocation; the S4 helix can translocate a few ångstroms up or down without strongly affecting the activation. Indeed, the transition state ensemble is primarily distinguished by S4 rotation of nearly 50° and the presence of a salt bridge between R226 and D129. In general, however, to adequately describe the dynamics of activation requires multiple coordinates, and the free energy landscape along local coordinates such as salt bridge distances is rough. On the other hand, we find that the dynamics along the displacement charge, an experimentally-relevant coordinate capturing system-wide charge transfer, are relatively smooth and closely track the activation progress.

II. RESULTS AND DISCUSSION

As described in Sec. IV (Theory and Methods), we used unbiased MD simulations to generate trajectories that each sampled a portion of the transition between the down and up states and used a framework based on dynamical operators (dynamical Galerkin approximation, henceforth DGA) to compute potentials of mean force (PMFs), committors, and reactive currents from these data. These statistics provide information about the relative stability of key states (reactants, products, intermediates), the progress of the reaction (reaction coordinate), and reaction pathways, respectively. By combining them, we can identify metastable intermediates and transition states as well as the pathways connecting them. For visualization, we project these statistics onto pairs of physically intuitive features, which we identify by modeling the committor with sparse regression. The workflow is summarized in Fig. 2.

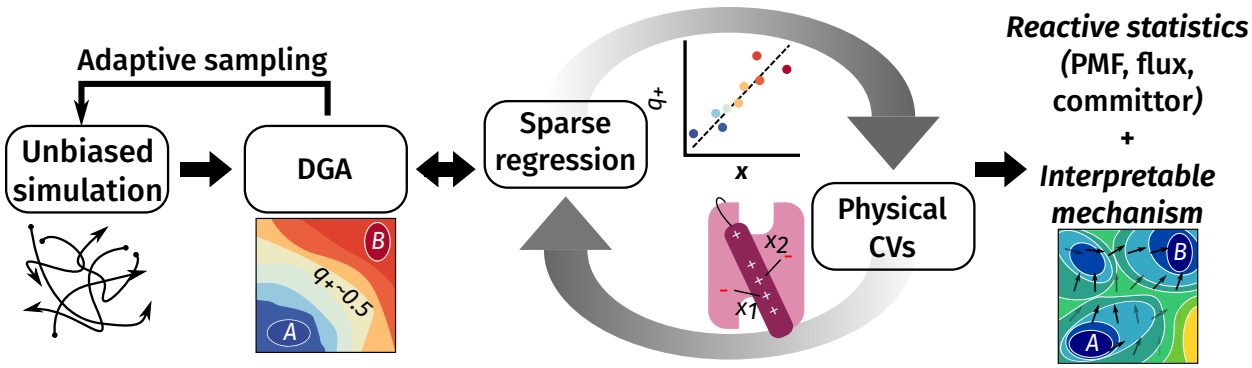

FIG. 2. Computational workflow for this study.

A. Decoupling of S4 helix and sensing arginine conformational changes

In light of previous studies focusing on the S4 helix[20,32,36,37], we first investigated its role in the down-up transition. The down state crystal structure lies approximately at $(-4.5\ \text{Å}, -60°)$ in the translocation-rotation plane (Fig. 3), indicating that during the down-up transition, the S4 helix translocates 4.5 Å upwards and rotates 60° clockwise (as seen from the extracellular side of the membrane). Shen *et al.*[20] computed the PMF as a function of the translocation and rotation of the S4 helix using replica exchange umbrella sampling (REUS), and we compare the PMF that we obtained with DGA to theirs in Fig. 3a. The two PMFs are overall qualitatively similar but diverge near the down state, where the DGA PMF lacks a local minimum. Moreover, the DGA PMF contains a shallow, wide minimum with a depth of about 1–2 kcal/mol near $(-5\ \text{Å}, -10°)$. This intermediate contains structures in which the S4 helix is rotated and R226 is located above the hydrophobic plug. Because the REUS simulations are biased to sample the PMF as a function of these coordinates, while the DGA PMF comes from a dataset of unbiased simulations with starting points distributed between the down– and up+ states (homology models that shift the sensing arginines one click past the down and up states[16,20]), we expect the REUS PMF to be more reliable in the region around the down state. By the same token, the region at $(-5\ \text{Å}, -10°)$ may represent a previously unappreciated metastable state of the model, though the REUS PMF does have a similar free-energy minimum between the up and up+ states (which persists as the membrane potential is varied[20]). Below, we proceed with the analysis, but these differences are important to keep in mind while evaluating the results.

As a first step toward understanding the dynamics of the transition, we computed committors for all structures and projected the values onto the translocation and rotation variables (Fig. 3b). By definition, the structures near the down (circle) and up (triangle) states have $q_+ = 0$ and $q_+ = 1$, respectively. The intermediate with R226 above the hydrophobic plug described above has $q_+ \approx 0.4$, which means that it is on the down side of the transition state. Although the committor generally increases from the lower left to the upper right of Fig. 3b, it rises rapidly just below the up state. As a result, the transition state ensemble (TSE)—that is, where $q_+ = 0.5$—lies relatively close to the up state. We also find that the variance of the committor is high in this region (Supplementary Fig. 1), suggesting that additional variables play important roles.

Another feature of note is that the normal vector to the $q_+ = 0.5$ isocommittor surface (i.e., ∇q_+) points nearly vertically in Fig. 3b, that is, parallel to the rotation axis. Because the gradient of the committor points in the direction of "productive" flux[38], this suggests that S4 rotation is more important at that point than S4 translocation for determining the progress of the reaction. This idea is reinforced by the reactive current (Supplementary Fig. 2a), which also accounts for the likelihoods of states.

If rotation is more important, then what is the role of S4 translocation? One possible explanation is that the down-up transition follows a "sliding-helix" mechanism[7], in which conversion to a 3_{10} helix, which is narrower than an α-helix, allows the side chains to move. However, we find that the propensity of S4 helix residues to convert from an α-helix to a 3_{10} helix is low (Supplementary Fig. 3) across the range of committor values, consistent with prior analyses[20]. We thus favor the "helical-screw" model instead. However, we observe in our simulations that vertical translocation of the S4 helix frequently precedes or lags behind the exchange of salt bridge partners by the sensing arginines (Fig. 3c, d and Supplementary Fig. 4). The observation that the S4 helix and sensing arginine side chains can move independently argues against a model where the entire S4 helix (backbone and side chains) moves concertedly, as pictured in the simplified sliding helix or helical-screw mechanisms. Put together, our computational analysis suggests that to model activation accurately it is necessary to account for both the global movements of the S4 helix and the local dynamics of the sensing arginines.

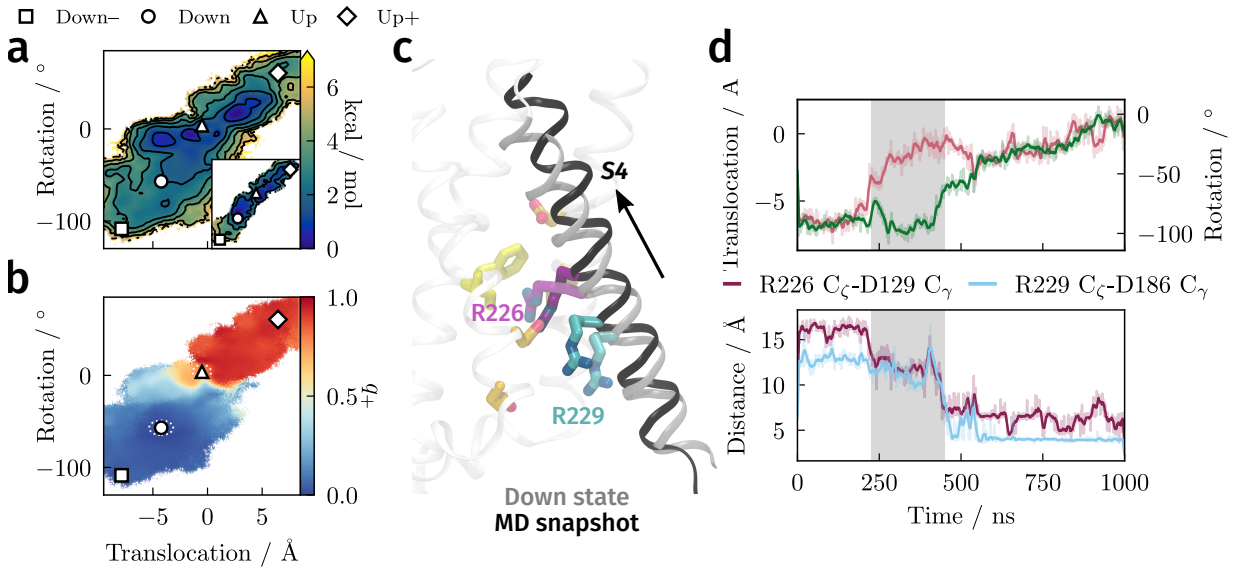

FIG. 3. Helical-screw model of voltage-activation and lag of side chain movements. **a**, PMF projected onto S4 helix translocation and rotation computed using DGA from unbiased trajectories. (Inset) Reference PMF computed using umbrella sampling simulations reported in ref. 20. **b**, Forward committor with approximate cutoffs for down and up states marked as dotted white ellipses. **c**, A frame from a simulation (opaque) superimposed against the down state crystal structure (translucent) after aligning the S1–S3 helices. Only the S4 helix, R226, and R229 of the MD simulation frame are shown for clarity. **d**, Top, translocation (pink) and rotation (green) of S4 helix; bottom, salt-bridge distances during a representative 1 µs trajectory showing translocation of the helix ahead of rotation and movement of the side chains. Light lines show instantaneous values; dark lines show moving averages over ten frames.

B. Sparse modeling of the committor

Out of the 60 inter-residue distances used as input to the DGA committors, we selected several to track the movements of the side chains of the sensing arginines, as well as those of their acidic salt-bridge partners. These include distances between the sensing arginine C_ζ atoms and the C_γ/C_δ atoms of D129, E183, and D186 (Fig. 3d). We also used the distance between the C_ζ atom of R226 and the center of mass of F161's side chain to track the movement of R226 through the hydrophobic plug, which is thought to present a barrier between the deactivated and activated states[22,23]; we chose F161 instead of I126 or I190 because it is highly conserved among voltage-sensitive proteins (corresponding to F290 in

the *Shaker* potassium channel for example)[22].

Following previous work[39,40], we constructed an interpretable statistical model of the committor using LASSO regression and identified physically-intuitive collective variables (CVs) that predict whether the reaction proceeds. We initially fit the model to all points with $q_+ \in [0.2, 0.8]$. However, in analyzing the model, we found that some CVs that we expected to contribute had small or zero-valued coefficients because the committor varies non-monotonically with them (e.g., the distance between R226 and F161 decreases and then increases as R226 goes through the hydrophobic plug in between the down and up states). Consequently, we divided the reaction into two stages, one from the down (reactant) state to the TSE, and another from the TSE to the up (product) state.

We found that there are nine CVs that contribute significantly to at least one of the two models. They are the S4 helix translocation and rotation and distances involving R226, R229, and R232 (Fig. 4a, b). To verify that these CVs are indeed sufficient to express the committor (and hence represent a useful subspace of variables to analyze), we constructed coordinates from linear combinations of input variables using the integrated variational approach to conformational dynamics (IVAC)[41], a method that robustly identifies slow modes of a system. The committor monotonically follows the first nontrivial IVAC coordinate (Fig. 4c), suggesting that the down-up transition is fully captured using our chosen subset of variables. Moreover, the committor projection on the IVAC coordinates has much lower variance compared with the projection on the S4 coordinates alone (Supplementary Fig. 1), reflecting the information gained from distances tracking the sensing arginine movements.

C. Roles of the S4 helix and associated salt bridges

We interpreted the results of the LASSO regression with the aid of the distributions of our selected CVs in the down, up, and transition states, as well as projections of the PMF, committors, and reactive currents onto those CVs. Below, we detail in turn the roles of the S4 helix and sensing arginines in the down-up transition.

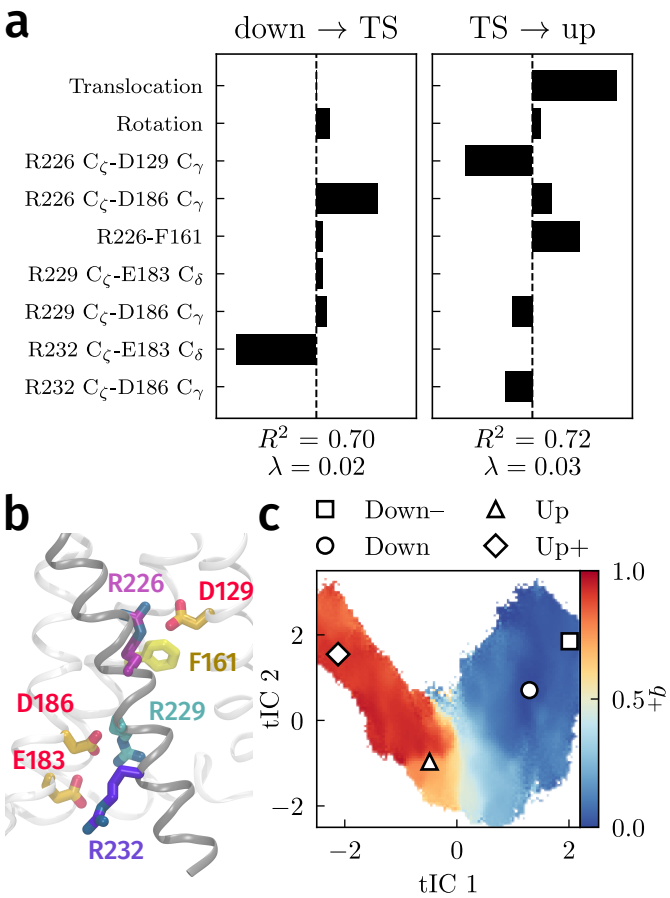

FIG. 4. Dependence of the committor on collective variables (CVs). **a**, Coefficients computed from LASSO regression for points with $q_+ \leq 0.5$ and $q_+ \geq 0.5$. **b**, Residues used to define intramolecular distances for regression. **c**, Committor projected onto the first two nontrivial IVAC coordinates (tICs) constructed from the CVs in panel **a**.

1. S4 helix rotation is more tightly constrained than translocation in the TSE

For both $q_+ \leq 0.5$ and $q_+ \geq 0.5$, the rotation of the S4 helix contributes a small but statistically significant positive coefficient to the fit (Fig. 4a). In contrast, the coefficient corresponding to translocation of the S4 helix is nonzero only for $q_+ \geq 0.5$. Consistent with this finding and the projection of statistics onto these CVs (in particular, the nearly horizontal line representing the TSE in Fig. 3a), the distributions of translocation in the down and transition states overlap. By contrast, rotation values in the TSE are relatively tightly clustered and close to those of the up state (Fig. 5a and Supplementary Fig. 2b).

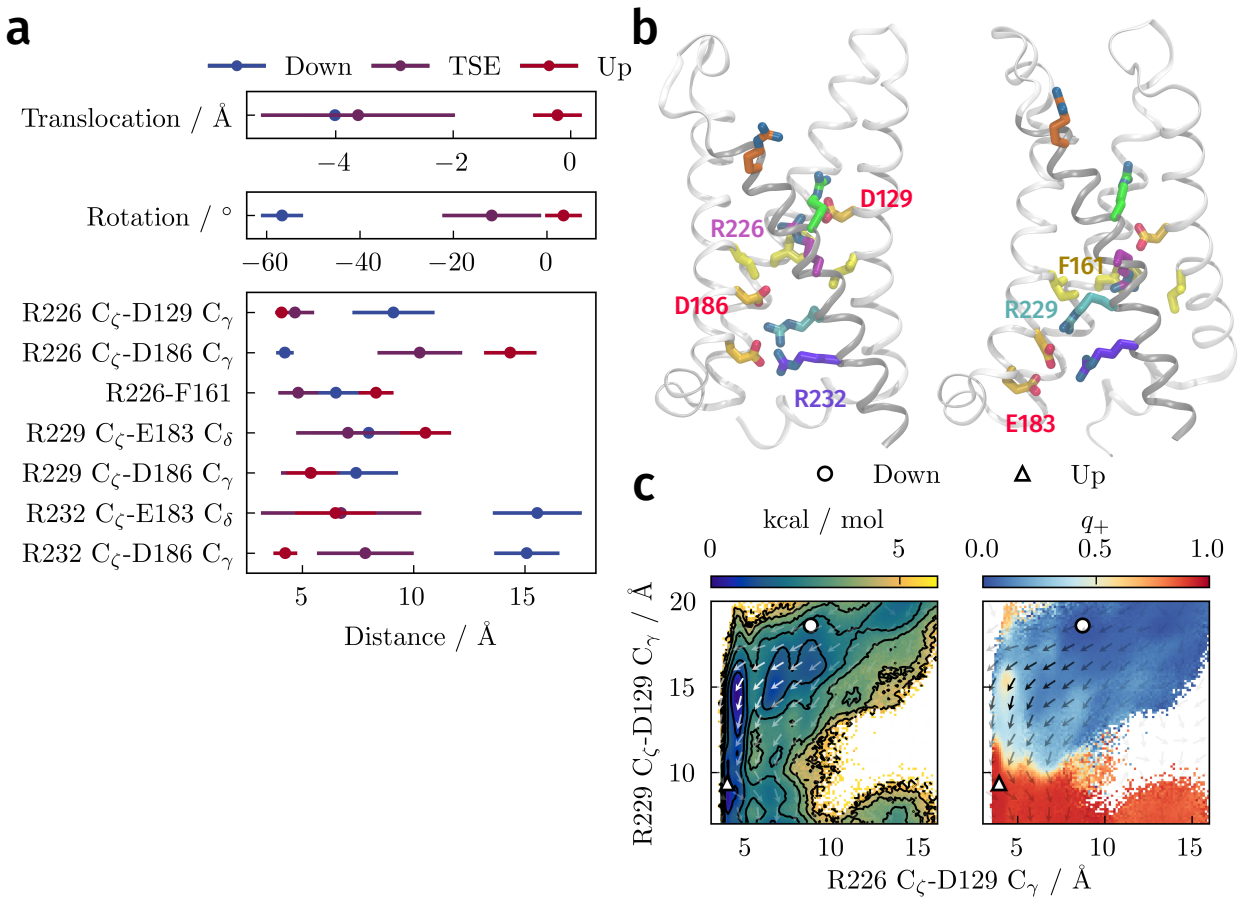

FIG. 5. Transition state ensemble and role of sensing arginine salt bridges. **a**, Average of selected CVs in the down state, in the TSE ($q_+ \in [0.45, 0.55]$), and in the up state. **b**, Selected structures in the TSE from k-medoids clustering. **c**, Reactive current as a function of selected salt-bridge distances plotted on PMF (left) and committor (right). Opacity of the arrows corresponds to the magnitude of the current.

The wide range of translocation values is consistent with our previous observation that the helix can move up and down a few ångstroms without affecting the progress of the reaction significantly (see Fig. 5b for two TSE conformations with similar rotation states and different translocations). Beyond this range, translocation continues to increase to the up state.

2. Arginines exchange electrostatic interactions with acidic residues and lipid phosphates

The LASSO regression yielded nonzero coefficients for a number of distances involving sensing arginines, in particular, R226, R229, and R232. We consider each in turn. In the

down state, R226 forms a salt bridge with D186 (Fig. 1). Based on the distributions of the R226 C_ζ-D186 C_γ, R226-F161, and R226 C_ζ-D129 C_γ distances (Fig. 5a and Supplementary Fig. 5a), we interpreted the coefficients for these distances as tracking R226's movement away from D186 when $q_+ \leq 0.5$, through the hydrophobic plug (as evidenced by the coefficients for the R226-F161 distance and the fact that its the R226-F161 distance is smallest in the TSE), and toward D129 when $q_+ \geq 0.5$. As early as $q_+ < 0.3$, the R226 C_ζ-D129 C_γ distance decreases (Fig. 5c, right) and the number of hydrogen bonds between R226 and D129 increases (Supplementary Fig. 5b). In the TSE, the guanidinium group of R226 is very close to the phenyl side chain of F161 (Fig. 5a and Supplementary Fig. 6a, b). Although cation-π interactions and π-π stacking have been suggested to stabilize interactions between arginine and phenylalanine residues[42–44], such interactions are not explicitly included in the force field. Rather, R226 can make a salt bridge with D129 at the same time (Supplementary Fig. 6c), and this interaction stabilizes the transition state. Consistent with the discussion of the S4 helix, R226 can interact with F161 and D129 irrespective of S4 vertical translocation. That said, as discussed above, the rotation of the S4 helix is tightly defined in the TSE, so R226 movement and S4 rotation must be coupled.

The LASSO coefficients for distances involving R229 for $q_+ \leq 0.5$ are relatively small. This is because the R229-E183 distance exhibits two populations: one centered around 4 Å and another from 6–10 Å (Supplementary Fig. 5a). In the former, R229 is coordinated to both E183 and D186, whereas in the latter R229 is only coordinated to D186 (Fig. 5b). Consistent with the distances, R229 forms hydrogen bonds mainly with D186, but also with E183 and lipid phosphate groups when $q_+ < 0.5$ (Supplementary Fig. 5a, b and Supplementary Fig. 7). Moreover, there is a trough in the free-energy at a R229 C_ζ-D186 C_γ distance of 4.5 Å (Supplementary Fig. 8). At $q_+ \geq 0.5$, R229 no longer interacts with E183 (Supplementary Fig. 5a), and the LASSO coefficients indicate that it moves toward D186. It maintains this salt bridge even as its side chain otherwise occupies the hydrophobic plug in the up state.

In Fig. 5c, we examine reactive statistics for R226 and R229 together to clarify their relationship. Note that the R229-D129 distance is not analyzed in the LASSO regression since it does not correspond to a salt bridge involved in the down-up transition. We use it nonetheless since we found the projection to be more informative than others (Supplementary Fig. 8). In Fig. 5c (left), we observe a free-energy minimum corresponding to a R226

C_ζ-D129 C_γ distance of 4.5 Å, consistent with the fact that R226 passes through the plug to interact favorably with D129 in initial stages of the reaction. Fig. 5c (right) shows that the TSE lies near the up state, consistent with the S4 helix CVs. By plotting the vector field of reactive flux on top of the PMF and committor (Fig. 5c, arrows), we see that R226 moves towards D129 before R229 moves toward its position in the up state. This presumably reflects the fact that steric hindrance prevents R229 from moving before R226 has already vacated its original location.

Finally, we consider R232. In the down state, the guanidinium group of R232 forms hydrogen bonds to phosphate groups of lipids (Supplementary Fig. 5b and Supplementary Fig. 7a, c), a role previously noted in studies of VSDs[25,32,45–47]. Over the course of the down-up transition, R232 loses these interactions (Supplementary Fig. 5b). In the TSE, the R232 C_ζ-E183 C_δ distance distribution overlaps with that of the up state (Fig. 5b). After the TSE, R232 moves toward E183 and then D186, forming stable hydrogen bonds with both (Supplementary Fig. 5c). We therefore interpret the negative coefficients for the R232-E183 and R232-D186 distances as tracking the movement of R232 from the lipid bilayer's inner leaflet to interacting with E183 when $q_+ \leq 0.5$ and, from there, to interacting with D186 when $q_+ \geq 0.5$.

D. Dynamics along the displacement charge coordinate

Next, we relate the displacement charge, Q_d, to our reactive statistics. The displacement charge represents the energetic coupling of the membrane protein system with an external potential[48]. The change in the displacement charge upon a conformational transition of the voltage-sensor, ΔQ_d, is a quantity that can be measured with voltage-clamp electrophysiology[3,49] and is commonly referred to as a sensing charge (gating charge in VG-ICs). The displacement charge of different conformational states of the VSD can also be calculated using MD simulations[31,48]. Ci-VSD was previously found[16,20] to transfer approximately 1.1 e across the potential from the down state to the up state; our calculations yield a sensing charge of $\Delta Q_\mathrm{d} = 0.9 \pm 0.3\ e$, agreeing with previous estimates (Supplementary Fig. 9a).

The committor closely tracks with Q_d (Fig. 6a and Supplementary Fig. 9b), which indicates

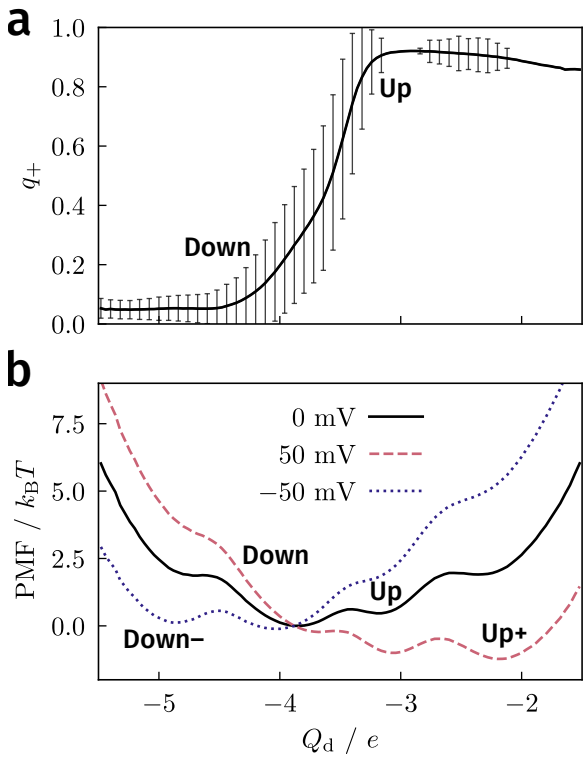

FIG. 6. Displacement charge. **a**, The committor averaged as a function of Q_d with the bars indicating the associated variance in q_+. The down and up states correspond to Q_d values of -4.2 ± 0.2 and -3.3 ± 0.2 e, respectively. **b**, PMF computed from DGA along Q_d. The free-energy profiles at nonzero voltages were obtained by adding a linear component $Q_d V$ to the profile obtained under 0 mV.

that the microscopic charge movement from the down to the up state is a good reporter of the progress of the conformational transition. The variance is relatively high in the transition region, though, reflecting the presence of other degrees of freedom. We are unable to decompose the gating charge into per-residue values due to the nature of the simulations, but the increase in Q_d through the transition region most likely is related to upward movements of R226 and R229 (Supplementary Fig. 9a), given their proximity to the hydrophobic plug where the membrane electric field is focused[21–23].

Using DGA, we also computed the PMF as a function of Q_d. At zero potential, neither the down nor the up state is strongly favored, and the two associated minima are separated by a barrier of less than 1 $k_B T$ (Fig. 6b, black line). In the presence of an external potential V, the free energy shifts by $Q_d V$[48]. In Ci-VSD, relatively mild hyperpolarizing or depolarizing voltages on the order of ∓ 50 mV are sufficient to tilt the PMF in favor of the down or up

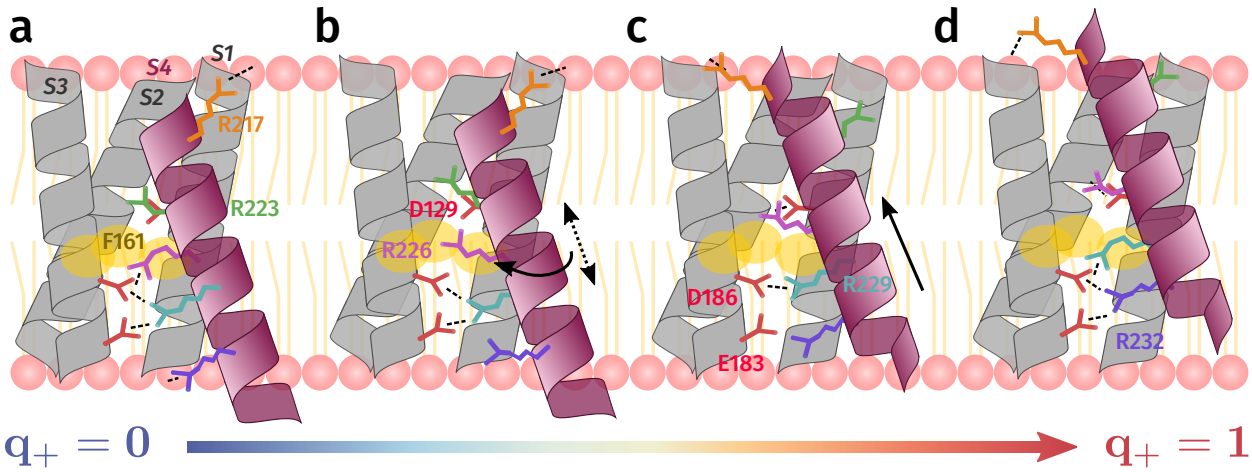

FIG. 7. Schematic of the mechanism of the down-up transition in Ci-VSD. **a**, In the down state, R223 forms a salt bridge with D129 above the hydrophobic plug. R226 is hydrated, while R229 interacts with E183/D186 and R232 coordinates to POPC headgroups. **b**, First, R226 rotates upwards along with the S4 helix which allows it to cross the hydrophobic plug and form a salt bridge with D129. **c**, Near the transition state, R226 interacts with D129, while R229 interacts mostly with D186 and E183. **d**, Beyond the transition state, the S4 helix translocates to its final up state, while R229 detaches from its coordination to E183, and R232 forms a salt bridge with D186 and E183.

states, respectively (Fig. 6b). The double-well shape of the PMF between the down and up states is consistent with a previous study of a single step ("click") transition in the Kv1.2 channel[31], although the free-energy barrier is much lower. At the same time, the profile along Q_d is much smoother than the projections on variables tracking the arginine-countercharge salt bridges (Fig. 5c and Supplementary Fig. 8).

Finally, to estimate the timescale of the down-up transition, we computed a coordinate-dependent diffusion constant D along Q_d. The approximate diffusion constant 3×10^{-4} e^2/ns (Supplementary Fig. 9c) yields a timescale on the order of 2 µs, after using the relation $\Delta Q_\mathrm{d} = \sqrt{2Dt}$ and assuming ΔQ_d between the down and up states is 1 e. As this rough estimate does not account for barriers in the PMF that would slow down transitions, the true timescale is likely to be longer.

III. CONCLUSION

In this paper, we used advances in the dynamical analysis of slow conformational transitions to elucidate the mechanism of the down-up transition of Ci-VSD. We find that S4 helix translocation and rotation are not as tightly coupled as suggested by previous simulations[20], and the S4 helix can translocate a significant amount without breaking its salt bridges with neighboring helices. As depicted in Fig. 7, the down-up transition proceeds through a series of global S4 helix movements together with local side chain rearrangements associated with salt bridges breaking and forming. Instead of a highly concerted movement, the activation of Ci-VSD is stepwise, with the overall displacement of the helix and the fine-grained rearrangements of the side chain occurring in an uncoordinated fashion. Near the transition state, rotation of the S4 helix allows R226 to detach from D186, move through the hydrophobic plug, and form a salt bridge with D129 (Fig. 7b). R229 follows and forms salt bridges with E183 and D186, ultimately favoring the latter (Fig. 7b, c). Similarly, R232 detaches from the lipid phosphate head groups and forms salt bridges with E183 and D186. The movements of R226, R229, and R232 do not follow a strict sequence, though they are also not completely independent, in part because of steric interference between their bulky side chains. To complete the transition, the S4 helix must ultimately translocate to its up state position (Fig. 7c, d). While our simulations are consistent with the important dynamics traditionally attributed to the S4 helix, they shed new light on the critical role of the salt bridges involving the arginine side chains in modulating the conformational transition.

Two sensing arginines not previously highlighted are R217 and R223. We find that R217 interacts with lipid phosphate groups in the extracellular crevice; interactions of this form have been observed in studies of other voltage sensing proteins[46]. R217's interactions with negative countercharges such as D129, D136, and D154 (which are located on the extracellular side of Ci-VSD) correlate only weakly with the committor and are heterogeneous in the TSE (Fig. 5b and Supplementary Fig. 7b, d). Our results are consistent with prior suggestions that R217 primarily participates in activation by modulating the electric field rather than through salt-bridge formation[17], although its role may be different in the transition between the down– and down states owing to its proximity to D129 and the hydrophobic plug in the down– state. Turning to R223, we find that it tends to form salt bridges with D129. Relative to R226, R223's salt-bridge distances are less predictive of the committor

and its contribution to the gating charge is small, consistent with exposure to solvent in the extracellular crevice of Ci-VSD. Prior studies using cysteine and histidine mutagenesis[17,20] have suggested that R223 may play a larger role in transitions involving the down– and up+ states of Ci-VSD.

While it is reasonable to expect that the stepwise mechanism described here to extend to mildly depolarizing and hyperpolarizing voltages, stronger applied fields could elicit a different response from the voltage sensor. It is thus unclear whether the differences between our mechanism and that observed for the Kv1.2 channel in ref. 32 reflect differences in the applied potential or differences in sequence. The authors describe the "clicks" as relatively fast, concerted events that punctuate hundreds of microseconds-long periods of stability. On the other hand, it was noted that during a single activation event in Kv1.2, the gating arginine closest to the hydrophobic plug moved across first[32]. This behavior is consistent with our simulations, in which R226 moves first. Besides the steric considerations previously mentioned, another possible explanation for this phenomenon is that the electric field reaches its maximum strength in the hydrophobic plug region. Given Ci-VSP's sequence divergence and unusual macroscopic behavior compared to VGICs[12,16,17,23,50,51] (slower gating kinetics and smaller magnitude of charge transfer during electrophysiology measurements), Ci-VSD activation may behave differently from other VSDs at the microscopic scale. Indeed, a previous study suggested that the activation of Ci-VSD is qualitatively unlike that of VGICs[23]. Compared with free-energy calculations of the Kv1.2 VSD in ref. 31, the free-energy profile along the displacement charge coordinate for a single activation step in Ci-VSD is much flatter. It would be interesting to apply the methods used in the present study to other VSDs to determine whether they have similar activation mechanisms[3,15].

Studies using fluorescently-labeled VSDs[13,51–54] have indicated that the time course of fluorescence possesses a fast component with a time constant of tens of milliseconds, tracking the charge movement (sensing current), as well as a slow component on the order of hundreds of milliseconds. However, these studies are typically performed with a full-length Ci-VSP containing the phosphatase domain and linker, which likely slow down the voltage-activation process. Although our simulations were performed using specialized, high-performance hardware that enabled us to obtain hundreds of microsecond-long trajectories, our dataset is unfortunately insufficient to obtain converged estimates of the transition rates for direct

comparison with the measured kinetics. Nevertheless, our simulations provide the first microscopic information about voltage-sensing kinetics at a depolarized membrane potential. Experimentally, it may be possible to validate the proposed mechanism by correlating movements of a fluorescent label (e.g., attached to a G214C mutation on the S3-S4 linker) to S4 helix movement and/or displacement charge changes. Computationally, it would be interesting to study the Ci-VSD down-up transition in the presence of applied fields. It may also be possible to further analyze the proposed mechanism by increasing the accuracy of the DGA estimates through inclusion of non-Markovian effects or by applying nonequilibrium sampling methods.

IV. THEORY AND METHODS

We compute statistics from MD simulations using the dynamical Galerkin approximation (DGA)[33,34]. In DGA, statistics of interest are cast as solutions to operator equations that are solved via a basis-set expansion. DGA can be viewed as a generalization of Markov state models (MSMs). The resulting estimates of the statistics depend on many variables (the input molecular features to the basis functions) and so to interpret them, we use sparse regression to discover the most important collective variables (CVs) for the transition between the down and up states.

A. Key statistics

As described in Section II (Results and Discussion), we use three statistics to understand the mechanism of the down-up transition: the potential of mean force (PMF), the committor, and the reactive current. The transition operator $\mathcal{T}_t$ encodes the stochastic dynamics underlying these statistics and is defined as

$$\mathcal{T}_t f(x) = \mathbf{E}[f(X(t)) \mid X(0) = x], \tag{1}$$

where f is an arbitrary function of the coordinates $X(t)$, and $\mathbf{E}$ denotes an expectation. The transition operator evolves the expectation of f over a lag time t, conditioned on starting at x.

The PMF is the reversible work required to constrain variables to particular values, i.e., the

free energy as a function of those variables. We can obtain the free energy from the negative logarithm of the equilibrium distribution $\pi(x)$ (which is proportional to the Boltzmann weight in the canonical ensemble). In turn, we obtain $\pi(x)$ by appealing to the fact that the equilibrium distribution must not change upon application of the transition operator. Specifically, we introduce a weight factor $w(x)$ that relates the sampled distribution $\mu(x)$ (which need not be drawn from π) to $\pi(x)$:

$$\begin{aligned}\int f(x)\pi(x)\,\mathrm{d}x &= \int f(x)w(x)\mu(x)\,\mathrm{d}x \\ &= \int \mathcal{T}_t f(x)w(x)\mu(x)\,\mathrm{d}x \\ &= \int f(x)(\mathcal{T}_t)^\dagger_\mu w(x)\mu(x)\,\mathrm{d}x,\end{aligned} \tag{2}$$

where $(\mathcal{T}_t)^\dagger_\mu$ is the μ-weighted adjoint of $\mathcal{T}_t$. The weight factor thus satisfies $(\mathcal{T}_t)^\dagger_\mu w(x) = w(x)$[34,55]. We solve this equation using DGA as described below.

The committor $q_+(x)$ is defined as the probability of reaching the product state B (here the up state) before returning to the reactant state A (here the down state), conditioned on starting at x. In this sense, it is a perfect reaction coordinate (measure of progress) for a stochastic reaction[35]. Because $q_+(x) = 0$ in A and $q_+(x) = 1$ in B, it is governed by a modified transition operator

$$\mathcal{S}_t f(x) = \mathbf{E}[f(X(t \wedge T_{A\cup B})) \mid X(0) = x], \tag{3}$$

where $T_{A\cup B} = \min\{t \geq 0 \mid t \in A \cup B\}$ is the first time X enters $A \cup B$ and $t \wedge T_{A\cup B} \equiv \min\{t, T_{A\cup B}\}$. The transition operator $\mathcal{S}_t$ corresponds to terminating the dynamics when A or B is entered. The committor satisfies[34,56]

$$q_+(x) = \begin{cases} \mathcal{S}_t q_+(x) & x \in (A \cup B)^c \\ 0 & x \in A \\ 1 & x \in B. \end{cases} \tag{4}$$

Finally, the reactive current measures the flux of reactive probability density. The reactive current is most easily interpreted after projecting onto a subspace of CVs, $z(x)$, yielding a vector field $J^z_{AB}(s)$ that summarizes how reactive trajectories flow through $z(x) = s$. It was shown previously that the projected reactive probability current satisfies[34,57]

$$J^z_{AB}(s) = \int J_{AB} \cdot \nabla z(x) \delta(z(x) - s) \pi(x) \, \mathrm{d}x, \tag{5}$$

where J_{AB} is the reactive flux in the space of original coordinates. For example, the "productive" reactive flux from state A to B is obtained by choosing $z(x) = q_+(x)$[38]. As in previous work, we smooth all reactive currents with a Gaussian kernel density estimate of width 1 bin.

B. Dynamical Galerkin approximation

In DGA, we write the statistic of interest, such as w or q_+, as a linear combination of basis functions $\phi_i(x)$ with coefficients v_i:

$$f(x) = \gamma(x) + \sum_{i=1}^{N} v_i \phi_i(x). \tag{6}$$

Here, $\gamma(x)$ is a "guess" function chosen to satisfy the boundary conditions, so $\phi_i(x) = 0$ for $x \in A \cup B$. Note that solving for the weight factor w, unlike q_+, does not require any explicit boundary conditions. Substituting the ansatz (6) into the appropriate operator equation [e.g., (4) for q_+], multiplying by another basis function ϕ_j, and integrating over the sampled distribution $\mu(x)$, we obtain the linear system

$$(C_t - C_0)v = -(r_t - r_0), \tag{7}$$

where

$$C_t^{ij} = \int \phi_i(x) \mathcal{S}_t \phi_j(x) \mu(x) \, \mathrm{d}x \tag{8a}$$

$$r_t^j = \int \phi_j(x) \mathcal{S}_t \gamma(x) \mu(x) \, \mathrm{d}x. \tag{8b}$$

Importantly, (4) holds for any lag time t, so we can estimate the integrals in (8) from averages over an ensemble of simulations that are much shorter than the transition time (i.e., $t \ll T_{A \cup B}$). Indexing simulations by $m = 1, 2, ..., M$,

$$C_t^{ij} \approx \frac{1}{M} \sum_{m=1}^{M} \phi_i(X^m(0)) \phi_j(X^m(t \wedge T_{A \cup B})) \tag{9a}$$

$$r_t^j \approx \frac{1}{M} \sum_{m=1}^{M} \phi_j(X^m(0)) \gamma(X^m(t \wedge T_{A \cup B})). \tag{9b}$$

Given these quantities, (7) can be solved for v, which can be used to reconstruct f through (6).

C. Simulation details

The starting point for our simulations was the crystal structure of the up state (PDB ID 4G7V)[16]. We added hydrogens to this structure and embedded it into a palmitoyl oleyl phosphatidylcholine (POPC) lipid bilayer approximately 78 Å × 78 Å in area, with 86 and 88 lipids in the outer and inner leaflets, respectively. Titratable residues were assigned their default protonation states at pH 7. The orientation and relative position of the protein inside the membrane was adjusted according to the prediction from the orientations of proteins in membranes (OPM) database[58]. We then used the program VMD[59] to solvate the system in 0.1 M NaCl aqueous solution so that it was electrically neutral. There were 20 Na^+ and 19 Cl^- ions, for a total of 56,582 atoms.

In simulations, we used the CHARMM36m force field[60] for the protein, phospholipids, and ions, and the TIP3P model for water molecules[61]. Initial conditions were drawn uniformly from replica exchange umbrella sampling windows in the space of S4 helix CVs[20]. The coordinates, corresponding velocities and periodic boundary parameters were used to initialize Anton 2 simulations[62]. We used the hydrogen mass repartitioning scheme[63] and a time step of 4 fs. All simulations were carried out in an isothermal-isobaric ensemble (300 K, 1 atm) with periodic boundary conditions and zero applied voltage. The temperature and pressure were controlled using the Nosé–Hoover thermostat and the semi-isotropic MTK barostat, respectively[64–66]. Long-range electrostatic interactions were calculated using the u-series method[67], and van der Waals interactions were truncated at a cutoff distance of 9 Å.

We initially performed 230 simulations of 1 µs and 7 simulations of 10 µs for a total of 300 µs. Because these data were insufficient to converge the committor, we further sampled adaptively. Specifically, we manually selected structures with estimated committor values near $q_+ = 0.5$ and with under-sampled values of S4-helix and salt-bridge CVs. In the initial dataset, we observed that many simulations did not diverge significantly from their initial structures on the timescale of hundreds of nanoseconds, suggesting that there were degrees of freedom preventing fast equilibration of the desired CVs. Steering the system toward the the down or up state and back prior to running unbiased dynamics seemed to relax these degrees of freedom and ameliorate this issue slightly. These steered molecular dynamics simulations[68,69] were based on a mild harmonic restraint on selected salt-bridge distances

and on the root-mean-square deviation (RMSD) to either the down or up state. From the final structures we initialized additional 1 µs simulations (Fig. 2). In total, we accumulated 415 µs of trajectories, with coordinates saved every 0.1 ns.

D. Collective variables

We computed the intramolecular distances used as inputs to DGA using PyEMMA[70] and MDTraj[71]. Hydrogen bonds were computed with a donor-acceptor maximum heavy-atom distance of 3.5 Å and a minimum angle cutoff of 120° using the HydrogenBondAnalysis module in MDAnalysis[72,73]. For salt-bridge distances between sensing arginines and acidic residues, we compute the distance between the C_ζ of arginine (the terminal carbon in the guanidinium group) and the C_γ/C_δ of aspartate/glutamate (the carbon in the carboxylate group). Between arginines and countercharges, we take the N_ϵ and both N_η atoms of the guanidinium group as donors and O_δ/O_ε atoms of the carboxylate group in aspartate/glutamate as acceptors. For lipids, we choose the oxygens of the phosphatidylcholine headgroup in POPC as hydrogen bond acceptors.

The S4 helix translocation (d) and rotation (θ) were computed with respect to the up state after aligning the C_α atoms of helices S1–S3 to their positions in the up state crystal structure so as to minimize the RMSD. The translocation was computed using the distanceZ function in the Colvars module of NAMD[74]. This function computes the distance along a projection of one vector onto another. In our case, the first vector was one connecting the centers of mass of C_α atoms of residues 217–233 in the structure of interest and the up state crystal structure, and the second vector was the principal inertial axis of these atoms in the up state crystal structure. The rotation was computed using the function spinAngle, measuring the angle of the rotation around the local helical axis of the C_α atoms of residues 217–233, with respect to to the positions of these atoms in the up state crystal structure.

We computed the displacement charge, Q_d, according to the "Q-route"[48]:

$$Q_\mathrm{d} = \sum_i q_i \left(\frac{z_i^{(u)} + L_z/2}{L_z} \right), \tag{10}$$

where q_i are charges, $z_i^{(u)}$ are unwrapped z-coordinates (i.e., without applying periodic boundary conditions), and L_z is the length of the simulation box along the z-axis (with

the membrane parallel to the xy-plane).

E. State definitions

To perform committor calculations, we need to define which structures are in A and B, or the down and up states, respectively. We defined the down state as

$$\left(\frac{d + 4.24\ \text{Å}}{1.1\ \text{Å}}\right)^2 + \left(\frac{\theta + 56.95^\circ}{8^\circ}\right)^2 < 1$$

and the up state as

$$\left(\frac{d + 0.506\ \text{Å}}{0.84\ \text{Å}}\right)^2 + \left(\frac{\theta - 3.94^\circ}{7.6^\circ}\right)^2 < 1.$$

The centers and radii of the ellipses are obtained from several nanoseconds-long MD simulations initialized from the down and up states, rather than the crystal structures, since we want to include typical fluctuations. Visual inspection of the structures and trajectories showed that S4 helix movements often precede or lag behind the movements of arginine side chains, so we chose to include cutoffs on these distances as well. These are listed in Supplementary Table 1. To test for the possibility that our results depended on the choice of cutoffs, we altered the cutoffs and found that the qualitative features of the committor were preserved.

F. Basis set construction

For the committor calculations, we used a basis set built from pairwise intramolecular distances[34]. The distances are based on two groups of residues: (1) R217, R223, R226, R229, and R232 and (2) D129, D136, D151, E183, and D186. We include all intergroup distances between the C_α atoms, as well as all distances between the arginine C_ζ atoms and the aspartate/glutamate C_δ/C_ε atoms of these residues. This yields $(5 \times 6) + (5 \times 6) = 60$ distances. We also tested including other pairwise distances, such as those involving residues in the hydrophobic plug, but these did not change the committors significantly. Briefly, we constructed the basis functions by first computing the Euclidean distance for each configuration x to the boundary of state A or state B, yielding d_A and d_B (the dependence on x is left implicit). Then, we define a smoothing function $h(x) = d_A d_B/(d_A + d_B)^2$, which obeys the boundary conditions by construction, and construct $\phi_i(x) = x_i h(x)$ where x_i is the ith

component of the 60-dimensional coordinate vector. Finally, we whiten the basis functions with SVD so they are orthogonal. The guess function is $\gamma(x) = d_A^2/(d_A^2 + d_B^2)$, which also satisfies the boundary conditions by construction.

For change of measure calculations, a basis of indicator functions was used, since the distance basis tended to yield noisier results. Instead of the 60 salt-bridge distances, we featurized our data using the minimum heavy-atom distance between every residue in the S4 helix and every residue in the S1–S3 helices (1924 distances in total). These features were reduced in dimensionality to 10 dimensions using the integrated variational approach to conformational dynamics (IVAC)[41] and then used to cluster the MD structures with mini-batch k-means ($k = 200$), as implemented in scikit-learn[75]. Each cluster was converted into an indicator basis function.

G. Sparse regression

To identify a small number of variables that are important for the dynamics, we modeled the committor using LASSO[39,76] (least absolute shrinkage and selection operator) as implemented in scikit-learn[75]. LASSO is a form of linear regression in which the loss is the sum of a least-squares term and a penalty term that encourages sparsity of the model coefficients β_i:

$$\mathcal{L}_{\text{LASSO}} = \sum_k (\beta^\top x^{(k)} - q_+(x^{(k)}))^2 + \lambda \|\beta\|_1, \tag{11}$$

where $\|\zeta\|_1 = \sum_i |\zeta_i|$ is the ℓ_1 norm and λ is a parameter controlling the penalty strength. In (11), $x^{(k)}$ is a vector of features for the kth data point and $q_+(x^{(k)})$ is the corresponding committor value. Prior to fitting, the data are standardized by removing the mean and normalizing to unit variance.

To overweight the transition region, where we expect the committor to be most informative, we sampled 100,000 points weighted by $\pi q_+(1 - q_+)$, which has a maximum at $q_+ = 0.5$. As discussed in Section II (Results and Discussion), our initial analysis suggested that the committor depends non-monotonically on some variables. To capture these relationships with LASSO, we split our dataset into points with $q_+ \leq 0.5$ and $q_+ \geq 0.5$ and then use an inverse sigmoid transform to map the DGA-estimated committors to the entire range $(-\infty, \infty)$ so that the LASSO-predicted committors are constrained to the appropriate range

[either $(0, 0.5)$ or $(0.5, 1.0)$]. For the former, we chose the transformation $\ln[2q_+/(1-2q_+)]$, and, for the latter, we instead used the transformation $\ln[(2q_+ - 1)/(2 - 2q_+)]$. Although DGA can yield q_+ values outside the range $[0, 1]$, we clipped them to this range before applying the transformation.

As λ varies, there is a tradeoff between sparsity and accuracy. Here, we chose $\lambda = 0.02$ for points $q_+ \leq 0.5$ and $\lambda = 0.03$ for points $q_+ \geq 0.5$ because they appeared to yield reasonably sparse representations while maintaining relatively good fits of the committor ($R^2 \approx 0.7$) given the linear form (Supplementary Fig. 10).

DATA AVAILABILITY

Raw simulation data generated on Anton 2, as well as derived data (featurized MD trajectories and calculations) are available upon request from the authors. Code to reproduce the analysis and plots in this work can be found at https://github.com/spencercguo/ci-vsd.

REFERENCES

[1]Aggarwal, S. K. & MacKinnon, R. Contribution of the S4 Segment to Gating Charge in the Shaker K+ Channel. *Neuron* **16**, 1169–1177 (1996).

[2]Sigg, D. & Bezanilla, F. Total Charge Movement per Channel : The Relation between Gating Charge Displacement and the Voltage Sensitivity of Activation. *J. Gen. Physiol.* **109**, 27–39 (1997).

[3]Bezanilla, F. How membrane proteins sense voltage. *Nat. Rev. Mol. Cell Biol.* **9**, 323–332 (2008).

[4]Long, S. B., Campbell, E. B. & MacKinnon, R. Crystal Structure of a Mammalian Voltage-Dependent Shaker Family K+ Channel. *Science* **309**, 897–903 (2005).

[5]Wu, J. *et al.* Structure of the voltage-gated calcium channel Cav1.1 complex. *Science* **350** (2015).

[6]Jiang, Y. *et al.* X-ray structure of a voltage-dependent K+ channel. *Nature* **423**, 33–41 (2003).

[7]Henrion, U. *et al.* Tracking a complete voltage-sensor cycle with metal-ion bridges. *Proc. Natl. Acad. Sci.* **109**, 8552–8557 (2012).

[8]Catterall, W. A. Ion Channel Voltage Sensors: Structure, Function, and Pathophysiology. *Neuron* **67**, 915–928 (2010).

[9]Wisedchaisri, G. *et al.* Resting-State Structure and Gating Mechanism of a Voltage-Gated Sodium Channel. *Cell* **178**, 993–1003.e12 (2019).

[10]Ye, F. *et al.* Voltage-gating and cytosolic Ca2+ activation mechanisms of Arabidopsis two-pore channel AtTPC1. *Proc. Natl. Acad. Sci.* **118** (2021).

[11]Gao, S., Yao, X. & Yan, N. Structure of human Cav2.2 channel blocked by the painkiller ziconotide. *Nature* 1–5 (2021).

[12]Murata, Y., Iwasaki, H., Sasaki, M., Inaba, K. & Okamura, Y. Phosphoinositide phosphatase activity coupled to an intrinsic voltage sensor. *Nature* **435**, 1239–1243 (2005).

[13]Kohout, S. C., Ulbrich, M. H., Bell, S. C. & Isacoff, E. Y. Subunit organization and functional transitions in Ci-VSP. *Nat. Struct. Mol. Biol.* **15**, 106–108 (2008).

[14]Kohout, S. C. *et al.* Electrochemical coupling in the voltage-dependent phosphatase Ci-VSP. *Nat. Chem. Biol.* **6**, 369–375 (2010).

[15]Okamura, Y., Kawanabe, A. & Kawai, T. Voltage-Sensing Phosphatases: Biophysics, Physiology, and Molecular Engineering. *Physiol. Rev.* **98**, 2097–2131 (2018).

[16]Li, Q. *et al.* Structural mechanism of voltage-dependent gating in an isolated voltage-sensing domain. *Nat. Struct. Mol. Biol.* **21**, 244–252 (2014).

[17]Villalba-Galea, C. A., Frezza, L., Sandtner, W. & Bezanilla, F. Sensing charges of the Ciona intestinalis voltage-sensing phosphatase. *J. Gen. Physiol.* **142**, 543–555 (2013).

[18]DeCaen, P. G., Yarov-Yarovoy, V., Zhao, Y., Scheuer, T. & Catterall, W. A. Disulfide locking a sodium channel voltage sensor reveals ion pair formation during activation. *Proc. Natl. Acad. Sci.* **105**, 15142–15147 (2008).

[19]DeCaen, P. G., Yarov-Yarovoy, V., Sharp, E. M., Scheuer, T. & Catterall, W. A. Sequential formation of ion pairs during activation of a sodium channel voltage sensor. *Proc. Natl. Acad. Sci.* **106**, 22498–22503 (2009).

[20]Shen, R., Meng, Y., Roux, B. & Perozo, E. Mechanism of voltage gating in the voltage-sensing phosphatase Ci-VSP. *Proc. Natl. Acad. Sci.* **119**, e2206649119 (2022).

[21]Ahern, C. A. & Horn, R. Focused Electric Field across the Voltage Sensor of Potassium Channels. *Neuron* **48**, 25–29 (2005).

[22]Tao, X., Lee, A., Limapichat, W., Dougherty, D. A. & MacKinnon, R. A Gating Charge Transfer Center in Voltage Sensors. *Science* **328**, 67–73 (2010).

[23]Lacroix, J. J. & Bezanilla, F. Tuning the Voltage-Sensor Motion with a Single Residue. *Biophys. J.* **103**, L23–L25 (2012).

[24]Gumbart, J., Khalili-Araghi, F., Sotomayor, M. & Roux, B. Constant electric field simulations of the membrane potential illustrated with simple systems. *Biochim. Biophys. Acta* **1818**, 294–302 (2012).

[25]Jogini, V. & Roux, B. Dynamics of the Kv1.2 Voltage-Gated K+ Channel in a Membrane Environment. *Biophys. J.* **93**, 3070–3082 (2007).

[26]Nishizawa, M. & Nishizawa, K. Molecular Dynamics Simulation of Kv Channel Voltage Sensor Helix in a Lipid Membrane with Applied Electric Field. *Biophys. J.* **95**, 1729–1744 (2008).

[27]Bjelkmar, P., Niemelä, P. S., Vattulainen, I. & Lindahl, E. Conformational Changes and Slow Dynamics through Microsecond Polarized Atomistic Molecular Simulation of an Integral Kv1.2 Ion Channel. *PLoS Comput. Biol.* **5**, e1000289 (2009).

[28]Treptow, W., Tarek, M. & Klein, M. L. Initial Response of the Potassium Channel Voltage Sensor to a Transmembrane Potential. *Journal of the American Chemical Society* **131**, 2107–2109 (2009).

[29]Khalili-Araghi, F. *et al.* Calculation of the Gating Charge for the Kv1.2 Voltage-Activated Potassium Channel. *Biophys. J.* **98**, 2189–2198 (2010).

[30]Delemotte, L., Tarek, M., Klein, M. L., Amaral, C. & Treptow, W. Intermediate states of the Kv1.2 voltage sensor from atomistic molecular dynamics simulations. *Proc. Natl. Acad. Sci.* **108**, 6109-6114 (2011).

[31]Delemotte, L., Kasimova, M. A., Klein, M. L., Tarek, M. & Carnevale, V. Free-energy landscape of ion-channel voltage-sensor–domain activation. *Proc. Natl. Acad. Sci.* **112**, 124–129 (2015).

[32]Jensen, M. *et al.* Mechanism of Voltage Gating in Potassium Channels. *Science* **336**, 229–233 (2012).

[33]Thiede, E. H., Giannakis, D., Dinner, A. R. & Weare, J. Galerkin approximation of dynamical quantities using trajectory data. *J. Chem. Phys.* **150**, 244111 (2019).

[34]Strahan, J. *et al.* Long-time-scale predictions from short-trajectory data: A benchmark analysis of the trp-cage miniprotein. *J. Chem. Theory Comput.* **17**, 2948–2963 (2021).

[35]Vanden-Eijnden, E. Transition path theory. In *Computer Simulations in Condensed Matter Systems: From Materials to Chemical Biology Volume 1*, 453–493 (Springer, 2006).

[36]Pathak, M. M. *et al.* Closing In on the Resting State of the Shaker K+ Channel. *Neuron* **56**, 124–140 (2007).

[37]Vargas, E. *et al.* An emerging consensus on voltage-dependent gating from computational modeling and molecular dynamics simulations. *J. Gen. Physiol.* **140**, 587–594 (2012).

[38]Roux, B. Transition rate theory, spectral analysis, and reactive paths. *J. Chem. Phys.* **156**, 134111 (2022).

[39]Finkel, J., Webber, R. J., Abbot, D. S., Gerber, E. P. & Weare, J. Learning forecasts of rare stratospheric transitions from short simulations. *Mon. Weather Rev.* **149**, 3647–3669 (2021).

[40]Ma, A. & Dinner, A. R. Automatic method for identifying reaction coordinates in complex systems. *J. Phys. Chem. B* **109**, 6769–6779 (2005).

[41]Lorpaiboon, C., Thiede, E. H., Webber, R. J., Weare, J. & Dinner, A. R. Integrated Variational Approach to Conformational Dynamics: A Robust Strategy for Identifying Eigenfunctions of Dynamical Operators. *J. Phys. Chem. B* **124**, 9354–9364 (2020).

[42]Gallivan, J. P. & Dougherty, D. A. Cation-π interactions in structural biology. *Proc. Natl. Acad. Sci.* **96**, 9459-9464 (1999).

[43]Armstrong, C. T., Mason, P. E., Anderson, J. L. R. & Dempsey, C. E. Arginine side chain interactions and the role of arginine as a gating charge carrier in voltage sensitive ion channels. *Sci. Rep.* **6**, 21759 (2016).

[44]Infield, D. T. *et al.* Cation-π interactions and their functional roles in membrane proteins. *J. Mol. Biol.* **433**, 167035 (2021).

[45]Freites, J. A., Tobias, D. J., Heijne, G. v. & White, S. H. Interface connections of a transmembrane voltage sensor. *Proc. Natl. Acad. Sci.* **102**, 15059–15064 (2005).

[46]Schmidt, D., Jiang, Q.-X. & MacKinnon, R. Phospholipids and the origin of cationic gating charges in voltage sensors. *Nature* **444**, 775–779 (2006).

[47]Li, S., Su, Y., Luo, W. & Hong, M. Water-Protein Interactions of an Arginine-Rich Membrane Peptide in Lipid Bilayers Investigated by Solid-State Nuclear Magnetic Resonance Spectroscopy. *J. Phys. Chem. B* **114**, 4063–4069 (2010).

[48]Roux, B. The membrane potential and its representation by a constant electric field in computer simulations. *Biophys. J.* **95**, 4205–4216 (2008).

[49]Bezanilla, F. Gating currents. *J. Gen. Physiol.* **150**, 911–932 (2018).

[50]Murata, Y. & Okamura, Y. Depolarization activates the phosphoinositide phosphatase Ci-VSP, as detected in Xenopus oocytes coexpressing sensors of PIP2. *J. Physiol.* **583**, 875–889 (2007).

[51]Villalba-Galea, C. A., Sandtner, W., Starace, D. M. & Bezanilla, F. S4-based voltage sensors have three major conformations. *Proc. Natl. Acad. Sci.* **105**, 17600–17607 (2008).

[52]Lundby, A., Mutoh, H., Dimitrov, D., Akemann, W. & Knöpfel, T. Engineering of a Genetically Encodable Fluorescent Voltage Sensor Exploiting Fast Ci-VSP Voltage-Sensing Movements. *PLoS ONE* **3**, e2514 (2008).

[53]Villalba-Galea, C. A. *et al.* Charge Movement of a Voltage-Sensitive Fluorescent Protein. *Biophys. J.* **96**, L19–L21 (2009).

[54]Villalba-Galea, C. Voltage-Controlled Enzymes: The New JanusBifrons. *Front. Pharmacol.* **3** (2012).

[55]Wu, H. *et al.* Variational koopman models: Slow collective variables and molecular kinetics from short off-equilibrium simulations. *J. Chem. Phys.* **146**, 154104 (2017).

[56]Gardiner, C. *Stochastic Methods: A Handbook for the Natural and Social Sciences*. Springer Series in Synergetics (Springer-Verlag, Berlin Heidelberg, 2009), 4 edn.

[57]Vani, B. P., Weare, J. & Dinner, A. R. Computing transition path theory quantities with trajectory stratification. *J. Chem. Phys.* **157**, 034106 (2022).

[58]Lomize, M. A., Pogozheva, I. D., Joo, H., Mosberg, H. I. & Lomize, A. L. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* **40**, D370–376 (2012).

[59]Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J. Mol. Graphics* **14**, 33–38, 27–28 (1996).

[60]Huang, J. *et al.* CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nature methods* **14**, 71–73 (2017).

[61]Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).

[62]Shaw, D. E. *et al.* Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer. In *SC14: International Conference for High Performance Computing, Networking, Storage and Analysis*, 41–53 (IEEE, New Orleans, LA, USA, 2014).

[63]Hopkins, C. W., Le Grand, S., Walker, R. C. & Roitberg, A. E. Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **11**, 1864–1874 (2015).

[64]Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101**, 4177–4189 (1994).

[65]Martyna, G. J., Klein, M. L. & Tuckerman, M. Nosé–Hoover chains: The canonical ensemble via continuous dynamics. *J. Chem. Phys.* **97**, 2635–2643 (1992).

[66]Lippert, R. A. *et al.* Accurate and efficient integration for molecular dynamics simulations at constant temperature and pressure. *J. Chem. Phys.* **139**, 164106 (2013).

[67]Predescu, C. *et al.* The u-series: A separable decomposition for electrostatics computation with improved accuracy. *J. Chem. Phys.* **152**, 084113 (2020).

[68]Isralewitz, B., Gao, M. & Schulten, K. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.* **11**, 224–230 (2001).

[69]Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C. & Bussi, G. PLUMED 2: New feathers for an old bird. *Comp. Phys. Comm.* **185**, 604–613 (2014).

[70]Scherer, M. K. *et al.* PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **11**, 5525–5542 (2015).

[71]McGibbon, R. T. *et al.* MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys. J.* **109**, 1528–1532 (2015).

[72]Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319–2327 (2011).

[73]Smith, P., Ziolek, R. M., Gazzarrini, E., Owen, D. M. & Lorenz, C. D. On the interaction of hyaluronic acid with synovial fluid lipid membranes. *Phys. Chem. Chem. Phys.* **21**, 9845–9857 (2019).

[74]Fiorin, G., Klein, M. L. & Hénin, J. Using collective variables to drive molecular dynamics simulations. *Mol. Phys.* **111**, 3345–3362 (2013).

[75]Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

[76]Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. Series B Stat. Methodol.* **58**, 267–288 (1996).

ACKNOWLEDGEMENTS

The authors thank Eduardo Perozo, Adam Antoszewski, John Strahan, Chatipat Lorpaiboon, and Jonathan Weare for discussions and suggestions on the paper. S.C.G. acknowledges support by the National Science Foundation Graduate Research Fellowship under Grant No. 2140001. A.R.D. acknowledges National Institutes of Health (NIH) award R35 GM136381. B.R. acknowledges NIH/NIGMS award R01 GM062342. Anton 2 computer time was provided by the Pittsburgh Supercomputing Center (PSC) through Grant R01 GM116961 from NIH. The Anton 2 machine at PSC was generously made available by D. E. Shaw Research. This work was completed in part with resources provided by the University of Chicago Research Computing Center and we are grateful for their assistance with the calculations.

AUTHOR CONTRIBUTIONS

S.C.G., B.R., and A.R.D. conceptualized the study. S.C.G. and R.S. performed the MD simulations. S.C.G. performed the dynamical analysis and regression. S.C.G., B.R., and A.R.D. interpreted the results and wrote the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

MATERIALS AND CORRESPONDENCE

Correspondence and material requests should be addressed to Aaron Dinner (dinner@uchicago.edu) or Benoît Roux (roux@uchicago.edu).