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

Spencer C. Guo,[1,2] Rong Shen,[3] Benoît Roux,[1,3,4,*)] and Aaron R. Dinner[1,2,4,†)]

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

[2)] *James Franck Institute, The University of Chicago, Chicago, Illinois 60637*

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

[4)] *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-gated ion channels responsible for cellular excitability. Two crystal structures of Ci-VSD in putative resting and active conformations suggest a helical-screw voltage sensing mechanism in which the S4 helix translocates and rotates to enable exchange of salt-bridge partners. By combining extensive molecular dynamics simulations with a computational framework based on dynamical operators, we elucidate the microscopic mechanism of the resting-active transition at physiological membrane potential. Sparse regression reveals a small set of coordinates that distinguish intermediates hidden from electrophysiological measurements. The intermediates arise from a noncanonical helical-screw mechanism in which translocation, rotation, and side-chain movement of the S4 helix are only loosely coupled. These results provide new insights into existing experimental and computational findings on voltage sensing and suggest ways of further probing its mechanism.

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[7–15].

Like VGICs, the voltage-sensing phosphatase from *Ciona intestinalis* (Ci-VSP) exhibits sensing currents in response to changes in transmembrane potential, as well as voltage-regulated enzymatic activity[16–20]. 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[21]. 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[4,6,20], 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[3,16,20–22]; 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[7,8,23–25]. Finally, there is a group of three highly-conserved, bulky hydrophobic residues, I126, F161, and 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[26–29]. 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 translocation of about 5 Å to a rotation of 60° around the helix axis[21,25]. 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)[17,25,30].

Although there have been previous MD studies of VSDs[25,31–40], direct simulation of the transition between resting and active states at physiological voltages remains computationally intractable. In ref. 39, the authors were able to recapitulate the gating cycle of the Kv1.2/2.1 chimera by applying hyperpolarizing or depolarizing potentials of magnitudes over 300 mV. In these simulations the S4 helices translocated 15 Å and rotated about 120° while the sensing arginines moved past a cen-

*)Electronic mail: roux@uchicago.edu
†)Electronic mail: dinner@uchicago.edu

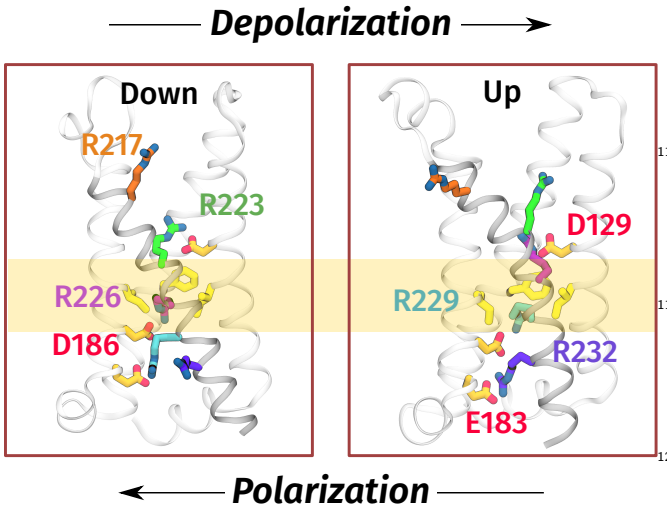

FIG. 1. Structure of Ci-VSD. Left, down (PDB: 4G80) and right, up (PDB: 4G7V) states. The side chains of sensing arginines on the S4 helix (gray) which participate in voltage-sensing, R217 (orange), R223 (lime), R226 (magenta), R229 (cyan), and R232 (purple), as well as of residues that form salt bridges 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.

tral hydrophobic residue in discrete steps. However, such strongly applied voltages can impact the mechanism, and the statistical significance of these observations remains unclear from the few transitions observed during milliseconds of simulation time. Enhanced sampling methods have been used to compute free energy landscapes for Ci-VSD[25] and the VSD of the Kv1.2 channel[40], but such calculations do not directly provide information about kinetics. Furthermore, they assume *a priori* that certain coordinates are important for describing the mechanism.

Here, we dissect the down-up transition of Ci-VSD at physiological voltages by analyzing 415 μs of molecular dynamics simulation. By applying recently introduced computational methods[41,42] within the context of transition path theory[43], we elucidate the microscopic events underlying activation. Despite capturing the apparent two-state behavior of the displacement charge, the analysis reveals multiple intermediates and deviations from a canonical helical-screw mechanism. These findings provide new insights into the origins of the complex kinetics of activation measured for wild-type and mutant Ci-VSPs, which we set in the context of results for other voltage-sensing proteins. We conclude by suggesting experiments that can test the computational predictions.

II. RESULTS

Our goal is to compute kinetic statistics for the transition between the down and up states and then to relate these statistics to specific microscopic events. We rely on a framework based on transition path theory[43] that assumes the dynamics are Markovian after a relatively short time. The method, which is called the dynamical Galerkin approximation (DGA), can be viewed as a generalization of Markov state models (MSMs)[41,42]. The essential idea is that long-time statistics can be estimated by combining information from a dataset of short unbiased molecular dynamics trajectories that each samples a portion of the VSD activation event. In other words, we can learn the mechanism of the transition even when no single trajectory connects the down and up states, so long as the transition region between them is well sampled (as discussed further in Section IV). This dramatically reduces the computational cost because we do not need to wait for fluctuations that give rise to such a trajectory, let alone the many that would be needed to achieve the same statistical confidence. We exploit the specialized Anton 2 supercomputer[44] to generate a dataset of 415 μs of aggregate simulation at zero (depolarized) membrane potential. The dataset consists of short trajectories initiated from starting points distributed between the down– and up+ states (homology models that shift the sensing arginines one click past the down and up states[21,25]).

Here we describe three statistics that are particularly important for molecular studies and how we use them to learn mechanisms. We define them mathematically and discuss how we compute them in Section IV.

- *Equilibrium probabilities of states* define relative free energies, which can be used to compute equilibrium averages and potentials of mean force (PMFs) along selected coordinates.
- The *committor* is the probability of completing the transition (i.e., proceeding to the up state before returning to the down state).
- The *reactive current* shows the fluxes between states. We use it to visualize the transition pathways and quantify their relative contributions.

These statistics provide key insights into complex conformational transitions. For example, the PMF allows us to visualize metastable states and the barriers separating them. The committor, because it varies monotonically between zero at the reactant state and one at the product state, represents a natural coordinate for tracking the progress of a conformational transition and can be used to identify the molecular features that correlate with that progress. Finally, the reactive current yields information about the dynamical propagation of molecular features during a conformational transition.

We first connect our simulations to experimental measurements via the displacement charge. Next, we show

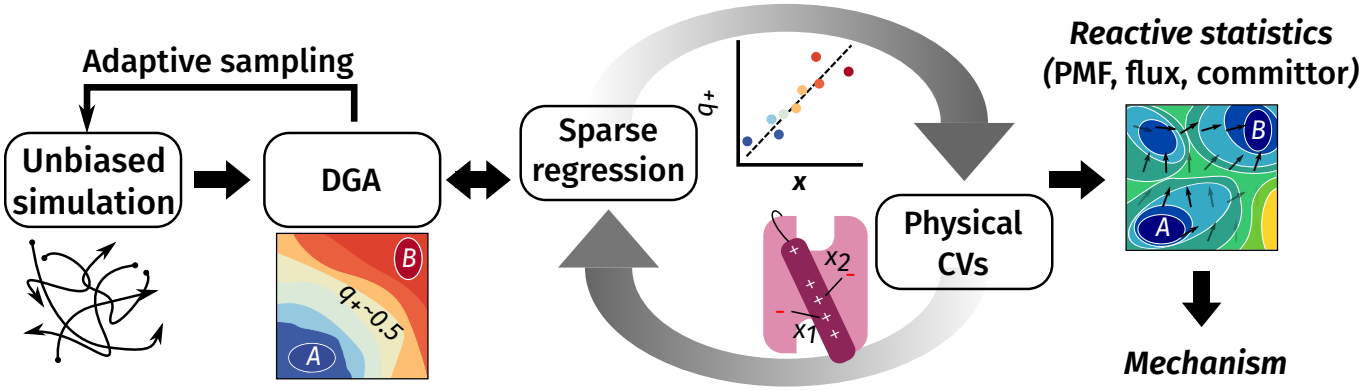

FIG. 2. Computational workflow for this study.

how the committor provides additional microscopic insight into the kinetics and use a sparse regression procedure to identify a small set of molecular collective variables (CVs) useful for modeling activation. Projections of the reactive currents and free energies onto these CVs enable us to characterize the transition pathways and thermodynamic forces stabilizing them. The overall workflow is summarized in Fig. 2.

A. The committor reveals intermediates hidden within a single "click" of the displacement charge

The displacement charge Q_d represents the energetic coupling of the membrane protein system with an external potential[45]. The change in displacement charge upon a conformational transition of the voltage-sensor, ΔQ_d, also known as the sensing charge (or, in VGICs, the gating charge), can be measured with voltage-clamp electrophysiology[3,46]. Ci-VSD was previously found[16,21,25] to transfer approximately 1.1 e in going from the down state to the up state. Our calculations yield a sensing charge of $\Delta Q_\mathrm{d} = 0.9 \pm 0.3\ e$, in agreement with previous estimates.

Using DGA, we computed the average Q_d as a function of the S4 helix translocation and rotation, and it exhibits discrete values (Supplementary Fig. 1a), consistent with the idea that the S4 helix moves by discrete "clicks" (these results can be compared directly with Fig. 1B of Ref. 25). In further support of this idea, the PMF as a function of Q_d (Fig. 3a, black line) exhibits minima at Q_d values corresponding to the down and up states; these minima are nearly identical in free energy and are separated by a barrier of less than 0.5 kcal/mol, suggesting both states may be populated in the absence of an applied potential. Although the simulations are performed at 0 mV, we can mimic the effect of an applied potential, V, by adding a linear term $Q_\mathrm{d}V$ to the free energy profile[45]. Relatively mild hyperpolarizing or depolarizing voltages on the order of $\mp$50 mV are sufficient to tilt the PMF in favor of the down (and down–) or up (and up+) states, respectively (Fig. 3a).

To connect with measured kinetics[17,18,47–50], we computed an equilibrium time-correlation function for Q_d (Fig. 3b). It can be well fit by a biexponential function with slow and fast time constants of approximately 59 μs and 0.04 μs, respectively. The slow time constant is consistent with our estimates of the mean first passage times between down and up states ($\sim$100 μs, Supplementary Fig. 2). These values are about an order of magnitude faster than the fast component of fluorescence relaxation of dye-labeled VSDs, which tracks the charge movement (sensing current)[17,47–50]. However, the experimental studies were mainly performed with full-length Ci-VSP containing the phosphatase domain and linker, which slows down the voltage-activation process[47]. With this in mind, our results suggest both the equilibrium averages and the kinetics are consistent with available data and previous simulations for Ci-VSD, so we now turn to dissecting the microscopic dynamics during activation.

As a first step toward this end, we computed the committor (q_+) for all structures in our dataset. To recap, the committor is the probability that trajectories initialized from a conformation lead to the up (i.e., active) state before the down (i.e., resting) state, so the down state corresponds to $q_+ = 0$ and the up state corresponds to $q_+ = 1$. We used the committor to compute a two-dimensional PMF (Fig. 3c), revealing a qualitatively different picture than the PMF of the displacement charge alone. There are minima at $q_+ \approx 0.25$, 0.50, and 0.75, all with Q_d between $-3.5\ e$ and $-4\ e$. The simulations thus suggest that, while the displacement charge dynamics are essentially two-state, the microscopic dynamics are five-state. Moreover, the highest barrier is about 1.5 kcal/mol relative to the down state. To understand the differences between the PMFs and to characterize the intermediates and barriers between the down and up states, we sought to identify molecular features that correlated with the committor.

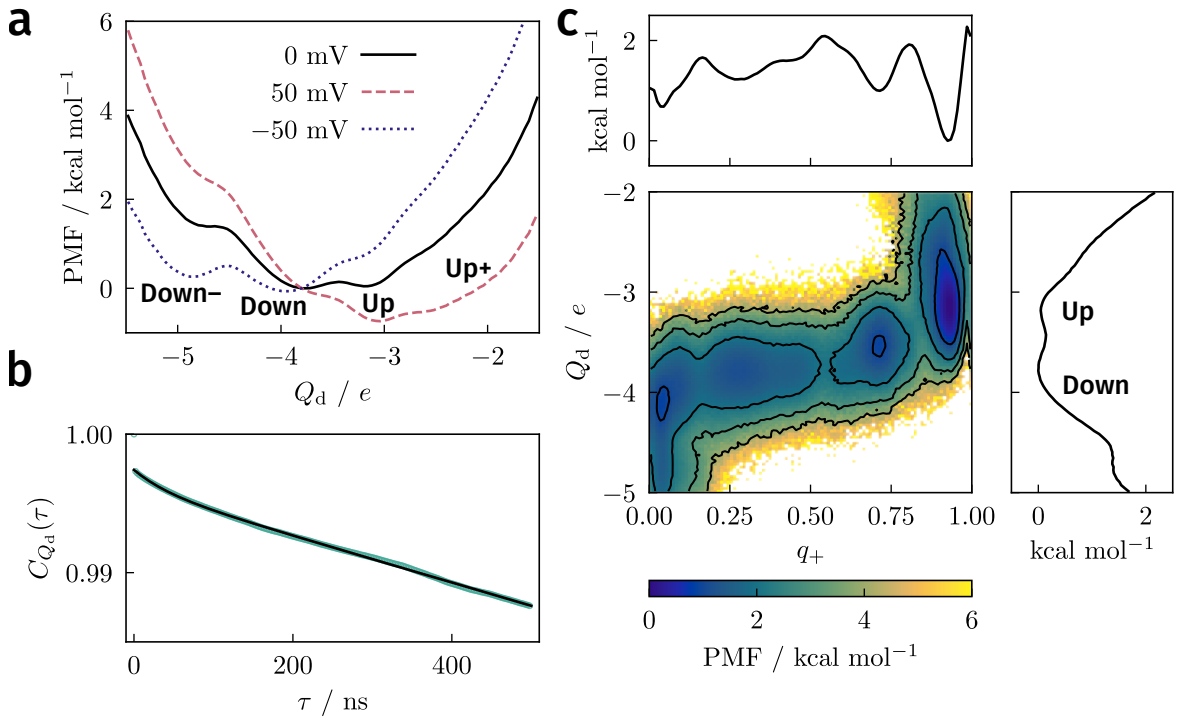

FIG. 3. Displacement charge. **a**, PMF computed from DGA along Q_d. The free-energy profiles at nonzero voltages were obtained by adding $Q_\mathrm{d}V$ to the profile obtained under 0 mV (note that differences in free energies between states depend only on their relative rather than absolute Q_d values). The down and up states correspond to Q_d values of -4.2 ± 0.2 e and -3.3 ± 0.2 e, respectively. **b**, Time correlation function $C_{Q_\mathrm{d}}(\tau) = \langle Q_\mathrm{d}(0)Q_\mathrm{d}(\tau)\rangle / \langle Q_\mathrm{d}(0)Q_\mathrm{d}(0)\rangle$ of the displacement charge. Green circles were computed from short trajectories by reweighting to equilibrium. Solid line represents a biexponential fit with time constants of 59 μs and 0.04 μs. **c**, PMF projected onto q_+ and Q_d and along each individual coordinate (above and right). Contours on the two-dimensional PMF are drawn every 1 kcal/mol.

B. Sparse modeling of the committor identifies physically interpretable variables that describe the transition

Following previous work[51,52], we sought to identify combinations of physically-intuitive collective variables (CVs) that can describe the dynamics by modeling the committor in terms of a small number of interpretable molecular features with LASSO regression (Section IV G). Out of the 60 inter-residue distances used as input to the committor (Section IV D), we manually selected several to track the movements of the side chains of the sensing arginines (R217, R223, R226, R229, and R232), as well as those of their acidic salt-bridge partners. These were distances between the sensing arginine C_ζ atoms and the C_γ/C_δ atoms of D129, E183, and D186 (Fig. 1). 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 resting and active states[27,28,38]. 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)[27]. Finally, we included the S4 helix translocation and rotation, given their known role in VSD activation[25,36,39,53–55].

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 R226-F161 distance decreases and then increases as R226 goes through the hydrophobic plug). Consequently, we divided the reaction into two stages, one from the down (reactant) state to $q_+ = 0.5$ (i.e., the transition state ensemble in the sense that conformations with $q_+ = 0.5$ have an equal likelihood of leading to the down and up states), and another from $q_+ = 0.5$ to the up (product) state (Section IV G).

We found that nine CVs contributed significantly to at least one of the two models. They were 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)[56], 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

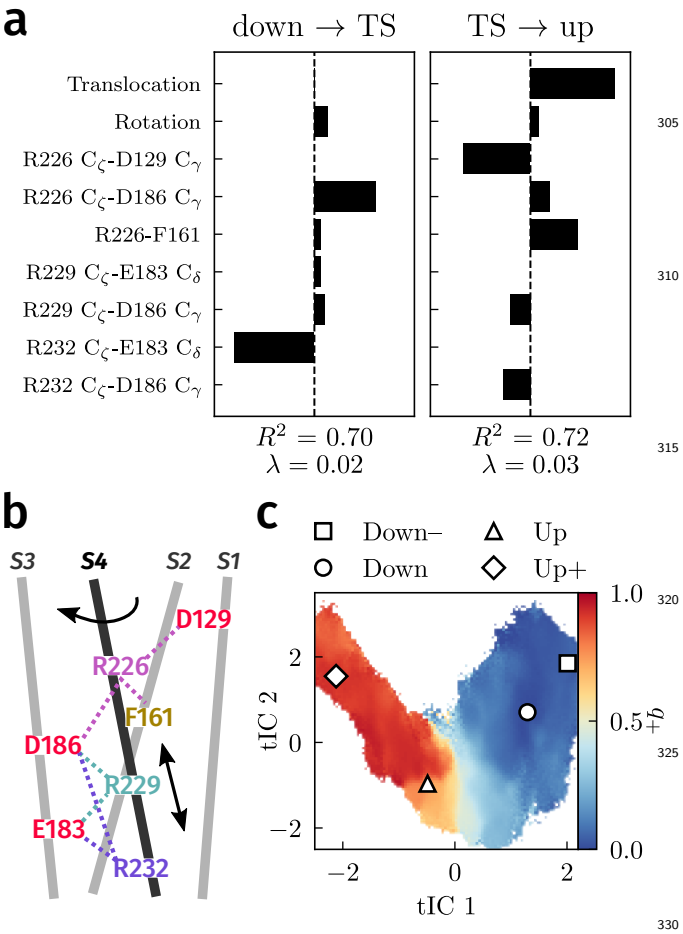

FIG. 4. Dependence of the committor on collective variables (CVs). **a**, Coefficients computed from LASSO regression for points with $q_+ \leq 0.5$ and $q_+ \geq 0.5$. **b**, Schematic of CVs (S4 helix variables and intramolecular distances) used for regression. **c**, Committor projected onto the first two nontrivial IVAC coordinates (tICs) constructed from the CVs in panel **a**.

fully captured using our chosen subset of variables. In addition, the committor projection on the IVAC coordinates has much lower variance compared with the projection on the S4 coordinates (Supplementary Fig. 3), reflecting the information gained from tracking the sensing arginines. Given this, we now interpret the PMF in Fig. 3c in terms of these variables. We first discuss the S4 helix translocation and rotation and then the side chain dynamics.

C. S4-helix translocation and rotation are loosely coupled

The up state crystal structure corresponds to the origin of the translocation-rotation plane, and the down state crystal structure lies at approximately (−4.5 Å, −60°) (Fig. 5). We can compute the PMF along these coordinates from our dataset and compare it directly with the PMF in Ref. 25 obtained using replica exchange umbrella sampling (REUS). The two PMFs are similar overall but differ slightly near the down state (Fig. 5a and Supplementary Fig. 4), and there is a prominent minimum with a depth of about 1–2 kcal/mol near (−5 Å, −10°). This minimum is also present in the REUS simulations at positive applied potentials, when up-like states are stabilized[25]. Structures in this minimum have the R226 side chain located above the hydrophobic plug. Because the sampling in these CVs is controlled directly in REUS, we expect the REUS PMF to be more accurate—DGA slightly underestimates the stability of the down state and the barrier to the up state.

We show histograms of translocation and rotation as functions of committor in Fig. 5c. These indicate that both variables have considerable freedom in the down state ($q_+ = 0$). Prior to the transition state ($q_+ < 0.5$), translocation remains unrestricted, but rotation becomes restricted and steadily increases towards 0° with the committor. After the transition state ($q_+ > 0.5$), translocation rapidly increases to its value in the up state ($q_+ = 1$). These results are consistent with the average committor, the PMF, and the reactive currents in this plane (Fig. 5a, b). The transition state ($q_+ = 0.5$) appears as a nearly horizontal white line just below the up state, and the reactive current is in the vertical direction.

Historically, two mechanisms have been proposed for S4 helix movement: a helical-screw mechanism in which translocation and rotation advance together[53–55,57,58] and a sliding-helix mechanism in which the α-helix converts to a narrower 3_{10}-helix, allowing it to translocate independently of rotation[8,47]. In contrast to the helical-screw mechanism, we find that the S4 helix is relatively free to translocate back and forth prior to the transition state (Fig. 5c), and it is the restriction of the rotational degree of freedom that determines when the transition state is crossed. At the same time, we find little 3_{10}-helix content to support the sliding-helix mechanism (Supplementary Fig. 5), consistent with prior analyses[25]. The mechanism that we observe combines the rotation of the helical-screw mechanism with the translocation of the sliding-helix mechanism, resembling the action of a barrel slide bolt. In the next section, we elaborate this mechanism by considering the side-chain dynamics.

D. Transitions between intermediates involve exchange of arginine salt-bridge partners

In addition to the S4 helix variables, the sparse model from LASSO regression (Fig. 4a) shows that the movements of R226, R229, and R232 relative to their salt bridge partners and the hydrophobic plug are important, and we find that they can account for the structure in Fig. 3c. Moreover, we observe in our simulations that vertical translocation of the S4 helix can precede or lag behind the exchange of salt bridge partners by the sensing arginines (Supplementary Fig. 6). That is, changes in

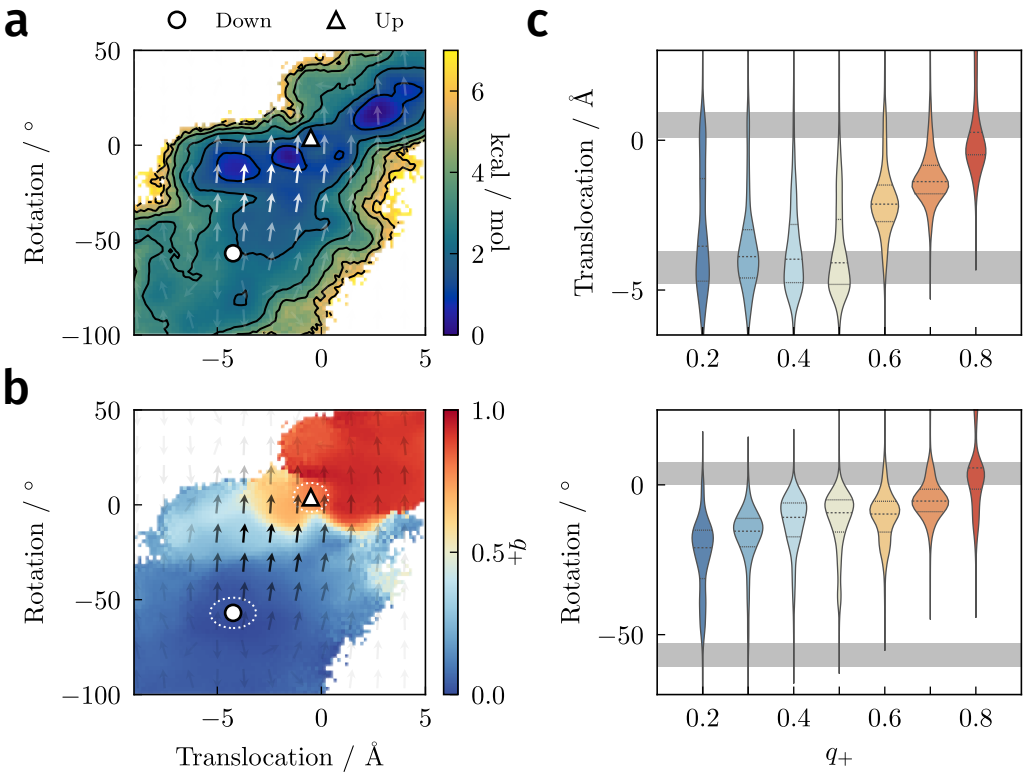

FIG. 5. Role of the S4 helix. **a**, PMF as a function of S4 helix translocation and rotation. Reactive currents are shown as a vector field with opacity indicating the magnitude of the vector. Contours are drawn every 1 kcal/mol. **b**, Average committor as function of S4 helix translocation and rotation. Approximate definitions of the down and up states for the committor calculations are marked with white ellipses. **c**, Distributions (violin plots) of S4 helix translocation and rotation as a function of committor. Each violin is obtained by binning structures with the indicated $q_+ \pm 0.05$ (i.e., the first violin contains structures with $q_+ \in [0.15, 0.25]$). The median and upper/lower quartiles are denoted with dashed and solid lines, respectively. Approximate values of translocation or rotation in the down and up states marked by gray bars.

the side chains of R226, R229 (and to some extent R232), translocation, and rotation are quasi-independent. Here, we describe the successive movements from the down state to the minima at $q_+ \approx 0.25$, 0.50, and 0.75, and finally to the up state.

In the down state, R226 and R229 form salt bridges with D186 and E183, respectively (Fig. 1), while the guanidinium group of R232 forms hydrogen bonds to phosphate groups of lipids (Supplementary Figs. 7a and 8a, c), a role previously noted in studies of VSDs[31,39,59–61]. The aliphatic part of the R226 side chain is in contact with the hydrophobic plug.

The transition to the minimum at $q_+ \approx 0.25$ involves the movement of R226's guanidinium group into the plug so as to interact with the phenyl side chain of F161 (Fig. 6a, b and Supplementary Fig. 9a, b). Although cation-π interactions and π-π stacking have been suggested to stabilize interactions between arginine and phenylalanine residues[27,62–64], the fixed charges in the force field do not directly represent these (quadrupole) interactions. Instead, this minimum is primarily stabilized by R226 forming a salt bridge with D129 from the hydrophobic plug (Supplementary Fig. 9c). Consistent with the discussion above, making these interactions relies on S4 helix rotation but not translocation (Supplementary Fig. 10).

The next barrier between the minima at $q_+ \approx 0.25$ and $q_+ \approx 0.50$ is very low, and R229 interacts with both E183 and D186 to varying degrees in this range of committor values (Fig. 6a and Supplementary Fig. 7a). The transition from the first intermediate minimum to the second reflects the movement of R232 away from the phosphate groups of the lipids to interact with E183 (Fig. 6a, Supplementary Figs. 7a and 10). In the intermediate near $q_+ \approx 0.5$ (which can be considered the transition state ensemble, as noted above), we find significant heterogeneity in R229's conformations: the R229-E183 distance, for example, exhibits two populations, one centered around 4 Å and another from 6–10 Å (Fig. 6a and Supplementary Fig. 7b). In the former, R229 is coordinated to both E183 and D186, whereas in the latter R229 is only coordinated to D186 (Fig. 6c). Accordingly, R229 forms hydrogen bonds mainly with D186, but also with E183 and lipid phosphate groups when $q_+ < 0.5$ (Supplementary Figs. 7a and 8). As discussed earlier, translocation of the S4 helix is relatively unconstrained in the intermediate stages of the transition (Fig. 5), with heterogeneity

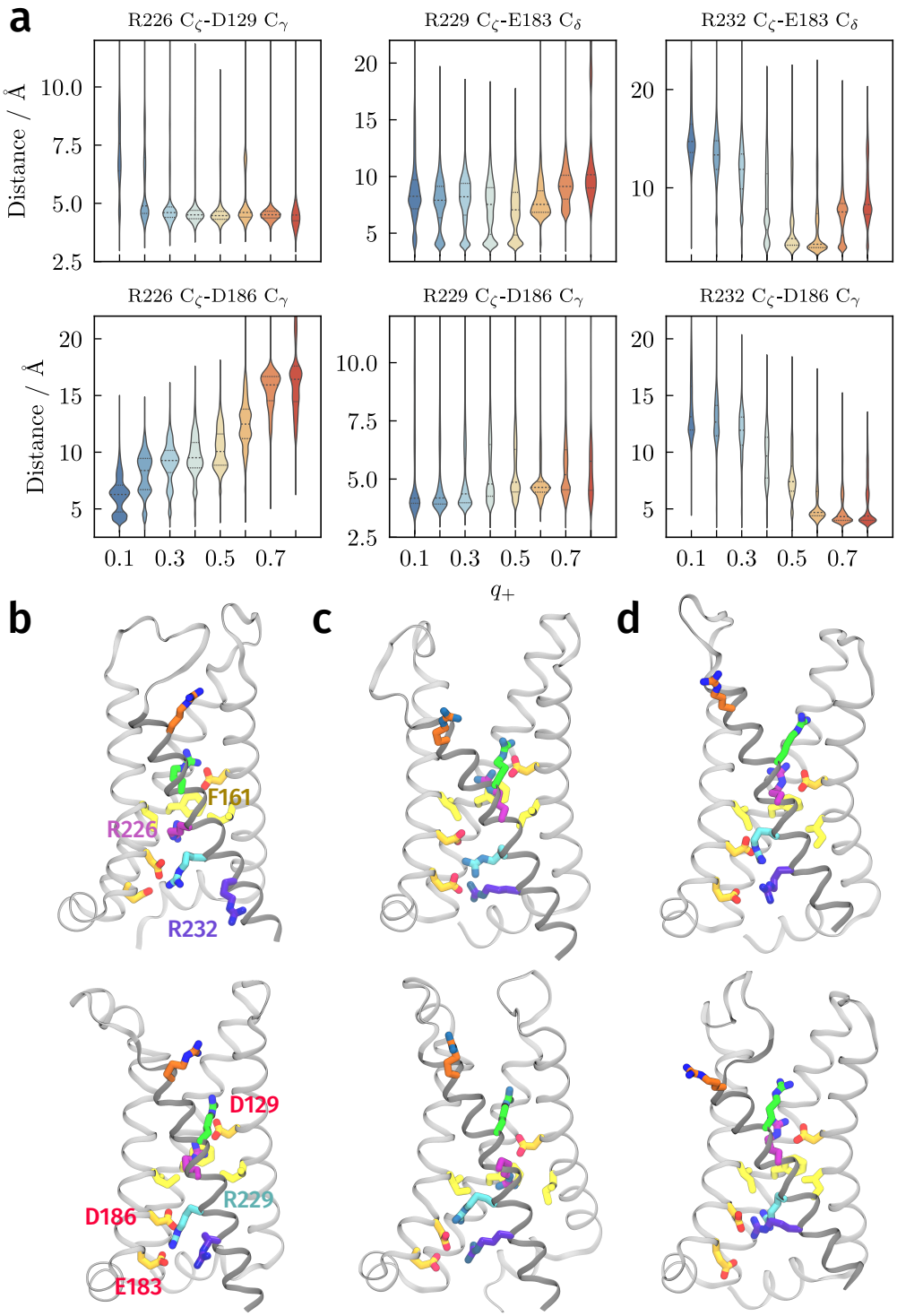

FIG. 6. Role of sensing arginine salt bridges and structure of the intermediates. **a**, Distribution of salt-bridge distances as a function of committor. Each violin is obtained by binning structures with the indicated $q_+ \pm 0.05$ (e.g., the first violin contains structures with $q_+ \in [0.05, 0.15]$). The median and upper/lower quartiles are denoted with dashed and solid lines, respectively. **b, c, d**, Selected structures from the intermediates at $q_+ \approx 0.25$, $q_+ \approx 0.5$, and $q_+ \approx 0.75$ from k-medoids clustering.

seen in transition state structures.

The transition to the minimum at $q_+ \approx 0.75$ requires the dissociation of R229 from E183 and the movement of R232 to interact with D186 (Fig. 6a, d and Supplementary Fig. 10). The barrier separating the minima at $q_+ \approx 0.50$ and $q_+ \approx 0.75$ results from the exchange of the two salt bridges with R229 and R232. Furthermore, after the transition state $q_+ \approx 0.5$, S4 helix translocation begins increasing (Fig. 5b and Supplementary Fig. 10), and the completion of this process results in the final transition to $q_+ = 1$. In some structures with high committor values, R226 extends to interact with S158 and T197 (Supplementary Fig. 7c).

III. CONCLUDING DISCUSSION

In this paper, we used recent advances in methods for computing kinetic statistics to elucidate the mechanism of the down-up transition of Ci-VSD. In contrast to previous studies of voltage-sensing proteins, which have focused mainly on metastable states of the voltage-sensing domain[25,36,39,40], our work dissects the microscopic events underlying a single "click".

As summarized in Fig. 7, the down-up transition proceeds through a stepwise mechanism analogous to a caterpillar traversing a rugged landscape (Supplementary Movie 1). The overall displacement of the helix and the rearrangements of the side chains are only loosely coordinated. The movements of R226, R229, and R232 do not follow a strict sequence, though steric interference prevents them from being totally independent. From the down state to $q_+ \approx 0.25$, rotation of the S4 helix allows R226 to detach from D186, move through the hydrophobic plug, and form a salt bridge with D129 (Fig. 7a, b). At $q_+ \approx 0.5$, where the system has an equal likelihood of proceeding forward and backward, R229 delocalizes and interacts with both E183 and D186, while R232 detaches from the lipid phosphate head groups and forms a salt bridge with E183 (Fig. 7c). At $q_+ \approx 0.75$, R229 forms a salt bridge to D186 and R232 interacts with both E183 and D186 (Fig. 7d). Finally, the S4 helix must translocate to its up state position while R229 enters the plug to occupy the position of R226 in the down state (Fig. 7d, e).

Although the free-energy landscape along the displacement charge appears two-state, our analysis based on the committor reveals the presence of metastable intermediates. These intermediates may help explain the multi-exponential kinetics of fluorescence relaxation measured for labeled VSDs[17,47–50]. Given that the PMFs are relatively shallow, even under applied transmembrane potentials of ±150 mV[25], there could also be contributions from transitions involving the down− and up+ states, which could introduce additional timescales. Regarding the kinetics, it is important to note that our estimates of the activation timescales are systematically too fast. This may reflect a breakdown in the Markov assumption underlying the DGA calculations, limitations of the choice of basis, insufficient sampling, and limitations of the force field. The first two could be addressed by including non-Markovian effects[65] in the DGA calculations or by applying nonequilibrium sampling methods[66].

Despite the potential issues, we expect that our simulations represent the microscopic structures and dynamics of Ci-VSD reasonably well. We consequently expect the "caterpillar" mechanism that we describe to hold for mildly depolarizing and hyperpolarizing voltages. This mechanism represents a significant deviation from a canonical helical-screw model in that translocation, rotation, and side-chain movements are only loosely coupled. While we are not aware of a previous study that explicitly describes such a mechanism, some decoupling of the S4 translocation, rotation, and side-chain movements has been noted in passing for simulations of Ci-VSD[25] and Kv1.2[33,36].

Another striking feature of our simulations is the relatively low barrier between the down and up states in Ci-VSD compared with, for example, the barrier between the δ and ε states of the Kv1.2 channel (5 kcal/mol)[40]. Given Ci-VSP's sequence divergence from VGICs and comparatively slower gating kinetics[16,21,22,28,47,67], we believe the smaller barrier is genuine and not an artifact of different computational approaches. The lower barrier could furthermore be partially attributed to the lack of 3_{10} helix formation[8,25].

We expect that R226's movement through the hydrophobic plug is facilitated by its ability to make a salt bridge with D129 while in contact with the I126 and F161 side chains (Supplementary Fig. 9c). In contrast, VGICs generally have a polar but uncharged amino acid at the site corresponding to D129[20,68], preventing formation of an analogous interaction and leading to a significant dielectric barrier[69]. Perhaps relatedly, in VGICs, reducing the hydrophobicity of gating-pore residues by mutation accelerates activation upon depolarization[69], whereas in Ci-VSP kinetics are more closely correlated with the sizes of side chains of hydrophobic-plug residues[28,70]. The closer spacing of the acidic residues in Ci-VSD compared with VGICs may also explain how the arginine side chains can make salt bridges without the S4 helix needing to convert to a 3_{10}-helix near the hydrophobic plug, as in the sliding-helix mechanism[8,25,37,71].

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[17,18,30,47], or while perturbing residues implicated in key rearrangements. For example, the importance of R226 side-chain reorientation during early stages of activation suggests that limiting rotation of the χ_1 dihedral angle (perhaps by introducing an unnatural arginine congener[72]) might alter the kinetics of activa-

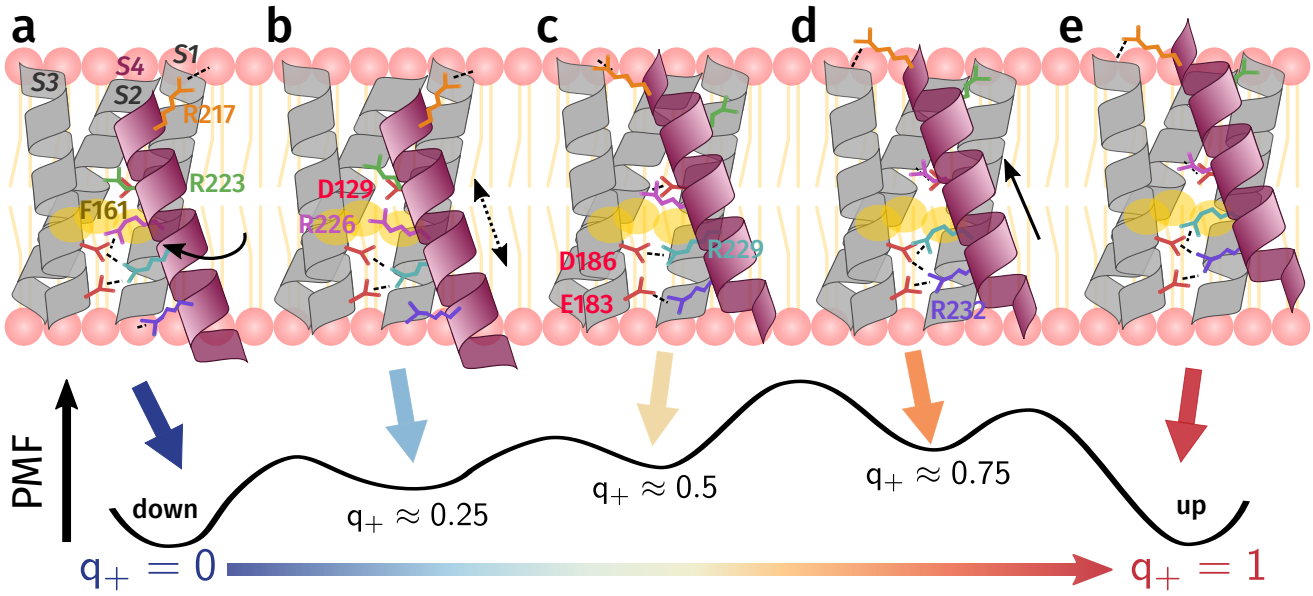

FIG. 7. Schematic of the mechanism of the down-up transition in Ci-VSD. **a**, In the down state ($q_+ = 0$), R223 forms a salt bridge with D129 above the hydrophobic plug. R226 is hydrated, while R229 interacts with mainly with E183 and D186, and R232 coordinates to POPC headgroups. **b**, First ($q_+ \approx 0.25$), R226 rotates upwards as the S4 helix rotates, which allows it to cross the hydrophobic plug and begin forming a salt bridge with D129. **c**, Near the transition state ($q_+ \approx 0.5$), R226 interacts with D129, while R229 interacts with both E183 and D186. **d**, Beyond the transition state ($q_+ \approx 0.75$), R229 detaches from E183, and R232 interacts with both E183 and D186. **e**, The S4 helix translocates to its final up state position ($q_+ = 1$).

tion or block it entirely. Other residues not analyzed in detail, such as S158 and T197 (Supplementary Fig. 7c), may modulate intermediate states of the activation process by interacting with the sensing arginines, and their role could be tested by observing the effect of mutations on activation kinetics.

Notably, the physically interpretable low-dimensional model of the committor obtained from our LASSO procedure did not require distances to the sensing arginines R217 and R223. R217's interactions with negative countercharges such as D129, D136, and D154 correlate only weakly with the committor and are heterogeneous in the transition state (Supplementary Fig. 7b). Our results are consistent with prior suggestions that R217 primarily participates in activation by modulating the electric field rather than through salt-bridge formation[22], 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. On the other hand, R223 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 displacement charge is small, consistent with its exposure to solvent in the extracellular crevice of Ci-VSD. Prior studies that mutated this amino acid to cysteine or histidine[22,25] suggested that R223 may play a larger role in transitions involving the down– and up+ states of Ci-VSD. Computationally, it would therefore be interesting to study these other transitions, as well as the down-up transition in the presence of applied fields.

IV. THEORY AND METHODS

We compute statistics from MD simulations using the dynamical Galerkin approximation (DGA)[41,42]. The aim of DGA is to formulate and solve a problem for these statistics as a linear system of equations. The theoretical framework is based on transition path theory[43] and only assumes that the dynamics are Markovian within a user defined time interval termed the lag time (t below). The key advantage of this approach is that estimating the terms in the linear equations requires only short trajectories (on the order of the lag time) rather than ones recapitulating full transition paths from the down state to the up state. The main challenge is ensuring that the short trajectories in the aggregated simulation dataset adequately sample all parts of the desired transition (see Section IV E.)

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 identify the most important collective variables (CVs) for describing the conformational change. Readers who wish to obtain an overview of the DGA framework without delving into the mathematical details need only skim Sections IV A, IV B, and IV D.

A. Key statistics

As described in Section II, we use three statistics to understand the mechanism of the down-up transition: the equilibrium probability, the committor, and the reactive current. The equilibrium distribution (or more precisely, its ratio to the sampling distribution, w) and the committor, q_+, satisfy equations of transition operators, which describe how averages of functions of the configuration evolve under the dynamics. The reactive current, J^{ξ}_{AB}, can be computed from the equilibrium distribution and the committor. In this section, we define the operators mathematically and present key equations for these statistics.

For a Markov process $X(t)$, the transition operator $\mathcal{T}_t$ 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 average of f (over multiple random realizations of X) for a lag time t, conditioned on starting at x.

The PMF is the free energy as a function of specific coordinates. 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). Specifically, we introduce a weight factor $w(x)$ that relates the sampled data distribution $\mu(x)$ (which need not be drawn from π) to $\pi(x)$:

$$\int f(x)\pi(x)\,\mathrm{d}x = \int f(x)w(x)\mu(x)\,\mathrm{d}x. \tag{2}$$

Because $\pi(x)$ is the equilibrium distribution, by definition it must not change over time or, in other words, upon application of the transition operator. The weight factor can be shown[42,73] to satisfy

$$(\mathcal{T}_t)^{\dagger}_{\mu} w(x) = w(x), \tag{3}$$

where $(\mathcal{T}_t)^{\dagger}_{\mu}$ is the μ-weighted adjoint of $\mathcal{T}_t$. 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[43]. Because of the boundary conditions $q_+(x) = 0$ in A and $q_+(x) = 1$ in B, the committor is governed by a modified transition operator

$$\mathcal{S}_t f(x) = \mathbf{E}[f(X(t \wedge T_{A\cup B})) \mid X(0) = x], \tag{4}$$

where $T_{A\cup B} = \min\{t \geq 0 \mid X(t) \in A \cup B\}$ 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[42]

$$q_+(x) = \begin{cases} \mathcal{S}_t q_+(x) & x \in (A \cup B)^{\mathsf{c}} \\ 0 & x \in A \\ 1 & x \in B. \end{cases} \tag{5}$$

Crucially, (5) holds for *any* lag time t, so in principle we can obtain q_+ from simulations that are much shorter than the transition time (i.e., $t \ll T_{A\cup B}$).

Finally, the reactive current measures the flux of reactive probability density, the probability of observing a reactive trajectory transitioning from A to B[43,74]. The reactive current is most easily interpreted after projecting onto a subspace of CVs, $\xi(x)$, yielding a vector field $J^{\xi}_{AB}(s)$ that summarizes how reactive trajectories flow through $\xi(x) = s$. It was shown previously that the projected reactive probability current satisfies[42,66]

$$J^{\xi}_{AB}(s) = \int J_{AB}(x) \cdot \nabla\xi(x)\delta(\xi(x) - s)\pi(x)\,\mathrm{d}x, \tag{6}$$

where J_{AB} is the reactive flux in the space of original coordinates. 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:

$$f(x) = \gamma(x) + \sum_{i=1}^{N} v_i \phi_i(x). \tag{7}$$

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. After substituting the ansatz in equation (7) into the appropriate operator equation [e.g., equation (5) for q_+], multiplying by another basis function ϕ_j, and integrating over the sampled distribution $\mu(x)$, we obtain the linear system

$$(C_t - C_0)v = -(r_t - r_0), \tag{8}$$

where

$$C^{ij}_t = \int \phi_i(x)\mathcal{S}_t\phi_j(x)\mu(x)\,\mathrm{d}x \tag{9a}$$

$$r^j_t = \int \phi_j(x)\mathcal{S}_t\gamma(x)\mu(x)\,\mathrm{d}x. \tag{9b}$$

We can then estimate the integrals in equation (9) from averages over an ensemble of simulations (that can be

much shorter than the transition time, as mentioned above). Indexing simulations by $m = 1, 2, ..., M$,

$$C_t^{ij} \approx \frac{1}{M} \sum_{m=1}^{M} \phi_i(X^m(0))\phi_j(X^m(t \wedge T_{A\cup B})) \tag{10a}$$

$$r_t^j \approx \frac{1}{M} \sum_{m=1}^{M} \phi_j(X^m(0))\gamma(X^m(t \wedge T_{A\cup B})). \tag{10b}$$

Given these quantities, equation (8) can be solved for v, which can be used to reconstruct f through equation (7).

When the basis functions are indicator functions on sets of configurations, DGA is equivalent to using an MSM with appropriate boundary conditions to compute dynamical statistics. DGA can thus be viewed as a generalization of MSMs to arbitrary basis sets. As such, the data requirements for DGA can never be more than for MSMs, and they can be less if a basis better than indicator functions can be found[41,42]. As we discuss further in Section IV D, we employ the MSM (indicator) basis set for computing the equilibrium probability, and a modified pairwise distance basis set[42] for the committor.

C. 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

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

where d and θ are the S4 helix translocation and rotation, respectively. The centers and radii of the ellipses were obtained from several nanoseconds-long MD simulations initialized from the down and up states, rather than the crystal structures, since we wanted to include typical fluctuations. Visual inspection of the structures and trajectories showed that S4 helix movements often preceded or lagged 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.

D. Basis set construction

For the committor calculations, we used a basis set built from pairwise intramolecular distances[42]. The distances are based on two groups of residues: (1) R217, R223, R226, R229, and R232 and (2) D129, D136, D151, D164, 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 committor 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 function $h(x) = d_A d_B/(d_A + d_B)^2$, which obeys the boundary conditions by construction, and set $\phi_i(x) = x_i h(x)$ where x_i is the ith component of the 60-dimensional coordinate vector. Finally, we use singular value decomposition to whiten the basis functions. The guess function is $\gamma(x) = d_A^2/(d_A^2 + d_B^2)$, which also satisfies the boundary conditions by construction.

For the weight factor, w, a basis of indicator functions (corresponding to an MSM) was used, since the distance basis tended to yield noisier results. Instead of the 60 salt-bridge distances, we used the minimum heavy-atom distance between each residue in the S4 helix and each residue in the S1–S3 helices (1924 distances in total). We reduced the feature vector to 10 dimensions using the integrated variational approach to conformational dynamics (IVAC)[56] and then clustered 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.

E. Simulation details and adaptive sampling procedure

The starting point for our simulations was the crystal structure of the up state (PDB ID 4G7V)[21]. 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[76]. We then used the program VMD[77] 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[78] for the protein, phospholipids, and ions, and the TIP3P model[79] for water molecules. We used the hydrogen mass repartitioning scheme[80] 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

without an applied voltage. The temperature and pressure were controlled using the Nosé–Hoover thermostat and the semi-isotropic MTK barostat, respectively[81–83]. Long-range electrostatic interactions were calculated using the u-series method[84], and van der Waals interactions were truncated at a cutoff distance of 9 Å.

To construct a dataset that adequately sampled the transition, we employed a two stage approach (Fig. 2). First, we drew initial conditions uniformly from REUS windows in the space of S4 helix CVs[25]. The coordinates, corresponding velocities and periodic boundary parameters were used to initialize Anton 2 simulations[44]. From these initial conditions, we 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 expanded our dataset by sampling adaptively. Specifically, we selected structures with estimated committor values near $q_+ = 0.5$ and with under-sampled values of CVs tracking the S4 helix or arginine side chain motions.

In the initial dataset, we observed that some 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 ameliorate this issue. These steered molecular dynamics simulations[85,86] were performed with 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. In total, we accumulated 415 μs of trajectories, with coordinates saved every 0.1 ns.

F. Collective variables

We computed the intramolecular distances used as inputs to DGA using PyEMMA[87] and MDTraj[88]. 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[89,90]. 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[91]. 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"[45]:

$$Q_\mathrm{d} = \sum_i q_i \left(\frac{z_i^{(u)} + L_z/2}{L_z} \right), \tag{11}$$

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).

G. Sparse regression

To identify a small number of variables that are important for the dynamics, we modeled the committor using LASSO[51,92] (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:

$$\mathcal{L}_\mathrm{LASSO} = \sum_k (\beta^\top x^{(k)} - q_+(x^{(k)}))^2 + \lambda \|\beta\|_1, \tag{12}$$

where $\|\zeta\|_1 = \sum_i |\zeta_i|$ is the ℓ_1 norm and λ is a parameter controlling the penalty strength. In (12), $x^{(k)}$ 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, 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 used an inverse sigmoid transform to map the DGA-estimated committor to the entire range $(-\infty, \infty)$ so that the LASSO-predicted committors were 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. 11). Non-linear functions of the features, such as a polynomial or a neural network function, likely would better describe the committor. However, we used a linear model in the interest of maximizing interpretability.

DATA AVAILABILITY

Simulation data generated on Anton 2, as well as derived data (featurized MD trajectories and calculations) are available upon request from the authors.

CODE AVAILABILITY

Code to reproduce the analysis and plots in this work, as well as instructions for running calculations with example data, can be found at https://github.com/dinner-group/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] Catterall, W. A. Ion channel voltage sensors: Structure, function, and pathophysiology. *Neuron* **67**, 915–928 (2010).

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

[9] Kintzer, A. F. & Stroud, R. M. Structure, inhibition and regulation of two-pore channel TPC1 from Arabidopsis thaliana. *Nature* **531**, 258–264 (2016).

[10] Guo, J. *et al.* Structure of the voltage-gated two-pore channel TPC1 from Arabidopsis thaliana. *Nature* **531**, 196–201 (2016).

[11] Wisedchaisri, G. *et al.* Resting-state structure and gating mechanism of a voltage-gated sodium channel. *Cell* **178**, 993–1003.e12 (2019).

[12] Xu, H. *et al.* Structural basis of Nav1.7 inhibition by a gating-modifier spider toxin. *Cell* **176**, 702–715.e14 (2019).

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

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

[15] Mandala, V. S. & MacKinnon, R. Voltage-sensor movements in the EagKv channel under an applied electric field. *Proc. Natl Acad. Sci. USA* **119**, e2214151119 (2022).

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

[17] 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).

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

[19] Hobiger, K., Utesch, T., Mroginski, M. A. & Friedrich, T. Coupling of Ci-VSP modules requires a combination of structure and electrostatics within the linker. *Biophys. J.* **102**, 1313–1322 (2012).

[20] Okamura, Y., Kawanabe, A. & Kawai, T. Voltage-sensing phosphatases: biophysics, physiology, and molecular engineering. *Physiol. Rev.* **98**, 2097–2131 (2018).

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

[22] 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).

[23] 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. USA* **105**, 15142–15147 (2008).

[24] 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. USA* **106**, 22498–22503 (2009).

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

[26] Ahern, C. A. & Horn, R. Focused electric field across the voltage sensor of potassium channels. *Neuron* **48**, 25–29 (2005).

[27] Tao, X., Lee, A., Limapichat, W., Dougherty, D. A. & MacKinnon, R. A gating charge transfer center in voltage sensors. *Science* **328**, 67–73 (2010).

[28] Lacroix, J. J. & Bezanilla, F. Tuning the voltage-sensor motion with a single residue. *Biophys. J.* **103**, L23–l25 (2012).

[29] 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).

[30] Grimm, S. S. & Isacoff, E. Y. Allosteric substrate switching in a voltage-sensing lipid phosphatase. *Nat. Chem. Biol.* **12**, 261–267 (2016).

[31] Jogini, V. & Roux, B. Dynamics of the Kv1.2 voltage-gated K+ channel in a membrane environment. *Biophys. J.* **93**, 3070–3082 (2007).

[32] 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).

[33] 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).

[34] Treptow, W., Tarek, M. & Klein, M. L. Initial response of the potassium channel voltage sensor to a transmembrane potential. *J. Am. Chem. Soc.* **131**, 2107–2109 (2009).

[35]Khalili-Araghi, F. *et al.* Calculation of the gating charge for the Kv1.2 voltage-activated potassium channel. *Biophys. J.* **98**, 2189–2198 (2010).

[36]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. USA* **108**, 6109–6114 (2011).

[37]Schwaiger, C. S., Bjelkmar, P., Hess, B. & Lindahl, E. 310-helix conformation facilitates the transition of a voltage sensor S4 segment toward the down state. *Biophys. J.* **100**, 1446–1454 (2011).

[38]Schwaiger, C. S. *et al.* The free energy barrier for arginine gating charge translation is altered by mutations in the voltage sensor domain. *PLoS One* **7**, e45880 (2012).

[39]Jensen, M. *et al.* Mechanism of voltage gating in potassium channels. *Science* **336**, 229–233 (2012).

[40]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. USA* **112**, 124–129 (2015).

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

[42]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).

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

[44]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).

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

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

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

[48]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).

[49]Villalba-Galea, C. A. *et al.* Charge movement of a voltage-sensitive fluorescent protein. *Biophys. J.* **96**, L19–l21 (2009).

[50]Villalba-Galea, C. Voltage-controlled enzymes: The new Janus Bifrons. *Front. Pharmcol.* **3** (2012).

[51]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).

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

[53]Pathak, M. M. *et al.* Closing in on the resting state of the Shaker K+ channel. *Neuron* **56**, 124–140 (2007).

[54]Campos, F. V., Chanda, B., Roux, B. & Bezanilla, F. Two atomic constraints unambiguously position the S4 segment relative to S1 and S2 segments in the closed state of Shaker K+ channel. *Proc. Natl Acad. Sci. USA* **104**, 7904–7909 (2007).

[55]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).

[56]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).

[57]Guy, H. R. & Seetharamulu, P. Molecular model of the action potential sodium channel. *Proc. Natl Acad. Sci. USA* **83**, 508–512 (1986).

[58]Catterall, W. A. Molecular properties of voltage-sensitive sodium channels. *Annu. Rev. Biochem.* **55**, 953–985 (1986).

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

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

[61]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).

[62]Gallivan, J. P. & Dougherty, D. A. Cation-π interactions in structural biology. *Proc. Natl. Acad. Sci. USA* 6 (1999).

[63]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).

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

[65]Dominic III, A. J. *et al.* Building insightful, memory-enriched models to capture long-time biochemical processes from short-time simulations. *Proc. Natl Acad. Sci. USA* **120**, e2221048120 (2023).

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

[67]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).

[68]Palovcak, E., Delemotte, L., Klein, M. L. & Carnevale, V. Evolutionary imprint of activation: The design principles of VSDs. *J. Gen. Physiol.* **143**, 145–156 (2014).

[69]Catacuzzeno, L., Sforna, L., Franciolini, F. & Eisenberg, R. S. Multiscale modeling shows that dielectric differences make Nav channels faster than Kv channels. *J. Gen. Physiol.* **153**, e202012706 (2021). URL `https://doi.org/10.1085/jgp.202012706`.

[70]Lacroix, J. J., Hyde, H. C., Campos, F. V. & Bezanilla, F. Moving gating charges through the gating pore in a Kv channel voltage sensor. *Proc. Natl Acad. Sci. USA* **111**, E1950–e1959 (2014). URL `https://www.pnas.org/doi/full/10.1073/pnas.1406161111`.

[71]Lee, C.-H. & MacKinnon, R. Voltage sensor movements during hyperpolarization in the HCN channel. *Cell* **179**, 1582–1589.e7 (2019).

[72]Leisle, L., Valiyaveetil, F., Mehl, R. A. & Ahern, C. A. Incorporation of non-canonical amino acids. *Adv. Exp. Med. Biol.* **869**, 119–151 (2015).

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

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

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

[76]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).

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

[78]Huang, J. *et al.* CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2017).

[79] 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).

[80] 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).

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

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

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

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

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

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

[87] 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).

[88] McGibbon, R. T. *et al.* MDTraj: A modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.* **109**, 1528–1532 (2015).

[89] 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).

[90] 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).

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

[92] Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc., B: Stat. Methodol.* **58**, 267–288 (1996).

ACKNOWLEDGEMENTS

The authors would like to 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. The "Beagle-3: A Shared GPU Cluster for Biomolecular Sciences" is supported by the NIH under the High-End Instrumentation (HEI) grant program award 1S10OD028655-0.

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).