

# Multiscale Responsive Kinetic Modeling: Quantifying Biomolecular Reaction Flux under Varying Electrochemical Conditions

*Hannah Weckel-Dahman<sup>1</sup>, Ryan Carlsen<sup>1</sup> and Jessica M.J. Swanson<sup>1\*</sup>*

<sup>1</sup>Department of Chemistry, University of Utah, Salt Lake City, UT, 84112 – United States of America

## **ABSTRACT**

Attaining a complete thermodynamic and kinetic characterization for processes involving multiple interconnected rare-event transitions remains a central challenge in molecular biophysics. This challenge is amplified when the process must be understood under a range of reaction conditions. Herein, we present a condition-responsive kinetic modeling framework that can combine the strengths of bottom-up rate quantification from multiscale simulations with top-down solution refinement using experimental data. Although this framework can be applied to any process, we demonstrate its use for electrochemically driven transport through channels and transporters. Using the Cl<sup>-</sup>/H<sup>+</sup> antiporter CIC-ec1 as a model system, we show how robust and predictive kinetic solutions can be obtained when the solution space is grounded by thermodynamic constraints, seeded through multiscale rate quantification, and further refined with experimental data, such as

electrophysiology assays. Turning to the Shaker  $K^+$  channel, we demonstrate that robust solutions and biophysical insights can also be obtained with sufficient experimental data. This multi-pathway method proves capable of identifying single-pathway dominant mechanisms but also highlights that competing and off-pathway flux is still essential to replicate experimental findings and to describe concentration-dependent channel rectification.

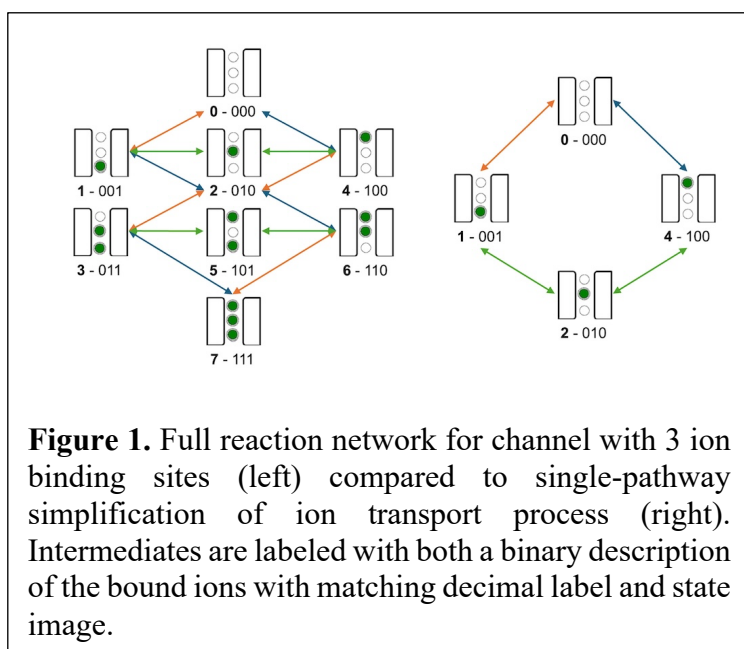
## 1. INTRODUCTION

The mechanisms of most biomolecular processes are well described by reactant and product states that are separated by a finite number of metastable intermediates (states) connected by rare-event transitions. As long as rates of transitions are known, the reactive flux through these intermediates in response to non-equilibrium starting populations is easily described by a set of ordinary differential equations according to the kinetic master equation. Connecting this coarse-grained kinetic description with a molecular-level understanding of the involved rare-event transition ensembles lends unprecedented insight into the mechanism and how it responds to non-equilibrium conditions. However, defining a robust and accurate kinetic model is challenging as it requires identifying all relevant intermediates, determining which intermediates are connected by kinetically relevant transitions, and quantifying the transition rates for the range of conditions of interest. To help address these challenges, we present multiscale responsive kinetic modeling (MsRKM), a framework that combines the strengths of computational characterization of intermediates and rare-event transitions with experimental data to identify robust kinetic solutions. This enables the description of any biomolecular processes involving multiple rare event transitions in terms of the underlying mechanistic reaction network and the ensemble flux through intermediates in response to initial nonequilibrium and equilibrium conditions. MsRKM includes

the dynamic response of the underlying rate coefficients to the influence of transmembrane voltage, enabling the description of electrochemically driven transport through channels and transporters.

Part of the motivation for this work is an increasing recognition of the importance of kinetic selection and mechanistic heterogeneity.<sup>3-9</sup> Although some processes are certainly dominated by a single pathway (sequence of intermediates), many others involve off-pathway transitions and condition-varying flux through alternative pathways that can play essential roles in mechanistic

outcomes (Figure 1). A prime example is kinetic proofreading, as originally envisioned by Hopfield<sup>10</sup> and Ninio<sup>11</sup>, in which competing pathways increase substrate selectivity when a slow step follows substrate association. This decreases the error rates in processes requiring high fidelity, such as replication,



**Figure 1.** Full reaction network for channel with 3 ion binding sites (left) compared to single-pathway simplification of ion transport process (right). Intermediates are labeled with both a binary description of the bound ions with matching decimal label and state image.

translation, and transcription<sup>12</sup>, and has also been demonstrated in signal transduction<sup>13</sup> and pathogen recognition<sup>14, 15</sup>. Kinetic selection can also enable condition-specific mechanistic outcomes, as demonstrated in molecular machines that transform input-free energy (e.g., from an abundance of ATP) into various cellular processes. For example, branched networks play a central role in explaining both chemomechanical coupling<sup>16</sup> and the different mechanistic steps observed in motor proteins such as myosin and kinesin under different ATP/ADP ratios<sup>17, 18</sup>.

Another domain where kinetic selection is proving important is the controlled transport of solutes across membranes. Membrane channels and transporters play pivotal roles in a myriad of cellular processes (maintaining concentration gradients, triggering action potentials or cell signaling, importing nutrients, exporting waste and toxins, and regulating intracellular signaling) through the translocation of ions and substrates across cellular membranes. In transporters for which consistent stoichiometry must be obeyed across a range of conditions, such as the  $\text{Na}^+/\text{K}^+$  antiporter, exact precision via a single transport pathway may indeed be important. For others, however, kinetic selection between competing pathways explains non-integral and variable stoichiometries and enables essential function. For example, the sodium/glucose transporter was shown to use a substrate slip pathway to exchange some degree of efficiency for toxin discrimination.<sup>8, 19</sup> Additionally, the multidrug efflux pump EmrE was shown to employ multiple transport pathways to enable different transport regimes, including symport, antiport, and uniport, under different electrochemical driving forces.<sup>20</sup> These two examples demonstrate how the electrochemical conditions can be a controlling factor in substrate transport, as opposed to the more commonly recognized ligand binding, voltage gating, or ATP hydrolysis. Collectively, these findings demonstrate the importance of modeling transporters with a network description that includes the influences of competing pathways. However, it remains an outstanding question if ion channels also involve competing transport pathways to the degree that a network description is essential to understanding their transport mechanisms. The methods described herein were designed to help address this question.

Kinetic modeling can generally be approached in a ‘top-down’ physics-based theoretical manner using system knowledge and experimental data or in a ‘bottom-up’ tour-de-force quantification of rate coefficients for every transition, typically with simulations. Markov state modeling has

become a powerful approach of the latter form, stitching together simulation data to define metastable states and quantifying their transition probabilities.<sup>21, 22</sup> MsRKM can be used in any combination of purely top-down, as demonstrated for the Shaker Kv channel herein, or purely bottom-up, using simulation-based rate coefficients to describe system behavior under non-equilibrium conditions and test their ability to replicate experimental data. However, it is more powerfully employed by combining the two—narrowing the kinetic solution space based on calculated rate coefficients and refining the solution based on experimental data. This enables an experimentally driven theoretical framework that iteratively tests for consistency between molecular-level characterization and experimental behavior.

The choice of the word ‘responsive’ in the acronym deserves explanation. The ability of a kinetic model to describe a range of conditions requires the underlying rate coefficients to adapt to the conditions of interest. Since unimolecular rate coefficients typically remain stable despite changes in reactant and product concentrations, a single set of rate coefficients can generally capture non-equilibrium conditions by accommodating shifts in concentrations with bimolecular uptake coefficients. In contrast, rate coefficients can shift significantly in response to pH and voltage. MsRKM was developed to explicitly ‘respond to’ changes in voltage with voltage-dependent rate coefficients. This does not, however, include large voltage-induced conformational changes such as those seen in voltage-gated channels. Nor does the current implementation account for significant changes due to pH, which are often associated with a conformational change. Rather, distinct kinetic solutions should be developed for significantly different conformational ensembles. Future efforts will focus on an explicit inclusion of pH-dependent conformational changes and the corresponding rate changes. With the inclusion of voltage-dependent rate coefficients, processes driven by electrochemical gradients can be described by a single set of base

( $\Delta\Psi = 0\text{mV}$ ) rate coefficients, capturing non-equilibrium flux under varying electrochemical driving forces. This enables a direct comparison to typical electrophysiology assays.

MsRKM is demonstrated on two membrane-bound proteins: ClC-ec1, a secondary-active exchanger, and the Shaker Kv channel. ClC-ec1 is a well-studied  $\text{Cl}^-/\text{H}^+$  antiporter from *Escherichia coli* that is involved in cellular ion homeostasis and plays an important, though incompletely understood, role in the cell's acid resistance mechanism.<sup>23, 24</sup> Structurally, ClC-ec1 operates as a homodimer, with each subunit containing an independent ion-conducting pore.<sup>25</sup> Based on its measured reversal potential, ClC-ec1 was shown to exchange ions with a 2.2  $\text{Cl}^-$  to 1  $\text{H}^+$  stoichiometry.<sup>26</sup> This exchange is influenced by pH-dependent conformational changes near and within the pore<sup>26-29</sup>. Structure-function studies have identified the key residues involved in ion transport, and electrophysiology assays have revealed a range current-voltage behaviors under different electrochemical conditions.<sup>24-26, 29-32</sup> Simulations have characterized the proton transport mechanisms<sup>33, 34</sup> and pH-dependent conformational changes<sup>27-29</sup>. Additionally, previous kinetic models have suggested that ClC-ec1 uses multiple ion exchange pathways, each contributing to the total flux with varying magnitudes under different pH conditions.<sup>6, 9, 31</sup> Collectively, the extensive experimental and simulation-based analyses make ClC-ec1 an ideal, though still very challenging, model system for MsRKM—testing the method's ability to capture nuanced transporter behavior in response to changing electrochemical conditions.

Shaker is a voltage-gated potassium channel originally found in *Drosophila funebris*<sup>35</sup> that is essential for repolarizing the neuronal membrane after an action potential, thereby regulating neuronal excitability and firing patterns.<sup>36, 37</sup> Shaker is comprised of four subunits that form a central pore through which potassium ions pass when unobstructed by the activation gate.<sup>38, 39</sup> While Shaker has two distinct inactivation modes,<sup>40</sup> we focus on the fully open channel. Several

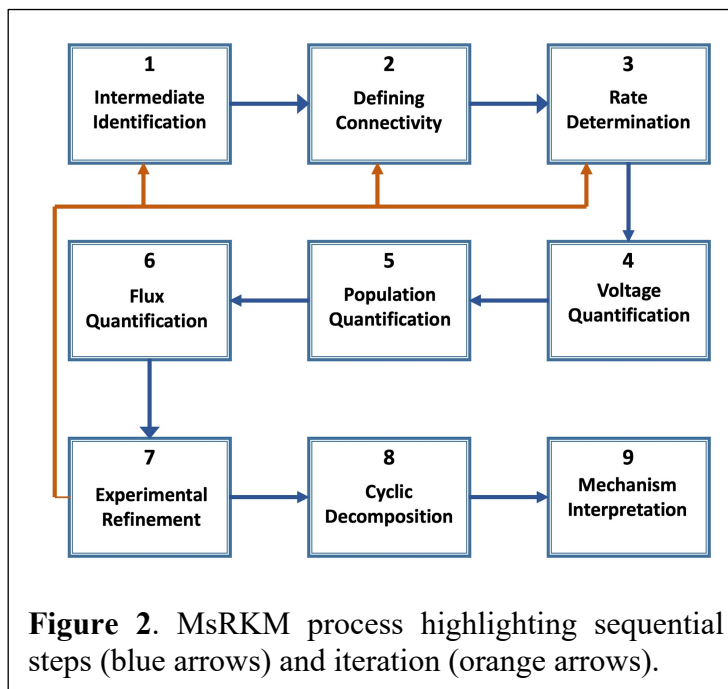
mechanisms have been proposed for Shaker that vary in ion occupancy, the order of uptake and release,<sup>41-44</sup> and the presence or absence of water in the selectivity filter.<sup>45</sup> However, mechanistic cycles proposed for the open-conformation of Shaker are primarily represented by single-pathway dominant mechanism regardless of the experimental conditions tested, though there is debate on the number of cycles present for homologous proteins.<sup>43, 46</sup> Additionally, two interesting studies using molecular dynamics simulations run with applied voltages and combined with ion binding graph representations showed many contributing pathways and a mechanistic shift between those observed under chemical gradients and voltages.<sup>47, 48</sup> However, the applied voltages were far larger than biologically or experimentally relevant magnitudes, which perhaps explains why the findings have not been followed up on. Collectively, these facets make Shaker ideal for testing ground for MsRKM—particularly testing if a full network kinetic characterization will retain the presumed mechanistic insensitivity to electrochemical conditions and if it supports a mechanism dominated by a single mechanistic pathway or if multiple pathways contribute.

For CIC-ec1 we find that a combined bottom-up/top-down approach integrating sufficient simulation and experimental data in MsRKM produces robust kinetic models that are predictive of experimental behavior across a range of electrochemical conditions. For Shaker, MsRKM identifies robust solutions using a purely top-down approach integrating increasing types of experimental data and biophysical insight. Oversimplification of the involved intermediates proves incapable of reproducing all experimental trends, demonstrating how the method can help refine flawed network descriptions. Interestingly, through continued solution refinement MsRKM iteratively rules out many multi-pathway solutions to converge on a single-pathway dominant mechanism that does not vary significantly under electrochemical conditions. Yet even with this

mechanism, off pathway flux and a network description of the Shaker channel proves essential for replicating experimental observations.

## 2. THEORY AND METHODS

The essential ingredients for any kinetic model are the relevant intermediate states and rates of transitions between those connected states. Framed in network language, the states comprise the nodes and the rates define the edges (Figure 1). The reactive flux through the network is determined by the populations (normalized probability) of the nodes



**Figure 2.** MsRKM process highlighting sequential steps (blue arrows) and iteration (orange arrows).

and the weights of the edges (transition rates). The relative flux through different reaction pathways can be considered the percentage of single molecules that carry out the process of interest by traveling through a particular series of intermediates (a particular pathway). Once defined, such a kinetic model maps the flux of population changes, such as ion flow in a channel, under any set of conditions, including equilibrium, steady-state, and non-equilibrium relaxation from a given set of starting populations. The MKM procedure can be broken into 9 steps, which will be successively described below in each section, but iteration through these steps is frequently essential (Figure 2).

### 2.1. Intermediate Identification (State Space):

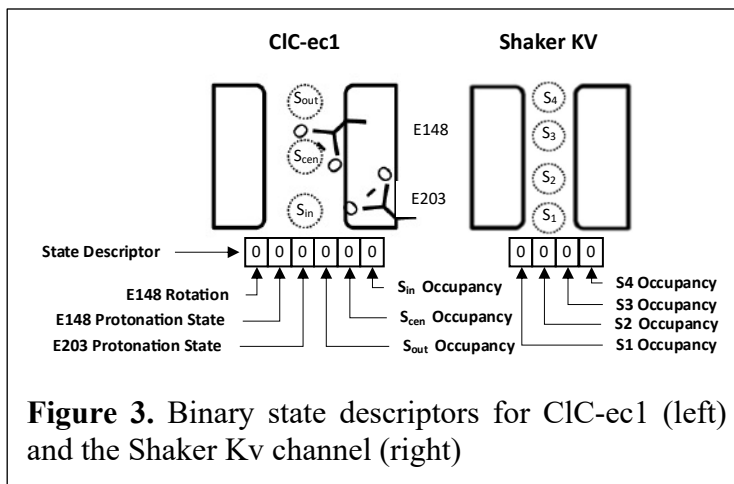


In principle, the selected state space should include all metastable intermediates that are separated from others by rate-influencing barriers, and which contribute above a given threshold value to the ensemble flux for the range of conditions of interest.

This can be a challenging aspect of

kinetic modeling since identifying all metastable intermediates requires extensive system characterization—often through a combination of experiments and simulations.<sup>49</sup> If correctly identified, a kinetic network constructed from these intermediates and their transitions can describe protein behavior under a wide range of experimental conditions. Even when the intermediate state space is not precisely resolved, relevant trends and insights can still be obtained, ideally by comparing several kinetic networks with different state space definitions as demonstrated for the Shaker channel below.

In the models described herein, intermediate states are assigned a binary descriptor (0/1) that is helpful for tracking each feature that can change, such as ion/ligand binding (empty/bound), conformational changes (conformation 1/conformation 2), or chemical state (reactant/product). The number of unique states can of course be increased for features with more than two possible definitions, such as the three competing conformations in a canonical alternating access mechanism (inward-facing, outward-facing, and occluded)<sup>50</sup>. For example, an ion channel with three ion binding sites would be defined as 000 when empty and 111 when bound by three ions. The binary can be reduced to a single number (e.g., 0 for 000 and 7 for 111) for an abbreviated



state definition. Retaining the binary description, this yields  $N = 2^{\# \text{descriptors}}$  total states, although some may not be relevant (as demonstrated for ClC-ec1 below). Though not defined in the binary descriptor, MsRKM also allows competing orientations of the protein embedded in a membrane to replicate data from liposomal assays lacking protein orientation. The model assumes a 1:1 ratio of biological to opposite orientations for simplicity. This ratio can be modified for proteins that exhibit preferential membrane orientation or removed altogether for oriented systems such as Shaker.

In this study, the intermediates of ClC-ec1 are defined by three  $\text{Cl}^-$  binding sites ( $S_{\text{out}}$ ,  $S_{\text{cen}}$ , and  $S_{\text{in}}$ ), two  $\text{H}^+$  loading sites (the lower E203 and upper E148 residues), and the rotation orientation of E148 (up/down) (Figure 3) as detailed in previous studies.<sup>6, 9, 33, 34</sup> The Shaker channel is modeled with either three or four  $\text{K}^+$  binding sites (Figure 3) located at the S6 ‘gate’ region ( $S_1$ )<sup>39</sup> and within the selectivity filter at T441( $S_2$ ), G443( $S_3$ ), and G445( $S_4$ )<sup>38, 51</sup>. All intermediates are included in the Shaker model, but only 48 of the 64 total intermediates are physiologically relevant in ClC-ec1 model since the  $S_{\text{cen}}$  site cannot be bound by  $\text{Cl}^-$  when E148 is in the ‘down’ conformation<sup>34</sup>.

It is also possible to simplify a reaction network by grouping states together that exchange more quickly than a user-defined frequency. This can help alleviate numerical challenges and reveal important pathways. Considerable effort has been put into developing robust methods to accomplish this type of network reduction. Since they are not explored herein, the interested reader is directed to spectral clustering<sup>52, 53</sup>, model reduction<sup>54</sup>, and graph theory<sup>55</sup> for examples of commonly used network reduction approaches.

## 2.2. Defining Connectivity (Transitions)

The reaction network is bidirectional, such that each rare event transition connecting two states/nodes has a forward and reverse directional edge and corresponding forward and reverse rate coefficient. For a fully connected network, this results in  $N*(N-1)$  edges and rate coefficients. However, for most systems only certain states will be connected by physically meaningful transitions that contribute to ensemble flux. For example, an ion cannot tunnel through a binding site from Sout to Sint (Figure 3), and transitions involving simultaneous rare-event crossings (e.g., proton uptake to E203 and  $Cl^-$  release from Sout) are excluded since the lack of coupling between these processes makes the probability of their simultaneous occurrence negligibly small.<sup>34</sup> In cases where coupling introduces a unique rate for simultaneous transitions, these rates should be included.

For ClC-ec1, this reduces  $48 \times (48-1)$  (2256) transitions to 248 transitions. Additionally, some transitions may have degenerate rate coefficients. For example,  $Cl^-$  uptake to Sout is assumed to be independent of the protonation state of E203 in the presented model of ClC-ec1. This further simplifies the network to 68 unique rate coefficients that describe ClC-ec1. The Shaker model assumes the open conformation of the protein and has no such conformational constraints within the region of the selectivity filter.<sup>56</sup> However, ions still cannot tunnel (e.g., from S1 to S3) and the current model assumes transitions are not coupled. Future work will explore the role of coupled transitions in Shaker. This reduces the number of transitions from 240 ( $16*(16-1)$  transitions from 16 ( $2^4$ ) states) to 56 transition rates that were modeled explicitly with no degeneracy assumed between the transitions.

### **2.3. Rate Determination in the Absence of Voltage**

Rate coefficients can occasionally be directly measured in experiment<sup>57</sup>, but are most commonly obtained through either bottom-up in silico quantification or top-down refinement based on

experimental data. There is a wide array of approaches available to calculate rates from simulations, ranging from quantification of the transition free energy profiles, often called potentials of mean force (PMFs), combined with transition state theory, to the direct tracking of transition probabilities in, for example, Markov state modeling. Often, these methods are accompanied by methods to identify the proper reaction coordinates and to improve the statistical convergence via adaptive or enhanced free energy sampling<sup>58</sup>. The choice of the method depends on the nature of the transition. For example, when chemical reactions are involved, this step can require multiscale treatment, e.g., via quantum mechanics/molecular mechanics (QM/MM)<sup>59</sup> simulations or reactive MD, such as the multiscale reactive molecular dynamics (MS-RMD),<sup>60</sup> particularly developed for long-range proton solvation and dynamics. When highly interfacial or ionic systems are involved, polarization effects can be important, requiring polarizable MD. Taking CIC-ec1 as an example, multiple computational methods were used to calculate rates for Cl<sup>-</sup>/H<sup>+</sup> exchange. MS-RMD was used to characterize the PMF of proton transport (PT) through the central region between E148 and E203,<sup>34</sup> as well as between E148 and bulk solution<sup>61</sup>. In both cases, the influence of Cl<sup>-</sup> occupancy at the S<sub>cen</sub> binding site was explicitly accounted for. All-atom MD with a DRUDE polarizable force field<sup>62, 63</sup> was used to determine multidimensional PMFs for Cl<sup>-</sup> transport to capture the influence of ion-ion repulsion.<sup>6</sup> Once the PMF has been characterized, transition state theory<sup>64</sup> can then be used to convert the free energy barrier between metastable states and the dynamics of the system in the reactant well into the rate coefficient through the Eyring-Polanyi equation<sup>65</sup>:

$$k_{rxn} = \frac{\omega_0}{2\pi} e^{\frac{-\Delta G^\ddagger}{k_B T}} \quad (1)$$

where  $k_B$  is Boltzmann's constant,  $T$  is temperature, and  $\Delta G^\ddagger$  is the free energy barrier height. The fundamental frequency  $\omega_0$  is specific to the transition and quantifies the frequency with which the

reactant attempts to cross the transition barrier. It can be approximated from the second derivative of the free energy profile in the reactant well:

$$\omega_0 = \sqrt{\frac{\partial^2 G(r)/\partial r^2}{m_{eff}}}\bigg|_{r=r_0} \quad (2)$$

Here  $\frac{\partial^2 G(r)}{\partial r^2}$  is the second derivative of the free energy  $G(r)$  with respect to the reaction coordinate  $r$ , evaluated at the local minimum in the PMF ( $r_0$ ). The effective mass ( $m_{eff}$ ) can be calculated using the thermal velocity of an ion,  $m_{eff}\langle v^2 \rangle = k_B T/2$ , where  $\langle v^2 \rangle$  is calculated at  $r_0$ . While the thermal velocity is not directly dependent on the reaction coordinate, the effective behavior of the ion along the reaction coordinate can vary due to variations in the local environment and potential energy surface. This is particularly important for transitions involving high-energy barriers, where the local thermal environment can influence the rate and mechanism of the reaction.<sup>66</sup>

Finally, a few rate coefficient values were estimated from alternative methods. BD simulations were performed to estimate  $\text{Cl}^-$  uptake/release rates to/from  $S_{in}$  and  $S_{out}$ .<sup>6</sup> E148 and E203 protonation and deprotonation rates when  $S_{out}$  or  $S_{int}$  is occupied<sup>6</sup> were estimated using the relative change in  $\text{pK}_a$  calculated from PropKa.<sup>67, 68</sup>

Regardless of how they are obtained, these initial rate coefficients act as a crucial ‘seeds’, helping to identify the correct region of parameter space while the error in their estimation can be used to set bounds for parameter optimization. Since biological systems operate in thermodynamic cycles, an additional thermodynamic constraint on the rate coefficients can help ensure good seed values. At chemical equilibrium, the rate constants for a thermodynamic cycle have the following relationship:

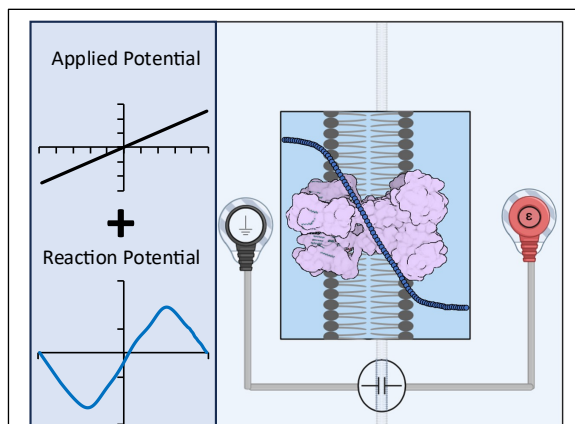
$$\mathbf{1} = \frac{\prod k_{for}}{\prod k_{rev}}, \quad (3)$$

where  $k_{\text{for}}$  are all rate constants taken in one direction around the cycle and  $k_{\text{rev}}$  are taken in the opposite direction. Since rate constants do not change as concentrations shift away from equilibrium, this relationship remains valid away from equilibrium as well, assuming no voltage, pH, or conformational changes occur. Using this relationship to precondition values for rate coefficients helps to provide a variety of optimization initiation points within the error of the rate coefficients while guaranteeing that none stray from thermodynamic feasibility.

## 2.4 Voltage Quantification

One important consequence of kinetic selection can be the shifting of mechanistic

pathways due to reaction conditions, as demonstrated by molecular motors under high/low ATP/ADP concentrations.<sup>17, 18</sup> A few candidate proteins<sup>20, 49</sup> have demonstrated that this also happens outside of the realm of chemomechanical coupling. Most notable for the purposes of this paper, ClC-ec1 has shown that different pH<sup>9</sup> and Cl<sup>-</sup> gradients<sup>6</sup> generate significant mechanistic pathway changes. However, this work did not explicitly factor in the influence of voltage. The transmembrane potential includes both chemical and electrical gradients and the latter must be included to understand the role of kinetic selection in ion channels and transporters. With MsRKM, we seek to capture the influence of the transmembrane electrochemical gradient on the rates of ion transport in a system- and voltage-specific manner. Significant work has gone into developing and



**Figure 4.** Membrane voltage captured by treating simulation box as a subsystem surrounded by a bath with ion exchange electrodes (right) inducing an applied potential (upper left). The system responds by rearranging charged species to create its own reaction potential (lower left). The summation of the applied and reaction potential localizes the total potential drop to the membrane region (right), mimicking the real membrane voltage created by a charge imbalance localized at the membrane interface.

validating computational methods to quantify the influence of voltage on substrate transport.<sup>69-72</sup>

We focus herein on methods that combine the calculation of free energy profiles for ion permeation with the dimensionless coupling profile that captures the influence of voltage as a function of the permeation reaction coordinate (often simply the z-axis) across a given channel or transporter of interest.

#### **2.4.1 The constant electric field method**

The transmembrane voltage, or electric potential, originates from an asymmetric distribution of ions that is actively maintained by ion pumps and transporters. This charge imbalance creates an electric field that pulls complementary ions to the membrane-solution interface, localizing the voltage drop to the transmembrane and interfacial region. Unfortunately, capturing a charge imbalance in MD simulations that employ periodic boundary conditions (PBC) isn't possible since the solution on either side of the membrane is continuous.

A commonly implemented solution to this problem is to impose a uniform electric field,  $E$ , perpendicular to the membrane plane (conventionally through the z-axis) creating a voltage difference over the length of the simulation.<sup>73-76</sup> The theoretical foundation of this approach has been well described<sup>77</sup> and validated<sup>76</sup>. In essence, it mimics experimental ion exchange electrodes holding the solutions on either side of the membrane at different voltages via an electromotive force. The system is conceptually broken down into bath surrounding a subsystem, where the electromotive force in the bath creates a linear “applied potential” and thus constant electric field across the subsystem close to the membrane (i.e., the simulation box). The response of the system, redistributing ions to either side of the membrane, creates the “reaction potential”. The resulting total potential across the box, calculated as a sum of the reaction and applied potentials, concentrates the change in potential across the membrane<sup>76, 77</sup> (Figure 4).

## 2.4.2 System-specific voltage

The dimensionless coupling factor  $\varphi(z)$ , also called the fraction of the membrane potential, couples the charge movement of the region close to the membrane and the transmembrane voltage applied by the electromotive force in the bulk solution.  $\varphi(z)$  can be determined from multiple methods; some using explicit solvent<sup>77</sup> and some using the linearized PB-V continuum approximation<sup>78</sup>. The method used in this paper, the linear coupling method<sup>79</sup>, calculates  $\varphi(z)$  by analyzing the ensemble averaged electrostatic potential with PMEpot<sup>75</sup> module in VMD<sup>80</sup>. This yields a 3-D electrostatic potential map ( $\Delta\phi_{elec}(x, y, z)$ ) which is then integrated over the x and y dimensions to yield  $\Delta\phi_{elec}(z)$ .  $\varphi(z)$  can then be calculated as the difference between two  $\Delta\phi_{elec}(z)$  values:

$$\varphi(z) = \frac{\Delta\phi_{elec}^{-V}(z) - \Delta\phi_{elec}^{+V}(z)}{V_T} \quad (4)$$

where  $\Delta\phi_{elec}^{-V}(z)$  and  $\Delta\phi_{elec}^{+V}(z)$  are the ensemble averaged electrostatic potentials for simulations run with a negative and positive applied voltage of equal magnitude;  $V_T$  normalizes the expression<sup>79</sup>. This approximation provides a useful description of the voltage profile along the z-axis of a simulated protein and captures the shift in an ion permeation free energy profile for model systems surprisingly well<sup>76, 79</sup>. However, it cannot describe significant voltage-induced conformational changes for which other methods are more appropriate<sup>77-79</sup>. Alternative approaches to calculating  $\varphi(z)$  and including voltage-induced conformational changes in the rate response will be the focus of future work. For this work, we limit our analysis to conditions under which we do not expect large voltage-induced conformational changes, or equivalently optimize unique solution sets for different conformations.

## 2.4.3 Obtaining a voltage profile via simulation

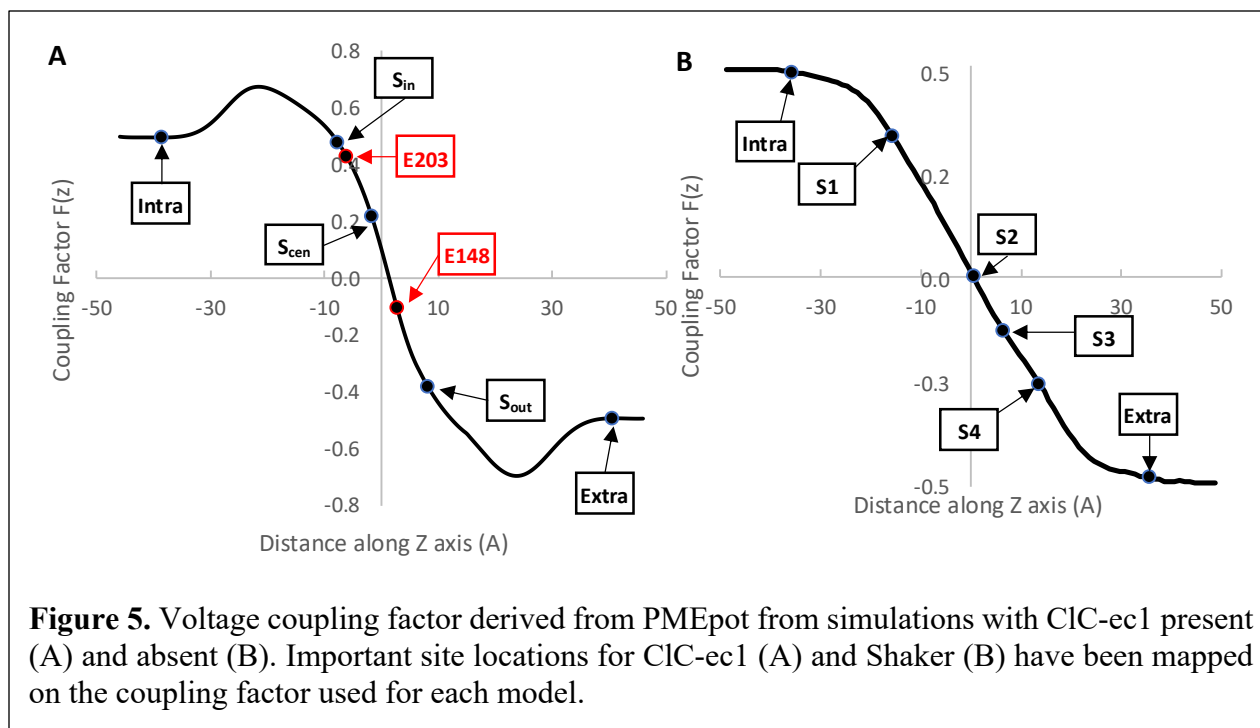


Two simulation systems were used to determine  $\phi(z)$ : one that consists of the ClC-ec1 dimer (PBD:1OTS)<sup>25</sup> with the internally bound Cl<sup>-</sup> ions removed from each monomer saturated with 163 POPE lipids and another of a protein-less membrane bilayer comprised of 231 POPE lipids used to represent a generic membrane-bound protein in the simple channel system. Both systems were saturated with 300mM NaCl and solvated with TIP3P<sup>81</sup> water in a 92Å x 92Å x 100Å box with periodic boundary conditions. The CHARMM36<sup>82</sup> forcefield was used for protein, and ions, while CHARMM36m<sup>83</sup> was used for the phospholipid membrane. The protonation states of all residues were determined based on previous pKa calculations<sup>84</sup> on the same crystal structure (E148, E113, and D417 were protonated), while standard protonation states were chosen for all other residues. The influence of pH on ClC-ec1 titration states and the complementary kinetic descriptions will be the focus of other work.

Minimization and equilibration were carried out with NAMD2.14<sup>85</sup> using a seven-step protocol suggested by CHARMM-GUI<sup>86-88</sup>. Both systems were minimized for 10,000 steps of conjugate gradient energy minimization with restraints on the protein's heavy atoms (10 kcal mol<sup>-1</sup> Å<sup>-2</sup> harmonic positional restraint) without SHAKE<sup>89</sup>. A six-step equilibration was then carried out, gradually removing position restraints on the lipid, protein backbone atoms, and ions. To prevent the spontaneous uptake of Cl<sup>-</sup>, additional wall potentials were placed at each side of a 35 Å x 30 Å rectangular box located at the channel mouth on each monomer, which was large enough to cover the entire area of the pore. The SHAKE algorithm<sup>89</sup> was introduced to constrain hydrogens in the 2<sup>nd</sup> equilibration step. The first two steps used the NVT ensemble, with the last four switching to the NPT ensemble. The first three steps were 125 ps, each with a timestep of 1 fs. The final three steps, each lasting for 0.5 ns with a timestep of 2 fs, decreased the force constant for backbone restraints from 1.0 kcal mol<sup>-1</sup> Å<sup>-2</sup> to 0.5 kcal mol<sup>-1</sup> Å<sup>-2</sup> to no positional restraints. The

temperature was maintained at 300 K using a Langevin thermostat with a relaxation time constant of 1 ps in the NVT ensemble. Anisotropic pressure scaling with a Langevin piston barostat was applied to maintain a constant pressure in the NPT simulations, with a damping coefficient of 1 ps. Long-range electrostatics were evaluated using the particle-mesh Ewald method<sup>90</sup>. The cutoff of the van der Waals and short-range electrostatic interactions was set to 12 Å which was smoothed with switching functions.

Each system (CIC-ec1 and the membrane only) next underwent two separate voltage equilibrations using a constant electric field (equal to a +500 mV and –500 mV voltage drop across the membrane) applied perpendicular to the membrane along the z-axis. The voltage equilibrations employed the same NVT ensemble as described above. Harmonic restraints with a spring constant of 2 kcal mol<sup>-1</sup> Å<sup>-2</sup> were applied to protein backbone atoms to prevent large conformational changes. Each system underwent 10 ns of simulation, with the last 2 ns used for analysis. Snapshots taken every 2 ps were analyzed in the VMD<sup>80</sup> PMEpot plugin<sup>75</sup> to compute  $\phi_{elec}(x, y, z)$ , and  $\varphi(z)$  was calculated using equation 4 (See figure 5).



#### 2.4.4 Voltage-responsive rates (Derivation)

An alternative way to derive a dimensionless voltage coupling profile,  $\varphi(z)$ , is to take the difference between ion translocation PMFs run at different voltages. This so-called *W-route*, which has been detailed in previous studies<sup>77, 78</sup>, can be used to describe an ion transport PMF under the influence of a transmembrane potential as:

$$G(z, V) = \varphi(z)qV + G(z, 0) \quad (5)$$

Rates of rare-event transitions are defined by the difference in free energy between meta-stable intermediates and transition barriers ( $\Delta G^\ddagger$ ). Using equation 5, we can estimate the voltage-induced changes in  $\Delta G^\ddagger$  using the relationship:

$$\Delta G_{ij}^\ddagger(z, V) = [\varphi_{ij}(z^t)qV + G_{ij}(z^t, 0)] - [\varphi_{ij}(z^r)qV + G_{ij}(z^r, 0)] \quad (6)$$

where  $G_{z;0}^t$  and  $G_{z;0}^r$  represent the free energy at the transition barrier and reactant basin, respectively. Placing equation 6 into equation 1, we can derive a relationship between the rate coefficients under different transmembrane potentials:

$$k_{ij}(V) = k_{ij}^0 e^{\frac{-(\phi_{ij}(z^t) - \phi_{ij}(z^r))qV}{k_b T}} \quad (7)$$

where  $k_{ij}^V$  and  $k_{ij}^0$  represent the rate coefficients for a transition from state  $i$  to  $j$  with and without the influence of voltage and  $k_b$  is in units of eV K<sup>-1</sup>.

## 2.5. Population Quantification

Once states and transitions are defined and rate constants are obtained, the time-dependent state populations can be quantified with a kinetic master equation:

$$\frac{dP_i}{dt} = -\sum_{i \neq j} P_i k_{ij} + \sum_{j \neq i} P_j k_{ji} , \quad (8)$$

where  $P_i$  is the population of state  $i$ , and the summation includes all other states  $j$  connected via a transition to state  $i$ . This set of linear differential equations can be cast in matrix form:

$$\frac{dP}{dt} = -\mathbf{K}P , \quad (9)$$

where  $\mathbf{K}$ , the rate matrix, has negative off-diagonal components  $\mathbf{K}_{ji} = -k_{ij}$  and positive diagonal components  $\mathbf{K}_{ii} = \sum_{j \neq i} k_{ji}$  for all states  $j$  connected to state  $i$  and  $\mathbf{K}_{ji} = 0$  otherwise.  $\mathbf{K}$  describes the physical flow or motion of ions and molecules in continuous time.  $\mathbf{K}$  is directly related to  $\mathbf{T}$ , the transition matrix used in Markov State Models (MSM).  $\mathbf{T}$  describes the abstract flow of probability between states in discrete time and is the matrix exponent of  $\mathbf{K}$ :

$$\mathbf{T} = e^{-\mathbf{K}\tau} , \quad (10)$$

where  $\tau$  is the timestep size used in the MSM.

By setting  $dP/dt = 0$ , the system can be solved for the equilibrium population of states. This equilibrium solution is required to be microscopically reversible with detailed balance such that  $P_i k_{ij} = P_j k_{ji}$ . In biological systems, where ion/substrate concentrations are often maintained physiologically, the system may not reach equilibrium due to constant influx to and efflux from the local environment surrounding the protein of interest. In this situation, the system may instead approach a non-equilibrium steady state with constant populations and flux:

$$\mathbf{0} = -\mathbf{K}P_{SS}, \quad (11)$$

where  $\mathbf{0}$  is a zero column vector with rows equal to the rank of the rate matrix and  $P_{SS}$  is the column vector containing the steady state populations. Steady-state solutions do not obey detailed balance. Adding these sources and sinks into the system of equations reduces the number of independent equations, necessitating the replacement of one state's rate equation with a normalization condition, and the equivalent row in the zero vector with the desired norm. The steady-state solution corresponds to the solution of an eigenvalue-eigenvector problem with an eigenvalue of zero, the corresponding normalized eigenvector providing the steady-state populations. Other eigenvalues-eigenvector pairs correspond to the dynamics of the system as it relaxes to equilibrium or steady-state, with large eigenvalues corresponding to slower relaxation modes.

The kinetic master equation also permits general time-dependent solution, allowing the evolution of the populations in time to be observed. If the environmental conditions are permitted to vary, these solutions show populations evolving toward equilibrium. If they are held constant, the populations will instead evolve toward a non-equilibrium steady state.

## 2.6 Flux Quantification

The net number of ions transferred through any transition is computed as:

$$J_{ij} = P_i k_{ij} - P_j k_{ji} \quad (12)$$

Where  $J_{ij}$  represents the flow in  $\frac{\text{ions}}{\text{ms}}$  going from state  $i$  to  $j$ . The net flow going through the protein is determined by summing all transitions to/from the inner-most or outer-most ion binding sites as:

$$J_{net} = \sum \alpha J_{binding} - \alpha J_{release} \quad (13)$$

where  $\alpha$  defines the directionality of the transitions as either outflux ( $\alpha = -1$ ) or influx ( $\alpha = 1$ ). For example, when analyzing  $\text{Cl}^-$  influx of ClC-ec1, equation 13 becomes:

$$J_{in} = \sum J_{release}^{Sint} - J_{binding}^{Sint} \quad (14)$$

where we have defined negative net flux to be ions flowing out of the cell and positive net flux as ions flowing into the cell.

## 2.7 Experimental Refinement

### 2.7.1 Limitations of bottom-up kinetic modeling

Although the rigor and accuracy of simulation-based thermodynamic and kinetic analyses are impressive and rapidly improving, systematic<sup>91</sup> and statistical<sup>92</sup> error is inevitable. In principle, obtaining a well-behaved kinetic model from bottom up is possible and has been demonstrated for select systems<sup>28</sup>. Frequently, however, compounding errors<sup>93-96</sup> will quickly throw off a kinetic solution. The same challenge has been observed in microkinetic modeling, an arguably more advanced field focused on modeling chemical reactions and heterogeneous catalysis<sup>97, 98</sup>. To remedy this, a top-down quantification of the kinetic parameters by optimization with experimental results is performed with optimization boundaries set to the calculated error of MD simulations.

### 2.7.2 Integrating experimental data

The solution space of kinetic networks tends to be vast, with the number of unknowns (rate coefficients) typically outnumbering the number of known data points with which to refine them. The ability to refine a solution thus hinges on two things: 1) good initial rate coefficients (seeds) with well-defined boundary conditions, and 2) the amount and quality of data that can be used as optimization constraints. High quality experimental data is not only that with small error bars, but also data that uniquely constrains the optimization. For example, five sets of IV curves under similar conditions will not refine the solution as well as three IV curves under highly different conditions and two liposomal flux assays. In general, adding additional experimental data will further reduce the solution space and improve the solution.<sup>49</sup> An overdetermined system tends not to be a problem when extracting generalizable trends if the results are consistent. In such a case, some equations running within the optimization may end up superfluous, resulting in poor optimization performance. If the overdetermined model produces inconsistencies or a lack of solution, overfitting can be handled with regression techniques to reduce the total number of constraints or equations.

However, it is important to note that it is possible to get optimization solutions that show different protein behaviors or a lack of optimization solutions regardless of the system type. In our experience, this has often been due to some experimental inconsistencies or clashing physical realities with other experimental results placed in the optimization. For example, the presence of an undetected ion leak could lead to an I-V curve that appears similar in profile to other I-V curves but has designated a reversal potential where an efflux assay, under identical experimental conditions, has non-zero flux. Difficulty finding a solution may also be due to assumptions made in the model-building process, especially if one is reducing to the minimum possible states to describe a network<sup>99</sup>.

### 2.7.3 Loss function design

The choice of optimization methodology is critical to the success of the optimization. To allow different methodologies in the fitting of experimental data, SpotPy<sup>100</sup> global, SciPy<sup>101</sup> global and local and user-defined optimization algorithms can be used. Though not the focus of this study, extensive work has been done to examine the best optimization algorithms for biological-based networks.<sup>102-104</sup> The selected optimization method changes the intrinsic rate coefficient values while minimizing a least squares loss function incorporating experimental data. Each experimental data point reflects different pH and ion concentrations for the intra- and extracellular sides of the membrane and different values of  $\Delta\Psi$  that produce an experimental observable to be compared to. We currently support several experimental observables in the loss function, such as liposomal flux assays, I-V curves, and orientation-based blockage studies. Optimizations based upon flux through the liposomal system also allow for the optimization of net flux through different orientations separately, i.e., net flux through the biological orientation only. The optimization also allows for flux component differentiation to be accounted for within the loss function, such as differential orientation influx/outflux, total influx (defined as all flow moving inwards from both orientations), total outflux (defined as all flow moving outwards), and the ability to turn the microscopic reversibility term on or off for functionally irreversible systems. The model also allows for the user to optimize to normalized flux values such as normalized IV curves, normalized IV curves with differential orientation preferences, or normalized flow assays. Microscopic reversibility has been represented as a mean squared error function normalized over the total number of transitions within the loss function:

$$Objective = \frac{\sum_{ij} \left( \frac{J_{ij^-} - J_{ji^+}}{\langle J_{ij^+ji^-} \rangle} \right)^2}{\#_{ij}} + \sum_{sys} W (J_{calc} - J_{def})^2 \quad (15)$$



where microscopic reversibility is included in the first term of the objective function, the second term calculates the least squares difference between the flow values calculated through MsRKM and the user-defined values for every experimental system input. MsRKM also allows for the weighting of different experimental values within the loss function dependent upon the experimental type represented by  $W$  in the equation above.

## 2.8 Cyclic Pathway Decomposition

Once a kinetic solution is obtained, the relative contribution of different pathways must be quantified. This is a nontrivial task as the flux into and out of any given node can contribute to multiple pathways. Although it is more straightforward to identify the dominant path by tracking the largest minimal flux transition,<sup>105</sup> this does not quantify the relative contributions of all pathways. To accomplish this goal, we previously developed CycFlowDec<sup>106</sup>, an algorithm that quantifies the expected flux through simple cycles (each containing any given node no more than once) in a closed network. Based on established methods in Markov circulation theory,<sup>107, 108</sup> the algorithm in its simplest form decomposes the total flux into relative cycle contributions based on random walks generated from Markov transition probabilities. It is then made efficient for complex networks by using a percolation with burn-in and minimum contribution tolerance. Herein we implemented CycFlowDec<sup>106</sup> for CIC-ec1 and Shaker with a 100,000-step percolating algorithm with a burn-in of 4500 steps, a flow tolerance of  $1 \times 10^{-7} \frac{\text{ions}}{\text{ms}}$ , and an MRE cutoff of  $9 \times 10^{-3}$ .





## 2.9 Mechanism Interpretation

The above analyses provide a full network description of the system's flux through various intermediates, which can be tracked under the equilibrium, steady state, or non-equilibrium conditions for which the rates remain applicable. Thus, the mechanism can be compared under

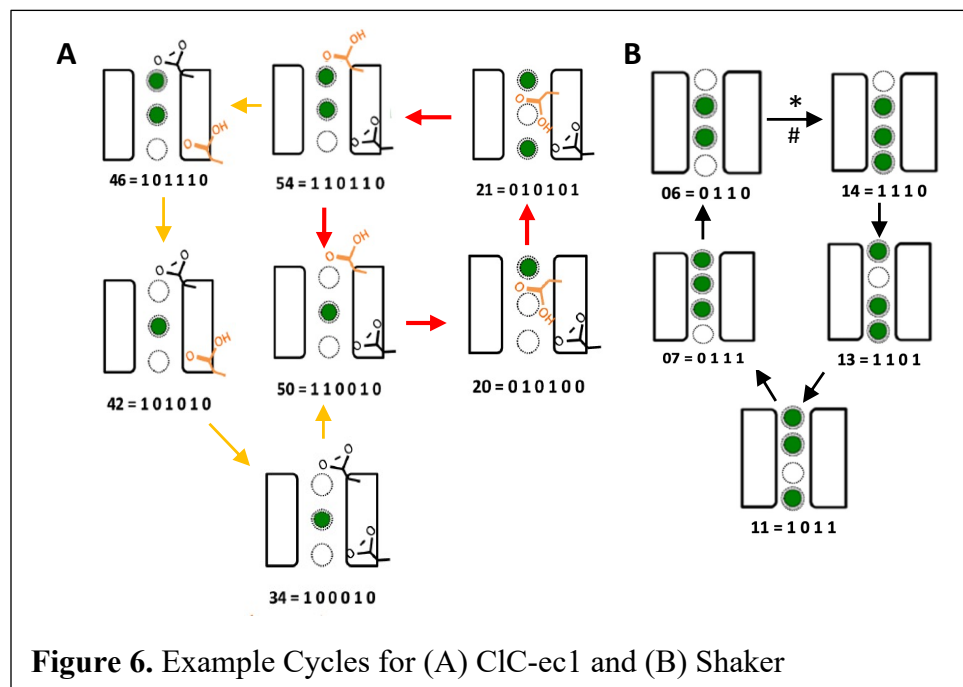
varying reaction conditions and the role of competing pathways can be delineated. Understanding the mechanism at the molecular level, and hence the structure-dynamics-function relationships that control the process efficiency and selection between competing pathways, is of course benefitted by simulations probing the transition ensembles. Simulations are also best suited to probe the impacts of point mutations on rate-influencing transitions. In addition, MsRKM can be used to probe and compare the network descriptions that differentiate wild type from mutant behavior, as demonstrated for Shaker below.

### 2.9.1 Pathway diagram visualization

Network flux can be tracked graphically as a function of time, or via network diagrams for equilibrium or steady-state conditions. For the latter, all significant pathways of a solution set can be plotted in a corresponding full network diagram or a subset of dominant or related pathways can be plotted in a pathway diagram. In this work, each diagram uses a series of cartoon representations of either ClC-ec1 or the Shaker channel states connected by arrows that represent transitions. The depictions include (Figure 6):

1. the non-interacting portion of the protein represented by two empty rectangles,
2. three dotted circles representing the ion binding sites that are filled ● when occupied or empty ○ when unoccupied,
3. H<sup>+</sup> binding sites for ClC-ec1, E148 and E203 either protonated  or deprotonated ,
4. E148 rotating “up”  towards the extracellular side or “down”  towards the central Cl<sup>-</sup> binding site (S<sub>cen</sub>) of ClC-ec1,

5. the ‘Biological’ orientation of the protein with the extracellular solution on top and intracellular solution on bottom and vice versa for the “Opposite” orientation,
6. Flux Limiting Steps (FLS) of the cycle denoted with \* and Rate Limiting Steps (RLS) denoted with #.

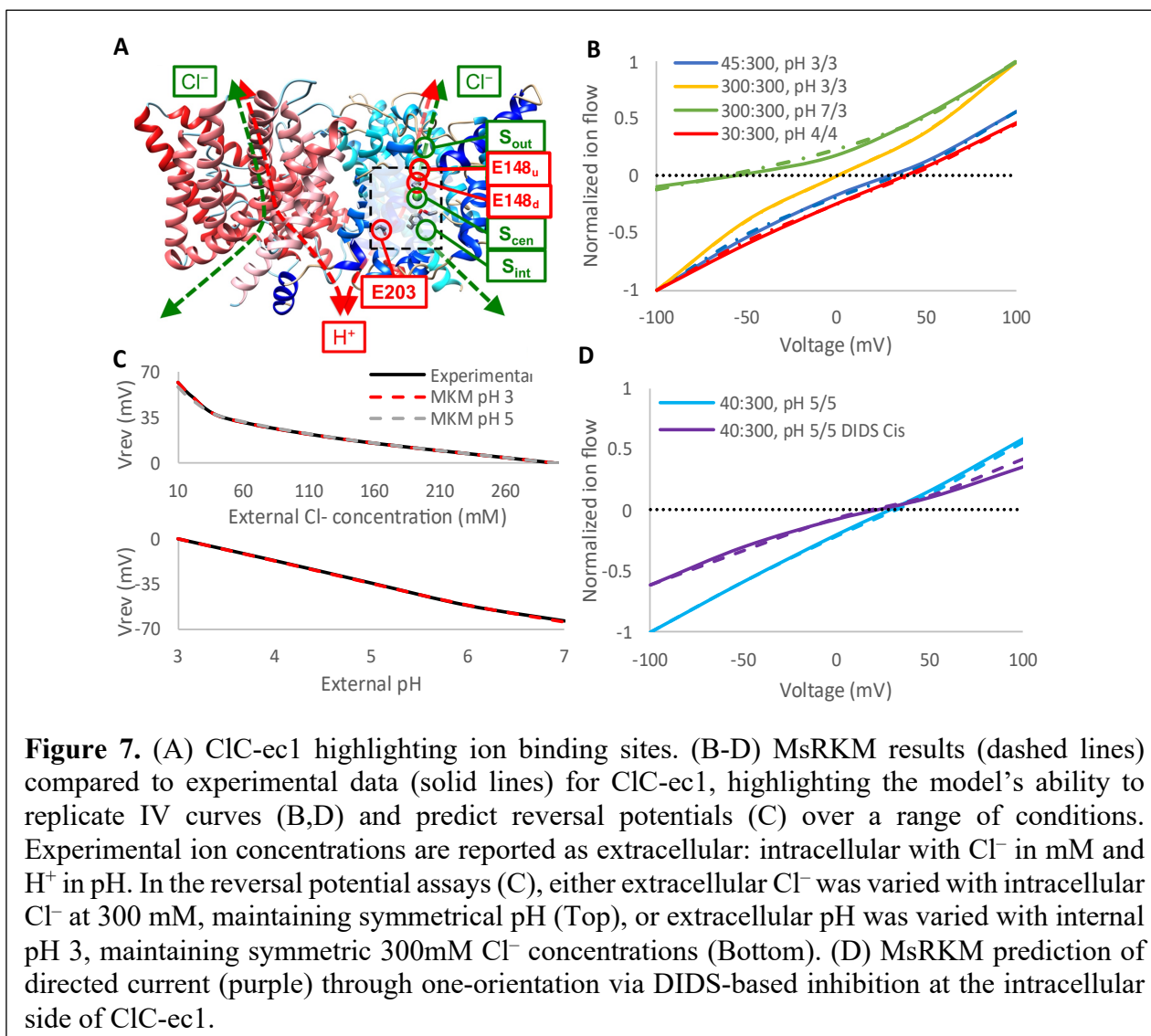


### 3. RESULTS AND DISCUSSION

#### 3.1 Model Robustness and Predictions

A central challenge in kinetic modeling is narrowing the solution space. Typically, the problem is underdetermined with far more parameters (rate coefficients) than known data to fit. Thus, kinetic modeling hinges on the quantity and quality of data available. For systems with a range of experimental results (often top-down macroscopic data, but sometimes rates for specific transitions) and simulation-derived kinetic parameters (bottom-up molecular data), MsRKM can

display robust predictive power—not only narrowing the solution space but also demonstrating when the underlying network (state space definition) is insufficient (as demonstrated below).



CIC-ec1 as a case study demonstrates the need for experimental and simulation-based data to determine robust solutions. For this system, previous studies employing state-of-the-art multiscale and polarizable simulations quantified most of the ion transition rates with reported errors in the corresponding free energy barriers ranging from 0.8 to 2 kcal/mol<sup>6, 33, 34</sup>. The remaining rates were estimated based on pKa calculations. Despite significant effort to quantify these rates as accurately as possible, they produced unphysical results when placed into a kinetic model due to

compounding errors. Note that even small errors in free energy profiles lead to larger errors when exponentiated to estimate rates. Additionally, new work has pointed to a significant conformational change at low pH values that was not accounted for in the simulations<sup>29</sup>. During solution refinement, MsRKM accounts for these errors by adjusting rate coefficients within user-defined boundaries of their reported error to produce a kinetic model consistent with macroscopic and molecular data. These adjustments can be essential, both to meet the requirement of microscopic reversibility and to fit experimental data. Interestingly, even before the conformational change had been published, the solutions that were able to fit the data at low pH absolutely required opening up the error bounds on several rates in a manner that is consistent with the low pH structure. At the time, it was not clear why. The details of these models and their implications will be the focus of separate work. However, the process of determining these models demonstrates how integrating data from both simulations and experiment can be essential to identify robust kinetic solutions.

Starting from the simulation-based seed rates, the solution refinement for ClC-ec1 included constraints of: 1) microscopic reversibility at 0mV with pH 3:3 and Cl<sup>-</sup> 300:300; 2) five current values at pH 4:4 and a Cl<sup>-</sup> gradient of 30:300 (Figure 7B, red IV curve<sup>109</sup>) and at pH 5:5 with a Cl<sup>-</sup> gradient of 40:300 (Figure 7D, light blue IV curve)<sup>31</sup>; 3) five efflux assay values at pH 4, 4.5, and 5<sup>109</sup>; 4) five influx assay values at pH 4, 4.5, and 5<sup>109</sup>; and 5) only two specific reversal potential<sup>26</sup> values—one at symmetric pH 5:5 and a Cl<sup>-</sup> gradient of 45mM:300mM (from Figure 7C, top) and the other at pH 5:3 with symmetric Cl<sup>-</sup> concentrations 300mM:300mM (from Figure 7C, bottom). This combination of experimental data limited the solution space significantly, resulting in an MsRKM model that then successfully predicts: 1) the remaining normalized I-V curves under different conditions (Figure 7B, blue, green, and yellow IV curves)<sup>30</sup>, 2) the remaining reversal

potential values for concentration gradients evaluated at pH 3 and 5 (Figure 7C Top), and 3) pH gradients at symmetric 300mM  $\text{Cl}^-$  concentration (Figure 7C Bottom). Since these assays are different from the assays included in the objective function, this offers the first essential validation check for solution robustness.

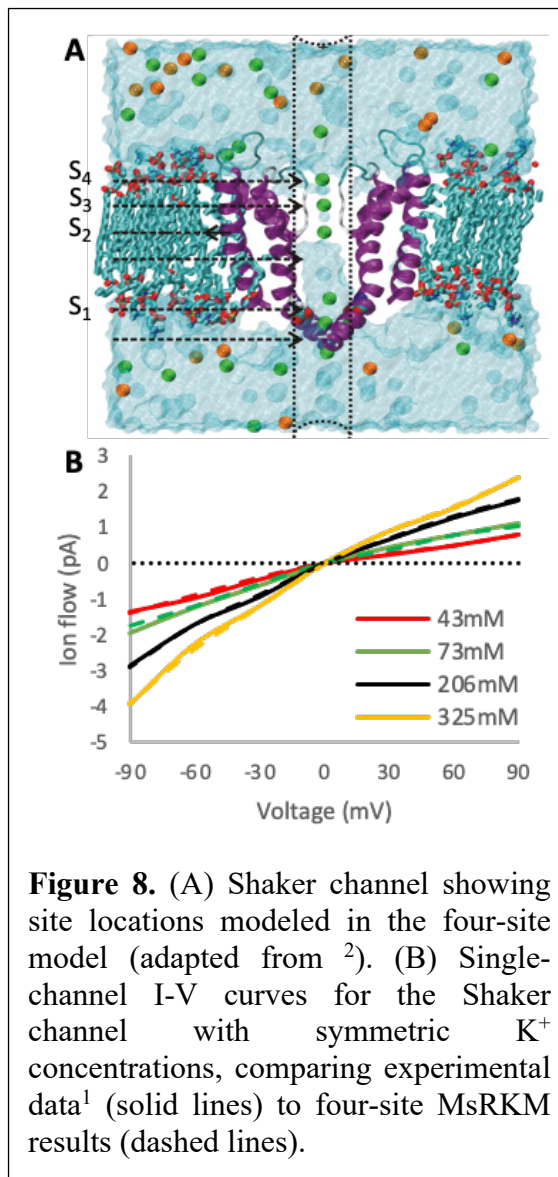
To further benchmark MsRKM solutions, it is important to test their ability to replicate experimental results for assays that are increasingly different. As a word of caution, the model should only be extended within the conditions under which the rates and rate responses are valid (see Methods). Thus, to further test the ClC-ec1 solutions we modeled a blockage assay<sup>31</sup> in which stilbenedisulfonate 4,4-diisothiocyanatostilbene-2,2'-disulfonic acid (DIDS) incompletely blocks the internal side of ClC-ec1, providing the only reported estimate of directional flux through the transporter. The opposite orientation is unaffected, so 100% of opposite flow is included, while the biological side is incompletely blocked, so 6% of the biological flow is included. With this single correction to the flows, MsRKM replicated the DIDS-blockage assay normalized I-V curves (Figure 7D). This is a striking result. Although further model testing is always warranted, the ability to fit a weighted unidirectional flow supports the model's validity and ability to predict ClC-ec1 current under the broad range of conditions tested. This example demonstrates the potential robust and predictive power of MsRKM when the solution space is sufficiently refined and the response of the rates to the range of conditions modeled is properly included.

### 3.2 Top-Down Modeling with MsRKM

Even when specific rate constants are not known, MsRKM can identify robust and predictive models as long as there is sufficient molecular insight to define the state space and experimental data to narrow the solution space. Although the most insight is gained by combining top-down modeling from experimental data and bottom-up characterization of the dynamics and rates of individual transitions, purely top-down modeling can still generate new biological insight, reveal mechanistic control, and inform experimental design. It additionally provides a framework to fold in increasing bottom-up characterization and test contrasting predictions. We demonstrate the process of top-down

modeling with the Shaker voltage-activated potassium channel (Figure 8A), modeling only the open conformation.

We first constructed a four-binding site model of Shaker with one site located at the internal pore entrance (S1) and three binding sites in the region of the selectivity filter, (Figure 5B and 8A) consistent with MD generated  $K^+$  density plots<sup>2</sup>. We then generated 50 starting rate coefficient sets using a random-number generator to enable the optimizer to broadly sample the solution space. Only  $K^+$  uptake rate coefficients were constrained with an upper boundary of  $1 \times 10^8 \text{ M}^{-1} \text{ s}^{-1}$  ( $1 \times 10^5$



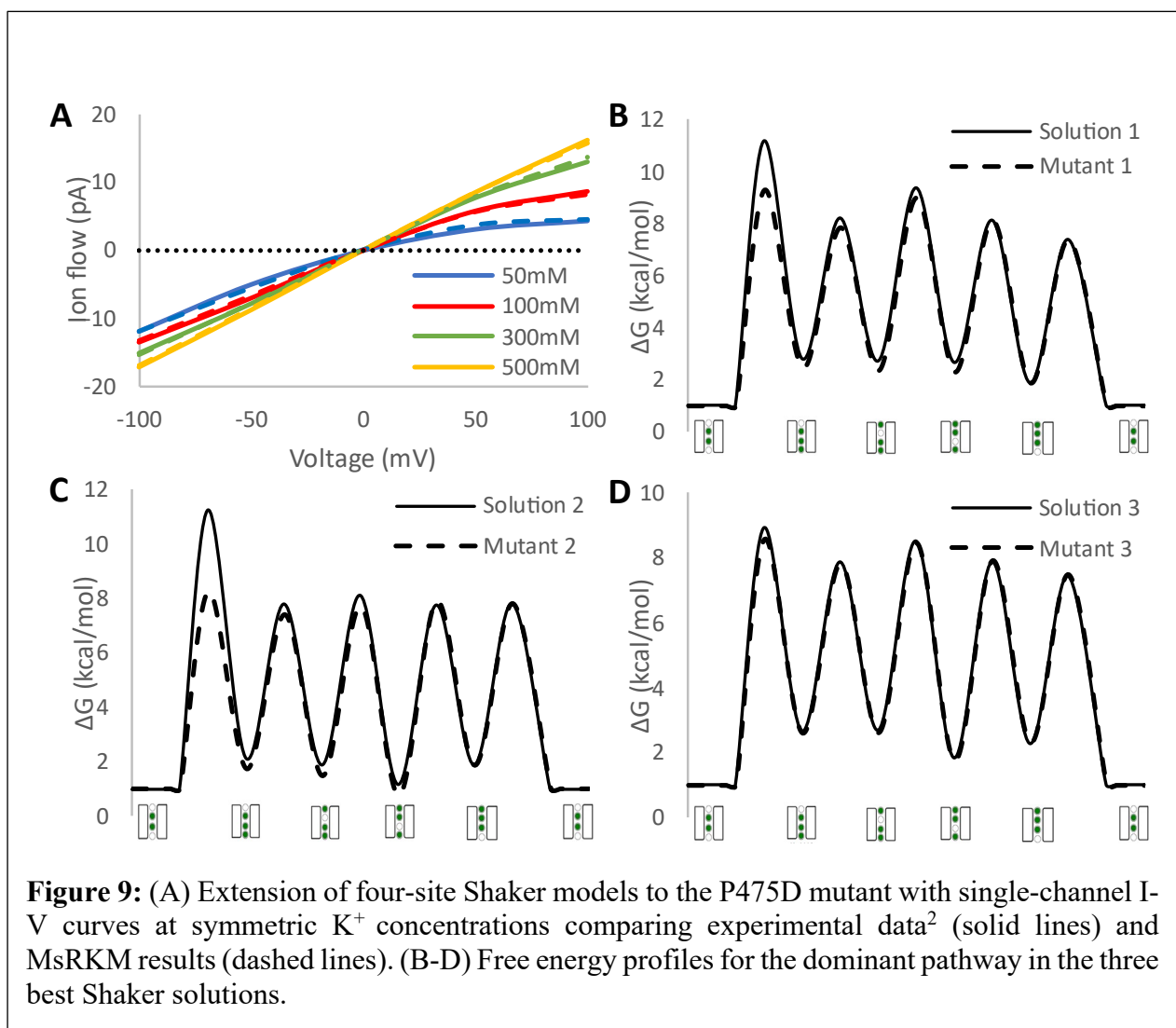
**Figure 8.** (A) Shaker channel showing site locations modeled in the four-site model (adapted from <sup>2</sup>). (B) Single-channel I-V curves for the Shaker channel with symmetric  $K^+$  concentrations, comparing experimental data<sup>1</sup> (solid lines) to four-site MsRKM results (dashed lines).

$\text{mM}^{-1}\text{ms}^{-1}$ ), approximating the diffusion limit. Sets of rate coefficients were considered successful solutions if they deviated by less than 0.3 pA from experimental single-channel I-V curves (Figure 8B). This criterion was met by 52% (26/50) of optimization runs. However, there was considerable heterogeneity in the resulting 26 Shaker solutions, including different voltage and chemical gradient current profiles, demonstrating the need for additional refinement.

One strategy for solution refinement is to use biophysical insight. Since all  $\text{K}^+$  channels share a strictly conserved selectivity filter but have a wide range in current (from 1.5 pA for the Shaker KV channel<sup>1</sup> to 25 pA in the BK Channel<sup>110</sup>), it has been suggested that the RLS for slower  $\text{K}^+$  channels like Shaker must be outside of the selectivity filter<sup>2</sup>. Specifically, the RLS should be uptake to S1 or transfer from the vestibule into the selectivity filter (S1 to S2). Although it seems unlikely for Shaker, release from S4 could also be considered if structural elements beyond the selectivity filter contribute to the barrier for  $\text{K}^+$  release. Based on these considerations, 13 of the 26 Shaker solutions could potentially be ruled out since their dominant pathways have a RLS in the selectivity filter. The remaining 13 solutions, however, remain viable solutions—11 with uptake to S1 as their RLS, 1 with S1-S2 transfer, and 1 with release from S4.



An alternative strategy for solution refinement is to test the solutions on new experimental data. Following this path, we further refined our solution space by modeling the Shaker-P475D mutation<sup>2</sup>. This mutant is particularly informative because the Pro to Asp mutation lies just inside the inner vestibule near the activation gate and significantly speeds up ion flux, promoting a 6 to 8-fold higher single-channel current. We modeled the P475D mutation using the rate coefficient sets of all 26 possible solutions as seeds in a secondary optimization fitting P475D's single-channel I-V curves (Figure 9A). Since the mutation is 17 Å below the selectivity filter, we only allowed rate coefficients near the mutation to change between the wildtype (WT) seeds and mutant. This

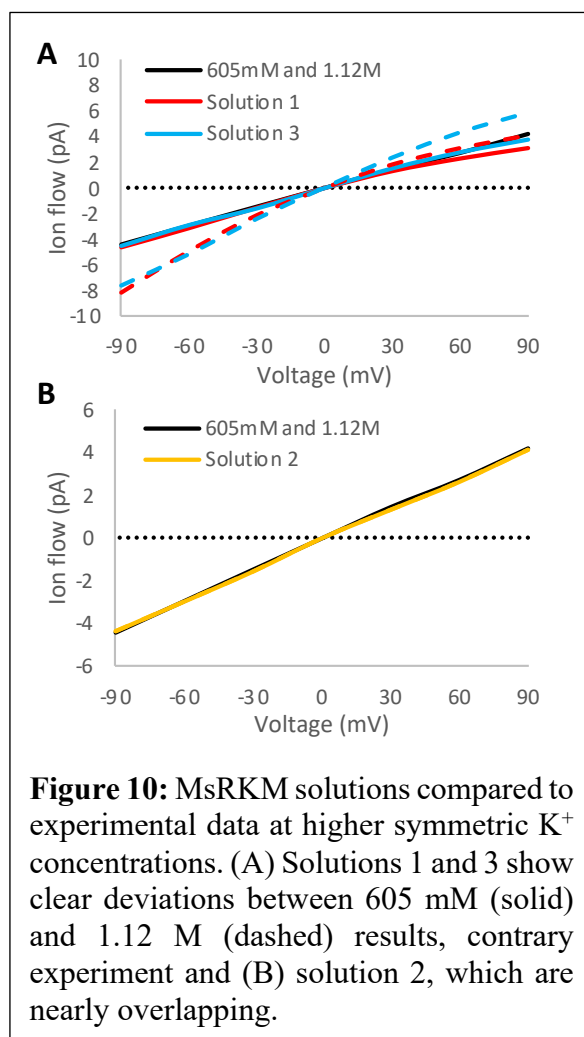


**Figure 9:** (A) Extension of four-site Shaker models to the P475D mutant with single-channel I-V curves at symmetric K<sup>+</sup> concentrations comparing experimental data<sup>2</sup> (solid lines) and MsRKM results (dashed lines). (B-D) Free energy profiles for the dominant pathway in the three best Shaker solutions.

included uptake to S1, S1 release to bulk, and S1 → S2 transfer. Notably, only 3 solutions were able to fit the mutant I-V curves within the <math><0.3\text{pA}</math> acceptance criterion (Figure 9A dashed lines). As expected, the WT RLS of these three solutions is outside the selectivity filter (two are uptake to S1 and one is S1 to S2 transfer at 100 mM  $\text{Cl}^-$ ). As a less stringent alternative test, we performed a secondary optimization allowing all rates to vary to fit the mutant data. This found an additional 7 solutions. The WT RLS of each of these seven additional solutions remains outside of the selectivity filter. Thus, both the WT and mutant results support the notion that the RLS for Shaker is either ion uptake or transfer into the selectivity filter.

We analyzed their dominant mechanistic pathways at 1:1 mM symmetric  $\text{K}^+$  equilibrium conditions (Figure 9B, C, and D) to differentiate between the three solutions. Analyzing pathways in the outward direction, solutions 1, 2, and 3 all have a RLS of uptake to S1 at this low concentration. Since uptake is rate limiting, the RLSs of the solutions are sensitive to the internal concentration of  $\text{K}^+$ . Solution 1 switches its RLS at 400mM from S1 uptake to S2 → S3 transfer. Solutions 2 and 3 switch their RLSs from S1 uptake to S1 → S2 transfer at 496mM and 25mM, respectively. The mutant solutions still have RLSs

of S1 uptake, but they reduce the uptake barrier heights by ~1.8, ~3, and ~0.4 kcal/mol for solutions



**Figure 10:** MsRKM solutions compared to experimental data at higher symmetric  $\text{K}^+$  concentrations. (A) Solutions 1 and 3 show clear deviations between 605 mM (solid) and 1.12 M (dashed) results, contrary experiment and (B) solution 2, which are nearly overlapping.

1, 2 and 3, respectively. Thus, the RLS moves at much lower concentrations. The RLS of solution 1 RLS shifts to  $S2 \rightarrow S3$  at 9mM, while the RLSs of solutions 2 and 3 shift to  $S1 \rightarrow S2$  at 2mM and 23mM, respectively. Collectively, the three solutions are unified in RLS at low  $K^+$  concentration, but they diverge significantly both as  $K^+$  concentrations increase and as they fit the mutant.

We turned to additional experimental data to further delineate which of the three solutions is best. Like other  $K^+$  channels, Shaker exhibits saturation of the channel, where increased concentration does not result in increased conduction. Saturation of the channel occurs between 325mM and 600mM, with concentrations  $>600$ mM minimally increasing flux through the transporter<sup>1</sup>. We tested the prediction capabilities of the three remaining solutions using the WT saturated I-V curves (Figure 10). All solutions correctly predict the 600mM single-channel I-V curve, but only solution 2 exhibits saturation at 1.15M.

The aforementioned process demonstrates how gradually integrating experimental insights and observations can narrow the solution space, enabling us to identify a single set of rate coefficients (solution 2) that is consistent with experimental data and biophysical insight. However, we caution against assuming additional solutions do not exist. The global optimization procedure explores the phase space stochastically, with a deterministic local optimization to provide locally optimal solutions, but it provides no guarantee that we have found the true global optimum. As solution 2 best describes all experimental data to date, we focus our mechanistic analysis below on solution 2.

### **3.3 MsRKM Can Identify Mechanisms Dominated by a Single Pathway**

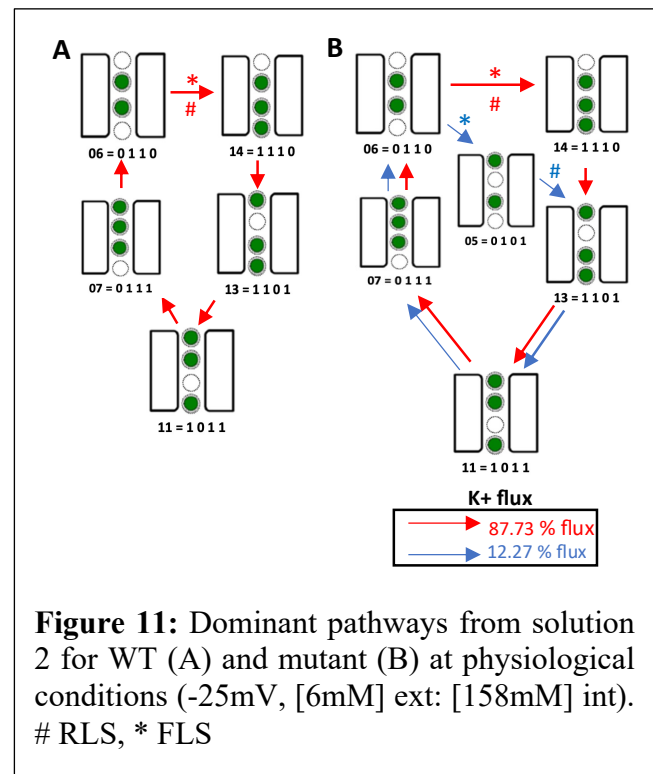
All previously published MKM models displayed mechanisms with multiple pathways contributing to the net flux<sup>6,9</sup> that shifted under different reaction conditions. Thus, a critical test

of the MsRKM methodology is not in its ability to identify competing pathways, which it is designed to do, but rather in its ability to eliminate superfluous mechanisms in single-pathway dominant systems. The Shaker four-site model is a good test in this sense. Most of the solutions identified displayed some degree of multiple contributing pathways. However, solution 2 is almost entirely single-pathway dominant, with >99% of ion flow at physiologically and experimentally relevant voltages and concentrations going through the pathway shown in Figure 11A.

Interestingly, this mechanism is consistent with that proposed by previously through double membrane computational electrophysiology simulations.<sup>70, 71, 111</sup> Flux through other pathways only becomes significant at high voltages or concentrations. Although additional mechanistic pathways appear at high concentrations, we observe no significant increase in total flux through the protein at concentrations above 500mM.

In contrast, solution 2 for the P475D mutant

does not retain this mechanistic homogeneity (Figure 11B). The secondary pathway of the mutant solution contributes 12.27% to the net flux at -25mV, within normal physiological conditions (Figure 6B blue). This system also shows more voltage-dependent kinetic selection, with the secondary pathway's contribution to the net flux increasing and decreasing with increasing and decreasing voltage, respectively. Thus, MsRKM correctly identified both a single-pathway



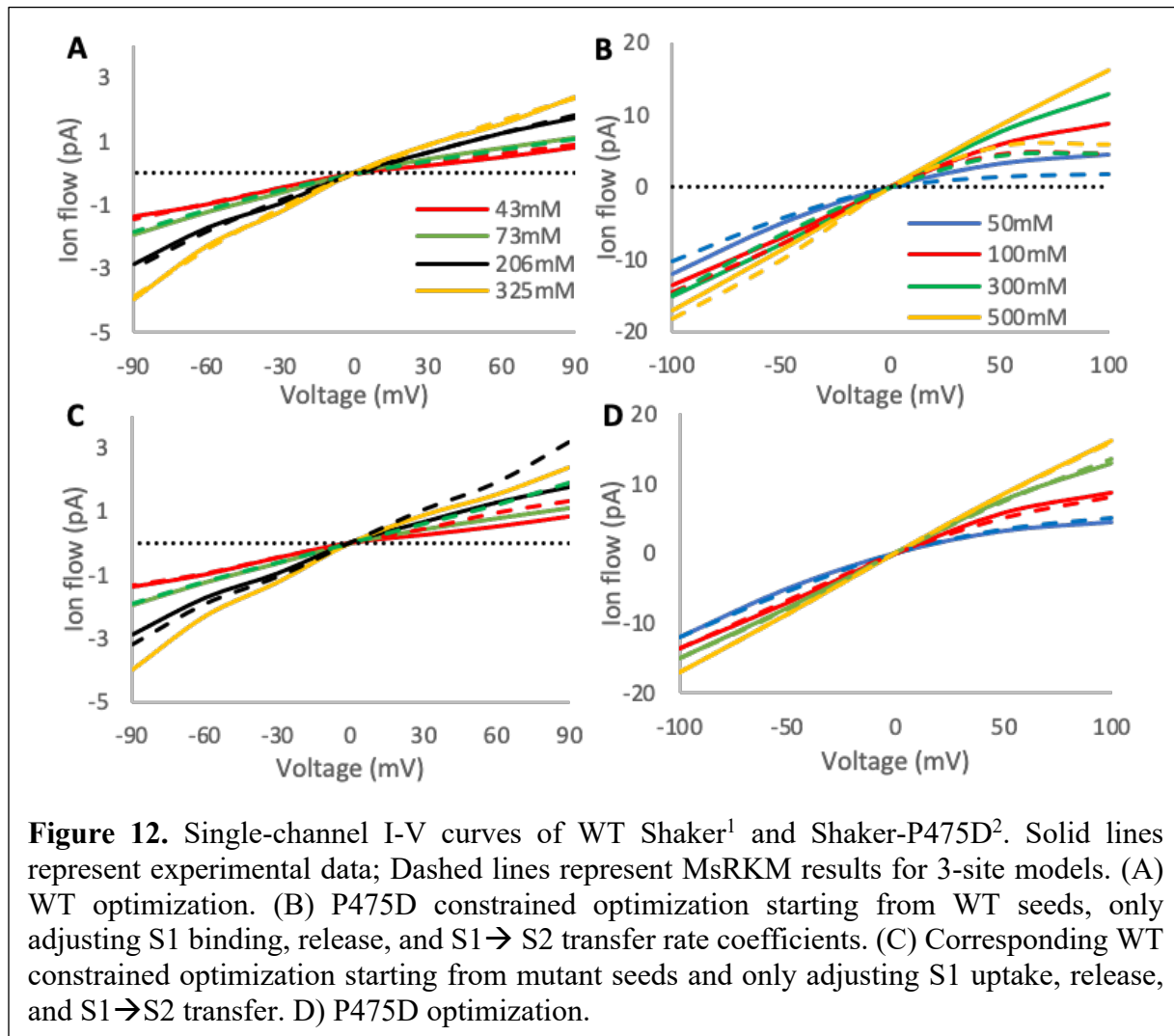
dominant mechanism and a multi-pathway mechanism in closely related systems, suggesting that MsRKM is sufficiently flexible to capture the nuances of real data.

While the mechanistic differences between this study's WT and mutant models are enlightening in exploring MsRKM capabilities, we caution against drawing Shaker-specific mechanistic conclusions. We constructed a simplified Shaker model to find the minimum number of states necessary to replicate experimental data. Most MD and experimentally derived models show additional binding sites<sup>43,45,47,111</sup> that likely need to be included to draw physiological conclusions. We also used a model membrane voltage coupling factor for Shaker instead of a system-specific profile, such as that shown for CIC-ec1 (Figure 5A). A more detailed MsRKM model of Shaker will be the goal of future work, including more sites, the role of water, and coupled ion transitions. However, if the results described above for our simple four-site model hold in more detailed models, it suggests that Shaker evolved a kinetic network that supports a robust, single-pathway dominant mechanism for K<sup>+</sup> transport in the 'open' configuration.

### **3.4 Model Refinement Provides Clear Feedback between Acceptable Simplification and Oversimplification**

While the four-site MsRKM model was capable of fitting all Shaker experimental data to date, theoretical models using fewer sites have provided valuable insights. For example, Mosoco et al. used three potassium ion binding sites in their Eyring model, an early kinetic model that altered barrier heights and well stabilities of a mechanistic pathway,<sup>64</sup> to explain the P475D mutant's increased conductance<sup>2</sup>. This suggests that further simplification from the four-site MsRKM could produce viable solutions. To test this, we constructed a three-site Shaker model that combines S2 and S3 (Figure 8A) to form a single site. We then fit the WT Shaker single-channel I-V curves

with the same procedure, boundaries, and acceptance criterion detailed above. Out of 100 optimizations, 62 (62%) resulted in a solution that met the acceptance criteria. The I-V curves for one such solution is shown in Figure 12A. These 62 WT solutions were then used as seeds for a secondary optimization to the mutant data, as detailed above. Surprisingly, this secondary optimization did not produce a single viable solution, with the best fit having a 2.96pA average deviation from experimental data (Figure 12B).



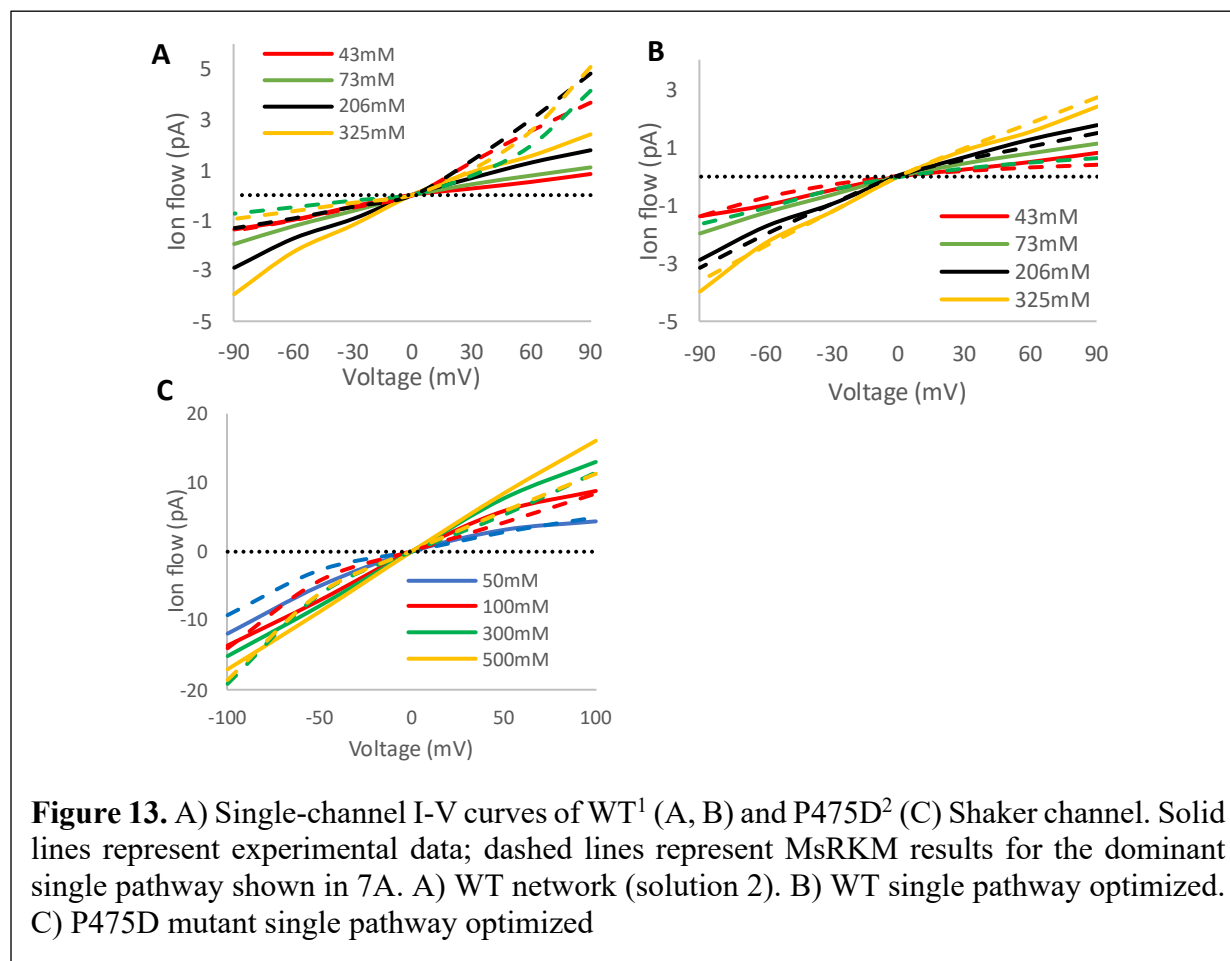
A critical difference between the three-site MsRKM model and the previously proposed P475D Eyring model is the assumed impact of electrostatics across the entire pore region. The MsRKM model assumes that electrostatic changes at the channel mouth would not dramatically impact the free energy of transitions  $>17$  Å away in the selectivity filter. Meanwhile, the Eyring model assumed the mutant would alter the stability of all the binding sites and barriers, including that of S4, which is  $\sim 35$  Å away. To test if the MsRKM could replicate the results of the Eyring model, we did an initial optimization of mutant experimental data using the same boundaries and acceptance criterion as the WT optimization. This optimization produced 95 (95%) mutant solutions that met the acceptance criterion. We repeated the secondary optimization using the 95 mutant solutions as seeds to test if the new mutant solutions could then back-map to reproduce the WT single-channel I-V curves. Again, only S1 uptake and release, and S1  $\rightarrow$  S2 transfer were allowed to change. None of the 95 mutants  $\rightarrow$  WT optimizations produced a viable solution within the acceptance criterion (Figure 12C).

Thus, the 3-site MsRKM model could not find a solution that accurately describes both WT and mutant experimental data. This is distinct from the four-site model and suggests that limiting state space to only 3 binding sites is an oversimplification for the Shaker channel. Future work will explore the ramifications of additional sites and the optimal representation for Shaker. Regardless of the optimal representation, these results demonstrate that the number of binding sites directly impacts the model's ability to replicate experimental data.

### **3.5 A Network Representation is Still Essential in Single-Pathway Dominant Mechanisms**

Previous studies have demonstrated that non-dominant mechanistic pathways contribute to ion flux in ClC-ec1 and can eventually shift to be dominant under different electrochemical gradients. This suggests that kinetic selection in ClC-ec1 changes mechanistic pathways as a function of

reaction conditions. However, understanding how the reaction network, including off-pathway and competing pathway flux, influences these dominant pathways remains relatively unknown. We



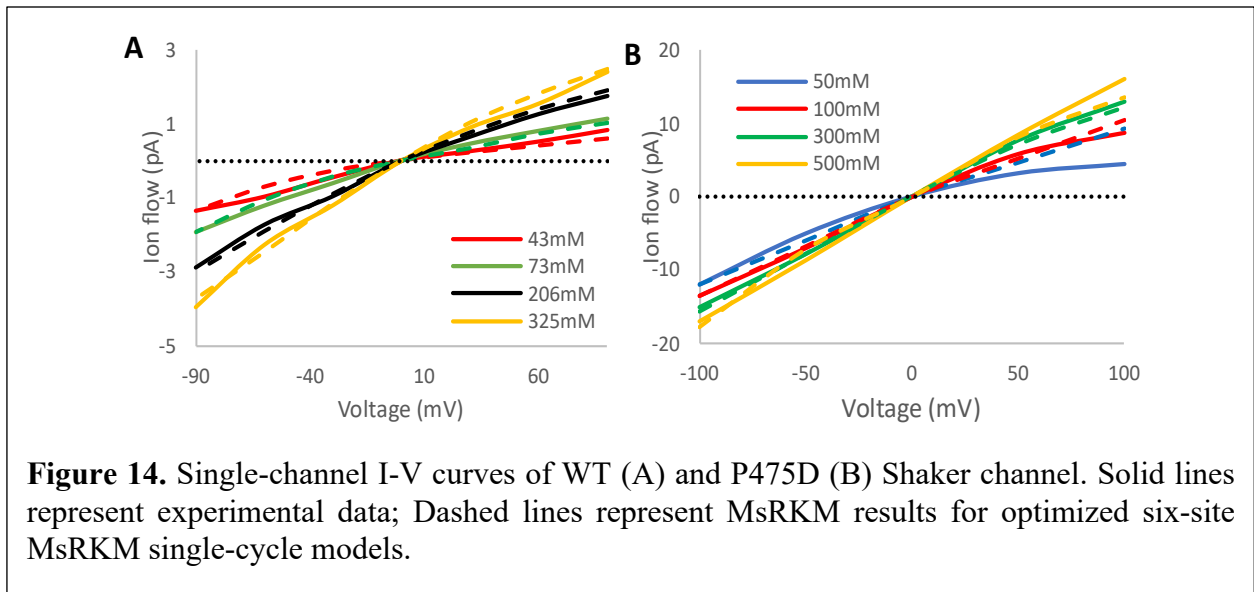
initially assumed that off-pathway cycles would not impact the dominant pathway in Shaker since a single mechanistic cycle prevails under a wide range of experimental conditions (solution 2). However, when we pruned the network down to the dominant pathway alone (Figure 11A), we saw completely different model behavior (Figure 13A), indicating that the off-pathway flux remains essential in this solution.

Since many published mechanisms of ion channels and secondary active transporters depict a single-cycle dominant mechanism, as seen for potassium channels<sup>43, 45, 111</sup> and CIC-ec1,<sup>27, 32, 112</sup> it is important to understand whether or not a network description of such mechanisms is important.



Thus, to test if the contribution of off-pathway flux in solution 2 is simply an artifact of the MsRKM network solution refinement, we constructed a single-cycle kinetic model of the dominant pathway in solution 2 and optimized it to fit the WT experimental data using the same refinement methods used in the network implementation. However, the single-cycle kinetic model struggled to replicate the experimental data, deviating  $>0.3\text{pA}$  at every non-0 point along the I-V curve (Figure 13B). We then optimized the model to mutant data with a similar unconstrained optimization. Despite the mutant sharing the same dominant mechanistic pathway as the WT (Figure 11A, B), we observed significantly larger deviations (Figure 13C), potentially due to the larger flux observed in the mutant.

As noted in the three-site MsRKM, the number of intermediates strongly influences model behavior. Further, the single-cycle mechanism that we used was derived from the multi-pathway MsRKM solution 2, potentially biasing the system towards a multiple-cycle description. To remove these biases, we modeled the six-site mechanism described from MD simulations of the KcsA channel, a similar potassium channel<sup>43</sup> (Figure 14). The fit of the six-site single-cycle model to the experimental data improves relative to the four-site single-cycle model, with the greatest single-point deviation just outside our acceptance criteria for the WT solution (Figure 14A). Notably, these deviations are still larger than the network optimized solution 2. The mutant six-site also performs better in an unconstrained mutant optimization (Figure 14B) compared to the four-site single-pathway model, but significant deviations remain compared to both experiment and the network optimized four-site model (Figure 11A).



To further test the validity of a single-cycle representation, we examined the rectification properties of the P475D mutant, which change as a function of  $K^+$  concentration. Specifically, the degree of rectification decreases as concentration increases: the rectification ratio is  $\sim 0.94$  for  $+100\text{mV}/-100\text{mV}$  at  $500\text{mM}$  but drops to  $0.36$  for  $+100\text{mV}/-100\text{mV}$  at  $50\text{mM}$ . The WT data does not display this behavior, with rectification maintaining a ratio of roughly  $0.60$  for  $+90\text{mV}/-90\text{mV}$  from  $43\text{mM}$  to  $325\text{mM}$ . Surprisingly, none of the single-cycle models tested capture the changing rectification behavior of the P475D mutant. Instead, they maintain a rectification ratio close to that of the WT experimental data ( $\sim 0.64$  for  $\pm 100\text{mV}$  at all concentrations tested). The six-site, single-pathway mutant solution shows small changes in rectification as a function of concentration, which allows for a slightly better fit, but is still far off from the  $\sim 0.58$  rectification ratio change in the experimental data (Figure 14B). This problem was not observed for the multi-pathway model (Figure 12D), which implies that multiple mechanistic pathways may play an important role in concentration-dependent rectification. However, changing rectification as a function of concentration may be affected by site location and the dimensionless coupling factor. Future studies with a more detailed Shaker model will explore these relationships further. However, if the

results found herein hold for more detailed models, a network representation that includes competing and off-pathway flux may be essential to describe concentration-dependent rectification properties.

## CONCLUSIONS

We have presented a new multiscale kinetic modeling framework that describes biomolecular processes involving multiple rare event transitions with reaction networks and the reactive flux through them under a range of equilibrium and non-equilibrium conditions. The framework was designed to combine the strengths of bottom-up rate quantification from multiscale simulations with top-down numerical solution refinement with experimental data to iteratively converge kinetic networks that not only describe known observables but are consistent with our best microscopic understanding. The obtained network descriptions can reveal when and how competing mechanistic pathways and off-pathway flux influence mechanistic outcomes. Motivated by the challenge of describing channels and transporters, the framework was designed to be responsive to electrochemical reaction conditions via voltage-dependent rate coefficients. This enables the use of and comparison to electrophysiology assays. Although MsRKM captures a system's non-equilibrium response to a range of conditions, care should be taken to apply it to conditions for which the influence on underlying rate coefficients remains valid. In present form, this does not include significant conformational changes, such as those demonstrated in voltage-gated channels' gating and inactivation.

The methods are demonstrated on the CIC-ec1 antiporter and the Shaker channel. For the more complex antiporter, CIC-ec1, determining a robust MsRKM solution requires integrating molecular-level data from simulations<sup>6, 33, 34</sup> with ensemble data from experiments<sup>26, 30, 31, 109</sup>. Specifically, multiscale simulations provide estimated rate coefficients that help limit the vast

solution space but require refinement based on experimental data. Increasing the variety in the experimental data further refines the solution space until robust and consistent solutions emerge. Importantly, the kinetic modeling predicted that certain rates require larger error bounds for low pH behavior than originally obtained in simulations. This prediction is consistent with recent cryoEM data demonstrating a pH-dependent conformational change.<sup>29</sup> This is an example of how kinetic modeling can combine simulation and experimental data to iteratively converge on a consistent mechanistic description and to identify inconsistencies between the two.

The Shaker channel is modeled to test MsRKM in a purely top-down kinetic modeling approach based on experimental data. We demonstrate that robust solutions and physical insight can still be obtained in the absence of simulation-based rates. By iteratively increasing the integrated experimental data,<sup>1,2</sup> MsRKM converges to a solution that best predicts channel behavior, while revealing several interesting biophysical features. First, solutions that best fit WT Shaker and the P475D mutant had rate limiting steps before the selectivity filter, consistent with previous predictions<sup>2</sup> for low-conductance K<sup>+</sup> channels. Second, despite the inclusion of the entire reaction network, MsRKM converged on a single-pathway dominant mechanism (solution 2), verifying that the method can eliminate noise to identify single-pathway dominant processes. Third, the MsRKM process can reveal limitations in the state space definition. Reducing the four-site Shaker model to three binding sites failed to fit all known experimental data. Fourth, despite being dominated by a single pathway, the network representation, including competing and off-pathway flux, is still essential. The dominant pathway alone, with any rate definitions, fails to replicate experimental data. Even with the addition of binding sites and the use of a dominant pathway identified in simulations<sup>43</sup>, the six-site single-pathway model remains insufficient. In particular, single-pathway models fail to describe concentration-dependent rectification, suggesting that

variable rectification ratios may require a network description. This is supported by the full four-site mutant solution, which revealed pathways sensitive to changing  $K^+$  and voltage gradients, resulting in variable rectification ratios.

While this paper focuses on channels and transporters, the MsRKM formalism is apt to treat any process involving multiple rare event transitions. It serves as a helpful bridge between experiment and simulation, enabling self-consistent data integration to identify and remedy inconsistencies. It additionally describes the non-equilibrium reaction flux that is often both biologically relevant and experimentally measured. This provides a consistent framework to describe and interpret a system's equilibrium characterization and nonequilibrium behavior. Perhaps most importantly, it captures a complete picture of the reaction network and quantifies how competing mechanistic pathways collectively define the reactive flux and, thus, mechanism.

Future work will explore the incorporation rates responsive to pH-driven conformational changes. We will also test the hypotheses presented herein on more detailed Shaker models with more ion binding sites, the inclusion of competing water molecules, and system-specific voltage coupling factors.

## **AUTHOR INFORMATION**

### **Corresponding Author**

\* **Jessica M. J. Swanson** - Department of Chemistry, University of Utah, Salt Lake City, UT, 84112 – United States, Email: [j.swanson@utah.edu](mailto:j.swanson@utah.edu); <https://orcid.org/0000-0002-9820-1307>

### **Authors**

**Hannah Weckel-Dahman** - Department of Chemistry, University of Utah, Salt Lake City, UT, 84112 – United States; <https://orcid.org/0000-0003-2834-8386>

**Ryan Carlsen** – Department of Chemistry, University of Utah, Salt Lake City, UT, 84112 – United States; <https://orcid.org/0009-0005-6213-1182>

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENT

The authors acknowledge Dimitry Vlachos and Benoit Roux for helpful discussions. This work was supported by NIH NIGMS (R35GM143117) and the computational resources provided by Expanse at the San Diego Supercomputing Center through the Advanced Cyberinfrastructure Coordination Ecosystem: Services and Support (ACCESS) program (allocation MCB200018) supported by NSF (grant nos. 2138259, 2138286, 2138307, 2137603, and 2138296), as well as the Center for High-Performance Computing (CHPC) at the University of Utah.

## REFERENCES

- (1) Heginbotham, L.; Mackinnon, R. Conduction Properties of the Cloned Shaker K<sup>+</sup> Channel. *Biophys J* **1993**, *65* (5), 2089-2096. DOI: 10.1016/S0006-3495(93)81244-X.
- (2) Moscoso, C.; Vergara-Jaque, A.; Márquez-Miranda, V.; Sepúlveda, R. V.; Valencia, I.; Díaz-Franulic, I.; González-Nilo, F.; Naranjo, D. K<sup>+</sup> Conduction and Mg<sup>+</sup> Blockade in a Kv-Channel Single Point Mutant with an Unusually High Conductance. *Biophys J* **2012**, *103* (6), 1198-1207. DOI: 10.1016/j.bpj.2012.08.015.

(3) Schuldiner, S.; Granot, D.; Mordoch, S. S.; Ninio, S.; Rotem, D.; Soskin, M.; Tate, C. G.; Yerushalmi, H. Small is mighty: EmrE, a multidrug transporter as an experimental paradigm. *News Physiol Sci* **2001**, *16*, 130-134. DOI: 10.1152/physiologyonline.2001.16.3.130.

(4) Robinson, A. E.; Thomas, N. E.; Morrison, E. A.; Balthazor, B. M.; Henzler-Wildman, K. A. New free-exchange model of EmrE transport. *Proc Natl Acad Sci U S A* **2017**, *114* (47), E10083-E10091. DOI: 10.1073/pnas.1708671114.

(5) Bazzone, A.; Zabadne, A. J.; Salisowski, A.; Madej, M. G.; Fendler, K. A Loose Relationship: Incomplete H(+)/Sugar Coupling in the MFS Sugar Transporter GlcP. *Biophys J* **2017**, *113* (12), 2736-2749. DOI: 10.1016/j.bpj.2017.09.038.

(6) Mayes, H. B.; Lee, S.; White, A. D.; Voth, G. A.; Swanson, J. M. J. Multiscale Kinetic Modeling Reveals an Ensemble of Cl/H Exchange Pathways in ClC-ec1 Antiporter. *J Am Chem Soc* **2018**, *140* (5), 1793-1804. DOI: 10.1021/jacs.7b11463.

(7) Grabe, M.; Zuckerman, D. M.; Rosenberg, J. M. EmrE reminds us to expect the unexpected in membrane transport. *J Gen Physiol* **2020**, *152* (1). DOI: 10.1085/jgp.201912467.

(8) Bisignano, P.; Lee, M. A.; George, A.; Zuckerman, D. M.; Grabe, M.; Rosenberg, J. M. A kinetic mechanism for enhanced selectivity of membrane transport. *PLoS Comput Biol* **2020**, *16* (7), e1007789. DOI: 10.1371/journal.pcbi.1007789.

(9) Yue, Z.; Bernardi, A.; Li, C.; Mironenko, A. V.; Swanson, J. M. J. Toward a Multipathway Perspective: pH-Dependent Kinetic Selection of Competing Pathways and the Role of the Internal Glutamate in Cl(-)/H(+) Antiporters. *J Phys Chem B* **2021**, *125* (29), 7975-7984. DOI: 10.1021/acs.jpcc.1c03304.

(10) Hopfield, J. J. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc Natl Acad Sci U S A* **1974**, *71* (10), 4135-4139. DOI: 10.1073/pnas.71.10.4135.

(11) Ninio, J. Kinetic amplification of enzyme discrimination. *Biochimie* **1975**, *57* (5), 587-595. DOI: 10.1016/s0300-9084(75)80139-8.

(12) Unarta, I. C.; Zhu, L.; Tse, C. K. M.; Cheung, P. P.; Yu, J.; Huang, X. Molecular mechanisms of RNA polymerase II transcription elongation elucidated by kinetic network models. *Curr Opin Struct Biol* **2018**, *49*, 54-62. DOI: 10.1016/j.sbi.2018.01.002.

(13) Swain, P. S.; Siggia, E. D. The role of proofreading in signal transduction specificity. *Biophys J* **2002**, *82* (6), 2928-2933. DOI: 10.1016/S0006-3495(02)75633-6.

(14) Yousefi, O. S.; Gunther, M.; Horner, M.; Chalupsky, J.; Wess, M.; Brandl, S. M.; Smith, R. W.; Fleck, C.; Kunkel, T.; Zurbriggen, M. D.; et al. Optogenetic control shows that kinetic proofreading regulates the activity of the T cell receptor. *Elife* **2019**, *8*. DOI: 10.7554/eLife.42475.

(15) Lo, W. L.; Weiss, A. Adapting T Cell Receptor Ligand Discrimination Capability via LAT. *Front Immunol* **2021**, *12*, 673196. DOI: 10.3389/fimmu.2021.673196.

(16) Liepelt, S.; Lipowsky, R. Kinesin's Network of Chemomechanical Motor Cycles. *Phys Rev Lett* **2007**, *98* (25), 258102-258106. DOI: 10.1103/PhysRevLett.98.258102.

(17) Svoboda, K.; Schmidt, C. F.; Schnapp, B. J.; Block, S. M. Direct observation of kinesin stepping by optical trapping interferometry. *Nature* **1993**, *365* (6448), 721-727. DOI: 10.1038/365721a0.



- (18) Sumi, T. Myosin V: Chemomechanical-coupling ratchet with load-induced mechanical slip. *Sci Rep* **2017**, 7 (1), 13489. DOI: 10.1038/s41598-017-13661-0.
- (19) George, A.; Bisignano, P.; Rosenberg, J. M.; Grabe, M.; Zuckerman, D. M. A systems-biology approach to molecular machines: Exploration of alternative transporter mechanisms. *PLoS Comput Biol* **2020**, 16 (7), e1007884. DOI: 10.1371/journal.pcbi.1007884.
- (20) Hussey, G. A.; Thomas, N. E.; Henzler-Wildman, K. A. Highly coupled transport can be achieved in free-exchange transport models. *J Gen Physiol* **2020**, 152 (1). DOI: 10.1085/jgp.201912437.
- (21) Husic, B. E.; Pande, V. S. Markov State Models: From an Art to a Science. *J Am Chem Soc* **2018**, 140 (7), 2386-2396. DOI: 10.1021/jacs.7b12191.
- (22) Noe, F.; Rosta, E. Markov Models of Molecular Kinetics. *J Chem Phys* **2019**, 151 (19), 190401. DOI: 10.1063/1.5134029.
- (23) Iyer, R.; Iverson, T. M.; Accardi, A.; Miller, C. A biological role for prokaryotic ClC chloride channels. *Nature* **2002**, 419 (6908), 715-718. DOI: 10.1038/nature01000.
- (24) Jentsch, T. J.; Pusch, M. CLC Chloride Channels and Transporters: Structure, Function, Physiology, and Disease. *Physiol Rev* **2018**, 98 (3), 1493-1590. DOI: 10.1152/physrev.00047.2017.
- (25) Dutzler, R.; Campbell, E. B.; Cadene, M.; Chait, B. T.; MacKinnon, R. X-ray structure of a ClC chloride channel at 3.0 Å reveals the molecular basis of anion selectivity. *Nature* **2002**, 415 (6869), 287-294. DOI: 10.1038/415287a.

(26) Accardi, A.; Miller, C. Secondary active transport mediated by a prokaryotic homologue of CLC Cl channels. *Nature* **2004**, *427* (6977), 803-807. DOI: 10.1038/nature02314.

(27) Chavan, T. S.; Cheng, R. C.; Jiang, T.; Mathews, II; Stein, R. A.; Koehl, A.; McHaourab, H. S.; Tajkhorshid, E.; Maduke, M. A CLC-ec1 mutant reveals global conformational change and suggests a unifying mechanism for the CLC Cl(-)/H(+) transport cycle. *Elife* **2020**, *9*. DOI: 10.7554/eLife.53479.

(28) Yue, Z.; Li, C.; Voth, G. A. The role of conformational change and key glutamic acid residues in the CLC-ec1 antiporter. *Biophys J* **2023**, *122* (6), 1068-1085. DOI: 10.1016/j.bpj.2023.01.025.

(29) Fortea, E.; Lee, S.; Chadda, R.; Argyros, Y.; Sandal, P.; Mahoney-Kruszka, R.; Ciftci, H. D.; Falzone, M. E.; Huysmans, G.; Robertson, J. L.; et al. Structural basis of pH-dependent activation in a CLC transporter. *Nat Struct Mol Biol* **2024**, *31* (4), 644-656. DOI: 10.1038/s41594-023-01210-5.

(30) Accardi, A.; Kolmakova-Partensky, L.; Williams, C.; Miller, C. Ionic currents mediated by a prokaryotic homologue of CLC Cl channels. *J Gen Physiol* **2004**, *123* (2), 109-119. DOI: 10.1085/jgp.200308935.

(31) Matulef, K.; Maduke, M. Side-dependent inhibition of a prokaryotic CIC by DIDS. *Biophys J* **2005**, *89* (3), 1721-1730. DOI: 10.1529/biophysj.105.066522.

(32) Feng, L.; Campbell, E. B.; MacKinnon, R. Molecular mechanism of proton transport in CLC Cl-/H+ exchange transporters. *Proc Natl Acad Sci U S A* **2012**, *109* (29), 11699-11704. DOI: 10.1073/pnas.1205764109.

(33) Lee, S.; Mayes, H. B.; Swanson, J. M. J.; Voth, G. A. The Origin of Coupled Chloride and Proton Transport in a Cl/H Antiporter. *J Am Chem Soc* **2016**, *138* (45), 14923-14930. DOI: 10.1021/jacs.6b06683.

(34) Lee, S.; Swanson, J. M. J.; Voth, G. A. Multiscale Simulations Reveal Key Aspects of the Proton Transport Mechanism in the ClC-ec1 Antiporter. *Biophys J* **2016**, *110* (6), 1334-1345. DOI: 10.1016/j.bpj.2016.02.014.

(35) Lüers, H. Shaker, eine erbliche Bewegungsstörung bei *Drosophila funebris*. *Zeitschrift für Induktive Abstammungs- und Vererbungslehre* **1937**, *72* (1), 119-126. DOI: 10.1007/BF01850146.

(36) Jan, Y. N.; Jan, L. Y.; Dennis, M. J. Two mutations of synaptic transmission in *Drosophila*. *Proc R Soc Lond B Biol Sci* **1977**, *198* (1130), 87-108. DOI: 10.1098/rspb.1977.0087.

(37) Tanouye, M. A.; Ferrus, A.; Fujita, S. C. Abnormal action potentials associated with the Shaker complex locus of *Drosophila*. *Proc Natl Acad Sci U S A* **1981**, *78* (10), 6548-6552. DOI: 10.1073/pnas.78.10.6548.

(38) Doyle, D. A.; Morais Cabral, J.; Pfuetzner, R. A.; Kuo, A.; Gulbis, J. M.; Cohen, S. L.; Chait, B. T.; MacKinnon, R. The structure of the potassium channel: molecular basis of K<sup>+</sup> conduction and selectivity. *Science* **1998**, *280* (5360), 69-77. DOI: 10.1126/science.280.5360.69.

(39) Yifrach, O.; MacKinnon, R. Energetics of pore opening in a voltage-gated K<sup>(+)</sup> channel. *Cell* **2002**, *111* (2), 231-239. DOI: 10.1016/s0092-8674(02)01013-9.

(40) Hoshi, T.; Zagotta, W. N.; Aldrich, R. W. Two types of inactivation in Shaker K<sup>+</sup> channels: effects of alterations in the carboxy-terminal region. *Neuron* **1991**, *7* (4), 547-556. DOI: 10.1016/0896-6273(91)90367-9.

- (41) Hodgkin, A. L.; Keynes, R. D. The potassium permeability of a giant nerve fibre. *J Physiol* **1955**, *128* (1), 61-88. DOI: 10.1113/jphysiol.1955.sp005291.
- (42) Schumaker, M. F.; MacKinnon, R. A simple model for multi-ion permeation. Single-vacancy conduction in a simple pore model. *Biophys J* **1990**, *58* (4), 975-984. DOI: 10.1016/S0006-3495(90)82442-5.
- (43) Berneche, S.; Roux, B. Energetics of ion conduction through the K<sup>+</sup> channel. *Nature* **2001**, *414* (6859), 73-77. DOI: 10.1038/35102067.
- (44) Nelson, P. H. A permeation theory for single-file ion channels: one- and two-step models. *J Chem Phys* **2011**, *134* (16), 165102. DOI: 10.1063/1.3580562.
- (45) Mironenko, A.; Zachariae, U.; de Groot, B. L.; Kopec, W. The Persistent Question of Potassium Channel Permeation Mechanisms. *J Mol Biol* **2021**, *433* (17), 167002. DOI: 10.1016/j.jmb.2021.167002.
- (46) Lam, C. K.; de Groot, B. L. Ion Conduction Mechanisms in Potassium Channels Revealed by Permeation Cycles. *J Chem Theory Comput* **2023**, *19* (9), 2574-2589. DOI: 10.1021/acs.jctc.3c00061.
- (47) Kasahara, K.; Shirota, M.; Kinoshita, K. Ion concentration-dependent ion conduction mechanism of a voltage-sensitive potassium channel. *PLoS One* **2013**, *8* (2), e56342. DOI: 10.1371/journal.pone.0056342.
- (48) Kasahara, K.; Shirota, M.; Kinoshita, K. Ion Concentration- and Voltage-Dependent Push and Pull Mechanisms of Potassium Channel Ion Conduction. *PLoS One* **2016**, *11* (3), e0150716. DOI: 10.1371/journal.pone.0150716.

- (49) Swanson, J. M. Multiscale kinetic analysis of proteins. *Curr Opin Struct Biol* **2022**, *72*, 169-175. DOI: 10.1016/j.sbi.2021.11.005.
- (50) Drew, D.; Boudker, O. Shared Molecular Mechanisms of Membrane Transporters. *Annu Rev Biochem* **2016**, *85*, 543-572. DOI: 10.1146/annurev-biochem-060815-014520.
- (51) Heginbotham, L.; MacKinnon, R. The aromatic binding site for tetraethylammonium ion on potassium channels. *Neuron* **1992**, *8* (3), 483-491. DOI: 10.1016/0896-6273(92)90276-j.
- (52) Gfeller, D.; Chappelier, J. C.; De Los Rios, P. Finding instabilities in the community structure of complex networks. *Phys Rev E Stat Nonlin Soft Matter Phys* **2005**, *72* (5 Pt 2), 056135. DOI: 10.1103/PhysRevE.72.056135.
- (53) Li, A.; Cornelius, S. P.; Liu, Y. Y.; Wang, L.; Barabasi, A. L. The fundamental advantages of temporal networks. *Science* **2017**, *358* (6366), 1042-1046. DOI: 10.1126/science.aai7488.
- (54) Daniel, T.; Casenave, F.; Akkari, N.; Ryckelynck, D. Model order reduction assisted by deep neural networks (ROM-net). *Advanced Modeling and Simulation in Engineering Sciences* **2020**, *7* (1), 16. DOI: 10.1186/s40323-020-00153-6.
- (55) Pavlopoulos, G. A.; Secrier, M.; Moschopoulos, C. N.; Soldatos, T. G.; Kossida, S.; Aerts, J.; Schneider, R.; Bagos, P. G. Using graph theory to analyze biological networks. *BioData Min* **2011**, *4*, 10. DOI: 10.1186/1756-0381-4-10.
- (56) Tan, X. F.; Bae, C.; Stix, R.; Fernandez-Marino, A. I.; Huffer, K.; Chang, T. H.; Jiang, J. S.; Faraldo-Gómez, J. D.; Swartz, K. J. Structure of the Shaker Kv channel and mechanism of slow C-type inactivation. *Sci Adv* **2022**, *8* (11). DOI: 10.1126/sciadv.abm7814.

(57) Adam, Y.; Tayer, N.; Rotem, D.; Schreiber, G.; Schuldiner, S. The fast release of sticky protons: kinetics of substrate binding and proton release in a multidrug transporter. *Proc Natl Acad Sci U S A* **2007**, *104* (46), 17989-17994. DOI: 10.1073/pnas.0704425104.

(58) Bernardi, R. C.; Melo, M. C. R.; Schulten, K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Bba-Gen Subjects* **2015**, *1850* (5), 872-877. DOI: 10.1016/j.bbagen.2014.10.019.

(59) Warshel, A.; Levitt, M. Theoretical Studies of Enzymic Reactions - Dielectric, Electrostatic and Steric Stabilization of Carbonium-Ion in Reaction of Lysozyme. *J Mol Biol* **1976**, *103* (2), 227-249. DOI: 10.1016/0022-2836(76)90311-9.

(60) Knight, C.; Lindberg, G. E.; Voth, G. A. Multiscale reactive molecular dynamics. *J Chem Phys* **2012**, *137* (22). DOI: 10.1063/1.4743958.

(61) Lee, S.; Liang, R. B.; Voth, G. A.; Swanson, J. M. J. Computationally Efficient Multiscale Reactive Molecular Dynamics to Describe Amino Acid Deprotonation in Proteins. *J Chem Theory Comput* **2016**, *12* (2), 879-891. DOI: 10.1021/acs.jctc.5b01109.

(62) Anisimov, V. M.; Lamoureux, G.; Vorobyov, I. V.; Huang, N.; Roux, B.; MacKerell, A. D. Determination of electrostatic parameters for a polarizable force field based on the classical Drude oscillator. *J Chem Theory Comput* **2005**, *1* (1), 153-168. DOI: 10.1021/ct049930p.

(63) Harder, E.; Anisimov, V. M.; Vorobyov, I. V.; Lopes, P. E. M.; Noskov, S. Y.; MacKerell, A. D.; Roux, B. Atomic level anisotropy in the electrostatic modeling of lone pairs for a polarizable force field based on the classical Drude oscillator. *J Chem Theory Comput* **2006**, *2* (6), 1587-1597. DOI: 10.1021/ct600180x.

- (64) Eyring, H. The activated complex in chemical reactions. *J Chem Phys* **1935**, 3 (2), 107-115.  
DOI: 10.1063/1.1749604.
- (65) Chandler, D. *Introduction to modern statistical mechanics*; Oxford University Press, 1987.
- (66) Ghysbrecht, S.; Keller, B. G. Thermal isomerization rates in retinal analogues using Ab-Initio molecular dynamics. *J Comput Chem* **2024**, 45 (16), 1390-1403. DOI: 10.1002/jcc.27332.
- (67) Olsson, M. H. M.; Sondergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J Chem Theory Comput* **2011**, 7 (2), 525-537. DOI: 10.1021/ct100578z.
- (68) Sondergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J Chem Theory Comput* **2011**, 7 (7), 2284-2295. DOI: 10.1021/ct200133y.
- (69) Roux, B.; Allen, T.; Berneche, S.; Im, W. Theoretical and computational models of biological ion channels. *Q Rev Biophys* **2004**, 37 (1), 15-103. DOI: 10.1017/s0033583504003968.
- (70) Kutzner, C.; Grubmuller, H.; de Groot, B. L.; Zachariae, U. Computational electrophysiology: the molecular dynamics of ion channel permeation and selectivity in atomistic detail. *Biophys J* **2011**, 101 (4), 809-817. DOI: 10.1016/j.bpj.2011.06.010.
- (71) Kutzner, C.; Kopfer, D. A.; Machtens, J. P.; de Groot, B. L.; Song, C.; Zachariae, U. Insights into the function of ion channels by computational electrophysiology simulations. *Biochim Biophys Acta* **2016**, 1858 (7 Pt B), 1741-1752. DOI: 10.1016/j.bbamem.2016.02.006.

(72) Flood, E.; Boiteux, C.; Lev, B.; Vorobyov, I.; Allen, T. W. Atomistic Simulations of Membrane Ion Channel Conduction, Gating, and Modulation. *Chem Rev* **2019**, *119* (13), 7737-7832. DOI: 10.1021/acs.chemrev.8b00630.

(73) Suenaga, A.; Komeiji, Y.; Uebayasi, M.; Meguro, T.; Saito, M.; Yamato, I. Computational observation of an ion permeation through a channel protein. *Bioscience Rep* **1998**, *18* (1), 39-48. DOI: 10.1023/A:1022292801256.

(74) Zhong, Q. F.; Husslein, T.; Moore, P. B.; Newns, D. M.; Pattnaik, P.; Klein, M. L. The M2 channel of influenza A virus: a molecular dynamics study. *Febs Lett* **1998**, *434* (3), 265-271. DOI: 10.1016/S0014-5793(98)00988-0.

(75) Aksimentiev, A.; Schulten, K. Imaging  $\alpha$ -hemolysin with molecular dynamics:: Ionic conductance, osmotic permeability, and the electrostatic potential map. *Biophys J* **2005**, *88* (6), 3745-3761. DOI: 10.1529/biophysj.104.058727.

(76) Gumbart, J.; Khalili-Araghi, F.; Sotomayor, M.; Roux, B. Constant electric field simulations of the membrane potential illustrated with simple systems. *Bba-Biomembranes* **2012**, *1818* (2), 294-302. DOI: 10.1016/j.bbamem.2011.09.030.

(77) Roux, B. The Membrane Potential and its Representation by a Constant Electric Field in Computer Simulations. *Biophys J* **2008**, *95* (9), 4205-4216. DOI: 10.1529/biophysj.108.136499.

(78) Roux, B. Influence of the membrane potential on the free energy of an intrinsic protein. *Biophys J* **1997**, *73* (6), 2980-2989. DOI: 10.1016/S0006-3495(97)78327-9.



(79) Castillo, J. P.; Rui, H.; Basilio, D.; Das, A.; Roux, B.; Latorre, R.; Bezanilla, F.; Holmgren, M. Mechanism of potassium ion uptake by the Na<sup>+</sup>/K<sup>+</sup> ATPase. *Nat Commun* **2015**, *6*. DOI: 10.1038/ncomms8622.

(80) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J Mol Graph Model* **1996**, *14* (1), 33-38. DOI: 10.1016/0263-7855(96)00018-5.

(81) Mark, P.; Nilsson, L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J Phys Chem B* **2001**, *105* (43), 24a-24a.

(82) Huang, J.; MacKerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J Comput Chem* **2013**, *34* (25), 2135-2145. DOI: 10.1002/jcc.23354.

(83) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmuller, H.; MacKerell, A. D., Jr. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat Methods* **2017**, *14* (1), 71-73. DOI: 10.1038/nmeth.4067.

(84) Faraldo-Gómez, J. D.; Roux, B. Electrostatics of ion stabilization in a ClC chloride channel homologue from. *J Mol Biol* **2004**, *339* (4), 981-1000. DOI: 10.1016/j.jmb.2004.04.023.

(85) Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Hénin, J.; Jiang, W.; et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J Chem Phys* **2020**, *153* (4). DOI: 10.1063/5.0014475.

(86) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. Software news and updates - CHARNIM-GUI: A web-based graphical user interface for CHARMM. *J Comput Chem* **2008**, *29* (11), 1859-1865. DOI: 10.1002/jcc.20945.

(87) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J Phys Chem B* **2010**, *114* (23), 7830-7843. DOI: 10.1021/jp101759q.

(88) Wu, E. L.; Cheng, X.; Jo, S.; Rui, H.; Song, K. C.; Dávila-Contreras, E. M.; Qi, Y. F.; Lee, J. M.; Monje-Galvan, V.; Venable, R. M.; et al. CHARMM-GUI Toward Realistic Biological Membrane Simulations. *J Comput Chem* **2014**, *35* (27), 1997-2004. DOI: 10.1002/jcc.23702.

(89) Kräutler, V.; Van Gunsteren, W. F.; Hünenberger, P. H. A fast SHAKE:: Algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J Comput Chem* **2001**, *22* (5), 501-508.

(90) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald - an N.Log(N) Method for Ewald Sums in Large Systems. *J Chem Phys* **1993**, *98* (12), 10089-10092. DOI: 10.1063/1.464397.

(91) Wan, S. Z.; Sinclair, R. C.; Coveney, P. V. Uncertainty quantification in classical molecular dynamics. *Philos T R Soc A* **2021**, *379* (2197). DOI: 10.1098/rsta.2020.0082.

(92) Miao, Y. L.; McCammon, J. A. Unconstrained enhanced sampling for free energy calculations of biomolecules: a review. *Mol Simulat* **2016**, *42* (13), 1046-1055. DOI: 10.1080/08927022.2015.1121541.

(93) Hansen, N.; van Gunsteren, W. F. Practical Aspects of Free-Energy Calculations: A Review. *J Chem Theory Comput* **2014**, *10* (7), 2632-2647. DOI: 10.1021/ct500161f.

(94) Leimkuhler, B.; Matthews, C. *Molecular Dynamics : With Deterministic and Stochastic Numerical Methods*, 1st ed.; Springer International Publishing : Imprint: Springer,, 2015. DOI: 10.1007/978-3-319-16375-8.

(95) Coveney, P. V.; Wan, S. Z. On the calculation of equilibrium thermodynamic properties from molecular dynamics. *Phys Chem Chem Phys* **2016**, *18* (44), 30236-30240. DOI: 10.1039/c6cp02349e.

(96) van Gunsteren, W. F.; Daura, X.; Hansen, N.; Mark, A. E.; Oostenbrink, C.; Riniker, S.; Smith, L. J. Validation of Molecular Simulation: An Overview of Issues. *Angew Chem Int Edit* **2018**, *57* (4), 884-902. DOI: 10.1002/anie.201702945.

(97) Matera, S.; Schneider, W. F.; Heyden, A.; Savara, A. Progress in Accurate Chemical Kinetic Modeling, Simulations, and Parameter Estimation for Heterogeneous Catalysis. *ACS Catalysis* **2019**, *9* (8), 6624-6647. DOI: 10.1021/acscatal.9b01234.

(98) Wittreich, G. R.; Alexopoulos, K.; Vlachos, D. G. Microkinetic Modeling of Surface Catalysis. In *Handbook of Materials Modeling: Applications: Current and Emerging Materials*, Andreoni, W., Yip, S. Eds.; Springer International Publishing, 2020; pp 1377-1404.

(99) Engelhardt, B.; Frohlich, H.; Kschischo, M. Learning (from) the errors of a systems biology model. *Sci Rep-Uk* **2016**, *6*. DOI: 10.1038/srep20772.

(100) Houska, T.; Kraft, P.; Chamorro-Chavez, A.; Breuer, L. SPOTting Model Parameters Using a Ready-Made Python Package. *PLoS One* **2015**, *10* (12), e0145180. DOI: 10.1371/journal.pone.0145180.

(101) Virtanen, P.; Gommers, R.; Oliphant, T. E.; Haberland, M.; Reddy, T.; Cournapeau, D.; Burovski, E.; Peterson, P.; Weckesser, W.; Bright, J.; et al. SciPy 1.0: fundamental algorithms for scientific computing in Python (vol 33, pg 219, 2020). *Nat Methods* **2020**, *17* (3), 352-352. DOI: 10.1038/s41592-020-0772-5.

(102) Thomas, S. A.; Jin, Y. Reconstructing biological gene regulatory networks: where optimization meets big data. *Evolutionary Intelligence* **2014**, *7* (1), 29-47. DOI: 10.1007/s12065-013-0098-7.

(103) Liu, G.; Liu, Q.; Ma, L.; Shao, Z. Evolutionary Algorithms for Applications of Biological Networks: A Review. Cham, 2021; Springer International Publishing: pp 102-114.

(104) Ashraf, C.; Pfaendtner, J. Assessing the Performance of Various Stochastic Optimization Methods on Chemical Kinetic Modeling of Combustion. *Industrial & Engineering Chemistry Research* **2020**, *59* (43), 19212-19225. DOI: 10.1021/acs.iecr.0c04009.

(105) Liu, Y.; Li, C.; Gupta, M.; Stroud, R. M.; Voth, G. A. Kinetic network modeling with molecular simulation inputs: A proton-coupled phosphate symporter. *Biophys J* **2024**. DOI: 10.1016/j.bpj.2024.03.035.

(106) Bernardi, A.; Swanson, J. M. J. CycFlowDec: A Python module for decomposing flow networks using simple cycles. *SoftwareX* **2021**, *14*. DOI: 10.1016/j.softx.2021.100676.

(107) Minping, Q.; Min, Q. Circulation for recurrent markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* **1982**, *59* (2), 203-210. DOI: 10.1007/BF00531744.

(108) Kalpazidou, S. L. *Cycle representations of Markov processes*; Springer Science & Business Media, 2007.

(109) Lim, H. H.; Miller, C. Intracellular Proton-Transfer Mutants in a CLC Cl<sup>-</sup>/H<sup>+</sup> Exchanger. *J Gen Physiol* **2009**, *133* (2), 131-138. DOI: 10.1085/jgp.200810112.

(110) Geng, Y.; Wang, X.; Magleby, K. L. Lack of negative slope in I-V plots for BK channels at positive potentials in the absence of intracellular blockers. *J Gen Physiol* **2013**, *141* (4), 493-497. DOI: 10.1085/jgp.201210955.

(111) Köpfer, D. A.; Song, C.; Gruene, T.; Sheldrick, G. M.; Zachariae, U.; de Groot, B. L. Ion permeation in K<sup>+</sup> channels occurs by direct Coulomb knock-on. *Science* **2014**, *346* (6207), 352-355. DOI: 10.1126/science.1254840.

(112) Miller, C.; Nguitragool, W. A provisional transport mechanism for a chloride channel-type Cl<sup>-</sup>/H<sup>+</sup> exchanger. *Philos Trans R Soc Lond B Biol Sci* **2009**, *364* (1514), 175-180. DOI: 10.1098/rstb.2008.0138.

For Table of Contents only

