Estimating kinetic mechanisms with prior knowledge I: Linear parameter constraints

Autoosa Salari,* Marco A. Navarro,* Mirela Milescu, and Lorin S. Milescu

Division of Biological Sciences, University of Missouri, Columbia, MO

To understand how ion channels and other proteins function at the molecular and cellular levels, one must decrypt their kinetic mechanisms. Sophisticated algorithms have been developed that can be used to extract kinetic parameters from a variety of experimental data types. However, formulating models that not only explain new data, but are also consistent with existing knowledge, remains a challenge. Here, we present a two-part study describing a mathematical and computational formalism that can be used to enforce prior knowledge into the model using constraints. In this first part, we focus on constraints that enforce explicit linear relationships involving rate constants or other model parameters. We develop a simple, linear algebra–based transformation that can be applied to enforce many types of model properties and assumptions, such as microscopic reversibility, allosteric gating, and equality and inequality parameter relationships. This transformation converts the set of linearly interdependent model parameters into a reduced set of independent parameters, which can be passed to an automated search engine for model optimization. In the companion article, we introduce a complementary method that can be used to enforce arbitrary parameter relationships and any constraints that quantify the behavior of the model under certain conditions. The procedures described in this study can, in principle, be coupled to any of the existing methods for solving molecular kinetics for ion channels or other proteins. These concepts can be used not only to enforce existing knowledge but also to formulate and test new hypotheses.

Journal of General Physiology

INTRODUCTION

Ion channels are highly adapted to perform specific functions in the cell. To give just one example, voltage-gated sodium (Nav) channels have finely tuned kinetic properties that allow neurons and other excitable cells to generate action potentials of specific shape and frequency (Bean, 2007). The properties that enable Nav and other channels to perform such complex and well-calibrated behavior are encoded in the kinetic mechanism, defined as a set of conformational and functional states, interconnected by a network of allowed state transitions that may depend on ligand concentration, membrane potential, or other physical variables (Colquhoun and Hawkes, 1995a,b). To understand how ion channels function, one must decrypt the kinetic mechanism. The same is true for all proteins, from ion channels and receptors to enzymes and molecular motors (Popescu and Auerbach, 2003; Milescu et al., 2006; Müllner et al., 2010; Syed et al., 2010).

A kinetic mechanism can be solved by fitting experimental data with a mathematical model. However, decades of ion channel research have shown that kinetic mechanisms cannot be fully captured by any single type of experiment. Instead, to update or construct a new model, one must fit a comprehensive data collection (Horn and Lange, 1983; Hawkes et al., 1990, 1992; Vandenberg and Bezanilla, 1991; Hoshi et al., 1994; Zagotta et al., 1994a,b; Schoppa and Sigworth, 1998a,b; Rothberg and Magleby, 2000; Milescu et al., 2005), ideally generated by multiple experimental paradigms (Vandenberg and Bezanilla, 1991; Akk et al., 2005; Milescu et al., 2008). For example, we know that Nav and other channels have four voltage sensors that gate with different timing and voltage sensitivity (Bezanilla, 2000; Pantazis et al., 2014). These fundamental aspects cannot be easily resolved by single-channel or whole-cell recordings alone, but they can be addressed in combination with other types of experiments, such as patch-clamp fluorometry (Chanda and Bezanilla, 2002; Zaydman et al., 2013; Pantazis et al., 2014).

Optimizing a model against multiple types of data is difficult in itself. A further complication is that some results—quantitative or qualitative—cannot be added to the data collection that is used for fitting. The number of voltage sensors, the existence of open-state block, and numerical relationships between parameters due to allosterism, etc., are examples of such results. Instead, this prior knowledge about the channel must be encoded directly into the model. In this way, the model will explain the new data but will also remain consistent with what is already known.

How do we introduce prior knowledge into a model? We present here some strategies for addressing this issue. At the most basic level, structural assumptions

*A. Salari and M.A. Navarro contributed equally to this paper.
Correspondence to Lorin S. Milescu: MilescuL@missouri.edu

The Rockefeller University Press
J. Gen. Physiol. 2018 Vol. 150 No. 2 323–338
https://doi.org/10.1085/jgp.201711911

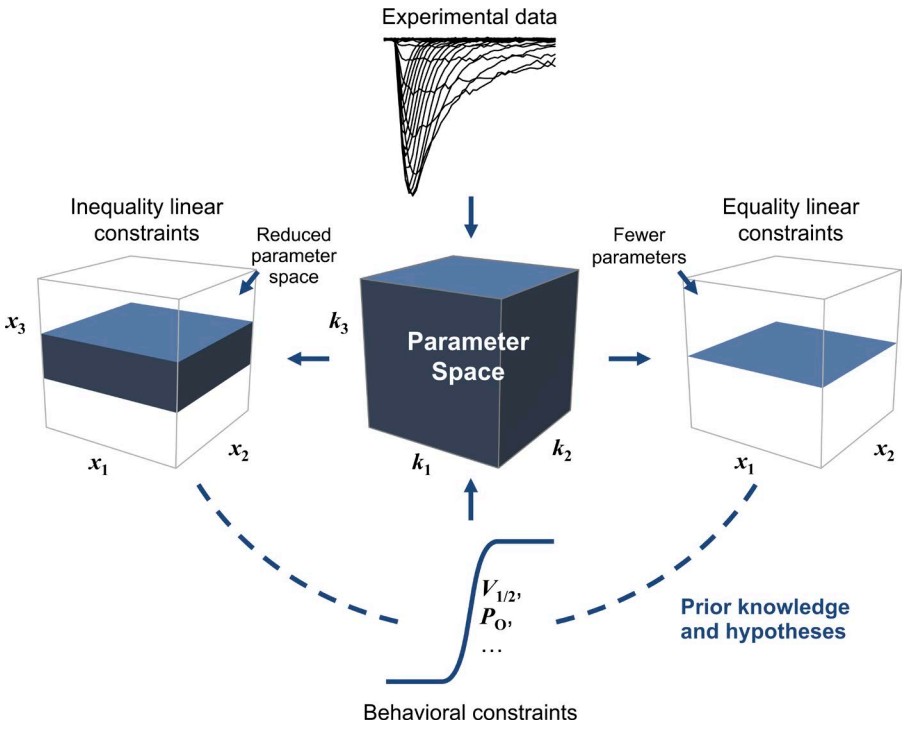

Figure 1. **Estimating kinetic mechanisms with prior knowledge.** A model can be made to fit experimental data while also satisfying user-defined constraints that establish explicit relationships between model parameters or that define specific model behaviors. In the absence of constraints, a fitting algorithm will search a potentially large parameter space to find a solution that best explains the data. Adding constraints to the model not only enforces prior knowledge but also accelerates the fitting procedure by narrowing the search space and reducing the number of free parameters. Furthermore, constraints can be used as a mechanism for testing hypotheses against experimental data.

about the kinetic mechanism can be stated implicitly by choosing a specific set of states and connectivity, as we explain with an example from the literature (Kuo and Bean, 1994). Further, quantitative or qualitative assumptions can be introduced by defining a set of constraints that the model has to satisfy, while also explaining the new data. These model constraints can be formulated as explicit mathematical relationships between rate constants or other model parameters, or they can specify the behavior of the model under certain conditions (Fig. 1).

To implement these ideas, we developed new computational tools with a focus on parameter estimation. First, we build upon an existing method for enforcing linear constraints between rate constants (Qin et al., 1996; Colquhoun et al., 2004; Milescu et al., 2005) and extend it to cover arbitrary linear constraints between model parameters, including allosteric factors. Furthermore, we provide a new formalism for handling both equality and inequality relationships. In the companion article (see Navarro et al. in this issue), we test a method for implementing behavioral constraints, as well as arbitrary parameter relationships, by adding a penalty term to the cost function of the fitting algorithm. The theory and computational procedures described here can be coupled, in principle, to any of the existing methods for solving molecular kinetics, for ion channels or other proteins. These concepts can be used not only to enforce existing knowledge but also to formulate and test new hypotheses.

MATERIALS AND METHODS

All the mathematical and computational algorithms described in this study were implemented and tested with the freely available MLab edition of the QuB program.

Theoretical background

Kinetic mechanisms. Ion channel kinetic mechanisms are well described by Markov models, which reduce the continuum of molecular conformations that can be assumed by the protein to a small set of discrete states that can be detected experimentally or inferred statistically (Colquhoun and Hawkes, 1995a,b; Colquhoun and Sigworth, 1995). These states correspond to various conformations of functional and structural elements, such as resting or activated voltage sensors, bound or unbound ligands, closed or open pore, and inactivated or noninactivated channel. Direct transitions are permitted between certain states, and the frequency of these transitions is quantified by rate constants, which can be functions of ligand concentration, membrane potential, tension, or other physical variables. The topology (or structure) of a kinetic mechanism is defined by the set of states and their transition connectivity, including information on which rates are ligand dependent, voltage dependent, etc.

Here, we assume that all microscopic rate constants follow the Eyring formalism (Eq. 1; Eyring, 1935), with the implication that complexity in kinetic behavior should be explained with more elaborate state mod-

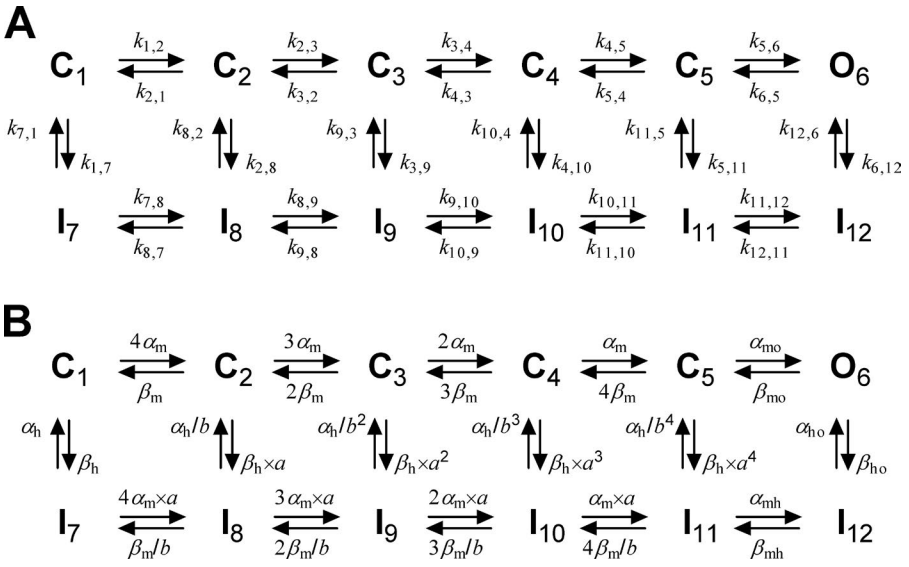

Figure 2. **Expressing prior knowledge via model topology and parameter relationships. (A)** An example model that captures the kinetic properties of neuronal sodium channels (Kuo and Bean, 1994; Milescu et al., 2010). **(B)** Various assumptions about the structural and functional elements of the channel are contained in the structure of the model (the states and their connectivity) and in the quantitative relationships between rate constants. The parameter constraints resulting from these assumptions are explained in the text. The α and β quantities are voltage-dependent rate constants, whereas *a* and *b* are multiplicative factors expressing allosteric relationships. State labels denote closed (C), closed and inactivated (I), and open (O) states.

els, rather than through over-parameterized and ad hoc rate constant expressions. Accordingly, voltage-dependent rate constants are simple exponential functions of voltage:

$$k_{ij} = k_{ij}^0 \times e^{k_{ij}^1 \times V}, \quad (1)$$

where k_{ij} is the rate constant of the transition from state i to state j, and V is the membrane potential. The k_{ij}^0 value is the rate constant at zero membrane potential, whereas the k_{ij}^1 value is the voltage-sensitivity factor, which can be expanded as follows:

$$k_{ij}^1 = (\delta_{ij} \times z_{ij} \times F)/(R \times T), \quad (2)$$

where z_{ij} is the electrical charge moving over the fraction δ_{ij} of the electric field, F is Faraday's constant, R is the gas constant, and T is the absolute temperature. For voltage-insensitive rates, $k_{ij}^1 = 0$. Rate constants that depend on other physical variables, such as membrane tension, have similar exponential expressions (Gnanasambandam et al., 2017). For state transitions that represent the binding of a ligand, rate constants have the following expression:

$$k_{ij} = k_{ij}^0 \times [L], \quad (3)$$

where k_{ij}^0 is the rate constant at unitary ligand concentration $[L]$. When a model lumps several states together, some rates become macroscopic and contain a statistical factor in their expression (e.g., the transition between C_1 and C_2 in the model shown in Fig. 2 B).

The set of k_{ij}^0 and k_{ij}^1 values are the main parameters of the kinetic mechanism. Together, the kinetic parameters and the topology of the model fully specify the mechanism. In turn, the kinetic mechanism describes the operation of the channel within the membrane, under stationary conditions or in response to stimuli. Markov models, computational algorithms, and software have been adapted and developed to extract the kinetic mechanism from experimental data (Ball and Sansom, 1989; Hawkes et al., 1990; Qin et al., 1996, 2000a,b; Venkataramanan and Sigworth, 2002; Celentano and Hawkes, 2004; Qin and Li, 2004; Milescu et al., 2005; Csanády, 2006; Moffatt, 2007; Stepanyuk et al., 2011, 2014), with two interrelated aims: to find an appropriate topology and to estimate the kinetic parameters.

Formulating the topology of the model. The first step in building a kinetic model is to identify a particular topology that defines the structural and functional elements of the channel and their connecting pathways. The topology can be used to specify the number of voltage sensors, the identity of voltage-sensitive transitions, the number of inactivated states, the presence of multiple open states, the existence of allosteric relationships, etc. The model shown in Fig. 2 B illustrates how a topology can be formulated to capture the key features of a Nav channel kinetic mechanism, as detailed in the Results. This model was based on a large body of knowledge accumulated in the field and, not surprisingly, has provided a flexible enough framework that explained voltage-clamp recordings of sodium currents in several neuronal preparations (Kuo and Bean, 1994; Raman and Bean, 2001; Taddese and Bean, 2002; Milescu et al., 2010).

In contrast, when little is known about the channel, one must take a purely data-driven approach and build a parsimonious topology that explains the data reasonably well. Of course, if one wishes a realistic model, the Eyring theory and related concepts must still be obeyed. Some kinetic properties are intuitive enough and can be easily translated into model features (Salari et al., 2016). For example, whole-cell recordings in which the

current first rises and then decays require a model with one conducting and at least two nonconducting states. In general, searching for the right topology can be approached as an iterative process, where one tests a series of models of increasing complexity, adding more and more states and connections. For each tested model, one must determine whether the topology is compatible with the data. If no parameter values can be found that result in a good fit, the topology must be reformulated and the parameters reestimated. Because larger models can inherently fit better, one should take into account the number of degrees of freedom when ranking models. Thus, unless a larger model improves the fit substantially, one should give preference to a model with fewer free parameters. The search across topologies can be terminated when the fit no longer improves.

This model search process is not trivial and relies heavily on the experience of the investigator. The number of nonequivalent (Kienker, 1989) connectivity schemes can be prohibitively large, even for models with a relatively small state count (Bruno et al., 2005). A possible solution is to use a smart optimization algorithm that not only estimates parameters for a given topology but also searches efficiently across topologies at the same time (Gurkiewicz and Korngreen, 2007; Menon et al., 2009). Furthermore, one may be able to reduce the searched state space by using some information contained in the data. For example, statistical analysis of single-channel electrical recordings can provide reasonable estimates on the number of conductance levels (through visual inspection or amplitude histogram analysis) and the minimum number of kinetic states in each conductance level (through dwell-time histogram analysis; Colquhoun and Hawkes, 1982; Hawkes et al., 1990). Other methods can provide more direct evidence about the structural conformations and transition pathways of the channel, such as the number of voltage sensors, the number of inactivated states, or the identity of voltage- or ligand-dependent transitions (Grosman et al., 2000; Ahern et al., 2016). In principle, an automated search across model topologies can incorporate this information.

Parameter estimation. A computational procedure for finding the "best" parameters for a proposed model topology combines an algorithm that measures how well a given model explains the data with an optimization engine that searches the parameter space for the "best" solution (Fletcher, 2013). This optimal solution minimizes the error between the data and the prediction of the model (e.g., the sum of square errors) or maximizes a probability function (e.g., the likelihood that the experimental data were generated by the model or the Bayesian posterior probability; Horn and Lange, 1983; Hawkes et al., 1990; Qin et al., 1996, 2000a; Celentano and Hawkes, 2004; Milescu et al., 2005; Csanády, 2006; Moffatt, 2007; Calderhead et al., 2013; Stepanyuk et al., 2014; Epstein et al., 2016). Intuitively, the first approach can be described as minimizing a "cost function," whereas the second, as maximizing a "goodness of fit." Throughout this study, we will use the "cost function" term, but with the understanding that it could refer to either minimizing the sum of square errors or, equivalently, minimizing the negative log-likelihood. When one also searches for an appropriate topology, the value of the cost function can be used as a score to rank the model. Even though the kinetic mechanism is fully characterized by the k_{ij}^{0} and k_{ij}^{1} parameters, the experimental data typically depend on some other parameters as well, such as the number of active ion channels, the single-channel conductance, and the ionic concentrations. To extract the kinetic mechanism from the data, these other parameters may need to be coestimated (Colquhoun et al., 1996; Qin et al., 2000b; Milescu et al., 2005).

RESULTS

Once a model topology is selected to appropriately express what is known or hypothesized about the channel, thus encoding prior knowledge, the next step is to find a set of parameters that explain the data well. However, these parameters can also contain prior knowledge. In fact, the parameter-estimation procedure itself can be designed to enforce prior knowledge, by making it generate parameter values that are in agreement with a set of model constraints. We classify these model constraints in two categories: (1) parameter constraints, discussed in this study, and (2) behavioral constraints, discussed in the companion article (Navarro et al., 2018). A parameter constraint is formulated as an explicit mathematical relationship that involves rate constants or other model parameters. An example is the scaling of one rate constant to another or restricting the range of a parameter to positive values. In contrast, a behavioral constraint quantifies the behavior of the model under certain conditions, without explicitly referring to rate constants or other model parameters. An example is enforcing the maximum open probability (P_O) reached by the channel during a specific voltage-clamp stimulation protocol.

The mathematical and computational procedures discussed here for enforcing parameter constraints are limited to linear relationships. However, nonlinear relationships can be enforced using the mechanism developed for behavioral constraints, presented in the second part of this study (Navarro et al., 2018). As illustrated in Fig. 1, linear parameter constraints that enforce an equality relationship reduce the dimensionality of the parameter space, eliminating one dimension for each relationship. In contrast, both inequality parameter constraints and behavioral

constraints preserve dimensionality but reduce the size of the parameter space. To describe it intuitively, inequality parameter constraints present the optimizer with a reduced road map, whereas behavioral constraints guide the optimizer through toll-free roads only.

Implementing prior knowledge with linear parameter constraints

In this section, we discuss the implementation of prior knowledge via linear parameter constraints. To illustrate this concept, we use the Nav channel kinetic mechanism shown in Fig. 2 A. We chose this model because it covers many of the parameter constraints that our formalism can enforce. The model was originally formulated (Kuo and Bean, 1994) with several mechanistic assumptions in mind, which are reflected in the number of states and connections, and in the mathematical relationships between various kinetic parameters (Fig. 2 B). These assumptions can be regarded and expressed as parameter constraints.

Model assumptions. The first assumption is that channel activation involves four identical and independent voltage sensors, and all four must be activated to open the pore. Thus, to simplify the kinetic mechanism, all closed states with the same number n of resting sensors are lumped into a single compound state. The result is the five-state activation pathway $C_1 \ldots C_5$. The frequency of activation transitions for any of the compound states $C_1 \ldots C_5$ is equal to n times the frequency of the activation transition for a single sensor, where n is a statistical factor. The same rule applies to deactivation transitions. For example, when the channel resides in a closed state that has three resting voltage sensors (C_2, $n = 3$), the compound activation rate ($k_{2,3}$) is three times the activation rate of a single sensor ($k_{4,5}$ or α_m). Thus, if $k_{4,5}$ and $k_{2,1}$ are the microscopic transition rates of a single sensor activating or deactivating, respectively, the assumption of identical and independent voltage sensors is expressed by the following mathematical relationships, where one rate is scaled to another by a constant factor:

$$\begin{aligned} k_{3,4} &= 2 \times k_{4,5}, \quad k_{2,3} = 3 \times k_{4,5}, \quad k_{1,2} = 4 \times k_{4,5}, \\ k_{3,2} &= 2 \times k_{2,1}, \quad k_{4,3} = 3 \times k_{2,1}, \quad k_{5,4} = 4 \times k_{2,1}. \end{aligned} \tag{4}$$

Any deviation from the condition of identical and independent voltage sensors will require a model with a different number of states, connections, and statistical factors along the activation pathway. In fact, these constant multiplication factors (2, 3, and 4) could be replaced with unknown cooperativity factors, to be determined from the experimental data, similar to the inactivation allosteric factors introduced next. The same principles apply to ligand-gated ion channel mechanisms. In this case, the statistical factors can be used to describe the relationships between the ligand-binding sites.

Another assumption is that the channel can inactivate not only from the open state O_6 but also from any of the $C_1 \ldots C_5$ closed states into the $I_7 \ldots I_{12}$ inactivated states. However, the transition into inactivation from the closed states depends on the degree of activation: as more voltage sensors are activated, the C to I transitions become faster, whereas the I to C transitions become slower. As envisioned in the original model, this property is implemented with the allosteric factors a and b. Thus, the rate of inactivation from a closed state C_n is equal to the rate of inactivation from the previous closed state C_{n-1}, multiplied by the a factor. For example, $k_{2,8} = a \times k_{1,7}$, $k_{3,9} = a \times k_{2,8}$, etc. The opposite is true for the return rates: $k_{8,2} = b^{-1} \times k_{7,1}$, $k_{9,3} = b^{-1} \times k_{8,2}$, etc. Taking $k_{1,7}$ and $k_{7,1}$ as the reference microscopic rates, this assumption is expressed by the following mathematical relationships, where one rate is scaled to another by an undetermined factor:

$$\begin{aligned} &k_{2,8} = a \times k_{1,7}, \quad k_{3,9} = a^2 \times k_{1,7}, \quad k_{4,10} = a^3 \times k_{1,7}, \\ &k_{5,11} = a^4 \times k_{1,7}, \quad k_{8,2} = b^{-1} \times k_{7,1}, \quad k_{9,3} = b^{-2} \times k_{7,1}, \\ &k_{10,4} = b^{-3} \times k_{7,1}, \quad k_{11,5} = b^{-4} \times k_{7,1}. \end{aligned} \tag{5}$$

Furthermore, the voltage sensors can also activate when the channel is inactivated, along the $I_7 \ldots I_{11}$ pathway, but with different kinetics. This is also encoded by the allosteric factors a and b, resulting in another set of similar mathematical relationships:

$$\begin{aligned} &k_{10,11} = a \times k_{4,5}, \quad k_{9,10} = a \times k_{3,4}, \quad k_{8,9} = a \times k_{2,3}, \\ &k_{7,8} = a \times k_{1,2}, \quad k_{8,7} = b^{-1} \times k_{2,1}, \quad k_{9,8} = b^{-1} \times k_{3,2}, \\ &k_{10,9} = b^{-1} \times k_{4,3}, \quad k_{11,10} = b^{-1} \times k_{5,4}. \end{aligned} \tag{6}$$

Overall, this allosteric coupling between activation and inactivation can explain the apparently contradictory findings that inactivation appears strongly voltage sensitive but only minimal electrical charge is detected to move within the channel during inactivation (Armstrong and Bezanilla, 1977). Generally, allosteric factors, such as the a and b quantities in Eqs. 5 and 6, are unknown and need to be determined from the data.

Finally, the last assumption is that the channel satisfies the condition of microscopic reversibility, i.e., no energy input is required for gating and opening. Under this condition, for any reaction loop in the model, the clockwise product of rates around the loop must equal the counterclockwise product (Song and Magleby, 1994; Rothberg and Magleby, 2001; Colquhoun et al., 2004). As the model in Fig. 2 A has five independent loops, the following mathematical relationships must hold true:

$$
\begin{aligned}
&k_{1,2} \times k_{2,8} \times k_{8,7} \times k_{7,1} = \\
&k_{2,1} \times k_{8,2} \times k_{7,8} \times k_{1,7} \ (\text{for the } C_1\text{-}C_2\text{-}I_8\text{-}I_7\text{-}C_1 \text{ loop}), \\
&k_{2,3} \times k_{3,9} \times k_{9,8} \times k_{8,2} = \\
&k_{3,2} \times k_{9,3} \times k_{8,9} \times k_{2,8} \ (C_2\text{-}C_3\text{-}I_9\text{-}I_8\text{-}C_2), \\
&k_{3,4} \times k_{4,10} \times k_{10,9} \times k_{9,3} = \\
&k_{4,3} \times k_{10,4} \times k_{9,10} \times k_{3,9} \ (C_3\text{-}C_4\text{-}I_{10}\text{-}I_9\text{-}C_3), \\
&k_{4,5} \times k_{5,11} \times k_{11,10} \times k_{10,4} = \\
&k_{5,4} \times k_{11,5} \times k_{10,11} \times k_{4,10} \ (C_4\text{-}C_5\text{-}I_{11}\text{-}I_{10}\text{-}C_4), \\
&k_{5,6} \times k_{6,12} \times k_{12,11} \times k_{11,5} = \\
&k_{6,5} \times k_{12,6} \times k_{11,12} \times k_{5,11} \ (C_5\text{-}O_6\text{-}I_{12}\text{-}I_{11}\text{-}C_5).
\end{aligned} \tag{7}
$$

Voltage- and ligand-dependent rate constants. Some of the mathematical relationships used to express parameter constraints may involve rate constants that are functions of membrane potential. Unless otherwise specified, all these mathematical relationships must be true for any membrane potential value. For example, the scaling relationship $k_{3,4} = 2 \times k_{4,5}$ in Eq. 4 can be expanded as follows:

$$
k^0_{3,4} \times \exp\left(k^1_{3,4} \times V\right) = 2 \times k^0_{4,5} \times \exp\left(k^1_{4,5} \times V\right). \tag{8}
$$

A logarithm transformation can be applied to convert products into sums:

$$
\ln\left(k^0_{3,4}\right) + k^1_{3,4} \times V = \ln(2) + \ln\left(k^0_{4,5}\right) + k^1_{4,5} \times V. \tag{9}
$$

Rearranging the terms gives the following:

$$
\ln\left(k^0_{3,4}\right) - \ln\left(k^0_{4,5}\right) + V \times \left(k^1_{3,4} - k^1_{4,5}\right) = \ln(2). \tag{10}
$$

This relationship must be true when $V = 0$, in which case it simplifies to

$$
\ln\left(k^0_{3,4}\right) - \ln\left(k^0_{4,5}\right) = \ln(2). \tag{11}
$$

Using this result in Eq. 10, we obtain the following:

$$
k^1_{3,4} - k^1_{4,5} = 0. \tag{12}
$$

Thus, to enforce a scaling relationship between two voltage-dependent rate constants, the two mathematical relationships above (Eqs. 11 and 12) must be simultaneously satisfied. The same reasoning can be applied to other types of constraints. For example, after taking the logarithm and rearranging the terms, the first loop balance constraint in Eq. 7 becomes

$$
\begin{aligned}
&\ln\left(k^0_{1,2}\right) - \ln\left(k^0_{2,1}\right) + \ln\left(k^0_{2,8}\right) - \ln\left(k^0_{8,2}\right) + \\
&\ln\left(k^0_{8,7}\right) - \ln\left(k^0_{7,8}\right) + \ln\left(k^0_{7,1}\right) - \ln\left(k^0_{1,7}\right) + \\
&V \times \left(k^1_{1,2} - k^1_{2,1} + k^1_{2,8} - k^1_{8,2} + k^1_{8,7} - k^1_{7,8} + k^1_{7,1} - k^1_{1,7}\right) = 0.
\end{aligned} \tag{13}
$$

For Eq. 13 to be true, two mathematical relationships must be simultaneously satisfied:

$$
\begin{aligned}
&\ln\left(k^0_{1,2}\right) - \ln\left(k^0_{2,1}\right) + \ln\left(k^0_{2,8}\right) - \ln\left(k^0_{8,2}\right) + \\
&\ln\left(k^0_{8,7}\right) - \ln\left(k^0_{7,8}\right) + \ln\left(k^0_{7,1}\right) - \ln\left(k^0_{1,7}\right) = 0 \\
&\text{and } k^1_{1,2} - k^1_{2,1} + k^1_{2,8} - k^1_{8,2} + k^1_{8,7} - k^1_{7,8} + k^1_{7,1} - k^1_{1,7} = 0.
\end{aligned} \tag{14}
$$

Some kinetic mechanisms involve state transitions associated with the binding of a ligand (Grosman et al., 2000; Burzomato et al., 2004; Akk et al., 2005). For example, a relationship where one ligand-dependent rate constant k_{ij} is scaled by a constant factor c to another ligand-dependent rate constant k_{kl} can be expanded as follows:

$$
\begin{aligned}
&k^0_{ij} \times [L] = c \times k^0_{kl} \times [L] \Rightarrow \\
&\ln\left(k^0_{ij}\right) + \ln([L]) = \ln(c) + \ln\left(k^0_{kl}\right) + \ln([L]),
\end{aligned} \tag{15}
$$

with the final solution:

$$
\ln\left(k^0_{ij}\right) - \ln\left(k^0_{kl}\right) = \ln(c). \tag{16}
$$

Relationships involving voltage- or ligand-dependent rate constants have some special restrictions that have a simple mathematical provenance: (a) if a rate is scaled to another rate, their voltage sensitivities must be equal (or trivially zero); thus, a voltage-dependent rate cannot be scaled to a voltage-independent rate, except for a single voltage value; (b) if a loop involves only one voltage-dependent transition, then the forward and backward voltage-sensitivity factors for that transition must be equal (or zero); a more typical and useful scenario would require at least two voltage-dependent transitions in the loop; and (c) a mathematical relationship that involves ligand-dependent transitions cannot be satisfied for all ligand concentrations, unless the algebraic sum of all the $\ln([L])$ terms is equal to zero. Thus, a ligand-dependent rate cannot be scaled to a ligand-independent rate, except for a single concentration value. In the case of microscopic reversibility, this condition requires that the clockwise and counterclockwise transitions around the loop involve the same number of ligand-dependent steps. When the channel binds multiple types of ligands, each type must satisfy these conditions. Models formulated without taking these precautions are, in principle, physically unrealistic.

Allosteric, statistical, and other multiplicative factors. Multiplicative factors can be introduced in the rate constant equation to formulate macroscopic rates and to express a variety of parameter constraints. One obvious application is to implement allosteric model behavior, as previously discussed, where the *a* and *b* factors multiply the rate constant pre-exponential term k^0_{ij}. However, multiplicative factors can also be introduced within the exponential in Eq. 1. So far, we lumped the voltage sen-

sitivity as a single factor, k_{ij}^1, but, in fact, we may need to consider the other quantities in Eq. 2. For example, one may want to introduce explicit temperature dependence for a rate. In this case, k_{ij}^1 can be factorized by the following constraint expression:

$$k_{ij}^1 = a_k \times C_k, \tag{17}$$

where a_k is a multiplicative factor that stands for $\delta_{ij} \times z_{ij}$, and C_k is a numerical constant equal to F/RT, as in Eq. 2. This approach would make it possible to mix data collected at different temperatures, in the same way as we can already account for different voltages or ligand concentrations.

As another example, one may want to enforce a relationship between the δ_{ij} and δ_{ji} values for a given transition, such as $\delta_{ij} = 1 - \delta_{ji}$. In this case, assuming that z_{ij} and z_{ji} are known quantities, we would write the following constraint expressions:

$$k_{ij}^1 = a_{k1} \times C_k, \quad k_{ji}^1 = a_{k2} \times C_k, \quad a_{k1} = 1 - a_{k2}, \tag{18}$$

where a_{k1} and a_{k2} are multiplicative factors that stand for δ_{ij} and δ_{ji}, respectively, and C_k is a numerical constant equal to zF/RT, where $z = z_{ij} = z_{ji}$.

As the multiplicative factors are logarithmically transformed, they are subject to some restrictions. Thus, a pre-exponential parameter k_{ij}^0 can only be constrained to an unlimited product of multiplicative factors and other pre-exponential parameters, each raised to an arbitrary power:

$$k_{ij}^0 = C \times \prod_k a_k^{C_k} \times \prod_{m,n} k_{mn}^{0\,C_{mn}}, \tag{19}$$

where C is a positive numerical constant. Taking the logarithm from k_{ij}^0 will convert this product into a linear sum. In contrast, an exponential parameter k_{ij}^1 can only be constrained to an unlimited sum of multiplicative factors and other exponential parameters, each multiplied by an arbitrary numerical constant:

$$k_{ij}^1 = C + \sum_k C_k \times a_k + \sum_{m,n} C_{mn} \times k_{mn}^1, \tag{20}$$

where C is an arbitrary constant. As explained further, a given multiplicative factor can only be used as a pre-exponential-type factor, as in Eq. 19, or as an exponential-type factor, as in Eq. 20, but not as both simultaneously.

Inequality constraints. So far, we have only considered parameter constraints that are formulated as mathematical equalities. However, prior knowledge may also be expressed through inequality parameter constraints. First, there is a physical requirement that all pre-exponential rate parameters k_{ij}^0 must be greater than zero because transition frequencies are positive numbers. Likewise, quantities that multiply rate constants, such as the a and b allosteric factors in Eqs. 5 and 6, must also be restricted to positive values in order to keep rates positive. Both of these constraints are automatically handled by the logarithm transformation of variable, as explained further down. In contrast, the exponential factors k_{ij}^1 are in principle free to take any value in the $-\infty$ to $+\infty$ range, but they can also be restricted by a variety of inequality constraints. For example, we may want the voltage sensitivity parameter to be greater than zero for voltage-sensor activation, and less than zero for deactivation:

$$k_{ij}^1 \geq 0 \quad \text{(activation)}, \quad k_{ji}^1 \leq 0 \quad \text{(deactivation)}. \tag{21}$$

Applying any of these constraints could be a useful working hypothesis during the initial stages of formulating a model. Subsequently, these constraints could be relaxed. In reality, the forward and backward values can both have the same sign: as long as the activation value is more positive than the deactivation value ($k_{ij}^1 \geq k_{ji}^1$), the channel will be more activated at more positive membrane potentials, as is the case with Nav and other voltage-gated channels.

To give another hypothetical example, we may want the ratio between two rate constants at a certain membrane potential V_0 to be smaller than a numerical constant c:

$$\begin{aligned} &k_{ij}/k_{kl} \leq c \Rightarrow \\ &\left[k_{ij}^0 \times \exp\left(k_{ij}^1 \times V_0\right)\right] / \left[k_{kl}^0 \times \exp\left(k_{kl}^1 \times V_0\right)\right] \leq c \Rightarrow \\ &\ln\left(k_{ij}^0\right) - \ln\left(k_{kl}^0\right) + V_0 \times \left(k_{ij}^1 - k_{kl}^1\right) \leq \ln(c). \end{aligned} \tag{22}$$

All of these "≤" or "≥" inequalities can be converted to equality relationships by subtracting or adding, respectively, a positive quantity to the right side of the inequality. For example, in Eq. 22, we can subtract z^2, a quantity that, by definition, is positive:

$$\ln\left(k_{ij}^0\right) - \ln\left(k_{kl}^0\right) + V_0 \times \left(k_{ij}^1 - k_{kl}^1\right) \leq \ln(c) - z^2. \tag{23}$$

As long as the inequality condition in Eq. 23 is satisfied for $z = 0$, we can find a value for z that converts the inequality into an equality:

$$\ln\left(k_{ij}^0\right) - \ln\left(k_{kl}^0\right) + V_0 \times \left(k_{ij}^1 - k_{kl}^1\right) = \ln(c) - z^2. \tag{24}$$

If we want the above ratio between two rate constants to be smaller than a numerical constant at any voltage V, not just at V_0, then we have

$$\ln\left(k_{ij}^0\right) - \ln\left(k_{kl}^0\right) + V \times \left(k_{ij}^1 - k_{kl}^1\right) = \ln(c) - z^2. \tag{25}$$

Because this is an equality, we follow the same logic as for equality constraints: the above relationship must

also be true when $V = 0$, and thus, we obtain two simultaneous relationships:

$$\ln\left(k_{ij}^{0}\right) - \ln\left(k_{kl}^{0}\right) = \ln(c) - z^2, \quad k_{ij}^{1} - k_{kl}^{1} = 0. \tag{26}$$

For "≥" inequalities, we must add z^2 to the right side, rather than subtract it.

In the jargon of optimization theory, z is called a "slack" variable (Fletcher, 2013). With equality constraints, one has to find a set of model parameters that satisfy a set of relationships. When inequalities are added to the model and transformed into equalities using slack variables, one has to find both a set of model parameters and a set of slack variables that together satisfy the constraint relationships. A slack variable is not a true parameter of the model, but merely a variable that is temporarily used to handle inequality constraints during the search for optimum parameters. Very importantly, the quantity added or subtracted via slack variables must take positive values, which is why we use z^2 and not z. The reason for converting inequality relationships to equalities is to have all linear constraints handled by the same linear algebra mathematical formalism, as explained further.

Model parameters. As discussed in the previous section, the core parameters of a kinetic model are k_{ij}^{0} and k_{ij}^{1}, together with some optional multiplicative factors a_k that describe allosteric coupling or other properties (e.g., the *a* and *b* allosteric factors in the model shown in Fig. 2 B) or help parameterizing the rate constants in more detail. However, other parameters q_l may also be added to the modeling framework, depending on the particular application. These external parameters are not necessarily present in any of the rate constant expressions. Instead, they may describe the data or experimental variables. For example, when fitting macroscopic currents, one may also need to estimate the number of channels in the record or the single channel conductance (Milescu et al., 2005). Thus, we define a set **K**, of size N_K, which contains all of these model parameters:

$$\mathbf{K} = \left\{k_{ij}^{0}, k_{ij}^{1}, a_k, q_l\right\}. \tag{27}$$

All of these quantities, which we term "rate constant parameters" (pre-exponential k_{ij}^{0} and exponential k_{ij}^{1}), "multiplicative factors" (a_k), and "external parameters" (q_l), may be involved in the mathematical relationships that express parameter constraints, as discussed further.

A general equation for linear parameter constraints. All the mathematical relationships that were used to implement the assumptions made for the Nav model in Fig. 2 A have something in common: regardless of type (scaling, microscopic reversibility, etc.), each of these equality and inequality parameter constraints result in one or two equations involving $\ln\left(k_{ij}^{0}\right)$, k_{ij}^{1}, a_k, and z^2, each multiplied by a constant. Although not shown in these examples, those relationships can also contain external parameters q_l. Thus, a general form that covers all these examples can be written as follows:

$$\sum_{i,j}\left[C_{ij}^{0} \times \ln\left(k_{ij}^{0}\right)\right] + \sum_{i,j}\left(C_{ij}^{1} \times k_{ij}^{1}\right) + \sum_{k}\left[C_k \times f_k(a_k)\right] + \sum_{l}\left[C_l \times f_l(q_l)\right] = C, \quad \text{(equality)} \tag{28}$$

$$\sum_{i,j}\left[C_{ij}^{0} \times \ln\left(k_{ij}^{0}\right)\right] + \sum_{i,j}\left(C_{ij}^{1} \times k_{ij}^{1}\right) + \sum_{k}\left[C_k \times f_k(a_k)\right] + \sum_{l}\left[C_l \times f_l(q_l)\right] = C + C_z \times z^2, \quad \text{(inequality)} \tag{29}$$

where f_k is an invertible function of the multiplicative factor a_k (e.g., $f_k(a_k) \equiv \ln(a_k)$ or $f_k(a_k) \equiv a_k$), and f_l is an invertible function of the external model parameter q_l. The C_{ij}^{0}, C_{ij}^{1}, C_k, C_l, C, and C_z quantities are numerical constants, with $C_z = 1$ for a "≥" inequality, and $C_z = -1$ for a "≤" inequality.

Specific parameter constraints (e.g., scaling one rate constant to another) can be obtained from the general equation by selecting a subset of $\ln\left(k_{ij}^{0}\right)$, k_{ij}^{1}, $f_k(a_k)$, and $f_l(q_l)$ via nonzero multiplication constants C_{ij}^{0}, C_{ij}^{1}, C_k, or C_l. As discussed throughout the article, a variety of useful constraints can be implemented using this mechanism, such as making a rate equal to a constant, scaling two rates by a constant factor, scaling two rates by a variable factor, constraining the total charge for a set of transitions, enforcing microscopic reversibility, constraining a reaction loop out of microscopic balance, restricting a model parameter to a range, expressing explicit temperature dependence, etc. Some of these constraints will require a single mathematical relationship, whereas others will require two. We note that using multiplicative factors in constraints generally makes sense when the same factor is used in multiple relationships. Otherwise, these factors can be simply calculated after the parameters are estimated.

Converting between model parameters and free parameters. One can easily verify that the model in Fig. 2 B was parameterized in such a way as to implicitly satisfy most assumptions: identical and independent voltage sensors, allosteric coupling of inactivation to activation, and microscopic reversibility. For example, the condition of identical and independent voltage sensors is enforced by the 4, 3, 2, 1 or 1, 2, 3, 4 statistical factors multiplying the α_m or β_m quantities, respectively. In other words, any values can be assigned to the model parameters, α_m^0, α_m^1, *a*, *b*, etc., and the assumptions will be automatically satisfied. There is only one exception: the C_5–O_6–I_{12}–I_{11}–C_5 reaction loop is not implicitly balanced. In this case, the balance equation (i.e., $\alpha_{mo} \times \beta_{ho} \times \beta_{mh} \times b^{-4} \times \alpha_h = \beta_{mo} \times a^4 \times \beta_h \times \alpha_{mh} \times \alpha_{ho}$) is not true by definition. Instead, it must be enforced by choosing an appropriate set of numerical values for all the parame-

ters involved. In contrast, all the other loops are automatically balanced (e.g., $4\alpha_m \times \beta_h \times a \times \beta_m \times b^{-1} \times \alpha_h = \beta_m \times \beta_h \times 4\alpha_m \times a \times \alpha_h \times b^{-1}$).

To formulate a parameterization that implicitly satisfies all parameter constraints, a commonly used strategy is to identify a subset of independent model parameters that can be estimated by the optimization engine and a subset of dependent parameters that can be simply derived from the independent ones. This is exactly how the model in Fig. 2 B was formulated. However, finding this parameterization is not trivial in some cases (Colquhoun et al., 2004). Moreover, it is not clear to us how constraints that are defined by inequality relationships would be handled by this type of parameterization. A potentially easier and certainly more flexible strategy is to define the constraints as an invertible transformation f_c between the set of interdependent model parameters **K** and a set of independent or "free" parameters **X**:

$$\mathbf{X} = f_C(\mathbf{K}), \quad \mathbf{K} = f_C^{-1}(\mathbf{X}). \tag{30}$$

Thus, the model is defined by the **K** parameters, which are interdependent and thus cannot take arbitrary values but only those values that satisfy the user-defined parameter constraints. In contrast, the **X** parameters are independent of each other and are "free" to take any value in the $-\infty$ to $+\infty$ range. We emphasize that the **X** parameters are not simply a subset of **K**, as explained in the following paragraphs. These free parameters are passed to a model-blind optimizer that can search without any constraint in the parameter space defined by **X**, where it finds a solution that best explains the data. This optimal solution can be translated from the free parameter space back into the model parameter space, via the f_C^{-1} transformation.

If we want to implement the linear parameter constraints defined by Eq. 28 or 29, how do we define the f_C and f_C^{-1} transformations that translate the model parameters **K** into the free parameters **X** and vice versa? In preparation for this, we need to recognize that the left side of the generalized Eq. 28 or 29 is nonlinear with respect to k_{ij}^0 and a_k, and perhaps to some external parameter q_l. However, we can make the following invertible transformations of variable:

$$\begin{aligned} &\varepsilon_{ij}^0 = \ln(k_{ij}^0), \quad k_{ij}^0 = \exp(\varepsilon_{ij}^0), \quad \phi_k = f_k(a_k), \\ &a_k = f_k^{-1}(\phi_k), \quad \varphi_l = f_l(q_l), \quad q_l = f_l^{-1}(\varphi_l). \end{aligned} \tag{31}$$

If the multiplicative factor a_k is an allosteric factor or a similar quantity that multiplies a rate constant k_{ij}, then $f_k(a_k)$ is $\ln(a_k)$, which is invertible. If a_k is a factor that multiplies a voltage-sensitivity parameter k_{ij}^1, then $f_k(a_k)$ is a_k, which is obviously invertible as well. Similar logic applies to the external parameters q_l. For example, if q_l refers to the number of channels, we can also use the logarithm transformation (Milescu et al., 2005). In all cases, the logarithm has two desirable effects: it restricts the variables to positive values, and it scales the parameters to more similar values relative to each other, helping the optimization engine to find the solution.

We can rewrite the generalized Eqs. 28 and 29 with these transformations of variable:

$$\begin{aligned} &\textstyle\sum_{i,j}(C_{ij}^0 \times \varepsilon_{ij}^0) + \sum_{i,j}(C_{ij}^1 \times k_{ij}^1) + \\ &\textstyle\sum_k(C_k \times \phi_k) + \sum_l(C_l \times \varphi_l) = C \quad \text{(equality)}, \end{aligned} \tag{32}$$

$$\begin{aligned} &\textstyle\sum_{i,j}(C_{ij}^0 \times \varepsilon_{ij}^0) + \sum_{i,j}(C_{ij}^1 \times k_{ij}^1) + \\ &\textstyle\sum_k(C_k \times \phi_k) + \sum_l(C_l \times \varphi_l) = C + C_z \times z^2 \quad \text{(inequality)}. \end{aligned} \tag{33}$$

The left side of these equations is now linear with respect to ε_{ij}^0, k_{ij}^1, ϕ_k, and φ_l. Next, we define a vector **R**, of dimension N_R, with elements that correspond to ε_{ij}^0, k_{ij}^1, ϕ_k, and φ_l. For a more intuitive notation, we refer to an element of **R** as r_i, when its type is unspecified, or as r_{ij}^0, r_{ij}^1, r_k, or r_l, when we emphasize its identity as a specific type of model parameter (k_{ij}^0, k_{ij}^1, a_k, or q_l, respectively). **R** has the same size as **K** ($N_R = N_K$). Thus, a parameter constraint is expressed as a linear relationship between the elements r_i of **R**, as follows:

$$\textstyle\sum_i C_i \times r_i = C, \quad \text{(equality)} \tag{34}$$

$$\textstyle\sum_i C_i \times r_i = C + C_z \times z^2, \quad \text{(inequality)} \tag{35}$$

where C_i stands for one of the numerical constants C_{ij}^0, C_{ij}^1, C_k, or C_l, respectively. Then, assuming that we have N_C constraint relationships, we can write the generalized constraint from Eqs. 34 or 35 in a more compact matrix form (Qin et al., 1996; Fletcher, 2013), as follows:

$$\mathbf{M} \times \mathbf{R} = \mathbf{V}, \tag{36}$$

where **M** is a matrix of dimension $N_C \times N_R$, and **V** is a vector of dimension N_C. Each row of **M** corresponds to the numerical constants on the left side of the generalized Eqs. 34 or 35, whereas each element of **V** represents the right side of Eqs. 34 or 35:

$$\begin{pmatrix} \cdots & (C_{ij}^0)_1 & (C_{ij}^1)_1 & \cdots & (C_k)_1 & \cdots & (C_l)_1 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & (C_{ij}^0)_{N_C} & (C_{ij}^1)_{N_C} & \cdots & (C_k)_{N_C} & \cdots & (C_l)_{N_C} & \cdots \end{pmatrix}_{(\mathbf{M})} \times \begin{pmatrix} \cdots \\ \varepsilon_{ij}^0 \\ k_{ij}^1 \\ \cdots \\ \phi_k \\ \cdots \\ \varphi_l \\ \cdots \end{pmatrix}_{(\mathbf{R})} = \begin{pmatrix} (C)_1 \quad \text{or} \quad (C + C_z \times z^2)_1 \\ \cdots \\ (C)_{N_C} \quad \text{or} \quad (C + C_z \times z^2)_{N_C} \end{pmatrix}_{(\mathbf{V})}. \tag{37}$$

Thus, each linear relationship between the ε_{ij}^0, k_{ij}^1, ϕ_k, and φ_l variables is encoded by row c of matrix **M**

and element c of vector $\mathbf{V}$. Eq. 36 encapsulates, in matrix form, all the linear parameter constraints imposed on the model, including both equality and inequality relationships.

Although they linearize the constraint relationships, the transformed model parameters $\mathbf{R}$ are still interdependent through the constraint relationships. How do we remove the interdependence and convert $\mathbf{R}$ into the set of free parameters $\mathbf{X}$? Ignoring, for now, that $\mathbf{V}$ is not a constant vector because it depends (nonlinearly) on the optional slack variables z, we can take advantage of the linear form of the matrix equation $\mathbf{M} \times \mathbf{R} = \mathbf{V}$ and express $\mathbf{R}$ as a linear function of $\mathbf{X}$ (Qin et al., 1996) and vice versa:

$$\mathbf{R} = \mathbf{A} \times \mathbf{X} + \mathbf{B}, \tag{38}$$

$$\mathbf{X} = \mathbf{A}^{-1} \times \mathbf{R}, \tag{39}$$

where the vector $\mathbf{X}$, of dimension $N_X = N_R - N_C$, contains the independent parameters. Note that Eq. 39 is obtained from $\mathbf{X} = \mathbf{A}^{-1} \times (\mathbf{R} - \mathbf{B})$, because $\mathbf{A}^{-1}$ multiplied by $\mathbf{B}$ is equal to a zero vector. The matrix $\mathbf{A}$, of dimension $N_R \times N_X$, and the vector $\mathbf{B}$, of dimension N_R, can be determined from $\mathbf{M}$ and $\mathbf{V}$ using the singular value decomposition (Golub and Reinsch, 1970). First, $\mathbf{M}$ is decomposed as follows:

$$\mathbf{M} = \mathbf{U}_M \times \mathbf{S}_M \times \mathbf{V}_M{}^T, \tag{40}$$

where $\mathbf{U}_M$ is an orthogonal matrix of dimension $N_C \times N_C$, $\mathbf{V}_M$ is an orthogonal matrix of dimension $N_R \times N_R$, and $\mathbf{S}_M$ is a diagonal matrix of dimension $N_C \times N_R$ that contains the singular values of matrix $\mathbf{M}$. Then, $\mathbf{A}$ can be extracted as a submatrix of $\mathbf{V}_M$:

$$\mathbf{A}_{i=1\ldots N_R,\, j=1\ldots N_X} = \mathbf{V}_{M\ i=1\ldots N_R,\, j=N_C\ldots N_R}. \tag{41}$$

The inverse of the $\mathbf{A}$ matrix, $\mathbf{A}^{-1}$, is similarly obtained from $\mathbf{V}_M{}^{-1}$. Because $\mathbf{V}_M$ is orthogonal, $\mathbf{V}_M{}^{-1}$ is simply equal to $\mathbf{V}_M$ transposed ($\mathbf{V}_M{}^T$). Then, $\mathbf{B}$ can be calculated as follows:

$$\mathbf{B} = \mathbf{M}^+ \times \mathbf{V}, \tag{42}$$

where the matrix $\mathbf{M}^+$, of dimension $N_R \times N_C$, is the pseudoinverse of $\mathbf{M}$ and can be calculated as follows:

$$\mathbf{M}^+ = \mathbf{V}_M \times \mathbf{S}_M{}^+ \times \mathbf{U}_M{}^T, \tag{43}$$

where $\mathbf{S}_M{}^+$ is obtained from $\mathbf{S}_M$ by replacing all nonzero diagonal elements (the singular values) with their inverse. With the $\mathbf{A}$ and $\mathbf{B}$ matrices obtained as in Eqs. 41 and 42, we can now calculate $\mathbf{R}$ from $\mathbf{K}$, and then $\mathbf{X}$ from $\mathbf{R}$, and vice versa.

How do we deal with inequality constraints and slack variables? We found a solution that is actually quite simple, although, perhaps, not immediately obvious. First, we define a vector $\mathbf{Z}$, of dimension N_Z equal to the number of inequality constraints. This vector contains all the slack variables, one for each inequality constraint. Then, we define another vector $\bar{\mathbf{X}}$, which is the union of $\mathbf{X}$ and $\mathbf{Z}$:

$$\bar{\mathbf{X}} = \mathbf{X} \cup \mathbf{Z}. \tag{44}$$

The size of $\bar{\mathbf{X}}$ is equal to $N_X + N_Z$. The slack variables $\mathbf{Z}$ are arbitrary and thus independent of each other and of the free parameters $\mathbf{X}$. Hence, the elements of $\bar{\mathbf{X}}$ are also independent of each other and represent the free parameters given to the optimizer. Each time the optimizer tries a new set of free-parameters $\bar{\mathbf{X}}$, the corresponding $\mathbf{Z}$ is used to recalculate $\mathbf{V}$ (Eq. 37), which, in turn, is used to recalculate $\mathbf{B}$ (Eq. 42). The $\mathbf{A}$ matrix remains the same because the coefficient matrix $\mathbf{M}$ contains only constants. Thus, we can calculate the transformed model parameters $\mathbf{R}$ as follows:

$$\mathbf{B}_Z = \mathbf{M}^+ \times \mathbf{V}_Z, \tag{45}$$

$$\mathbf{R} = \mathbf{A} \times \mathbf{X} + \mathbf{B}_Z, \tag{46}$$

where $\mathbf{B}_z$ and $\mathbf{V}_z$ are the $\mathbf{B}$ and $\mathbf{V}$ quantities calculated for a given vector $\mathbf{Z}$.

During optimization, the slack variables in $\mathbf{Z}$ are provided by the optimizer, together with $\mathbf{X}$. However, $\mathbf{Z}$ must be initialized at the beginning of the optimization from a given set of transformed model parameters $\mathbf{R}$ and the appropriate relationships in Eq. 37. Let z_c be the slack variable introduced by the inequality relationship defined by row c of the constraint matrix $\mathbf{M}$. Then, z_c can be calculated with the following equation:

$$\mathbf{M}_c \times \mathbf{R} = C_c + C_{zc} \times z_c{}^2, \tag{47}$$

where $\mathbf{M}_c$ is a vector corresponding to row c of $\mathbf{M}$. This equation has the obvious solution:

$$z_c = \sqrt{\left(\frac{\mathbf{M}_c \times \mathbf{R} - C_c}{C_{zc}}\right)}. \tag{48}$$

The two-way conversion between the model parameters $\mathbf{K}$ and the free parameters $\bar{\mathbf{X}}$ is summarized in Fig. 3.

Redundant constraints. One should take care to prevent redundancy and use only the minimum number of mathematical relationships that are necessary to implement the assumptions of the model. Intuitively, a constraint relationship is redundant if its intended consequence is already enforced by other relationships. With our example model, one could use either the scal-

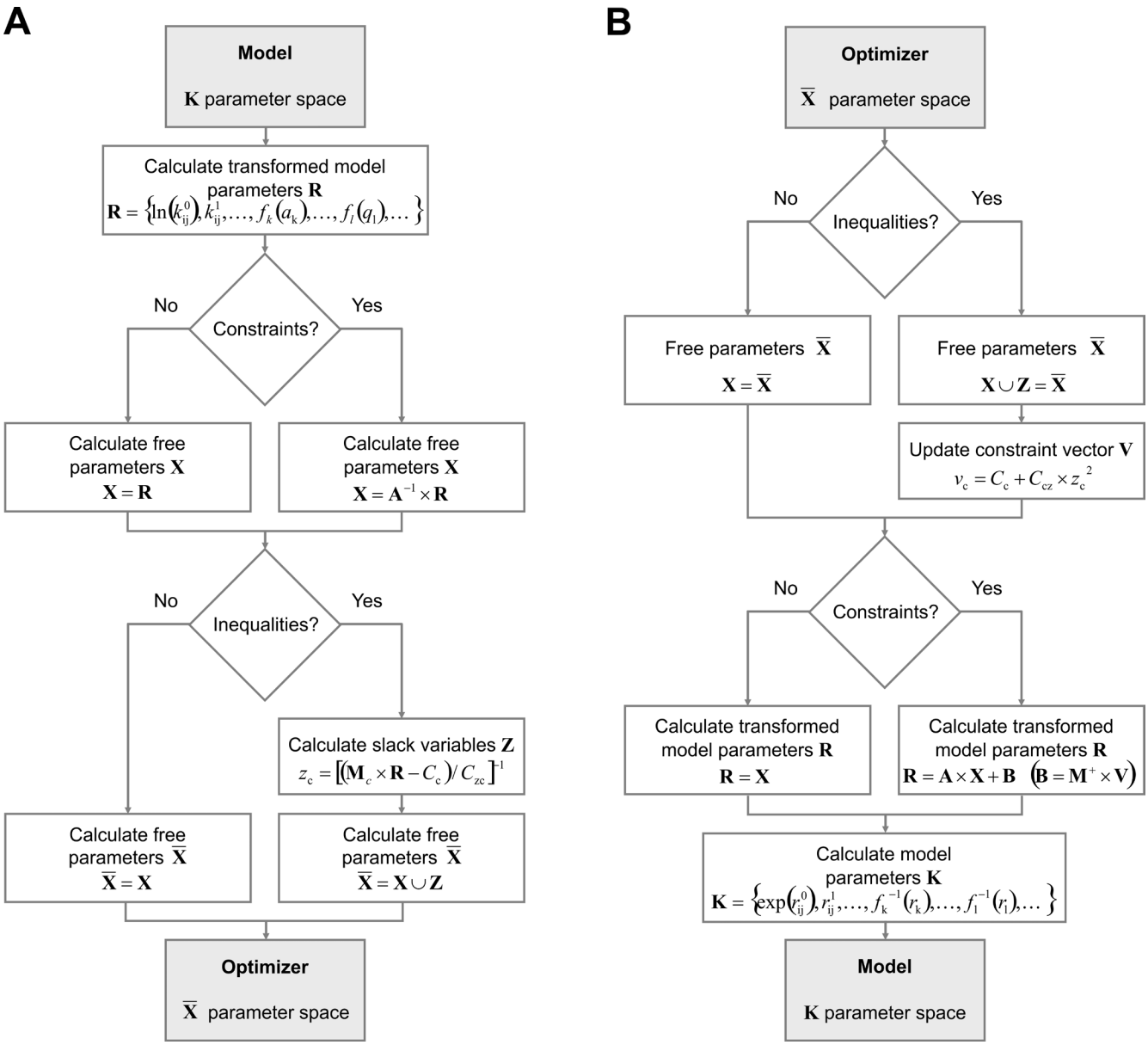

Figure 3. **Transformations between model parameters and free parameters.** The model is defined by a set of interdependent parameters **K**, whereas prior knowledge is expressed as a set of linear parameter constraints. **K** contains pre-exponential and exponential kinetic parameters (k_{ij}^0 and k_{ij}^1), multiplicative factors (a_k), and external parameters (q_l). To enable more types of constraints, **K** is transformed into **R** by applying the logarithm or other functions to some of the parameters in **K**. The linear constraints are reduced via the singular value decomposition to obtain a set of free parameters **X**. Inequality constraints are handled by a set of slack variables **Z**. The constraints reduce the number of free parameters in **X** by one for each mathematical relationship, although each inequality relationship increases the size of **Z** by one. An overall set of free parameters $\bar{\mathbf{X}}$ is formed from **X** (equality constraints only) or from **X** and **Z** (equality and inequality constraints). $\bar{\mathbf{X}}$ is given to the model-blind optimizer to search for an optimal solution, which can be converted back into a set of model parameters **K**. **(A)** Conversion from **K** to $\bar{\mathbf{X}}$. **(B)** Reverse conversion from $\bar{\mathbf{X}}$ to **K**. These conversions can be applied to any kinetic mechanism, regardless of the number of states and connections. All the quantities in the figure are explained in the main text.

ing $k_{7,8} = a \times k_{1,2}$, or the scaling $k_{7,8} = 4 \times a \times k_{4,5}$, but not both because that would create an additional relationship between $k_{1,2}$ and $k_{4,5}$. Similarly, the condition in which the algebraic sum of the voltage sensitivities around a reaction loop is equal to zero may already be enforced by some rate-scaling relationships and should not be duplicated. When in doubt, one could check the rank of the **M** matrix, which will be reduced by redundant constraints.

Redundant constraints may also arise from inequalities. Although tempting, using inequality relationships to enforce a range constraint on a model parameter

is not possible, unfortunately, because it would result in two equality relationships that are redundant. However, this limitation can be overcome through the same mechanism that handles behavioral constraints (Navarro et al., 2018). Furthermore, although inequality constraints add slack variables to the overall set of free parameters, the total number of equality and inequality constraints must still be strictly smaller than the number of model parameters **K**.

Calculating the cost function and its gradients. In a typical scenario, the cost function F is an explicit function of the rate constants k_{ij} and of some external model parameters q_l:

$$F = f(k_{ij}, q_l). \tag{49}$$

Typically, F would not depend explicitly on the multiplicative factors a_k, which are generally used to establish relationships between other parameters. Because the model parameters **K** can be obtained from the free parameters $\bar{\mathbf{X}}$, F can also be written as a function of the free parameters $\bar{\mathbf{X}}$:

$$F = f(\bar{x}_k). \tag{50}$$

Thus, the optimization algorithm can be model blind. In other words, even though it eventually generates a set of optimum model parameters $\mathbf{K}^*$, it actually searches for a solution in the space defined by the free parameters $\bar{\mathbf{X}}$. As it searches for the solution, the optimizer requires the cost function F to be calculated for each proposed $\bar{\mathbf{X}}$. The specific expression of the cost function F depends on the type of application; it could be a sum of square errors, a likelihood, or a Bayesian posterior probability or it could be a mixture of these expressions, when multiple types of data are bundled together (e.g., single-channel and whole-cell traces).

The optimizer may also require the gradients of F with respect to the free parameters $\bar{\mathbf{X}}$, as in the case of gradient descent optimization methods (Fletcher, 2013). These gradients can be calculated by numerical approximations, but analytical calculations are more accurate and may actually be faster in some instances. To calculate the gradient of F with respect to a free parameter $\bar{x}_k$, we have to consider that the rate constants k_{ij} are functions of k_{ij}^0 and k_{ij}^1. In turn, k_{ij}^0 is a function of ε_{ij}^0. We also have to consider that q_l is a function of φ_l. These ε_{ij}^0, k_{ij}^1, and φ_l quantities are entries in the **R** vector and, thus, are explicit functions of a free parameter $\bar{x}_k$. To calculate a gradient, we apply the chain differentiation rule, as follows:

$$\frac{\partial F}{\partial \bar{x}_k} = \sum_{i,j}\left[\frac{\partial F}{\partial k_{ij}} \times \left(\frac{\partial k_{ij}}{\partial k_{ij}^0} \times \frac{\partial k_{ij}^0}{\partial r_m} \times \frac{\partial r_m}{\partial \bar{x}_k} + \frac{\partial k_{ij}}{\partial k_{ij}^1} \times \frac{\partial k_{ij}^1}{\partial r_n} \times \frac{\partial r_n}{\partial \bar{x}_k}\right)\right] + \sum_l\left[\frac{\partial F}{\partial q_l} \times \frac{\partial q_l}{\partial r_p} \times \frac{\partial r_p}{\partial \bar{x}_k}\right]. \tag{51}$$

In Eq. 51, r_m, r_n, and r_p are the elements of the **R** vector that correspond to ε_{ij}^0, k_{ij}^1, and φ_l, respectively. The $\partial F/\partial k_{ij}$ and $\partial F/\partial q_l$ quantities depend on the specific application, e.g., the maximum interval likelihood (Qin et al., 1996) or the maximum point likelihood of single-channel data (Qin et al., 2000a) or the maximum likelihood of macroscopic currents (Milescu et al., 2005). The other partial derivatives can be calculated as follows:

$$\frac{\partial k_{ij}}{\partial k_{ij}^0} = \frac{k_{ij}}{k_{ij}^0}, \quad \frac{\partial k_{ij}^0}{\partial r_m} = k_{ij}^0, \quad \frac{\partial k_{ij}}{\partial k_{ij}^1} = k_{ij} \times V, \quad \frac{\partial k_{ij}^1}{\partial r_n} = 1. \tag{52}$$

The $\partial q_l/\partial r_p$ partial derivative depends on the specific transformation between q_l and φ_l, as illustrated in Eq. 53 for the logarithm and identity transformations:

$$\begin{aligned} &\frac{\partial q_l}{\partial r_p} = q_l \quad \text{if } \varphi_l = \ln(q_l), \\ &\frac{\partial q_l}{\partial r_p} = 1 \quad \text{if } \varphi_l = q_l. \end{aligned} \tag{53}$$

Finally, the partial derivative of any r_i with respect to any $\bar{x}_k$ takes the following form:

$$\begin{aligned} &\frac{\partial r_i}{\partial \bar{x}_k} = a_{i,k} \quad \text{if } \bar{x}_k \in \mathbf{X}, \\ &\frac{\partial r_i}{\partial \bar{x}_k} = 2 \times m_{i,c}^+ \times \bar{x}_k \quad \text{if } \bar{x}_k \in \mathbf{Z} \text{ and "}\geq\text{"}, \\ &\frac{\partial r_i}{\partial \bar{x}_k} = 2 \times m_{i,c}^+ \times \bar{x}_k \quad \text{if } \bar{x}_k \in \mathbf{Z} \text{ and "}\leq\text{"}. \end{aligned} \tag{54}$$

where $a_{i,k}$ and $m_{i,c}^+$ are elements in the **A** and $\mathbf{M}^+$ matrices, respectively. In the last two expressions of Eq. 54, the c subscript is the index of the inequality constraint relationship that uses $\bar{x}_k$ as the slack variable (the index in **V**; Eq. 45).

Using all these quantities, the overall analytical derivative of F with respect to $\bar{x}_k$ becomes

$$\begin{aligned} &\frac{\partial F}{\partial \bar{x}_k} = \sum_{ij}\left[\frac{\partial F}{\partial k_{ij}} \times k_{ij} \times (a_{m,k} + V \times a_{n,k})\right] + \sum_l\left[\frac{\partial F}{\partial q_l} \times q_l \times a_{p,k}\right] \quad \text{if } \bar{x}_k \in \mathbf{X}, \\ &\frac{\partial F}{\partial \bar{x}_k} = 2 \times \bar{x}_k \times \left\{\begin{array}{l} \sum_{ij}\left[\frac{\partial F}{\partial k_{ij}} \times k_{ij} \times (m_{m,c}^+ + V \times m_{n,c}^+)\right] + \\ \sum_l\left[\frac{\partial F}{\partial q_l} \times q_l \times m_{p,c}^+\right] \end{array}\right\} \quad \text{if } \bar{x}_k \in \mathbf{Z}, \text{ "}\geq\text{"}, \\ &\frac{\partial F}{\partial \bar{x}_k} = -2 \times \bar{x}_k \times \left\{\begin{array}{l} \sum_{ij}\left[\frac{\partial F}{\partial k_{ij}} \times k_{ij} \times (m_{m,c}^+ + V \times m_{n,c}^+)\right] + \\ \sum_l\left[\frac{\partial F}{\partial q_l} \times q_l \times m_{p,c}^+\right] \end{array}\right\} \quad \text{if } \bar{x}_k \in \mathbf{Z}, \text{ "}\leq\text{"}. \end{aligned} \tag{55}$$

Eq. 55 is for the case of $\varphi_l = \ln(q_l)$. The subscripts m, n, and p used for the a and m^+ quantities are equal to the indices in the **R** vector that correspond to ε_{ij}^0, k_{ij}^1, and φ_l, respectively.

Calculating the error of the estimates. When estimating the parameters of a model, it is important to have a measure of confidence in those estimates. The variance of a free parameter estimate measures the curvature of the cost function with respect to that parameter. Intuitively, the variance tells us how much the calculated prediction of the model will change when the value of a free parameter $\bar{x}_k$ is changed by a small amount. We emphasize that this change must be interpreted in the context of the specific data used in the analysis. Thus, parameters estimated with a large variance are generally poorly determined because of insufficient data, whereas a small variance denotes a well-defined parameter.

One could calculate the variance of a free parameter estimate, $Var(\bar{x}_k)$, from the second-order partial derivative of the cost function or could use the variance provided by some optimization engines, as in the case of the Davidon-Fletcher-Powell optimizer (Fletcher, 2013). However, when using the parameter constraints described here, the free parameters $\bar{\mathbf{X}}$ must be converted back to model parameters $\mathbf{K}$, and some transformation must be applied to the variance. Thus, the variance of a model parameter, $Var(k_i)$, can be calculated using the following approximation (Qin et al., 2000a; Milescu et al., 2005):

$$Var(k_i) = \sum_p \left[Var(\bar{x}_p) \times \left(\frac{\partial k_i}{\partial \bar{x}_p} \right)^2 \right], \tag{56}$$

where $\bar{x}_p$ is a free parameter in $\bar{\mathbf{X}}$. To calculate the variance for each type of model parameter (k_{ij}^0, k_{ij}^1, a_k, and q_l), we use the chain differentiation rule. For rate constant parameters k_{ij}^0 and k_{ij}^1, we obtain the following:

$$\begin{aligned} Var(k_{ij}^0) &= \sum_p \left[Var(\bar{x}_p) \times \left(k_{ij}^0 \times \frac{\partial r_{ij}^0}{\partial \bar{x}_p} \right)^2 \right], \\ Var(k_{ij}^1) &= \sum_p \left[Var(\bar{x}_p) \times \left(\frac{\partial r_{ij}^1}{\partial \bar{x}_p} \right)^2 \right]. \end{aligned} \tag{57}$$

For pre-exponential and exponential multiplicative factors a_k, we have the following expressions:

$$\begin{aligned} Var(a_k) &= \sum_p \left[Var(\bar{x}_p) \times \left(a_k \times \frac{\partial r_k}{\partial \bar{x}_p} \right)^2 \right] \\ &\text{if } a_k \text{ is pre-exponential;} \\ Var(a_k) &= \sum_p \left[Var(\bar{x}_p) \times \left(\frac{\partial r_k}{\partial \bar{x}_p} \right)^2 \right] \\ &\text{if } a_k \text{ is exponential.} \end{aligned} \tag{58}$$

Finally, for external parameters q_l, the expression depends on the transformation function. For the logarithm and the identity function, we have the following expressions:

$$\begin{aligned} Var(q_l) &= \sum_p \left[Var(\bar{x}_p) \times \left(q_l \times \frac{\partial r_l}{\partial \bar{x}_p} \right)^2 \right] \\ &\text{for logarithm transformation;} \\ Var(q_l) &= \sum_p \left[Var(\bar{x}_p) \times \left(\frac{\partial r_l}{\partial \bar{x}_p} \right)^2 \right] \\ &\text{for identity transformation.} \end{aligned} \tag{59}$$

In Eqs. 56, 57, 58, and 59, the partial derivative of r (r_{ij}^0, r_{ij}^1, r_k, and r_l) with respect to $\bar{x}_p$ is calculated as in Eq. 54, depending on whether $\bar{x}_p$ is an element of $\mathbf{X}$ or a slack variable in $\mathbf{Z}$.

DISCUSSION

Enforcing prior knowledge when fitting new data is not trivial, and one reason is that prior knowledge may take different forms. For example, it can be a linear mathematical relationship between two sequential, ligand-binding transitions or it can describe the dynamics of the channel during complex episodes of action-potential firing. The first example could be easily handled through model parameterization: the independent parameters are identified by the user and passed on to a search engine, whereas the remaining (dependent) parameters are simply derived from the first set, whenever necessary. However, a more elegant and flexible solution, in our opinion, is the method of reduction (Fletcher, 2013), first applied to kinetic modeling algorithms some 20 years ago (Qin et al., 1996, 2000a; Milescu et al., 2005). However, even the reduction method, despite its reach, is not the universal solution to enforcing prior knowledge. Although very powerful, this method is limited to constraints that can be formulated as explicit linear equality relationships between model parameters. Thus, it cannot handle inequalities, nonlinear relationships, and implicit constraints that describe a model property or behavior, such as the maximum open state occupancy during an action potential.

In this two-part study, we proposed a comprehensive set of mathematical and computational tools that address all these limitations and greatly expand the range of prior knowledge that can be enforced. First, as described in this article, we enhanced the reduction method to handle both equality and inequality linear parameter constraints. Furthermore, we expanded the range of parameters that can be constrained to include not only rate constant parameters but also allosteric and other similar factors and external parameters that describe the data or the experiment. Any relationship between these parameters can now be enforced, as long as it is linear. Second, as described in the companion article (Navarro et al., 2018), any other types of model constraints, such as range constraints, nonlinear parameter relationships, or model properties and behavior, are handled by applying a penalty to the cost function. To-

gether, the reduction method and the penalty method can handle virtually any type of model constraint that is likely to be encountered in the field.

The constraining methods described here and in the companion article are available through the freely available QuB software, as maintained by our laboratory. These methods are also easy to implement by interested readers. The only high-level mathematical operation involved is the singular value decomposition, which is readily available from many free, linear algebra packages. As illustrated in Fig. 3, the code can be implemented as a pair of functions: one that converts a set of interdependent model parameters into a set of free parameters, and a second function that performs the reverse operation. The first function is called only once, when the optimization is started, to initialize the free parameters from the model parameters. Any optimization package has one user-customizable callback function that is called each time the search engine needs the cost function for a given set of parameters. The function that converts free parameters into model parameters can be inserted at the beginning of this callback function. For interested users, a step-by-step numerical example is given in the companion article.

ACKNOWLEDGMENTS

We thank the members of the Milescu laboratories for their constructive comments and suggestions. This work was supported by American Heart Association grants 13SDG16990083 to L.S. Milescu and 13SDG14570024 to M. Milescu and a training grant fellowship from the Graduate Assistance in Areas of National Need Initiative/Department of Education to M.A. Navarro.

The authors declare no competing financial interests.

Author contributions: L.S. Milescu developed the mathematical and computational algorithms and implemented them in software. All authors contributed to designing and testing the algorithms and software and to writing the manuscript.

Richard W. Aldrich served as editor.

Submitted: 28 September 2017
Accepted: 6 December 2017

REFERENCES

Ahern, C.A., J. Payandeh, F. Bosmans, and B. Chanda. 2016. The hitchhiker's guide to the voltage-gated sodium channel galaxy. *J. Gen. Physiol.* 147:1–24. https://doi.org/10.1085/jgp.201511492

Akk, G., L.S. Milescu, and M. Heckmann. 2005. Activation of heteroliganded mouse muscle nicotinic receptors. *J. Physiol.* 564:359–376. https://doi.org/10.1113/jphysiol.2004.078535

Armstrong, C.M., and F. Bezanilla. 1977. Inactivation of the sodium channel. II. Gating current experiments. *J. Gen. Physiol.* 70:567–590. https://doi.org/10.1085/jgp.70.5.567

Ball, F.G., and M.S. Sansom. 1989. Ion-channel gating mechanisms: model identification and parameter estimation from single channel recordings. *Proc. R. Soc. Lond. B Biol. Sci.* 236:385–416. https://doi.org/10.1098/rspb.1989.0029

Bean, B.P. 2007. The action potential in mammalian central neurons. *Nat. Rev. Neurosci.* 8:451–465. https://doi.org/10.1038/nrn2148

Bezanilla, F. 2000. The voltage sensor in voltage-dependent ion channels. *Physiol. Rev.* 80:555–592.

Bruno, W.J., J. Yang, and J.E. Pearson. 2005. Using independent open-to-closed transitions to simplify aggregated Markov models of ion channel gating kinetics. *Proc. Natl. Acad. Sci. USA.* 102:6326–6331. https://doi.org/10.1073/pnas.0409110102

Burzomato, V., M. Beato, P.J. Groot-Kormelink, D. Colquhoun, and L.G. Sivilotti. 2004. Single-channel behavior of heteromeric α1β glycine receptors: an attempt to detect a conformational change before the channel opens. *J. Neurosci.* 24:10924–10940. https://doi.org/10.1523/JNEUROSCI.3424-04.2004

Calderhead, B., M. Epstein, L. Sivilotti, and M. Girolami. 2013. Bayesian approaches for mechanistic ion channel modeling. In silico systems biology. *Methods Mol. Biol.* 1021:247–272. https://doi.org/10.1007/978-1-62703-450-0_13

Celentano, J.J., and A.G. Hawkes. 2004. Use of the covariance matrix in directly fitting kinetic parameters: application to GABAA receptors. *Biophys. J.* 87:276–294. https://doi.org/10.1529/biophysj.103.036632

Chanda, B., and F. Bezanilla. 2002. Changes in voltage dependent fluorescence in domain IV follow other domains in the skeletal muscle sodium channel. *Biophys. J.* 82:174a–174a.

Colquhoun, D., and A.G. Hawkes. 1982. On the stochastic properties of bursts of single ion channel openings and of clusters of bursts. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 300:1–59. https://doi.org/10.1098/rstb.1982.0156

Colquhoun, D., and A.G. Hawkes. 1995a. The Principles of the Stochastic Interpretation of Ion-channel mechanisms. *In* Single-channel recording. B. Sakmann, and E. Neher, editors. Plenum Press, New York. 397–479. https://doi.org/10.1007/978-1-4419-1229-9_18

Colquhoun, D., and A.G. Hawkes. 1995b. A Q-matrix cookbook: how to write only one program to calculate the single-channel and macroscopic predictions for any kinetic mechanism. *In* Single-channel recording. B. Sakmann, and E. Neher, editors. Plenum Press, New York. 589–636. https://doi.org/10.1007/978-1-4419-1229-9_20

Colquhoun, D., and F. Sigworth. 1995. Fitting and statistical analysis of single-channel records. *In* Single-channel recording. B. Sakmann, and E. Neher, editors. Plenum Press, New York. 483–587. https://doi.org/10.1007/978-1-4419-1229-9_19

Colquhoun, D., A.G. Hawkes, and K. Srodzinski. 1996. Joint distributions of apparent open times and shut times of single ion channels and the maximum likelihood fitting of mechanisms. *Philos. Trans. A. Math. Phys. Eng. Sci.* 354:2555–2590. https://doi.org/10.1098/rsta.1996.0115

Colquhoun, D., K.A. Dowsland, M. Beato, and A.J.R. Plested. 2004. How to impose microscopic reversibility in complex reaction mechanisms. *Biophys. J.* 86:3510–3518. https://doi.org/10.1529/biophysj.103.038679

Csanády, L. 2006. Statistical evaluation of ion-channel gating models based on distributions of log-likelihood ratios. *Biophys. J.* 90:3523–3545. https://doi.org/10.1529/biophysj.105.075135

Epstein, M., B. Calderhead, M.A. Girolami, and L.G. Sivilotti. 2016. Bayesian statistical inference in ion-channel models with exact missed event correction. *Biophys. J.* 111:333–348. https://doi.org/10.1016/j.bpj.2016.04.053

Eyring, H. 1935. The activated complex in chemical reactions. *J. Chem. Phys.* 3:107–115. https://doi.org/10.1063/1.1749604

Fletcher, R. 2013. Practical Methods of Optimization. Second edition. John Wiley & Sons, Hoboken, NJ. 436 pp.

Gnanasambandam, R., P. Gottlieb, and F. Sachs. 2017. The kinetics and the permeation properties of PIEZO channels. *In* Current Topics in Membranes. P. Gottlieb, editor. Elsevier Science Publishing Co. Inc., New York. 275–307.

Golub, G.H., and C. Reinsch. 1970. Singular value decomposition and least squares solutions. *Numer. Math.* 14:403–420. https://doi.org/10.1007/BF02163027

Grosman, C., F.N. Salamone, S.M. Sine, and A. Auerbach. 2000. The extracellular linker of muscle acetylcholine receptor channels is a gating control element. *J. Gen. Physiol.* 116:327–340. https://doi.org/10.1085/jgp.116.3.327

Gurkiewicz, M., and A. Korngreen. 2007. A numerical approach to ion channel modelling using whole-cell voltage-clamp recordings and a genetic algorithm. *PLOS Comput. Biol.* 3:e169. https://doi.org/10.1371/journal.pcbi.0030169

Hawkes, A.G., A. Jalali, and D. Colquhoun. 1990. The distributions of the apparent open times and shut times in a single channel record when brief events cannot be detected. *Philos. Trans. A Math. Phys. Eng. Sci.* 332:511–538. https://doi.org/10.1098/rsta.1990.0129

Hawkes, A.G., A. Jalali, and D. Colquhoun. 1992. Asymptotic distributions of apparent open times and shut times in a single channel record allowing for the omission of brief events. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 337:383–404. https://doi.org/10.1098/rstb.1992.0116

Horn, R., and K. Lange. 1983. Estimating kinetic constants from single channel data. *Biophys. J.* 43:207–223. https://doi.org/10.1016/S0006-3495(83)84341-0

Hoshi, T., W.N. Zagotta, and R.W. Aldrich. 1994. Shaker potassium channel gating. I: Transitions near the open state. *J. Gen. Physiol.* 103:249–278. https://doi.org/10.1085/jgp.103.2.249

Kienker, P. 1989. Equivalence of aggregated Markov models of ion-channel gating. *Proc. R. Soc. Lond. B Biol. Sci.* 236:269–309. https://doi.org/10.1098/rspb.1989.0024

Kuo, C.C., and B.P. Bean. 1994. Na+ channels must deactivate to recover from inactivation. *Neuron.* 12:819–829. https://doi.org/10.1016/0896-6273(94)90335-2

Menon, V., N. Spruston, and W.L. Kath. 2009. A state-mutating genetic algorithm to design ion-channel models. *Proc. Natl. Acad. Sci. USA.* 106:16829–16834. https://doi.org/10.1073/pnas.0903766106

Milescu, L.S., G. Akk, and F. Sachs. 2005. Maximum likelihood estimation of ion channel kinetics from macroscopic currents. *Biophys. J.* 88:2494–2515. https://doi.org/10.1529/biophysj.104.053256

Milescu, L.S., A. Yildiz, P.R. Selvin, and F. Sachs. 2006. Maximum likelihood estimation of molecular motor kinetics from staircase dwell-time sequences. *Biophys. J.* 91:1156–1168. https://doi.org/10.1529/biophysj.105.079541

Milescu, L.S., T. Yamanishi, K. Ptak, M.Z. Mogri, and J.C. Smith. 2008. Real-time kinetic modeling of voltage-gated ion channels using dynamic clamp. *Biophys. J.* 95:66–87. https://doi.org/10.1529/biophysj.107.118190

Milescu, L.S., T. Yamanishi, K. Ptak, and J.C. Smith. 2010. Kinetic properties and functional dynamics of sodium channels during repetitive spiking in a slow pacemaker neuron. *J. Neurosci.* 30:12113–12127. https://doi.org/10.1523/JNEUROSCI.0445-10.2010

Moffatt, L. 2007. Estimation of ion channel kinetics from fluctuations of macroscopic currents. *Biophys. J.* 93:74–91. https://doi.org/10.1529/biophysj.106.101212

Müllner, F.E., S. Syed, P.R. Selvin, and F.J. Sigworth. 2010. Improved hidden Markov models for molecular motors, part 1: basic theory. *Biophys. J.* 99:3684–3695. https://doi.org/10.1016/j.bpj.2010.09.067

Navarro, M.A., A. Salari, M. Milescu, and L.S. Milescu. 2018. Estimating kinetic mechanisms with prior knowledge II: Behavioral constraints and numerical tests. *J. Gen. Physiol.* https://doi.org/10.1085/jgp.201711912

Pantazis, A., N. Savalli, D. Sigg, A. Neely, and R. Olcese. 2014. Functional heterogeneity of the four voltage sensors of a human L-type calcium channel. *Proc. Natl. Acad. Sci. USA.* 111:18381–18386. https://doi.org/10.1073/pnas.1411127112

Popescu, G., and A. Auerbach. 2003. Modal gating of NMDA receptors and the shape of their synaptic response. *Nat. Neurosci.* 6:476–483. https://doi.org/10.1038/nn1044

Qin, F., and L. Li. 2004. Model-based fitting of single-channel dwell-time distributions. *Biophys. J.* 87:1657–1671. https://doi.org/10.1529/biophysj.103.037531

Qin, F., A. Auerbach, and F. Sachs. 1996. Estimating single-channel kinetic parameters from idealized patch-clamp data containing missed events. *Biophys. J.* 70:264–280. https://doi.org/10.1016/S0006-3495(96)79568-1

Qin, F., A. Auerbach, and F. Sachs. 2000a. A direct optimization approach to hidden Markov modeling for single channel kinetics. *Biophys. J.* 79:1915–1927. https://doi.org/10.1016/S0006-3495(00)76441-1

Qin, F., A. Auerbach, and F. Sachs. 2000b. Hidden Markov modeling for single channel kinetics with filtering and correlated noise. *Biophys. J.* 79:1928–1944. https://doi.org/10.1016/S0006-3495(00)76442-3

Raman, I.M., and B.P. Bean. 2001. Inactivation and recovery of sodium currents in cerebellar Purkinje neurons: evidence for two mechanisms. *Biophys. J.* 80:729–737. https://doi.org/10.1016/S0006-3495(01)76052-3

Rothberg, B.S., and K.L. Magleby. 2000. Voltage and Ca2+ activation of single large-conductance Ca2+-activated K+ channels described by a two-tiered allosteric gating mechanism. *J. Gen. Physiol.* 116:75–99. https://doi.org/10.1085/jgp.116.1.75

Rothberg, B.S., and K.L. Magleby. 2001. Testing for detailed balance (microscopic reversibility in ion channel gating. *Biophys. J.* 80:3025–3026. https://doi.org/10.1016/S0006-3495(01)76268-6

Salari, A., M.A. Navarro, and L.S. Milescu. 2016. Modeling the kinetic mechanisms of voltage-gated ion channels. *In* Advanced Patch-Clamp Analysis for Neuroscientists. Vol. 113. A. Korngreen, editor. Humana Press, New York. 267–304. https://doi.org/10.1007/978-1-4939-3411-9_13

Schoppa, N.E., and F.J. Sigworth. 1998a. Activation of shaker potassium channels. I. Characterization of voltage-dependent transitions. *J. Gen. Physiol.* 111:271–294. https://doi.org/10.1085/jgp.111.2.271

Schoppa, N.E., and F.J. Sigworth. 1998b. Activation of Shaker potassium channels. II. Kinetics of the V2 mutant channel. *J. Gen. Physiol.* 111:295–311. https://doi.org/10.1085/jgp.111.2.295

Song, L., and K.L. Magleby. 1994. Testing for microscopic reversibility in the gating of maxi K+ channels using two-dimensional dwell-time distributions. *Biophys. J.* 67:91–104. https://doi.org/10.1016/S0006-3495(94)80458-8

Stepanyuk, A.R., A.L. Borisyuk, and P.V. Belan. 2011. Efficient maximum likelihood estimation of kinetic rate constants from macroscopic currents. *PLoS One.* 6:e29731. https://doi.org/10.1371/journal.pone.0029731

Stepanyuk, A., A. Borisyuk, and P. Belan. 2014. Maximum likelihood estimation of biophysical parameters of synaptic receptors from macroscopic currents. *Front. Cell. Neurosci.* 8:303. https://doi.org/10.3389/fncel.2014.00303

Syed, S., F.E. Müllner, P.R. Selvin, and F.J. Sigworth. 2010. Improved hidden Markov models for molecular motors, part 2: extensions

and application to experimental data. *Biophys. J.* 99:3696–3703. https://doi.org/10.1016/j.bpj.2010.09.066

Taddese, A., and B.P. Bean. 2002. Subthreshold sodium current from rapidly inactivating sodium channels drives spontaneous firing of tuberomammillary neurons. *Neuron.* 33:587–600. https://doi.org/10.1016/S0896-6273(02)00574-3

Vandenberg, C.A., and F. Bezanilla. 1991. Single-channel, macroscopic, and gating currents from sodium channels in the squid giant axon. *Biophys. J.* 60:1499–1510. https://doi.org/10.1016/S0006-3495(91)82185-3

Venkataramanan, L., and F.J. Sigworth. 2002. Applying hidden Markov models to the analysis of single ion channel activity. *Biophys. J.* 82:1930–1942. https://doi.org/10.1016/S0006-3495(02)75542-2

Zagotta, W.N., T. Hoshi, and R.W. Aldrich. 1994a. Shaker potassium channel gating. III: Evaluation of kinetic models for activation. *J. Gen. Physiol.* 103:321–362. https://doi.org/10.1085/jgp.103.2.321

Zagotta, W.N., T. Hoshi, J. Dittman, and R.W. Aldrich. 1994b. Shaker potassium channel gating. II: Transitions in the activation pathway. *J. Gen. Physiol.* 103:279–319. https://doi.org/10.1085/jgp.103.2.279

Zaydman, M.A., J.R. Silva, K. Delaloye, Y. Li, H. Liang, H.P. Larsson, J. Shi, and J. Cui. 2013. Kv7.1 ion channels require a lipid to couple voltage sensing to pore opening. *Proc. Natl. Acad. Sci. USA.* 110:13180–13185. https://doi.org/10.1073/pnas.1305167110