



Cite this article: Siekmann I, Fackrell M, Crampin EJ, Taylor P. 2016 Modelling modal gating of ion channels with hierarchical Markov models. *Proc. R. Soc. A* **472**: 20160122. <http://dx.doi.org/10.1098/rspa.2016.0122>

Received: 18 February 2016

Accepted: 22 July 2016

Subject Areas:

applied mathematics, biophysics

Keywords:

ion channels, modal gating, continuous-time hierarchical Markov model, inositol-trisphosphate receptor

Author for correspondence:

Ivo Siekmann

e-mail:

ivo.siekmann@mathematik.uni-goettingen.de

[†]Present address: Felix Bernstein Institute for Mathematical Statistics, Georg August University of Göttingen, Goldschmidtstraße 7, 37077 Göttingen, Germany.

Modelling modal gating of ion channels with hierarchical Markov models

Ivo Siekmann^{1,2,†}, Mark Fackrell³, Edmund J. Crampin^{1,2,3,4,5} and Peter Taylor^{3,6}

¹Systems Biology Laboratory, Melbourne School of Engineering,

²Centre for Systems Genomics, ³School of Mathematics and

Statistics, and ⁴School of Medicine, University of Melbourne, Melbourne, Australia

⁵Australian Research Council Centre of Excellence in Convergent Bio-Nano Science and Technology, Melbourne, Australia

⁶Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers, Melbourne, Australia

 IS, 0000-0002-0706-6307

Many ion channels spontaneously switch between different levels of activity. Although this behaviour known as modal gating has been observed for a long time it is currently not well understood. Despite the fact that appropriately representing activity changes is essential for accurately capturing time course data from ion channels, systematic approaches for modelling modal gating are currently not available. In this paper, we develop a modular approach for building such a model in an iterative process. First, stochastic switching between modes and stochastic opening and closing within modes are represented in separate aggregated Markov models. Second, the continuous-time hierarchical Markov model, a new modelling framework proposed here, then enables us to combine these components so that in the integrated model both mode switching as well as the kinetics within modes are appropriately represented. A mathematical analysis reveals that the behaviour of the hierarchical Markov model naturally depends on the properties of its components. We also demonstrate how a hierarchical Markov model can be parametrized using experimental data and show that it provides a better representation than a previous model of the same dataset. Because evidence

© 2016 The Authors. Published by the Royal Society under the terms of the Creative Commons Attribution License <http://creativecommons.org/licenses/by/4.0/>, which permits unrestricted use, provided the original author and source are credited.

is increasing that modal gating reflects underlying molecular properties of the channel protein, it is likely that biophysical processes are better captured by our new approach than in earlier models.

1. Introduction

Ion channels regulate the flow of ions across the cell membrane by stochastic opening and closing. As soon as it became possible to detect currents generated by the movement of charged ions through the channel via the patch-clamp technique [1], Colquhoun & Hawkes [2] developed the theory of modelling single ion channels with continuous-time Markov models which describe the time-course of opening and closing that is reflected in single-channel currents by stochastic jumps between zero (closed) and one or more small non-zero current levels in the pA range (open). The activity of an ion channel is usually measured by its open probability P_O . But by 1979, Patlak *et al.* [3] had already observed spontaneous changes of channel activity in glutamate-activated channels, shortly afterwards followed by Magleby & Pallotta [4,5], who made similar observations in the calcium-activated potassium channel. Since then this phenomenon, known as modal gating, has been ubiquitously observed across a wide range of ion channels but the significance of modal gating has remained unclear. See Siekmann *et al.* [6] for a more comprehensive review of the experimental literature. Colquhoun & Hawkes [7] modified their general theory from Colquhoun & Hawkes [2] for the analysis of bursts. Bursts are defined as ‘closely spaced openings, separated by longer shut periods’ [7, p. 4], which means that they are related to modes with a high level of activity. Thus, the papers Colquhoun & Hawkes [2,7] contain a comprehensive theory for calculating various statistical properties of the channel kinetics from a given Markov model. However, the problem of constructing models that capture spontaneous changes of channel activity in a systematic way has, so far, not been addressed in the literature.

In this study, we present a general framework for building data-driven models of ion channels that account for modal gating. This is essential for accurately representing the dynamics of an ion channel—instead of producing a misleading constant intermediate open probability P_O , a model should represent the switching between highly different levels of activity characteristic of each mode. This is illustrated in figure 1 where data points labelled M^1 form a segment characterized by a low open probability, whereas the segment labelled M^2 is characterized by a high open probability. In a realistic time series, the changes between M^1 and M^2 occur on a time scale so slow that directly fitting a model (even if they have a sufficient number of open and closed states) to the data will not be able to resolve the infrequent switching between high and low open probabilities but instead will most likely lead to a model with a constant intermediate open probability. Moreover, modes of an ion channel have been associated with biophysical properties of the channel protein [6]. Therefore, a model accounting for modal gating is more likely to appropriately relate the dynamics of ion channels to underlying biophysical states of the channel protein.

Blatz & Magleby [8] presented an early modelling study of three modes observed in a chloride channel. They chose segments representative of an inactive, an active and a flicker mode and went through a thorough model selection process. In this way, they obtained models for each of the three modes. They estimated the order of magnitude of the transitions between these modes and presented a qualitative model structure that illustrates the transitions between the three modes. The model that will be developed here can be regarded as a quantitative development of the idea by Blatz & Magleby [8].

After this early study, modal gating has only rarely been considered in ion channel models. But recently, shortly after the discovery of modal gating in the inositol-trisphosphate receptor (IP₃R) by Ionescu *et al.* [9]—an observation that has been received with great interest in the IP₃R community, Mak & Foskett [10]—Ullah *et al.* [11] and Siekmann *et al.* [12] independently proposed two different models that represent modal gating in the IP₃R. Both models are discussed in more detail in §5. The model by Ullah *et al.* [11] has most recently been used for investigating

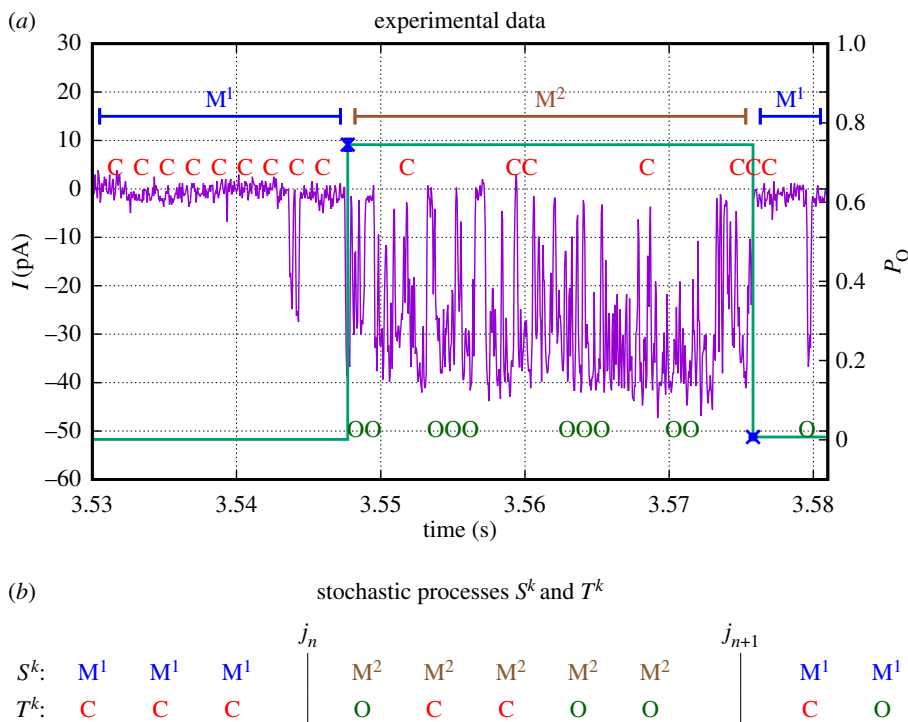


Figure 1. After a statistical analysis of modal gating [6], experimental data are partitioned into segments based on different levels of open probability P_0 by inferring changepoints j_n . For the small section of data shown in (a), the channel spontaneously jumps at $t \approx 3.55$ s from a low P_0 close to zero (M^1) to a high level of activity with $P_0 \approx 75\%$ (M^2). At $t \approx 3.575$ s, the channel leaves the highly active mode M^2 and returns to the low level of activity characteristic for M^1 . Through this segmentation, the original stochastic process T^k of open (O) and closed (C) events has been augmented by the additional information S^k of the mode (M^1, M^2, \dots) that the channel is in for a given point in time. The two coupled stochastic processes S^k and T^k will be represented by the continuous-time hierarchical Markov model developed in this study. (a) Experimental data and (b) stochastic processes S^k and T^k .

the influence of modal gating on calcium puffs [13] and for studying the impact of increased IP₃R activity in Alzheimer's disease [14]. One difficulty in appropriately representing modal gating of ion channels in a model is the fact that for a time series of measurements collected from an ion channel, it is impossible to infer directly in which mode the channel is at a given point in time. However, Siekmann *et al.* [6] have shown how this information can be obtained by statistical changepoint analysis (figure 1). Previously, segments representative of different modes were either selected by visual inspection or by estimating the open probability using moving averages. Ionescu *et al.* [9] presented a heuristic algorithm that segments the data based on an analysis of burst durations and burst-terminating gaps. Siekmann *et al.* [6] detected mode changes by identifying significant changes of the open probability between adjacent segments in a time series recorded from an ion channel. In contrast with previous approaches, the uncertainty of the inferred changepoints where mode switching has supposedly occurred can be comprehensively assessed because Siekmann *et al.* [6] calculated probability distributions for the changepoint locations.

As a result, after this analysis has been carried out, for each point in the time series it is not only known if the channel is open (O) or closed (C), but also—with an associated level of uncertainty calculated by the method—in which of the modes M^1, M^2, \dots the channel is. Previously, we observed stochastic switching between a nearly inactive mode M^1 and a highly active mode M^2 in data from the IP₃R [6]. In this paper, we will represent the stochastic process of switching

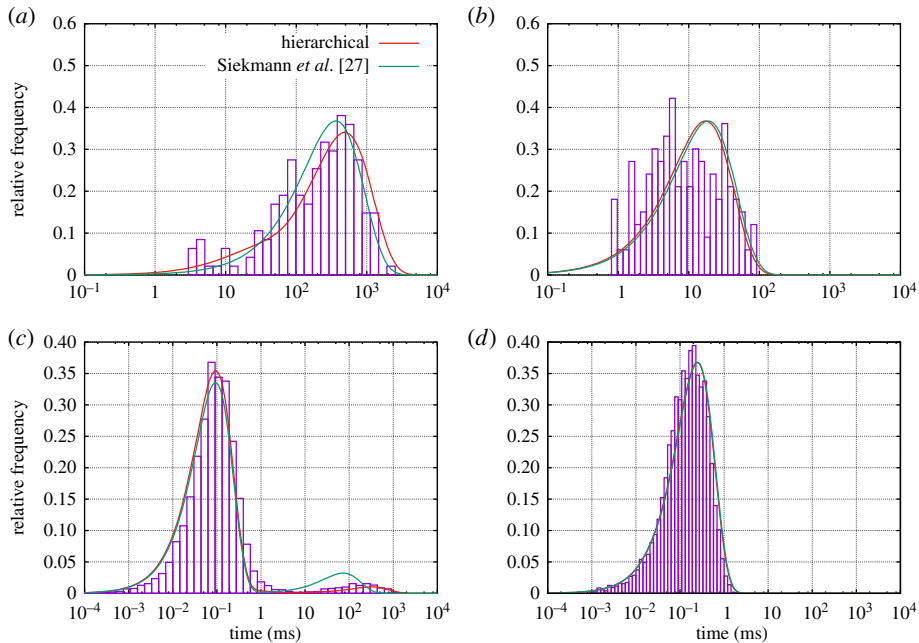


Figure 2. The model from Siekmann *et al.* [12] and the new hierarchical model are compared for a dataset from type I IP₃R for 10 μM IP₃, 5 mM ATP and 0.01 μM Ca²⁺. (a) The fit of the new model to the empirical sojourn time density in mode M¹ (shown in red) is slightly improved in comparison with the original model (shown in green). This improved fit of the modal kinetics clearly improves the fit to the closed time densities shown in (c). (a) Sojourn time density in M¹, (b) sojourn time density in M², (c) closed time density and (d) open time density.

between an arbitrary number of different modes M^i by a continuous-time Markov model with infinitesimal generator \tilde{M} . For data by Wagner & Yule [15], empirical histograms suggest that the sojourn time distribution $f_{M^1}(t)$ within mode M^1 is not exponential (see figs 5 and 6 in Siekmann *et al.* [6] and figures 2a and 5). For this reason, in general, more than one state is needed for accurately representing the process of switching between modes. This means that modal sojourn times are represented by phase-type distributions, a class of distributions which is defined by the time a Markov chain spends in a set of transient states until exiting to an absorbing state [16,17]. We assume that the infinitesimal generator \tilde{M} representing the switching between modes M^i , $i = 1, \dots, n_M$, has the following block structure:

$$\tilde{M} = \begin{pmatrix} \tilde{M}^{1,1} & \tilde{M}^{1,2} & \dots & \tilde{M}^{1,n_M} \\ \tilde{M}^{2,1} & \tilde{M}^{2,2} & \dots & \tilde{M}^{2,n_M} \\ \vdots & & \ddots & \vdots \\ \vdots & & & \vdots \\ \tilde{M}^{n_M,1} & \dots & \dots & \tilde{M}^{n_M,n_M} \end{pmatrix}, \quad (1.1)$$

where the block matrices $\tilde{M}^{i,i} \in \mathbb{R}^{m_i \times m_i}$, $m_i \in \mathbb{N}$, on the diagonal describe transitions between states that represent the same mode M^i , whereas the off-diagonal blocks $\tilde{M}^{i,j} \in \mathbb{R}^{m_i \times m_j}$ represent transitions between states representing different modes M^i and M^j , $i \neq j$. An example for a model for switching between two modes M^1 and M^2 is shown in figure 3a.

Our modal gating analysis illustrated in figure 1 not only enables us to represent the stochastic process of switching *between* modes M^i but by studying the dynamics within representative

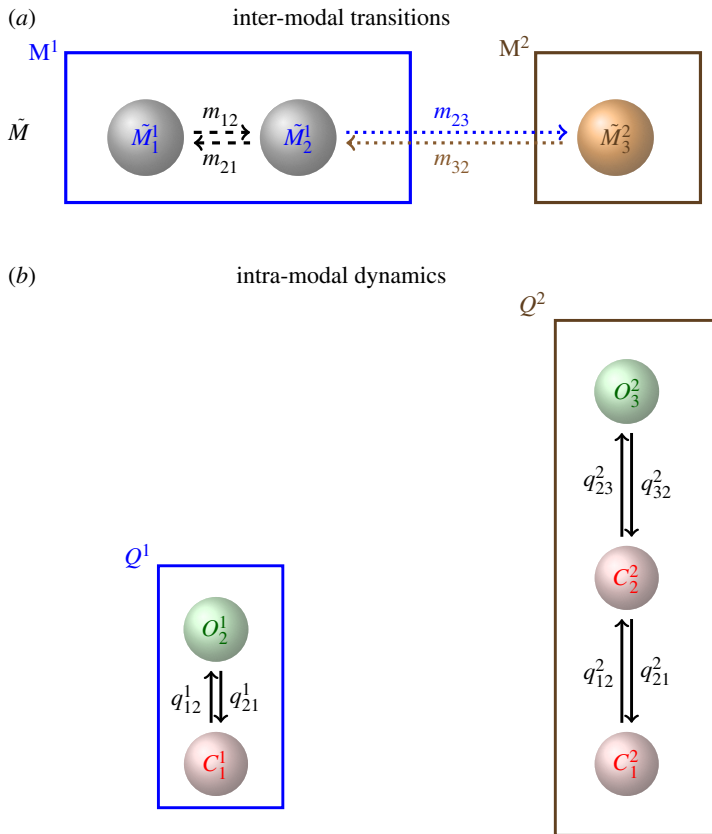


Figure 3. Modular components of a model for modal gating. (a) An example for an aggregated Markov model \tilde{M} representing inter-modal dynamics, the stochastic switching between two modes, M^1 and M^2 . M^1 is modelled by an aggregate of two states, whereas M^2 is represented by one state. The rates m_{23} and m_{32} stand for transitions between both modes. Note that \tilde{M} may, in general, represent transitions between more than two modes, therefore, the states \tilde{M}_j^i are numbered consecutively by subscripts j , whereas the superscripts i indicate the mode M^i . (b) Models Q^1 and Q^2 representing the stochastic opening and closing that is characteristic of mode M^1 or M^2 , respectively. The states C_k^i and O_k^i are numbered similar to the \tilde{M}_j^i . Note that $k = 1, \dots, n_i$ for each mode M^i in contrast with the states \tilde{M}_j^i where the index j runs from 1 to the total number of states. In figure 4, we show how \tilde{M} and the Q^i s are combined in a model that accurately represents both inter-modal transitions as well as intra-modal kinetics. (a) Inter-modal transitions and (b) intra-modal dynamics.

segments we can investigate the processes of stochastic opening and closing characteristic of each mode. For the example in figure 1, the dynamics *within* mode M^2 can be analysed by considering the sequence of open and closed events between j_k and j_{k+1} . The dynamics within a mode M^i can be represented by a Markov model with infinitesimal generator Q^i which is obtained by fitting to representative segments of the same mode [12]. Similar to the sojourn times in the modes M^i , the open and closed time distributions $f_O(t)$ and $f_C(t)$, respectively, are non-exponential and more than one open or closed state may be needed for accurately representing the dynamics. For the example shown in figure 1, we obtain two models with infinitesimal generators Q^1 and Q^2 (figure 3b).

In this paper, we develop a new mathematical model, the continuous-time hierarchical Markov model, that accounts simultaneously for both transitions *between* modes as well as the stochastic opening and closing *within* modes. A hierarchical Markov model in discrete time has been previously described by Fine *et al.* [18] but because we are not aware of a continuous-time version discussed in the literature, we develop the mathematical theory in detail and prove some fundamental properties. For ion channel modelling a continuous-time representation of the

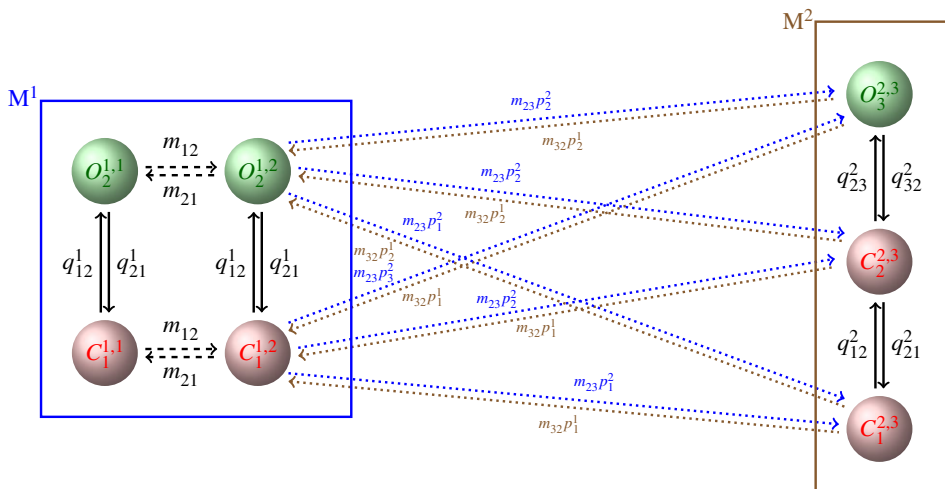


Figure 4. Aggregated Markov model that represents both transitions between modes M^1 and M^2 according to model \tilde{M} (figure 3a) as well as stochastic opening and closing consistent with models Q^1 and Q^2 (figure 3b). The open and closed states are O_k^j and C_k^j , respectively, where the superscripts ij refer to the state \tilde{M}_j^i in the model shown in figure 3a, whereas the subscript k is the index of the state within a model Q^i shown in figure 3b. This illustrates that the state set of the full model is obtained by the Cartesian product of states representing the modes M^i with the states of the model Q^i . Owing to the transitions m_{12} and m_{21} between the two states representing M^1 , in the full model there are two copies of model Q^1 connected by transition rates m_{12} and m_{21} . For transitions between modes it is decided stochastically in which state the target mode is entered. The transitions are determined by initial distributions over the states of the models Q^i . Thus, for our example, we have to choose two stochastic vectors $p^1 = (p_1^1, p_2^1)$ and $p^2 = (p_1^2, p_2^2, p_3^2)$ that give the initial distributions over the states of Q^1 and Q^2 . In order to ensure that the states are indeed entered with the chosen initial distribution, the rates m_{23} exiting M^1 and m_{32} exiting M^2 are weighted with p^1 and p^2 .

dynamics is more appropriate because it is commonly assumed that ion channels are able to make faster transitions than currently resolved by experiments. For the example of modal gating, we assume that switching between modes M^i is a top-level process that regulates the bottom-level process, the opening and closing of the channel characteristic of a particular mode M^i . This is illustrated in figure 3.

The states \tilde{M}_j^i are numbered consecutively by subscripts j , whereas the superscripts i indicate the mode M^i . While the model is in mode M^1 or analogously within one of the states \tilde{M}_1^1 or \tilde{M}_2^1 (figure 3a), its opening and closing is described by the infinitesimal generator Q^1 (figure 3b). As soon as M^1 is left to state \tilde{M}_3^2 , the current state of model Q^1 is vacated and a state of model Q^2 is entered. Now, opening and closing is accounted for by Q^2 until the state \tilde{M}_3^2 and mode M^2 is left and state \tilde{M}_2^2 is entered.

The transitions between modes described via \tilde{M} and the dynamics within modes captured by Q^i illustrated in figure 3 can be represented in a Markov model with infinitesimal generator M that is derived from the individual components \tilde{M} and Q^i . The idea is illustrated in figure 4 and developed formally in §2.

In order to account for the states \tilde{M}_j^i as well as the states O_k^i and C_k^i representing the opening and closing within M^i , the state space of the full model consists of the Cartesian products of the \tilde{M}_j^i with the O_k^i and C_k^i . Thus, the state space of the full model consists of open and closed states O_k^{ij} and C_k^{ij} , respectively, where the superscripts ij refer to the state \tilde{M}_j^i in the model shown in figure 3a, whereas the subscript k is the index of the state within a model Q^i shown in figure 3b. For the example shown in the figure, the closed states $C_1^{1,1}$ and $C_1^{1,2}$ as well as the open states $O_2^{1,1}$ and $O_2^{1,2}$ are connected by the transition rates m_{12} and m_{21} . Because M^1 is modelled by two states

\tilde{M}_1^1 and \tilde{M}_2^1 , two ‘copies’ of Q^1 appear in the full model, whereas there is only one ‘copy’ of Q^2 which is represented by only one state in \tilde{M} .

For transitions between modes, it is decided stochastically in which state the target mode is entered. The transitions are determined by initial distributions over the states of the models Q^i . Thus, for our example, we have to choose two stochastic vectors $p^1 = (p_1^1, p_2^1)$ and $p^2 = (p_1^2, p_2^2, p_3^2)$ that give the initial distributions over the states of Q^1 and Q^2 , respectively. For simplicity, we assume that this initial distribution does not depend on the state from which the transition originates so that—independent of the originating state—each state in the target model is entered with the same probability. In order to ensure that the states of Q^i are indeed entered with the chosen initial distribution, the transition rates have to be ‘split’ accordingly. For our example, the rates m_{23} exiting M^1 and m_{32} exiting M^2 are weighted with the stochastic vectors p^1 and p^2 . The mathematical details of the construction of this model are presented in §2.

It is a strength of our approach that it enables us to build data-driven models of modal gating in a modular way. After segmenting ion channel data with the method by Siekmann *et al.* [6], we obtain a stochastic sequence of events M^i that describes the time course of transitions between different modes. The infinitesimal generators \tilde{M} and the Q^i can then be parametrized from these data. We demonstrate the practical implementation of this approach in §3 using experimental data by Wagner & Yule [15] and compare the results with our previously published model of the same dataset [12].

We investigate the mathematical structure of the continuous-time hierarchical Markov model in more detail in §4. In particular, we show that many important properties of the infinitesimal generator M of the full model can be derived from the generators \tilde{M} and Q^i . We expect that similar to its discrete-time counterpart [18], the continuous-time hierarchical Markov model will have a variety of applications beyond the modelling of modal gating considered here.

We discuss our approach to modal gating in §5. In particular, we explain why our new modelling framework provides a representation of ion channel dynamics that is likely to provide a structure that realistically captures biophysical processes.

2. Material and methods

(a) Preliminaries

We now develop formally the hierarchical Markov model illustrated graphically in figures 3 and 4. First, let us describe the structure of the probability distribution p over the states of the hierarchical Markov model. Let $v = (v^1; v^2; \dots; v^{n_M})$ denote a state probability distribution of the model \tilde{M} . That is, for $i = 1, \dots, n_M$, v^i is the probability distribution of the states in mode M^i . In general, we will allow \tilde{M} to be an aggregated Markov model so that each of the components v^i of the vector v may itself be a vector. With the term aggregated Markov model, we refer to a model where possibly multiple rather than one Markov states are used for representing the same experimental observation. Multiple states of the same aggregate cannot be directly distinguished based on experimental observations. Aggregated Markov models are capable of accounting for observations whose dwell times are distributed according to a mixture of exponentials rather than the exponentially distributed sojourns of single Markov states. We make the convention that components v^i and v^j that are meant to refer to a vector are separated by semicolons, whereas components of a vector are separated by commas. Let us first assume for simplicity that all modes M^i are represented by only one state so that the components v^i are scalars. Then the distribution p over the states of the full model M is a weighting of the distributions w^i over the distributions over the states of the models Q^i . Thus, we obtain $p := (v^1 \cdot w^1; \dots; v^i \cdot w^i; \dots; v^{n_M} \cdot w^{n_M})$. Here ‘ \cdot ’ denotes scalar multiplication of vectors w^i with scalars v^i . If more than one state is needed for representing the modes M^i , we must generalize appropriately the ‘weighting’ of a vector w^i with a vector v^i . Such a generalization is provided by the tensor product ‘ \otimes ’.

Definition 2.1 (Kronecker product \otimes). We will only need the special case of the tensor product for matrices, the Kronecker product. Let $A \in \mathbb{R}^{m \times n}$, $B \in \mathbb{R}^{p \times r}$. Then

$$A \otimes B := (a_{ij} \cdot B)_{1 \leq i \leq m, 1 \leq j \leq n} = \begin{pmatrix} a_{11}B & \dots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \dots & a_{mn}B \end{pmatrix} \in \mathbb{R}^{mp \times nr}. \quad (2.1)$$

The Kronecker product also applies to vectors by identifying column vectors with $(m \times 1)$ - and row vectors with $(1 \times m)$ -matrices.

Definition 2.2 (Kronecker sum \oplus). The Kronecker sum of square matrices $A \in \mathbb{R}^{m \times m}$ and $B \in \mathbb{R}^{n \times n}$ is

$$A \oplus B := A \otimes \text{id}_n + \text{id}_m \otimes B \in \mathbb{R}^{mn \times mn}, \quad (2.2)$$

where id_m and id_n are the identity matrices of the respective dimensions.

For some properties of Kronecker product and sum that we require for our analysis of the hierarchical Markov model (§4), we refer to appendix A. For a distribution v over the states of an aggregated Markov model, subvectors that represent the distributions over the states of the same mode M^i can be naturally described by partitions.

Definition 2.3 (Partitioned vectors, multi-indices). A multi-index is any vector $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}^d$. We define the absolute value $|\alpha| = \sum_{i=1}^d \alpha_i$ and denote $\text{dim}(\alpha) = d$ the dimension of α .

A vector v is partitioned by a multi-index α if

$$v_\alpha := (v^1, \dots; v^i, \dots; v^{\text{dim}(\alpha)})$$

and for each i we have $v^i \in \mathbb{R}^{\alpha_i}$.

Selection of the i th partition of v_α is written as

$$v_\alpha(i) = v^i.$$

The vector space of α -partitioned vectors v_α is denoted \mathbb{R}^α .

How distributions p over the states of a hierarchical Markov model relate to distributions over the states of \tilde{M} and Q^i can be clarified by the tensor product of partitioned vector spaces.

Definition 2.4 (Tensor product $\mathbb{R}^m \otimes_{m,n} \mathbb{R}^n$ of d -partitioned vector spaces). Let $m, n \in \mathbb{N}^d$, $v_m \in \mathbb{R}^m$, $w_n \in \mathbb{R}^n$ be d -partitioned vectors. Then the tensor product $u_{m,n}$ of d -partitioned vectors v_m and w_n is defined by

$$u_{m,n} := v_m \otimes_{m,n} w_n := (v^1 \otimes w^1; \dots; v^i \otimes w^i; \dots; v^d \otimes w^d), \quad (2.3)$$

with the component-wise product $m \cdot n$ of m and n . With the tensor product ' $\otimes_{m,n}$ ' we obtain the vector space

$$\mathbb{R}^m \otimes_{m,n} \mathbb{R}^n$$

of the d -partitioned vector spaces \mathbb{R}^m and \mathbb{R}^n .

Remark 2.1. We make some remarks regarding the interpretation of definition 2.4:

- It can be easily verified that ' $\otimes_{m,n}$ ' fulfils the properties of a tensor product on the vector space $\mathbb{R}^m \otimes_{m,n} \mathbb{R}^n$.

— Vectors $u_{m,n} \in \mathbb{R}^m \otimes_{m,n} \mathbb{R}^n$ can be written as linear combinations

$$u_{m,n} = \sum_{k=1}^d \sum_{i=1}^{m_k} \sum_{j=1}^{n_k} a_{ij}^k (v_m^{k,i} \otimes_{m,n} w_n^{k,j}), \quad a_{ij}^k \in \mathbb{R}, \quad (2.4)$$

where $d = \dim m = \dim n$. By choosing bases $\{v_m^{k,i}, i = 1, \dots, m_k\}$, $\{w_n^{k,j}, j = 1, \dots, n_k\}$, we obtain systems of linearly independent vectors

$$v_m^{k,i} = (0; \dots; v_m^{k,i}; \dots; 0) \in \mathbb{R}^m$$

and

$$w_n^{k,j} = (0; \dots; w_n^{k,j}; \dots; 0) \in \mathbb{R}^n.$$

Thus, from (2.4) it is easy to see that

$$\mathbb{R}^m \otimes_{m,n} \mathbb{R}^n \cong \mathbb{R}^{m \cdot n},$$

where $m \cdot n$ again denotes the component-wise product of m and n .

(b) A hierarchical Markov model for modal gating

Based on the block structure (1.1) of \tilde{M} , we now show how a transition matrix for the full model can be calculated from its components $((\tilde{m}_0, \tilde{M}), (p^i, Q^i)_{i=1}^{n_M})$. Let m and n be the multi-indices defined above. The transitions within the modes M^i are represented in the full model by block matrices $M^{i,j} = \tilde{M}^{i,j} \oplus Q^i \in \mathbb{R}^{m_i n_i \times m_i n_i}$. It follows that $\dim M^{i,j} = m_i n_i$. Moreover, we define the matrix of initial conditions for a transition from Q^i to Q^j by

$$P^{i,j} = u_{n_i}^T \otimes p^j = p^j \otimes u_{n_i}^T, \quad (2.5)$$

where the row vector $p^j \in \mathbb{R}^{1 \times n_j}$ is the initial condition for Q^j from definition 2.5, and $u_{n_i}^T \in \mathbb{R}^{n_i \times 1}$ is a column vector of ones. We observe that $P^{i,j} \in \mathbb{R}^{n_i \times n_j}$ so that, for $i \neq j$ we have $M^{i,j} = \tilde{M}^{i,j} \otimes P^{i,j} \in \mathbb{R}^{m_i n_i \times m_j n_j}$. We can now define the components of a continuous-time hierarchical Markov model and calculate its infinitesimal generator:

Definition 2.5 (Components of a continuous-time hierarchical Markov model). A continuous-time hierarchical Markov model (with a two-level hierarchy) is specified by the components $((\tilde{m}_0, \tilde{M}), (p^i, Q^i)_{i=1}^{n_M})$:

- An infinitesimal generator \tilde{M} of a Markov model with initial distribution \tilde{m}_0 with aggregates of states M^i , $i = 1, \dots, n_M$. The M^i are referred to as modes.
- For each mode, a Markov model with infinitesimal generator Q^i and initial distribution p^i .

Then the infinitesimal generator M of the aggregated model for modal gating is calculated as follows:

$$M = \begin{pmatrix} \tilde{M}^{1,1} \oplus Q^1 & \tilde{M}^{1,2} \otimes P^{1,2} & \dots & \tilde{M}^{1,n_M} \otimes P^{1,n_M} \\ \tilde{M}^{2,1} \otimes P^{2,1} & \tilde{M}^{2,2} \oplus Q^2 & \dots & \tilde{M}^{2,n_M} \otimes P^{2,n_M} \\ \vdots & & \ddots & \vdots \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ \tilde{M}^{n_M,1} \otimes P^{n_M,1} & \dots & \dots & \tilde{M}^{n_M,n_M} \oplus Q^{n_M} \end{pmatrix}. \quad (2.6)$$

It is straightforward to generalize this definition recursively to an arbitrary number of hierarchies. From definition 2.4 and (2.3), we know that an arbitrary distribution p over the states of the full model can be represented by a linear combination of tensor products of the form (2.3). We now require for initial distributions that they should arise from a single tensor product of initial distributions over the states of \tilde{M} and initial distributions over the states of the Q^i .

Definition 2.6 (Initial distribution over the states of a hierarchical Markov model). Let v_m be the initial distribution over the states of the top-level model \tilde{M} and w_n , a vector whose components w^i are initial distributions over the states of the models Q^i . Then the initial distribution $p_{m,n}^0$ over the states of the full model M is calculated by the tensor product ' $\otimes_{m,n}$ ' introduced in definition 2.4:

$$p_{m,n}^0 = v_m \otimes_{m,n} w_n = (v^1 \otimes w^1; \dots; v^i \otimes w^i; \dots; v^{mM} \otimes w^{mM}). \quad (2.7)$$

Remark 2.2. We make some remarks regarding the interpretation of definition 2.6:

- Note that whereas v_m is a stochastic vector, w_n is not. It is easy to see that $p_{m,n}^0$ is a stochastic vector.
- Algebraically, definition 2.6 constrains initial distributions to so-called pure tensors which can be written as a single tensor product rather than a linear combination of tensor products.
- Statistically, definition 2.6 says that for the initial distribution the probabilities of being in a state \tilde{M}_j^i and a state Q_k^i are stochastically independent: the joint probability of being in \tilde{M}_j^i and Q_k^i is the product of the individual probabilities (2.7).

It is an interesting question if the time-dependent solution $p_{m,n}(t)$ or the stationary distribution of the full model M remain in the form $p_{m,n}(t) = v_m(t) \otimes_{m,n} w_n(t)$ for $t > 0$. In fact, this is generally not the case.

Remark 2.3 (Caution). In most situations, $p_{m,n}(t)$ cannot be written as a pure tensor $p_{m,n}(t) = v_m(t) \otimes_{m,n} w_n(t)$ for $t > 0$. As discussed in proposition 4.4, we obtain a solution $(v_m(t) \otimes_{m,n} \pi_n)$ for a solution $v_m(t)$ of \tilde{M} and a vector π_n of stationary solutions π^i of Q^i if and only if we choose initial conditions $p^i = \pi^i$ for all Q^i .

(c) Example

As an example for the construction of the infinitesimal generator M from the components $((\tilde{m}_0, \tilde{M}), (p^i, Q^i)_{i=1}^{mM})$, we present a model that will be used in §3 for experimental data from the inositol trisphosphate receptor (IP₃R).

Let the infinitesimal generator for the switching between modes be

$$\tilde{M} = \left(\begin{array}{cc|cc} -m_{13} & 0 & m_{13} & \\ 0 & -m_{23} & m_{23} & \\ \hline m_{31} & m_{32} & -m_{31} - m_{32} & \end{array} \right) \quad (2.8)$$

and the models representing the intra-modal kinetics

$$Q^1 = \begin{pmatrix} -q_{12}^1 & q_{12}^1 \\ q_{21}^1 & -q_{21}^1 \end{pmatrix} \quad \text{and} \quad Q^2 = \begin{pmatrix} -q_{12}^2 & q_{12}^2 & 0 & 0 \\ q_{21}^2 & -q_{21}^2 - q_{23}^2 - q_{24}^2 & q_{23}^2 & q_{24}^2 \\ 0 & q_{32}^2 & -q_{32}^2 & 0 \\ 0 & q_{42}^2 & 0 & -q_{42}^2 \end{pmatrix} \quad (2.9)$$

with initial conditions

$$p^1 = (p_1^1, p_2^1) \quad \text{and} \quad p^2 = (p_1^2, p_2^2, p_3^2, p_4^2). \quad (2.10)$$

Then

$$M = \left(\begin{array}{cc|cc} \tilde{M}_2^{1,1} \oplus Q^1 & & \tilde{M}_2^{1,2} \otimes P^{1,2} & \\ \tilde{M}_2^{2,1} \otimes P^{2,1} & & \tilde{M}_2^{2,2} \oplus Q^2 & \end{array} \right)$$

$$= \left(\begin{array}{cccc|cccc} -m_{13}-q_{12}^1 & q_{12}^1 & 0 & 0 & m_{13}p_1^2 & m_{13}p_2^2 & m_{13}p_3^2 & m_{13}p_4^2 \\ q_{21}^1 & -m_{13}-q_{21}^1 & 0 & 0 & m_{13}p_1^2 & m_{13}p_2^2 & m_{13}p_3^2 & m_{13}p_4^2 \\ 0 & 0 & -m_{23}-q_{12}^1 & q_{12}^1 & m_{23}p_1^2 & m_{23}p_2^2 & m_{23}p_3^2 & m_{23}p_4^2 \\ 0 & 0 & q_{21}^1 & -m_{23}-q_{21}^1 & m_{23}p_1^2 & m_{23}p_2^2 & m_{23}p_3^2 & m_{23}p_4^2 \\ \hline m_{31}p_1^1 & m_{31}p_2^1 & m_{32}p_1^1 & m_{32}p_2^1 & -R-q_{12}^2 & q_{12}^2 & 0 & 0 \\ m_{31}p_1^1 & m_{31}p_2^1 & m_{32}p_1^1 & m_{32}p_2^1 & q_{21}^2 & -R-q_{21}^2-q_{23}^2-q_{24}^2 & q_{23}^2 & q_{24}^2 \\ m_{31}p_1^1 & m_{31}p_2^1 & m_{32}p_1^1 & m_{32}p_2^1 & 0 & q_{32}^2 & -R-q_{32}^2 & 0 \\ m_{31}p_1^1 & m_{31}p_2^1 & m_{32}p_1^1 & m_{32}p_2^1 & 0 & q_{42}^2 & 0 & -R-q_{42}^2 \end{array} \right) \quad (2.11)$$

with $R := m_{31} + m_{32}$.

(d) Parametrizing the model with experimental data

In order to parametrize the components $((\tilde{m}_0, \tilde{M}), (p^i, Q^i)_{i=1}^{n_M})$ of our model, the infinitesimal generators \tilde{M} and Q^i have to be inferred from ion channel data. We assume that the original data, a sequence of current measurements recorded with a constant sampling interval τ , have been statistically analysed so that they have the form of figure 1. Apart from visual inspection, mode changes have been investigated based on calculating the open probability within a window of a certain number of data points. One problem with these methods based on moving averages is that—depending on the window size—instantaneous jumps are transformed to gradual transitions so that the transitions between modes cannot be localized very accurately. By contrast, the heuristic method by Ionescu *et al.* [9] localizes switching events at specific data points but the uncertainty of the segmentation into different modes cannot be quantified. By contrast, the method by Siekmann *et al.* [6] calculates probability distributions for the position of each transition between different modes so that for each detected transition between different modes comprehensive information on the uncertainty is available. After a time series has been segmented each measurement is classified as open (O) or closed (C) and it has also been determined in which mode M^i the channel was at this point in time. From the results of a probabilistic method such as Siekmann *et al.* [6] rather than assigning a particular mode to each data point, it is possible to calculate a probability distribution for the different modes. This may improve the results for datasets where mode changes cannot be localized very accurately. The Markov model \tilde{M} is then inferred from the sequence S^k of modes M^i , whereas the models Q^i are parametrized from sequences of T^k that are representative of a particular mode. For example, in figure 1, the five data points between j_n and j_{n+1} could be used for inferring the model Q^2 representing the stochastic opening and closing within mode M^2 .

All models are parametrized with the Bayesian method developed in Siekmann *et al.* [19,20] or, alternatively, any other algorithm for fitting Markov models to single channel data. For inferring the infinitesimal generator \tilde{M} the likelihood has the form

$$\mathbb{P}((S^k) | \tilde{M}) = \tilde{\mu} \cdot P_{S^1} \cdot \exp(\tilde{M}\tau) \cdot P_{S^2} \cdots \exp(\tilde{M}\tau) \cdot P_{S^N} \cdot u^T, \quad (2.12)$$

where (S^k) is a sequence of observations of modes M^i separated by the sampling interval τ , \tilde{M} is the infinitesimal generator of an aggregated Markov model, $\tilde{\mu}$ is the stationary distribution of \tilde{M} and u^T is a column vector of ones. The matrices P_{S^k} project to the states of the model that represent the mode observed at data point k . For example,

$$P_{M^1} = \left(\begin{array}{c|ccc} \text{id}_{m_1} & 0 & \dots & 0 \\ 0 & \dots & & \\ \vdots & & & \\ 0 & \dots & \dots & 0 \end{array} \right) \quad (2.13)$$

with the same block structure as in (1.1) projects to states representing mode M^1 , the other projection matrices P_{S^i} are defined equivalently. The likelihood for inferring the infinitesimal generators Q^i from representative segments of T^k of open (O) and closed (C) events (figure 1) is analogous to (2.12). Missed events, see Hawkes and co-workers [21–23] and the references therein, are not considered because they are not relevant for this approach. The method is discussed in detail in Siekmann *et al.* [19,20].

3. Data-driven modelling of modal gating

Our new framework enables us to easily construct and parametrize models for modal gating following a transparent iterative process:

- (i) Infer the stochastic process S^k of switching between modes M^i (figure 1).
- (ii) Model the process S^k of mode switching by parametrizing an infinitesimal generator \tilde{M} (figure 3a).
- (iii) From segments of T^k representative for the opening or closing within each of the modes M^1, M^2, \dots (figure 3b) parametrize infinitesimal generators Q^1, Q^2, \dots .
- (iv) Choose initial distributions \tilde{m}_0 and p^i and combine all components $((\tilde{m}_0, \tilde{M}), (p^i, Q^i)_{i=1}^{n_M})$ by calculating the infinitesimal generator M of the full model (figure 4).

Inferring \tilde{M} and Q^i using the Bayesian approach briefly described in §2d ensures that the resulting model will be highly parsimonious because at each step a model with the optimal number of parameters for representing stochastic switching between modes, and opening and closing within modes, is determined. We demonstrate the practical implementation of this process using data collected by Wagner & Yule [15] and compare the results with our previously published model of the same dataset [12].

(a) Step (i): statistical analysis of modal gating

Previously, we have statistically analysed mode switching exhibited in the data by Wagner & Yule [15] and found two modes, the nearly inactive mode M^1 with a very low open probability and the highly active mode M^2 with $P_O \approx 70\%$ (see Siekmann *et al.* [6] for details). As illustrated in figure 1, we have a stochastic sequence of events M^1 and M^2 that are separated by a sampling interval $\tau = 0.05$ ms. We have results from two types of the inositol trisphosphate receptor (type I IP₃R and type II IP₃R) for various calcium concentrations (Ca^{2+}), 0.01 μM , 0.05 μM and 5 μM , at fixed concentrations of 10 μM inositol trisphosphate (IP₃) and 5 mM adenosine trisphosphate (ATP). Empirical histograms of the sojourn times in M^1 and M^2 for all except one dataset indicate that whereas time spent in the active mode M^2 may be represented satisfactorily by one state, accurately representing sojourn times in the nearly inactive mode M^1 seems to require at least two states (e.g. figure 2). Whereas one state accounts for the support of the sojourn time density in mode M^2 (figure 2b), the more widespread sojourn time density in mode M^1 is better approximated by two states (figure 2a). Thus, for five of our six datasets we parametrize \tilde{M} with the structure of (2.8). For one dataset (type II IP₃R at 0.05 μM Ca^{2+}), the histograms suggest that we need a model with two states representing M^1 and two states representing M^2 (figure 5). Thus, for these data we use the following infinitesimal generator:

$$\tilde{M}_{\text{Type II IP}_3\text{R, } 0.05 \mu\text{M Ca}^{2+}} = \left(\begin{array}{cc|cc} -m_{12} - m_{13} & m_{12} & m_{13} & 0 \\ m_{21} & -m_{21} - m_{24} & 0 & m_{24} \\ \hline m_{31} & 0 & -m_{31} & 0 \\ 0 & m_{42} & 0 & -m_{42} \end{array} \right). \quad (3.1)$$

It may seem that the mode switching dynamics of type II IP₃R is represented here with two different model structures. But, in fact, we can obtain the model structure from (3.1) by simply adding an additional M^2 state to the models for 0.01 μM Ca^{2+} and 5 μM Ca^{2+} such that transition

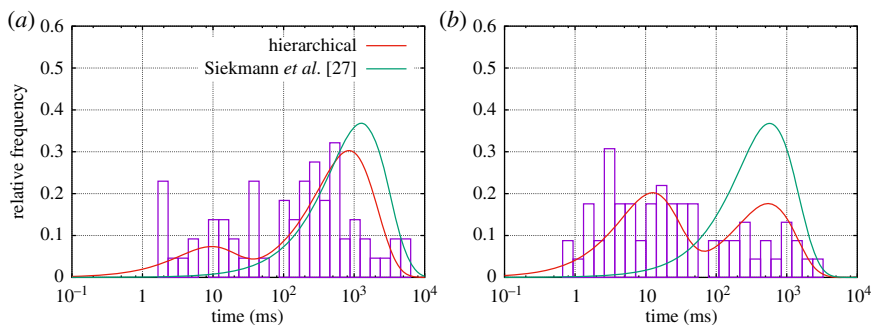


Figure 5. Empirical sojourn time distributions for both modes M^1 and M^2 for type II IP_3R for for $10 \mu\text{M } IP_3$, 5 mM ATP and $0.05 \mu\text{M } Ca^{2+}$. Whereas the hierarchical model can resolve (by using a four-state model) the widespread distributions of both M^1 and M^2 , the model from Siekmann *et al.* [12] can only capture one characteristic sojourn time due to the fact that only one pair of transition rates has been used to connect the submodels for mode M^1 and M^2 . Sojourn time distribution in (a) M^1 and (b) M^2 .

rates entering this state vanish. The interpretation of this is that the additional state representing long sojourns in M^2 observed for $0.05 \mu\text{M } Ca^{2+}$ —although present in the model—is never visited at the other ligand concentrations.

(b) Step (ii): parametrizing \tilde{M}

Fitting \tilde{M} to a time series S^k of M^1 and M^2 using our MCMC method [19,20] is a challenging problem. Because in a time series of a few hundred thousand up to about a million data points, the number of transitions between the two modes is only in the order of hundreds, the data from which the rate constants have to be inferred are effectively very limited—despite the large number of data points. An example of a convergence plot shown in electronic supplementary material, figure S1, demonstrates that values of the two rates, m_{13} and m_{23} , alternate. This is due to symmetry in the model structure chosen for the model \tilde{M} where the two states M_1^1 and M_2^1 can be swapped without changing the model. This effect can be removed by considering only one mode of the multi-modal posterior, in this case by considering only samples where m_{31} exceeds a certain threshold. Nevertheless, even after this correction some parameters such as the rate m_{23} show a high degree of uncertainty indicated by a widespread marginal distribution (electronic supplementary material, figure S1). Mean values and standard deviations of the distributions of the model parameters are summarized in electronic supplementary material, tables S1 and S2.

(c) Step (iii): parametrizing Q^1 and Q^2

In our previous study [12], we have already fitted a model with two states to representative segments of the inactive mode M^1 and a model with four states for representing M^2 , see (2.9) for the form of the infinitesimal generators Q^1 and Q^2 . Interestingly, we could show that Q^1 and Q^2 were independent of the concentrations of IP_3 , ATP and Ca^{2+} . The parameter values from the Supplementary Material of Siekmann *et al.* [12] are reproduced here for convenience (electronic supplementary material, table S3).

(d) Step (iv): the generator M of the full model

After the models \tilde{M} , Q^1 and Q^2 have been obtained, we finally need to specify the initial distributions \tilde{m}_0 , p^1 and p^2 . Consistent with the experimental assumption that recording of the data was started when the channel had reached steady state, we set $\tilde{m}_0 = \tilde{\mu}$, $p^1 = \pi^1$ and $p^2 = \pi^2$, where $\tilde{\mu}$, π^1 and π^2 are the stationary distributions of \tilde{M} , Q^1 and Q^2 , respectively. After all

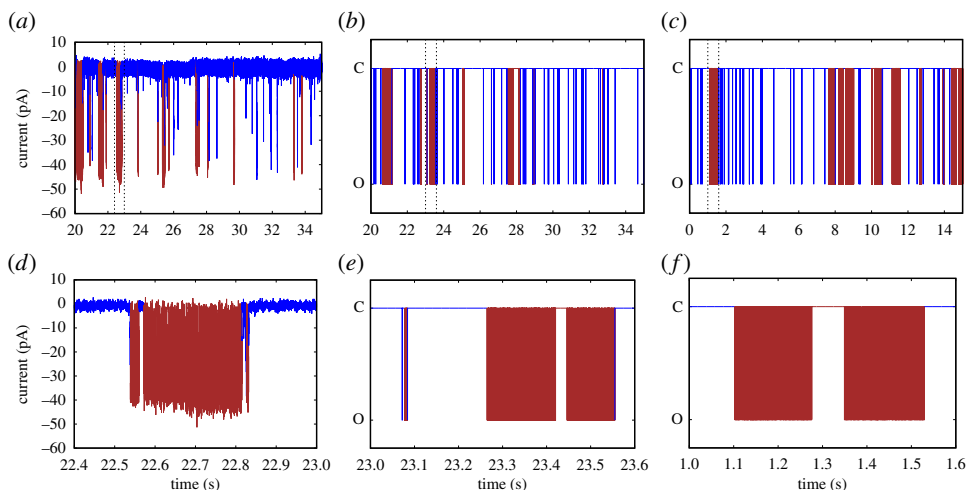


Figure 6. Comparison of a segment of data from type II IP₃R recorded at recorded at 10 μM IP₃, 5 mM ATP and 0.05 μM (a,d) with simulations of the hierarchical model presented here and the model from Siekmann *et al.* [12]. The colour of the line indicates if the channel is in the nearly inactive mode M¹ or the active mode M². As expected from the dwell time distributions of the two modes (figure 2), the model from Siekmann *et al.* [12] shows too many long sojourns in the active mode M² as well as in the inactive mode M¹ ((c) and (f)). By contrast, both long as well as short visits to both modes are seen in the sample path generated for the hierarchical model which is closer to what is observed in the data ((b) and (e)). (a) Data, (b) hierarchical model (HM), (c) Siekmann *et al.* [12] (SM), (d) data (detail), (e) HM (detail) and (f) SM (detail).

components $((\tilde{m}_0, \tilde{M}), (p^i, Q^i)_{i=1}^{n_{HM}})$ of our model have been specified, the infinitesimal generator M of the full model can be calculated using (2.6).

(e) Results

Owing to the problems with fitting the infinitesimal generator \tilde{M} (2.8) mentioned in §3b, one may ask if a simpler two-state model representing the dynamics of modal gating would be preferable. However, the ability of a three-state model to approximate the sojourn distribution of the nearly inactive mode M¹ more accurately (figure 2a) was found to be crucial for obtaining a better fit of the closed time distribution in comparison with the model from Siekmann *et al.* [12] (figure 2c). That the model structure of the hierarchical model proposed here is better able to capture the properties of the entire time series data seems even more convincing because it has—unlike the original model from Siekmann *et al.* [12]—been built without directly fitting to the time series at any step of its construction.

In electronic supplementary material, figure S2, we show that the bimodal closed time distribution observed for some combinations of ligand concentrations arises due to the mixing of the closed time distributions within nearly inactive mode M¹ and active mode M² both of which only have one distinct maximum.

Stronger differences between both models are observed for a dataset collected from type II IP₃R for 10 μM IP₃, 5 mM ATP and 0.05 μM Ca²⁺. For this experimental condition, the effect of modal gating can be observed without statistical analysis (electronic supplementary material, figure S3a). Figure 5 shows that both modes M¹ and M² exhibit a widespread distribution of sojourn times which can only approximately be captured by a four-state model with two states each for both M¹ and M². Whereas the new hierarchical model can approximate the empirical distributions of both modes relatively well, the model from Siekmann *et al.* [12] fails due to the fact that only one characteristic sojourn time for each mode can be captured by the pair of transition rates accounting for modal gating in this model (figure 5).

Owing to the failure to account for the modal sojourn time distributions, we expect the model from Siekmann *et al.* [12] to reproduce the kinetics observed in the data much less accurately than the new hierarchical model. In order to illustrate this, we simulated both the Siekmann *et al.* [12] model and the new model and compared them with a segment of experimental data of the same length (figure 6). The sample path was plotted in blue when the channel was in mode M^1 , whereas it was plotted in brown when the channel was in mode M^2 . The same colours were used for colouring the data based on the results of the statistical analysis from Siekmann *et al.* [6]. In the data segment shown here, both dwell times in the active mode M^2 of about 0.2–0.5 s are observed as well as very brief sojourns of a few milliseconds. Consistent with the dwell time distribution (figure 2), the long but not the short sojourns in the active mode M^2 are captured by the model from Siekmann *et al.* [12], whereas the hierarchical model developed in this study reproduces both long and short sojourns in this mode. Interestingly, as we show in the electronic supplementary material, for this particular dataset the channel seems to change its behaviour at an even slower time scale by spontaneously increasing the observed prevalence in M^2 for an extended period of time before returning to the initial level of activity (electronic supplementary material, figure S3a).

4. Mathematical analysis of the hierarchical Markov model

In the previous section, we demonstrated that the hierarchical Markov model introduced in §2 provides a statistically efficient framework for systematically building models for modal gating. Now, we focus on some interesting aspects of the mathematical structure of the hierarchical Markov model and show that many important properties of the infinitesimal generator M of the full model can be derived from the components $((\tilde{m}_0, \tilde{M}), (p^i, Q^i)_{i=1}^{N_M})$ of the model.

In §4a, we calculate the eigenvalues of M . The spectrum of M consists of two parts: the eigenvalues of \tilde{M} and a subset of the eigenvalues of the blocks $M^{i,i} = \tilde{M}^{i,i} \oplus Q^i$. But whereas the eigenvalues of the submatrices $\tilde{M}^{i,i}$ appear in the spectrum of the submatrices $M^{i,i}$, they are not eigenvalues of the full model M .

From a modelling point of view, it is an important question if properties of the components $((\tilde{m}_0, \tilde{M}), (p^i, Q^i)_{i=1}^{N_M})$ are preserved when they are combined in the full model. In §4b, we demonstrate that the sojourn time distribution in the states representing a particular mode in the model \tilde{M} is preserved for the analogous distribution calculated for the augmented state space of M .

When the initial distributions p^i coincide with the stationary distributions, $p^i = \pi^i$, we calculate the full time-dependent solution and the stationary distribution of M from the components $((\tilde{m}_0, \tilde{M}), (p^i, Q^i)_{i=1}^{N_M})$ of the hierarchical Markov model (§4c).

(a) Eigenvalues

Before we calculate the eigenvalues for general infinitesimal generators M of the full model, we remark that in most cases relevant for ion channel modelling we may assume that the matrices \tilde{M} and Q^i appearing in our model are diagonalizable—this is implied by the so-called detailed balance conditions:

$$\pi^i q_{ij} = \pi^j q_{ji}, \quad (4.1)$$

where π is the stationary distribution of an infinitesimal generator $Q = (q_{ij})$. A matrix $Q = (q_{ij})$ with (4.1) is diagonalizable with real eigenvalues because by choosing the transformation matrix $\text{diag}(\pi)^{1/2}$ it is similar to a symmetric real matrix. Detailed balance is usually assumed to hold for ion channel models because it can be related to thermodynamic reversibility of the transitions between different states in the model. Note that (4.1) holds automatically if the adjacency graph of the states of a Markov model is acyclic. This follows from Kolmogorov's criterion [24], see theorem 1.8 of [25] for a more recent statement of the continuous-time version. Thus, in particular, all infinitesimal generators \tilde{M} and Q^i considered in this article satisfy detailed balance.

Proposition 4.1 (Eigenvalues and eigenvectors of M assuming detailed balance). *We assume that the matrices \tilde{M} and Q^i of a hierarchical Markov model fulfil the detailed balance conditions (4.1).*

- (i) Let ζ be an eigenvalue of the matrix \tilde{M} and v_m^T a right eigenvector associated with ζ . Then ζ is also an eigenvalue of the full model M with associated right eigenvector $v_m^T \otimes u_n^T$, where u_n^T is a vector of $|n|$ ones.
- (ii) Moreover, all $v = \tilde{\zeta} + \lambda$, where $\tilde{\zeta}$ is an eigenvalue of $\tilde{M}^{i,i}$ and $\lambda \neq 0$ is an eigenvalue of Q^i , are eigenvalues of the full model M . If \tilde{w}^i is a left eigenvector of the submatrix $M^{i,i}$ associated with the eigenvalue v , $w_m = (0; \dots; 0; \tilde{w}^i; 0; \dots; 0)$ with $w(i) = \tilde{w}^i$ and $w(j) = 0$, $i \neq j$ is a left eigenvector of M associated with v .

Proof. Detailed balance implies that \tilde{M} and the Q^i are diagonalizable with real eigenvalues. In particular, all matrices have full sets of eigenvectors. This enables us to construct eigenvectors of the infinitesimal generator M of the full model from the eigenvectors of \tilde{M} and the Q^i .

- (i) We need to show that $M(v_m^T \otimes_{m,n} u_n^T) = \zeta(v_m^T \otimes_{m,n} u_n^T)$. Let $[M(v_m^T \otimes_{m,n} u_n^T)]^i$ denote the i th component of the partitioned vector. Here, $v_m^T \otimes_{m,n} u_n^T$ is a tensor product that is consistent with the partitions m and n as in (2.3) (definition 2.3). We calculate

$$[M(v_m^T \otimes_{m,n} u_n^T)]^i = (\tilde{M}^{i,i} \oplus Q^i)((v^i)^T \otimes u_{n_i}^T) + \sum_{k \neq i} (\tilde{M}^{i,k} \otimes P^{i,k})((v^k)^T \otimes u_{n_k}^T).$$

Using the compatibility condition of matrix multiplication and tensor product (A 2) we calculate

$$[M(v_m^T \otimes_{m,n} u_n^T)]^i = (\tilde{M}^{i,i}(v^i)^T \otimes u_{n_i}^T + (v^i)^T \otimes Q^i u_{n_i}^T) + \sum_{k \neq i} (\tilde{M}^{i,k}(v^k)^T \otimes P^{i,k} u_{n_k}^T).$$

Noting that $Q^i u_{n_i}^T = 0$ and $P^{i,k} u_{n_k}^T = u_{n_i}^T$, we finally get

$$[M(v_m^T \otimes_{m,n} u_n^T)]^i = \sum_{k=1}^{n_M} \tilde{M}^{i,k}(v^k)^T \otimes u_{n_i}^T = \zeta((v^i)^T \otimes u_{n_i}^T).$$

Because this holds for all blocks we obtain the desired result.

- (ii) All except for the i th block of w are zero, so we get

$$wM = (\tilde{w}^i[\tilde{M}^{i,1} \otimes P^{1,i}]; \dots; \tilde{w}^i[\tilde{M}^{i,i} \oplus Q^i]; \dots; \tilde{w}^i[\tilde{M}^{i,n_M} \otimes P^{n_M,i}]).$$

Because \tilde{w}^i is an eigenvector of $\tilde{M}^{i,i} \oplus Q^i$ we know that $\tilde{w}^i(\tilde{M}^{i,i} \oplus Q^i) = v\tilde{w}^i$. For w to be an eigenvector, it remains to be shown that all other blocks vanish. Let u be a left eigenvector of $\tilde{M}^{i,i}$ associated with the eigenvalue $\tilde{\zeta}$ and v a left eigenvector of Q^i associated with the eigenvalue λ . Then \tilde{w}^i can be written as $\tilde{w}^i = u \otimes v$ according to (A 3). Substituting this and $P^{i,k} = p^k \otimes u_{n_i}^T$, $k \neq i$, we calculate

$$(u \otimes v)[\tilde{M}^{1,k} \otimes p^k \otimes u_{n_i}^T] = u(\tilde{M}^{1,k} \otimes p^k) \otimes v u_{n_i}^T. \quad (4.2)$$

The term $v u_{n_i}^T$ is the standard scalar product $\langle v^T, u_{n_i}^T \rangle$ of the vectors v^T and $u_{n_i}^T$. Because the row sums of Q^i are zero, $u_{n_i}^T$ is in the right nullspace of Q^i . By assumption, v is an eigenvector associated with any eigenvalue $\lambda \neq 0$. This means that v is not in the left nullspace of Q^i , so it must be orthogonal to any vector in the right nullspace. It follows that (4.2) vanishes as required. ■

For the general case where the infinitesimal generators of the model \tilde{M} and the submatrices $M^{i,i}$ may not necessarily be diagonalizable we need the Schur decomposition (proposition A.2). The Schur decomposition ensures that the matrix M can be transformed to an upper-triangular matrix by a unitary matrix. In the following, we construct a unitary matrix S from the components $((\tilde{m}_0, \tilde{M}), (p^i, Q^i)_{i=1}^{n_M})$ of our model.

Lemma 4.1 (Unitary matrix S). For the components $((\tilde{m}_0, \tilde{M}), (p^i, Q^i)_{i=1}^{n_M})$ of a hierarchical Markov model, let

$$T_{\tilde{M}} = \Theta^* \tilde{M} \Theta, \quad T_{\tilde{M}^{i,j} \oplus Q^i} = (V_i \otimes W_i)^* \tilde{M}^{i,j} \oplus Q^i (V_i \otimes W_i),$$

be the Schur decompositions of \tilde{M} and $\tilde{M}^{i,j} \oplus Q^i$. Let $\tilde{u}_{n_i}^T = 1/\sqrt{n_i} u_{n_i}^T$ be the vectors obtained by normalizing the vectors of ones $u_{n_i}^T$.

- (i) The matrices W_i may be chosen so that they have the form $W_i = (\tilde{u}_{n_i}^T | \tilde{W}_i)$ with $\tilde{W}_i \in \mathbb{C}^{n_i \times (n_i-1)}$.
(ii) Let

$$\Theta = \begin{pmatrix} \Theta^1 \\ \vdots \\ \Theta^{n_M} \end{pmatrix}$$

be row-partitioned according to the block structure of \tilde{M} from (1.1). Then the matrix

$$S = \begin{pmatrix} \Theta^1 \otimes \tilde{u}_{n_1}^T & V_1 \otimes \tilde{W}_1 & 0 & \dots & 0 \\ \vdots & 0 & V_2 \otimes \tilde{W}_2 & \dots & \vdots \\ \vdots & \vdots & \dots & \ddots & 0 \\ \Theta^{n_M} \otimes \tilde{u}_{n_{n_M}}^T & 0 & \dots & \dots & V_{n_M} \otimes \tilde{W}_{n_M} \end{pmatrix} \quad (4.3)$$

is unitary.

Proof.

- (i) Because the row sums of Q^i vanish, the vector $\tilde{u}_{n_i}^T$ is a right eigenvector of Q^i associated with the eigenvalue zero. Without loss of generality, we can choose $\tilde{u}_{n_i}^T$ as the first column of W_i .
(ii) By construction, all column vectors of S are normalized. Thus, it remains to show that they are also pairwise orthogonal. By definition, any two distinct column vectors appearing in the same block of S are orthogonal. It is trivial that column vectors from different blocks are orthogonal unless one of the two appears in the first block of S . Thus, let θ^T be a column vector of Θ and $v_i^T \otimes \tilde{w}_i^T$ be a column vector of any $V_i \otimes W_i$. With the shorthand for tensor products consistent with partitions (2.3) introduced in definition 2.3, the scalar product $\langle \cdot, \cdot \rangle$ of the two columns is

$$\langle \theta_m^T \otimes_{m,n} \tilde{u}_n^T, (0; \dots; v_i \otimes \tilde{w}_i; \dots; 0)^T \rangle = \langle \theta_m^T(i) \otimes \tilde{u}_{n_i}^T, v_i^T \otimes \tilde{w}_i^T \rangle$$

and due to the zeroes in all except for the i th block, all other summands vanish. Noting that $\langle u, v \rangle = u(v^*)^T = u^T v^*$ can be interpreted as a special case of matrix multiplication (where $'^*'$ denotes component-wise complex conjugation) we can use (A 2):

$$\langle \theta_m^T(i) \otimes_{m,n} \tilde{u}_{n_i}^T, v_i^T \otimes \tilde{w}_i^T \rangle = \langle \theta_m^T(i), v_i^T \rangle \langle \tilde{u}_{n_i}^T, \tilde{w}_i^T \rangle.$$

But because $\tilde{u}_{n_i}^T$ appeared as a column in the original unitary matrix W_i , the \tilde{w}_i^T are all orthogonal to $\tilde{u}_{n_i}^T$ so that the above scalar product vanishes. Thus, the matrix S is unitary. ■

Proposition 4.2 (Eigenvalues of the full model M). Let $\tilde{\zeta}$ be an eigenvalue of the model \tilde{M} . Then $\tilde{\zeta}$ is also an eigenvalue of the full model M . Moreover, all $v = \tilde{\zeta} + \lambda$, where $\tilde{\zeta}$ is an eigenvalue of $\tilde{M}^{i,j}$ and $\lambda \neq 0$ is an eigenvalue of Q^i , are eigenvalues of the full model M .

Proof. We demonstrate that with the matrix S from (4.3), we obtain a Schur decomposition of the matrix M . We need to show that $A = S^* M S$ is upper triangular. The block structure of S is rectangular with $n_M \times (n_M + 1)$ blocks which means that S^* has an $(n_M + 1) \times n_M$ block structure.

Thus, the resulting matrix A will have $(n_M + 1) \times (n_M + 1)$ blocks and its diagonal will consist of the eigenvalues of \tilde{M} in the upper left block followed by the remaining eigenvalues from the submatrices $\tilde{M}^{i,i}$. We show that all blocks $A^{i,j}$ are upper triangular which implies that A is indeed upper triangular. First, a lengthy calculation shows that $A^{1,1}$ is a block-wise expanded form of $\Theta^* \tilde{M} \Theta$ and thus upper triangular. One can see directly that the remaining elements on the block diagonal are

$$A^{i,i} = (V_i \otimes \tilde{W}_i)^* (\tilde{M}^{i,i} \oplus Q^i) (V_i \otimes \tilde{W}_i)$$

and, therefore, all upper triangular.

It remains to show that the lower diagonal blocks $A^{i,j}$ with $i > j$ vanish. We will demonstrate that the $A^{i,j}$ vanish provided that

$$\tilde{W}_i^* \tilde{u}_{n_i}^T = 0. \quad (4.4)$$

Equation (4.4) is just another way of saying that $\tilde{u}_{n_i}^T$ is orthogonal to all columns of \tilde{W}_i . But this is true because from Lemma 4.1(i) we know that $\tilde{u}_{n_i}^T$ is the first column of W_i , so it must be orthogonal to all column vectors of \tilde{W}_i .

We now calculate the subdiagonal blocks $A^{i,j}$, $i > j$. First, we calculate the blocks $A^{1,1}$ on the first block column. We observe that

$$(M \cdot S)^{k,1} = (\tilde{M}^{k,k} \oplus Q^k) (\Theta^k \otimes \tilde{u}_{n_k}^T) + \sum_{j \neq k} (\tilde{M}^{k,j} \oplus P^{k,j}) (\Theta^j \otimes \tilde{u}_{n_j}^T).$$

Because S^* is block diagonal below the first row, we can calculate

$$\begin{aligned} A^{k+1,1} &= (S^* \cdot M \cdot S)^{k+1,1} = (V_k \otimes \tilde{W}_k)^* (\tilde{M}^{k,k} \oplus Q^k) (\Theta^k \otimes \tilde{u}_{n_k}^T) \\ &\quad + \sum_{j \neq k} (V_k \otimes \tilde{W}_k)^* (\tilde{M}^{k,j} \oplus P^{k,j}) (\Theta^j \otimes \tilde{u}_{n_j}^T) \end{aligned}$$

because in the row $(k+1)$ th row of S^* for $k=1, \dots, n_M$ only the k th block is non-zero. By taking advantage of (A 2), we obtain

$$\begin{aligned} A^{k+1,1} &= (V_k \otimes \tilde{W}_k)^* (\tilde{M}^{k,k} \Theta^k \otimes \tilde{u}_{n_k}^T + \Theta^k \otimes Q^k \tilde{u}_{n_k}^T) + \sum_{j \neq k} (V_k \otimes \tilde{W}_k)^* (\tilde{M}^{k,j} \Theta^j \otimes P^{k,j} \tilde{u}_{n_j}^T) \\ &= (V_k \otimes \tilde{W}_k)^* (\tilde{M}^{k,k} \Theta^k \otimes \tilde{u}_{n_k}^T) + \sum_{j \neq k} (V_k \otimes \tilde{W}_k)^* (\tilde{M}^{k,j} \Theta^j \otimes \tilde{u}_{n_k}^T), \end{aligned}$$

where we have used $Q^k \tilde{u}_{n_k}^T = 0$ and $P^{k,j} \tilde{u}_{n_j}^T = \tilde{u}_{n_k}^T$. Again using (A 2), we calculate

$$A^{k+1,1} = V_k^* \tilde{M}^{k,k} \Theta^k \otimes \tilde{W}_k^* \tilde{u}_{n_k}^T + \sum_{j \neq k} V_k^* \tilde{M}^{k,j} \Theta^j \otimes \tilde{W}_k^* \tilde{u}_{n_k}^T.$$

This vanishes due to (4.4) as explained above.

For the remaining blocks $A^{k+1,l+1}$, $k > l = 1, \dots, n_M - 1$, we simply calculate

$$\begin{aligned} A^{k+1,l+1} &= (V_k \otimes \tilde{W}_k)^* (\tilde{M}^{k,l} \oplus P^{k,l}) (V_l \otimes \tilde{W}_l) \\ &= (V_k^* \tilde{M}^{k,l} \otimes \tilde{W}_k^* P^{k,l}) (V_l \otimes \tilde{W}_l) \\ &= (V_k^* \tilde{M}^{k,l} V_l) \otimes (\tilde{W}_k^* P^{k,l} \tilde{W}_l). \end{aligned}$$

Replacing $P^{k,l}$ by $\tilde{u}_{n_k}^T \otimes p^l$ (2.5), we get

$$A^{k+1,l+1} = (V_k^* \tilde{M}^{k,l} V_l) \otimes (\tilde{W}_k^* \tilde{u}_{n_k}^T) \otimes (p^l \tilde{W}_l),$$

where—due to the term $\tilde{W}_k^* \tilde{u}_{n_k}^T$ —we again conclude with (4.4) that $A^{k+1,l+1}$ vanishes. ■

(b) Sojourn times in modes

We will now investigate the sojourn times within the states that represent the modes M^i . The switching between modes is represented by a model with infinitesimal generator \tilde{M} and

one can ask if the dynamics is preserved after \tilde{M} is combined with the other components $((\tilde{m}_0, \tilde{M}), (p^i, Q^i)_{i=1}^{n_M})$ to the generator M of the full model. We denote by $f_{M^i}(t)$, the density function of the sojourn time in mode M^i represented by \tilde{M} and by $f_{M^i}(t)$ the sojourn time densities of M^i in the augmented state space of the generator M of the full model. If the mode switching dynamics is preserved, the sojourn time densities should be equal and we will show that indeed $f_{M^i}(t) = f_{\tilde{M}^i}(t)$.

Proposition 4.3 (Modal sojourn times). For $f_{M^i}(t)$, sojourn time densities within mode M^i with an initial distribution p^0 as in definition 2.6, we have $f_{M^i}(t) = f_{\tilde{M}^i}(t)$.

Proof. For simplicity we only treat the case of two aggregates of states, M^1 and M^2 . For the sojourn time within M^1 we have

$$f_{M^1}(t) = p^0 M^{2,1} \exp(M^{1,1}t) M^{1,2} u_{m_2 n_2}^T,$$

where $p^0 = p_{M^2}^0 \otimes p_{Q^2}^0$ is a suitably normalized initial state probability distribution. Substituting from (2.6), we obtain for

$$\begin{aligned} \exp(M^{1,1}t) M^{1,2} &= \exp([\tilde{M}^{1,1} \oplus Q^1]t) M^{1,2} \\ &= [\exp(\tilde{M}^{1,1}t) \otimes \exp(Q^1t)] (\tilde{M}^{1,2} \otimes P^{1,2}), \end{aligned}$$

where we have used (A 4) for calculating the matrix exponential. Now,

$$[\exp(\tilde{M}^{1,1}t) \otimes \exp(Q^1t)] (\tilde{M}^{1,2} \otimes P^{1,2}) = [\exp(\tilde{M}^{1,1}t) \tilde{M}^{1,2}] \otimes P^{1,2}$$

according to the compatibility of tensor and matrix product (A 2) which will be used repeatedly below. Also note that $\exp(Q^1t) \cdot P^{1,2} = P^{1,2}$. Multiplying this on the right by $u_{m_2 n_2}^T = u_{m_2}^T \otimes u_{n_2}^T$ leads to

$$[\exp(\tilde{M}^{1,1}t) \tilde{M}^{1,2}] \otimes P^{1,2} (u_{m_2}^T \otimes u_{n_2}^T) = [\exp(\tilde{M}^{1,1}t) \tilde{M}^{1,2} u_{m_2}^T] \otimes u_{n_2}^T,$$

where we have evaluated $P^{1,2} u_{n_2}^T = u_{n_1}^T$ in the right-most term. Analogous calculations will be carried out automatically below. The above result is now multiplied on the left by $M^{2,1} = \tilde{M}^{2,1} \otimes P^{2,1}$:

$$(\tilde{M}^{2,1} \otimes P^{2,1}) [\exp(\tilde{M}^{1,1}t) \tilde{M}^{1,2} u_{m_2}^T] \otimes u_{n_1}^T = [\tilde{M}^{2,1} \exp(\tilde{M}^{1,1}t) \tilde{M}^{1,2} u_{m_2}^T] \otimes u_{n_2}^T.$$

Finally, we multiply the preceding result on the left by $p^0 = p_{M^2}^0 \otimes p_{Q^2}^0$ and compute

$$\begin{aligned} f_{M^1}(t) &= (p_{M^2}^0 \otimes p_{Q^2}^0) [\tilde{M}^{2,1} \exp(\tilde{M}^{1,1}t) \tilde{M}^{1,2} u_{m_2}^T] \otimes u_{n_2}^T \\ &= [p_{M^2}^0 \tilde{M}^{2,1} \exp(\tilde{M}^{1,1}t) \tilde{M}^{1,2} u_{m_2}^T] \otimes (p_{Q^2}^0 u_{n_2}^T). \end{aligned}$$

Now, because $(p_{Q^2}^0 u_{n_2}^T) = 1$, we obtain the desired result:

$$f_{M^1}(t) = p_{\tilde{M}^2}^0 \tilde{M}^{2,1} \exp(\tilde{M}^{1,1}t) \tilde{M}^{1,2} u_{m_2}^T = f_{\tilde{M}^1}(t). \quad \blacksquare$$

(c) Full solution for $p^i = \pi^i$

If we choose initial conditions $p^i = \pi^i$, where the π^i are stationary distributions of the models Q^i , the solution of the full model has a particularly simple form.

Proposition 4.4 (Full solution for $p^i = \pi^i$). Let $v_m(t)$ be the time-dependent solution for the initial condition w_n^0 and $\tilde{\mu}_n$ be the stationary solution of the infinitesimal generator \tilde{M} with their partition \mathbf{m} . Let $\pi^i, i = 1, \dots, n_M$ be the stationary distributions of Q^i or written as a partitioned vector, π_n with its partition \mathbf{n} . If for each generator Q^i we set $p^i = \pi^i$ and we choose an initial distribution $p_{\mathbf{m} \cdot \mathbf{n}}^0 = v_{\mathbf{m}}^0 \otimes_{\mathbf{m}, \mathbf{n}} \pi_n$ consistent with definition 2.6, the solution $p_{\mathbf{m} \cdot \mathbf{n}}(t)$ of the full model is

$$p_{\mathbf{m} \cdot \mathbf{n}}(t) = v_m(t) \otimes_{\mathbf{m}, \mathbf{n}} \pi_n = (v^1(t) \otimes \pi^1; \dots; v^i(t) \otimes \pi^i; \dots; v^{n_M}(t) \otimes \pi^{n_M}). \quad (4.5)$$

By taking the limit $t \rightarrow \infty$, we obtain the stationary distribution

$$\mu_{\mathbf{m} \cdot \mathbf{n}} = \tilde{\mu}_{\mathbf{m}} \otimes_{\mathbf{m}, \mathbf{n}} \pi_n = (\tilde{\mu}^1 \otimes \pi^1; \dots; \tilde{\mu}^i \otimes \pi^i; \dots; \tilde{\mu}^{n_M} \otimes \pi^{n_M}). \quad (4.6)$$

Remark 4.1. The stationary distribution (4.6) is independent of the initial distribution $p_{m,n}^0$, so, for $p^i = \pi^i$, we converge to the stationary distribution (4.6) also for $p_{m,n}^0 = (v_m^0 \otimes_{m,n} w_n^0)$ with $w_n^0 \neq \pi_n$ and even for arbitrary initial conditions $p_{m,n}^0$ that are inconsistent with definition 2.6.

Proof. That (4.5) is a solution can be shown by substituting $p_{m,n}(t) = v_m(t) \otimes_{m,n} \pi_n$ into

$$\frac{dp(t)}{dt} = p(t)M, \quad (4.7)$$

where M is the generator of the full model (2.6). First, we calculate the left-hand side:

$$\begin{aligned} \frac{dp_{m,n}(t)}{dt} &= \frac{d(v_m(t) \otimes_{m,n} \pi_n)}{dt} \\ &= \left(\frac{dv_m(t)}{dt} \right) \otimes_{m,n} \pi_n \\ &= (v_m(t)\tilde{M}) \otimes_{m,n} \pi_n, \end{aligned} \quad (4.8)$$

where the last equality (4.8) follows because $v_m(t)$ is a solution of the model generated by \tilde{M} .

We now show that we also obtain (4.8) from the right-hand side of (4.7). For the i th component $[p_{m,n}(t) \cdot M]^i$, we calculate

$$[p_{m,n}(t) \cdot M]^i = (v^i(t) \otimes \pi^i)(\tilde{M}^{i,i} \oplus Q^i) + \sum_{j \neq i} (v^j(t) \otimes \pi^j)(\tilde{M}^{j,i} \otimes P^{j,i}).$$

For the first summand, the contribution of Q^i vanishes because of $\pi^i Q^i = 0$

$$(v^i(t) \otimes \pi^i)(\tilde{M}^{i,i} \oplus Q^i) = (v^i(t)\tilde{M}^{i,i}) \otimes \pi^i + v^i(t) \otimes \pi^i Q^i = (v^i(t)\tilde{M}^{i,i}) \otimes \pi^i. \quad (4.9)$$

Because of $\pi^j P^{j,i} = \pi^i$, the second summand simplifies to

$$\sum_{j \neq i} (v^j(t) \otimes \pi^j)(\tilde{M}^{j,i} \otimes P^{j,i}) = \sum_{j \neq i} (v^j(t)\tilde{M}^{j,i}) \otimes \pi^i. \quad (4.10)$$

With (4.9) and (4.10), we derive for each component:

$$[p_{m,n}(t) \cdot M]^i = \sum_{i=1}^{n_M} (v^j(t)\tilde{M}^{j,i}) \otimes \pi^i.$$

This means that the right-hand side of (4.7) is indeed of the form (4.8) which confirms that (4.5) is a solution. \blacksquare

5. Conclusion

We have proposed a new model for representing modal gating, the spontaneous switching of ion channels between different levels of activity. The model is suitable for modelling channels with an arbitrary number of modes and is capable of representing both the probabilistic opening and closing within modes as well as the stochastic switching between modes that regulates these dynamics.

(a) Modular representation of modal gating

In comparison with previous studies, the model presented here incorporates modal gating in a much more transparent way. Ullah *et al.* [11] developed their model of the IP₃R from a binding scheme. First, the authors determined the set of open and closed model states from a statistical model selection criterion. Second, they determined which of these states should account for which of the three modes observed by Ionescu *et al.* [9]. The decision that a particular open or closed state should account for the mode showing a low, intermediate or high level of activity was based on heuristic inspection of the ligand-dependency of modal gating. The model was parametrized

by optimizing a likelihood that accounted for various sources of single channel data including statistics of modal gating. This treats the parameter space of their model as a black box from which a suitable set of parameters capable of accounting for all datasets is selected by optimization. We expect such an approach to be statistically less efficient than a model whose structure incorporates modal gating more explicitly.

Siekmann *et al.* [12] used modal gating as the underlying construction principle of their model by separating the inference of parameters related to dynamics within modes from estimation of parameters related to switching between modes. First, models for the inactive mode M^1 and the active mode M^2 were inferred by fitting segments of data representative of each of the two modes—in fact, the same models were reused in the present study. However, because at that time rigorous statistical techniques for segmenting ion channel data by modes were not available, the time scales of the switching between both modes was inferred by connecting the submodels for M^1 and M^2 with a pair of transition rates whose values were then determined from a fit to complete traces of single channel data. Similar to Ullah *et al.* [11] modal gating was thus incorporated into the model without explicitly considering its stochastic dynamics apparent in the data.

In this study using our previously developed method Siekmann *et al.* [6], we were able to explicitly account for transitions between different modes inferred from experimental data. The method partitions a time series into segments based on the open probability P_O of the channel. In this way, spontaneous changes of channel activity can be detected. Because the analysis is based on Bayesian statistics, comprehensive information on the uncertainty of the results is available via the posterior distribution. For the IP₃R data used here the inferred times at which the channel made a transition to another mode had very low estimated standard deviations (less than one data point up to a few data points). Whereas for this study, it was therefore sufficient to use point estimates of the change times, the full posterior distribution can be used for channels whose modes cannot be distinguished with similar accuracy.

Thus, the statistical method from Siekmann *et al.* [6] enables us to fit a model \tilde{M} directly to the stochastic process of mode switching inferred from the experimental data instead of arbitrarily introducing transition rates between modes as in our previous study [12]. Therefore, we can accurately represent mode switching, only adding exactly as many parameters as required. In comparison with our previous model, the new model described here requires only two additional parameters. Inspection of the sojourn time histograms show that these two parameters are essential in order to account for the fact that sojourns in the nearly inactive mode M^1 exhibit two different time scales which cannot be represented by a model with less parameters.

Because Siekmann *et al.* [6] distinguished modes based on their open probabilities, it may be difficult to distinguish modes with similar characteristic open probabilities P_O but with different kinetics, like, for example, an active mode with high P_O and a flicker mode with similar P_O but faster transitions between open and closed states. This possibility can be excluded by fitting models Q^i to the observed open and closed events in segments representative for each mode. For the IP₃R dataset used for this study we could not only show that Q^1 and Q^2 , respectively, did not differ significantly for segments from the same time series but were also ligand-independent.

The hierarchical Markov model developed in this study allows us to combine the ‘modules’ Q^i accounting for opening and closing within modes with \tilde{M} representing mode switching to a representation of both aspects of the single channel dynamics. It is important to note that none of the components $((\tilde{m}_0, \tilde{M}), (p^i, Q^i)_{i=1}^{n_M})$ of our model were determined by directly fitting to the sequence of open and closed events observed in experiments—the models Q^i are inferred from segments of the data and the model \tilde{M} is parametrized from transitions between the modes M^i . Thus, the open and closed time distributions $f_O(t)$ and $f_C(t)$, respectively, can be considered a prediction of our hierarchical model M . That the hierarchical model M performs better at predicting $f_O(t)$ and $f_C(t)$ than our previous model whose transition rates were inferred from a direct fit to complete traces of open and closed events provides additional evidence that our new approach does not suffer from possible sources of error in our statistical analysis of modal gating but is, in fact, a superior representation of the data.

For analysing statistical properties of modal gating, an advantage of our model is that in addition to representing the channel being open or closed each state is also associated with a mode. The analysis of bursts according to Colquhoun & Hawkes [7] depends on the selection of open and 'short-lived' closed states that represent a burst and a class of 'long-lived' closed states that account for gaps between bursts. This not only requires additional assumptions but it is also unclear how the state space of an unstructured Markov model should be partitioned if a channel has multiple modes. No such difficulties arise in our model because each mode is represented by an aggregate of open and closed states. It is, therefore, clear how the relevant states should be chosen so that various properties such as, for example, the open and closed times of each mode, can be calculated using the theory of Colquhoun & Hawkes [2] or Colquhoun & Hawkes [7]. Because here, mode switching is defined as spontaneous changes between different open probabilities rather than clusters of open events, the methods from Colquhoun & Hawkes [2] seem more suitable. The theory of Colquhoun & Hawkes [7] leads to more complicated calculations due to the assumption that bursts must commence with an opening of the channel, whereas this does not necessarily have to be the case for a sojourn in a mode. In summary, this means that an additional benefit of our modelling approach is that statistical properties of modes can be calculated more easily from our model than from a general Markov model.

The modular structure of our hierarchical model which separates the representation of transitions between modes (inter-modal kinetics) from the dynamics within modes (intra-modal kinetics) not only provides a more parsimonious representation than previous models but, most notably, evidence is accumulating that in channels that exhibit different modes the switching between modes may be more important for their physiological function than intra-modal kinetics. This is strongly suggested by recent studies of the IP₃R. In a study of insect type I IP₃R, Ionescu *et al.* [9] observed three modes with essentially identical kinetic properties across different ligand concentrations, whereas the overall dynamics of the channel was determined by the highly ligand-dependent prevalence of the channel in these modes. Thus, Ionescu *et al.* [9] proposed that modal gating is the major mechanism of ligand regulation in the IP₃R. This was confirmed for mammalian type I and type II IP₃R data by Siekmann *et al.* [6] and led to the interpretation that ion channel kinetics is restricted to a fixed repertoire of modes which have to be mixed appropriately in order to respond to given ligand concentrations. Ligand-dependent switching between ligand-independent modes suggests that physiological function may depend more strongly on the slow time scale of switching between modes rather than the fast opening and closing of the channel within a mode. This was indeed recently shown in two studies of the role of IP₃R in intracellular calcium dynamics. Cao *et al.* [26] showed that the essential features of calcium oscillations in airway smooth muscle could be preserved after iteratively simplifying the model from Siekmann *et al.* [12] to a two-state model that only accounted for switching between the two modes neglecting the kinetics of transitions between multiple open and closed states within the modes. Siekmann *et al.* [27] applied similar reduction techniques to demonstrate that also the stochastic dynamics of small clusters of IP₃R s can be captured by a two-state model reduced to the dynamics of mode switching. In our new hierarchical model, inter-modal and intra-modal kinetics are represented separately so that the model representation with the right level of detail can be chosen based on the requirements of a specific application.

(b) Biophysical implications of modal gating

Although modal gating has been observed for a long time it has rarely been accounted for in ion channel models. The crucial importance of modal gating has only recently been appreciated among investigators of the IP₃R channel and it is now widely recognized in the community [10]. Various independent sources of evidence indicate that modal gating must be accounted for, both for understanding IP₃R function as well as for gaining insight into biophysical properties of the channel molecule. As mentioned in the previous section, the role of IP₃R in intracellular calcium dynamics is defined by its behaviour on the slow time scale of transitions between different modes rather than the fast time scale of opening and closing [26,27]. Previously,

Ionescu *et al.* [9] discovered that the IP₃R adjusts its level of activity depending on ligands such as calcium by regulating the proportion of time that the channel spends in different modes. This was subsequently confirmed by the statistical analysis by Siekmann *et al.* [6]. These results reveal the major functional implications of modal gating, so one may ask if any insight can be gained into the underlying biophysics. In their early model of modal gating in a chloride channel, Blatz & Magleby [8] postulated that different modes may be related to different conformations of the channel protein. Direct experimental evidence into how different modes arise from biophysical constraints of the channel protein is accumulating. Two examples include a thorough analysis of the potassium channel KscA discussed in more detail below [28–30] and a more recent study by Vij *et al.* [31] on the acetylcholine receptor. Also see the commentary by Geng & Magleby [32]. This suggests that modes form a fixed repertoire of possible behaviours defined by the molecular properties of the channel. Being constrained to a few different modes, ion channels overcome these limitations by switching between modes.

This implies that methods for identifying different modes in single channel data not only provide us with more accurate insight into the channel dynamics but may also reveal the transitions between different biophysical states of the channel. As mentioned above there are strong indications that each mode is reflected by a different three-dimensional arrangement of the channel protein, known as a conformational state. Thus, the aggregates of states in ion channel models that account for different modes M^i correspond to different conformations at the level of the channel protein. In such a model, the transitions between states representing different modes reflect the rates of conformational changes.

This direct correspondence between aggregates of states and underlying biophysics is important to note because interpreting individual states in Markov models for ion channels is problematic in general, at least without additional experiments. For the simplest possible representation of a gating ion channel is a two-state Markov model with only one open and one closed state it is, of course, obvious that these two different model states at the same time correspond to different biophysical states of the channel protein. This ‘mechanistic’ interpretation explains the popularity of this type of model. The Markov assumption implies that the open and closed times of a two-state model are exponentially distributed which means that durations of channel openings and closings both have characteristic time scales τ_O and τ_C given by the parameters of the exponential sojourn time distributions $f_O(t)$ and $f_C(t)$. However, many ion channels exhibit multiple characteristic open and closed times that cannot be represented by exponential distributions. Whereas an open ion channel must be in a different conformation than a closed ion channel distinguishing only two conformational states is a very coarse description of the complicated deformations of channel proteins that can be identified by molecular dynamics models. Nevertheless, if our goal is to base our models on rigorous statistical analysis, for some data we may not be able to identify more than two states.

Non-exponential open and closed times can often be represented satisfactorily by aggregated continuous-time Markov models where more than one state is used for representing the channel being open or closed. These models provide a simple generalization of the two-state Markov model and account for more than just one characteristic open or closed time scale τ_O and τ_C . By definition, the sojourn times in the open or closed class of an aggregated Markov model are distributed according to a phase-type distribution, a class of distributions representing the time a Markov chain spends in a set of transient states until exiting to an absorbing state [16,17]. As with the two-state model it is tempting to also associate the individual states of an aggregated Markov model with different biophysical states of the channel protein. The multiple open and closed states of an aggregated Markov model could be interpreted to resolve in more detail the series of conformational changes that the channel goes through while it opens. If this interpretation was valid one could hope to discover details of the molecular structure of ion channels beyond the trivial distinction between an open and a closed state once the ‘best’ aggregated Markov model for a given dataset has been found.

Unfortunately, this ‘mechanistic’ interpretation of aggregated Markov models has several flaws. Whereas it can be directly inferred from single channel data if the channel is open or

closed and in which of its modes M^i it is, distinguishing different open or closed Markov states requires additional experiments and is possibly ill-defined. First, the only reason that a particular model consists of multiple open and closed states is that multiple characteristic open and closed times were observed. It is an assumption to be empirically confirmed that for each observed exponentially distributed sojourn time the channel must necessarily be in a distinct conformational states—so more Markov states may appear in the model than can be distinguished biophysically at the level of the channel protein. By contrast, it is likely that some conformational states may not have a strong enough influence on the dynamics that they are represented by a state in a model inferred from the data. But even if we assume that each Markov state should, in principle, reflect a distinct underlying biophysical state, it is challenging both experimentally as well as theoretically to identify, for example, a three-dimensional configuration of the channel protein that corresponds to a model state with a short open time and distinguish it from another conformational state that is characterized by a long open time.

Second, and more importantly, aggregated Markov models are only defined up to equivalence [20,33–36] with other models having the same number of open and closed states. In particular, it can be shown that models with completely different adjacency matrices can describe the same process [35] although there is a canonical phase-type description, given, for example, by its Laplace–Stieltjes transform. Thus, interpreting the graphical structure of an aggregated Markov model as a description of possible transitions between different conformational states is not necessarily meaningful without further data. A related problem is the fact that some adjacency matrices lead to non-identifiable models, in particular, certain types of cyclic models are non-identifiable. Whereas it is unlikely that transitions between conformational states underlie any fundamental restrictions of this kind, only some of these transitions would be identifiable from experimental data. It is important to note that the described challenge of relating aggregated Markov models with biophysical processes does not restrict in any way their capability of statistically capturing the stochastic dynamics of ion channels. This only demonstrates that aggregated Markov models are a more abstract representation than they may appear to be at first glance.

In summary, because it is much less problematic to associate aggregates of states with different underlying biophysical states than individual states within an aggregate, interpreting mode switching as transitions between distinct biophysical states does not suffer from these difficulties. Chakrapani *et al.* [28–30] were able to restrict the KscA channel to one of its normally four modes by mutating a particular site of the amino acid sequence of the channel protein. Combining crystallography imaging and molecular dynamics modelling they could further demonstrate that the four modes were related to different conformational states of the channel. It is therefore likely that switching between distinct characteristic dynamical patterns in single channel data can be directly associated with the transition from one to another conformation of the channel protein. This implies that models which accurately represent mode switching can also be used to infer the time scales of transitions between biophysical states associated with these modes. This opens up the exciting possibility that we can gain insight into biophysical processes involved in ion channel gating by statistical analysis and modelling of single channel data rather than having to rely on more time-consuming experimental techniques such as crystallography or more laborious modelling techniques such as molecular dynamics.

Data accessibility. Data and code used for this study have been published on [github](https://github.com/merlinthemagician/icmcstat). The most current version can be obtained from <https://github.com/merlinthemagician/icmcstat>.

Authors' contributions. I.S., M.F., P.T. and E.J.C. designed the study and developed the mathematical theory. I.S. analysed the data. All authors wrote the paper and approved the final version.

Competing interests. We declare we have no competing interests.

Funding. This research was in part conducted and funded by the Australian Research Council Centre of Excellence in Convergent Bio-Nano Science and Technology (project number CE140100036). I.S. gratefully acknowledges funding from the German Academic Exchange Service (DAAD). P.T. is supported by the Australian Research Council (ARC) Laureate Fellowship FL130100039 and the ARC Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS).

Acknowledgements. The authors thank three anonymous reviewers for their helpful suggestions which greatly improved this article. I.S. completed the last stages of this study at the Felix Bernstein Institute for Mathematical Statistics (FBMS), Göttingen, Germany. He cordially thanks Axel Munk and his group for their friendly hospitality. The authors thank Larry Wagner and David Yule for making available their data for this study [15].

Appendix A. Mathematical background

The results presented in the main text are derived from the following properties of the Kronecker product and sum and some well-known results from linear algebra.

Proposition A.1 (Properties of Kronecker product \otimes and Kronecker sum \oplus). *The following properties of the Kronecker product and sums can all be found in Horn & Johnson [37].*

(i) *Transposition and conjugate transpose (Properties 4.2.4 and 4.2.5):*

$$(A \otimes B)^T = A^T \otimes B^T, \quad (A \otimes B)^* = A^* \otimes B^*. \quad (\text{A } 1)$$

(ii) *Compatibility of tensor product and matrix multiplication (Lemma 4.2.10): Let $A \in \mathbb{R}^{k_1 \times m_1}$, $C \in \mathbb{R}^{m_1 \times n_1}$, $B \in \mathbb{R}^{k_2 \times m_2}$, $D \in \mathbb{R}^{m_2 \times n_2}$.*

$$(A \otimes B)(C \otimes D) = (AC) \otimes (BD) \in \mathbb{R}^{k_1 k_2 \times n_1 n_2}. \quad (\text{A } 2)$$

(iii) *Eigenvalues of Kronecker sums $A \oplus B$ (Theorem 4.4.5): Let α, β denote eigenvalues of the square matrices A and B . Then the eigenvalues of $M = A \oplus B$ are*

$$\gamma = \alpha + \beta. \quad (\text{A } 3)$$

(iv) *Matrix exponentials of Kronecker sums (ch. 6, Problem 14): For square matrices $A \in \mathbb{R}^{m \times m}$ and $B \in \mathbb{R}^{n \times n}$:*

$$\exp(A \oplus B) = \exp(A) \otimes \exp(B) \in \mathbb{R}^{m \times m \times n \times n}. \quad (\text{A } 4)$$

If we cannot assume that a matrix has a complete set of eigenvectors so that it may not be diagonalizable we can still triangularize this matrix over the complex numbers \mathbb{C} . The process of triangulation can be described by the Schur decomposition.

Proposition A.2 (Schur decomposition). *For a square matrix $A \in \mathbb{R}^{m \times m}$ there exists a unitary matrix $\Theta \in \mathbb{C}^{m \times m}$ and an upper triangular matrix T such that*

$$T = \Theta^* A \Theta, \quad (\text{A } 5)$$

where Θ^* is the conjugate transpose of Θ ; (A 5) is known as the Schur decomposition.

Let $A \in \mathbb{R}^{m \times m}$ and $B \in \mathbb{R}^{n \times n}$ with Schur decompositions

$$T_A = V^* A V \quad \text{and} \quad T_B = W^* B W.$$

Schur decompositions for the Kronecker product $A \otimes B$ and the Kronecker sum $A \oplus B$ can then be obtained via

$$T_{A \otimes B} = (V \otimes W)^* A \otimes B (V \otimes W), \quad T_{A \oplus B} = (V \otimes W)^* A \oplus B (V \otimes W). \quad (\text{A } 6)$$

Proof. See Horn & Johnson [38], theorem 2.3.1. For (A 6), we refer to the proofs of Theorems 4.2.12 and 4.4.5 in [37]. ■

References

1. Neher E, Sakmann B. 1976 Single-channel currents recorded from membrane of denervated frog muscle fibres. *Nature* **260**, 799–802. (doi:10.1038/260799a0)
2. Colquhoun D, Hawkes AG. 1981 On the stochastic properties of single ion channels. *Proc. R. Soc. Lond. B* **211**, 205–235. (doi:10.1098/rspb.1981.0003)

3. Patlak JB, Gration KAF, Usherwood PNR. 1979 Single glutamate-activated channels in locust muscle. *Nature* **278**, 643–645. (doi:10.1038/278643a0)
4. Magleby KL, Pallotta BS. 1983b Calcium dependence of open and shut interval distributions from calcium-activated potassium channels in cultured rat muscle. *J. Physiol. Lond.* **344**, 585–604. (doi:10.1113/jphysiol.1983.sp014957)
5. Magleby KL, Pallotta BS. 1983a Burst kinetics of single calcium-activated potassium channels in cultured rat muscle. *J. Physiol. Lond.* **344**, 605–623. (doi:10.1113/jphysiol.1983.sp014958)
6. Siekmann I, Sneyd J, Crampin EJ. 2014 Statistical analysis of modal gating in ion channels. *Proc. R. Soc. A* **470**, 20140030. (doi:10.1098/rspa.2014.0030)
7. Colquhoun D, Hawkes AG. 1982 Properties of bursts of single ion channel openings and of clusters of bursts. *Phil. Trans. R. Soc. Lond. B* **300**, 1–59. (doi:10.1098/rstb.1982.0156)
8. Blatz AL, Magleby KL. 1986 Quantitative description of 3 modes of activity of fast chloride channels from rat skeletal-muscle. *J. Physiol. Lond.* **378**, 141–174. (doi:10.1113/jphysiol.1986.sp016212)
9. Ionescu L, White C, Cheung K-H, Shuai J, Parker I, Pearson JE, Foskett JK, Mak D-OD. 2007 Mode switching is the major mechanism of ligand regulation of InsP₃ receptor calcium release channels. *J. Gen. Physiol.* **130**, 631–645. (doi:10.1085/jgp.200709859)
10. Mak DO, Foskett JK. 2015 Inositol 1,4,5-trisphosphate receptors in the endoplasmic reticulum: a single-channel point of view. *Cell Calcium* **58**, 67–78. (doi:10.1016/j.ceca.2014.12.008)
11. Ullah G, Mak D-OD, Pearson JE. 2012 A data-driven model of a modal gated ion channel: the inositol 1,4,5-trisphosphate receptor in insect Sf9 cells. *J. Gen. Physiol.* **140**, 159–173. (doi:10.1085/jgp.201110753)
12. Siekmann I, Wagner II LE, Yule D, Crampin EJ, Sneyd J. 2012 A kinetic model of type I and type II IP₃R accounting for mode changes. *Biophys. J.* **103**, 658–668. (doi:10.1016/j.bpj.2012.07.016)
13. Ullah G, Ullah A. In press. Mode switching of inositol 1,4,5-trisphosphate receptor channel shapes the spatiotemporal scales of Ca²⁺ signals. *J. Biol. Phys.*
14. Mak D-OD, Cheung K-H, Toglia P, Foskett JK, Ullah G. 2015 Analyzing and quantifying the gain-of-function enhancement of IP₃ receptor gating by familial Alzheimer's disease-causing Mutants in Presenilins. *PLoS Comput. Biol.* **11**, e1004529. (doi:10.1371/journal.pcbi.1004529)
15. Wagner LE, Yule DI. 2012 Differential regulation of the InsP₃ receptor type-1 and -2 single channel properties by InsP₃, Ca²⁺ and ATP. *J. Physiol.* **590**, 3245–3259. (doi:10.1113/jphysiol.2012.228320)
16. Neuts MF. 1975 Probability distribution of phase type. In *Liber Amicorum Professor Emeritus H Florin*. Leuven, Belgium: Department of Mathematics, University of Louvain.
17. Neuts MF. 1981 *Matrix-geometric solutions in stochastic models: an algorithmic approach*. New York, NY: Dover Publications.
18. Fine S, Singer Y, Tishby N. 1998 The hierarchical hidden Markov model: analysis and applications. *Mach. Learn.* **32**, 41–62. (doi:10.1023/A:1007469218079)
19. Siekmann I, Wagner II LE, Yule D, Fox C, Bryant D, Crampin EJ, Sneyd J. 2011 MCMC estimation of Markov models for ion channels. *Biophys. J.* **100**, 1919–1929. (doi:10.1016/j.bpj.2011.02.059)
20. Siekmann I, Crampin EJ, Sneyd J. 2012 MCMC can detect non-identifiable models. *Biophys. J.* **103**, 1275–1286. (doi:10.1016/j.bpj.2012.08.014)
21. Hawkes AG, Jalali A, Colquhoun D. 1990 The distributions of the apparent open times and shut times in a single channel record when brief events cannot be detected. *Phil. Trans. R. Soc. Lond. A* **332**, 511–583. (doi:10.1098/rsta.1990.0129)
22. Hawkes AG, Jalali A, Colquhoun D. 1992 Asymptotic distributions of apparent open times and shut times in a single channel record allowing for the omission of brief events. *Phil. Trans. R. Soc. Lond. B* **337**, 383–404. (doi:10.1098/rstb.1992.0116)
23. Colquhoun D, Hawkes AG, Srodzinski K. 1996 Joint distributions of apparent open and shut times of single-ion channels and maximum likelihood fitting of mechanisms. *Phil. Trans. R. Soc. Lond. A* **354**, 2555–2590. (doi:10.1098/rsta.1996.0115)
24. Kolmogorov AN. 1936 Zur Theorie der Markoffschen Ketten. *Math. Ann.* **112**, 155–160. (doi:10.1007/BF01565412)
25. Kelly FP. 2011 *Reversibility and stochastic networks*, 2nd edn. Cambridge, UK: Cambridge University Press.

26. Cao P, Tan X, Donovan G, Sanderson MJ, Sneyd J. 2014 A deterministic model predicts the properties of stochastic calcium oscillations in airway smooth muscle cells. *PLoS Comput. Biol.* **10**, e1003783. (doi:10.1371/journal.pcbi.1003783)
27. Siekmann I, Cao P, Sneyd J, Crampin EJ. 2015 Data-driven modelling of the inositol trisphosphate receptor (IP₃R) and its role in calcium induced calcium release (CICR). In *Computational glioscience*, ch. 2 (eds MD Pittà, H Berry). Berlin, Germany: Springer.
28. Chakrapani S, Cordero-Morales JF, Perozo E. 2007 A quantitative description of KscA gating I: macroscopic currents. *J. Gen. Physiol.* **130**, 465–478. (doi:10.1085/jgp.200709843)
29. Chakrapani S, Cordero-Morales JF, Perozo E. 2007 A quantitative description of KscA gating II: single-channel currents. *J. Gen. Physiol.* **130**, 479–496. (doi:10.1085/jgp.200709844)
30. Chakrapani S, Cordero-Morales JF, Jogini V, Pan AC, Cortes DM, Roux B, Perozo E. 2011 On the structural basis of modal gating behaviour in K⁺ channels. *Nat. Struct. Mol. Biol.* **18**, 67–75. (doi:10.1038/nsmb.1968)
31. Vij R, Purohit P, Auerbach A. 2015 Modal affinities of endplate acetylcholine receptors caused by loop C mutations. *J. Gen. Physiol.* **146**, 375–386. (doi:10.1085/jgp.201511503)
32. Geng Y, Magleby KL. 2015 Modal gating of endplate acetylcholine receptors: a proposed mechanism. *J. Gen. Physiol.* **146**, 435–439. (doi:10.1085/jgp.201511534)
33. Fredkin DR, Montal M, Rice JA. 1985 Identification of aggregated Markovian models: application to the nicotinic acetylcholine receptor. In *Proc. of the Berkeley Conf. in Honor of Jerzy Neyman and Jack Kiefer*, vol. 1 (eds LML Cam, RA Olshen), pp. 269–289. Belmont, CA: Wadsworth.
34. Fredkin DR, Rice JA. 1986 On aggregated Markov processes. *J. Appl. Probab.* **23**, 208–214. (doi:10.1017/S0021900200106412)
35. Kienker P. 1989 Equivalence of aggregated Markov models of ion-channel gating. *Proc. R. Soc. Lond. B* **236**, 269–309. (doi:10.1098/rspb.1989.0024)
36. Bruno WJ, Yang J, Pearson JE. 2005 Using independent open-to-closed transitions to simplify aggregated Markov models for ion channel gating kinetics. *Proc. Natl Acad. Sci. USA* **102**, 6326–6331. (doi:10.1073/pnas.0409110102)
37. Horn RA, Johnson CR. 1994 *Topics in matrix analysis*. Cambridge, UK: Cambridge University Press.
38. Horn RA, Johnson CR. 1985 *Matrix analysis*. Cambridge, UK: Cambridge University Press.