A data-driven model of a modal gated ion channel: The inositol 1,4,5-trisphosphate receptor in insect Sf9 cells

Ghanim Ullah,[1] Don-On Daniel Mak,[2] and John E. Pearson[1]

[1]Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, NM 87544
[2]Department of Physiology, University of Pennsylvania, Philadelphia, PA 19104

The inositol 1,4,5-trisphosphate (IP_3) receptor (IP_3R) channel is crucial for the generation and modulation of intracellular Ca^{2+} signals in animal cells. To gain insight into the complicated ligand regulation of this ubiquitous channel, we constructed a simple quantitative continuous-time Markov-chain model from the data. Our model accounts for most experimentally observed gating behaviors of single native IP_3R channels from insect Sf9 cells. Ligand (Ca^{2+} and IP_3) dependencies of channel activity established six main ligand-bound channel complexes, where a complex consists of one or more states with the same ligand stoichiometry and open or closed conformation. Channel gating in three distinct modes added one complex and indicated that three complexes gate in multiple modes. This also restricted the connectivity between channel complexes. Finally, latencies of channel responses to abrupt ligand concentration changes defined a model with specific network topology between 9 closed and 3 open states. The model with 28 parameters can closely reproduce the equilibrium gating statistics for all three gating modes over a broad range of ligand concentrations. It also captures the major features of channel response latency distributions. The model can generate falsifiable predictions of IP_3R channel gating behaviors and provide insights to both guide future experiment development and improve IP_3R channel gating analysis. Maximum likelihood estimates of the model parameters and of the parameters in the De Young–Keizer model yield strong statistical evidence in favor of our model. Our method is simple and easily applicable to the dynamics of other ion channels and molecules.

The Journal of General Physiology

INTRODUCTION

Ca^{2+} is one of the most important signaling ions in that it controls numerous cellular functions such as neurotransmitter release, muscle contraction, and nuclear transcription (Berridge et al., 1998). Ca^{2+} performs such a wide range of tasks through a hierarchy of signaling patterns ranging from single-channel events to waves sweeping over entire organs (Berridge et al., 1998). In many cell types, the inositol 1,4,5-trisphosphate (IP_3) receptor (IP_3R) Ca^{2+} release channel plays a central role in orchestrating the hierarchical Ca^{2+} patterns. Thus, it is crucial to understand the mechanisms regulating the IP_3R. Considerable efforts and expertise have gone into modeling the modulation of IP_3R channel gating by its main physiological ligands: IP_3 and Ca^{2+} (De Young and Keizer, 1992; Sneyd and Dufour, 2002; Dawson et al., 2003; Mak et al., 2003; Shuai et al., 2007; Gin et al., 2009; Swaminathan et al., 2009). Most of these models can explain the Ca^{2+} dependence of the open probability (P_o) of IP_3R. Only the model in Mak et al. (2003) can also accurately reproduce the IP_3 dependence of P_o, but it required 14 parameters to do so. One of us (Mak) has conducted several studies of the native IP_3R channel from insect Sf9 cells, which has a primary sequence most closely related to the type 1 IP_3R isoform in vertebrate cells (Foskett et al., 2007), to elucidate various behaviors of the IP_3R channel, including the ligand dependence of its P_o and mean open- and closed-channel dwell times (τ_o and τ_c, respectively), its modal gating behaviors in various ligand concentrations, and the transient response of the channel to rapid changes in ligand concentrations. To date, no model of the IP_3R has accounted for all these observations. Here, we propose such a model.

MATERIALS AND METHODS

The many observations mentioned above have enabled us to develop a highly constrained mathematical model for the IP_3R channel gating kinetics, with the model structure and all of the parameters inferred directly from experimental data. The model consists of a Markov chain with 3 open states and 9 closed states, and requires 28 parameters. It reproduces the ligand-dependent equilibrium P_o, τ_o, and τ_c over wide ranges of cytoplasmic free Ca^{2+} ($\mathcal{C}$) and IP_3 ($\mathcal{I}$) concentrations (Ionescu et al., 2006); the Ca^{2+} dependence of various modal gating features, including the relative prevalence (the probability of finding the channel in a particular gating mode) and lifetimes of the gating modes of the IP_3R channel, and channel P_o in each of the gating modes reported

Correspondence to John E. Pearson: pearson@lanl.gov

Abbreviations used in this paper: CM, cooperative model; DYKM, De Young–Keizer model; IP_3, inositol 1,4,5-trisphosphate; IP_3R, IP_3 receptor.

The Rockefeller University Press $30.00
J. Gen. Physiol. Vol. 140 No. 2 159–173
www.jgp.org/cgi/doi/10.1085/jgp.201110753

in Ionescu et al. (2007); as well as the distributions of the channel response latencies observed after various abrupt changes in the concentrations of one or both of its ligands (Mak et al., 2007).

We accomplished this working under the following major assumptions: (a) the model is a finite-state Markov chain, (b) the ligand-binding kinetics obey the law of mass action, (c) the chain is fully connected, (d) the channel must be able to unbind ligands, and (e) the model should be as simple as possible. We show in the supplemental text, section 1.1 (see also Fig. S1), that the gating and binding kinetics of IP_3R are consistent with the detailed balance hypothesis.

Structurally, the IP_3R channel is an oligomer. However, we do not assume that the IP_3R subunits in the channel are independent or that the ligand-binding sites in each subunit are independent. We make no a priori assumption about sequential or independent binding of IP_3 and Ca^{2+} to the channel and therefore are consistent with experimental observations demonstrating that IP_3 and Ca^{2+} bind to the channel cooperatively but with no specific sequential requirements (Mak et al., 2007), unlike many previous models that assumed sequential (Tang et al., 1996; Marchant and Taylor, 1997) or independent (Swillens et al., 1994; Tang et al., 1996; Hirose et al., 1998; Fraiman and Dawson, 2004) IP_3 and Ca^{2+} binding.

Our one structural constraint is to search only over models with four IP_3-binding sites. Structure is at most an a priori weak constraint on dynamics. Beece et al. (1980) note that "the protein is not a rigid system in which a ligand moves in a fixed potential. Rather there is a strong mutual interaction between ligand and protein . . . A protein is not like a solid house into which the visitor (the ligand) enters by opening doors without changing the structure. Rather it is like a tent into which a cow strays." Another difficulty with structurally motivated models is that they tend to have large numbers of states that cannot be deduced from patch-clamp recordings. Eigen's "general allosteric model" has 55 states for a tetrameric channel in which each subunit can bind a single ligand and has two major conformations (Eigen, 1968). There is no a priori reason to believe that the IP_3R subunits have a particular number of conformations available to them. The IP_3R is known to have at least one IP_3-binding site and two Ca^{2+}-binding sites per monomer. The same reasoning that led to Eigen's model would result in an IP_3R model with thousands of states. A large Markov chain will not have rates that can be learned from the data because of the large number of parameters that would need to be estimated from data. We believe that there are probably numerous small Markov chains that all yield similar likelihoods for the observed data. Based on the quality of the fit, one can use Markov chain selection criteria to at least select between specific categories of Markov chains. Here, we construct a model with cooperativity in the IP_3-binding dynamics. We refer to this model as the "cooperative model" (CM). We compare the likelihood of the data given the CM to the likelihood of the same data given the most frequently used De Young–Keizer model (DYKM) (De Young and Keizer, 1992). The DYKM yields a somewhat higher likelihood for the P_o data than the CM (although such was not the case for *Xenopus laevis* oocyte P_o data). We also performed maximum likelihood fits to dynamic data: latency measurements as well as stationary time series. Fits to the latency data only and the combined latency and current trace data indicated that the CM outperforms the DYKM. Thus, the DYKM does a very good job reproducing the P_o data from Sf9 cells. However, it does not compete well with the CM when the recent kinetic data are taken into account.

We use the theory of aggregated binary reversible Markov chains (Colquhoun and Hawkes, 1981; Fredkin et al. 1985. Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer; Fredkin and Rice, 1986; Qin et al., 1997; Bruno et al., 2005) to derive a model via an iterative, data-driven approach. First, we consider the ligand dependencies of the equilibrium P_o, which provides evidence for the main ligand-binding stoichiometric complexes that must be included in a model Markov chain. A stoichiometric complex is specified by the number of Ca^{2+} ions (m) and IP_3 molecules (n) bound to the channel, and by the gating state of the channel (open or closed). The C_{32} (O_{32}) complex refers to a closed (open) channel bound to three Ca^{2+} ions and two IP_3 molecules. Such complexes can comprise more than one Markovian state (a state in a Markov chain from

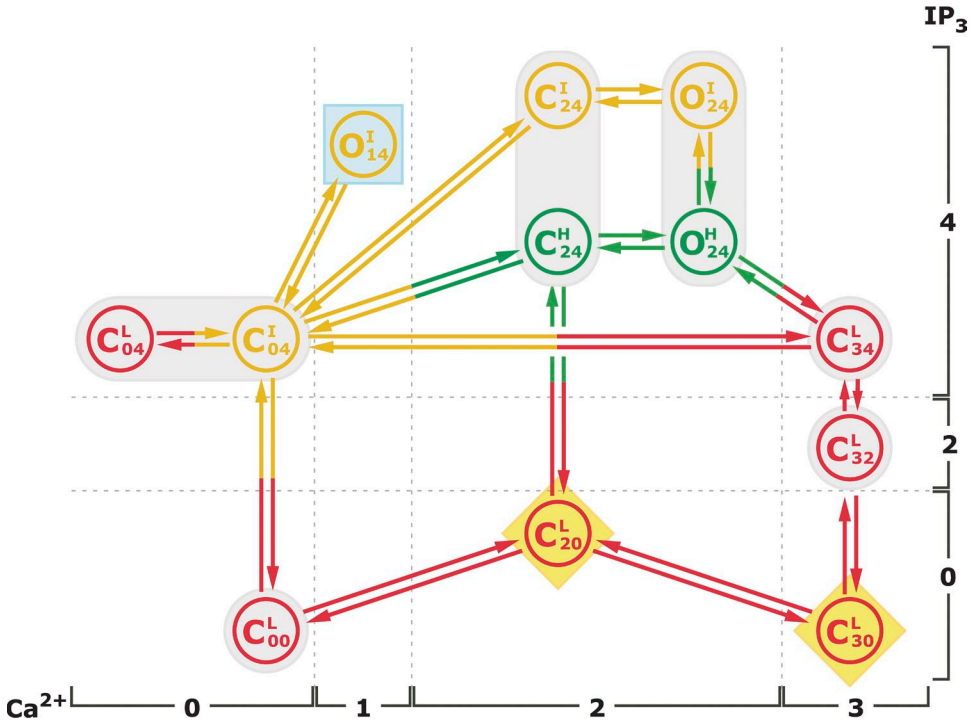

Figure 1. The kinetic scheme for the proposed model. The superscript of each state denotes the gating mode that state is in. The two numbers in the subscripts indicate the number of Ca^{2+} and IP_3, respectively, that is bound to the channel in that state. The model has nine closed states (C^L_{00}, C^L_{20}, C^L_{30}, C^L_{32}, C^L_{34}, C^L_{04}, C^I_{04}, C^I_{24}, and C^H_{24}) and three open states (O^I_{14}, O^I_{24}, and O^H_{24}). The elliptical shadings highlight complexes needed to account for the ligand dependencies of the channel P_o. The square shading highlights the complex required to fit the modal gating statistics, and the diamond shadings highlight complexes introduced as a result of IP_3 activation latency distribution. The transition rates between various states and the related flux parameters are listed in Tables S2 and S3, respectively.

which the channel exits with time-independent rate constant(s) so that it has a mono-exponential lifetime distribution). The number of such states in a complex cannot be ascertained from ligand dependencies of the P_o data, but in the absence of data to the contrary, assumption (e) demands one state per complex. The parameters that can be inferred from P_o are the equilibrium occupancies of complexes and contain no information regarding network topology (supplemental text, section 1.2). Nor can the total number of complexes necessarily be inferred from P_o. At this stage, if the ligand dependencies of channel P_o do not require a particular complex, simplicity demands that the complex not be explicitly included in the chain. Then we consider the Ca^{2+} dependence of the modal gating behaviors: the relative prevalence of and the channel P_o in each of the modes. These data provide evidence for the inclusion of one more complex and suggest that some complexes can be in more than one gating state.

Furthermore, the Ca^{2+} dependence of the lifetimes of the gating modes provides some broad constraints on the connectivity between the various gating states. Once we have an estimate of the important states in the model and their occupancies, we deduce the full model with additional states and connections, and the rates related to the connections by fitting the response latencies of the channel to abrupt changes in ligand concentrations and current trace data. The full model that we constructed is shown in Fig. 1.

Occupancy and flux

Instead of parameterizing our model with reaction rates, as done in most previous efforts to model the IP_3R channel gating, we use "state occupancy" and "flux" parameters as suggested in Yang et al. (2006). The two approaches are mathematically equivalent, but we find the occupancy–flux parameter approach simpler and more intuitive because it separates thermodynamic quantities (equilibrium occupancies) from kinetic quantities (reaction fluxes), or equivalently, it separates reaction rates from equilibrium constants. Occupancy and flux parameters are used also because occupancy determination is far more robust than state lifetime measurements. Failure to detect brief events (channel opening and closing or changing gating modes) can have a large impact on estimates of mean lifetimes of the states but does not affect state occupancy estimates significantly (supplemental text, section 1.3, and Fig. S2). Furthermore, the normalized occupancies are just rational functions of ligand. Thus, once the occupancies are learned by fitting the P_o function to the data (described below), they are fixed throughout the subsequent data-fitting process. This constraint is trivial to apply using occupancy parameters: we simply do not alter the occupancy parameters unless it can be done without significantly affecting channel P_o. Details of occupancy and flux parameters are given in the supplemental text, section 1.4.

Low occupancy states

Another concept that we required is the notion of short-lived low occupancy states (Ullah et al., 2012). The P_o data suggest that there are complexes with no IP_3 bound and complexes with four IP_3 bound. Clearly there must be complexes with one, two, and three IP_3 bound, but our analysis of channel P_o provides little evidence for them. We conclude that these states exist, but their occupancies are not high enough for them to be detected in the P_o data. Such low occupancy transition states are not included explicitly in the model and their occupancies are set to zero to satisfy the maximal simplicity assumption. Similarly, some complexes with no Ca^{2+} bound are linked directly to complexes with two Ca^{2+} bound in the model. However, the effects of these transition states are manifested in the particular functional form for the fluxes passing through them imposed by mass action and detailed balance. Proper treatment of low occupancy states is crucial, both for getting the rates right in the simplified model and for obtaining a model structure with learnable parameters. Applying such simplifications by excluding low occupancy states can result in a substantial reduction in the number of parameters required. More details regarding low occupancy states are given in the supplemental text, section 1.5. In the supplemental text, section 1.6, we derive the functional form of the transition rates between various model states (Fig. 1) and simplify the branches that involve low occupancy states by eliminating these states.

Complex occupancies

We identify the relevant (high occupancy) complexes from the ligand dependencies of channel P_o. Relative to the reference unliganded closed complex C_{00}, the occupancies of the C_{mn} and O_{mn} complexes are proportional to $\mathcal{C}^m\mathcal{I}^n$, with occupancy parameters

$$K_{C_{mn}} \text{ and } K_{O_{mn}},$$

respectively. One can choose any complex as a reference complex, but we found the unliganded complex to be the simplest. The occupancy parameter of a complex is the sum over all the states contained in the complex of the products of equilibrium association constants along any path connecting C_{00} to each state ($K_{C_{00}} = 1$). The normalized occupancy of C_{mn} and O_{mn} are

$$K_{C_{mn}}\,\mathcal{C}^m\mathcal{I}^n / Z \text{ and } K_{O_{mn}}\,\mathcal{C}^m\mathcal{I}^n / Z,$$

respectively, where Z, the total occupancy over all states, is given by

$$Z = Z_o + Z_c = \sum_{m=0}^{m_{max}} \sum_{n=0}^{n_{max}} K_{O_{mn}}\,\mathcal{C}^m\mathcal{I}^n + \sum_{m=0}^{m_{max}} \sum_{n=0}^{n_{max}} K_{C_{mn}}\,\mathcal{C}^m\mathcal{I}^n, \quad (1)$$

with m_{max} and n_{max} being the maximum number of Ca^{2+} and IP_3 that the channel can bind. The equilibrium P_o function is:

$$P_o(\mathcal{C},\mathcal{I}) = \frac{Z_o}{Z}. \quad (2)$$

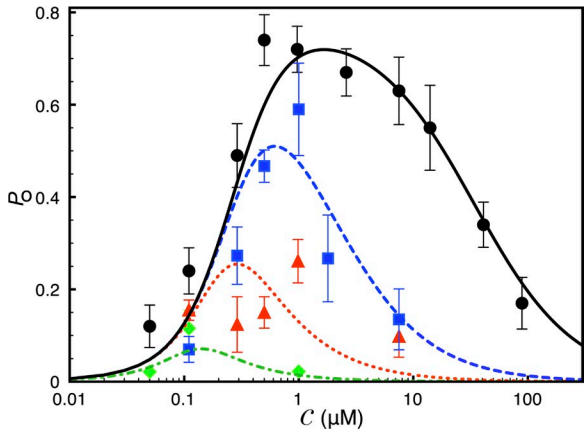

Figure 2. Ca^{2+} dependence of IP_3R channel P_o. Channel P_o under IP_3 concentration of 10 μM (bullets and solid line), 100 nM (squares and dashed line), 33 nM (triangles and dotted line), and 10 nM (diamonds and dotted-dashed line). The symbols represent mean experimental P_o, error bars represent standard errors of the mean, and lines are theoretical fits. The experimental data are taken from Ionescu et al. (2006) (supplemental text, section 1.10.1).

Our initial P_o-based search resulted in a best-fit (according to the AIC score; Mcquarrie, 1998) model with five closed complexes (C_{00}, C_{04}, C_{24}, C_{32}, and C_{34}) and one open complex (O_{24}) (Fig. 1). We first used the Mathematica routine "NonlinearModelFit," which does least squares and assumes a constant variance, meaning that it assumed the P_o function was Gaussian distributed with constant variance. We also assumed a variance proportional to P_o (1-P_o) and found little difference. Both distribution functions for the P_o data returned similar models. The fit to the P_o data are shown in Fig. 2. Although not essential to the P_o fit, the O_{14} complex is necessary to account for the Ca^{2+} dependence of channel P_o in the I gating mode (mode with intermediate P_o; see Modal occupancies). The C_{20} and C_{30} complexes are required to account for the latency distribution observed (see Model development using latencies), but the occupancies of these two complexes are so low that they do not contribute significantly to P_o or affect the quality of the P_o fits in Fig. 2. Theoretically, the P_o can be expressed as a function of the eight occupancy parameters of the nine complexes (occupancy of the reference C_{00} complex is 1). However, the terms 1,

$$K_{C_{20}}\,\mathcal{C}^2\text{, and }K_{C_{30}}\,\mathcal{C}^3$$

are all small relative to the other terms in Z for the values of $\mathcal{I}$ used in the P_o measurements, and O_{14} complex is only required for the channel P_o in I mode. This implies that the P_o is effectively determined by the five occupancy parameters for the complexes with IP_3 bound

$$(K_{C_{04}},\ K_{C_{24}},\ K_{C_{32}},\ K_{C_{34}},\text{ and }K_{O_{24}}).$$

We took the occupancy parameters to be as small as possible without altering our estimate of the P_o.

Subsequent searches yielded other models with better AIC scores (for the fit to Sf9 P_o data) than the one we discuss here. However, we were unable to account for the modal gating statistics using these other models. One such model that gave a better AIC score on the Sf9 P_o data was DYKM. However, the CM performed better than the DYKM when fitted to the P_o data from *Xenopus* oocytes (see Comparison between the CM and the DYKM). The value of

$$K_{C_{04}}$$

that we ended up with gives the geometric mean for the IP_3 affinities along the path connecting C_{00} to C_{04} of ~8 nM. The numerical values of the occupancy parameters are given in Table S1. In the remainder of this article, we construct a model based on these complexes.

Modal occupancies

The Sf9 IP_3R channel exhibits three distinct gating patterns or modes: a quiescent L mode with low P_o in which brief openings are separated by long quiescent periods with long τ_o, an I mode with intermediate P_o in which the channel opens and closes rapidly with short τ_c and τ_o, and an H mode with high P_o in which the channel gates in bursts (Ionescu et al., 2007). All three modes are observed under all conditions for which the channel gates, and the channel spontaneously changes among all three modes, even under constant ligand conditions (Fig. 3 A). The gating characteristics of the channel (P_o, τ_c, and τ_o) remain remarkably consistent in each gating mode in all ligand conditions so the ligand dependencies of overall channel P_o, τ_c, and τ_o come from the ligand dependencies of the relative prevalence (or normalized occupancies), π^M, of the gating modes (M can be L for low mode, I for intermediate mode, and H for high mode) (Ionescu, et al., 2007). The key modal statistical data at our disposal are shown in Fig. 4. Each of the three plots shows nine data points corresponding to $\mathcal{C}$ = 0.1, 1, and 89 µM ($\mathcal{I}$ = 10 µM for all data). The squares, bullets, and triangles in Fig. 4 represent the L, M, and H modes, respectively. To account for these observed modal gating behaviors, we have to extend our minimal model by splitting some complexes into two states when required, without adding new complexes other than O_{14}, which is required for the channel P_o in I mode.

Fig. 4 A shows the conditional open probabilities, $P_o^M = Z_o^M / Z^M$, within each mode. P_o^L and P_o^H are nearly independent of $\mathcal{C}$ (≈0.007 and 0.8, respectively), whereas P_o^I exhibits biphasic Ca^{2+} dependence: rising from ~0.2 at $\mathcal{C}$ = 100 nM to 0.3 at $\mathcal{C}$ = 1 µM, and then dropping back to 0.2 at $\mathcal{C}$ = 89 µM.

Because of the ligand independence of P_o^H, we postulated that the H mode comprises two states, O_{24}^H and C_{24}^H, so

$$P_o^H = K_{O_{24}^H}\,\mathcal{C}^2\mathcal{I}^4 / (K_{C_{24}^H}\,\mathcal{C}^2\mathcal{I}^4 + K_{O_{24}^H}\,\mathcal{C}^2\mathcal{I}^4) = K_{O_{24}^H} / (K_{C_{24}^H} + K_{O_{24}^H})$$

where we have split the C_{24} complex into C_{24}^H and C_{24}^I states, and the O_{24} complex into O_{24}^H and O_{24}^I states (also see next paragraph). Any combination of open and closed states (at least one open and one closed state) with the same number of $\mathcal{C}$ and $\mathcal{I}$ bound would give ligand-independent P_o^H. The only such

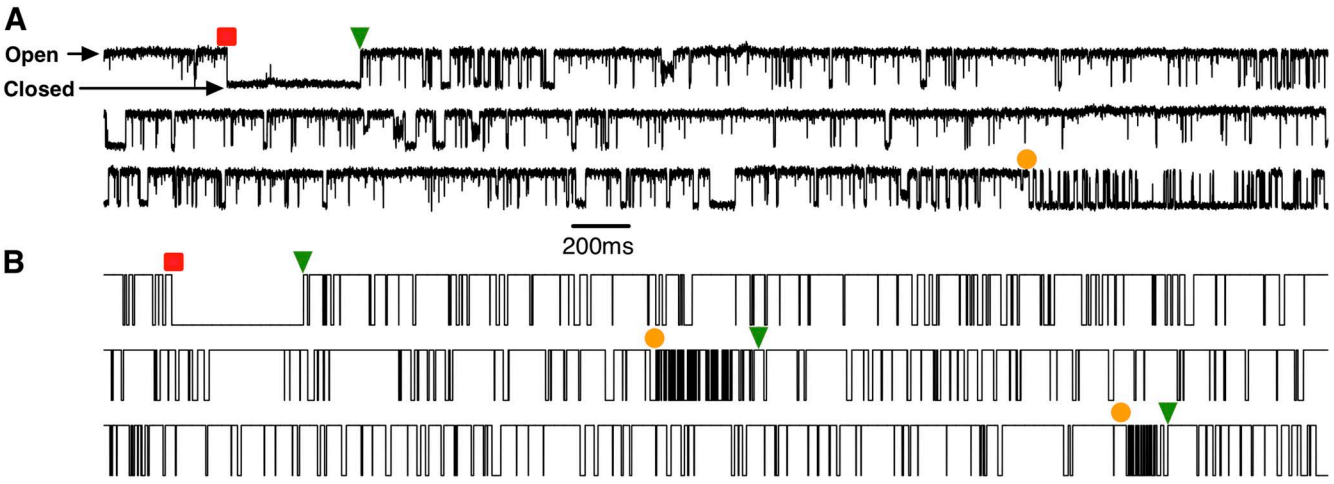

Figure 3. Modal gating time series. (A) Single-channel current record of IP_3R channel gating in the presence of $\mathcal{C}$ = 1 µM and $\mathcal{I}$ = 10 µM. (B) Simulated binary on/off single-channel current record generated from the model. The squares, bullets, and triangles indicate a transition to the L, I, and H gating mode, respectively.

combination available in the model is of states with two $\mathcal{C}$ and four $\mathcal{I}$ bound. Considering any other combination would require adding more states to the model, which is not warranted by our simplicity assumption.

If the I mode contains only states with two Ca^{2+} and four IP_3 bound, P_o^I will also be ligand independent. Thus, we postulate the O_{14} complex besides the C_{24} and O_{24} complexes are in the I mode. However, this will only give P_o^I monotonically rising as $\mathcal{C}$ is increased. Adding a C_{34}^I state can generate a biphasic dependence of P_o^I on $\mathcal{C}$, but will cause the relative prevalence of the I mode (π^I) to increase with increasing $\mathcal{C}$ (Fig. S4), which is not observed (Fig. 4 B). Thus, the peak in P_o^I necessitates the C_{04}^I state besides C_{24}^I, O_{14}^I, and O_{24}^I, so

$$P_o^I = (K_{O_{14}^I}\mathcal{C} + K_{O_{24}^I}\mathcal{C}^2) / [K_{C_{04}^I} + K_{O_{14}^I}\mathcal{C} + (K_{O_{24}^I} + K_{C_{24}^I})\mathcal{C}^2].$$

Here, we have split the C_{04} complex into C_{04}^I and C_{04}^L states, as required by the observation regarding I and L modes, and $O_{04}^I = O_{14}$. Adding the O_{14} complex worsens the AIC score for the model fit to the P_o data (Fig. 2) by 1 but improves the quality of the fit by improving the log likelihood (see Eq. 5 below). Adding any other open states such as O_{04} or O_{34} worsens both the AIC and log-likelihood scores by a factor much higher than 1.

According to the AIC and BIC criteria (Mcquarrie, 1998), there is insufficient justification from experimental evidence to assume that the C_{34}, C_{32}, and C_{00} complexes consist of more than one state. For example, splitting the C_{34} complex into two states will add one more parameter to the model and will increase the AIC score by a factor of 2 (see the factor $2k$ in Eq. 5 below), without changing the P_o given by the model (Eq. 2). So we assign them as C_{34}^L, C_{32}^L, and C_{00}^L to satisfy the simplicity assumption.

Fig. 4 B shows the Ca^{2+} dependence of $\pi^M = Z^M/Z$. A fit of the P_o data in various $\mathcal{C}$ and $\mathcal{I}$ (curves in Fig. 2) followed by fits to the P_o^M and π^M data in different $\mathcal{C}$ (curves in Fig. 4, A and B, respectively) using the Mathematica routine NonlinearModelFit (see Complex occupancies) determines the values of the occupancy parameters for the various states in the three gating modes.

By this point, the number of states in the model and the occupancy parameters have been determined. We will determine the connections between these states guided by the observed modal lifetimes and latency distributions. We will evaluate all possible connections and choose only those that lead to the observed behavior of the channel.

Modal lifetimes

Some, but not all, connectivities between the various states can be inferred from the observed modal gating behaviors. Channel opening and closing events were observed as the channel remained in the same mode; therefore, there are direct connections between C_{24}^I and O_{24}^I, and between C_{24}^H and O_{24}^H. Furthermore, the three modes are all connected to allow modal changes among the three modes (Ionescu et al., 2007).

The mean H mode lifetime, τ^H, has a maximum as a function of $\mathcal{C}$ (Fig. 4 C, triangles). This suggested that the H mode is connected to C_{34}^L to yield a short τ^H at high $\mathcal{C}$. We also connect the H mode to C_{04}^I so the rate from the H mode to C_{04}^I increases and τ^H decreases with decreasing $\mathcal{C}$. It should be noted that although moving from the H mode to the C_{04}^I state involves Ca^{2+} dissociation from the channel, the effective rate of the reaction is $\mathcal{C}$ dependent (see Low occupancy states). The C_{04}^I state is connected to the C_{04}^L state to account for the τ^L and τ^I in low $\mathcal{C}$. The long τ^L (Fig. 4 C, $\sim$ 3 s) indicates a low rate from C_{04}^L to C_{04}^I.

Finally, we connect the I and H mode open states via Ca^{2+}-independent rates to get the appropriate H mode lifetime under optimal ligand conditions. Further determination of the connectivity of the model requires consideration of channel response latencies (see Model development using latencies).

The remaining connections and the final values of the flux parameters will be determined below using the fit to the latency and time-series data.

Model development using latencies

$\mathcal{C}$ and $\mathcal{I}$ on the cytosolic side of the IP_3R channel were abruptly changed in nuclear patch-clamp experiments in the cytoplasmic-side-out configuration using rapid perfusion solution exchange techniques to observe the latencies for the response of the channel to ligand concentration changes (Mak et al., 2007). $\mathcal{C}$ was changed among low (<10 nM, which we will refer to as 0 for notational simplicity), optimal (2 µM) and inhibitory (300 µM) levels, and $\mathcal{I}$ between 0 and high (10 µM) levels. ($\mathcal{C}$, $\mathcal{I}$) were switched from nonactivating (2 µM, 0), (0, 10 µM), and (0, 0) to optimal (2 µM, 10 µM) to measure IP_3 activation, Ca^{2+} activation, and Ca^{2+} and IP_3 activation latencies, respectively. To measure corresponding deactivation latencies, ($\mathcal{C}$, $\mathcal{I}$) were changed from optimal back to various nonactivating combinations. Vertebrate IP_3R channels exhibit spontaneous activity in (0, 0) (Mak et al., 2003). However, the insect Sf9 IP_3R does not exhibit such spontaneous

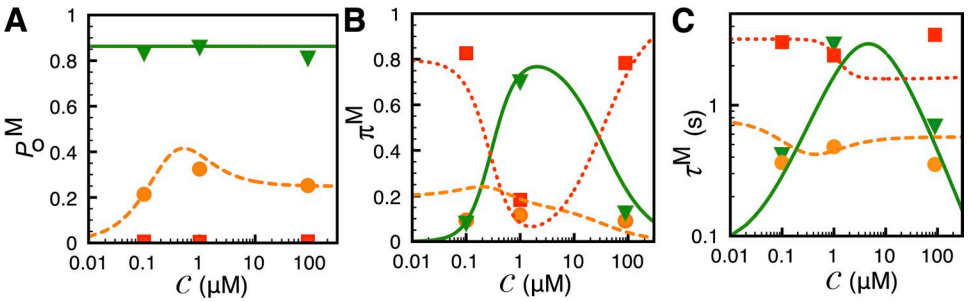

Figure 4. Ca^{2+} dependence of IP_3R channel modal behavior. (A) Channel open probability P_o^M in various gating modes, (B) relative prevalence π^M, and (C) mean lifetimes τ^M of the various modes (M can be L, I, or H). Symbols and curves represent the mean modal gating characteristics and theoretical fits, respectively. Symbols and curves for different modes are: triangles and solid lines, H; bullets and dashed lines, I; squares and dotted lines, L. Error bars in A–C are smaller than the symbols. The experimental data are taken from Ionescu et al. (2007) (supplemental text, section 1.10.2).

activity (Mak et al., 2007), so the Ca^{2+} and Ca^{2+} and IP_3 activation and deactivation latencies can be unambiguously determined. ($\mathcal{C}$, $\mathcal{I}$) were changed from optimal to inhibitory (300 µM, 10 µM) and from (300 µM, 10 µM) to optimal to measure latencies of inhibition and recovery from inhibition, respectively. Latencies after activating or recovery solution switches are the durations between solution switches and the first observed openings of the channel in response to the changes, whereas latencies after deactivating or inhibitory solution switches are the durations from the switches to the last observed closings of the channel as its activity terminated. Histograms of these latencies were built up from repeated solution switches in multiple experiments as shown in Fig. 5.

Additional experiments were performed in which ($\mathcal{C}$, $\mathcal{I}$) was switched from (0, 10 µM) to (300 µM, 10 µM) and back. In most of these switches, a burst of channel activity was observed before the channel was inhibited by high $\mathcal{C}$. But in 9 out of 103 switches, no channel open activity was observed. When $\mathcal{C}$ was decreased from 300 µM to 0, the channel was transiently activated in only 6 out of 94 experiments (Mak et al., 2007). These observations indicate that C_{34}^{L} has a direct connection to C_{04}^{I} without passing through the O_{24} complex.

Both the distributions of the inhibitory and inhibition recovery latencies after ($\mathcal{C}$, $\mathcal{I}$) switching between (2 µM, 10 µM) and (300 µM, 10 µM) can be fitted with single-exponential curves with no deficit of short latencies (Mak et al., 2007). This suggests that the O_{24}^{H} state is directly connected to the C_{34}^{L} state.

The Ca^{2+} and IP_3 activation latencies are short (Fig. 5 E), indicating that C_{00}^{L} is directly connected to C_{04}^{I} without going through C_{04}^{L} because rate from C_{04}^{L} to C_{04}^{I} is too slow, as revealed by long τ^{L}. Moreover, deficit in short latencies (<20 ms) for Ca^{2+} and IP_3 activation and Ca^{2+} activation (Mak et al., 2007) suggests that C_{04}^{I} is not directly connected to O_{24}^{H}. Thus, it should be connected to C_{24}^{H}.

Furthermore, to satisfy assumption (d) and allow IP_3 unbinding from the C_{34}^{L} and C_{32}^{L} states when $\mathcal{I}$ is dropped to 0 in $\mathcal{C}$ = 300 µM, we postulate that C_{30}^{L} is not a transition state, and it must be connected to C_{34}^{L} through C_{32}^{L}. Similarly, to allow Ca^{2+} to unbind from C_{30}^{L} in the absence of IP_3, it must be connected to C_{00}^{L} through C_{20}^{L}; and to allow Ca^{2+} to unbind from O_{14}^{I}, it must be connected to C_{04}^{I}.

The three distinguishable peaks in the IP_3 activation latency distribution (Fig. 5 A) require at least three pathways connecting the closed states with no IP_3 bound to the O_{24}^{H} state. Thus, we include in our model one more closed state, C_{20}^{L}, besides the C_{00}^{L} and C_{30}^{L} states. C_{20}^{L} is connected to C_{24}^{H} to maintain the deficit in short latencies for IP_3 activation.

Finally, to provide two pathways for the open states to reach C_{04}^{I} to account for the two peaks in the logarithmically binned latency histogram for Ca^{2+} deactivation (Fig. 5 D), C_{24}^{I} is connected to C_{04}^{I}. At this point, we have inferred the occupancies of all 12 Markovian states and the connections shown in our final model (Fig. 1) based on experimental observations. This completes the determination of states and the connections between various states in the model.

Next, we will determine the values of flux parameters involved in the transition rates between various states by fitting the model to the latency and time-series data. First, we set the initial values of flux parameters in all reactions, as described in the supplemental text, section 1.7. We will use these values as initial conditions for the global fit to the latency and time-series data described below. Although we estimate the initial values of flux parameters from observations, the initial guess proved unnecessary for the fit. The fit converged to the same final parameter values for random initial conditions.

Maximum likelihood and AIC calculations

To determine various flux parameters, we performed maximum likelihood fits on current traces of channel gating under constant conditions of $\mathcal{I}$ = 10 µM and $\mathcal{C}$ = 0.1, 1, and 89 µM, and on the latency data. To accomplish this task, we used our own implementation and calculated the relevant "log-likelihood" functions for the CM in the open-source programming language "Octave." We used the function "nelder_mead_min" in the optimization tool kit to minimize the likelihood scores ($-\log(\text{likelihood}(\text{data}))$) of the latency and time-series data by varying the 18 free flux parameters (while holding the occupancy parameters fixed).

The (negative) of the logarithm of the likelihood function for the latency data are minus the sum of the logarithms of the likelihoods of all the latencies:

$$-\sum_i \log(f_{lat}(t_i)),$$

where $f_{lat}(t_i)$ is the probability density function for latency and is explained in the next section. Rather than varying the rates and constraining them to be positive, we varied the log base 10 of the parameters without constraint. As discussed in the next section, $f_{lat}(t_i)$ involves the generator matrix. It is shown in the supplemental text, section 1.2, that when detailed balance is obeyed, the generator matrix is similar to a symmetric matrix. We performed the symmetrizing similarity transformation on the generator matrix whenever all the occupancies relevant to a given experiment were nonzero. This improves the condition number for eigenvalue calculations (see section 8.2 in Cheney and Kincaid, 2007).

For the current traces, we calculated the log of the likelihood function:

$$\begin{aligned} log(f(tc_1,to_1,tc_2,to_2,\ldots tc_n,to_n)) = log(\pi_C \exp(Q_{CC}tc_1)Q_{CO}\exp(Q_{OO}to_1) \\ Q_{OC}\exp(Q_{CC}tc_2)Q_{CO}\exp(Q_{OO}to_2)\ldots \\ \exp(Q_{OO}to_n)u_O), \end{aligned} \tag{3}$$

where π_C is the initial probability of closed states being occupied at equilibrium, to_i and tc_i are the ith opening and closing in the time series, respectively, and Q_{OC} is a matrix of the transition rates from all open to all closed states, etc. For the CM, this was accomplished as described in Qin et al. (1997). We could not use QUB (Qin et al., 1997) for this because of the complex form of our rates. Nor could we perform a brute force search through the trillions of candidate model topologies using QUB. Similarly for the DYKM, we could not use QUB because of the complex form of the 120-state model (see Comparison between the CM and the DYKM for details about DYKM). For data obtained at equilibrium, $\pi_C = W_O Q_{OC}/J$ with $J = W_O Q_{OC} u_C$, where u_C is an N_C (the number of close states) component vector of all 1's. w_C and w_O are the equilibrium occupancies of the closed and open states, respectively. For the CM, $N_C = 9$ and the number of open states, $N_O = 3$. Thus Q_{CC}, Q_{OO}, Q_{CO}, and Q_{OC} are 9 × 9, 3 × 3, 9 × 3, and 3 × 9 matrices, respectively.

We did two different fits: one in which we fit to all the latency data and none of the time-series data, and a second in which we fit both the latency data and the time-series data. The total log likelihood of all data used in the fit was calculated as

$$log(likelihood(data)) = \sum_{i=1}^{N_{exp}} log(likelihood(data_i)), \tag{4}$$

where N_{exp} is the number of experiments, and $data_i$ is the dataset (latencies or time series) from experiment i. There are eight

latency experiments where each experiment consists of hundreds of trials and each trial represents a latency measurement. The time-series data consists of three experiments (each experiment results in one time series), one each for $\mathcal{C}$ = 0.1, 1, and 89 μM. Finally, the AIC score is given by

$$AIC = -2log(likelihood(data)) + 2k, \quad (5)$$

where k is the number of parameters in the model. The CM has 10 occupancy and 18 flux parameters (k = 28).

To minimize the number of parameters, we assume that the effective rates for the reactions $C_{04}^{I} \rightarrow C_{24}^{I}$, $C_{00}^{L} \rightarrow C_{20}^{L}$, and $C_{04}^{I} \rightarrow C_{24}^{H}$ are the same. Similarly, the effective rates for the reactions $C_{04}^{I} \leftarrow C_{24}^{I}$, $C_{00}^{L} \leftarrow C_{20}^{L}$, and $C_{04}^{I} \leftarrow C_{24}^{H}$ are the same. Relaxing these constraints does not improve the fit to the data. The functional form of various transition rates for the CM is given in Table S2. Parameter values obtained from the fits to the latency data alone and latency plus time-series data are given in Table S3.

Online supplemental material

The supplemental text provides a more in-depth discussion and derivation of various statements made in the main text. In sections 1.1, 1.2, and 1.3, respectively, we show that the IP_3R obeys detailed balance (Fig. S1), the equilibrium occupancy data contain no information about the network connectivity, and missing events cause more error in τ_o and τ_c (Fig. S2). Sections 1.4, 1.5, 1.6, and 1.7 discuss the concepts of occupancy and probability flux, low occupancy states, simplification of the model, and parameter estimation, respectively, in detail. The derivation of mathematical expressions for τ_o and τ_c, and dwell-time distributions are given in sections 1.8 and 1.9. The experimental methods are described in section 1.10. Fig. S3 plots the ligand dependence of some rates in the model, and Fig. S4 proves that state C_{34}^{L} cannot be included in the I mode. Fig. S5 shows that the mean IP_3 activation latency has a minimum at intermediate $\mathcal{C}$ value. The parameters and transition rates for the CM are given in Tables S1–S3, whereas those for the DYKM are given in Tables S4 and S5. The supplemental material is available at http://www.jgp.org/cgi/content/full/jgp.201110753/DC1.

RESULTS

Fitting modal lifetimes with the model

To reproduce the observed modal lifetimes (Fig. 4 C, τ^M) with our model, τ^M are expressed in terms of the occupancy and flux parameters in the model. In general, the lifetime τ^X of any aggregate X (complex, mode, or other combinations of Markovian states) in an aggregated Markov chain is given by

$$\tau^X = \frac{Z^X}{J_X},$$

where Z^X is the unnormalized occupancy of aggregate X, and J_X is the unnormalized flux out of that aggregate. J_X is the sum of all the fluxes from all reactions from all Markovian states contained in the aggregate X to Markovian states not contained in X, so

$$J_X = \sum\nolimits_S J_S = \sum\nolimits_S \left(Z_S \sum\nolimits_U r_{SU} \right),$$

where $\sum_s$ is summing over all Markovian states S in the aggregate X, $\sum_U$ is summing over all Markovian states U that are not in X, and r_{SU} is the rate from a state S to a state U (supplemental text, section 1.6, and Table S2). The modal lifetimes given by the model are shown by the lines in Fig. 4 C.

Fitting latency distributions with the model

To use the CM to account for the distributions of latencies observed in the rapid perfusion solution exchange experiments, we express the latencies in terms of occupancy and flux parameters. When the perfusion solution to which the channel is exposed is changed at time t = 0, the system is initially in an aggregate X containing N_X states. As time goes by, it will transition into another aggregate Y with N_Y states, which is disjoint from X. We assume that $N_X + N_Y = N$, the total number of states in the system. If Q is the generator matrix for the full system, we partition Q as

$$Q = \begin{pmatrix} Q_{XX} & Q_{XY} \\ Q_{YX} & Q_{YY} \end{pmatrix}.$$

Using the formulation developed above, the ij element in Q is J_{ij}/Z_i. At time t = 0, the occupancies of initial states (all states in X) are specified in the vector of initial normalized occupancies $\pi_{(0)}$. Let u_X and u_Y be the N_X and N_Y component vectors, respectively, in which all elements are 1, so $\pi_{(0)}\ u_X = 1$. The probability density function for the latency can be expressed as (Colquhoun and Hawkes, 1981; Fredkin et al. 1985. Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer; Qin et al., 1997; Bruno et al., 2005):

$$f_{lat}(t) = \pi_{(0)} e^{Q_{XX} t} Q_{XY} u_Y, \quad (6)$$

and the probability value, $P_{[t_{min}, t_{max}]}$, for each bin (between $t = t_{min}$ and t_{max}) in the model latency histograms $= \int_{t_{min}}^{t_{max}} f_{lat}(t)dt$. The best fits to the latency distributions obtained by maximizing the likelihood of the latency data are shown by solid lines in Fig. 5. We also performed a maximum likelihood estimate of the parameters using the latency data and stationary time series shown by dotted lines in Fig. 5.

We apply this to our measurement of IP_3 channel response latencies, taking care to choose states X and Y, as well as the initial occupancy vector $\pi_{(0)}$ correctly, i.e., knowing the set of states in which the channel gates before and after the jump in the solution to which the channel is exposed. For activation and recovery from inhibition latencies, the choices of X and Y are more straightforward. The experiment starts with the channel being closed, and the latency is the time until the first channel opening as the channel enters an open state after the change of $\mathcal{C}$ and/or $\mathcal{I}$. Thus, for activation and recovery from inhibition experiments, the

aggregate X contains the nine closed states in the CM, tabulated in related matrices and vectors in this order: C^{L}_{00}, C^{L}_{20}, C^{L}_{30}, C^{L}_{32}, C^{L}_{34}, C^{L}_{04}, C^{I}_{04}, C^{H}_{24}, C^{I}_{24}. The aggregate Y contains the three open states tabulated in matrices and vectors in this order: O^{I}_{14}, O^{I}_{24}, and O^{H}_{24}. $\pi_{(0)}$ is a vector having the initial normalized occupancies of the closed states. For the deactivation latencies, X is composed of O^{I}_{14}, O^{I}_{24}, O^{H}_{24}, C^{I}_{24}, and C^{H}_{24}, and the remaining states are included in the aggregate Y.

For IP_3 activation in constant 2 μM Ca^{2+}, the logarithmically binned histogram for experimental latencies (Fig. 5 A, bars) has three peaks, the main one at ~80 ms and two minor ones at ~250 ms and 2 s (Mak et al., 2007), thus giving a mean IP_3 activation latency of ~500 ms, even though the majority of the activation latencies are shorter than 100 ms. This indicates that the Markovian model for ligand regulation of IP_3R channel has at least three distinct branches involved in IP_3 activation (Mak et al., 2007). This is a natural feature of the CM. With $\mathcal{C} = 2$ μM,

$$\pi_{(0)} = (1, K_{C^{L}_{20}}\mathcal{C}^2, K_{C^{L}_{30}}\mathcal{C}^3, 0, 0, 0, 0, 0, 0) / Z_C,$$

where

$$Z_C = 1 + K_{C^{L}_{20}}\mathcal{C}^2 + K_{C^{L}_{30}}\mathcal{C}^3,$$

because only the C^{L}_{00}, C^{L}_{20}, and C^{L}_{30} states are occupied in the absence of IP_3. The C^{L}_{20} state is the predominantly occupied one from which the channel can be activated

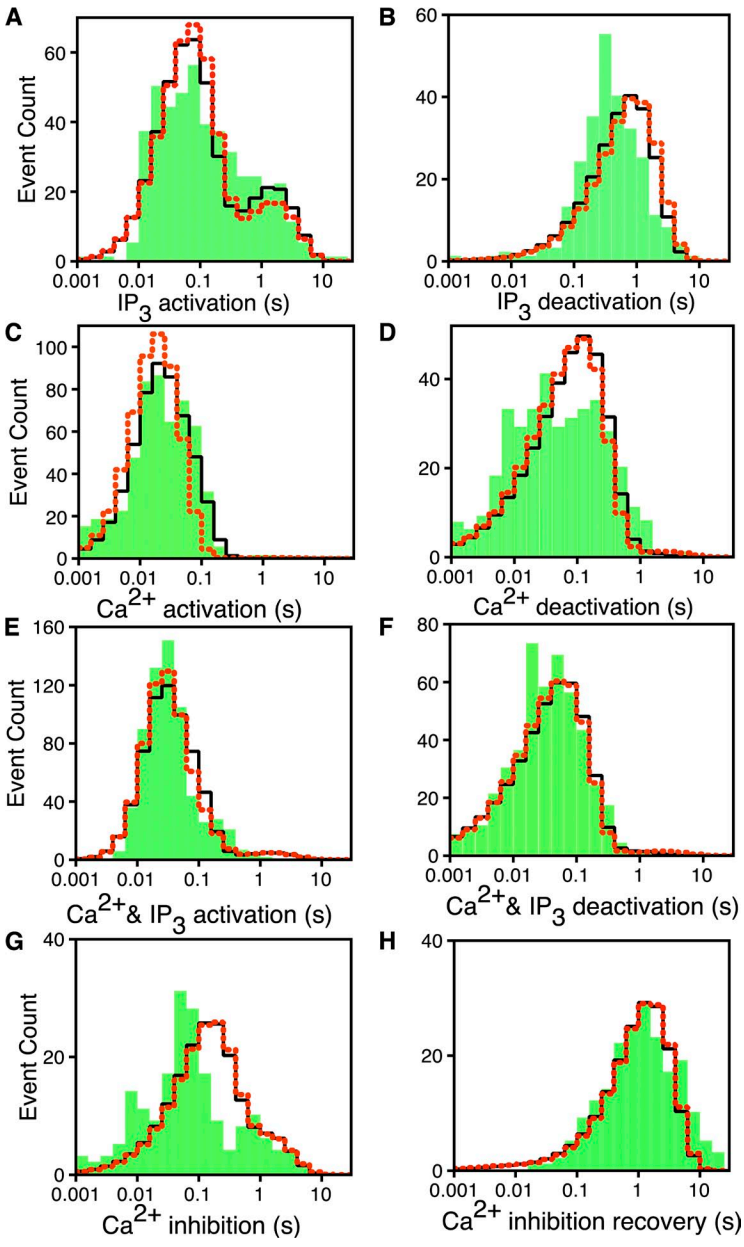

Figure 5. Logarithmically binned latency histograms of IP_3R channels responding to changes in $\mathcal{C}$ and $\mathcal{I}$. In each case, ($\mathcal{C}$, $\mathcal{I}$) are changed as indicated: (A) IP_3 activation: (2 μM, 0 μM) → (2 μM, 10 μM). (B) IP_3 deactivation: (2 μM, 10 μM) → (2 μM, 0 μM). (C) Ca^{2+} activation: (0 μM, 10 μM) → (2 μM, 10 μM). (D) Ca^{2+} deactivation: (2 μM, 10 μM) → (0 μM, 10 μM). (E) Ca^{2+} and IP_3 activation: (0 μM, 0 μM) → (2 μM, 10 μM). (F) Ca^{2+} and IP_3 deactivation: (2 μM, 10 μM) → (0 μM, 0 μM). (G) Ca^{2+} inhibition: (2 μM, 10 μM) → (300 μM, 10 μM). (H) Ca^{2+} inhibition recovery: (300 μM, 10 μM) → (2 μM, 10 μM). The experimental latency distributions given by the bars are taken from Mak et al. (2007). The solid and dotted lines, respectively, are from the CM when fit to the latency data alone and to the latency and time-series data simultaneously.

with short latencies by rapidly binding four IP_3, transiting through C^H_{24} to O^H_{24}. Longer latencies can be caused by channels initially occupying the C^L_{00} and C^L_{30} states being activated with longer latencies. Instead, it generates a distribution with one peak at ~80 ms for activating the channel from both C^L_{00} and C^L_{20} states, and another peak ~2 s corresponding to the channel going from C^L_{30} via either C^L_{34} or C^L_{20} to O^H_{24}, which evidently takes a long time as a result of the slow release of the third Ca^{2+}.

To fit the IP_3 deactivation latency histogram (Fig. 5 B), we assume that the initial occupancies of the states are their equilibrium occupancies in $\mathcal{I}$ = 10 µM and $\mathcal{C}$ = 2 µM. The case of Ca^{2+} activation in constant $\mathcal{I}$ (10 µM) is trickier and requires detailed consideration of the manner in which the experiments were performed. In an activation measurement, $\mathcal{C}$ was switched from <10 nM to 2 µM. $\mathcal{C}$ was not returned to <10 nM until ~2 s after channel activity was detected to ensure that the channel was really activated. Similarly, in a deactivation measurement, ~2 s was allowed after the last channel opening before $\mathcal{C}$ was switched to 2 µM again, to confirm that the channel was really deactivated. We assume that for all Ca^{2+} activation measurements, the channel only occupies the C^I_{04} state before the $\mathcal{C}$ jumps because it does not have time to transit from C^I_{04} to C^L_{04} before the solution switches; i.e., the mean transition time between states C^I_{04} and C^L_{04} is longer than the wait time for switching $\mathcal{C}$. There were instances in the Ca^{2+} activation experiments when the channel failed to go active.

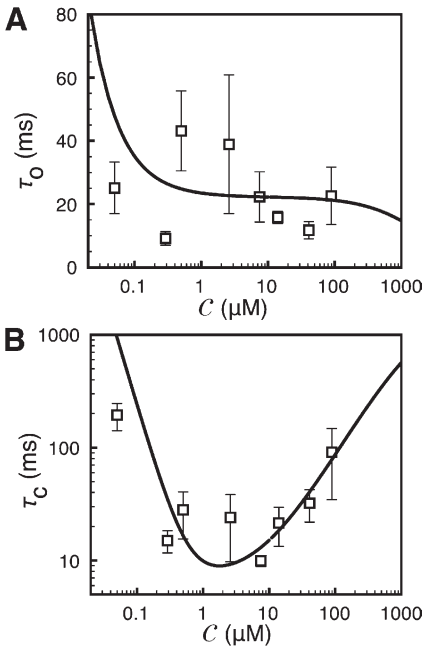

Figure 6. Ca^{2+} dependence of IP_3R channel mean open and closed times. Mean open (A) and closed times (B) for fixed $\mathcal{I}$ = 10 µM. The solid curves are theoretical fits, and squares are experimental values. The experimental data are taken from Ionescu et al. (2006) (supplemental text, section 1.10.1).

These failures to activate might have been caused by the IP_3R residing in state C^L_{04} at the time of the fluid switch. After the solution switch, the rates in Q_{XX} and Q_{XY} were set in accord with $\mathcal{C}$ = 2 µM and $\mathcal{I}$ = 10 µM (Fig. 5 C).

The Ca^{2+} deactivation latency histogram has two distinct peaks (Fig. 5 D). The CM fails to capture this feature even though it has two distinct pathways of deactivation: one from O^I_{24} and another from O^H_{24}. To attempt to improve the fit by relaxing the restriction of considering the rate from $C^I_{04} \rightarrow C^I_{24}$ = rate from $C^I_{04} \rightarrow C^H_{24}$ (see Maximum likelihood and AIC calculations in Materials and methods, and supplemental text, section 1.6) does not provide enough improvement in log likelihood for the deactivation latency fit to justify the extra parameters. It is possible that the two peaks corresponding to the two deactivation pathways in the CM cannot be resolved because the model peaks are broadened by having too many short latencies as a result of the exclusion of transition states used in the CM.

For Ca^{2+} and IP_3 activation, the channel initially can only occupy the C^L_{00} state with both $\mathcal{C}$ and $\mathcal{I}$ = 0. The CM generates Ca^{2+} and IP_3 activation and deactivation latency distributions that capture the general shape of the experimental distributions (Fig. 5, E and F).

For Ca^{2+} inhibition, $\pi_{(0)}$ is set to the normalized occupancies of states O^I_{14}, O^I_{24}, O^H_{24}, C^I_{24}, and C^H_{24} at $\mathcal{I}$ = 10 µM and $\mathcal{C}$ = 2 µM. For recovery from inhibition, only two inhibited states (C^L_{32} and C^L_{34}) are initially occupied at their equilibrium occupancy at $\mathcal{I}$ = 10 µM and $\mathcal{C}$ = 300 µM. Our fit to the Ca^{2+} inhibition latencies is unimpressive (Fig. 5 G). The observed inhibition latency distribution has at least two peaks, and the peaks are substantially sharper than the model distribution. The CM has only a single inhibition pathway that is characterized by a single flux parameter. Increasing the flux parameter to obtain a better fit for the inhibition latencies will shift the peak of the inhibition recovery latencies (Fig. 5 H) from the CM by the same amount in the same direction.

At this point, the inhibition pathways are opaque to us. Examination of experimental records reveals that after the majority (>80%) of $\mathcal{C}$ switches from 2 to 300 µM, the channel quickly (~30 ms) transits into a mode with low but nonzero P_o, followed by a transition into a different totally quiescent mode with P_o = 0. The inhibition latencies were evaluated as the durations the channel takes to enter the totally quiescent mode (Mak et al., 2007). Our present model does not distinguish the mode with low but nonzero P_o from the totally quiescent mode and therefore is not equipped to properly describe this gating behavior of the channel. Incorporating this feature in the model will require future experiments to gather enough data for modeling the kinetics of these two modes. We will then be able to devise a model distinguishing the two behaviors (with at least four gating modes) to more properly account for Ca^{2+} inhibition kinetics of the channel.

Channel gating kinetics from the model

The mean open- and closed-channel durations (τ_o and τ_c) were calculated for various $\mathcal{C}$ in $\mathcal{I}$ = 10 μM (supplemental text, section 1.8). The values from the model agree reasonably well with experimental measurements over a broad range of $\mathcal{C}$ (300 nM to 100 μM) (Fig. 6, A and B). The parameters were also used to generate simulated single-channel gating record exhibiting modal gating behavior that resembles experimental single IP_3R channel current record in 1 μM Ca^{2+} and 10 μM IP_3 (Fig. 3 B). The open and closed dwell-time distributions from the model are computed as in the supplemental text, section 1.9, and are shown by red lines in Fig. 7. The distributions obtained from the observed time-series data are shown by the green bars in Fig. 7.

The supplemental text, section 1.9, also describes the derivation of the open and closed dwell-time distributions in H and I modes from the model. The dwell-time distributions from the model for $\mathcal{C}$ = 1 μM in the I and H modes are given by the red lines in Fig. 8. The experimental data are presented by the green bars for comparison. The peaks near 1 ms in the observed closed-time distributions indicate that there are short-lived closed states available to both I and H modes (Fig. 8, B and D, respectively) that the model misses. There also seems to be a short-lived open state available to the I model (Fig. 8 A). Removing this discrepancy will require adding more states to the model. At this stage, we are not interested in complicating the model further. Nevertheless, our analysis about open and closed dwell-time distributions provides useful insight into the modal behavior of the channel. The discrepancy between the model and the observed dwell-time distributions in the I mode can also be removed by increasing the flux parameter involved in $C^I_{04} \rightleftarrows O^I_{14}$ transition (j^{II}_{0414}). However, increasing j^{II}_{0414} worsens the overall fit to the latency plus time-series data and therefore is not warranted.

Comparison between the CM and the DYKM

The one model of note that gave a better AIC score for P_o data only is the DYKM. Here, we compare the maximum likelihood fit of the CM (Fig. 1) to the DYKM using time-series and latency data. We find that the CM has a higher likelihood of producing the observations than does the DYKM. The DYKM has only a single open state and cannot produce modal gating, so we do not try to extract modal statistics from it.

The DYKM is a 120-state model of the IP_3R. The DYKM treats the channel as consisting of three rather than four identical subunits. It assumes that the subunits are independent and that each subunit has three binding sites, one for IP_3, a Ca^{2+} activation, and Ca^{2+} inhibition. The subunit states can be labeled by three binary digits (000, 001, . . . 111), with the first, second, and third bits representing the occupancy of the IP_3, Ca^{2+} activation, and Ca^{2+} inhibition binding sites, respectively. The channel is open if all three subunits are in the (110) state and closed otherwise. The channel open probability can be written $P_{o,dyk} = p_A^3$, where $p_A = K_{110}\mathcal{IC} / Z_{dyk}$ with

$$Z_{dyk} = 1 + (K_{010} + K_{001})\mathcal{C} + K_{011}\mathcal{C}^2 + K_{100}\mathcal{I} + K_{110}\mathcal{IC} + K_{101}\mathcal{IC} + K_{111}\mathcal{IC}^2,$$

where the K_{mno} are the occupancy parameters for the subunit states. The full-channel model has 120 states of which 119 are closed. The DYKM requires 7 occupancy and 12 flux parameters, or, equivalently, 12 independent reaction rates.

The DYKM had an AIC score ~10 better than the CM for the Sf9 P_o data. However, in the case of *Xenopus* P_o data, the DYKM had an AIC score of −92 as compared with −138 (better) for the CM. Furthermore, we show that P_o alone is inadequate to discriminate between models. To shed further light on the differences between the CM and the DYKM, we performed maximum

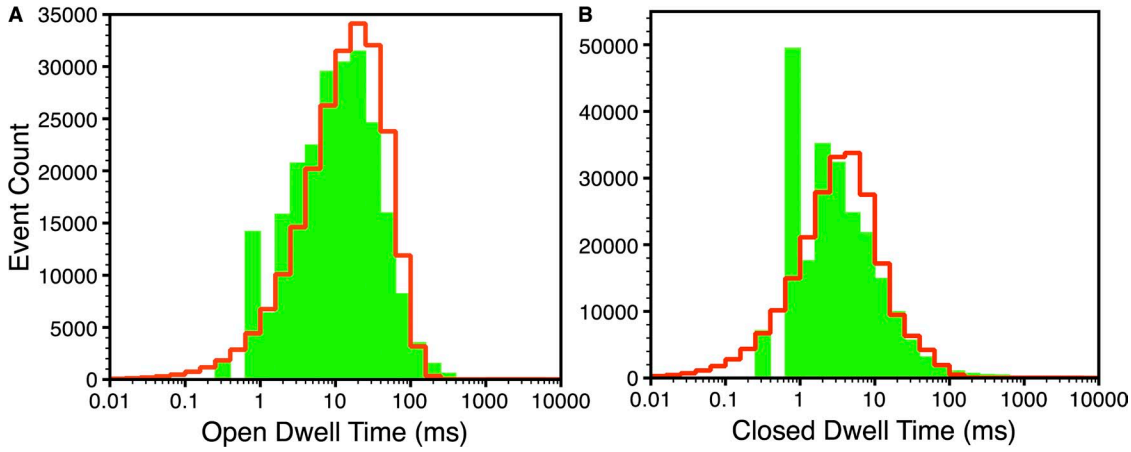

Figure 7. Open (A) and closed (B) dwell-time distributions at $\mathcal{C}$ = 1 μM and $\mathcal{I}$ = 10 μM. Green bars are the experimental values, and red lines are the model fits.

likelihood fits on the latency data and current traces of channel gating using DYKM as described in Maximum likelihood and AIC calculations in Materials and methods. We varied the 12 free rate constants of the DYKM while holding the occupancy parameters fixed. The likelihood ratio for CM and DYKM, which is interpreted as the relative probabilities of the two models, is the exponential of the difference between their AIC scores: $exp^{\delta AIC/2}$, where

$$\delta AIC \equiv AIC_{dYKM} - AIC_{CM} . \quad (7)$$

For the DYKM, we took advantage of the fact that there is only a single open state, thus Eq. 3 becomes:

$$log(f(tc_1, to_1, tc_2, to_2, \ldots tc_n, to_n)) = Q_{OO}\sum_i to_i + \sum_i log(Q_{OC}\exp(Q_{CC}tc_i)Q_{CO}) - log(Q_{OC}u_C). \quad (8)$$

In the case of DYKM, $N_C = 119$ and $N_O = 1$. Thus, w_O is a scalar from which it follows that $\pi_C = Q_{OC} / Q_{OC}u_C$. Thus, Q_{CC} is a 119 × 119 matrix, Q_{OO} is scalar (equal to minus the inverse lifetime of the single open state), and Q_{CO} and Q_{OC} are 119 × 1 and 1 × 119 matrices, respectively.

In both cases, fit to the latency data alone and fit to both the latency and time-series data (see Maximum likelihood and AIC calculations in Materials and methods), we calculated the likelihood score of the time-series data. In the first case, in which the time-series data were not used in the fit, the time-series data can be considered as "out-of-sample" data, with the differences in scores indicating which model does the best job of "predicting" the unseen data.

The results are summarized in Table 1, which demonstrate that there is strong statistical evidence for the CM. Adding more time-series data (up to 28 experiments for $\mathcal{C}$ = 0.1, 1, and 89 µM each) further increases the probability for the CM as compared with DYKM.

The occupancy parameters for the DYKM are given in Table S4. The rate constants are given in Table S5. The fits to the latency distributions are shown in Fig. 9.

DISCUSSION

Deficiencies of the CM

In a famously titled section heading, George Box wrote "All models are wrong . . ." (Box, 1979). Ours is no exception. One of the major deficiencies of the CM is its lack of distinction between the gating mode in which the IP_3R channel gates with low (∼0.05) but nonzero P_o and the mode in which the channel is quiescent with P_o = 0 (Q mode). In the study of IP_3R modal gating

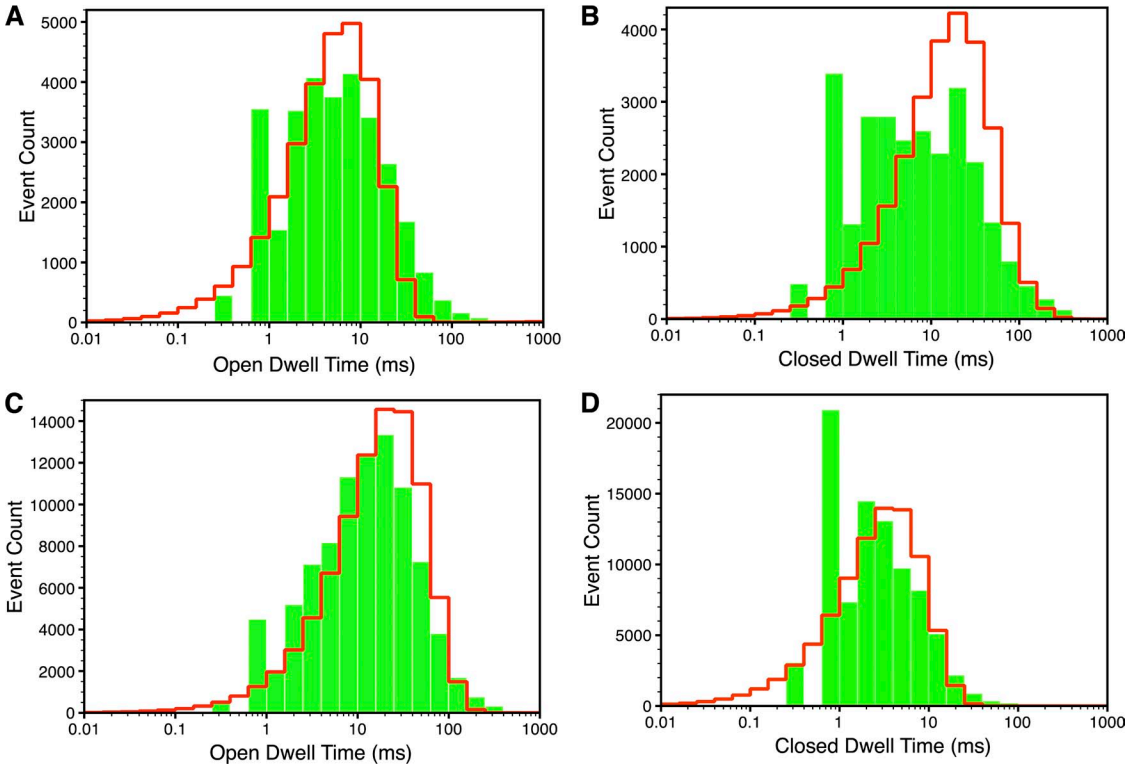

Figure 8. Dwell-time distributions in I and H modes at $\mathcal{C}$ = 1 µM and $\mathcal{I}$ = 10 µM. Open (A) and closed dwell-time distributions (B) in I mode. Open (C) and closed dwell-time distributions (D) in H mode. Green bars are the experimental values, and red lines are the model fits.

TABLE 1

Summary of the fits of the CM and the DYKM to data

Model	CM		DYKM		δAIC		$Exp^{\delta AIC/2}$	
Exp.	LAT	LAT+TS	LAT	LAT+TS	LAT	LAT+TS	LAT	LAT+TS
LAT	18213	18295	18632	21867	820	7126	10^{178}	10^{1547}
TS	33327	32570	69784	44610	72896	24062	10^{15029}	10^{5225}
Total	51540	50865	88416	66478	73734	31206	10^{16011}	10^{6776}

The first column denotes the experiment class (LAT, latency; TS, time series). The columns LAT and LAT+TS, respectively, indicate that the latency data only and latency plus time-series data were used in the fit. The values in the CM and DYKM columns contain -log(likelihood(data)) scores of the CM and the DYKM, respectively, for the data given in that row. The column δAIC contains the difference in the AIC scores for the two models, with positive values indicating that the CM had the lowest (best) AIC score. The last column gives $exp^{\delta AIC/2}$ where values >1 indicate that the CM is more probable than the DYKM for that data. The row Total contains the total -log(likelihood(data)) (for columns CM and DYKM), δAIC, and $exp^{\delta AIC/2}$ scores for the latency and time-series data.

behaviors under steady-state ligand conditions (Ionescu et al., 2007), it is difficult to tell if a channel remains closed for a long duration whether it stays in the L mode during that whole duration or has spontaneously entered into and exited from the Q mode during that time. However, reexamination of Ca^{2+} inhibition latency measurements reveals that the distinction between the L and Q mode is significant because the majority of such inhibition events proceed from channel in H mode, favored by the initial ligand conditions of ($\mathcal{C}$, $\mathcal{I}$) = (2 μM, 10 μM), through the L mode to the Q mode, instead of going directly from the H mode into the Q mode. Thus, the latencies consist of two contributions: the time for the channel to transit from the H mode to the L mode and the time for the channel to enter the Q mode from the L mode.

Furthermore, the Ca^{2+} inhibition latency distribution clearly shows two peaks (at ∼70 and 900 ms). A natural explanation for this derived from the CM is that the short latencies are caused by channel transiting directly from O_{24}^{H} to C_{34}^{L}, whereas the long latencies are a result of the channels initially in I mode that have to transit into the H mode, which takes a long time because of the long τ^{I}, before they can be inhibited. However, examination of data records indicates that channel initially in the I mode when $\mathcal{C}$ is changed enters the L mode very rapidly (∼3 ms). Thus, the long Ca^{2+} inhibition latencies are caused by the slow exit from the L mode into the Q mode. To properly model Ca^{2+} inhibition of IP_3R requires distinguishing the L mode from Q mode by including open O^{L} states for the channel in the L mode, which will increase the complexity of the model substantially. We believe that a more accurate description of the inhibition latencies is crucial for the understanding of how the channel functions and will be the subject of our future research.

The CM has a slightly higher number of short IP_3 activation latencies (<10 ms) compared with the latency data. This may be a consequence of the exclusion of low occupancy states during IP_3 binding (like the C_{i1}, C_{i2}, and C_{i3} states excluded from the $C_{i0} \leftrightharpoons C_{i4}$ connection) applied to simplify our model. However, if this is true, it suggests that the channel binds IP_3 at a rate of ∼0.4 ms^{-1}, substantially slower than the diffusion-limited rate of ∼20 ms^{-1} in $\mathcal{I}$ = 10 μM (assuming a reaction radius of 1 nm and diffusion coefficient of 0.5 μm^2 ms^{-1} for IP_3; Redner, 2001). This indicates that IP_3 probably needs to enter its binding site in IP_3R with a particular orientation, as suggested by the structure of the site in the receptor with residues from different parts of the site interacting with different phosphate groups in IP_3 through networks of hydrogen bonds (Bosanac et al., 2002). Alternately, this may be caused by slow conformation changes the channel has to undergo after each IP_3 binding that prolongs the activation latencies.

With activation and deactivation latencies considered only for $\mathcal{C}$ jumping between <10 nM and 2 μM and $\mathcal{I}$ between 0 and 10 μM, there are insufficient data about the dynamic behaviors of the IP_3R channel with one to three IP_3 or only one Ca^{2+} bound for the CM to include those states explicitly under the simplicity principle. The failure of the model to generate two resolved peaks for the Ca^{2+} deactivation latency distribution (Fig. 5 D) suggests that those processes probably involve multiple fast transitions through intermediate states, particularly for Ca^{2+} activation. Future latency measurements involving $\mathcal{I}$ jumping between 0 and subsaturating levels (100 nM, say) and/or $\mathcal{C}$ between <10 nM and suboptimal levels (300 nM, say) could provide more insights into these processes to improve the model. Finally, we cannot rule out the possibility of other simple models that might better describe our data.

Strengths of the CM

The rest of Box's section heading was ". . . some models are useful." The CM is the result of the synthesis of a large body of data gathered in several studies of many different aspects of ligand regulations of IP_3R channel gating in a novel, data-driven minimalist approach. The model uses five to six complex occupancy parameters to capture accurately not only the Ca^{2+} but also the IP_3 dependencies of the channel equilibrium P_o. The model does a fair job of accounting for the P_o, τ_o, τ_c, modal gating statistics, and the majority of latency distributions

examined. Moreover, our model makes concrete predictions regarding the single-channel response to rapid changes in the cytosolic ligand concentrations of IP_3 and Ca^{2+}, so its validity can be verified by future experiments. With jumps of ($\mathcal{C}$, $\mathcal{I}$) from (2 µM, 0) to (2 µM, 10 µM), the channel mostly enters the H mode from C^L_{20}. However, in lower $\mathcal{C}$, the initial occupancy of C^L_{00} will be higher so jump in $\mathcal{I}$ will activate the channel more frequently from $C^L_{00} \rightarrow C^I_{04}$, and then through a Ca^{2+}-dependent transition to either $C^I_{24} \rightarrow O^I_{24}$ or $C^H_{24} \rightarrow O^H_{24}$. The latency will increase with decreasing $\mathcal{C}$ because of the Ca^{2+}-dependent $C_{04} \rightarrow C_{24}$ step. In high $\mathcal{C}$, the probability of the channel being activated from C^L_{30} will increase with $\mathcal{C}$. Because the pathways out of C^L_{30} are both slow, the IP_3 activation latency should increase with increase in $\mathcal{C}$. Thus, our model predicts that the mean IP_3 activation latency has a minimum of ≈0.26 s at moderate $\mathcal{C}$ of ∼0.6 µM (Fig. S5).

The CM predicts that a higher number of longer Ca^{2+} activation latencies will be detected if the channel is exposed to $\mathcal{C} < 10$ nM longer than the 2 s used (Mak et al., 2007) before $\mathcal{C}$ is changed to 2 µM. This is because the relatively Ca^{2+}-independent τ^I (∼0.4 s) observed suggests that the channel may not be fully equilibrated between the C^I_{04} and C^L_{04} states during the time the channel remains exposed to $\mathcal{C} < 10$ nM. Longer exposure will increase the occupancy of C^L_{04}, from which the channel will take longer to exit, thus giving longer Ca^{2+} activation latencies.

As mentioned in Maximum likelihood and AIC calculations in Materials and methods, we minimize the number of parameters by assuming that the effective

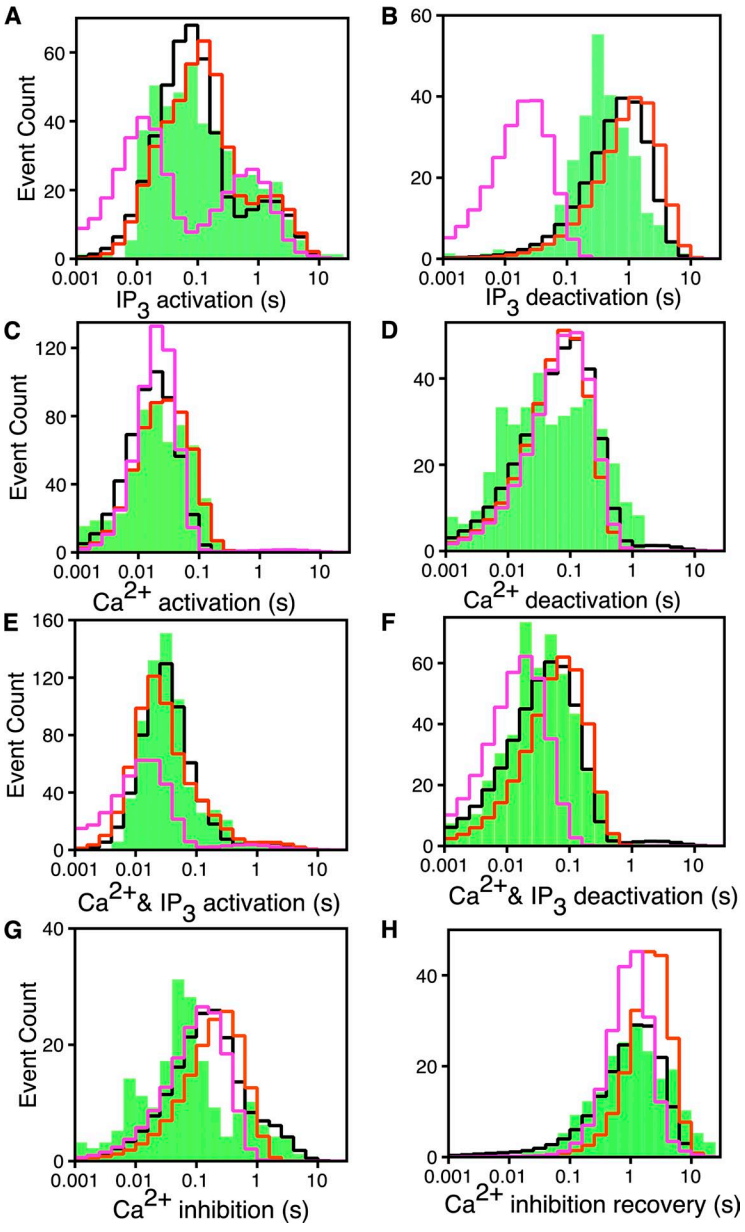

Figure 9. Fits from the DYKM. Logarithmically binned latency histograms of IP_3R channels responding to changes in $\mathcal{C}$ and $\mathcal{I}$. In each case, ($\mathcal{C}$, $\mathcal{I}$) are changed as indicated in Fig. 5. (A) IP_3 activation, (B) IP_3 deactivation, (C) Ca^{2+} activation, (D) Ca^{2+} deactivation, (E) Ca^{2+} and IP_3 activation, (F) Ca^{2+} and IP_3 deactivation, (G) Ca^{2+} inhibition, and (H) Ca^{2+} inhibition recovery. The experimental latency distributions given by the bars are taken from Mak et al. (2007). The red and pink lines, respectively, are from the DYKM when fit to the latency data alone and latency and times-series data simultaneously. The black line is from the CM developed in this paper when fit to the latency and times-series data simultaneously and is shown for comparison.

rates for the reactions from the states with no Ca^{2+} bound to the states with two Ca^{2+} bound are the same at $\mathcal{I} = 0$, 10 μM. Relaxing this constraint did not improve the fit to the data. This result indicates that the transition rate from the states with no Ca^{2+} bound to the states with two Ca^{2+} bound is independent of the number of IP_3 bound. Whether this assertion is correct or not can be investigated further by performing more ligand jump experiments, where $\mathcal{C}$ is jumped from 0 μM to suboptimal values, e.g., jumping ($\mathcal{C}$, $\mathcal{I}$) from (0 μM , 0 μM) to (0.5 μM , 0 μM) and (0 μM , 10 μM) to (0.5 μM , 10 μM).

The CM, especially its deficiencies, provides novel insights that are not immediately obvious from available data. The failure of the CM to adequately account for the Ca^{2+} inhibition latency distribution reveals the importance of drawing the distinction between the L mode with low but nonzero P_o and the quiescent Q mode with $P_o = 0$ for improving our understanding of Ca^{2+} regulation of the channel, which is not obvious from the modal analysis in Ionescu et al. (2007).

Although the development of the CM is directed by experimental observations, it also provides insights to guide the designs of future experiments and data analysis protocols. Our model suggests that experiments in which $\mathcal{I}$ is switched between 0 and subsaturating levels and $\mathcal{C}$ between physiological resting to suboptimal levels can provide valuable information to improve the characterization of the kinetics of the channel with one to three IP_3 bound, and channels with one Ca^{2+} bound. The model also reveals that the amount of time required for the channel to fully equilibrate in one combination of ($\mathcal{C}$, $\mathcal{I}$) before the ligand conditions can be changed in latency measurements, and that varying the time the channel stays in one ligand combination before perfusion solution switches can provide valuable information on kinetic rates. Furthermore, analysis of Ca^{2+} inhibition latencies should include measuring the duration between the time when $\mathcal{C}$ is changed and the time when the channel enters the L mode, and the duration when the channel stays in the L mode until its activity terminates as it enters the Q mode. We suspect that there are significant surprises in store regarding the mechanistic interpretation of inhibition.

Although the aim here was to model the IP_3R Ca^{2+} channel, our methods can be easily applied to other channels and single-molecule dynamics.

We would like to acknowledge the resources and support provided by J. Kevin Foskett and many useful conversations with Bill Bruno, Andy Fraser, Nick Hengartner, Ben McMahon, Paul Fenimore, Hans Frauenfelder, Joel Berenzden, Peter Hraber, Ian Parker, Jian-Wei Shuai, Silvina Ponce Dawson, Tony Auerbach, David Colquhoun, Fred Sachs, and Horia Vais.

This work was supported by National Institutes of Health (grant 5RO1GM065830-08).

Sharona E. Gordon served as editor.

Submitted: 13 December 2011
Accepted: 12 July 2012

REFERENCES

Beece, D., L. Eisenstein, H. Frauenfelder, D. Good, M.C. Marden, L. Reinisch, A.H. Reynolds, L.B. Sorensen, and K.T. Yue. 1980. Solvent viscosity and protein dynamics. *Biochemistry.* 19:5147–5157. http://dx.doi.org/10.1021/bi00564a001

Berridge, M.J., M.D. Bootman, and P. Lipp. 1998. Calcium—a life and death signal. *Nature.* 395:645–648. http://dx.doi.org/10.1038/27094

Bosanac, I., J.R. Alattia, T.K. Mal, J. Chan, S. Talarico, F.K. Tong, K.I. Tong, F. Yoshikawa, T. Furuichi, M. Iwai, et al. 2002. Structure of the inositol 1,4,5-trisphosphate receptor binding core in complex with its ligand. *Nature.* 420:696–700. http://dx.doi.org/10.1038/nature01268

Box, G.E.P. 1979. Robustness in the Strategy of Scientific Model Building. Academic Press, New York. 312 pp.

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

Cheney, E., and D. Kincaid. 2007. Numerical Mathematics and Computing. Sixth edition. Brooks/Cole Publishing Company, Pacific Grove, CA. 784 pp.

Colquhoun, D., and A.G. Hawkes. 1981. On the stochastic properties of single ion channels. *Proc. R. Soc. Lond. B Biol. Sci.* 211:205–235. http://dx.doi.org/10.1098/rspb.1981.0003

Dawson, A.P., E.J. Lea, and R.F. Irvine. 2003. Kinetic model of the inositol trisphosphate receptor that shows both steady-state and quantal patterns of Ca^{2+} release from intracellular stores. *Biochem. J.* 370:621–629. http://dx.doi.org/10.1042/BJ20021289

De Young, G.W., and J. Keizer. 1992. A single-pool inositol 1,4,5-trisphosphate-receptor-based model for agonist-stimulated oscillations in Ca^{2+} concentration. *Proc. Natl. Acad. Sci. USA.* 89:9895–9899. http://dx.doi.org/10.1073/pnas.89.20.9895

Eigen, M. 1968. New looks and outlooks on physical enzymology. *Q. Rev. Biophys.* 1:3–33. http://dx.doi.org/10.1017/S0033583500000445

Foskett, J.K., C. White, K.H. Cheung, and D.O.D. Mak. 2007. Inositol trisphosphate receptor Ca^{2+} release channels. *Physiol. Rev.* 87:593–658. http://dx.doi.org/10.1152/physrev.00035.2006

Fraiman, D., and S.P. Dawson. 2004. A model of IP_3 receptor with a luminal calcium binding site: stochastic simulations and analysis. *Cell Calcium.* 35:403–413. http://dx.doi.org/10.1016/j.ceca.2003.10.004

Fredkin, D., and J.A. Rice. 1986. On aggregated markov processes. *J. Appl. Probab.* 23:208–214. http://dx.doi.org/10.2307/3214130

Gin, E., M. Falcke, L.E. Wagner II, D.I. Yule, and J. Sneyd. 2009. A kinetic model of the inositol trisphosphate receptor based on single-channel data. *Biophys. J.* 96:4053–4062. http://dx.doi.org/10.1016/j.bpj.2008.12.3964

Hirose, K., S. Kadowaki, and M. Iino. 1998. Allosteric regulation by cytoplasmic Ca^{2+} and IP_3 of the gating of IP_3 receptors in permeabilized guinea-pig vascular smooth muscle cells. *J. Physiol.* 506:407–414. http://dx.doi.org/10.1111/j.1469-7793.1998.407bw.x

Ionescu, L., K.H. Cheung, H. Vais, D.O.D. Mak, C. White, and J.K. Foskett. 2006. Graded recruitment and inactivation of single $InsP_3$ receptor Ca^{2+}-release channels: implications for quantal Ca^{2+} release. *J. Physiol.* 573:645–662. http://dx.doi.org/10.1113/jphysiol.2006.109504

Ionescu, L., C. White, K.H. Cheung, J. Shuai, I. Parker, J.E. Pearson, J.K. Foskett, and D.O.D. Mak. 2007. Mode switching is the major mechanism of ligand regulation of $InsP_3$ receptor calcium release channels. *J. Gen. Physiol.* 130:631–645. http://dx.doi.org/10.1085/jgp.200709859

Mak, D.O.D., S.M. McBride, and J.K. Foskett. 2003. Spontaneous channel activity of the inositol 1,4,5-trisphosphate ($InsP_3$) receptor ($InsP_3R$). Application of allosteric modeling to calcium and

$InsP_3$ regulation of $InsP_3R$ single-channel gating. *J. Gen. Physiol.* 122:583–603. http://dx.doi.org/10.1085/jgp.200308809

Mak, D.O.D., J.E. Pearson, K.P.C. Loong, S. Datta, M. Fernández-Mongil, and J.K. Foskett. 2007. Rapid ligand-regulated gating kinetics of single inositol 1,4,5-trisphosphate receptor Ca^{2+} release channels. *EMBO Rep.* 8:1044–1051. http://dx.doi.org/10.1038/sj.embor.7401087

Marchant, J.S., and C.W. Taylor. 1997. Cooperative activation of IP_3 receptors by sequential binding of IP_3 and Ca^{2+} safeguards against spontaneous activity. *Curr. Biol.* 7:510–518. http://dx.doi.org/10.1016/S0960-9822(06)00222-3

Mcquarrie, A.D.R. 1998. Regression and Time Series Model Selection. World Scientific Publishing Company, Singapore. 455 pp.

Qin, F., A. Auerbach, and F. Sachs. 1997. Maximum likelihood estimation of aggregated Markov processes. *Proc. Biol. Sci.* 264:375–383. http://dx.doi.org/10.1098/rspb.1997.0054

Redner, S. 2001. A Guide to First-Passage Processes. Cambridge University Press, Cambridge. 312 pp.

Shuai, J., J.E. Pearson, J.K. Foskett, D.O.D. Mak, and I. Parker. 2007. A kinetic model of single and clustered IP_3 receptors in the absence of Ca^{2+} feedback. *Biophys. J.* 93:1151–1162. http://dx.doi.org/10.1529/biophysj.107.108795

Sneyd, J., and J.F. Dufour. 2002. A dynamic model of the type-2 inositol trisphosphate receptor. *Proc. Natl. Acad. Sci. USA.* 99:2398–2403. http://dx.doi.org/10.1073/pnas.032281999

Swaminathan, D., G. Ullah, and P. Jung. 2009. A simple sequential-binding model for calcium puffs. *Chaos.* 19:037109. http://dx.doi.org/10.1063/1.3152227

Swillens, S., L. Combettes, and P. Champeil. 1994. Transient inositol 1,4,5-trisphosphate-induced Ca^{2+} release: a model based on regulatory Ca^{2+}-binding sites along the permeation pathway. *Proc. Natl. Acad. Sci. USA.* 91:10074–10078. http://dx.doi.org/10.1073/pnas.91.21.10074

Tang, Y., J.L. Stephenson, and H.G. Othmer. 1996. Simplification and analysis of models of calcium dynamics based on IP_3-sensitive calcium channel kinetics. *Biophys. J.* 70:246–263. http://dx.doi.org/10.1016/S0006-3495(96)79567-X

Ullah, G., W.J. Bruno, and J.E. Peason. 2012. Simplification of reversible Markov chains by removal of states with low equilibrium occupancy. *J. Thor. Biol.* http://dx.doi.org/10.1016/j.jtbi.2012.07.007

Yang, J., W.J. Bruno, W.S. Hlavacek, and J.E. Pearson. 2006. On imposing detailed balance in complex reaction mechanisms. *Biophys. J.* 91:1136–1141. http://dx.doi.org/10.1529/biophysj.105.071852