Statistical Evaluation of Ion-Channel Gating Models Based on Distributions of Log-Likelihood Ratios

László Csanády
Department of Medical Biochemistry, Semmelweis University, and Neurochemical Group of the Hungarian Academy of Sciences, Budapest, Hungary

ABSTRACT The distributions of log-likelihood ratios (ΔLL) obtained from fitting ion-channel dwell-time distributions with nested pairs of gating models (Ξ, full model; Ξ_R, submodel) were studied both theoretically and using simulated data. When Ξ is true, ΔLL is asymptotically normally distributed with predictable mean and variance that increase linearly with data length (n). When Ξ_R is true and corresponds to a distinct point in full parameter space, ΔLL is Γ-distributed ($2\Delta LL$ is χ-square). However, when data generated by an l-component multiexponential distribution are fitted by $l+1$ components, Ξ_R corresponds to an infinite set of points in parameter space. The distribution of ΔLL is a mixture of two components, one identically zero, the other approximated by a Γ-distribution. This empirical distribution of ΔLL, assuming Ξ_R, allows construction of a valid log-likelihood ratio test. The log-likelihood ratio test, the Akaike information criterion, and the Schwarz criterion all produce asymmetrical Type I and II errors and inefficiently recognize Ξ, when true, from short datasets. A new decision strategy, which considers both the parameter estimates and ΔLL, yields more symmetrical errors and a larger discrimination power for small n. These observations are explained by the distributions of ΔLL when Ξ or Ξ_R is true.

INTRODUCTION

Conformational transitions of ion-channel proteins associated with gating are best described as a stochastic process governed by some discrete Markov model (1). Such models, called gating schemes, consist of sets of discrete open-channel and closed-channel conformations (states) connected to each other in a given pattern, and of rate constants which describe probabilities of transition along the allowed kinetic pathways. Many features of ion-channel gating, such as the durations of individual open- and closed-channel dwell times or the vector of the entire ordered sequence of open- and closed-time durations, are random variables whose distributions are defined by the underlying Markov model (2). These random variables can be experimentally sampled, as the timing of individual open-to-closed and closed-to-open transitions are observable in single-channel patch-clamp records (3). A steady-state current record of a single ion channel is first idealized to reconstruct the sequence of open and closed dwell times (the events list (2)) which is then fitted to an assumed gating scheme by maximum likelihood (ML) to obtain the set of parameters (rate constants, or time constants and fractional amplitudes of exponentials) which best describes the data—i.e., the set of dwell times (4), the set of pairs of consecutive open/closed dwell times (5), or the entire ordered series of dwell times (6–11).

Fitting pairs, or the entire series, of events reveals possible correlations between the durations of adjacent open and closed dwell times. Much valuable information has been obtained from such studies, e.g., for large-conductance Ca^{2+}-activated K^+ channels (5,12,13), several members of the pentameric ligand-gated ion channel family such as the glycine (14,15), the $GABA_A$ (16), and the nicotinic acetylcholine (17–19) receptor channel, and recently for the NMDA receptor (20). Nonetheless, traditional fitting of one-dimensional dwell times is still used to study channels for which such correlations are not apparent (21–25) or for which a gating scheme is not yet at hand (26,27); but also for a visual comparison of such fits with model predictions—even if the analysis itself relies on more sophisticated methods (12–16,18–20).

Typically, the true underlying gating scheme is unknown, and the aim of the experimenter is to find the minimal model that explains the experimental findings. Several possible models are therefore considered, and these are then ranked based on the likelihood, or its logarithm (Log-Likelihood, LL), of observing the data, given the model. When two models are compared that contain the same numbers of parameters, the one yielding a higher LL is accepted. Comparison of two models with different numbers of free parameters is more problematic, as more free parameters may result in slightly higher likelihood values even if the introduction of the additional parameters is not justified. A good strategy for model evaluation in this case depends on knowledge of the distribution of the log-likelihood ratio $\Delta LL = LL_2 - LL_1$. The present work addresses this question for the subset of cases in which the model with fewer parameters is a submodel of the other (nested models). The results are presented in the

Submitted September 28, 2005, and accepted for publication January 12, 2006.

Address reprint requests to László Csanády, Dept. of Medical Biochemistry, Semmelweis University, 1444 Budapest, Pf. 262, Hungary. Tel.: 36-1-266-2755, Ext. 4023; E-mail: csanady@puskin.sote.hu.

Abbreviations used: LL, log likelihood; ΔLL, log-likelihood ratio; $E(X)$, expectation value of the random variable X; $D^2(X)$, variance of X; $D(X)$, standard deviation of X; Ω, parameter space; Ξ, Θ, two different models, defined in terms of both scheme and parameter vector; $\boldsymbol{\xi}$, $\boldsymbol{\vartheta}$, parameter vectors for models Ξ and Θ.

0006-3495/06/05/3523/23 $2.00

doi: 10.1529/biophysj.105.075135

framework of ML-fitting of one-dimensional dwell-time distributions; the implications for ML-fitting of joint distributions of series of dwell times are discussed. The core of the article consists of two basic parts—the first, more theoretically and the second, more practically oriented.

The section ''Distributions of Log-Likelihood Ratios'' investigates the distributions of ΔLL under various situations in which either the submodel or the broader model is the true model. For the case in which the broader model is true, the asymptotic distribution of ΔLL is mathematically derived and confirmed using extensive simulations. The case in which the submodel is true has been studied in the past (28) and is discussed in standard statistics textbooks (e.g., (29–31)); $2\Delta LL$ is χ^2-distributed if certain regularity criteria apply. This assumption has been widely used for evaluation of ion-channel gating schemes (e.g., (2,32)), but the applicability of the required regularity criteria to such models has not been tested. These criteria are examined here and shown not to apply in some of the most common situations. The empirical distributions of ΔLL for such cases are obtained from large sets of simulated data.

The section ''Strategies for Model Discrimination'' evaluates several existing methods for model identification, including the log-likelihood ratio test (LLR test, e.g., (32)), the Akaike information criterion (AIC, (33)) and the Schwarz criterion (SC, (34)). Experimenters often put their faith in these statistical tests without first studying the applicability of the underlying theorems. E.g., the assumptions required for use of the canonical LLR test frequently do not apply (a valid LLR test for such cases is presented here). Moreover, even studies in which these methods were tested on simulated data (6,7,9) have examined their efficiencies only under limited sets of conditions. First, the ability to detect an extra component in a distribution is very dependent on the size of that component. It is inevitable in simulation studies that the results depend on the chosen values of the rate constants. Second, the value of ΔLL is a random variable. Therefore, obtaining a correct decision for a single simulated events list may provide little information. For a given model, a large set of independent events lists should be examined to obtain a population of decisions that will allow estimation of the probabilities of error. Third, the effect of the length of the data on the reliability of the decision lacks systematic evaluation. E.g., for each of the above statistical tests the Type I error (rejection of the null hypothesis although it is true) and Type II error (acceptance of the null hypothesis although it is wrong) depend differentially on the length of the data.

These issues are explored here for nested pairs of models using large sets of simulated data, and the findings explained in light of the distributions of ΔLL for both cases, i.e., when the submodel is correct (the null hypothesis) and when it is incorrect. In addition, a new decision strategy is proposed, which is not a priori biased toward either of the two models and produces relatively symmetrical Type I and Type II errors. Because it also results in a higher power of discrimination for relatively small data sets, the novel approach might prove a useful alternative of the LLR test, and of the AIC or SC, for such applications.

MATERIALS AND METHODS

Simulations

Simulated sequences of dwell times (events lists) were generated as described (e.g., (35)), using exponentially distributed random numbers obtained by a transformation on an evenly distributed random variable (36). To minimize simulation-related artifacts, evenly distributed random numbers were generated using the high-quality random generator ran2 (36). This function combines a long-period random sequence algorithm (37) with a shuffling procedure (38) which further removes serial correlations. The period of ran2 is $\sim 2 \times 10^{18}$, which allows generation of a random sequence of $\sim 10^{18}$ channel gating events, as each such event consumes two random numbers. For a given gating scheme, 1000 independent events lists of length $2n$ (containing n closed events) were produced by simulating one long events list of length $2000n$, which was then broken up into 1000 parts of length $2n$ each. For the longest individual events lists studied, n was 10^4; thus, simulation of 1000 such events lists consumed 4×10^7 random numbers, far smaller than the period of the generator.

Maximum likelihood fitting

Closed dwell times of simulated events lists were fitted to multiexponential probability density functions by unbinned ML (e.g., (4)). The LL was maximized with respect to the parameters of the fitted distribution using two optimizers, an uphill simplex algorithm (39) and the Davidon-Fletcher-Powell variable metric method (36), started from several different seed parameters. This was especially important when the fitted distribution contained more parameters than the true generating distribution. E.g., in the case in which single-exponentially distributed dwell times are fitted by a double-exponential distribution, the best unconstrained ML estimate $\hat{\vartheta}$ is expected to scatter in a broad region of the parameter space Ω (along a jointed line such as that depicted in Fig. 7). Thus, in such cases, each events list was fitted using both optimizer algorithms, both seeded from 40 different seed values along the expected jointed line. Out of the 80 parameters to which these attempts converged, $\hat{\vartheta}$ was chosen as the one that yielded the highest LL. However, it must be noted that despite all the effort to identify the best $\hat{\vartheta}$, it remains possible that for some of the events lists the true $\hat{\vartheta}$ was not found. Single-exponential distributions were fitted by solving the likelihood equation; i.e., the ML estimate of the time constant was obtained as the arithmetic average of the dwell times.

Evaluation of $\hat{h}(x)$

The parameters $\hat{m}$ and $\hat{\sigma}^2$ were calculated by numerically evaluating the integrals in Eqs. 16a and 16b using the simple trapezoidal rule and exponentially increasing interval widths. Limited computer dynamic range was corrected for using the algorithm described in Appendix B.

RESULTS AND DISCUSSION

Distributions of log-likelihood ratios

Ion channel gating models consist of a scheme, i.e., a number of closed and open states with a given connectivity among them, and a set of numeric parameters, i.e., the values of the transition rates (Fig. 1). Because of the underlying Markovian process the distributions of open and closed dwell times are mixtures of exponentials. As a usual first step in identifying a model, one is simply concerned about the probability

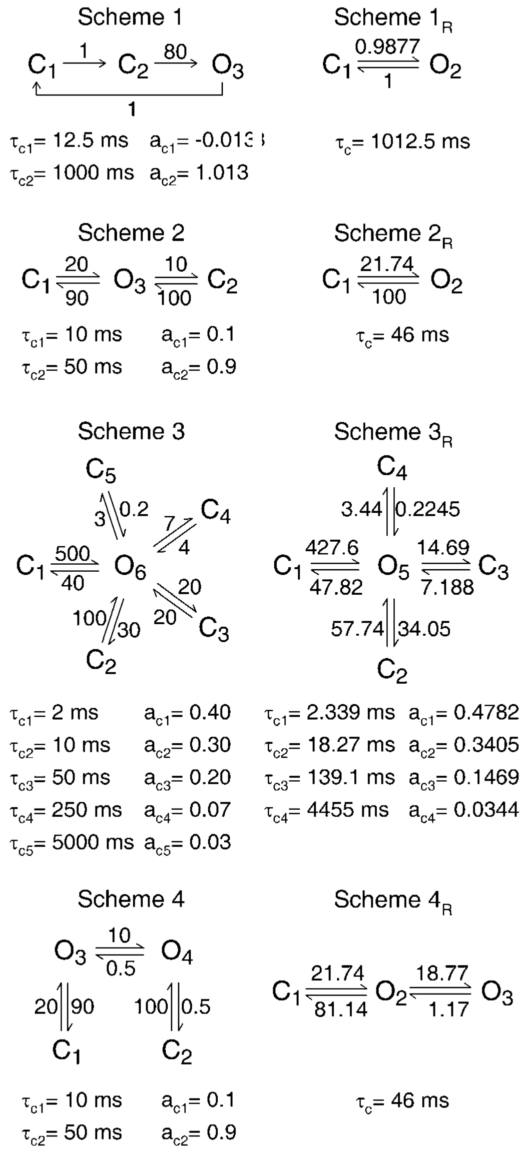

FIGURE 1 Gating schemes used for simulations. Rate constants are in s^{-1}. Time constants (τ_{cj}, in milliseconds) and fractional amplitudes (a_{cj}) describe the closed-time distributions of the respective schemes. The schemes are organized in pairs. Both schemes in each row have identical open-time distributions; the closed-time distribution of each scheme in the right column (Scheme Θ_R, for *reduced*) is a special case, under some constraint on the parameters, of the closed-time distribution of the scheme to its left (Scheme Θ). All reduced schemes are printed with parameter values corresponding to $\boldsymbol{\vartheta}_R^*$ (see text).

density function (pdf) of either closed or open dwell times. In this context the model Ξ is described by the shape, i.e., the number of exponential components, of the pdf (which reflects the number of relevant single-channel states), and the parameter vector x, which consists of the set of time constants $\tau_1, \tau_2, \ldots \tau_k$ and fractional amplitudes $a_1, a_2, \ldots a_{k-1}$. Usually, a particular number of exponential components is first assumed after visual inspection of the dwell-time histogram of interest, and then the parameter vector $\boldsymbol{\xi}$ found that maximizes the LL function, defined as follows. If a set of n dwell times $t_1, t_2, \ldots t_n$ is observed and the data are fitted to a model with pdf f, then $LL = \sum_{i=1}^{n} \ln f(t_i)$. Note that the dwell times are measured in some unit of time (e.g., seconds or milliseconds); t_i denotes here the dimensionless form of the dwell time, i.e., $t_i = (i^{th}$ dwell-time)/(unit of time). Therefore, the pdf f of the variable t_i is also dimensionless. In a typical situation the data are generated from a true (but unknown) model Θ described by parameter vector $\boldsymbol{\vartheta}^*$ (* will be used throughout this article to denote the true parameter vector which was used to generate the data); i.e., t_i values are drawn from a distribution with pdf $f_{\boldsymbol{\vartheta}^*}$. These data are then tentatively fitted to a model Ξ represented by the pdf $f_{\boldsymbol{\xi}}$. This LL will be denoted $LL(\boldsymbol{\xi}|\boldsymbol{\vartheta}^*)$ to emphasize that Θ (with parameter vector $\boldsymbol{\vartheta}^*$) is the true model and Ξ (with parameter vector $\boldsymbol{\xi}$) is the model assumed for fitting. $LL(\boldsymbol{\xi}|\boldsymbol{\vartheta}^*)$ is itself a random variable, and the aim of the present work is to examine its distribution under various conditions relevant in practice. Without constraining generality, the focus will be on the closed-time distributions of the schemes shown in Fig. 1 (the parameters of the closed-time distribution are printed below each scheme).

Distribution of ΔLL values obtained from describing the data by two alternative arbitrary schemes with fixed sets of parameters

This section lays the groundwork for addressing the full complexity of problems encountered in practice, by first solving a more simple situation. Thus, as a first step, assume that the data (generated from a true model Θ with vector $\boldsymbol{\vartheta}^*$) are tentatively described by two arbitrary schemes Ξ_1 and Ξ_2 such that not only the shapes (numbers of components) of the pdfs are assumed, but also the parameter vectors $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$ are fixed to any arbitrary values. The observed dwell times t_i are distributed according to $f_{\boldsymbol{\vartheta}^*}$ and, if Θ contains only one gateway state (see all schemes in Fig. 1, except for Scheme 4), t_i are also independent of each other. Thus, the random variables $X_i^{(1)} = \ln f_{\boldsymbol{\xi}_1}(t_i)$, $i = 1, 2, \ldots n$, are also identically distributed and independent, and the same is true for the two series $X_i^{(2)} = \ln f_{\boldsymbol{\xi}_2}(t_i)$ and $Y_i = \ln\left(f_{\boldsymbol{\xi}_2}(t_i)/f_{\boldsymbol{\xi}_1}(t_i)\right)$. As $LL(\boldsymbol{\xi}_1|\boldsymbol{\vartheta}^*) = \sum_{i=1}^{n} X_i^{(1)}$, $LL(\boldsymbol{\xi}_2|\boldsymbol{\vartheta}^*) = \sum_{i=1}^{n} X_i^{(2)}$, and $\Delta LL(\boldsymbol{\xi}_1, \boldsymbol{\xi}_2|\boldsymbol{\vartheta}^*) = LL(\boldsymbol{\xi}_2|\boldsymbol{\vartheta}^*) - LL(\boldsymbol{\xi}_1|\boldsymbol{\vartheta}^*) = \sum_{i=1}^{n} Y_i$, by the central limit theorem all three are asymptotically normally distributed, and the expectations and variances can be calculated as $nE(X_1^{(1)})$, $nE(X_1^{(2)})$, $nE(Y_1)$, and $nD^2(X_1^{(1)})$, $nD^2(X_1^{(2)})$, and $nD^2(Y_1)$, respectively. Thus, for any model Ξ and any arbitrary fixed vector $\boldsymbol{\xi}$,

$$E(LL(\boldsymbol{\xi}|\boldsymbol{\vartheta}^*)) = n\int_0^\infty \ln\left(f_{\boldsymbol{\xi}}(t)\right) f_{\boldsymbol{\vartheta}^*}(t)dt, \tag{1}$$

$$D^2(LL(\boldsymbol{\xi}|\boldsymbol{\vartheta}^*)) = n\left\{\int_0^\infty \ln^2(f_{\boldsymbol{\xi}}(t)) f_{\boldsymbol{\vartheta}^*}(t)dt - \left[\int_0^\infty \ln(f_{\boldsymbol{\xi}}(t)) f_{\boldsymbol{\vartheta}^*}(t)dt\right]^2\right\}, \tag{2}$$

while for any pair of models Ξ_1, Ξ_2 and arbitrary fixed vectors $\boldsymbol{\xi}_1$ and $\boldsymbol{\xi}_2$,

$$E(\Delta LL(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2|\boldsymbol{\vartheta}^*)) = n\int_0^\infty \ln\left(f_{\boldsymbol{\xi}_2}(t)/f_{\boldsymbol{\xi}_1}(t)\right) f_{\boldsymbol{\vartheta}^*}(t)dt, \quad (3)$$

$$D^2(\Delta LL(\boldsymbol{\xi}_1,\boldsymbol{\xi}_2|\boldsymbol{\vartheta}^*)) = n\left\{\int_0^\infty \ln^2(f_{\boldsymbol{\xi}_2}(t)/f_{\boldsymbol{\xi}_1}(t)) f_{\boldsymbol{\vartheta}^*}(t)dt - \left[\int_0^\infty \ln(f_{\boldsymbol{\xi}_2}(t)/f_{\boldsymbol{\xi}_1}(t)) f_{\boldsymbol{\vartheta}^*}(t)dt\right]^2\right\}. \quad (4)$$

The validity of Eqs. 1–4 was confirmed using simulated data. E.g., 1000 independent events lists (see Materials and Methods), 500 closed events each, were simulated using Scheme 2 (Fig. 1). The *LL* for observing the closed-time distribution of each events list assuming either Scheme 2_R or 2, with fixed parameters (in this case using the real parameters listed in Fig. 1), was evaluated for both cases. The histograms of the obtained *LL* (Fig. 2, *A* and *B*) and ΔLL (Fig. 2 *C*) values were well fit by Gaussian pdfs (*solid lines*) with mean and variance predicted by Eqs. 1 and 2 (Fig. 2, *A* and *B*) or Eqs. 3 and 4 (Fig. 2 *C*). Note that the distribution of ΔLL extends to the negative range (Fig. 2 *C*, *light bars*), although Ξ_2 was chosen to be the true scheme Θ. This is because, even though $\boldsymbol{\vartheta}^*$ is the true parameter vector, the *LL* function is typically at a maximum at some different parameter vector (the ML estimate $\hat{\boldsymbol{\vartheta}}$; see below). Therefore, the description by the correct distribution (Scheme 2) sometimes yields smaller *LL*-values than that using the incorrect (and even simpler) distribution (Scheme 2_R).

If t and $\hat{t}$ is time in two different units, $t = \lambda \cdot \hat{t}$, and f and $\hat{f}$ are the pdfs expressed using those two units, then $\hat{f}(\hat{t}) = \lambda f(\lambda \hat{t})$. If LL and $\hat{L}L$ are obtained using these two time units, it is easy to show—using an integral transform on the variable t—that

$$E\left(\hat{L}L(\boldsymbol{\xi}|\boldsymbol{\vartheta}^*)\right) = E(LL(\boldsymbol{\xi}|\boldsymbol{\vartheta}^*)) + n\ln\lambda,$$

and

$$D^2\left(\hat{L}L(\boldsymbol{\xi}|\boldsymbol{\vartheta}^*)\right) = D^2(LL(\boldsymbol{\xi}|\boldsymbol{\vartheta}^*)).$$

Thus, the expectation of *LL* depends on the time unit, while its variance does not. In contrast, neither the expectation, nor the variance of ΔLL, depends on the time unit, as can be shown by an integral transform on the variable t in Eqs. 3 and 4 (for Eq. 3 this also follows from the latter two equations).

The integrals in Eqs. 1 and 2 can be performed numerically (see Appendix B), but also explicitly if $f_{\boldsymbol{\xi}}$ is single-exponential. If both $f_{\boldsymbol{\vartheta}^*}$ and $f_{\boldsymbol{\xi}}$ are single-exponential pdfs with time constants τ_ϑ and τ_ξ, respectively, then $E(LL(\boldsymbol{\xi}|\boldsymbol{\vartheta}^*)) = -n(\ln\tau_\xi + \tau_\vartheta/\tau_\xi)$, $D^2(LL(\boldsymbol{\xi}|\boldsymbol{\vartheta}^*)) = n(\tau_\vartheta^2/\tau_\xi^2)$, and $D(LL(\boldsymbol{\xi}|\boldsymbol{\vartheta}^*)) = \sqrt{n}(\tau_\vartheta/\tau_\xi)$. E.g., if $\tau_\vartheta = 10\text{ms}$ and the data are fitted with the proper distribution, then $D(LL) = \sqrt{n}$, and $E(LL) = -n \cdot 3.303$ if the pdf is expressed in units of ms^{-1} or $E(LL) = n \cdot 3.605$ if the pdf is expressed in units of s^{-1} (see above, $\lambda = 1000$). If $f_{\boldsymbol{\xi}}$ is single-exponential but $f_{\boldsymbol{\vartheta}^*}$ is multiexponential, then Eqs. 1 and 2 can be written in the explicit form of

$$E(LL(\boldsymbol{\xi}|\boldsymbol{\vartheta}^*)) = -n\left(\ln\tau_\xi + \left(\textstyle\sum_k a_{\vartheta k}\tau_{\vartheta k}\right)/\tau_\xi\right)$$

and

$$D^2(LL(\boldsymbol{\xi}|\boldsymbol{\vartheta}^*)) = n\left[\left(2\left(\textstyle\sum_k a_{\vartheta k}\tau_{\vartheta k}^2\right) - \left(\textstyle\sum_k a_{\vartheta k}\tau_{\vartheta k}\right)^2\right)\Big/\tau_\xi^2\right].$$

Representation of nested pairs of models

A simpler model is a submodel of a broader model if it can be identified with the broader model under some restriction on the parameters of the latter. The schemes in Fig. 1 are organized in such pairs. The schemes in the right column, with notation R (for reduced), can be identified with the schemes

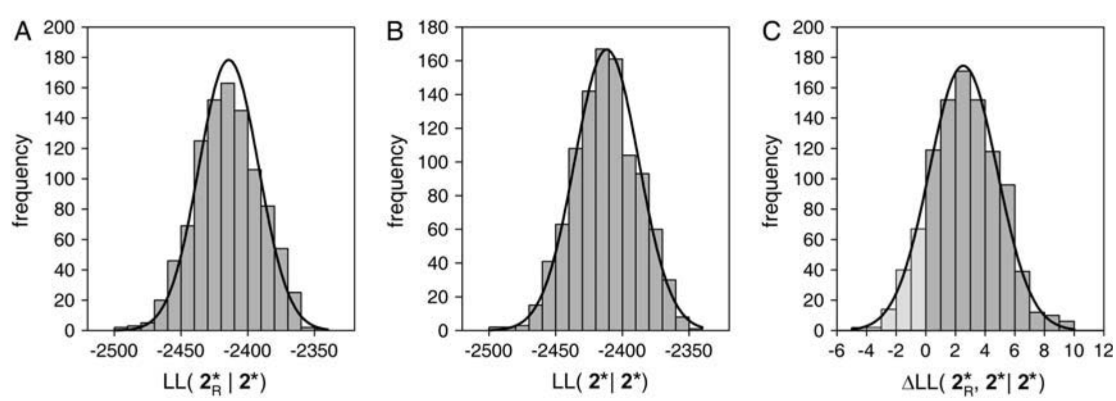

FIGURE 2 Distributions of *LL* and ΔLL values obtained from comparing two models with arbitrary fixed parameters. (*A–C*) One-thousand independent events lists, 500 closed events each, were simulated using Scheme 2. The *LL* of the observed closed-time distribution assuming either Scheme 2 or Scheme 2_R, with rates printed in Fig. 1, was determined for each events list. These *LL*-values for assuming (*A*) Scheme 2_R ($LL(\mathbf{2}_R^*|\mathbf{2}^*)$) and (*B*) Scheme 2 ($LL(\mathbf{2}^*|\mathbf{2}^*)$), as well as (*C*) their differences ($\Delta LL(\mathbf{2}_R^*, \mathbf{2}^*|\mathbf{2}^*) = LL(\mathbf{2}^*|\mathbf{2}^*) - LL(\mathbf{2}_R^*|\mathbf{2}^*)$), were binned to obtain the histograms shown. Light-colored bins in *C* identify cases in which ΔLL was negative. Solid fit lines are appropriately scaled (by the bin-width times the number of events lists) Gaussian pdfs with means and variances predicted by Eqs. 1 and 2 (*A* and *B*) or Eqs. 3 and 4 (*C*).

to their left under some particular restraint. The open-time distributions are identical for each pair of schemes, while the closed-time distributions of the schemes to the left are more complicated than those of the schemes to the right.

For instance, the closed-time distribution of Scheme 1 is defined by two free parameters, τ_{c1} and τ_{c2} (Fig. 3 *A*; for this particular closed-time distribution the fractional amplitudes a_{c1} and a_{c2} are functions of the two time constants τ_{c1} and τ_{c2}). Scheme 1_R is a special case of Scheme 1 in which, e.g., τ_{c1} is fixed to zero. Thus, the parameter space of Scheme 1_R (Ω_R) is a one-dimensional line within the two-dimensional full parameter space Ω of Scheme 1 (Fig. 3 *C*, *vertical shaded line*). (Equivalently, in terms of rates, the closed-time distribution of Scheme 1 is defined by k_{12} and k_{23}, and reduces to that of Scheme 1_R, e.g., if $k_{23} \rightarrow \infty$.)

Analogously, the closed-time distribution of Scheme 2 contains three free parameters, τ_{c1}, τ_{c2}, and a_{c1} (Fig. 3 *D*). Scheme 2_R is a special case of Scheme 2, e.g., under the constraints $\tau_{c1} = \tau_{c2}$, and a_{c1} is any fixed value. Thus, the parameter space Ω_R of Scheme 2_R is a one-dimensional line (one free parameter) within the three-dimensional full parameter space Ω of Scheme 2 (Fig. 3 *F*, *shaded line* identified by *arrow*).

Distribution of Δll *values obtained from fitting the data with the correct scheme and its subset, using the ML estimates of the parameters*

In practice, the parameters of any fitted scheme Ξ are left free and optimized by ML; i.e., the particular vector $\hat{\boldsymbol{\xi}}$, which maximizes the *LL*, is chosen. Fitting a submodel of Ξ to the data is identical to restricting the search for the best parameter vector to a subset Ω_R of the full parameter space Ω. This ML estimate within Ω_R will be denoted $\hat{\boldsymbol{\xi}}_R$.

Because the fitted parameter vectors are not fixed, the direct applicability of Eqs. 3 and 4 is limited. However, if the fitted model is the correct model Θ, and certain regularity criteria are met (Conditions I–VI in Appendix A), then the ML estimate $\hat{\boldsymbol{\vartheta}}$ is asymptotically normally distributed around the true parameter $\boldsymbol{\vartheta}^*$, and converges to $\boldsymbol{\vartheta}^*$ with probability 1 as *n* increases (29–31). Moreover, even if a set of incorrect restrictions is imposed on the true model, i.e., $\boldsymbol{\vartheta}^* \notin \Omega_R$, the

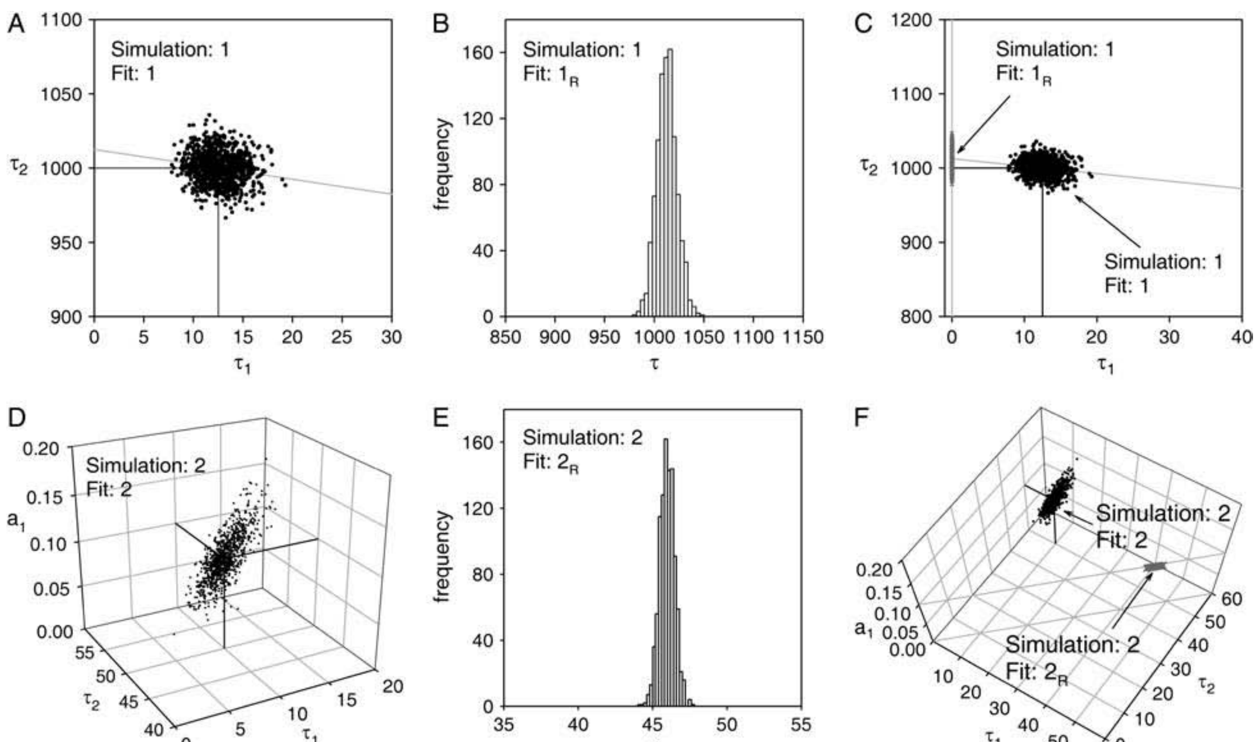

FIGURE 3 Asymptotic behavior of the ML estimate $\hat{\boldsymbol{\vartheta}}$ when the correct scheme or its subset is assumed. (*A*,*D*) One-thousand independent events lists, 10,000 closed events each, were simulated using Schemes 1 and 2, and the closed-time distribution of each events list was ML-fitted to the generating scheme. Obtained parameter vector estimates $\hat{\boldsymbol{\vartheta}}$ are plotted as solid dots in two-dimensional space for Scheme 1 (*A*) and in three-dimensional space for Scheme 2 (*D*). The intersection of the solid auxiliary lines marks the location of the true parameter vector $\boldsymbol{\vartheta}^*$. (*B*,*E*) Histograms of ML-estimates of τ_c obtained by fitting the same events lists with the respective reduced Schemes 1_R and 2_R (i.e., by single-exponential distributions). The histograms are centered on the mean closed times of Scheme 1 and 2; 1012.5 ms (*B*) and 46 ms (*E*). (*C*,*F*) Unconstrained ($\hat{\boldsymbol{\vartheta}}$, *solid dots*) and constrained ($\hat{\boldsymbol{\vartheta}}_R$, *shaded x-marks*, appear merged into *shaded lines*) parameter vector estimates plotted together in three-dimensional space for Scheme 1 (*C*) and in three-dimensional space for Scheme 2 (*F*). The vertical shaded line in *C* and the shaded line identified by the long arrow in *F* represent Ω_R for those two cases. The slanting shaded line in *A* and *C* contains all parameter vectors that define distributions with a mean identical to that of Scheme 1 (1012.5 ms).

restricted ML estimate $\hat{\boldsymbol{\vartheta}}_R$ converges in most cases to some particular point in Ω_R. (For instance, if a set of data is fitted by a single-exponential pdf, the ML estimate of the time constant converges to the mean of the true distribution; see the note in Appendix A.) This best parameter within Ω_R will be denoted $\boldsymbol{\vartheta}_R^*$. (All reduced schemes in Fig. 1 are printed with parameter values corresponding to $\boldsymbol{\vartheta}_R^*$.)

Fig. 3 provides a visual representation of these behaviors. One-thousand independent events lists, with 10,000 closed events each, were simulated using Schemes 1 and 2 (Fig. 1), and the closed-time distributions were fitted to the same schemes, and to the respective reduced schemes 1_R and 2_R, using ML. The solid dots in Fig. 3, *A* and *D*, show the unconstrained $\hat{\boldsymbol{\vartheta}}$ estimates in two-dimensional space for Scheme 1 (Fig. 3 *A*) and in three-dimensional space for Scheme 2 (Fig. 3 *D*)—in both cases these estimates scatter closely around their true values marked by the intercept of solid auxiliary lines. Fig. 3, *B* and *E*, show histograms of τ_c estimates obtained from the restricted fits to the submodels 1_R and 2_R; both approximate narrow Gaussians centered on the respective mean closed times of 1012.5 ms (Fig. 3 *B*) and 46 ms (Fig. 3 *E*) predicted by the true generating schemes (Fig. 1). The corresponding $\hat{\boldsymbol{\vartheta}}_R$ vectors are shown in Fig. 3, *C* and *F*, as shaded *X*-marks.

Because for large *n*, the value of $\hat{\boldsymbol{\vartheta}}$ becomes very similar to $\boldsymbol{\vartheta}^*$, and $\hat{\boldsymbol{\vartheta}}_R$ to $\boldsymbol{\vartheta}_R^*$, it might be expected that the distribution of *LL* values obtained from the free or restricted ML procedure will not be very different from that obtained when the fit parameters are fixed to $\boldsymbol{\vartheta}^*$ or $\boldsymbol{\vartheta}_R^*$, respectively. Thus, if the data are generated using model Θ with true parameter vector $\boldsymbol{\vartheta}^*$, Ω_R is a closed subset of Ω such that $\boldsymbol{\vartheta}^* \notin \Omega_R$, and the data are ML-fitted with the correct model and its subset, then $LL(\hat{\boldsymbol{\vartheta}}|\boldsymbol{\vartheta}^*)$, $LL(\hat{\boldsymbol{\vartheta}}_R|\boldsymbol{\vartheta}^*)$, and $\Delta LL(\hat{\boldsymbol{\vartheta}}_R, \hat{\boldsymbol{\vartheta}}|\boldsymbol{\vartheta}^*) = LL(\hat{\boldsymbol{\vartheta}}|\boldsymbol{\vartheta}^*) - LL(\hat{\boldsymbol{\vartheta}}_R|\boldsymbol{\vartheta}^*)$ are all asymptotically normally distributed with means and variances approximated as follows. For the ML fit to the true model,

$$E\Big(LL(\hat{\boldsymbol{\vartheta}}|\boldsymbol{\vartheta}^*)\Big) \approx n\int_0^\infty \ln\left(f_{\boldsymbol{\vartheta}^*}(t)\right) f_{\boldsymbol{\vartheta}^*}(t)dt \tag{5}$$

and

$$D^2\Big(LL(\hat{\boldsymbol{\vartheta}}|\boldsymbol{\vartheta}^*)\Big) \approx n\left\{\int_0^\infty \ln^2\left(f_{\boldsymbol{\vartheta}^*}(t)\right) f_{\boldsymbol{\vartheta}^*}(t)dt - \left[\int_0^\infty \ln\left(f_{\boldsymbol{\vartheta}^*}(t)\right) f_{\boldsymbol{\vartheta}^*}(t)dt\right]^2\right\}; \tag{6}$$

for the ML fit to the subset of the true model,

$$E\Big(LL(\hat{\boldsymbol{\vartheta}}_R|\boldsymbol{\vartheta}^*)\Big) \approx n\int_0^\infty \ln\Big(f_{\boldsymbol{\vartheta}_R^*}(t)\Big) f_{\boldsymbol{\vartheta}^*}(t)dt \tag{7}$$

and

$$D^2\Big(LL(\hat{\boldsymbol{\vartheta}}_R|\boldsymbol{\vartheta}^*)\Big) \approx n\left\{\int_0^\infty \ln^2\Big(f_{\boldsymbol{\vartheta}_R^*}(t)\Big) f_{\boldsymbol{\vartheta}^*}(t)dt - \left[\int_0^\infty \ln\Big(f_{\boldsymbol{\vartheta}_R^*}(t)\Big) f_{\boldsymbol{\vartheta}^*}(t)dt\right]^2\right\}; \tag{8}$$

and, finally, for the obtained log-likelihood ratio,

$$E\Big(\Delta LL(\hat{\boldsymbol{\vartheta}}_R, \hat{\boldsymbol{\vartheta}}|\boldsymbol{\vartheta}^*)\Big) \approx n\int_0^\infty \ln\left(f_{\boldsymbol{\vartheta}^*}(t)/f_{\boldsymbol{\vartheta}_{R^*}}(t)\right) f_{\boldsymbol{\vartheta}^*}(t)dt \tag{9}$$

and

$$D^2\Big(\Delta LL(\hat{\boldsymbol{\vartheta}}_R, \hat{\boldsymbol{\vartheta}}|\boldsymbol{\vartheta}^*)\Big) \approx n\left\{\int_0^\infty \ln^2\Big(f_{\boldsymbol{\vartheta}^*}(t)/f_{\boldsymbol{\vartheta}_R^*}(t)\Big) f_{\boldsymbol{\vartheta}^*}(t)dt - \left[\int_0^\infty \ln\Big(f_{\boldsymbol{\vartheta}^*}(t)/f_{\boldsymbol{\vartheta}_R^*}(t)\Big) f_{\boldsymbol{\vartheta}^*}(t)dt\right]^2\right\}. \tag{10}$$

Appendix A provides a mathematically exact formulation of the above intuitive statements together with the proofs; Eqs. 5 and 6 are formulated as Statement 3, Eqs. 7 and 8 as Statement 4, and Eqs. 9 and 10 as Statement 5 (for an equation on a related problem, see (30)). Appendix B provides technical advice for numerically calculating the above integrals. The online Supplementary Material shows that the conditions required for the proof of Statement 3 are met by the multiexponential pdfs relevant to ion-channel dwell-time distributions. Note that neither Eq. 9 nor Eq. 10 depends on the time unit.

An important condition for Eqs. 9 and 10 to hold (Statement 5, Appendix A) is $\boldsymbol{\vartheta}^* \notin \Omega_R$, which implies that $\boldsymbol{\vartheta}^*$ and $\boldsymbol{\vartheta}_R^*$ are two distinct points of the full parameter space Ω. (The opposite case, $\boldsymbol{\vartheta}^* \in \Omega_R$, means the restricted model is the correct model and $\boldsymbol{\vartheta}_R^* = \boldsymbol{\vartheta}^*$.) This condition is illustrated in Fig. 3, *C* and *F*, which replots, in the same panel, free $\hat{\boldsymbol{\vartheta}}$ estimates (*solid dots*) and $\hat{\boldsymbol{\vartheta}}_R$ estimates (*shaded x-marks*) for Scheme 1 (Fig. 3 *C*) and Scheme 2 (Fig. 3 *F*). Note the clear separation of the sets of solid and shaded symbols in parameter space.

Extensive simulations were done to verify the validity of Eqs. 9 and 10 (Statement 5), as well as to test how fast the distributions of $\Delta LL(\hat{\boldsymbol{\vartheta}}_R, \hat{\boldsymbol{\vartheta}}|\boldsymbol{\vartheta}^*)$ converge to their predicted shapes. Fig. 4 illustrates a subset of the results, for Scheme 1 versus 1_R (Fig. 4 *A*), 2 versus 2_R (Fig. 4 *B*), and 4 versus 4_R (Fig. 4 *C*), with *n* ranging from 125 to 10,000; it affords the following conclusions.

First, the description of the distributions of ΔLL is excellent for all three models when $n \geq 2000$, but already reasonably good for ~250 events. Naturally, no negative ΔLL values occur when nested models are compared using the ML estimates of the parameters, unlike for the case shown in Fig. 2 *C*, where ΔLL was obtained from comparison of fixed-parameter schemes. This is because $\boldsymbol{\vartheta}_R^*$ is also a member of the full parameter space Ω. Therefore, in the cases when $LL(\boldsymbol{\vartheta}_R^*|\boldsymbol{\vartheta}^*)$ happens to exceed $LL(\boldsymbol{\vartheta}^*|\boldsymbol{\vartheta}^*)$, the unconstrained ML procedure will find a *LL* maximum corresponding to a $\hat{\boldsymbol{\vartheta}} \approx \boldsymbol{\vartheta}_R^*$ (instead of $\hat{\boldsymbol{\vartheta}} \approx \boldsymbol{\vartheta}^*$ as assumed for Eqs. 9 and 10), and ΔLL will never be negative. Thus, the ΔLL distribution is truncated at zero, and the discrepancy between the histograms and predicted lines in Fig. 4 (normal distributions characterized by Eqs. 9 and 10) can be ascribed

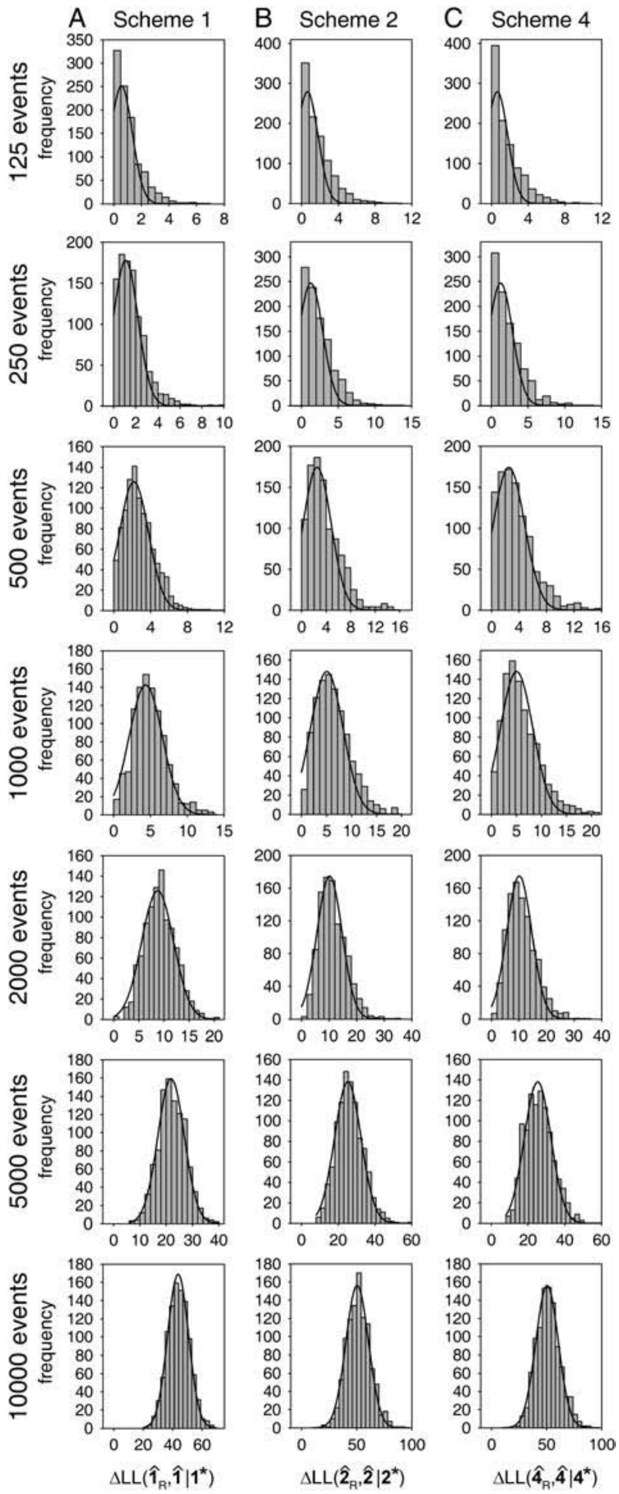

FIGURE 4 Distributions of free ΔLL values for the case when the broader scheme is true. (*A–C*) For each of Schemes 1, 2, and 4, events lists of increasing lengths, ranging from 125 to 10,000 closed events, were simulated, with 1000 independent events lists for each scheme and each length. The closed-time distribution of each events list was fitted by ML to both the appropriate generating scheme and its reduced pair. The resulting ΔLL values for Scheme 1 versus 1_R (*A*), Scheme 2 versus 2_R (*B*), and Scheme 4 versus 4_R (*C*), are plotted in the form of histograms. The solid lines are appropriately scaled Gaussian pdfs with mean and variance predicted by Eqs. 9 and 10.

to the data corresponding to the missing negative tail, spread over the positive part of the distribution. The fit improves as n increases and the lower tail of the Gaussian diverges away from zero (as predicted by Eqs. 9 and 10). Note that the distinction between all three pairs of closed-time distributions in Fig. 4 is relatively difficult; for easier cases the convergence of the distribution of ΔLL to the predicted Gaussian pdf will be faster.

Second, although Schemes 2 and 4 predict identical closed-time distributions (Fig. 1), Scheme 4 has two gateway states. Therefore, for Scheme 4 the durations of adjacent closed events are correlated (long closed events, transitions to state C_1, are clustered, as are the shorter closed events generated by state C_2), hence the independence of $t_1, t_2, \ldots t_n$, necessary for the proofs of asymptotically normal behavior and for Eqs. 2, 4, 6, 8, and 10, does not hold. Nevertheless, the simulations clearly show that the Gaussian prediction, together with the above equations, provides identically good descriptions of the distributions of LL (not shown) and ΔLL values for Scheme 4 (Fig. 4 *C*) as for Scheme 2 (Fig. 4 *B*). This need not be surprising, as ΔLL depends only on the distribution, not on the sequence, of observed closed durations, and the distributions are predicted to be identical for both schemes beyond a few tens of events.

Distribution of ΔLL values obtained from fitting the data with the correct scheme and its generalization, using the ML estimates of the parameters—the case of one extra parameter

The case that, between a more complicated scheme and its subset, the restricted model is the correct one, has long been studied (29–31). If a set of regularity criteria (Conditions I–VI in Appendix A) are met, $2\Delta LL$ is asymptotically χ^2_k distributed with a degree of freedom (k) equal to the difference in the number of free parameters between the two compared models (the χ^2-theorem). However, it has not yet been examined whether the required regularity criteria are satisfied in models that describe ion-channel gating. This and the following section examine this question for two cases; the case of one extra parameter ($k = 1$), and the more common case of one extra exponential component (two extra parameters, $k = 2$), in the more complicated scheme.

The pair of Schemes 1 and 1_R exemplifies the situation for one extra parameter; the closed-time distribution of Scheme 1 is defined by two (τ_{c1} and τ_{c2}), that of Scheme 1_R by one free parameter (τ_c). The full parameter space Ω is the subset of two-dimensional space bounded by $\tau_{c1} \geq 0$ and $\tau_{c2} \geq \tau_{c1}$; the reduced space Ω_R is the line $\tau_{c1} = 0$ (Fig. 5 *A*). When Scheme 1_R is true, $\boldsymbol{\vartheta}^* \in \Omega_R$. Thus, $\boldsymbol{\vartheta}^*$ lies on the boundary of Ω, while the χ^2-theorem requires Conditions I–VI to hold in an open region in Ω which contains $\boldsymbol{\vartheta}^*$. Although less well known, the distribution of $2\Delta LL$ has been studied for several specific geometries of Ω and Ω_R in which $\boldsymbol{\vartheta}^*$ is either a boundary point of Ω_R (but an interior point of Ω;

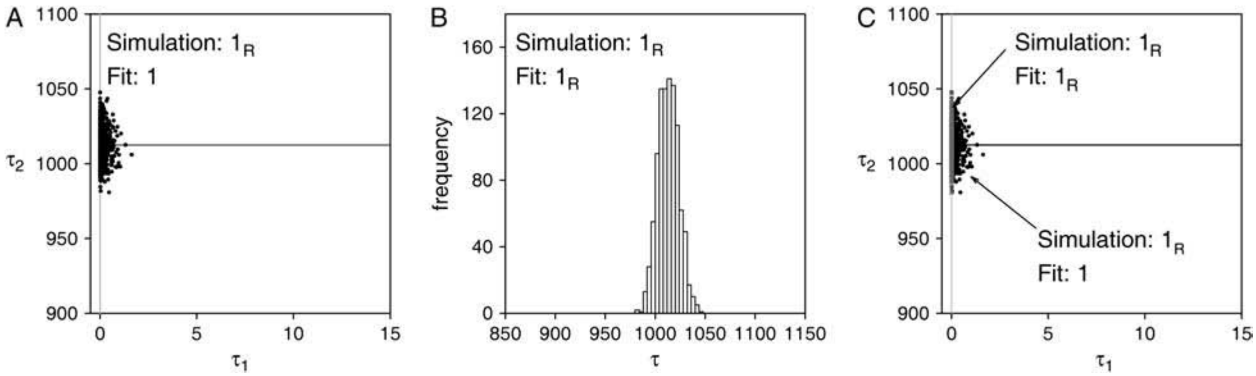

FIGURE 5 Asymptotic behavior of the ML estimate $\hat{\boldsymbol{\vartheta}}$ when the correct scheme or its generalization is assumed—the case of one extra parameter. (*A*) One-thousand independent events lists, 10,000 closed events each, were simulated using Scheme 1_R, and the closed-time distribution of each events list was ML-fitted to the more general Scheme 1. Obtained parameter vector estimates $\hat{\boldsymbol{\vartheta}}$ are plotted as solid dots in full (two-dimensional) parameter space Ω. The vertical shaded line represents Ω_R (i.e., $\tau_{c1} = 0$); the intersection of this line with the horizontal solid line marks the location of the true parameter vector $\boldsymbol{\vartheta}^*$. (*B*) Histogram of ML-estimates of τ_c obtained by fitting the same events lists with the true generating Scheme 1_R (i.e., with a single-exponential distribution). The histogram is centered on the true parameter $\tau_c = 1012.5$ ms. (*C*) Unconstrained ($\hat{\boldsymbol{\vartheta}}$, *solid dots*) and constrained ($\hat{\boldsymbol{\vartheta}}_R$, *shaded x-marks*, appear merged into a *shaded line*) parameter vector estimates plotted together in full, two-dimensional, parameter space Ω.

e.g., (40–43)), or, like here, a boundary point of Ω itself (44). However, the mathematical treatment of these situations still required in each case fulfillment of the regularity criteria on the boundary itself. Unfortunately, these criteria do not apply in the present situation for $\tau_{c1} = 0$. E.g., contrary to Condition V (Appendix A), the information matrix $\mathbf{I}(\boldsymbol{\vartheta})$ is infinite there (see Supplementary Material, Section 2).

To test whether this fact compromises any of the properties of the ML procedure, extensive simulations were done using Scheme 1_R, followed by ML fits to the closed-time distributions of both Schemes 1_R and 1. Fig. 5 shows the scatter of the parameter estimates for 1000 independent simulated events lists (10,000 closed events each). Although $\boldsymbol{\vartheta}^*$ lies on the boundary of Ω, the unconstrained ML estimates $\hat{\boldsymbol{\vartheta}}$ (Fig. 5 *A*, *solid dots*) are still approximately normally distributed on the half-plane. The restrained ML estimate $\hat{\boldsymbol{\vartheta}}_R$ is, of course, approximately normally distributed on the line (*histogram* in Fig. 5 *B*). As $\boldsymbol{\vartheta}_R^* = \boldsymbol{\vartheta}*$, the sets of $\hat{\boldsymbol{\vartheta}}$ and $\hat{\boldsymbol{\vartheta}}_R$ are merged in Ω space (*solid dots* and *shaded x-marks*, Fig. 5 *C*; compare to Fig. 3 *C*), and Eqs. 9 and 10 do not apply (see Statement 5, Appendix A).

Next, the distributions of $\Delta LL(\hat{\boldsymbol{\vartheta}}_R, \hat{\boldsymbol{\vartheta}}|\boldsymbol{\vartheta}_R^*) = LL(\hat{\boldsymbol{\vartheta}}|\boldsymbol{\vartheta}_R^*) - LL(\hat{\boldsymbol{\vartheta}}_R|\boldsymbol{\vartheta}_R^*)$ were examined. If $2\Delta LL$ were distributed as χ_k^2 then, equivalently, ΔLL should follow a Γ-distribution with shape parameter $k/2$ and scale parameter 1 ($\Gamma(k/2, 1)$. Thus, if theχ^2-theorem holds, the pdf of the distribution of ΔLL values should be given by

$$h_{\Delta LL(\hat{\vartheta}_R,\hat{\vartheta}|\vartheta_R^*)}(x) = \begin{cases} \Gamma(k/2)^{-1}x^{(k/2)-1}e^{-x}, & \text{for } x \geq 0 \\ 0, & \text{for } x < 0 \end{cases}, \quad (11)$$

where $\Gamma(p) = \int_0^\infty x^{p-1}e^{-x}dx$. Fig. 6 *A* shows histograms of $\Delta LL(\hat{\mathbf{1}}_R, \hat{\mathbf{1}}|\mathbf{1}_R^*)$ obtained from events lists with increasing n, simulated using Scheme 1_R. In contrast to the case in which the broader scheme is true (Fig. 4 *A*), the distribution of ΔLL does not shift with increasing n; instead, it is well fit, already for as few as 125 events, by a $\Gamma(k/2, 1)$ distribution with $k = 1$ (*solid line*; compare to $k = 2$, *thin line*). To better resolve the distribution at small x, the natural logarithm of ΔLL, $\ln\Delta LL$, was also binned to construct histograms (Fig. 6 *B*). This transformation converts the $\Gamma(k/2, 1)$ pdf into a new pdf,

$$g_{\ln\Delta LL(\hat{\vartheta}_R,\hat{\vartheta}|\vartheta_R^*)}(z) = \Gamma(k/2)^{-1}\exp((k/2)z - \exp(z)), \quad (12)$$

a function that peaks at $z = \ln(k/2)$. (The same transformation is routinely used for the display of multiexponential distributions of ion-channel dwell-times (4).) After this transformation, the $\ln\Delta LL$ distributions were still well fit with parameter $k = 1$ (Fig. 6 *B*, *solid lines*; compare to $k = 2$, *thin lines*). Finally, when the $\Delta LL(\hat{\mathbf{1}}_R, \hat{\mathbf{1}}|\,\mathbf{1}_R^*)$ values themselves were ML-fitted to the pdf in Eq. 11, with k as a free parameter, these fits returned $k \approx 1$ in each case (not shown). Thus, although $\boldsymbol{\vartheta}^*$ lies on the boundary of Ω, the ML estimates $\hat{\boldsymbol{\vartheta}}$ and $\hat{\boldsymbol{\vartheta}}_R$ are still asymptotically normally (or half-normally) distributed, and ΔLL is asymptotically distributed as $\Gamma(1/2, 1)$; i.e., $2\Delta LL$ approaches χ_1^2.

Distribution of ΔLL values obtained from fitting the data with the correct scheme and its generalization, using the ML estimates of the parameters—the case of one extra exponential component

A far more common problem in practice is whether a fit to a dwell-time distribution is significantly improved by allowing one extra exponential component (9,21–23,26,32,45–51). The broader model in this case contains two extra free parameters (a time constant and a fractional amplitude), i.e., the full parameter space Ω has two more dimensions than Ω_R. Fig. 7 illustrates Ω for the closed-time distribution of Scheme 2; the subset of two-dimensional space bounded by $\tau_{c1} \geq 0$, $\tau_{c2} \geq \tau_{c1}$, $a_{c1} \geq 0$, and $a_{c1} \leqq 1$. The term Ω_R, which describes Scheme 2_R, is a one-dimensional line in Ω (see Fig. 3 *F*),

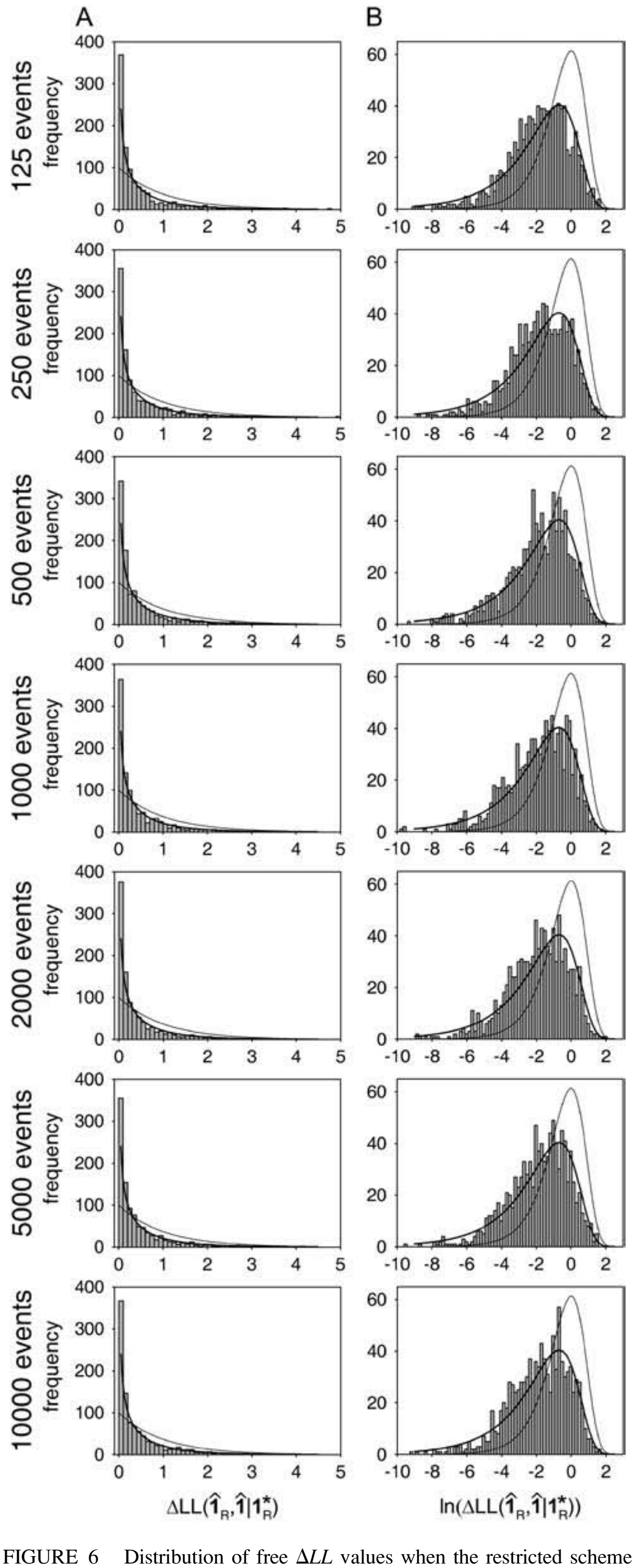

FIGURE 6 Distribution of free ΔLL values when the restricted scheme is true—the case of one extra parameter. (*A*,*B*) Events lists with numbers of closed events ranging from 125 to 10,000, with 1000 independent events lists each, were simulated using Scheme 1_R. The closed-time distributions were ML-fitted to both Schemes 1_R and 1. Obtained $\Delta LL(\hat{\mathbf{1}}_R, \hat{\mathbf{1}} | \mathbf{1}^*_R)$ (*A*) and $\ln \Delta LL(\hat{\mathbf{1}}_R, \hat{\mathbf{1}} | \mathbf{1}^*_R)$ values (*B*) were binned to construct histograms. The solid lines in *A* and *B* are plots of Eqs. 11 and 12, respectively, using $k = 1$. As a comparison, thin lines illustrate the case of $k = 2$.

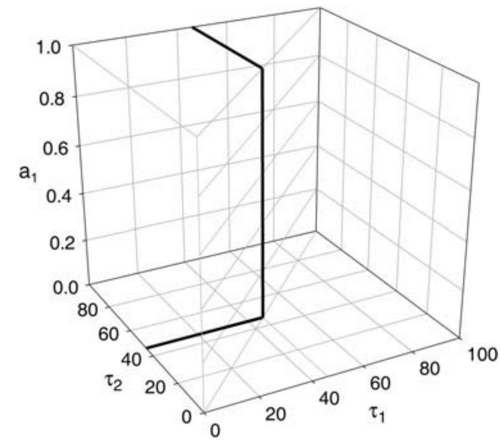

FIGURE 7 Full parameter space of the closed-time distribution of Scheme 2. The parameter space Ω is the subset of three-dimensional space bounded by $\tau_{c1} \geq 0$, $\tau_{c2} \geq \tau_{c1}$, $a_{cl} \leq 0$, and $a_{cl} \leq 1$. The reduced space Ω_R, which describes Scheme 2_R, is a one-dimensional line in Ω (see Fig. 3 *F*), defined, e.g., by the constraints $\tau_{c2} = \tau_{c1}$, and a_{cl} = fixed (to any value). The jointed solid line exemplifies a set of distinct points in Ω, which nevertheless all represent identical distributions.

defined, e.g., by the constraints $\tau_{c2} = \tau_{c1}$, and a_{c1} = fixed to any value. Thus, once again, Ω_R lies on the boundary of Ω.

However, a more important problem also arises in this case. This is that once $\tau_{c2} = \tau_{c1}$, the value of a_{c1} is immaterial—the dwell-time distributions are identical for any a_{c1}. The same problem arises for the other two ways in which Ω can be reduced to Ω_R (if $a_{c1} = 1$, τ_{c2} is immaterial, if $a_{c1} = 0$, τ_{c1} is immaterial). Thus, all points of Ω that lie on a jointed line like the one drawn in Fig. 7 (*solid line*) represent identical distributions. This is a serious violation of the regularity criteria which assume that different parameters $\boldsymbol{\vartheta}$ correspond to different distributions (Condition I, see Appendix A). $\ln\left(f_{(\tau_{c1},\tau_{c2},a_{c1})}(t)\right)$ is constant on any line segment defined by $\tau_{c2} = \tau_{c1}$ (fixed), $0 \leq a_{c1} \leq 1$ (Fig. 7, *vertical segment* of *solid line*); therefore $(\partial/\partial a_{c1})\ln\left(f_{(\tau_{c1},\tau_{c2},a_{c1})}(t)\right) = 0$ on the whole plane $\tau_{c2} = \tau_{c1}$. Consequently, the information matrix $\mathbf{I}(\boldsymbol{\vartheta})$ is singular for all $\boldsymbol{\vartheta} \in \Omega_R$ (also if Ω_R is defined as $a_{c1} = 0$ or $a_{c1} = 1$, see Supplementary Material, Section 3), rather than positive definite as required by Condition V.

Given the above irregularities, the properties of the ML procedure were extensively tested on events lists simulated using Scheme 2_R and fitted to the closed-time distributions of both Schemes 2_R and 2. Because $\hat{\boldsymbol{\vartheta}}$ is expected to scatter in a broad region of Ω, special care had to be taken to identify it (see Materials and Methods). As expected, the unconstrained $\hat{\boldsymbol{\vartheta}}$ estimates were scattered in the region of Ω surrounding the jointed line, which contains the points that all represent Scheme 2_R (Fig. 8 *A*, *solid dots*), while the constrained ML estimates were approximately normally distributed on the line around the true parameter τ_c (histogram in Fig. 8 *B*).

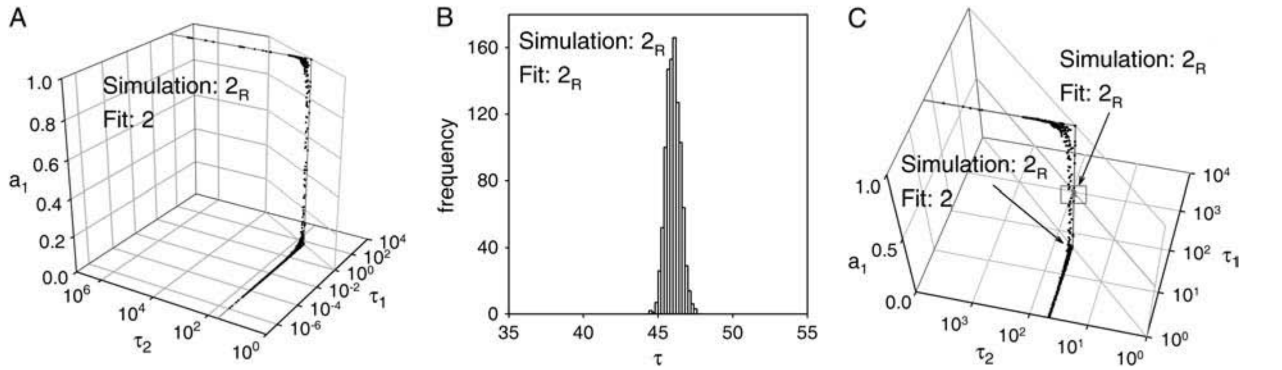

FIGURE 8 Asymptotic behavior of the ML estimate $\hat{\boldsymbol{\vartheta}}$ when the correct scheme or its generalization is assumed—the case of one extra exponential component. (*A*) One-thousand independent events lists, 10,000 closed events each, were simulated using Scheme 2_R, and the closed-time distribution of each events list was ML-fitted to the more general Scheme 2 (see Materials and Methods for details). Obtained parameter vector estimates $\hat{\boldsymbol{\vartheta}}$ are plotted as solid dots in three-dimensional parameter space Ω. (*B*) Histogram of ML-estimates of τ_c obtained by fitting the same events lists with the true generating Scheme 2_R (i.e., with a single-exponential distribution). The histogram is centered on the true parameter $\tau_c = 46$ ms. (*C*) Unconstrained ($\hat{\boldsymbol{\vartheta}}$, *solid dots*) and constrained ($\hat{\boldsymbol{\vartheta}}_R$, *shaded x-marks*, identified by *shaded box*) parameter vector estimates plotted together in full, three-dimensional, parameter space Ω. The subset Ω_R (*shaded line* identified by *upper arrow*) is arbitrarily drawn as the line defined by $\tau_{c1} = \tau_{c2}$, $a_{c1} = 0.5$.

Once again, as $\boldsymbol{\vartheta}^* \in \Omega_R$, the sets of $\hat{\boldsymbol{\vartheta}}_R$ and $\hat{\boldsymbol{\vartheta}}$ are merged in Ω space (Fig. 8 *C*, *shaded x-marks* identified by *shaded box*, and *solid dots*; compare to Fig. 3 *F*), and Eqs. 9 and 10 do not apply (see Statement 5, Appendix A).

To study the distributions of $\Delta LL(\hat{\boldsymbol{\vartheta}}_R, \hat{\boldsymbol{\vartheta}}|\boldsymbol{\vartheta}^*_R)$ for this situation, the obtained $\Delta LL(\hat{\mathbf{2}}_R, \hat{\mathbf{2}}|\mathbf{2}^*_R)$ were binned to form histograms (Fig. 9). Interestingly, these histograms were not well fit by the pdfs of $\Gamma(k/2, 1)$ distributions, whether $k = 1$ or $k = 2$ was assumed (Fig. 9 *A*, *thin lines*; compare to Fig. 6 *A*). In particular, a large fraction of all ΔLL values fell into the first bin. A similar deviation of $2\Delta LL$ from a χ^2 statistics was observed in a study using ML for fitting macroscopic ionic currents (52), and was attributed to the constraint that rate constants are nonnegative. However, as the χ^2-theorem clearly applies for the comparison of Schemes 1 versus 1_R (Fig. 6), a more likely reason for the unexpected behavior of $\Delta LL(\hat{\mathbf{2}}_R, \hat{\mathbf{2}}|\mathbf{2}^*_R)$ (Fig. 9) is the fact that Scheme 2_R cannot be identified with a unique point in the parameter space of Scheme 2 (Fig. 7).

To obtain a better resolution in the small-ΔLL range, histograms were again constructed from the natural logarithm of ΔLL (Fig. 9 *B*, *thin lines* show appropriately transformed $\Gamma(k/2, 1)$ pdfs with $k = 1$ and $k = 2$; see Fig. 6 *B*). This representation clearly shows that the ΔLL values belong to a mixture of two distinct distributions, one of which is clustered around machine-zero (the smallest positive number a computer algorithm can return) with a pdf approximating a δ-function. The fractional amplitude of this component decreases as n increases (Fig. 9 *B*). (The parallel apparent shift to the right of the δ-function is an artifact due to the decreasing floating-point precision to which ΔLL values are calculated, since ΔLL is obtained as the difference of two LL values which themselves increase linearly with n.)

To provide an empirical description of the two components of these distributions, an arbitrary cutoff value of $\Delta LL \geq 10^{-6}$ was set ($\ln(\Delta LL) = -13.8$). The values of $\Delta LL < 10^{-6}$ were assumed to belong to the component characterized by the δ-function, whereas the values of $\Delta LL \geq 10^{-6}$ were separately considered.

Fig. 10 *A* shows the fractional amplitude (α) of the latter component as a function of n, determined from an extensive series of experiments including the ones illustrated in Fig. 9. The solid line in Fig. 10 *A* is a least-squares fit to the observed $\alpha(n)$ of the empirical function

$$\alpha(n) = 1 - A_1/(A_2 + \log n)^4, \qquad (13a)$$

with best fit parameters $A_1 = 149.2$ and $A_2 = 2.927$ (log, 10-base logarithm).

The above results recall studies in which $2\Delta LL$ was shown to be a mixture of χ^2_0 (a component identically zero), χ^2_1, and χ^2_2 for cases in which $\boldsymbol{\vartheta}^*_R$ is on the boundary of Ω_R or Ω but the regularity criteria are satisfied (40–44). Therefore, as $\Delta LL(\hat{\boldsymbol{\vartheta}}_R, \hat{\boldsymbol{\vartheta}}|\boldsymbol{\vartheta}^*_R)$ values greater than 10^{-6} were still not well fit by either a $\Gamma(1/2, 1)$ or a $\Gamma(2/2, 1)$ distribution, it seemed natural to fit all $\Delta LL \geq 10^{-6}$ to a mixture of the above two distributions—with the fractional amplitude β of the latter component left as a free parameter ($0 \leq \beta \leq 1$). Such a pdf indeed improved the quality of the fit significantly, as compared to fitting with just one or the other component. Interestingly, the β-values returned by these ML-fits were also not constant, but increased with increasing n, from $\beta \approx 0.6$ (for $n = 125$) to $\beta \approx 0.9$ (for $n = 10{,}000$; Fig. 10 *B*, *solid dots*). The solid line in Fig. 10 *B* is a least-squares fit to the obtained $\beta(n)$ of the empirical function

$$\beta(n) = 1 - B_1/(B_2 + \log n)^4, \qquad (13b)$$

with best-fit parameters $B_1 = 212.2$ and $B_2 = 2.703$.

Thus, the empirical pdf of $\Delta LL(\hat{\mathbf{2}}_R, \hat{\mathbf{2}}|\mathbf{2}^*_R)$ can be assembled as

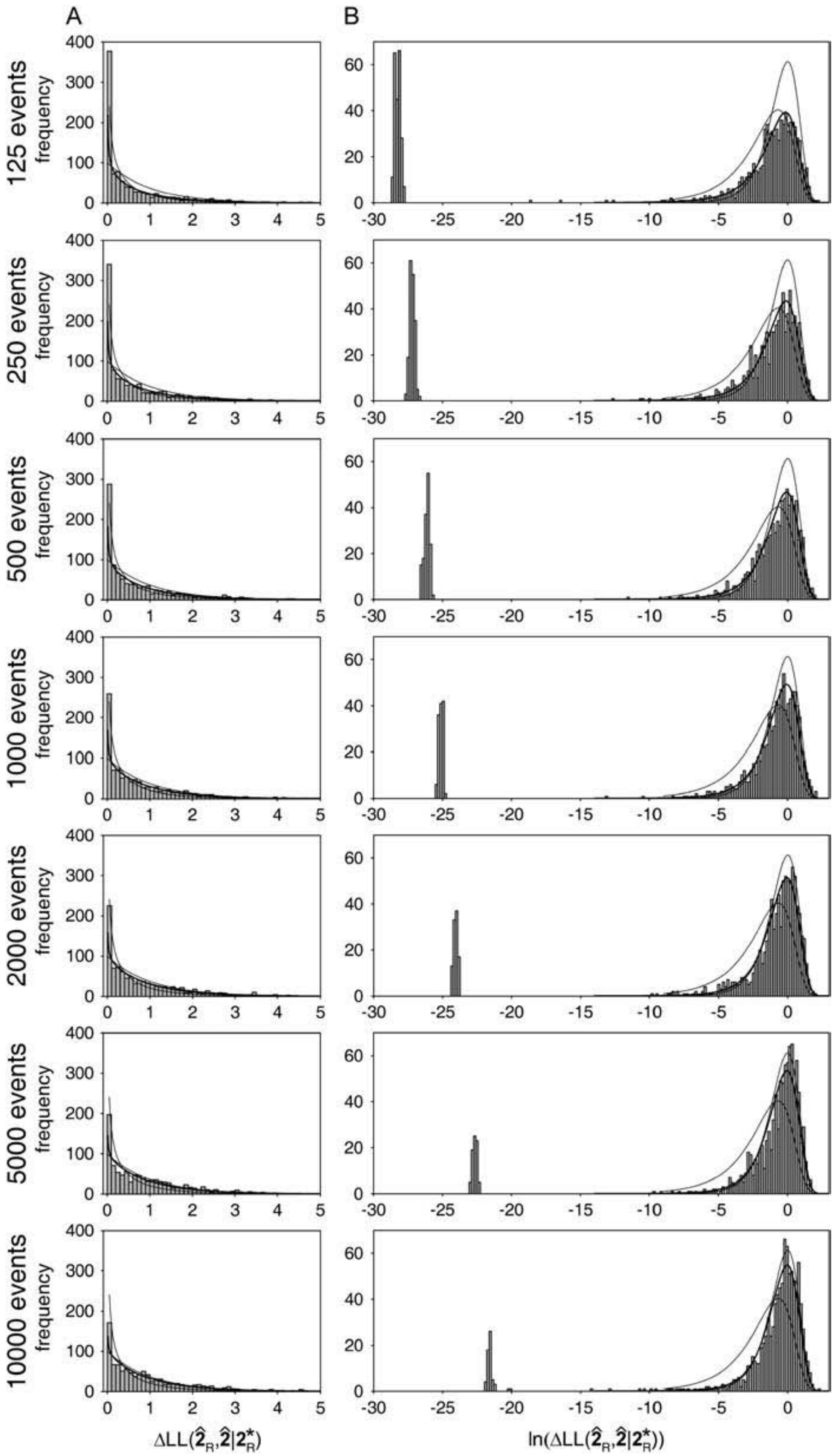

FIGURE 9 Distribution of free ΔLL values when the restricted scheme is true—the case of one extra exponential component. (*A*,*B*) Events lists with numbers of closed events ranging from 125 to 10,000, with 1000 independent events lists each, were simulated using Scheme 2_R. The closed-time distributions were ML-fitted to both Schemes 2_R and 2. Obtained $\Delta LL(\hat{\mathbf{2}}_R, \hat{\mathbf{2}}|\,\mathbf{2}_R^*)$ (*A*) and ln $\Delta LL(\hat{\mathbf{2}}_R, \hat{\mathbf{2}}|\mathbf{2}_R^*)$ values (*B*) were binned to construct histograms. Thin lines in *A* and *B* are plots of Eqs. 11 and 12, respectively, using $k = 1$ and $k = 2$ (see Fig. 6); solid lines are plots of Eq. 14 in *A*, and of its appropriate transform in *B*.

$$h_{\Delta LL(\hat{\vartheta}_R, \hat{\vartheta}|\vartheta_R^*)}(x) = \begin{cases} (1-\alpha(n))\delta_0(x) + \alpha(n)((1-\beta(n))(\pi x)^{-1/2}e^{-x} + \beta(n)e^{-x}), & \text{for } x \geq 0 \\ 0, & \text{for } x < 0 \end{cases}, \tag{14}$$

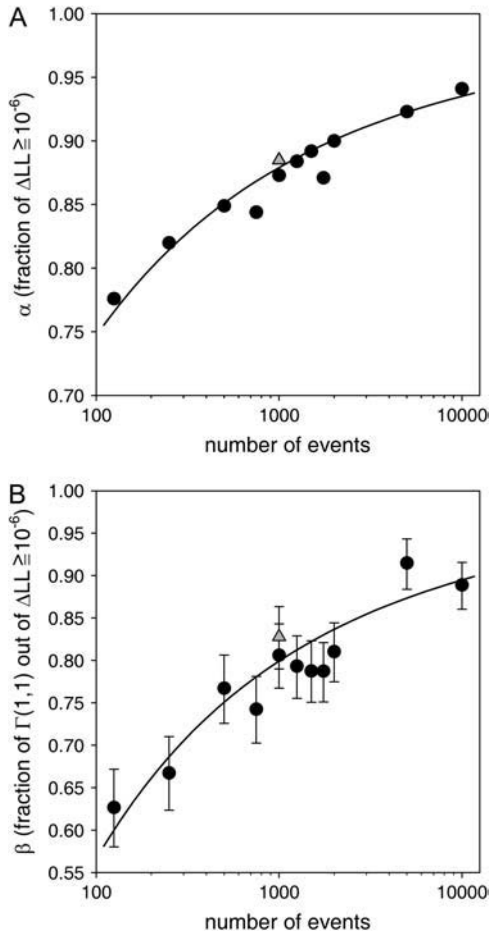

FIGURE 10 Parameters of the empirical distribution which best fits ΔLL—the case of one extra exponential component. Events lists with numbers of closed events (n) ranging from 125 to 10,000, 1000 independent events lists each, were simulated using Scheme 2_R. The closed-time distributions were ML-fitted to both Schemes 2_R and 2. (*A*) Fraction (α) of $\Delta LL(\hat{\mathbf{2}}_R, \hat{\mathbf{2}}|\mathbf{2}_R^*)$ values $\geq 10^{-6}$ plotted as a function of n (*solid circles*) and the fit line described by Eq. 13a (*solid line*). (*B*) The set of $\Delta LL(\hat{\mathbf{2}}_R, \hat{\mathbf{2}}|\mathbf{2}_R^*) \geq 10^{-6}$ was fitted by ML to a pdf of the form $(1-\beta)h_1 + \beta h_2$, with β as a free parameter, where h_1 and h_2 are $\Gamma(k/2, 1)$ pdfs (Eq. 11) with $k = 1$ and $k = 2$, respectively. Solid circles and error bars show estimates of β and 0.5-unit likelihood intervals. The solid line is a fit of the solid circles by Eq. 13b. Shaded triangles in *A* and *B* illustrate the results of an identical analysis for a set of 1000 independent events lists simulated using Scheme 3_R and fitted to both Schemes 3_R and 3 (see Fig. 11).

where $\alpha(n) = 1–149.2/(2.927 + \log n)^4$, and $\beta(n) = 1–212.2/(2.703 + \log n)^4$. ($\delta_0(x)$ is a δ-function centered on zero.) The solid lines in all panels of Fig. 9, *A* and *B*, illustrate the pdf predicted by Eq. 14 (transformed as in Eq. 12 for the panels in Fig. 9 *B*).

However, $\beta(n)$ from Eq. 13b did not provide a uniformly good description for the whole range of $\Delta LL \geq 10^{-6}$. When the ML fits to Eq. 14, with β as a free parameter, were repeated for $\Delta LL \geq 0.3$, the β-values converged to $\beta = 1$ in each case. Because in most statistical applications, in particular in log-likelihood ratio tests, a good prediction of the tail of the distribution is most important, a practically useful empirical description of the pdf of $\Delta LL(\hat{\mathbf{2}}_R, \hat{\mathbf{2}}|\mathbf{2}_R^*)$ is given by Eq. 14 with $\alpha(n)$ from Eq. 13a and $\beta = 1$. In this context, for $x > 0$, Eq. 14 simplifies to

$$h_{(\hat{\vartheta}_R, \hat{\vartheta}|\vartheta_R^*)}(x) = \alpha(n)e^{-x}. \tag{15}$$

How representative is the case of one versus two exponential components of the general case of one extra exponential component? The regularity violations (Fig. 7) apply identically for all those cases. But, most importantly, does Eq. 14 (or 15) provide a good description of the distribution of $\Delta LL(\hat{\vartheta}_R, \hat{\vartheta}|\vartheta_R^*)$ for every such case? To address this question, 1000 independent events lists (1000 closed events each) were simulated using Scheme 3_R, and the closed-time distributions were ML-fitted to both Schemes 3_R and 3, i.e., to pdfs with four and five exponential components. The distribution of obtained $\Delta LL(\hat{\mathbf{3}}_R, \hat{\mathbf{3}}|\mathbf{3}_R^*)$ values is displayed in Fig. 11 in the form of a linear (Fig. 11 *A*) and a logarithmic (Fig. 11 *B*) histogram. This distribution is remarkably similar to the one obtained from the comparison of Schemes 2_R and 2, 1000 events (Fig. 9 *A*, B, *row 4*). In particular, it was not well fit by $\Gamma(k/2, 1)$ pdfs whether $k = 1$ or $k = 2$ was used (Fig. 11, *A* and *B*, *thin lines*), but was well fit by the pdf described in Eq. 14 (Fig. 11, *A* and *B*, *solid line*). Both the fraction of ΔLL greater than 10^{-6} (Fig. 10 *A*, *shaded triangle*), and the best β-value for $\Delta LL \geq 10^{-6}$ (Fig. 10 *B*, *shaded triangle*) were closely similar to those obtained for one versus two exponential components. Also, when only the tail of the distribution ($\Delta LL \geq 0.3$) was fitted, the fit again converged to $\beta = 1$. Thus, the results illustrated in Figs. 9–11, as well as Eq. 14 or 15,

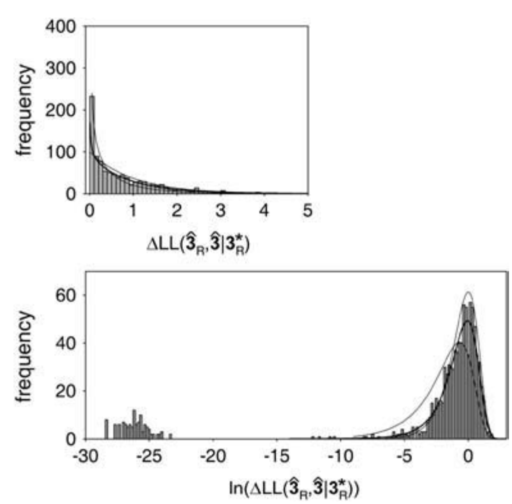

FIGURE 11 Distribution of free ΔLL values when the restricted scheme is true—the case of five versus four exponential components. (*A*,*B*) One-thousand independent events lists, 1000 closed events each, were simulated using Scheme 3_R. The closed-time distributions were ML-fitted to both Schemes 3_R and 3. Obtained $\Delta LL(\hat{\mathbf{3}}_R, \hat{\mathbf{3}}|\mathbf{3}_R^*)$ (*A*) and $\ln \Delta LL(\hat{\mathbf{3}}_R, \hat{\mathbf{3}}|\mathbf{3}_R^*)$ values (*B*) were binned to construct histograms. Thin lines in *A* and *B* are plots of Eq. 11 and Eq. 12, respectively, using $k = 1$ and $k = 2$; the solid line is a plot of Eq. 14 in *A*, and of its appropriate transform in *B*.

characterize the distribution of $\Delta LL(\hat{\boldsymbol{\vartheta}}_{\mathrm{R}}, \hat{\boldsymbol{\vartheta}}|\boldsymbol{\vartheta}^*_{\mathrm{R}})$ for the general case of one extra exponential component.

Strategies for model discrimination

Log-likelihood ratio test (LLR-test)

The log-likelihood ratio test (LLR test) is used to evaluate whether a model Θ or its submodel Θ_{R} is more appropriate to describe the data. If Θ_{R} is true (the null hypothesis) and the regularity criteria are satisfied $2\Delta LL$ is asymptotically χ^2_{k}-distributed (k is the difference in the number of free parameters (28,29)), and the chance of occurrence of a ΔLL greater than or equal to that observed is given by the integral of the tail of the χ^2_{k} pdf between $2\Delta LL_{\mathrm{observed}}$ and infinity. The submodel Θ_{R} is rejected (Θ is accepted) at significance level P (typically $P = 0.10$, 0.05, or 0.01) if $2\Delta LL_{\mathrm{observed}}$ is larger than the χ^2_{k}-value corresponding to P.

The LLR test has been used in the past for comparison of nested pairs of ion-channel gating models (e.g., (2,32,53)), but without verifying the validity of the required regularity criteria. As pointed out in the previous two sections, these criteria apply in the interior of Ω, but not at its boundary. Thus, a χ^2_{k} distribution for $2\Delta LL$ under the null hypothesis is guaranteed only if Θ_{R} is an interior point of Θ, e.g., when the restricted model consists of fixing a subset of the parameters or introducing linear constraints or microscopic reversibility (6). Interestingly, $2\Delta LL$ was found χ^2_1-distributed also for introduction of one superfluous parameter (Fig. 6) even if $\boldsymbol{\vartheta}^*_{\mathrm{R}}$ was a boundary point of Θ but distinct points $\boldsymbol{\vartheta}^*_{\mathrm{R}}$ corresponded to distinct pdfs (see Fig. 5).

However, for the most frequent application of judging the improvement of the fit of a dwell-time distribution by a multiexponential pdf upon introduction of one extra exponential component (e.g., (9,16,22,23,32,45–51)), the assumption that $2\Delta LL$ is χ^2_2-distributed under the null hypothesis is unwarranted and wrong (Fig. 9). A practically useful empirical pdf of ΔLL for that case is given by Eq. 15. The area under the tail of this pdf can be calculated; the cutoff value for ΔLL for a fixed P-value and n is $\Delta LL_{\mathrm{cutoff}} = \ln(\alpha(n)/P)$. Table 1 lists a set of such calculated $\Delta LL_{\mathrm{cutoff}}$ values for n ranging from 125 to 100,000, and for $P = 0.10$, 0.05, and 0.01. Thus, using Table 1, a valid LLR test can now be used for discriminating whether l or $l+1$ exponential components are present in a dwell-time distribution (corresponding to l vs. $l+1$ single-channel states in a gating model). Fig. 12 illustrates Type I errors obtained using the LLR test based on Table 1, and P-values of 0.10 (*circles*), 0.05 (*squares*), and 0.01 (*triangles*), for a large set of data simulated using Scheme 2_{R} and fitted by Schemes 2_{R} and 2 (Fig. 1). These errors scatter closely around their predicted values (*horizontal lines*). It must be noted that, although the overall distribution of $2\Delta LL$ differs substantially from a χ^2_2 for all but extremely large n, the tail of the distribution does not (note the weak dependence on n of $\Delta LL_{\mathrm{cutoff}}$ in Table 1; see also Fig. 9). Thus, fortunately, conclusions drawn in previous studies based on the unwarranted χ^2_2-assumption are unlikely to be far wrong.

TABLE 1 ΔLL cutoff values for the LLR test to distinguish l vs. $l+1$ exponential components

Number of events	$P < 0.10$	$P < 0.05$	$P < 0.01$
125	2.0359	2.7290	4.3385
250	2.0974	2.7906	4.4000
500	2.1414	2.8346	4.4440
750	2.1615	2.8546	4.4641
1000	2.1738	2.8669	4.4764
1250	2.1824	2.8755	4.4850
1500	2.1889	2.8820	4.4915
1750	2.1940	2.8871	4.4966
2000	2.1982	2.8914	4.5008
5000	2.2221	2.9152	4.5246
10,000	2.2356	2.9288	4.5382
20,000	2.2464	2.9395	4.5490
100,000	2.2641	2.9572	4.5667
⋮			
Infinite	2.3026	2.9957	4.6052

The empirical pdf of ΔLL for the case when the l-component distribution is true (Eq. 15) depends on the number of fitted events (n). For each n- and P-value the listed $\Delta LL_{\mathrm{cutoff}}$ was calculated as $\ln(\alpha(n)/P)$, with $\alpha(n)$ from Eq. 13a.

Akaike information criterion (AIC) and Schwarz criterion (SC)

The goodness of fit by any two alternative models, whether nested or not, is frequently compared using the Akaike information criterion (AIC, (33)) or the Schwarz criterion (SC, (34)). Although the SC was originally tested on LL estimates obtained from fitting the entire sequence of single-channel closed and open durations (7), it can be adapted to ML fits of a dwell-time distribution for gating schemes with

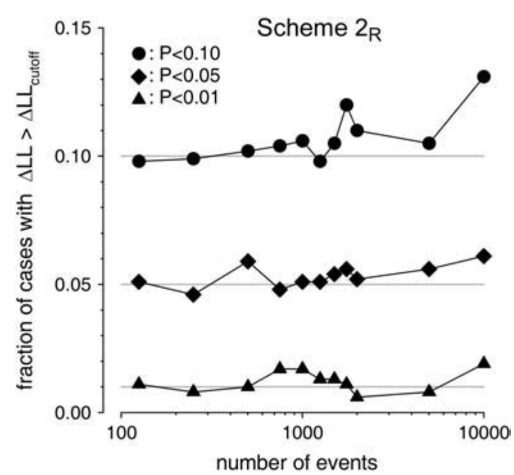

FIGURE 12 Type I errors obtained using the LLR test based on Table 1. Events lists with numbers of closed events ranging from 125 to 10,000, one-thousand independent events lists each, were simulated using Scheme 2_{R} and the closed-time distributions ML-fitted to both Schemes 2_{R} and 2. Circles, squares, and triangles show the fraction of cases in which the obtained ΔLL was larger than the cutoff specified in Table 1 for P-values of 0.10, 0.05, and 0.01, respectively. Horizontal lines indicate the expected values.

only one gateway state (21). In the present context, a larger model is considered better by the AIC if $\Delta LL > k$, or by the SC if $\Delta LL > 0.5k\ln(2n)$, where k is the difference in the number of free parameters, and n is the number of fitted, closed or open, events. Both methods penalize more free parameters, but the SC also considers the size of the data set. The AIC (7,12,13,53–57) and the SC (7,12,21,26,57) have been widely used in ion-channel model discrimination. Their relative merit is debatable, as different studies found the SC either more (7), similarly (57), or even less (12) efficient than the AIC in identifying the right model.

New nonbiased decision strategy based on the parameter estimates and the distributions of ΔLL

For the comparison of a pair of nested models the LLR test exploits the knowledge of the distribution of ΔLL under the null hypothesis, but it does not take into account the distribution of ΔLL for the alternative case. The philosophy of the LLR test is to always accept Θ_R unless the latter can be excluded with very high certainty. Thus, it is biased toward Θ_R by ensuring that the Type I error (rejection of Θ_R, although true) is kept at a constant (low) value, without trying to minimize the Type II error (rejection of Θ, although true). Although rewarding parsimony does have its merits (Occam's razor), identification of complexities using the LLR test will be inefficient for all but large amounts of data. An alternative philosophy, without an a priori bias, would be to always choose the model which seems more likely, i.e., to equalize the probabilities of the two types of error. Although this cannot be rigorously achieved, this section describes a new, semiempirical strategy attempting to approximate that aim. The decision is based on consideration of two aspects of the fit.

First, the fit parameter $\hat{\boldsymbol{\vartheta}}$ itself is inspected. When Θ_R is true, fitting with Θ in many cases yields parameters which are unreasonable. E.g., when a dwell-time distribution is ML-fitted with too many components, frequently the fit returns $\tau_i \approx \tau_j$ for $i \neq j$, or $a_i \approx 0$ (Figs. 5 *A* and 8 *A*). For many, but not all, such cases $\Delta LL \approx 0$. Such fits (called overparameterized) are excluded from further analysis, i.e., Θ_R is accepted as the better model. In practice, the following, somewhat arbitrary, set of criteria was adopted. A fit was defined as overparameterized, if either of the following three conditions applied:

1. For some $i \neq j$ $9/10 \leq \tau_i/\tau_j \leq 10/9$.
2. For any i $|a_i| < 0.005$.
3. $\Delta LL \leq 0.0015n$, where n is the number of dwell-times analyzed.

Second, if a fit returns a $\hat{\boldsymbol{\vartheta}}$ which itself seems reasonable, the predicted pdfs of ΔLL are compared for the two possible situations, which are that either Θ or Θ_R is true (Fig. 13). If Θ_R is true, $h_R(x) = \mathrm{pdf}_{\Delta LL(\hat{\boldsymbol{\vartheta}}_R, \hat{\boldsymbol{\vartheta}}|\boldsymbol{\vartheta}^*_R)}(x)$ is given either by Eq. 11 (if the χ^2-theorem holds; e.g., if $\boldsymbol{\vartheta}^*_R$ is in the interior of

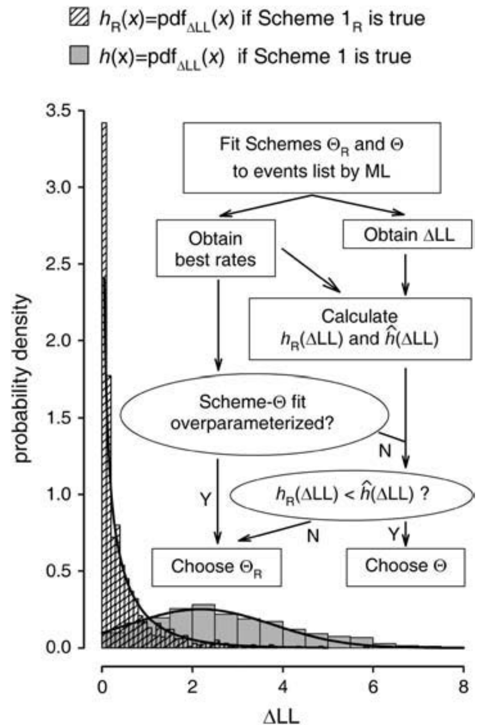

FIGURE 13 Direct comparison of pdfs of ΔLL distributions for the cases in which the submodel or the full model is true. ΔLL values binned to construct the two histograms were obtained by ML-fitting both Schemes 1_R and 1 to 1000 independent events lists (500 closed events each), simulated either using Scheme 1_R (*hatched bars*) or using Scheme 1 (*shaded bars*). Binwidths were 0.1 and 0.5, respectively; and both histograms were normalized by dividing them by 1000 · binwidth to obtain histograms that integrate to unity. Solid lines plot the theoretical pdfs $h_R(x)$ and $h(x)$ (Eqs. 11, 9, and 10, respectively). Flow chart (*inset*) summarizes the new decision algorithm.

Θ, or for a comparison involving one extra parameter such as that between Scheme 1 and 1_R) or by Eq. 15 (for a comparison involving one extra exponential component such as that of Scheme 2 with 2_R or 3 with 3_R). If Θ is true, $h(x) = \mathrm{pdf}_{\Delta LL(\hat{\boldsymbol{\vartheta}}_R, \hat{\boldsymbol{\vartheta}}|\boldsymbol{\vartheta}^*)}(x)$ is approximated by

$$h(x) \approx (1/\sqrt{2\pi\sigma^2})e^{-\frac{(x-m)^2}{2\sigma^2}},$$

where m and σ^2 are given by Eqs. 9 and 10, respectively. Fig. 13 allows visual comparison of appropriately scaled experimental histograms of $h_R(x)$ for the former case (*hatched bars*; obtained from fitting both Schemes 1_R and 1 to events lists simulated using Scheme 1_R), and of the corresponding $h(x)$ (*shaded bars*; from fitting both Schemes 1_R and 1 to simulated Scheme-1 events lists). Solid lines illustrate the theoretical pdfs predicted by Eq. 11 and Eqs. 9 and 10, respectively.

Unfortunately, $h(x)$ depends on $\boldsymbol{\vartheta}^*$ and $\boldsymbol{\vartheta}^*_R$ (Eqs. 9 and 10) which are not known. To approximate $h(x)$, the ML estimates $\hat{\boldsymbol{\vartheta}}$ and $\hat{\boldsymbol{\vartheta}}_R$ are substituted in place of $\boldsymbol{\vartheta}^*$ and $\boldsymbol{\vartheta}^*_R$. Thus, $\hat{h}(x)$ is defined as

$$\hat{h}(x) = \frac{1}{\sqrt{2\pi\hat{\sigma}^2}} e^{-\frac{(x-\hat{m})^2}{2\hat{\sigma}^2}}, \qquad (16)$$

where

$$\hat{m} = n \int_0^\infty \ln\left(f_{\hat{\vartheta}}(t)/f_{\hat{\vartheta}_R}(t)\right) f_{\hat{\vartheta}}(t)dt, \qquad (16a)$$

and

$$\hat{\sigma}^2 = n\left\{\int_0^\infty \ln^2\left(f_{\hat{\vartheta}}(t)/f_{\hat{\vartheta}_R}(t)\right) f_{\hat{\vartheta}}(t)dt - \left[\int_0^\infty \ln\left(f_{\hat{\vartheta}}(t)/f_{\hat{\vartheta}_R}(t)\right) f_{\hat{\vartheta}}(t)dt\right]^2\right\}. \qquad (16b)$$

The model Θ_R or Θ is then selected based on whether $h_R(x)$ or $\hat{h}(x)$ is larger for $x = \Delta LL_{observed}$.

Thus, the new strategy can be summarized as follows (Fig. 13, *inset*). The broader model Θ is accepted as the better model if

1. $\hat{\boldsymbol{\vartheta}}$ is not overparameterized based on criteria 1–3, above.
2. $\hat{h}(\Delta LL_{observed}) > h_R(\Delta LL_{observed})$.

For all other cases, the reduced model Θ_R is preferred.

Performance of various decision strategies for the case of one extra parameter or one extra exponential component

The performances of the LLR test, the AIC, the SC, and of the new strategy described in the previous section, were compared on large sets of simulated data. Both the case of one extra parameter (Scheme 1 versus 1_R) and that of one extra exponential component (Scheme 2 versus 2_R) in the broader model was examined. For both cases, events lists of various lengths were simulated using either the respective Scheme Θ_R or the respective Scheme Θ, 1000 independent events lists each. The closed-time distribution of each events list was ML-fitted to both schemes to produce $\hat{\boldsymbol{\vartheta}}$, $\hat{\boldsymbol{\vartheta}}_R$, and ΔLL values. Finally, a decision was made for each events list using all four strategies described above. The results are summarized in Fig. 14 *A* for Scheme 1 versus 1_R, and in Fig. 14 *B* for Scheme 2 versus 2_R. The panels plot, for each set of 1000 events lists, the fraction of cases in which a given strategy (the LLR test is shown for $P = 0.05$) opted for the broader Scheme Θ. Events lists simulated using Scheme Θ are denoted by solid symbols, and those simulated using Scheme Θ_R, by open symbols.

For Scheme 1 versus 1_R $h_R(x)$ (Eq. 11) does not depend on n. Therefore, any strategy that uses a fixed cutoff is expected to give a constant Type I error irrespective of n. Accordingly, both the LLR test (which uses a constant $\Delta LL_{cutoff} = 1.9207$ for $P = 0.05$) and the AIC (which uses a constant $\Delta LL_{cutoff} = 1$) misclassified a constant fraction, ~5% and ~13%, respectively, of all Scheme-1_R events lists (Fig. 14 *A*, *open diamonds* and *triangles*). For Scheme 2 versus 2_R $h_R(x)$ (approximated by Eq. 15) shifts to the right with increasing n. Because the

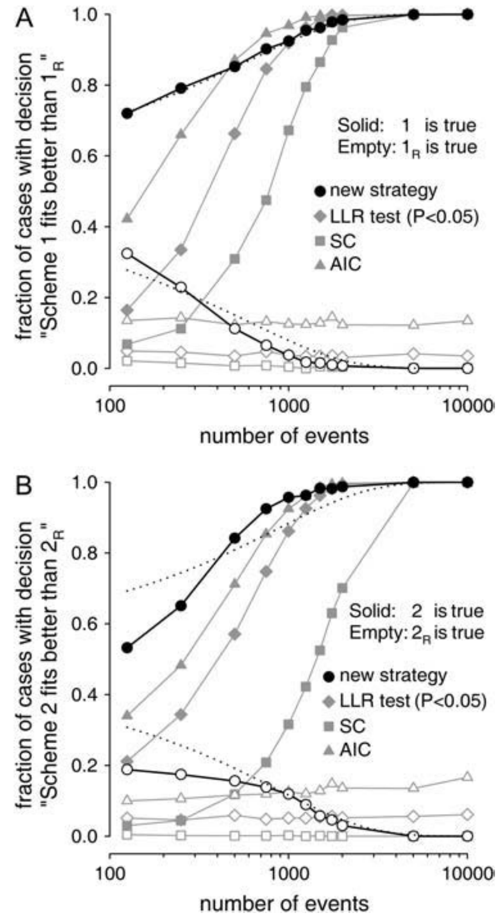

FIGURE 14 Performance of various decision strategies in identifying the correct model. The cases of (*A*) one extra parameter (Scheme 1 versus 1_R), and of (*B*) one extra exponential component (Scheme 2 versus 2_R) in the broader model are illustrated. (*A*,*B*) Events lists with numbers of closed events ranging from 125 to 10,000, one-thousand independent events lists each, were simulated using either the full model Θ (Scheme 1 in *A*, Scheme 2 in *B*) or its submodel Θ_R (Scheme 1_R in *A*, Scheme 2_R in *B*). For each events list the closed-time distribution was ML-fitted to both Θ_R and Θ to produce fit parameters and ΔLL values. A decision in favor of either Θ_R or Θ was then made using four different strategies. Symbols plot, for a given set of 1000 events lists, the fraction of cases in which a given strategy opted for Θ. Solid symbols denote events lists simulated using Θ, open symbols denote those simulated using Θ_R. LLR test for $P = 0.05$, diamonds; AIC, triangles; SC, squares; new strategy, circles. Dotted line*s* illustrate symmetrical Type I and Type II errors expected for using the respective ideal ΔLL_{cut} (see Fig. 15).

AIC uses a fixed cutoff (of 2 in this case; Fig. 15, *C* and *D*, *long-dash line*), the Type I error produced by the AIC increases slightly, from ~10% at 125 events to ~17% at 10,000 events (Fig. 14 *B*, *open triangles*; unlike in Fig. 14 *A*). Accordingly, the LLR test ($P = 0.05$) was performed in this case using Table 1, not a fixed cutoff value (Fig. 15, *C* and *D*, *solid line*). As a consequence, the Type I errors remain constant at ~5% (Fig. 14 *B*, *open diamonds*), as expected for this type of test.

On the other hand, for small n, both methods produced large Type II errors, i.e., failed to recognize Scheme-Θ

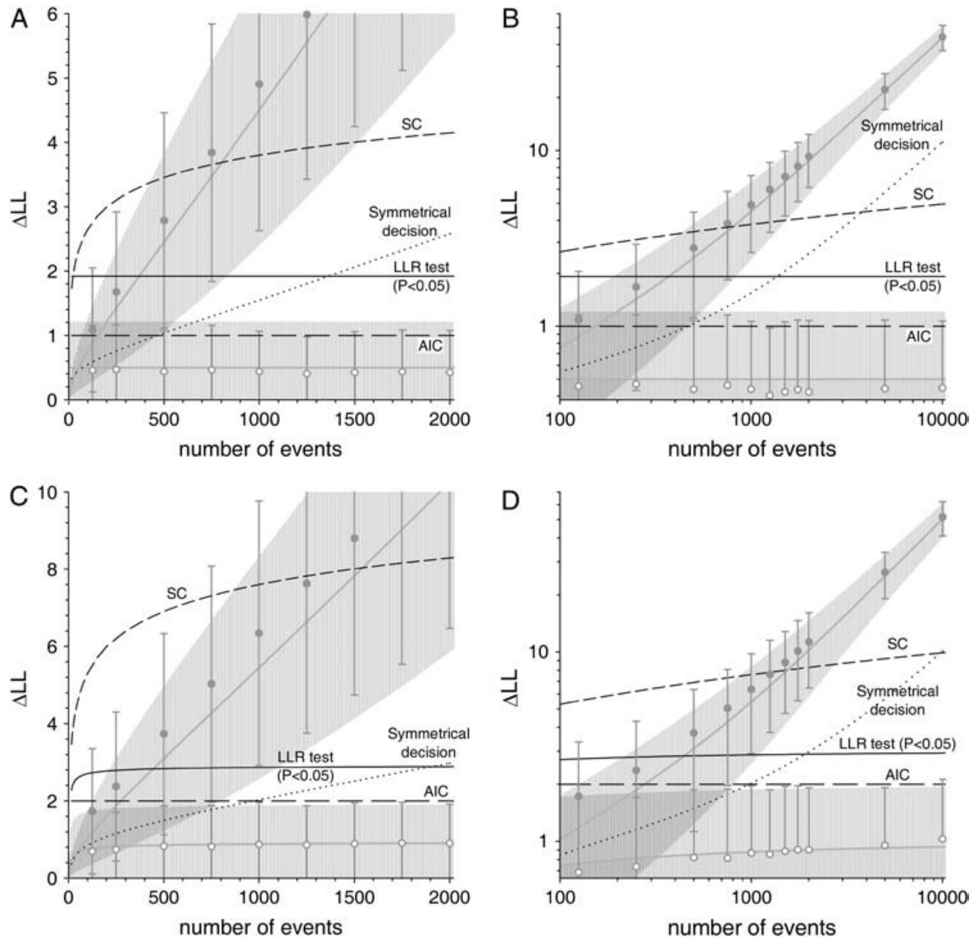

FIGURE 15 Comparison of ΔLL distributions with cutoff values used by various strategies for evaluation of a pair of nested models. The cases of (*A*,*B*) one extra parameter (Scheme 1 versus 1_R), and of (*C*,*D*) one extra exponential component (Scheme 2 versus 2_R) in the broader model are illustrated. The value of ΔLL is plotted as a function of the number of closed events (n) using linear axes in *A* and *C*, and double-logarithmic axes in *B* and *D*. Lower shaded lines and shaded areas plot the expectations and mean ± SD ranges of the pdfs $h_R(x)$ (mean = 0.5, SD = $2^{-1/2}$ (Eq. 11) in *A* and *B*; mean = $\alpha(n)$, SD = $\sqrt{\alpha(n)(2-\alpha)}$ (Eq. 15) in *C* and *D*). Open shaded circles and error bars show observed averages and SD of ΔLL for sets of 1000 independent events lists simulated using the submodels Θ_R (Scheme 1_R in *A* and *B*; 2_R in *C* and *D*) and ML-fitted with both Θ_R and the respective Θ (Scheme 1 in *A* and *B*; Scheme 2 in *C* and *D*). Upper shaded lines and shaded areas plot the expectations and mean ± SD ranges of the pdfs $h(x)$; i.e., Gaussian pdfs truncated at zero, with m and σ^2 from Eqs. 9 and 10 for the scheme-pairs $1_R/1$ (*A* and *B*) and $2_R/2$ (*C* and *D*), respectively. Shaded circles and error bars show observed averages and SD of ΔLL for sets of 1000 independent events lists simulated using the full schemes Θ (Scheme 1 in *A* and *B*; Scheme 2 in *C* and *D*), and ML-fitted with both the respective Θ_R and Θ. Black lines in *A–D* plot cutoff values used by the LLR test ($P = 0.05$, *solid line*), the AIC (*long-dash line*), and the SC (*short-dash line*). Dotted lines in *A–D* illustrate the respective ideal ΔLL_{cut} values satisfying the equation $\int_{\Delta LL_{cut}}^{\infty} h_R(x)dx = \int_0^{\Delta LL_{cut}} h(x)dx$.

events lists as such (*solid diamonds* and *triangles*). E.g., only ~33% of Scheme-1 events lists were correctly identified by the LLR test at $P = 0.05$ when 250 closed events were fitted, and only ~16% when 125 closed events were fitted.

The SC proved to be extremely conservative in both cases (Fig. 14, *A* and *B*; *squares*). It produced negligible Type I errors (*open squares*, <2% in Fig. 14 *A*, <0.5% in Fig. 14 *B*), but was extremely inefficient in correctly identifying Scheme-1 events lists for all but very large n (Fig. 14 *A*, *solid squares*); and recognition of Scheme 2 was even less efficient (Fig. 14 *B*; *solid squares*).

As expected, the new strategy yielded the most symmetrical results (Fig. 14, *A* and *B*; *circles*). Two advantages of this approach over other tests should be noted. First, at low event numbers the decrease in Type II error, produced by the new strategy relative to the other three tests, is much larger than the increase in Type I error, resulting in a larger overall power of discrimination (Fig. 14, *A* and *B*). E.g., the new strategy identifies 72% of true Scheme-1 events lists already from 125 closed events (Fig. 14 *A*, *solid circles*), an improvement by 56% relative to the LLR test, at a cost of misclassifying 32% of Scheme-1_R events lists of identical length (Fig. 14 *A*, *open circles*), a relapse of 27% relative to the LLR test. Second, at high event numbers the Type I error for the new strategy converges to zero (Fig. 14, *A* and *B*, *open circles*), unlike that for the LLR test or the AIC (or any fixed-cutoff strategy) for which Type I errors do not decrease even for very large n.

ΔLL distributions quantitatively explain observed efficiencies of various decision strategies

The performances (including the observed Type II errors) of the LLR test, the AIC, and the SC shown in Fig. 14 can be quantitatively understood (Fig. 15) in light of the knowledge of the distributions of ΔLL for both possibilities, i.e., when Θ_R is true or when Θ is true ($h_R(x)$ and $h(x)$).

Lower shaded lines and shaded areas in Fig. 15, *A–D*, represent the expectation values and calculated mean ± SD ranges, of the distributions characterized by $h_R(x)$ for Schemes 1 versus 1_R (Fig. 15, *A* and *B*) and 2 versus 2_R (Fig. 15, *C* and *D*). Open shaded circles and error bars represent the experimental averages and SD of ΔLL, obtained for sets of 1000 independent simulated Scheme-1_R (Fig. 15, *A* and *B*) or 2_R (Fig. 15, *C* and *D*) events lists of various lengths fitted to both the respective Θ_R and Θ. Upper shaded lines and shaded areas represent the expectation values and mean $\pm$ SD ranges of $h(x)$ for the Scheme pairs $1/1_R$ (Fig. 15, *A* and *B*) and $2/2_R$ (Fig. 15, *C* and *D*). As the means are asymptotically linear in n and the SDs increase as the square-root of the latter (Fig. 15 *A* and *C*), the relative width of both distributions decreases as $1/\sqrt{n}$ (Fig. 15, *B* and *D*; note double-logarithmic axes). Shaded symbols and error bars plot the mean and SD of ΔLL obtained for simulated Scheme-1 (Fig. 15, *A* and *B*) and 2 (Fig. 15, *C* and *D*) events lists fitted to both the respective Θ_R and Θ.

It is enlightening to compare the cutoff values used by the LLR test ($P = 0.05$), the AIC, and the SC (*solid*, *long-dash*, and *short-dash black lines*), with the shaded areas and lines. E.g., in Fig. 15, *A* and *B*, both solid and the short-dash black lines are above the upper shaded area at low n, predicting little chance for correct identification of Scheme-1 events lists by the LLR test for $<\sim$150 events and by the SC for $<\sim$400 events. The intersections of these black lines with the upper shaded line predict 50% recovery of Scheme-1 events lists by the AIC at $\sim$150 events, by the LLR test at $\sim$350 events, and by the SC at $\sim$800 events, respectively. Finally, efficient recovery of Scheme-1 events lists is expected where the black lines are found below the upper shaded area, i.e., at $>\sim$500 events for the AIC, at $>\sim$800 events for the LLR test, and at $>\sim$1500 events for the SC. All of these quantitative predictions are confirmed by the observations on simulated data (Fig. 14 *A*, *solid diamonds*, *triangles*, and *squares*). Fig. 15, *C* and *D*, allow analogous observations for Schemes $2/2_R$. E.g., the cutoff specified by the SC in this case (Fig. 15, *C* and *D*, *short-dash black line*) is even less adapted to the shape of $h(x)$. Comparison of the upper shaded line and shading with the short-dash black line predicts little chance for recognition by the SC of Scheme-2 events lists for $<\sim$800 events, 50% recognition for $\sim$1500 events, and efficient recognition only for $>\sim$2500 events; predictions which all match the experimental findings (Fig. 14 *B*, *solid squares*; compare to Fig. 15 *A*). Comparison of the long-dash black line with the lower shaded area (Fig. 15 *D*) explains the increasing (with n) Type I error of the AIC for Scheme 2 versus 2_R (Fig. 14 *B*, *open triangles*).

To what extent is symmetry inherent in the new method?

Imagine the situation that $\boldsymbol{\vartheta}^*$ and $\boldsymbol{\vartheta}_R^*$ are known; i.e., that a large set of events lists were generated by either one or the other of two possible models, those corresponding to a known $\boldsymbol{\vartheta}^*$ and $\boldsymbol{\vartheta}_R^*$, the task being to identify the generating scheme for each events list. In that case, the cutoff producing symmetrical errors can be numerically calculated by finding ΔLL_{cut}, which satisfies $\int_{\Delta LL_{cut}}^{\infty} h_R(x)dx = \int_0^{\Delta LL_{cut}} h(x)dx$. The dotted lines in Fig. 15, *A–D*, illustrate this ideal ΔLL_{cut} as a function of n for Scheme 1 versus 1_R and Scheme 2 versus 2_R; the dotted lines in Fig. 14, *A* and *B*, illustrate the corresponding expected Type I and Type II errors. Note that the SC, the only cutoff-based strategy attempting to compensate for the length of the data, nevertheless specifies cutoffs that are very different from the respective ideal ΔLL_{cut} (Fig. 15, *A–D*, compare *short-dash* and *dotted lines*).

In real situations $\boldsymbol{\vartheta}^*$ is not known. Therefore, because $h(x)$ depends on $\boldsymbol{\vartheta}^*$, it is not possible to specify a universally ideal symmetrical ΔLL_{cut}. Although the new strategy does not use a fixed cutoff, the results obtained were relatively symmetrical for Scheme 1 versus 1_R and, to a lesser degree, for Scheme 2 versus 2_R (Fig. 14, *A* and *B*, *circles*; compare to *dotted lines*). How do the individual decisions under the new strategy relate to the respective ideal symmetrical ΔLL_{cut} plotted in Fig. 15, *A–D* (*dotted lines*)? When, for a given set of simulated events lists, the ΔLL values from all Θ-decisions were compared with those from all Θ_R-decisions in retrospect, these two sets of ΔLL values were relatively separated in each case, with a small overlap in the region corresponding to the respective ideal ΔLL_{cut}.

Because $h_R(x)$ (Eq. 11 or Eq. 15) does not depend on the parameter values of the two compared models, the Type I errors, and in particular their dependences on n (Fig. 14, *A* and *B*; *open symbols*), are universal, parameter-independent, features of the new strategy. Conversely, as $h(x)$ is a function of $\boldsymbol{\vartheta}^*$, the magnitude of the Type II error is parameter-dependent, and its dependence on n (see Fig. 14, *A* and *B*; *solid symbols*) is expected to shift to the left for easier decisions and to the right for more difficult decisions, as illustrated by the examples below.

With the rate constants chosen for Schemes 1 and 2 (Fig. 1), the identification of these two schemes is relatively hard. Identifying a second exponential component in a distribution is an easier task for most other (reasonable) combinations of rates. E.g., if the magnitudes of the rates r_{31} and r_{32} in Scheme 2 are exchanged, the closed-time distribution of this new Scheme (Scheme 5) is described by $\tau_{c1} = 10$ ms, $\tau_{c2} = 50$ ms, $a_{c1} = 0.9$, and $a_{c2} = 0.1$. This distribution is far easier to distinguish from that of the corresponding submodel Scheme 5_R (identical to 2_R but with $r_{21} = 71.43\ s^{-1}$; $\tau_c = 14$ ms), because Eq. 9 predicts a much larger mean for $h(x)$ ($m = 0.005045n$ for Schemes $2/2_R$; $m = 0.06797n$ for Schemes $5/5_R$). Accordingly, the new method correctly recognized 92% of simulated Scheme-5 events lists even from as few as 125 closed events; and the efficiencies of the other three decision strategies were similarly increased.

On the other hand, distinguishing a five- from a four-component distribution is usually a harder task. E.g., for the pair of Schemes $3/3_R$, Eq. 9 predicts a smaller mean for $h(x)$ than for the pair of Schemes $2/2_R$ ($m = 0.003977n$) for Schemes $3/3_R$. Accordingly, the new method correctly recognized only 84% of simulated Scheme-3 data, even from events lists containing 1000 closed events (and the other three decision strategies were even less efficient with 57%, 74%, and 5.3% of all Scheme-3 events lists correctly identified by the LLR test at $P = 0.05$, the AIC, and the SC, respectively).

Confidence of the decision using the new method

An attractive feature of the LLR test is that it provides a measure of confidence, i.e., the probability P that the decision is wrong, at least for decisions in which the null hypothesis is rejected (the broader model Θ is accepted). It would be desirable for the new method to provide a similar measure of confidence, especially given its larger Type I errors for shorter events lists.

The ratio $h(x)/h_R(x)$ reports how many times more likely model Θ is relative to Θ_R, given the observed ΔLL. However, $h(x)$ is not known, only its approximation $\hat{h}(x)$ (Eq. 16). Therefore, the ratio $\hat{h}(x)/h_R(x)$, termed the decision ratio, was used to compile a database of correlations between experimentally obtained decision ratios and Type I errors, by reexamining the sets of events lists simulated using the reduced Schemes 1_R and 2_R. The new method was modified to accept the broader model (Scheme 1 and 2, respectively) only if the decision ratio was larger than or equal to some fixed value. Fig. 16 shows the resulting Type I errors for required values of 1, 1.2, 1.5, 2, 3, 5, and 10 of the decision ratio (d.r.) for comparing Schemes $1/1_R$ (Fig. 16 *A*) and $2/2_R$ (Fig. 16 *B*). Note that the plots corresponding to d.r. = 1 in Fig. 16, *A* and *B*, are identical to those in Fig. 14, *A* and *B* (*open circles*). Clearly, increasing decision ratios indicate smaller probabilities of Type I error. Because the Type I errors are independent of model parameters, Fig. 16 is a universal indicator of Type I errors for the case of evaluating the need for one extra parameter (Fig. 16 *A*) or one extra exponential component (Fig. 16 *B*). E.g., suppose an $l+1$ component distribution is preferred by the new method over an l-component distribution for the description of data consisting of only 125 dwell-times; then a decision ratio >2 indicates $<7\%$ chance for that decision to be wrong (Fig. 16 *B*, *solid diamonds*). There is a conceptual difference between the confidence estimates of the LLR test and of the new method. A large P-value of the LLR test indicates that the possibility Θ_R is true cannot be excluded with sufficient confidence, without telling whether Θ or Θ_R is more likely. In contrast, the new method both votes for one of the two models *and* provides the probability of error if Θ is chosen.

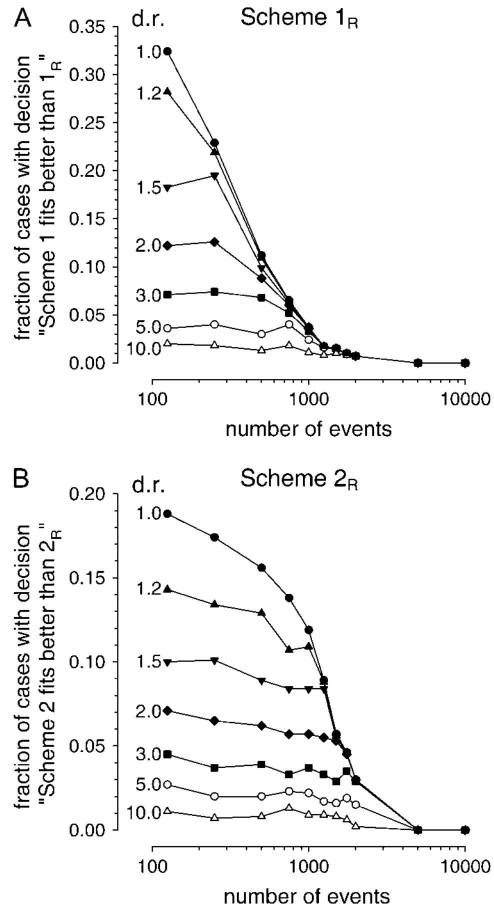

FIGURE 16 Confidence estimates for identification of one extra parameter or one extra exponential component using the new method. (*A*) Events lists with numbers of closed events ranging from 125 to 10,000, one-thousand independent events lists each, were simulated using Scheme 1_R. For each events list, the closed-time distribution was ML-fitted to both Schemes 1_R and 1. Decisions were made using the new method, modified to accept Scheme 1 only for decision ratios greater than or equal to a fixed value. Plots show the fractions of events-lists misclassified as Scheme 1 for the cases when the required decision ratio (d.r.) was set to 1, 1.2, 1.5, 2, 3, 5, and 10. The plot corresponding to d.r. = 1 (*solid circles*) is identical to the plot of open circles in Fig. 14 *A*. (*B*) Analogous to *A*, for the pair of Schemes $2_R/2$.

SIGNIFICANCE

To my knowledge, this work has, for the first time, directly examined the distributions of log-likelihood ratios obtained from comparison of ML-fits of nested pairs of Markov models to dwell-time distributions of single ion-channels.

In the section ''Distributions of Log-Likelihood Ratios'', above, the pdf of the asymptotic distribution of ΔLL was derived for the case in which the broader model is true (Fig. 4, Eqs. 9 and 10). Although the theory assumes independence of the durations of individual dwell times (strictly only true for one-gateway schemes), simulations confirm the intuitive expectation that the theory also applies in the presence of correlations (Fig. 4 *C*). For the case when the submodel is true, regularity criteria required by conventional theory were shown not to apply in common situations. For such cases the empirical distribution of ΔLL was determined from large simulated data sets (Figs. 6 and 9; Eqs. 11 and 15).

Frequently, only dwell times longer than some fixed t_{low} are included in the ML-fitting procedure (2), mostly to exclude from the analysis brief events distorted by limited bandwidth. This circumstance does not alter any of the results described here. Obviously, in such cases the integrals in

Eqs. 1–10 are taken between t_{low} and infinity, and the pdfs rescaled by dividing them with $P\ (t \geq t_{low})$ (4).

In the section ''Strategies for Model Discrimination'', above, a reliable LLR test (Fig. 12) was constructed (Table 1) for discrimination of distributions with l vs. $l+1$ exponential components. The LLR test, the Akaike information criterion and the Schwarz criterion were shown to produce strongly asymmetrical Type I and Type II errors, in full agreement with the distributions of ΔLL under the null hypothesis and its alternative. A new method yielding relatively symmetrical Type I and Type II errors and a higher power of discrimination was developed and characterized. Together with provided confidence estimates, this strategy might become a useful tool when the amount of data is limited, or the schemes are difficult to recognize.

Two immediate practical applications of this work are worth mentioning. First, ML-fitting of isolated (closed or open) dwell-time distributions, for which ΔLL values were studied here, is done in practice at the first stage of analysis, aimed at determining the minimum number of closed- and open-channel states required to describe the data (2,26,27). Second, although the entire sequence of events, rather than a dwell-time distribution, is ML-fitted in more advanced stages of analysis (6–11), some ion channels (e.g., CFTR (21,22) or K_{ATP} channels (23–25)) display only one gateway-state. For such models the LL of the sequence of events is the sum of the LL values obtained from fitting the isolated distributions of open and closed dwell-times (7). Thus, if two one-gateway schemes are compared (which differ either in the number of closed or open states), the ΔLL values obtained from fitting the entire sequence of events will be distributed like those in this work. Future work will need to address the distributions of ΔLL obtained from ML-fitting an entire sequence of events generated by a scheme with correlations.

APPENDIX A: REGULARITY CRITERIA, FORMULATION AND PROOF OF STATEMENTS 3, 4, AND 5

Let $X_n \xrightarrow{a.e.} X$, $X_n \xrightarrow{P} X$, and $X_n \xrightarrow{d} X$ denote convergence of a random variable almost everywhere, in probability, and in distribution, respectively. Let $E_{\boldsymbol{\vartheta}}(X)$ denote the expectation of X based on the distribution characterized by parameter $\boldsymbol{\vartheta}$. Let $t_1, t_2, \ldots t_n$ be identically distributed independent random variables, with pdf $f(\boldsymbol{\vartheta},t)$ depending on parameter vector $\boldsymbol{\vartheta}$ (both f and t_i are considered to be in a dimensionless form). Let $\boldsymbol{\vartheta}^*$ be the true parameter, and $\hat{\boldsymbol{\vartheta}}_n$ the ML estimate of $\boldsymbol{\vartheta}^*$ based on $t_1, t_2, \ldots t_n$.

$$LL\left(\hat{\boldsymbol{\vartheta}}_n|\boldsymbol{\vartheta}^*\right) = LL(\boldsymbol{\vartheta}^*|\boldsymbol{\vartheta}^*) + \left[\frac{\partial}{\partial\boldsymbol{\vartheta}}LL(\boldsymbol{\vartheta}^*|\boldsymbol{\vartheta}^*)\right]\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*\right) + \frac{1}{2}\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*\right)^T\left[\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}}LL(\boldsymbol{\vartheta}^*|\boldsymbol{\vartheta}^*)\right]\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*\right)$$
$$+\frac{1}{6}\sum_{p=1}^{k}\sum_{q=1}^{k}\sum_{r=1}^{k}\left[\frac{\partial^3}{\partial\vartheta_p\partial\vartheta_q\partial\vartheta_r}LL(\tilde{\boldsymbol{\vartheta}}_n|\boldsymbol{\vartheta}^*)\right]\left(\hat{\vartheta}_{n,p} - \vartheta_p^*\right)\left(\hat{\vartheta}_{n,q} - \vartheta_q^*\right)\left(\hat{\vartheta}_{n,r} - \vartheta_r^*\right),$$

Regularity criteria

For the proof of the χ^2 theorem (see text), as well as of Statements 3 and 5 below, the following regularity criteria (Conditions I–VI) are required to hold:

I. $\Omega \subset \mathbf{R}^k (k < \infty)$, and probability distributions defined by any two different $\boldsymbol{\vartheta} \in \Omega$ are distinct.
II. In an open convex neighborhood $S(\boldsymbol{\vartheta}^*) \subset \Omega$, containing the true parameter $\boldsymbol{\vartheta}^*$, there exist the first, second, and third partial derivatives of $f(\boldsymbol{\vartheta},t)$ with respect to $\boldsymbol{\vartheta}$.
III. $E_{\boldsymbol{\vartheta}}(((\partial/\partial\vartheta_j)\ln f(\boldsymbol{\vartheta},t))^2) < \infty$, $j = 1, \ldots, k$, for all $\boldsymbol{\vartheta} \in S(\boldsymbol{\vartheta}^*)$.
IV. $E_{\boldsymbol{\vartheta}}((\partial/\partial\vartheta_j)\ln f(\boldsymbol{\vartheta},t)) = 0$, $j = 1, \ldots, k$, for all $\boldsymbol{\vartheta} \in S(\boldsymbol{\vartheta}^*)$.
V. The components $I_{jl}(\boldsymbol{\vartheta})$ of the information matrix $\mathbf{I}(\boldsymbol{\vartheta}) = \left(I_{jl}(\boldsymbol{\vartheta})\right)_{k\times k}$, defined as $I_{jl}(\boldsymbol{\vartheta}) = E_{\boldsymbol{\vartheta}}\left(-(\partial^2/\partial\vartheta_j\partial\vartheta_l)\ln f(\boldsymbol{\vartheta},t)\right)$, are finite, and $\mathbf{I}(\boldsymbol{\vartheta})$ is a positive-definite for all $\boldsymbol{\vartheta} \in S(\boldsymbol{\vartheta}^*)$.
VI. There exist functions $M_{ijl}(t)$ for which $E_{\boldsymbol{\vartheta}^*}\left(M_{ijl}(t)\right) < \infty$, $i, j, l = 1, \ldots, k$, such that $|(\partial^3/\partial\vartheta_i\partial\vartheta_j\partial\vartheta_l)\ln f(\boldsymbol{\vartheta},t)| \leq M_{ijl}(t)$ for all $\boldsymbol{\vartheta} \in S(\boldsymbol{\vartheta}*)$.

Auxiliary statements

If the regularity criteria (Conditions I–VI) are satisfied, the following statements have been shown to apply (29–31):

Statement 1

$$\hat{\boldsymbol{\vartheta}}_n \xrightarrow{a.e.} \boldsymbol{\vartheta}^*.$$

Statement 2

$$\sqrt{n}\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*\right) \xrightarrow{d} N_k(\mathbf{0}, \mathbf{I}(\boldsymbol{\vartheta}^*)^{-1}).$$

(Normal distribution with $\mathbf{0}$ mean and covariance $\mathbf{I}(\boldsymbol{\vartheta}^*)^{-1}$.) A corollary of Statement 2 is that $\sqrt{n}\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}^*\right)$ is stochastically bounded.

Statements 3, 4, and 5

Statement 3

Suppose that regularity criteria (Conditions I–VI) are satisfied for all $\boldsymbol{\vartheta} \in S(\boldsymbol{\vartheta}^*)$. Then,

$$\frac{1}{\sqrt{n}}\left[LL(\hat{\boldsymbol{\vartheta}}_n|\boldsymbol{\vartheta}^*) - \sum_{i=1}^{n}\ln f(\boldsymbol{\vartheta}^*, t_i)\right] \xrightarrow{P} 0.$$

Proof

Expanding $LL(\hat{\boldsymbol{\vartheta}}_n|\boldsymbol{\vartheta}^*)$ into a Taylor series with respect to $\hat{\boldsymbol{\vartheta}}_n$ around the true parameter $\boldsymbol{\vartheta}^*$, where $\tilde{\boldsymbol{\vartheta}}_n$ is a point on the line connecting $\hat{\boldsymbol{\vartheta}}_n$ and $\boldsymbol{\vartheta}^*$. Because $LL(\boldsymbol{\vartheta}^*|\boldsymbol{\vartheta}^*) = \sum_{i=1}^{n} \ln f(\boldsymbol{\vartheta}^*, t_i)$,

$$\begin{aligned}\frac{1}{\sqrt{n}}\left[LL\left(\hat{\boldsymbol{\vartheta}}_n|\boldsymbol{\vartheta}^*\right) - \sum_{n=1}^{n}\ln f(\boldsymbol{\vartheta}^*,t_i)\right] &= \frac{1}{\sqrt{n}}\left[\frac{\partial}{\partial\boldsymbol{\vartheta}}LL(\boldsymbol{\vartheta}^*|\boldsymbol{\vartheta}^*)\right]\left(\hat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}^*\right) + \frac{1}{2\sqrt{n}}\left(\hat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}^*\right)^{T}\left[\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}}LL(\boldsymbol{\vartheta}^*|\boldsymbol{\vartheta}^*)\right]\left(\hat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}^*\right)\\ &\quad+\frac{1}{6\sqrt{n}}\sum_{p=1}^{k}\sum_{q=1}^{k}\sum_{r=1}^{k}\left[\frac{\partial^3}{\partial\vartheta_p\partial\vartheta_q\partial\vartheta_r}LL(\tilde{\boldsymbol{\vartheta}}_n|\boldsymbol{\vartheta}^*)\right]\left(\hat{\vartheta}_{n,p}-\vartheta_p^*\right)\left(\hat{\vartheta}_{n,q}-\vartheta_q^*\right)\left(\hat{\vartheta}_{n,r}-\vartheta_r^*\right)\\ &= S_1+S_2+S_3,\end{aligned}$$

and

$$\left|\frac{1}{\sqrt{n}}\left[LL\left(\hat{\boldsymbol{\vartheta}}_n|\boldsymbol{\vartheta}^*\right)-\sum_{n=1}^{n}\ln f(\boldsymbol{\vartheta}^*,t_i)\right]\right|\le|S_1|+|S_2|+|S_3|.$$

$$\begin{aligned}S_1&=\frac{1}{n}\left[\frac{\partial}{\partial\boldsymbol{\vartheta}}LL(\boldsymbol{\vartheta}^*|\boldsymbol{\vartheta}^*)\right]\sqrt{n}\left(\hat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}^*\right)\\ &=\frac{1}{n}\left[\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\vartheta}}\ln f(\boldsymbol{\vartheta}^*,t_i)\right]\sqrt{n}\left(\hat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}^*\right).\end{aligned}$$

Because the random variable $(\partial/\partial\boldsymbol{\vartheta})\ln f(\boldsymbol{\vartheta}^*,t)$ has zero mean (by Condition IV) and finite variance (by Condition III), the average $(1/n)\sum_{i=1}^{n}(\partial/\partial\boldsymbol{\vartheta})\ln f(\boldsymbol{\vartheta}^*,t_i)$ converges almost everywhere to zero, whereas $\sqrt{n}(\hat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}^*)$ is stochastically bounded (by Statement 2). Thus, $|S_1|\xrightarrow[n\to\infty]{P}0$.

$$\begin{aligned}S_2&=\frac{1}{2}(\hat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}^*)^{T}\frac{1}{n}\left[\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}}LL(\boldsymbol{\vartheta}^*|\boldsymbol{\vartheta}^*)\right]\sqrt{n}(\hat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}^*)\\ &=\frac{1}{2}(\hat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}^*)^{T}\frac{1}{n}\left[\sum_{i=1}^{n}\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}}\ln f(\boldsymbol{\vartheta}^*,t_i)\right]\sqrt{n}(\hat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}^*).\end{aligned}$$

Because the random matrix $(\partial^2/\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta})\ln f(\boldsymbol{\vartheta}_*,t)$ has finite mean (by Condition V), the average $\frac{1}{n}[\sum_{i=1}^{n}\frac{\partial^2}{\partial\boldsymbol{\vartheta}\partial\boldsymbol{\vartheta}}\ln f(\boldsymbol{\vartheta}_*,t_i)]$ converges in distribution to $-\mathbf{I}(\boldsymbol{\vartheta}^*)$; while $\sqrt{n}(\hat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}*)$ is stochastically bounded (by Statement 2) and $(1/2)(\hat{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}*)^{T}$ converges almost everywhere to zero (by Statement 1). Thus, $|S_2|\xrightarrow[n\to\infty]{P}0$.

$$\begin{aligned}|S_3|&\le\frac{1}{6\sqrt{n}}\sum_{p=1}^{k}\sum_{q=1}^{k}\sum_{r=1}^{k}|\hat{\vartheta}_{n,p}-\vartheta_p^*||\hat{\vartheta}_{n,q}-\vartheta_q^*||\hat{\vartheta}_{n,r}-\vartheta_r^*|\left|\frac{\partial^3}{\partial\vartheta_p\partial\vartheta_q\partial\vartheta_r}LL(\tilde{\boldsymbol{\vartheta}}_n|\boldsymbol{\vartheta}^*)\right|\\ &=\frac{1}{6}\sum_{p=1}^{k}\sum_{q=1}^{k}\sum_{r=1}^{k}(|\hat{\vartheta}_{n,p}-\vartheta_p^*||\hat{\vartheta}_{n,q}-\vartheta_p^*|)\left(\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\partial^3}{\partial\vartheta_p\partial\vartheta_q\partial\vartheta_r}\ln f(\tilde{\boldsymbol{\vartheta}}_n,t_i)\right|\right)(\sqrt{n}|\hat{\vartheta}_{n,r}-\vartheta_r^*|)\\ &\le\frac{1}{6}\sum_{p=1}^{k}\sum_{q=1}^{k}\sum_{r=1}^{k}(|\hat{\vartheta}_{n,p}-\vartheta_p^*||\hat{\vartheta}_{n,q}-\vartheta_q^*|)\left(\frac{1}{n}\sum_{i=1}^{n}M_{pqr}(t_i)\right)(\sqrt{n}|\hat{\vartheta}_{n,r}-\vartheta_r^*|).\end{aligned}$$

Because, by Condition VI, $E_{\boldsymbol{\vartheta}*}(M_{pqr}(t))<\infty$, the average $\frac{1}{n}\sum_{i=1}^{n}M_{pqr}(t_i)\xrightarrow[n\to\infty]{P}E_{\boldsymbol{\vartheta}*}(M_{pqr}(t))<\infty$; thus, both $\frac{1}{n}\sum_{i=1}^{n}M_{pqr}(t_i)$ and $|\sqrt{n}(\hat{\vartheta}_{n,r}-\vartheta_r^*)|$ are stochastically bounded, whereas $|\hat{\vartheta}_{n,p}-\vartheta_p^*||\hat{\vartheta}_{n,q}-\vartheta_q^*|$ converges almost everywhere to zero. Thus, $|S_3|\xrightarrow[n\to\infty]{P}0$, which concludes the proof.

Notes

Because, by the central limit theorem, $\sum_{i=1}^{n}\ln f(\boldsymbol{\vartheta}^*,t_i)$ is asymptotically normally distributed with mean and variance given by Eqs. 5 and 6, Statement 3 implies the same for $LL(\hat{\boldsymbol{\vartheta}}_n|\boldsymbol{\vartheta}^*)$.

The online Supplementary Material contains the demonstration that Conditions I–VI apply to the closed-time distributions of Schemes 1 and 2 (and 4), and 1_R (and 2_R, 4_R).

Statement 4

Let $\Omega\subset\mathbf{R}^k$, and let Ω_R be a closed subset of Ω such that $\boldsymbol{\vartheta}^*\notin\Omega_R$. Let $p{:}\Phi\to\Omega_R$ be a parameterization of Ω_R, where Φ is a closed subset of $\mathbf{R}^c$ $(c\le k)$. Let $\hat{\boldsymbol{\varphi}}_n\in\Phi$ denote the ML estimate based on $t_1,t_2,\ldots t_n$ restricted to Φ; i.e., $LL(p(\hat{\boldsymbol{\varphi}}_n)|\boldsymbol{\vartheta}^*)=\sup_{\boldsymbol{\varphi}\in\Phi}LL(p(\boldsymbol{\varphi})|\boldsymbol{\vartheta}^*)$; and let $\hat{\boldsymbol{\vartheta}}_{R,n}$ denote $p\left(\hat{\boldsymbol{\varphi}}_n\right)$.

Suppose that the following conditions apply:

VII. There exists a unique point $\boldsymbol{\varphi}^*\in\Phi$, such that $\sqrt{n}(\hat{\boldsymbol{\varphi}}_n-\boldsymbol{\varphi}^*)$ is stochastically bounded. (Let $\boldsymbol{\vartheta}_R^*$ denote $p(\boldsymbol{\varphi}^*)$. Note that Condition VII implies $\hat{\boldsymbol{\varphi}}_n\xrightarrow{P}\boldsymbol{\varphi}^*$.)

VIII. In an open convex neighborhood $G(\boldsymbol{\varphi}^*)\subset\Phi$, containing $\boldsymbol{\varphi}^*$, there exist continuous first partial derivatives of $f(p(\boldsymbol{\varphi}),t)$ with respect to $\boldsymbol{\varphi}$.

IX. $E_{\boldsymbol{\vartheta}*}(\frac{\partial}{\partial\varphi_j}\ln f(p(\boldsymbol{\varphi}*),t))=0$, $\varphi=1,\ldots,c$.

X. There exist functions $g:\mathbf{R}\to\mathbf{R}_0^+$ and $h:\mathbf{R}^c\to\mathbf{R}_0^+$ continuous, $E_{\boldsymbol{\vartheta}*}(g(t))<\infty$, $h(\mathbf{0})=0$, such that $|(\partial/\partial_{\varphi j})\ln f(p(\boldsymbol{\varphi}),t)-(\partial/\partial_{\varphi j})\ln f(p(\boldsymbol{\varphi}^*),t)|\le g(t)\cdot h(\boldsymbol{\varphi}-\boldsymbol{\varphi}^*)$ for $i,j=1,\ldots,k$, for all $\boldsymbol{\varphi}\in G(\boldsymbol{\varphi}^*)$.

Then, $\frac{1}{\sqrt{n}}\left[LL\left(\hat{\boldsymbol{\vartheta}}_{R,n}|\boldsymbol{\vartheta}^*\right)-\sum_{i=1}^{n}\ln f(\boldsymbol{\vartheta}_R^*t_i)\right]\xrightarrow[n\to\infty]{P}0$.

Proof

Differentiating $LL\left(\hat{\boldsymbol{\vartheta}}_{R,n}|\boldsymbol{\vartheta}^*\right)=LL\left(p(\hat{\boldsymbol{\varphi}}_n)|\boldsymbol{\vartheta}^*\right)$ with respect to $\hat{\boldsymbol{\varphi}}_n$ around the parameter $\boldsymbol{\varphi}^*$,

$$LL(\hat{\boldsymbol{\vartheta}}_{R,n}|\boldsymbol{\vartheta}^*)=LL(p(\boldsymbol{\varphi}^*)|\boldsymbol{\vartheta}^*)+\left[\frac{\partial}{\partial\boldsymbol{\varphi}}LL(p(\tilde{\boldsymbol{\varphi}}_n)|\boldsymbol{\vartheta}^*)\right](\hat{\boldsymbol{\varphi}}_n-\boldsymbol{\varphi}^*),$$

where $\tilde{\boldsymbol{\varphi}}_n$ is a point on the line connecting $\hat{\boldsymbol{\varphi}}_n$ and $\boldsymbol{\varphi}^*$. Thus,

$$\begin{aligned}LL(\hat{\boldsymbol{\vartheta}}_{R,n}|\boldsymbol{\vartheta}^*)&=\sum_{i=1}^{n}\ln f(p(\boldsymbol{\varphi}^*),t_i)\\ &\quad+\left[\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\varphi}}\ln f(p(\tilde{\boldsymbol{\varphi}}_n),t_i)\right](\hat{\boldsymbol{\varphi}}_n-\boldsymbol{\varphi}^*).\end{aligned}$$

Therefore,

$$\frac{1}{\sqrt{n}}\left[LL(\hat{\boldsymbol{\vartheta}}_{R,n}|\boldsymbol{\vartheta}^*)-\sum_{i=1}^{n}\ln f(\boldsymbol{\vartheta}_R^*,t_i)\right]=\frac{1}{\sqrt{n}}\left[\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\varphi}}\ln f(p(\tilde{\boldsymbol{\varphi}}_n),t_i)\right](\hat{\boldsymbol{\varphi}}_n-\boldsymbol{\varphi}^*)=\frac{1}{\sqrt{n}}\left[\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\varphi}}\ln f(p(\boldsymbol{\varphi}^*),t_i)\right.$$
$$\left.+\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\varphi}}\ln f(p(\tilde{\boldsymbol{\varphi}}_n),t_i)-\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\varphi}}\ln f(p(\boldsymbol{\varphi}^*),t_i)\right](\hat{\boldsymbol{\varphi}}_n-\boldsymbol{\varphi}^*)=\frac{1}{n}\left[\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\varphi}}\ln f(p(\boldsymbol{\varphi}^*),t_i)\right]\sqrt{n}(\hat{\boldsymbol{\varphi}}_n-\boldsymbol{\varphi}^*)$$
$$+\frac{1}{n}\left[\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\varphi}}\ln f(p(\tilde{\boldsymbol{\varphi}}_n),t_i)-\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\varphi}}\ln f(p(\boldsymbol{\varphi}^*),t_i)\right]\sqrt{n}(\hat{\boldsymbol{\varphi}}_n-\boldsymbol{\varphi}^*)=S_1+S_2.$$

Thus,

$$\left|\frac{1}{\sqrt{n}}\left[LL(\hat{\boldsymbol{\vartheta}}_{R,n}|\boldsymbol{\vartheta}^*)-\sum_{i=1}^{n}\ln f(\boldsymbol{\vartheta}_R^*,t_i)\right]\right|\le|S_1|+|S_2|.$$

Because the random variable $\frac{\partial}{\partial\boldsymbol{\varphi}}\ln f(p(\boldsymbol{\varphi}^*),t)$ has zero mean (by Condition IX), the average $\frac{1}{n}\sum_{i=1}^{n}\frac{\partial}{\partial\boldsymbol{\varphi}}\ln f(p(\boldsymbol{\varphi}^*),t_i)$ converges to zero in probability, whereas $\sqrt{n}(\hat{\boldsymbol{\varphi}}_n-\boldsymbol{\varphi}^*)$ is stochastically bounded (by Condition VII). Thus, $|S_1|\xrightarrow[n\to\infty]{P}0$.

On the other hand, applying Condition X,

$$|S_2|=\left|\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k}\left[\left(\frac{\partial}{\partial\varphi_j}\ln f(p(\tilde{\boldsymbol{\varphi}}_n),t_i)-\frac{\partial}{\partial\varphi_j}\ln f(p(\boldsymbol{\varphi}^*),t_i)\right)\sqrt{n}(\hat{\varphi}_{n,j}-\varphi_j^*)\right]\right|$$
$$\le\sum_{j=1}^{k}\frac{1}{n}\sum_{i=1}^{n}\left|\left(\frac{\partial}{\partial\varphi_j}\ln f(p(\tilde{\boldsymbol{\varphi}}_n),t_i)-\frac{\partial}{\partial\varphi_j}\ln f(p(\boldsymbol{\varphi}^*),t_i)\right)\right||\sqrt{n}(\hat{\varphi}_{n,j}-\varphi_j^*)|$$
$$\le\sum_{j=1}^{k}\frac{1}{n}\sum_{i=1}^{n}h(\tilde{\boldsymbol{\varphi}}_n-\boldsymbol{\varphi}^*)g(t_i)|\sqrt{n}(\hat{\varphi}_{n,j}-\varphi_j^*)|=h(\tilde{\boldsymbol{\varphi}}_n-\boldsymbol{\varphi}^*)\left(\frac{1}{n}\sum_{i=1}^{n}g(t_i)\right)\sum_{j=1}^{k}|\sqrt{n}(\hat{\varphi}_{n,j}-\varphi_j^*)|.$$

Because $E_{\boldsymbol{\vartheta}^*}(g(t))<\infty$, $\frac{1}{n}\sum_{i=1}^{n}g(t_i)\xrightarrow[n\to\infty]{P}E_{\boldsymbol{\vartheta}^*}(g(t))<\infty$; thus, both $\frac{1}{n}\sum_{i=1}^{n}g(t_i)$ and $\sum_{j=1}^{k}|\sqrt{n}(\hat{\varphi}_{n,j}-\varphi_j^*)|$ are stochastically bounded. Because $\tilde{\boldsymbol{\varphi}}_n\xrightarrow{P}\boldsymbol{\varphi}^*$ (by Condition VII), $h(\tilde{\boldsymbol{\varphi}}_n-\boldsymbol{\varphi}^*)\xrightarrow{P}0$. Thus, $|S_2|\xrightarrow[n\to\infty]{P}0$, which concludes the proof.

Notes

Because, by the central limit theorem, $\sum_{i=1}^{n}\ln f(\boldsymbol{\vartheta}_R^*,t_i)$ is asymptotically Gaussian distributed with mean and variance given by Eqs. 7 and 8, Statement 4 implies the same for $LL(\hat{\boldsymbol{\vartheta}}_{R,n}|\boldsymbol{\vartheta}^*)$.

It is easy to show that Conditions VII–X apply if the reduced scheme predicts a single-exponential distribution (such as Schemes 1_R, 2_R, and 4_R). The corresponding parameterization p will be $\varphi\mapsto(0,\varphi)$ for the case of Scheme 1 versus 1_R, and $\varphi\mapsto(\varphi,\varphi,a_{fixed})$ for the case of Scheme 2 versus 2_R, with Ω_R chosen as in Fig. 3 *F*.

Check VII. If the set of data $t_1,t_2,\ldots t_n$, regardless of its true distribution, is fitted to $f(t)=(1/\varphi)e^{-t/\varphi}$ by maximum likelihood, then $\hat{\varphi}_n=\frac{1}{n}\sum_{i=1}^{n}t_i=\langle t_i\rangle$ (2). By the strong law of large numbers (as the random variables t_i have finite variance), $\hat{\varphi}_n=\langle t_i\rangle$ converges almost everywhere to the mean $E(t)$ of the true distribution, thus $\varphi^*=E(t)$. By the central limit theorem, $\sqrt{n}(\hat{\varphi}_n-\varphi^*)=\sqrt{n}(\langle t_i\rangle-E(t))\xrightarrow{d}N(0,1)$ and is thus stochastically bounded.

Check VIII. $(\partial/\partial\varphi)f(p(\varphi),t)=(1/\varphi^2)e^{-t/\varphi}(t/\varphi-1)$, which exists for all $\varphi>0$.

Check IX. $(\partial/\partial\varphi)\ln f(p(\varphi),t)=-1/\varphi+t/\varphi^2$, therefore,

$$E_{\boldsymbol{\vartheta}^*}\left(\frac{\partial}{\partial\varphi}\ln f(p(\varphi^*),t)\right)=\int_0^{\infty}f(\boldsymbol{\vartheta}^*,t)\left(-\frac{1}{\varphi^*}+\frac{t}{\varphi^{*2}}\right)dt$$
$$=-\frac{1}{\varphi^*}+\frac{1}{\varphi^{*2}}\int_0^{\infty}tf(\boldsymbol{\vartheta}^*,t)dt=-\frac{1}{\varphi^*}+\frac{E(t)}{\varphi^{*2}}=0.$$

Check X. Let $\varepsilon<\varphi^*$, define $G(\varphi^*)$ as the open interval $(\varphi^*-\varepsilon,\varphi^*+\varepsilon)$. Then, for all $\varphi\in G(\varphi^*)$,

$$\left|\frac{\partial}{\partial\varphi}\ln f(p(\varphi),t)-\frac{\partial}{\partial\varphi}\ln f(p(\varphi^*),t)\right|=\left|\left(-\frac{1}{\hat{\varphi}}+\frac{t}{\hat{\varphi}^2}\right)\right.$$
$$\left.-\left(-\frac{1}{\varphi^*}+\frac{t}{\varphi^{*2}}\right)\right|\le\left|\frac{1}{\varphi^*}-\frac{1}{\hat{\varphi}}\right|+t\left|\frac{1}{\hat{\varphi}^2}-\frac{1}{\varphi^{*2}}\right|$$
$$\le|\hat{\varphi}-\varphi^*|\left(\frac{1}{\varphi^*(\varphi^*-\varepsilon)}+t\frac{2\varphi^*+\varepsilon}{\varphi^{*2}(\varphi^*-\varepsilon)^2}\right).$$

Thus, $h(x)=|x|$, and $g(t)$ has the form $g(t)=a+bt$. Trivially, $E_{\boldsymbol{\vartheta}*}(g(t))=a+bE(t)=3/(\varphi^*+\varepsilon)^2<\infty$.

If the reduced model still contains more than one exponential component, the demonstration of Condition VII is less straightforward. However, simulations using Scheme 3, followed by ML-fits of the closed-time distribution to Scheme 3_R, have confirmed that even if a five-component distribution is fitted with a four-component pdf the constrained fits converge to a unique point in Ω_R with increasing events list length. The set of parameters printed in Fig. 1 for Scheme 3_R corresponds to this experimentally determined $\boldsymbol{\varphi}^*$.

Statement 5

Let Ω_R be a closed subset of Ω such that $\boldsymbol{\vartheta}^*\notin\Omega_R$. Suppose that Conditions I–VI are satisfied on $S(\boldsymbol{\vartheta}^*)$, and Conditions VII–X are satisfied on $G(\boldsymbol{\varphi}^*)$. Then,

$$\frac{1}{\sqrt{n}}\left[\Delta LL(\hat{\boldsymbol{\vartheta}}_{R,n},\hat{\boldsymbol{\vartheta}}_n|\boldsymbol{\vartheta}^*)-\sum_{i=1}^{n}\ln(f(\boldsymbol{\vartheta}^*,t_i)/f(\boldsymbol{\vartheta}_R^*,t_i))\right]\xrightarrow[n\to\infty]{P}0.$$

Proof

$$\frac{1}{\sqrt{n}}\left[\Delta LL(\hat{\boldsymbol{\vartheta}}_{\mathrm{R,n}},\hat{\boldsymbol{\vartheta}}_{\mathrm{n}}|\boldsymbol{\vartheta}^*)-\sum_{\mathrm{i=1}}^{\mathrm{n}}\ln\big(f(\boldsymbol{\vartheta}^*,t_{\mathrm{i}})/f(\boldsymbol{\vartheta}^*_{\mathrm{R}},t_{\mathrm{i}})\big)\right]$$
$$=\frac{1}{\sqrt{n}}\left[(LL(\hat{\boldsymbol{\vartheta}}_{\mathrm{n}}|\boldsymbol{\vartheta}^*)-LL(\hat{\boldsymbol{\vartheta}}_{\mathrm{R,n}}|\boldsymbol{\vartheta}^*))-\left(\sum_{\mathrm{i=1}}^{\mathrm{n}}\ln(f(\boldsymbol{\vartheta}^*,t_{\mathrm{i}}))-\sum_{\mathrm{i=1}}^{\mathrm{n}}\ln(f(\boldsymbol{\vartheta}^*_{\mathrm{R}},t_{\mathrm{i}}))\right)\right]$$
$$=\frac{1}{\sqrt{n}}\left[LL(\hat{\boldsymbol{\vartheta}}_{\mathrm{n}}|\boldsymbol{\vartheta}^*)-\sum_{\mathrm{i=1}}^{\mathrm{n}}\ln(f(\boldsymbol{\vartheta}^*,t_{\mathrm{i}}))\right]-\frac{1}{\sqrt{n}}\left[LL(\hat{\boldsymbol{\vartheta}}_{\mathrm{R,n}}|\boldsymbol{\vartheta}^*)-\sum_{\mathrm{i=1}}^{\mathrm{n}}\ln(f(\boldsymbol{\vartheta}^*_{\mathrm{R}},t_{\mathrm{i}}))\right]=S_3-S_4.$$

As $S_3 \xrightarrow[n\to\infty]{P} 0$ by Statement 3, and $S_4 \xrightarrow[n\to\infty]{P} 0$ by Statement 4, this concludes the proof.

Notes

Because, by the central limit theorem, $\sum_{\mathrm{i=1}}^{\mathrm{n}}\ln\big(f(\boldsymbol{\vartheta}^*,t_{\mathrm{i}})/f(\boldsymbol{\vartheta}^*_{\mathrm{R}},t_{\mathrm{i}})\big)$ is asymptotically Gaussian-distributed with mean and variance given by Eqs. 9 and 10, Statement 5 implies the same for $\Delta LL(\hat{\boldsymbol{\vartheta}}_{\mathrm{R,n}},\hat{\boldsymbol{\vartheta}}_{\mathrm{n}}|\boldsymbol{\vartheta}*)$.

Statement 5 contains meaningful information only if $\boldsymbol{\vartheta}^*\notin\Omega_{\mathrm{R}}$. If $\boldsymbol{\vartheta}^*\in\Omega_{\mathrm{R}}$ (i.e., if the restricted model is the correct model), then $\boldsymbol{\vartheta}^*_{\mathrm{R}}=\boldsymbol{\vartheta}^*$, $f(\boldsymbol{\vartheta}^*,t_{\mathrm{i}})=f(\boldsymbol{\vartheta}^*_{R},t_{\mathrm{i}})$, and $\ln\big(f(\boldsymbol{\vartheta}^*,t_{\mathrm{i}})/f(\boldsymbol{\vartheta}^*_{\mathrm{R}},t_{\mathrm{i}})\big)=0$. Therefore, in this case, Statement 5 reduces to the trivial statement $(1/\sqrt{n})\Delta LL(\hat{\boldsymbol{\vartheta}}_{\mathrm{R,n}},\hat{\boldsymbol{\vartheta}}_{\mathrm{n}}|\boldsymbol{\vartheta}^*_{\mathrm{R}}) \xrightarrow[n\to\infty]{P} 0$. (This statement is obviously true as, by the χ^2-theorem, $\Delta LL(\hat{\boldsymbol{\vartheta}}_{\mathrm{R,n}},\hat{\boldsymbol{\vartheta}}_{\mathrm{n}}|\boldsymbol{\vartheta}^*_{\mathrm{R}}) \xrightarrow{d} \Gamma(k/2,\,1)$, and is thus stochastically bounded.)

APPENDIX B: NUMERICAL IMPLEMENTATION OF EQS. 1–10

In practice, the integrals in Eqs. 1–10 are performed numerically between $t=0$ (or $t=t_{\mathrm{low}}$) and some $t=t_{\mathrm{high}}$, where t_{high} is chosen such that the integrand vanishes for $t>t_{\mathrm{high}}$. Because the integrand has the form $f_{\boldsymbol{\vartheta}_1}(t)\cdot\ln(f_{\boldsymbol{\vartheta}_1}(t)/f_{\boldsymbol{\vartheta}_2}(t))$ or $f_{\boldsymbol{\vartheta}_1}(t)\cdot\ln^2(f_{\boldsymbol{\vartheta}_1}(t)/f_{\boldsymbol{\vartheta}_2}(t))$, a choice of t_{high} could be, e.g., 100 times the longest time constant of $f_{\boldsymbol{\vartheta}_1}(t)$. However, because the longest time constants of the two pdfs might be very different, evaluation of one of the two pdfs might result in an underflow at high t, i.e., return zero; thus causing the algorithm to crash when attempting to take the logarithm. To avoid this problem the following simple strategy can be used. Both pdfs involved have the form $f(t)=(\sum_{\mathrm{j=1}}^{\mathrm{k}}(a_{\mathrm{j}}/\tau_{\mathrm{j}})e^{-t/\tau_{\mathrm{j}}})/P(t\geq t_{\mathrm{low}})$, which can be rewritten in the form $f(t)=\sum_{\mathrm{j=1}}^{\mathrm{k}}\mathrm{sign}_{\mathrm{j}}(|a_{\mathrm{j}}|/\tau_{\mathrm{j}})e^{-t/\tau_{\mathrm{j}}}/P(t\geq t_{\mathrm{low}})$ to allow for possible negative components such as in the closed-time pdf of Scheme 1. Define $y_{\mathrm{j}}=\ln(|a_{\mathrm{j}}|/\tau_{\mathrm{j}})-t/\tau_{\mathrm{j}}$, $y_{\max}=\max\{y_1,\ldots,y_{\mathrm{k}}\}$, and $z_{\mathrm{low}}=\ln(P(t\geq t_{\mathrm{low}}))$. Then ln $f(t)$ can be rewritten in the form of $\ln f(t)=y_{\max}+\ln(\sum_{\mathrm{j=1}}^{\mathrm{k}}(\mathrm{sign}_{\mathrm{j}}\cdot e^{y_{\mathrm{j}}-y_{\max}}))-z_{\mathrm{low}}$. Because the term j corresponding to $y_{\max}$ always has a positive sign, the sum in the argument of the ln function will contain a term equal to $+1$, and thus will not underflow. Finally, $\ln\big(f_{\boldsymbol{\vartheta}*}(t)/f_{\boldsymbol{\vartheta}^*_{\mathrm{R}}}(t)\big)$ can be calculated as $\ln f_{\boldsymbol{\vartheta}*}(t)-\ln f_{\boldsymbol{\vartheta}^*_{\mathrm{R}}}(t)$.

SUPPLEMENTARY MATERIAL

An online supplement to this article can be found by visiting BJ Online at http://www.biophysj.org. Sections 1–3 of this Supplementary Material contain the demonstration that the regularity criteria described in Appendix A (Conditions I–VI) are satisfied in the interior of the parameter space Ω for the closed-time distributions of the following schemes shown in Fig. 1: Schemes 1_{R}, 2_{R}, and 4_{R} (Section 1), Scheme 1 (Section 2), and Schemes 2 and 4 (Section 3).

I am grateful to Drs. David C. Gadsby and Paola Vergani for discussions and for many valuable comments on the manuscript, as well as to the referee who brought the body of work referenced in Chernoff (40), Kudo (41), Nuesch (42), Shapiro (43), and Self and Liang (44) to my attention.

Supported by National Institutes of Health Fogarty International Center grant No. R03-TW05761.

REFERENCES

1. 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.
2. Colquhoun, D., and F. J. Sigworth. 1995. Fitting and statistical analysis of single-channel records. *In* Single Channel Recording. B. Sakmann and E. Neher, editors. Plenum Press, New York.
3. Hamill, O. P., A. Marty, E. Neher, B. Sakmann, and F. J. Sigworth. 1981. Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflugers Arch.* 391:85–100.
4. Sigworth, F. J., and S. M. Sine. 1987. Data transformations for improved display and fitting of single-channel dwell time histograms. *Biophys. J.* 52:1047–1054.
5. Rothberg, B. S., and K. L. Magleby. 1998. Kinetic structure of large-conductance Ca^{2+}-activated K^+ channels suggests that the gating includes transitions through intermediate or secondary states. A mechanism for flickers. *J. Gen. Physiol.* 111:751–780.
6. Horn, R., and K. Lange. 1983. Estimating kinetic constants from single channel data. *Biophys. J.* 43:207–223.
7. Ball, F. G., and M. S. Sansom. 1989. Ion-channel gating mechanisms: model identification and parameter estimation from single channel recordings. *Proc. R. Soc. Lond. B Biol. Sci.* 236:385–416.
8. Qin, F., A. Auerbach, and F. Sachs. 1996. Estimating single-channel kinetic parameters from idealized patch-clamp data containing missed events. *Biophys. J.* 70:264–280.
9. Fredkin, D. R., and J. A. Rice. 1992. Maximum likelihood estimation and identification directly from single-channel recordings. *Proc. Biol. Sci.* 249:125–132.
10. Colquhoun, D., A. G. Hawkes, and K. Srodsinski. 1996. Joint distributions of apparent open times and shut times of single ion channels and the maximum likelihood fitting of mechanisms. *Philos. Trans. R. Soc. Lond. A.* 354:2555–2590.
11. Colquhoun, D., C. J. Hatton, and A. G. Hawkes. 2003. The quality of maximum likelihood estimates of ion channel rate constants. *J. Physiol.* 547:699–728.
12. Rothberg, B. S., and K. L. Magleby. 1999. Gating kinetics of single large-conductance Ca^{2+}-activated K^+ channels in high Ca^{2+} suggest a two-tiered allosteric gating mechanism. *J. Gen. Physiol.* 114:93–124.
13. Rothberg, B. S., and K. L. Magleby. 2000. Voltage and Ca^{2+} activation of single large-conductance Ca^{2+}-activated K^+ channels described by a two-tiered allosteric gating mechanism. *J. Gen. Physiol.* 116:75–99.

14. Beato, M., P. J. Groot-Kormelink, D. Colquhoun, and L. G. Sivilotti. 2004. The activation mechanism of α1 homomeric glycine receptors. *J. Neurosci.* 24:895–906.

15. Burzomato, V., M. Beato, P. J. Groot-Kormelink, D. Colquhoun, and L. G. Sivilotti. 2004. Single-channel behavior of heteromeric $\alpha 1\beta$-glycine receptors: an attempt to detect a conformational change before the channel opens. *J. Neurosci.* 24:10924–10940.

16. Weiss, D. S., and K. L. Magleby. 1989. Gating scheme for single GABA-activated Cl^- channels determined from stability plots, dwell-time distributions, and adjacent-interval durations. *J. Neurosci.* 9:1314–1324.

17. Jackson, M. B., B. S. Wong, C. E. Morris, H. Lecar, and C. N. Christian. 1983. Successive openings of the same acetylcholine receptor channel are correlated in open time. *Biophys. J.* 42:109–114.

18. Hatton, C. J., C. Shelley, M. Brydson, D. Beeson, and D. Colquhoun. 2003. Properties of the human muscle nicotinic receptor, and of the slow-channel myasthenic syndrome mutant ε-L221F, inferred from maximum likelihood fits. *J. Physiol.* 547:729–760.

19. Shelley, C., and D. Colquhoun. 2005. A human congenital myasthenia-causing mutation (ε-L78P) of the muscle nicotinic acetylcholine receptor with unusual single channel properties. *J. Physiol.* 564:377–396.

20. Schorge, S., S. Elenes, and D. Colquhoun. 2005. Maximum likelihood fitting of single channel NMDA activity with a mechanism composed of independent dimers of subunits. *J. Physiol.* 569:395–418.

21. Csanády, L., K. W. Chan, D. Seto-Young, D. C. Kopsco, A. C. Nairn, and D. C. Gadsby. 2000. Severed channels probe regulation of gating of cystic fibrosis transmembrane conductance regulator by its cytoplasmic domains. *J. Gen. Physiol.* 116:477–500.

22. Carson, M. R., S. M. Travis, and M. J. Welsh. 1995. The two nucleotide-binding domains of cystic fibrosis transmembrane conductance regulator (CFTR) have distinct functions in controlling channel activity. *J. Biol. Chem.* 270:1711–1717.

23. Lin, Y. F., Y. N. Jan, and L. Y. Jan. 2000. Regulation of ATP-sensitive potassium channel function by protein kinase A-mediated phosphorylation in transfected HEK293 cells. *EMBO J.* 19:942–955.

24. Enkvetchakul, D., G. Loussouarn, E. Makhina, S. L. Shyng, and C. G. Nichols. 2000. The kinetic and physical basis of K(ATP) channel gating: toward a unified molecular understanding. *Biophys. J.* 78:2334–2348.

25. Yang, K., K. Fang, L. Fromondi, and K. W. Chan. 2005. Low temperature completely rescues the function of two misfolded K ATP channel disease-mutants. *FEBS Lett.* 579:4113–4118.

26. Csanády, L., and V. Adam-Vizi. 2003. Ca^{2+}- and voltage-dependent gating of Ca^{2+}- and ATP-sensitive cationic channels in brain capillary endothelium. *Biophys. J.* 85:313–327.

27. Mak, D. O., and J. K. Foskett. 1997. Single-channel kinetics, inactivation, and spatial distribution of inositol trisphosphate (IP_3) receptors in *Xenopus oocyte* nucleus. *J. Gen. Physiol.* 109:571–587.

28. Wald, A. 1943. Test of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Am. Math. Soc.* 54:426–482.

29. Rao, C. R. 1973. Linear Statistical Inference and its Applications. John Wiley & Sons, New York.

30. Cox, D. R., and D. V. Hinkley. 1973. Theoretical Statistics. Chapman & Hall, New York.

31. Móri, F. T., and J. G. Székely. 1986. The maximum-likelihood method. *In* Multivariate Statistical Analysis. Mûszaki Könyvkiadó, Budapest, Hungary. 297–317.

32. Horn, R. 1987. Statistical methods for model discrimination. Applications to gating kinetics and permeation of the acetylcholine receptor channel. *Biophys. J.* 51:255–263.

33. Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Autom. Control.* AC19:716–723.

34. Schwarz, G. 1978. Estimating the dimension of a model. *Ann. Stat.* 6:461–464.

35. Blunck, R., U. Kirst, T. Riessner, and U. Hansen. 1998. How powerful is the dwell-time analysis of multichannel records? *J. Membr. Biol.* 165:19–35.

36. Press, W. H., W. T. Vetterling, S. A. Teukolsky, and B. P. Flannery. 1992. Numerical Recipes in C. The Art of Scientific Computing. Cambridge University Press, Cambridge, UK.

37. L'Ecuyer, P. 1988. Efficient and portable combined random number generators. *Commun. ACM.* 31:742–749.

38. Knuth, D. E. 1981. Seminumerical algorithms. *In* The Art of Computer Programming. Addison-Wesley, Reading, MA.

39. Caceci, M. S., and W. P. Cacheris. 1984. Fitting curves to data. The simplex algorithm is the answer. *Byte* May:340–348.

40. Chernoff, H. 1954. On the distribution of the likelihood ratio. *Annu. Math. Stat.* 25:573–578.

41. Kudo, A. 1963. A multivariate analogue of the one-sided test. *Biometrika.* 50:403–418.

42. Nuesch, P. E. 1966. On the problem of testing location in multivariate populations for restricted alternatives. *Annu. Math. Stat.* 37:113–119.

43. Shapiro, A. 1985. Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika.* 72:133–144.

44. Self, S. G., and K. Y. Liang. 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Am. Stat. Assoc.* 82:605–610.

45. McManus, O. B., and K. L. Magleby. 1989. Kinetic time constants independent of previous single-channel activity suggest Markov gating for a large conductance Ca-activated K channel. *J. Gen. Physiol.* 94:1037–1070.

46. Neelands, T. R., L. J. Greenfield, Jr., J. Zhang, R. S. Turner, and R. L. Macdonald. 1998. GABAA receptor pharmacology and subtype mRNA expression in human neuronal NT2-N cells. *J. Neurosci.* 18: 4993–5007.

47. Obejero-Paz, C. A., S. W. Jones, and A. Scarpa. 1998. Multiple channels mediate calcium leakage in the A7r5 smooth muscle-derived cell line. *Biophys. J.* 75:1271–1286.

48. Obejero-Paz, C. A., M. Auslender, and A. Scarpa. 1998. PKC activity modulates availability and long openings of L-type Ca^{2+} channels in A7r5 cells. *Am. J. Physiol.* 275:C535–C543.

49. Li, J., and H. A. Lester. 1999. Single-channel kinetics of the rat olfactory cyclic nucleotide-gated channel expressed in *Xenopus* oocytes. *Mol. Pharmacol.* 55:883–893.

50. Haas, K. F., and R. L. Macdonald. 1999. GABAA receptor subunit γ-2 and δ-subtypes confer unique kinetic properties on recombinant GABAA receptor currents in mouse fibroblasts. *J. Physiol.* 514:27–45.

51. Hallermann, S., S. Heckmann, J. Dudel, and M. Heckmann. 2005. Short openings in high resolution single channel recordings of mouse nicotinic receptors. *J. Physiol.* 563:645–662.

52. Celentano, J. J., and A. G. Hawkes. 2004. Use of the covariance matrix in directly fitting kinetic parameters: application to GABAA receptors. *Biophys. J.* 87:276–294.

53. Horn, R., and C. A. Vandenberg. 1984. Statistical properties of single sodium channels. *J. Gen. Physiol.* 84:505–534.

54. Korn, S. J., and R. Horn. 1988. Statistical discrimination of fractal and Markov models of single-channel gating. *Biophys. J.* 54:871–877.

55. Ding, S., and F. Sachs. 1999. Single channel properties of P2X2 purinoceptors. *J. Gen. Physiol.* 113:695–720.

56. Michels, G., F. Er, I. Khan, M. Sudkamp, S. Herzig, and U. C. Hoppe. 2005. Single-channel properties support a potential contribution of hyperpolarization-activated cyclic nucleotide-gated channels and If to cardiac arrhythmias. *Circulation.* 111:399–404.

57. Milescu, L. S., G. Akk, and F. Sachs. 2005. Maximum likelihood estimation of ion channel kinetics from macroscopic currents. *Biophys. J.* 88:2494–2515.