Biophys J BioFAST, published on May 5, 2006 as doi:10.1529/biophysj.105.079541

Maximum likelihood estimation of molecular motor kinetics from staircase dwell time sequences

Lorin S. Milescu[1*], Ahmet Yildiz[2**], Paul R. Selvin[2,3], and Frederick Sachs[1]

Running title: Molecular motor kinetic analysis
Keywords: single molecule, Markov model, staircase, modeling, optimization, algorithm

[1]Department of Physiology and Biophysics, State University of New York at Buffalo, Buffalo, NY, USA.
[2]Physics Dept. and [3]Biophysics Center, 1110 W. Green St., University of Illinois at Urbana-Champaign, Urbana, Illinois, USA.
*Present address: National Institute for Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA
**Present address: Univ. of California, San Francisco, Genentech Hall, 600 16th Street CA San Francisco, CA 94107,

Abstract

Molecular motors, such as kinesin, myosin, or dynein, convert chemical energy into mechanical energy by hydrolyzing ATP. The mechanical energy is used for moving in discrete steps along the cytoskeleton and carrying a molecular load (1). High resolution single molecule recordings of motor steps appear as a stochastic sequence of dwells, resembling a staircase (2-4). Staircase data can also be obtained from other molecular machines such as F_1-ATPase (5, 6), RNA polymerase (7), or topoisomerase (8). We developed a maximum likelihood algorithm that estimates the rate constants between different conformational states of the protein, including motor steps. We model the motor with a periodic Markov model that reflects the repetitive chemistry of the motor step. The kinetics are estimated from the idealized dwell sequence (9) by numerical maximization of the likelihood function for discrete-time Markov models. This approach eliminates the need for missed event correction. The algorithm can fit kinetic models of arbitrary complexity, such as uniform or alternating step chemistry, reversible or irreversible kinetics, ATP concentration and mechanical force-dependent rates, etc. The method allows global fitting across stationary and nonstationary experimental conditions, and user-defined *a priori* constraints on rate constants. The algorithm was tested with simulated data, and implemented in the free QuB software (10).

Introduction

Processive motor proteins, such as kinesin, myosin, or dynein, convert chemical energy into mechanical energy by hydrolyzing ATP. The mechanical energy is used for moving in discrete steps along the cytoskeleton and carrying a molecular load (1). The mechano-chemistry of molecular motors is a repetitive chain of identical reaction units (11). Each unit corresponds to all the conformations – including ATP binding states – taken by the protein at each position. The reaction units in the chain are connected by motor step transitions, corresponding to forward or backward steps taken by the motor along its track. Myosin V, for example, is a dimeric motor protein walking with a hand-over-hand mechanism along actin filaments, taking a 37 nm step per ATP hydrolyzed. In the hand-over-hand motion, the motor alternately moves its heads to walk: first, the rear head moves 74 nm and becomes the leading head; next, the other head moves 74 nm and takes the lead position, and so on (12, 2) (Fig 1A).

The motor movement can be visualized by attaching a fluorescent probe to the molecule. Single molecule recordings of motor steps appear as a stochastic sequence of dwells (2-4) resembling a staircase (Fig. 1C). Each dwell corresponds to a defined position of a single motor. The duration of each dwell is random, with an exponential distribution determined by the kinetics. Generally, two consecutive dwells correspond to a step between two consecutive positions. However, due to the finite time resolution, the motor protein may take more than one step within the sampling interval, resulting in "missed events". We developed a maximum likelihood "idealization" algorithm to provide the dwell sequence for kinetic analysis (9). This procedure finds the motor's most likely position for each data point (implicitly detecting jump points), and estimates the step size distribution and transition probabilities. The idealization was tested successfully with different types of staircase data: uniform or alternating small and large steps, constant or variable step size, and reversible or irreversible kinetics.

Staircase data can be obtained from other molecular machines at the single molecule level. A typical example is the F_1-ATPase, which is a rotary molecular motor (Fig. 1B).

Unidirectional rotation of the central γ subunit is powered by ATP hydrolysis at three catalytic sites arranged 120° apart. At high resolution, each 120° step can be resolved into an 80° substep, driven by ATP binding, and another 40° substep (5, 6). Tracking the RNA polymerase position on the DNA template (7), or tracking the topoisomerase activity (8) also generates staircase data. What all these experiments have in common is the observation of a process with periodic chemistry through an increasing stochastic variable (i.e., position or rotational angle). In contrast, other single molecule data, such as the patch clamp recording of the current flowing through a single ion channel (13), observe a non-periodic stochastic variable with only a few states.

Stochastic data cannot be analyzed by fitting the mean. Instead, one must fit probability distributions, or maximize the likelihood. The simplest way to analyze dwell sequences is to construct a dwell time histogram (14) and fit it with a sum of exponentials, or directly with the probability density function (*pdf*). However, there are a few critical disadvantages to histogram fitting. Most importantly, the information provided by the correlation between dwell times is not utilized. For a model with N_S states, at equilibrium, a maximum number of $2N_S$ - 1 parameters can be extracted by histogram fitting: the time constant and the weight for each exponential component, minus one for the constraint that weights sum to one.

For example, the model shown in Fig. 2A has four rate constants but, since $N_S = 2$, only three parameters can be uniquely identified by histogram fitting. Thus, while the time constants of the underlying process can be estimated by histogram fitting, a detailed kinetic mechanism may be impossible to derive due to missing information. Furthermore, histogram fitting does not account for missed events, and requires a large amount of data to avoid counting errors. This can be a problem for single molecule fluorescence experiments, where photobleaching is limiting. Some of these problems are solved by maximizing the likelihood function. The maximum likelihood method utilizes all the information contained in the data and its estimates are asymptotically unbiased (15).

We present a maximum likelihood algorithm ("MIP", for Maximum Idealized Point likelihood) that estimates the rate constants of conformation and step transitions, from dwell time sequences. We model the molecular motor with an aggregated periodic Markov model. Each molecular conformation is assigned to a state in the model, and transitions between states are quantified by rate constants. The states of the Markov model must include not only the finite set of allosteric conformations, but also the position along the substrate, which may be large. However, we take advantage of the periodic chemistry, and reduce the Markov model to those core states and transitions that fully describe the kinetic mechanism. To reflect periodicity, certain constraints are imposed on these core transitions. In the calculation of the likelihood function, this reduced model is, in effect, recycled at each data point. Although computation with the truncated model is only asymptotically exact, the precision can be improved by increasing the size of the core model. For small data sets, optimization with the full model is possible (16), but it becomes too slow and numerically unstable for larger data sets that may require hundreds of states.

Stochastic single molecule data can be represented either with discrete-time or with continuous-time Markov models. A discrete-time algorithm maximizes the likelihood of a sequence of data points (17, 18), while a continuous-time algorithm maximizes the likelihood of a sequence of intervals (19, 20). Both algorithm types fully utilize the information contained in the data, and should give statistically equal estimates. For convenience, the likelihood function to be maximized by our MIP program is formulated for a discrete-time Markov model, with the added benefit that a correction for missed events (21) is no longer necessary.

The algorithm described here has advanced features (as described in (22)), such as global fitting across experimental conditions of arbitrary time course, linear constraints on rate constants (in addition to those reflecting the periodic step chemistry), etc. The algorithm can handle models with arbitrary kinetic complexity, including external driving forces, such as concentration, force, voltage, etc. The likelihood function and its derivatives are calculated analytically, permitting robust and fast (seconds to minutes) maximization. We tested the algorithm with a variety of kinetic models and simulated staircase data. Optimization with MIP was similar to the output from other maximum likelihood algorithms designed to estimate the rate constants of non-periodic Markov models (20, 18).

[Insert Fig. 1 about here]

Model and algorithms

Markov model

The behavior of a single molecule is well described with Markov models. At any position p along the substrate, the motor molecule is assumed to exist in one of N_S "aggregated" (i.e., experimentally undistinguishable) states. Thus, a staircase data set where the motor randomly walks across N_P positions requires a model with $N_S \times N_P$ states. Note that, due to missed events and reversible chemistry, the number of observed dwells may be different than N_P. In general, the rate constants of a Markov model are expressed as a rate matrix **Q** (23). For staircase data, the resulting rate matrix **Q** has a dimension $(N_S \times N_P) \times (N_S \times N_P)$. The **Q** matrix has each off-diagonal element q_{ij} equal to the rate constant between states $i \rightarrow j$, and each diagonal element q_{ii} equal to the negative sum of the off-diagonal elements of row i, so that the sum of each row is zero (this reflects the fact that the probability of being somewhere is unity). Hence, $-1/q_{ij}$ is equal to the mean lifetime of state i. Each rate constant q_{ij} has the following Eyring expression (24):

$$q_{ij} = k_{ij}^{\,0} \times C_{ij} \times e^{k_{ij}^{\,1} \times F_{ij}} \qquad (1)$$

where $k_{ij}^{\,0}$ is a pre-exponential factor, C_{ij} is the concentration of some ligand (such as ATP), and $k_{ij}^{\,1}$ is an exponential factor (Arrhenius), multiplied by the magnitude of the driving force F_{ij} (e.g., mechanical force). If the rate is not concentration dependent, then by definition $C_{ij} = 1$. Note that including C_{ij} in Eq. 1 relies on the assumption that C_{ij} is constant, in which case the kinetics are (pseudo-) first order. The units of the exponential factor $k_{ij}^{\,1}$ depend of the units of the driving force F_{ij}.

The occupancy of the $N_S \times N_P$ states is represented with a state probability vector **P**, of dimension $(N_S \times N_P)$. The Kolmogorov equation describes the dynamics of **P**:

$$\frac{\partial \mathbf{P}_t^{\,T}}{\partial t} = \mathbf{P}_t^{\,T} \cdot \mathbf{Q} \qquad (2)$$

where $\mathbf{P}_t$ is the state probability vector at time t, and the superscript ^{T} denotes vector transposition. The Chapman-Kolmogorov solution allows calculation of $\mathbf{P}_t$, given some initial value $\mathbf{P}_0$:

$$\mathbf{P}_t^T = \mathbf{P}_0^T \cdot e^{\mathbf{Q} \cdot t} \tag{3}$$

In the same way, the solution of $\mathbf{P}_t$ can be advanced over the sampling interval dt:

$$\mathbf{P}_{t+dt}^T = \mathbf{P}_t^T \cdot \mathbf{A} \tag{4}$$

where $\mathbf{A}$ is the transition probability matrix, of dimension $(N_S \times N_P) \times (N_S \times N_P)$. Each element a_{ij} is equal to the conditional probability that the process is in state j at time $t+dt$, given that it was in state i at time t. The $\mathbf{A}$ matrix is:

$$\mathbf{A} = e^{\mathbf{Q} \cdot dt} \tag{5}$$

By definition, the molecule cannot take more than one step at a time. Thus, the kinetic states at position p are directly connected only to the kinetic states at position $p + 1$ (forward step), and to the kinetic states at position $p - 1$ (backward step) (Fig. 2A). Therefore, $\mathbf{Q}$ is block-tridiagonal (Fig. 2B). Due to the periodicity of the kinetic model, the $\mathbf{Q}$ and $\mathbf{A}$ matrices have the following additional properties:

$$q_{ij} = q_{i \pm N_S, j \pm N_S} \tag{6}$$

$$a_{ij} = a_{i \pm N_S, j \pm N_S} \tag{7}$$

These two properties, together with the band-tridiagonal structure of the $\mathbf{Q}$ matrix, make the calculation of any q_{ij} element trivial. However, as further shown, the $\mathbf{A}$ matrix is also required to calculate the likelihood function in the discrete-time case. Even though the length of the reaction chain is practically limited by experimental conditions, the size of the corresponding Markov model could be very large, with possibly hundreds of states. How does one calculate the $\mathbf{A}$ matrix when $\mathbf{Q}$ is large?

Within a sampling interval dt, the motor can maintain position or can take any number of steps, back and forth. However, if the position at time t is known, the position at $t + dt$ is expected to be in the neighborhood, and the probability of finding the motor at a further away position decreases exponentially with the distance. If the sampling interval is short relative to the kinetics, the probability of the motor having undertaken more than a small number of steps approaches zero: $a_{ij} \to 0$, for j sufficiently far from i. While $\mathbf{Q}$ is band-tridiagonal, $\mathbf{A}$ is periodically banded. The size of any band of $\mathbf{A}$ is equal to N_S rows, and any two adjacent bands are identical but shifted horizontally by N_S columns (Fig. 2C). For a full representation of $\mathbf{A}$, it is therefore enough to calculate the elements a_{ij} within a band, such that j is within a certain distance from i, small enough to permit computation, yet large enough for precision. The other elements a_{ij}, where j is farther from i, can be approximated as zero. Any other band of $\mathbf{A}$ can be constructed simply by shifting left or right by a multiple of N_S columns.

A practical way of calculating the transition probabilities a_{ij} with minimal error is the following: first, a truncated rate matrix $\mathbf{Q}^r$ is constructed as if it were copied as a submatrix from

an infinite **Q**. The size of $\mathbf{Q}^r$ is $(2r + 1) \times (2r + 1)$ reaction units, where each unit has N_S states. The "truncation order" r is chosen to be greater than the highest-order step detected in the staircase data. A truncated $\mathbf{A}^r$ matrix is subsequently obtained from $\mathbf{Q}^r$ using Eq. 5. We have chosen the spectral decomposition technique (26, 27), which obtains the $\mathbf{A}^r$ matrix as follows:

$$\mathbf{A}^r = \sum_i \mathbf{A}_i \cdot e^{\lambda_i \cdot t} \tag{8}$$

where $\mathbf{A}_i$ are the spectral matrices obtained from the eigenvectors of $\mathbf{Q}^r$, and λ_i are the eigenvalues of $\mathbf{Q}^r$. A submatrix copied from the theoretically infinite **A** should be approximately equal to the corresponding submatrix copied from $\mathbf{A}^r$, if r is sufficiently large. For the example shown in Fig. 2C, we found that the approximation error becomes negligible for $r \geq 3$. Note that since the first and last rows of $\mathbf{Q}^r$ do not have zero sum, the $\mathbf{A}^r$ matrix is no longer row stochastic, i.e., the sum of each row is not one. In conclusion, all calculations can be done with a truncated model of small size.

[Insert Fig. 2 about here]

Dwell time probability density function

The probability density function for the lifetime of an aggregated Markov model in a given class is a sum of exponentials, with as many components as states in that class (28). To calculate the *pdf*, **Q** is partitioned into submatrices $\mathbf{Q}_{ab}$ of dimension $(N_a \times N_b)$, where N_a and N_b are the numbers of states in aggregation classes a and b, respectively. The *pdf* of the transitions between classes a and b is given by the matrix $\mathbf{G}_{ab}(t)$, as follows:

$$\mathbf{G}_{ab}(t) = e^{\mathbf{Q}_{aa} \cdot t} \cdot \mathbf{Q}_{ab} \tag{9}$$

where $\mathbf{Q}_{aa}$ is the submatrix of **Q** that contains only transitions within class a, and $\mathbf{Q}_{ab}$ is the submatrix containing transitions between classes a and b. The matrix $e^{\mathbf{Q}_{aa} \cdot t}$ represents the probabilities of transition within class a. $\mathbf{G}_{ab}(t)_{ij}$ is the *conditional pdf* that the process stays for time t in class a, given that it entered class a through state i and that it exited into class b through state j. To calculate the *unconditional pdf*, one must take into account the probabilities of entering through each state of class a, and all possible exit classes $b \neq a$. Thus, the *pdf* of dwelling for time t in class a, denoted $f_a(t)$, has the following expression (28, 29):

$$f_a(t) = \mathbf{P}_a{}^T \cdot e^{\mathbf{Q}_{aa} \cdot t} \cdot \sum_{b \neq a} (\mathbf{Q}_{ab} \cdot \mathbf{1}) \tag{10}$$

where the vector $\mathbf{P}_a$ (of dimension N_a) represents the entry probabilities in class a, and **1** is a vector of ones, of dimension matching $\mathbf{Q}_{ab}$. With the assumption that the process is at equilibrium, $\mathbf{P}_a$ can be calculated as follows (28):

$$\mathbf{P}_a{}^T = \frac{\sum_{b \neq a} \mathbf{P}_{b,eq}{}^T \cdot \mathbf{Q}_{ba}}{\sum_{b \neq a} \mathbf{P}_{b,eq}{}^T \cdot \mathbf{Q}_{ba} \cdot \mathbf{1}} \tag{11}$$

where $\mathbf{P}_{b,eq}$ is a vector of dimension N_b, containing the equilibrium probabilities for the states in class b. Essentially, the above expression calculates the entry probabilities into class a as the equilibrium probabilities in class $b \neq a$ multiplied by the rates between $b \rightarrow a$. The role of the denominator is to normalize the resulting probabilities to one. For a simple two state model, Eq. 10 reduces to $f_a(t) = k_{ab} \times e^{-k_{ab} \cdot t}$.

Likelihood function for the continuous-time model

The one-dimensional *pdf* (Eqs. 9 and 10) can be used to calculate the joint *pdf* of a sequence of dwells (**t**, **a**) as follows (30) :

$$f(\mathbf{t},\mathbf{a}) = \mathbf{P}_{a_1}{}^{T} \cdot e^{\mathbf{Q}_{a_1 a_1} \cdot t_1} \cdot \mathbf{Q}_{a_1 a_2} \cdot e^{\mathbf{Q}_{a_2 a_2} \cdot t_2} \cdot \mathbf{Q}_{a_2 a_3} \cdots e^{\mathbf{Q}_{a_T a_T} \cdot t_T} \cdot \sum_{b \neq a_T} \left(\mathbf{Q}_{a_T b} \cdot \mathbf{1} \right) \tag{12}$$

where **t** is the dwell time sequence **t** = [t_1, t_2, …, t_T], and **a** is the class sequence **a** = [a_1, a_2, …, a_T]. f(**t**, **a**) is in fact the *likelihood* of observing the dwell sequence (**t**, **a**), given the model and its parameter values.

The likelihood function above is formulated for an ideal dwell sequence, in which all transitions between different classes are observed. In practice, due to finite temporal resolution, some short dwells will always be missed. These missed events may result in observed steps between non-consecutive positions. Notice that the submatrix for transitions between non-consecutive positions is zero, which causes the likelihood to be zero too. Without a correction for missed events, dwell times are overestimated, and model parameters are biased. The distortion caused by missed events can be corrected by an exact but slow solution (31, 32), or by a fast approximation (33, 20). For example, in the first order correction (20), the *pdf* has exactly the same form, except it uses the matrices ${}^{e}\mathbf{Q}_{aa}$ and ${}^{e}\mathbf{Q}_{ab}$, corrected for missed events, as follows:

$${}^{e}\mathbf{Q}_{aa} = \mathbf{Q}_{aa} - \mathbf{Q}_{a\bar{a}} \cdot \left(\mathbf{I} - e^{\mathbf{Q}_{\bar{a}\bar{a}} \cdot t_d} \right) \cdot \mathbf{Q}_{\bar{a}\bar{a}}{}^{-1} \cdot \mathbf{Q}_{\bar{a}a} \tag{13}$$

$${}^{e}\mathbf{Q}_{ab} = \exp\left[t_d \cdot \mathbf{Q}_{a\bar{a}} \cdot \left(\mathbf{I} - e^{\mathbf{Q}_{\bar{a}\bar{a}} \cdot t_d} \right) \cdot \mathbf{Q}_{\bar{a}\bar{a}}{}^{-1} \cdot \mathbf{Q}_{\bar{a}a} \right] \cdot \left[\mathbf{Q}_{ab} - \mathbf{Q}_{ac} \cdot \left(\mathbf{I} - e^{\mathbf{Q}_{cc} \cdot t_d} \right) \cdot \mathbf{Q}_{cc}{}^{-1} \cdot \mathbf{Q}_{cb} \right] \tag{14}$$

where **I** is the identity matrix and t_d is the deadtime. $\bar{a}$ refers to those states that are not in class a, and c refers to those states that are neither a nor b.

In the case of processive molecular motors, there is generally only one aggregation class, and its dwell time *pdf* has N_S exponentials. The **Q** matrix can be partitioned using three submatrices $\mathbf{Q}_0$, $\mathbf{Q}_{-1}$ and $\mathbf{Q}_1$, as shown in Fig. 2B. Thus, $\mathbf{Q}_0$ is the submatrix of transitions within the same reaction unit, while $\mathbf{Q}_{-1}$ and $\mathbf{Q}_1$ are the submatrices of transitions between consecutive units, in the forward or backward direction, respectively. The dwell time *pdf* has the following expression:

$$f(t) = \mathbf{P}_0{}^{T} \cdot e^{\mathbf{Q}_0 \cdot t} \cdot \left(\mathbf{Q}_{-1} + \mathbf{Q}_1 \right) \cdot \mathbf{1} \tag{15}$$

where $\mathbf{P}_0$ is the vector of entry probabilities into one unit. $\mathbf{P}_0$ can be obtained as follows:

$$\mathbf{P}_0^{\,T} = \frac{\mathbf{P}_{eq}^{\,T} \cdot (\mathbf{Q}_{-1} + \mathbf{Q}_1)}{\mathbf{P}_{eq}^{\,T} \cdot (\mathbf{Q}_{-1} + \mathbf{Q}_1) \cdot \mathbf{1}} \tag{16}$$

where $\mathbf{P}_{eq}$ is the vector of equilibrium probabilities for the states in one reaction unit. The calculation of $\mathbf{P}_{eq}$ deserves some discussion: the equilibrium probability for any state i along the chain should be very small, theoretically zero for an infinite chain. However, since the process is periodic, dwells are indistinguishable: dwells corresponding to different positions have identical *pdf*. Therefore, the equilibrium probability for a state i inside a reaction unit, taken separately, is obtained by summing the probability of that state across all reaction units, which is a finite number. Furthermore, the same value is obtained whether the sum is over an open and infinite chain, or over a circular and finite chain. Thus, a practical way to calculate $\mathbf{P}_{eq}$ is by constructing a circular model that contains two reaction units connected in a loop. From the $\mathbf{Q}$ matrix of this circular model, equilibrium probabilities can be calculated with the method described in (23). From these, $P_{eq,i}$ is obtained as twice the equilibrium probability of state i in one unit of the circular model.

The likelihood L of a staircase dwell sequence can be written as follows:

$$L = \mathbf{P}_0^{\,T} \cdot e^{\mathbf{Q}_0 \cdot t_1} \cdot \mathbf{Q}_{i_1} \cdot e^{\mathbf{Q}_0 \cdot t_2} \cdot \mathbf{Q}_{i_2} \cdots e^{\mathbf{Q}_0 \cdot t_T} \cdot (\mathbf{Q}_{-1} + \mathbf{Q}_1) \cdot \mathbf{1} \tag{17}$$

where $\mathbf{Q}_i$ is equal to either $\mathbf{Q}_1$ or $\mathbf{Q}_{-1}$, depending on whether the step between the two consecutive dwells is in the forward or in the backward direction. An example of how to calculate the likelihood function for a given dwell time sequence is shown in Fig. 2E. Application of the missed events correction to staircase data is possible, but for details we refer the reader to (19, 20). Note that, due to missed events correction, the submatrix for transitions between non-consecutive positions is no longer zero.

Likelihood function for the discrete-time model

For easier understanding and without loss of generality, we assume that the maximum step detected in the data is of order ±1, i.e., any two consecutive measurements are not separated by more than one position. To better illustrate the method, we also assume the kinetic model is reversible, with two states per reaction unit, such as shown in Fig. 2A. We choose a truncation order $r = 1$, and thus the model has $2r+1 = 3$ units (referred to as "left", "center" and "right"), and a total of six states. Hence, the $\mathbf{Q}^r$ and $\mathbf{A}^r$ matrices have a dimension of 3×3 units, or 6×6 elements. Similarly, the state probability vector $\mathbf{P}$ and all other vectors used in the computation have a dimension of 3 units. For increased numerical accuracy, $\mathbf{A}^r$ should be obtained from a larger $\mathbf{Q}^r$ before truncation.

The likelihood L of a sequence [0…T] of idealized staircase data points is given in matrix form by the following expression:

$$L = \left(\mathbf{P}_0^{\,T} \cdot \mathbf{B}_0 \cdot \mathbf{S}\right) \cdot \left(\mathbf{A}^r \cdot \mathbf{B}_1 \cdot \mathbf{S}\right) \cdot \left(\mathbf{A}^r \cdot \mathbf{B}_2 \cdot \mathbf{S}\right) \cdots \left(\mathbf{A}^r \cdot \mathbf{B}_T \cdot \mathbf{S}\right) \cdot \mathbf{1} \tag{18}$$

The likelihood expression above is similar to that in (18). $\mathbf{P}_0$ is the initial state probability vector. All its entries are zero, except those corresponding to the center unit, which are non-zero and sum to one. $\mathbf{B}_t$ is a diagonal matrix, with entries equal to either 0 or 1 (Fig. 2D). For data points at the beginning of dwells, following a forward jump, the entries in the left block are 1 and all

the others are zero. For points corresponding to a backward jump, the entries in the right block are 1 and all the others are zero. For all other points, the entries in the center block are 1 and all the others are zero. At each point t, the appropriate $\mathbf{B}_t$ matrix is plugged into the likelihood expression (Fig. 2E).

In (18), $\mathbf{B}_t$ is also a diagonal matrix, but its diagonal entries are Gaussian conditional probability densities. There, these probability densities measure how likely it is that a noisy data point was obtained while the process was in a given state. Here, the probability densities are replaced with zero or one, since the signal is already separated from noise. Note that with data obtained from motors with high step variability (e.g., dynein), the step order may be ambiguous. In this case, $\mathbf{B}_t$ can be modified as to include all possible jump orders, each with a calculated weight. **S** is a permutation matrix (Fig. 2D) that sums, block-wise, the state probabilities, moves the sum into the center block and clears the left and right blocks. The role of the **S** matrix is to map the likelihood computation from the theoretically infinite state space into the finite, small state space of the truncated $\mathbf{A}^r$ matrix. This is permitted by the periodic chemistry and by the periodic properties of the **A** matrix (Eq. 7).

In short, the likelihood function in Eq. 18 propagates the state probabilities through the time sequence, according to Eq. 4, but it also compares the theoretical prediction with the data. The calculation starts with the initial state probabilities $\mathbf{P}_0$, which represent the *a priori* knowledge about the initial state occupancies. Then, for each time t, $\mathbf{P}_{t+1}$ is first predicted by post-multiplying $\mathbf{P}_t$ with the $\mathbf{A}^r$ matrix (as in Eq. 4). The prediction is then corrected by the evidence contained in the data, i.e., by post-multiplication with the diagonal matrix $\mathbf{B}_{t+1}$. This correction leaves unmodified the entries in the state probability vector corresponding to those states that match the actual data, while it zeroes all the others. The corrected $\mathbf{P}_{t+1}$ is the *a posteriori* estimate. The corrected state probabilities are then reset by post-multiplication with the permutation matrix **S**, so that only the states in the center block are occupied and the others are zero. Although the same result can be achieved without explicitly using the **S** matrix, this formulation permits a consistent matrix form of the likelihood function and of its derivatives, as shown below. Consequently, only the relative difference in position between two consecutive data points is used. Note that the **S** matrix is strictly necessary only at the beginning of dwells and has no effect otherwise. Finally, the *a posteriori* state probabilities at the last data point are summed over all states, by post-multiplication with a column vector of ones. This sum is equal to the likelihood. The more the prediction matches the actual data, the higher the likelihood. An example of how to calculate the likelihood for a point sequence is shown in Fig. 2E.

Maximization of the likelihood function

The objective is to find the parameter set $\boldsymbol{\theta}^{ML}$ that maximizes the likelihood L, or, equivalently, the log-likelihood LL:

$$\boldsymbol{\theta}^{ML} = \arg\max_{\boldsymbol{\theta}} LL \tag{19}$$

Either the continuous-time or the discrete-time likelihood function can be maximized. Here, we have chosen the discrete-time case, for two main reasons: it is a simpler computation, and it does not require correction for missed events. The likelihood function is maximized numerically. For details of implementation, we refer the interested reader to (22). The parameters to be estimated are the pre-exponential and the exponential factors k_{ij}^0 and k_{ij}^1, for each rate constant q_{ij}. A transformation of variable enforces the constraint of positive pre-exponential factors k_{ij}^0 (22).

Other constraints are the model periodicity (Eq. 6), and those derived from *a priori* knowledge (22), such as fixed or scaled rates, or microscopic reversibility of cycles. We implemented an efficient mechanism for imposing these constraints using the singular value decomposition, as previously reported (18, 22). The algorithm optimizes a set **x** of free parameters, from which the kinetic parameters k_{ij}^{0} and k_{ij}^{1} are subsequently calculated. Each constraint reduces by one the number of free parameters.

The likelihood is computed using recursive forward and backward vector variables $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, initialized and calculated as follows (18):

$$\begin{aligned}
\boldsymbol{\alpha}_0^{\,T} &= \mathbf{P}_0^{\,T} \cdot \mathbf{B}_0 \cdot \mathbf{S} \\
\boldsymbol{\alpha}_{t+1}^{\,T} &= \boldsymbol{\alpha}_t^{\,T} \cdot \left(\mathbf{A} \cdot \mathbf{B}_{t+1} \cdot \mathbf{S}\right) \\
\boldsymbol{\beta}_T &= \mathbf{1} \\
\boldsymbol{\beta}_t &= \left(\mathbf{A} \cdot \mathbf{B}_{t+1} \cdot \mathbf{S}\right) \cdot \boldsymbol{\beta}_{t+1}
\end{aligned} \tag{20-23}$$

Numerical underflow is avoided by using the scaling factors s_t to calculate the normalized probabilities $\breve{\boldsymbol{\alpha}}$ and $\breve{\boldsymbol{\beta}}$, as follows:

$$\begin{aligned}
s_0 &= \boldsymbol{\alpha}_0 \cdot \mathbf{1} \\
s_{t+1} &= \breve{\boldsymbol{\alpha}}_t^{\,T} \cdot \left(\mathbf{A} \cdot \mathbf{B}_{t+1} \cdot \mathbf{S}\right) \cdot \mathbf{1} \\
\breve{\boldsymbol{\alpha}}_0^{\,T} &= \boldsymbol{\alpha}_0^{\,T} / s_0 \\
\breve{\boldsymbol{\alpha}}_{t+1}^{\,T} &= \breve{\boldsymbol{\alpha}}_t^{\,T} \cdot \left(\mathbf{A} \cdot \mathbf{B}_{t+1} \cdot \mathbf{S}\right) / s_{t+1} \\
\breve{\boldsymbol{\beta}}_T &= \boldsymbol{\beta}_T \\
\breve{\boldsymbol{\beta}}_t &= \left(\mathbf{A} \cdot \mathbf{B}_{t+1} \cdot \mathbf{S}\right) \cdot \breve{\boldsymbol{\beta}}_{t+1} / s_{t+1}
\end{aligned} \tag{24-29}$$

Note that $\breve{\boldsymbol{\alpha}}_t^{\,T} \cdot \breve{\boldsymbol{\beta}}_t = 1$, since the same scaling factors were used for both $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. Hence, *LL* can be conveniently calculated using the scaling factors s_t, while its derivatives $\partial LL/\partial q_{ij}$ can be calculated using the chain rule of matrix differentiation, as follows:

$$LL = \sum_{t=0}^{T} \ln s_t \tag{30}$$

$$\frac{\partial LL}{\partial q_{ij}} = \frac{\partial \mathbf{P}_0^{\,T}}{\partial q_{ij}} \cdot \mathbf{B}_0 \cdot \mathbf{S} \cdot \breve{\boldsymbol{\beta}}_0 + \sum_{t=0}^{T-1} \left[\breve{\boldsymbol{\alpha}}_t^{\,T} \cdot \frac{\partial \mathbf{A}}{\partial q_{ij}} \cdot \mathbf{B}_{t+1} \cdot \mathbf{S} \cdot \breve{\boldsymbol{\beta}}_{t+1} \right] \tag{31}$$

The derivatives $\partial \mathbf{P}_0^{\,T}/\partial q_{ij}$, $\partial \mathbf{A}/\partial q_{ij}$, $\partial \mathbf{Q}/\partial q_{ij}$, and the **A** matrix itself are calculated as in (22), with the mention that $\partial \mathbf{Q}/\partial q_{ij}$ takes into account the periodic chemistry (Eq. 6). The derivative of *LL* with respect to a free parameter x_k is calculated with the chain rule, as follows:

$$\frac{\partial LL}{\partial x_k} = \sum_{i,j} \left[\frac{\partial LL}{\partial q_{ij}} \left(\frac{\partial q_{ij}}{\partial v_{ij}} \frac{\partial v_{ij}}{\partial x_k} + \frac{\partial q_{ij}}{\partial k_{ij}^{1}} \frac{\partial k_{ij}^{1}}{\partial x_k} \right) \right] \tag{32}$$

where $v_{ij} = \ln k_{ij}^0$, $\partial q_{ij} / \partial v_{ij} = q_{ij}$, and $\partial q_{ij} / \partial k_{ij}^1 = F_{ij} \times q_{ij}$ are obtained as in (22). In the above calculation, **A** and **Q** are the truncated matrices. The indexes *i* and *j* are for the off-diagonal entries in the **Q** matrix corresponding to non-zero rates in the kinetic model. Note that the likelihood function can be maximized without its analytical derivatives, but their availability allows for a significantly faster and more precise maximization. The optimization routine we used calculates the covariance matrix of the free parameters **x**. From this, we calculate the error estimates of the rates (22).

We accelerated the computation of the likelihood function and its derivatives by pre-computing multiplications of identical terms in the likelihood function (Eq. 18). For example, a dwell of length 73 points can be represented in the likelihood calculation either as a product of 73 identical terms, or more efficiently as a smaller product of powers, such as the following:

$$L = \cdots (\mathbf{A} \cdot \mathbf{B} \cdot \mathbf{S})^{64} \cdot (\mathbf{A} \cdot \mathbf{B} \cdot \mathbf{S})^{8} \cdot (\mathbf{A} \cdot \mathbf{B} \cdot \mathbf{S}) \cdots \qquad (33)$$

Note that the sum of the exponents 64 + 8 + 1 = 73. These exponents can be conveniently chosen as powers of two, e.g. 1, 2, 4, 8…, etc. Thus, one can first calculate the power with exponent one, which by multiplication with itself gives the power with exponent two, and so on. It is therefore enough to calculate the term $\mathbf{A} \cdot \mathbf{B} \cdot \mathbf{S}$ and then its 2^{nd}, 4^{th}, 8^{th}… powers. Thus, *K* terms in the series $(\mathbf{A} \cdot \mathbf{B} \cdot \mathbf{S})^{2^k}$ can be calculated in only *K* matrix multiplications, and only once for the whole data set. This treatment, applied to the likelihood function and its derivatives, results in a considerable speed improvement.

Materials and methods

All the computer work was done with the QuB program (10), running MS Windows XP on a 3.0 GHz Intel PC. Staircase and non-staircase noisy data and idealized dwell time sequences (subject to finite temporal resolution) were generated with the SIM routine. The simulator also provides the actual number of transitions for any pair of states, n_{ij}, and the actual time spent in each state, t_i. From these values, the true rate constants are calculated as $k_{ij} = n_{ij} / t_i$. Staircase data were generated with periodic Markov models, as utilized for processive molecular motors, while non-staircase data were generated with non-periodic Markov models, as utilized for ion channels. For estimation of rate constants, we used the MIP, MPL and MIL routines. MIP works by optimizing the discrete-time likelihood of either a staircase or a non-staircase dwell time sequence. MPL maximizes the discrete-time likelihood of a data sequence generated by a non-periodic Markov model (18). MIL maximizes the continuous-time likelihood of a dwell time sequence, and provides first order correction of missed events (20). All three algorithms calculate analytically the gradients of the likelihood function, and maximize the likelihood function using the same fast variable metric optimizer (*dfpmin)*, implemented as in (34) with modifications.

Results

Figure 3 shows a few examples of kinetic models that can be modeled by our algorithm. For each model, only motor steps can be observed (transitions between different units) but not

transitions between states inside the same reaction unit. All models implicitly include ATP binding and mechanical force-dependent steps. In Fig. 3, only model E is formulated with explicit ATP binding, in order to emphasize its special characteristic of having alternating steps with different chemistry. This model describes the two substeps of the F_1-ATPase: an 80° ATP binding substep and a 40° substep (5, 6). Alternating steps are handled by our method, with the simple modification that the permutation matrix **S** is applied only after a repeating pair of two alternating units. An illustration of the need for constraints on rate constants is model C, which describes a molecular motor, such as myosin V (2), with the fluorescent probe attached to the motor head. In this case, only every other motor step can be observed. Consequently, there are twice as many states per reaction unit but, due to constraints, the number of free parameters remains the same as for model B. Notice how dwell time histograms are truncated, since dwells shorter than the sampling interval (as shown, dt = 500 ms) are missed.

[Insert Fig. 3 about here]

Solving models with one state per reaction unit (Fig. 3A) is trivial: the two rates can be calculated by hand simply by dividing the number of forward and backward jumps (including missed events) by the length of the data. Although trivial, single state models are informative about possible bias in the maximum likelihood estimates. We tested the algorithm with both reversible and irreversible data generated by single state models. The rates obtained by the maximum likelihood algorithm were equal (within numerical precision) to the rates obtained by hand calculation. These results validate the algorithm, and show that the estimates are unbiased.

Next, we tested the algorithm with a non-trivial model with two states per reaction unit, as shown in Fig. 2A. First, data simulated with the two state model were fit with either the correct model, or with a single state model (Fig. 4A). The estimates obtained with the correct model are virtually identical to the true rates used by the simulator. Furthermore, the difference between the true and the estimated rates, however small, is well within the standard deviation of the estimates, as calculated by the optimizer. The two models have nearly identical *pdfs* (shown overlapped on the dwell time histogram), but their log-likelihoods are very different: -79561.27 (two state model) versus -83254.93 (one state model). Clearly, the correct model is favored by the likelihood function. In contrast, the histogram fitting, which ignores correlations in time, is not capable of distinguishing the two models.

Is the increase in likelihood for larger models spurious, simply the result of having more fitting parameters? The correct model has four parameters while the single state model has only two. To estimate the role of an increasing number of parameters on the increase in likelihood, we used the Akaike (*AIC*) (35) and the Bayesian (*BIC*) (36) information criteria, defined as follows:

$$AIC = -2(LL - k) \tag{34}$$

$$BIC = -2(LL - 0.5 \times k \times \ln N) \tag{35}$$

where k is the number of free parameters and N is the number of data points. For our example, both *AIC* and *BIC* favor the correct model by a large margin. We also simulated data with the single state model and fit them with both models (Fig. 4B). In this case, the log-likelihoods are virtually identical, and likewise the *pdf* curves. This indicates that the larger model is overfitting. Also indicative of overfitting is the large standard deviation of the estimates.

[Insert Fig. 4 about here]

Statistical distributions of kinetic estimates

To determine the statistical properties of the estimates, we simulated staircase data with the two state model shown in Fig. 3A. We found that the estimates for all rates have Gaussian distributions (Fig. 5A). However, obtaining a meaningful rate estimate requires that sufficient information about that transition is contained in the data. For example, k_B had a large variance and it was bimodal when only two data traces were globally fit (Fig. 5A, upper graphs). A large fraction of the estimates consisted of zero values (in the optimization, we constrained $k_B \geq 1\times10^{-5}$ s^{-1}), while the remaining fraction consisted of finite values, loosely centered on the true (simulated) value. The explanation is simple: in the simulated data file used for analysis, the following transitions were counted: $B_i \rightarrow A_i$ = 5,315, $A_i \rightarrow B_i$ = 98,532, $B_i \rightarrow A_{i+1}$ = 131,627, $A_{i+1} \rightarrow B_i$ = 39,477. As the file had 1000 segments of 5000 points each, the $B_i \rightarrow A_i$ transition (quantified by k_B) occurred about once per trace. While some traces had one or a few $B_i \rightarrow A_i$ transitions, others had none. This shortage of information explains the poor quality of k_B estimates (Fig. 5A, upper graphs). In comparison, the distribution of the k_B estimates improved considerably when 20 segments were globally fit (Fig. 5A, lower graphs).

We also observed a lack of cross-correlation between different pairs of estimates (Fig. 5B), which implies that the parameters are rather orthogonal. All four rate constants were estimated without bias (Fig. 5C, left graph) if at least 10 traces were globally fit. The standard error of all estimates was less than 10% (Fig. 5C, right graph) if the analyzed data contained at least a few hundred events of each transition. We emphasize that the source of this variance is the stochastic nature of the dwell sequence, and not the optimization program. One should regard these results as a measure of how much the estimates obtained from experimental data may depart from the true parameters of the generating process, under conditions of limited data.

We also tested the algorithm with irreversible data, generated with the same model but with either $k_B = 0$, or $j_B = 0$ (see Fig. 3C). As before, we constrained the estimated rates to be $\geq 1\times10^{-5}$ s^{-1}, to avoid numerical problems. The data were analyzed in groups of 20 segments, with the zero rate (k_B or j_B) as a free parameter, or constrained to 1×10^{-5}. In both cases, we obtained statistically correct estimates (results not shown). Data simulated with $j_B = 0$ contain direct evidence for a zero j_B, i.e., they lack backward steps. Thus, the algorithm theoretically should have obtained all j_B estimates equal to 1×10^{-5}. In practice, the optimizer estimated j_B as $2.01\times10^{-5} \pm 4.02\times10^{-5}$ (max = 4.6×10^{-4}), close to the expected value. In contrast, data simulated with $k_B = 0$ do not contain direct evidence for a zero k_B. While ideally k_B should also have been estimated as 1×10^{-5}, in practice j_B was estimated as $2.2\times10^{-4} \pm 8.6\times10^{-4}$ (max = 1×10^{-2}), a fairly good approximation.

[Insert Fig. 5 about here]

Comparison between MIP and other kinetic algorithms

The MIP algorithm may be applied to data generated by non-periodic Markov models, such as single channel data, with the removal of the **S** matrix from Eq. 18. We compared MIP against the MIL (20) and MPL (18) maximum likelihood algorithms used for single channel analysis. We simulated data with a three state "closed-open-closed" model. To test the effect of missed events, we intentionally chose two of the four rate constants to be comparable to the sampling time (2 s^{-1} and 5 s^{-1}, versus a sampling time dt = 0.1 s). Figure 6A shows that the three

methods are equivalent, and their estimates are statistically equal. As expected, the cross-correlation is almost perfect between MIP and MPL, as they are both point methods designed for discrete-time Markov models, as opposed to MIL, which is interval-based and designed for continuous-time Markov models.

The experiment also showed that it is legitimate to replace the Gaussian densities in the likelihood function with hard probabilities of zero and one, when the data are already idealized. MIL estimates may have greater variance than MIP or MPL estimates, which suggests that MIP requires less data to achieve the same precision. More importantly, the experiment demonstrates that a discrete-time method implicitly handles missed events as well as, or better than, the continuous-time method which explicitly requires correction for missed events. MIL estimates may depend critically on the choice of the dead time parameter (20). For this experiment, we found the best value to be ≈ 1.6 times the sampling interval (Fig. 6B), but the dead time is a function of the model and data.

The convergence time of the three programs (using a MIP pre-multiplication order of 16) scaled approximately as 1 (MIL): 1.6 (MIP): 8 (MPL). MIL is clearly the fastest, as it operates with dwells and not points (there are fewer dwells than points), but MIP is close. Thus, the potentially slower computation of point-based methods is no longer a problem. For example, fitting a two-state model to a data set with ~ one hundred dwells converged in seconds. Fitting a model with four states to a few thousands dwells takes minutes. All three programs converged in approximately the same number of iterations.

[Insert Fig. 6 about here]

Discussion

Molecular motors are mechano-chemical systems that *in vivo* may depart from a simple (pseudo-) first order kinetics. However, under the right *in vitro* experimental setup their intrinsic kinetics can be studied using Markov models, which can include dependence on external stimuli, such as ATP used to convert chemical into mechanical energy, or the generation of mechanical motion and force (11). The parameters thus obtained can be incorporated into more comprehensive *in vivo* models. In this paper we showed that: i). the potentially very large but periodic Markov model can be reduced to a small, truncated version; ii). the likelihood function of the stochastic staircase dwell sequence can be formulated in a compact form, for chemistry of arbitrary complexity; iii). unbiased kinetic parameters can be estimated by numerical maximization of the likelihood function.

Continuous-time versus discrete-time models

The discrete-time likelihood function makes no assumption about what happens between sampling points, and therefore does not "miss" events. In contrast, in the continuous-time model, it is assumed that the process does not change class during a dwell. Since this is not actually the case, a correction for missed events is necessary (31-33, 20). MIP does not require correction for missed events and thus has the potential of being more accurate than interval-based algorithms that have to deal with missed events (19, 20). It does not mean, however, that a point method such as MIP will work at any temporal resolution. Thus, when the sampling is slow relative to the kinetics, the process will reach equilibrium between samples, and information about

transitions will simply be lost. Neither MIP nor any other method can extract useful estimates in this case.

The rate estimates are unbiased and Gaussian

The maximum likelihood rate estimates are intrinsically unbiased, but the confidence level depends on the amount of available data (Fig. 5). Thus, a rate cannot be properly estimated when the corresponding transition occurs rarely. In this case, the distribution of the estimate may appear biased, or bimodal (Fig. 5A, upper graphs). While this is generally not a problem with single channel data which usually contain a large number of events, it can be a serious factor with data limited by photobleaching or by short substrate filaments. We recommend testing algorithm performance using simulated data to explore the errors in particular cases. We emphasize that this variability is determined by the stochastic nature of single molecule data and not due to the algorithm.

Parameter identifiability and model selection

Although the likelihood function (Eq. 18) fully utilizes the information contained in the data, any modeling study must answer two important questions: for a given model, how many kinetic parameters can be uniquely identified? Second, can two models be distinguished, according to some objective criterion? Regarding the first question, it is known that the maximum number of parameters that can be estimated from equilibrium single channel data is $2\times N_C\times N_O$ (30, 37), where N_C and N_O are the numbers of "closed" and "open" states, respectively. How do we define "closed" and "open" in the case of staircase data? In a sense, the molecular motor model has only one aggregation class; however, jump transitions can be observed. Therefore, a minimal representation of the periodic model presented in Fig. 2A can be formulated with four state transitions (denoted by k_F, k_B, j_F and j_B) connecting three states (e.g., A_i, B_i and A_{i+1}) that can be partitioned in two "conductance" classes: states A_i and B_i are "closed", and A_{i+1} is "open". From such a model, a maximum of four kinetic parameters should be uniquely identifiable. Our analysis (see Figs. 4 and 5) confirmed this prediction: all four rate constants were uniquely identified.

In practice, the maximum number of parameters that can be determined is a function of the kinetic model, the experimental protocol (stationary versus non-stationary, local versus global fitting, etc), and the availability of an adequate amount of data (see Fig. 5). If more complex models are to be studied, the parameter identifiability may be improved by manipulation of experimental conditions, e.g., global fitting across different ATP concentrations or mechanical force values (22). The algorithm presented here was implemented with this need in mind, and can globally model data obtained under different experimental protocols, including nonstationary stimuli (22).

With respect to model identifiability, it is known that two Markov models that are related by a similarity transform give identical likelihood (38, 39) and thus cannot be distinguished. In this case, global fitting across different conditions or using non-stationary stimuli will improve identifiability. Furthermore, due to the stochastic nature of the data, the likelihood estimator has an intrinsic variance proportional to the number of data points. When there are a small number of data points, the likelihood distributions for two different models may overlap. In the example shown in Fig. 4, the two models were correctly selected because we used a relatively large amount of data. Again, we recommend using simulations to check the statistical separation between likelihood distributions. Note that when models of different size are compared, their

likelihoods must be scaled for the number of free parameters and for the amount of data, as shown in Results.

Reversible models and truncated matrices do not pose numerical problems

Two kinds of numerical errors may in principle affect the kinetics algorithm. First, irreversible models may easily result in confluent (i.e., degenerate) **Q** matrix eigenvalues. In this case, the transition matrix **A** cannot be calculated using the convenient spectral decomposition (26, 27). A simple yet effective solution is to set a lower limit on rates, for example constraining the rates to >1/data length. Although this could potentially introduce bias in estimates, we found this bias to be negligible in practice.

The second source of numerical error is computing the truncated transition probability matrix $\mathbf{A}^r$. The error between the truncated and the theoretical **A** matrices decreases with the truncation order. However, since the computational speed scales quadratically with the size of the model, the truncation order is chosen as a compromise between accuracy and speed. We found that a model with $2r+1 = 7\ldots9$ units runs reasonably fast (seconds to minutes) and has an error that is negligible relative to the intrinsic variance of the estimates. Note that when the analyzed data lack backward steps (i.e., they are irreversible), the model can be simplified to only $r+1$ units, since the "left" block states will never be occupied in the likelihood chain.

Acknowledgements

We thank Asbed Keleshian and John E. Pearson for numerous and fruitful discussions, and Chris Nicolai and Harish Ananthakrishnan for help with programming. This work was supported by the National Institutes of Health (RR11114 for F.S.; AR44420 and GM068625 to P.R.S.).

References

1. Vale, R.D., and R.A. Milligan. 2000. The Way Things Move: Looking Under the Hood of Molecular Motor Proteins. *Science*. 288:88-95.
2. Yildiz, A., J.N. Forkey, S.A. McKinney, T. Ha, Y.E. Goldman, and P.R. Selvin. 2003. Myosin V walks hand-over-hand: Single fluorophore imaging with 1.5 nm localization. *Science.* 300:2061-2065.
3. Yildiz, A., M. Tomishige, R.D. Vale, and P.R. Selvin. 2004. Kinesin walks hand-over-hand. *Science.* 303:676-678.
4. Kural, C., H. Kim, S. Syed, G. Goshima, V.I. Gelfand, and P.R. Selvin. 2005. Kinesin and dynein move a peroxisome in vivo: a tug-of-war or coordinated movement? *Science* 308:1469-72.
5. Yasuda, R, H. Noji, M. Yoshida, K. Kinosita, and H. Itoh. 2001. Resolution of distinct rotational substeps by submilisecond kinetic analysis of F_1-ATPase. *Nature,* 410:898-904.
6. Nishizaka, T., K. Oiwa, H. Noji, S. Kimura, M. Yoshida, and K. Kinosita Jr. 2004. Chemomechanical coupling in F_1-ATPase revealed by simultaneous observation of nucleotide kinetics and rotation. *Nature.* 11:142-148.

7. Neuman, K.C., E.A. Abbondanzieri, R. Landick, J. Gelles, and S.M. Block. 2003. Ubiquitous Transcriptional Pausing Is Independent of RNA Polymerase Backtracking. *Cell.* 115, 437-447.
8. Charvin, G., T.R. Strick, D. Bensimon, and V. Croquette. 2005. Tracking Topoisomerase Activity at the Single-Molecule Level. *Annu. Rev. Biophys. Biomol. Struct.* 34:201-219.
9. Milescu, L.S., A. Yildiz, P.R. Selvin, and F. Sachs. Maximum likelihood restoration of dwell time sequences from molecular motor staircase data. *Submitted.*
10. QuB (www.qub.buffalo.edu, SUNY at Buffalo, Buffalo, NY).
11. Keller, D., and C. Bustamante. 2000. The Mechanochemistry of Molecular Motors. *Biophysical J.* 78:541-556.
12. Mehta, A.D., R.S. Rock, M. Rief, J.A. Spudich, M.S. Mooseker, R.E. and Cheney. 1999. Myosin-V Is A Processive Actin-Based Motor. *Nature* 400:590-593.
13. Neher, E., and B. Sakmann. 1992. The patch clamp technique. *Sci. Am.* 266:44-51.
14. 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.
15. Cox, D.R., and H.D. Miller. 1965. The Theory of Stochastic Processes. Methuen, London.
16. Jeney, S., F. Sachs, A. Pralle, E.L. Florin, and H.J.K. Horber. 2000. Statistical analysis of kinesin kinetics by applying methods from single chanel recordings. *Biophys J.* 717Pos.
17. Michalek, S., and J. Timmer. 1999. Estimating rate constants in hidden Markov models by the EM-algorithm. *IEEE Trans. Sign. Proc*. 47. 226-228.
18. Qin, F., A. Auerbach, and F. Sachs. 2000. A direct optimization approach to hidden Markov modeling for single channel kinetics. *Biophys. J.* 79:1915-1927.
19. Colquhoun, D., A.G. Hawkes, and K. Srodzinski. 1996. Joint distributions of apparent open times and shut times of single ion channels and the maximum likelihood fitting of mechanisms. *Phil. Trans. R. Soc. Lond. A.* 354:2555-2590.
20. 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.
21. Hawkes, A.G., A. Jalali, and D. Colquhoun. 1992. Asymptotic distributions of apparent open times and shut times in a single channel record allowing for the omission of brief events. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 337: 383-404.
22. Milescu, L.S., G. Akk, and F. Sachs. 2005. Maximum likelihood estimation of ion channel kinetics from macroscopic currents. *Biophysical J.* 88:2494-2515.
23. Colquhoun, D., and A.G. Hawkes. 1995. A Q-matrix cookbook: how to write only one program to calculate the single-channel and macroscopic predictions for any kinetic mechanism. *In* Single channel recording (2^{nd} ed). B. Sakmann and E. Neher, editors. Plenum Press, New York. 589-636.
24. Hille, B. 2001. Ionic Channels of Excitable Membranes, 3rd ed. Sinauer Associates, Sunderland, MA.
25. Elliott, R.J., L. Aggoun, and J.B. Moore. 1995. Hidden Markov Models: Estimation and Control. Springer-Verlag, New York.
26. Colquhoun, D., and A.G. Hawkes. 1977. Relaxation and fluctuations of membrane currents that flow through drug-operated channels. *Phil. Trans. R. Soc. Lond. B.* 199:231-262.
27. Najfeld, I., and T.F. Havel. 1995. Derivatives of the matrix exponential and their computation. *Adv. App. Math*. 16:321-375.

28. Colquhoun, D., and A.G. Hawkes. 1982. On the stochastic properties of bursts of single ion channel openings and of clusters of bursts. *Phil. Trans. R. Soc. Lond. B.* 300:1-59.
29. Qin, F., A. Auerbach, and F. Sachs. 1997. Maximum likelihood estimation of aggregated Markov processes. *Proc. R. Soc. Lond. B.* 264, 375–383.
30. Fredkin, D.R., and J.A. Rice. 1986. On aggregated Markov processes. *J. Appl. Prob.* 23:208-214.
31. Hawkes, A.G., A. Jalali, and D. Colquhoun. 1990. The distributions of the apparent open times and shut times in a single channel record when brief events can not be detected. *Phil. Trans. R. Soc. Lond. A.* 332:511-538.
32. Hawkes, A.G., A. Jalali, and D. Colquhoun. 1992. Asymptotic distributions of apparent open times and shut times in a single channel record allowing for the omission of brief events. *Phil. Trans. R. Soc. Lond. B.* 337:383-404.
33. Crouzy, S.C., and F.J. Sigworth. 1990. Yet another approach to the dwelltime omission problem of single-channel analysis. *Biophys. J.* 58:731-743.
34. Press, W.H., S.A. Teukolsky, W.T. Vetterling, and B. P. Flannery. 1992. Numerical Recipes in C. Cambridge University Press, Cambridge, UK.
35. Akaike, H. 1974. A new look at the statistical model identification. *IEEE Trans. Autom. Control* AC-19:716-723.
36. Schwartz, G. 1978. Estimating the dimension of a model. *Ann. Statist.* 6:461-464.
37. Bauer, R.J., B.F. Bowman, and J.L. Kenyon. 1987. Theory of the kinetic analysis of patch-clamp data. *Biophys. J.* 52:961-978.
38. Kienker, P. 1989. Equivalence of aggregated Markov models of ion-channel gating. *Proc. R. Soc. Lond. B. Biol. Sci.* 236:269-309.
39. Bruno, W.J., J. Yang, and J.E. Pearson. 2005. Using independent open-to-closed transitions to simplify aggregated Markov models of ion channel gating kinetics. *PNAS.* 102:6326-6331.

Figure legends

Figure 1. Myosin V and F_1-ATPase are typical molecular motors that convert chemical energy into mechanical energy by hydrolyzing ATP. **A**. Myosin V is a dimeric motor protein, walking with a hand-over-hand mechanism along the actin filament, with the stalk taking 37 nm steps per ATP hydrolyzed (2). **B**. F_1-ATPase is a rotary motor that unidirectionally turns a rotor (γ subunit) inside a stator ($\alpha_3\beta_3$ complex). The rotor turns in 120° steps, one for each ATP hydrolyzed. Each 120° step consists in an 80° ATP driven substep, followed by a 40° substep (5, 6). **C**. By attaching a probe to the motor protein, staircase data are constructed from single molecule measurements. For a linear motor (myosin V), each dwell in the staircase data represents the location of the protein at a given position along its tracks, whereas for a rotary motor (F_1-ATPase), each dwell corresponds to a given number of revolutions taken by the rotor. The duration of each dwell is random, with exponential distribution determined by kinetics and by experimental conditions, such as ATP concentration or applied mechanical force. Due to finite sampling time, more than one step may occur within the sampling interval, resulting in missed events. At each position, the motor undergoes transitions between two or more conformations. One of these transitions is ATP binding.

Figure 2. Molecular motors can be represented with reduced Markov models. **A**. The mechano-chemistry of molecular motors is a repetitive chain of identical reaction units. Each unit includes the conformations assumed by the protein while located at a given position along the cytoskeleton. The example shown is for a model with two states per reaction unit. Only transitions between states within different units can be detected experimentally. **B**. The rate matrix $\mathbf{Q}$ of the Markov model is block tridiagonal and periodic. Shown is a submatrix $\mathbf{Q}^r$ "copied" from the theoretically infinite $\mathbf{Q}$, and its block representation. Note that the first and last rows are not zero-sum. **C**. A truncated transition probability matrix $\mathbf{A}^r$ is calculated as $\mathbf{A}^r = e^{\mathbf{Q}^r \cdot dt}$. The example is shown for a sampling interval dt = 0.5 s. Note that $\mathbf{A}^r$ is also periodic. **D**. Auxiliary matrices ($\mathbf{B}_i$ and $\mathbf{S}$) used in the calculation of the discrete-time likelihood function. **E**. An example of likelihood calculation, for either the continuous-time or the discrete-time Markov model (see text for details).

Figure 3. Molecular motor kinetics can be represented by a variety of models. Shown are a few examples of kinetic models and their simulated staircase data and dwell time histograms. These models have one state per reaction unit (**A** and **E**), or two states (**B** and **C**), and have reversible (**A**, **B** and **D**), or irreversible (**C** and **E**) kinetics. Even when not explicitly stated, all models include an ATP binding state, and mechanical force-dependent transitions. Model **D** corresponds to, e.g., myosin V experiments where the fluorescent probe is attached to the motor head, resulting in double steps of 74 nm. The apparent kinetics are slower, as a single dwell consists of two consecutive steps. Model **E** corresponds to F_1-ATPase experiments, with an 80° ATP binding substep, followed by a 40° substep. Notice how the lifetime of those dwells following a 40° substep becomes shorter at higher ATP concentration. Due to finite sampling, some unitary steps are missed (e.g., the double step in trace **A**, marked with a star). Missed events could in principle occur anywhere between two samples. Notice the effect of missed events on dwell time histograms.

Figure 4. Maximum likelihood can select the correct model. Staircase data were simulated with the two state model shown in **A**, or with the single state model shown in **B**, sampled at $dt = 0.5$ s. For each simulation model, the true rate constants are given, calculated as $k_{ij} = n_{ij} / T_i$, where n_{ij} is the actual number of transitions $i \rightarrow j$, and T_i is the actual time spent in state i, as randomly chosen by the simulation routine. Each simulation was then maximum likelihood-fit with either model. The *pdf* curves for the ML estimates are shown overlapped on the dwell time histograms, without correction for missed events. In **A**, the log-likelihood *LL* is significantly greater for the correct model, even considering the additional two free parameters, while the *pdf* lines overlap almost perfectly. In **B**, the two models have virtually identical *LL* and *pdf*. Notice that estimates obtained with the correct model are very accurate. In **B**, estimates obtained with the wrong model have a very large SE. Notice how, due to missed events, the histogram appears to be slightly shifted to the right, relative to the ideal *pdf* curve.

Figure 5. Statistical distribution of maximum likelihood kinetic estimates. Five thousand dwell sequences, each 5000 points long, sampled at $dt = 0.5$ s, were simulated with the model shown in Fig. 2A and were globally fit in groups of 2…1000. **A**. Rate constants are Gaussian distributed with a width proportional to the number of data points globally fit (two traces in A1, and 20 traces in A2). The distribution of k_B has much higher variance, and is bimodal when only two traces were fit (A1). This is explained by the scarcity of k_B transitions: only ≈ one per segment, compared to more than ten per segment for all other rates. **B**. There is no apparent cross-correlation between the estimates of different parameters. The example shown is for the global fit of 20 traces. The implication is that if one parameter cannot be reliably estimated (i.e., k_B), it will not lower the precision of the other estimates. The solid circles mark the correct values. **C**. All four rate constants were estimated without bias, when at least 10 traces were globally fit, i.e., at least 10…100 transitions for each rate. The SE of all but k_B estimates was below 10%, when at least 10 traces were globally fit. Considering the actual number of simulated transitions, all parameters are estimated with similar precision.

Figure 6. Comparison between discrete-time and continuous-time maximum likelihood algorithms. One hundred data sets, each one second long, sampled at 10 Hz, were simulated with the non-periodic model "closed-open-closed" ($k_{12} = 0.1$ s^{-1}, $k_{21} = 0.05$ s^{-1}, $k_{23} = 5.0$ s^{-1}, and $k_{32} = 2.0$ s^{-1}), and were maximum likelihood-fit individually. We tested the following algorithms: MIP (discrete-time, presented in this paper), MPL (discrete-time) (18), and MIL (continuous-time, first order missed events correction) (20). **A**. Cross-correlation plots show that estimates obtained with MIP (x axis) match almost perfectly the estimates obtained from the same data set with MPL (y axis), and are well correlated with the estimates obtained with MIL (y axis). Notice that MIL's estimates have greater variance than those obtained with the other two algorithms. **B**. Estimates obtained with MIL may depend critically on the choice of dead time parameter for missed event correction. In the example shown, the best values were obtained with a dead time ≈ $1.6 \times dt$, where dt is the sampling interval. MIP and MPL do not require missed event correction. The dotted lines mark the true parameter values.

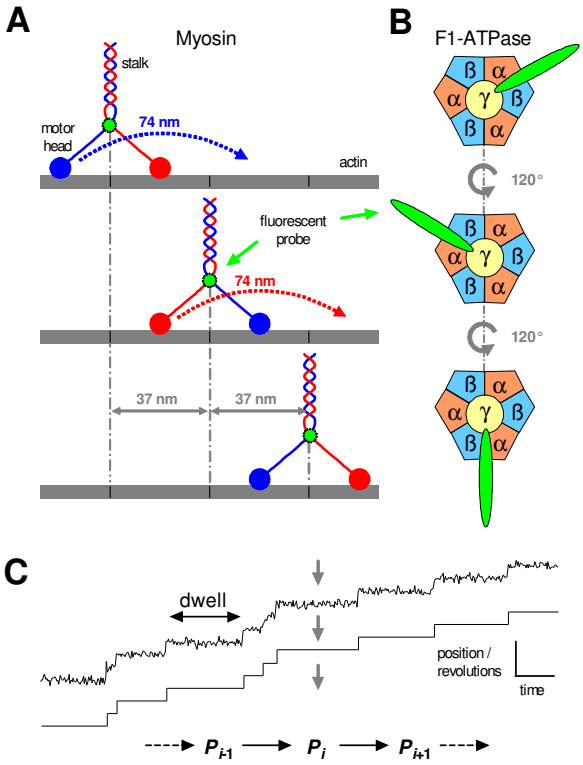
A
Myosin
stalk
motor
head
74 nm
actin
fluorescent
probe
74 nm
37 nm
37 nm
B
F1-ATPase
120°
120°
C
dwell
position /
revolutions
time
P_{i-1}
P_i
P_{i+1}

Figure 1.

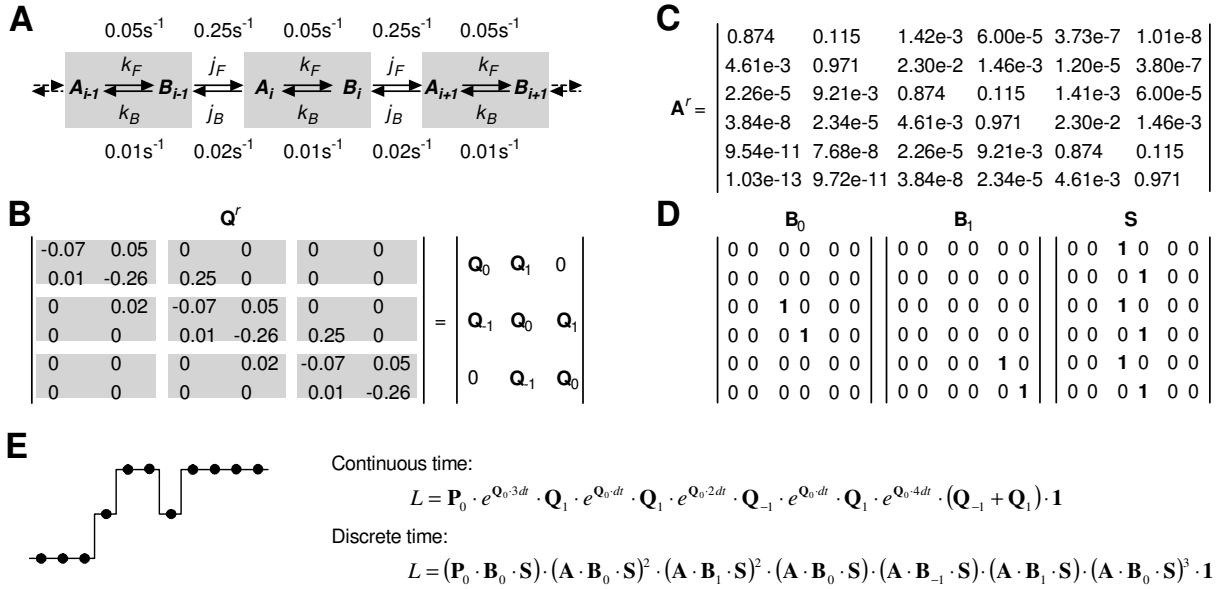
A
0.05s-1 0.25s-1 0.05s-1 0.25s-1 0.05s-1
kF jF kF jF kF
Ai-1 Bi-1 Ai Bi Ai+1 Bi+1
kB jB kB jB kB
0.01s-1 0.02s-1 0.01s-1 0.02s-1 0.01s-1
C
Ar =
0.874 0.115 1.42e-3 6.00e-5 3.73e-7 1.01e-8
4.61e-3 0.971 2.30e-2 1.46e-3 1.20e-5 3.80e-7
2.26e-5 9.21e-3 0.874 0.115 1.41e-3 6.00e-5
3.84e-8 2.34e-5 4.61e-3 0.971 2.30e-2 1.46e-3
9.54e-11 7.68e-8 2.26e-5 9.21e-3 0.874 0.115
1.03e-13 9.72e-11 3.84e-8 2.34e-5 4.61e-3 0.971
B
Qr
-0.07 0.05 0 0 0 0
0.01 -0.26 0.25 0 0 0
0 0.02 -0.07 0.05 0 0
0 0 0.01 -0.26 0.25 0
0 0 0 0.02 -0.07 0.05
0 0 0 0 0.01 -0.26
=
Q0 Q1 0
Q-1 Q0 Q1
0 Q-1 Q0
D
B0
0 0 0 0 0 0
0 0 0 0 0 0
0 0 1 0 0 0
0 0 0 1 0 0
0 0 0 0 0 0
0 0 0 0 0 0
B1
0 0 0 0 0 0
0 0 0 0 0 0
0 0 0 0 0 0
0 0 0 0 0 0
0 0 0 0 1 0
0 0 0 0 0 1
S
0 0 1 0 0 0
0 0 0 1 0 0
0 0 1 0 0 0
0 0 0 1 0 0
0 0 1 0 0 0
0 0 0 1 0 0
E
Continuous time:
$L = \mathbf{P}_0 \cdot e^{\mathbf{Q}_0 \cdot 3dt} \cdot \mathbf{Q}_1 \cdot e^{\mathbf{Q}_0 \cdot dt} \cdot \mathbf{Q}_1 \cdot e^{\mathbf{Q}_0 \cdot 2dt} \cdot \mathbf{Q}_{-1} \cdot e^{\mathbf{Q}_0 \cdot dt} \cdot \mathbf{Q}_1 \cdot e^{\mathbf{Q}_0 \cdot 4dt} \cdot (\mathbf{Q}_{-1} + \mathbf{Q}_1) \cdot \mathbf{1}$
Discrete time:
$L = (\mathbf{P}_0 \cdot \mathbf{B}_0 \cdot \mathbf{S}) \cdot (\mathbf{A} \cdot \mathbf{B}_0 \cdot \mathbf{S})^2 \cdot (\mathbf{A} \cdot \mathbf{B}_1 \cdot \mathbf{S})^2 \cdot (\mathbf{A} \cdot \mathbf{B}_0 \cdot \mathbf{S}) \cdot (\mathbf{A} \cdot \mathbf{B}_{-1} \cdot \mathbf{S}) \cdot (\mathbf{A} \cdot \mathbf{B}_1 \cdot \mathbf{S}) \cdot (\mathbf{A} \cdot \mathbf{B}_0 \cdot \mathbf{S})^3 \cdot \mathbf{1}$

Figure 2.

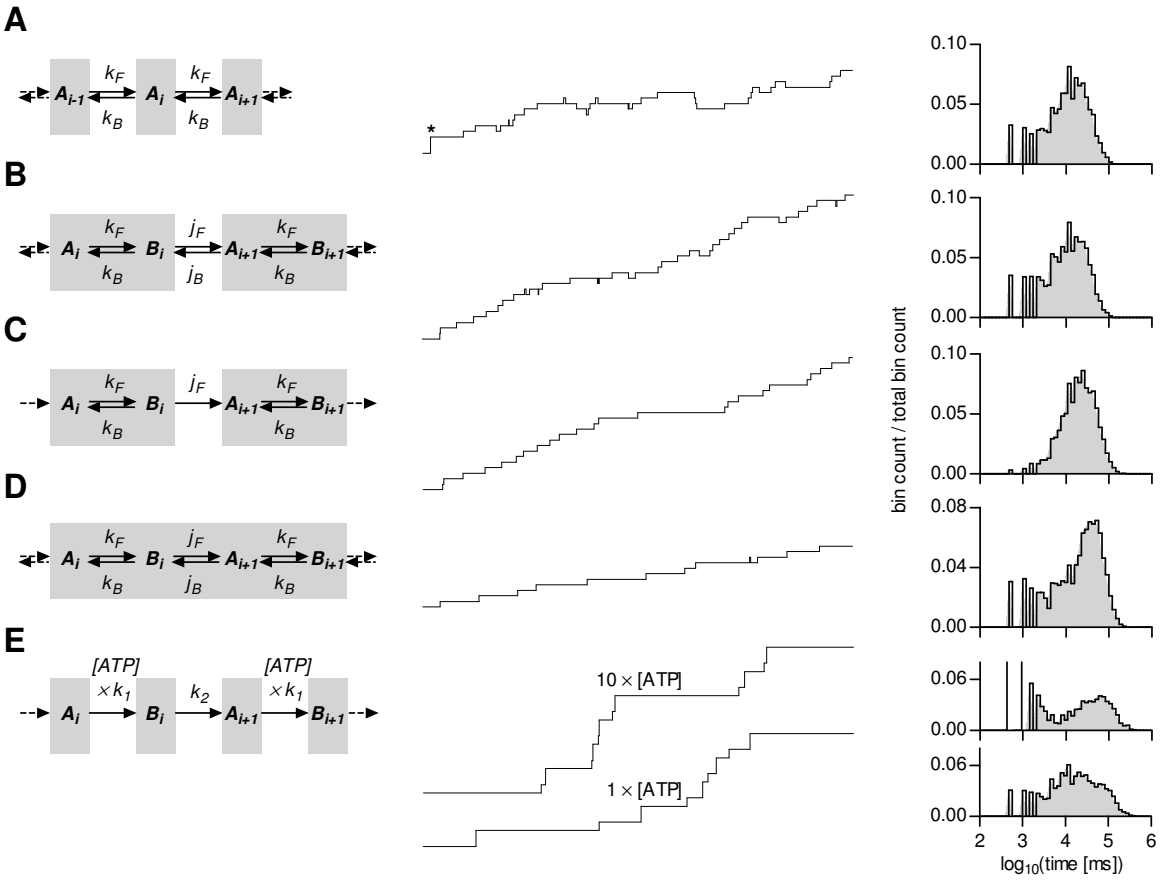
A
B
C
D
E
[ATP]
×k1
k2
10 × [ATP]
1 × [ATP]
bin count / total bin count
log10(time [ms])

Figure 3.

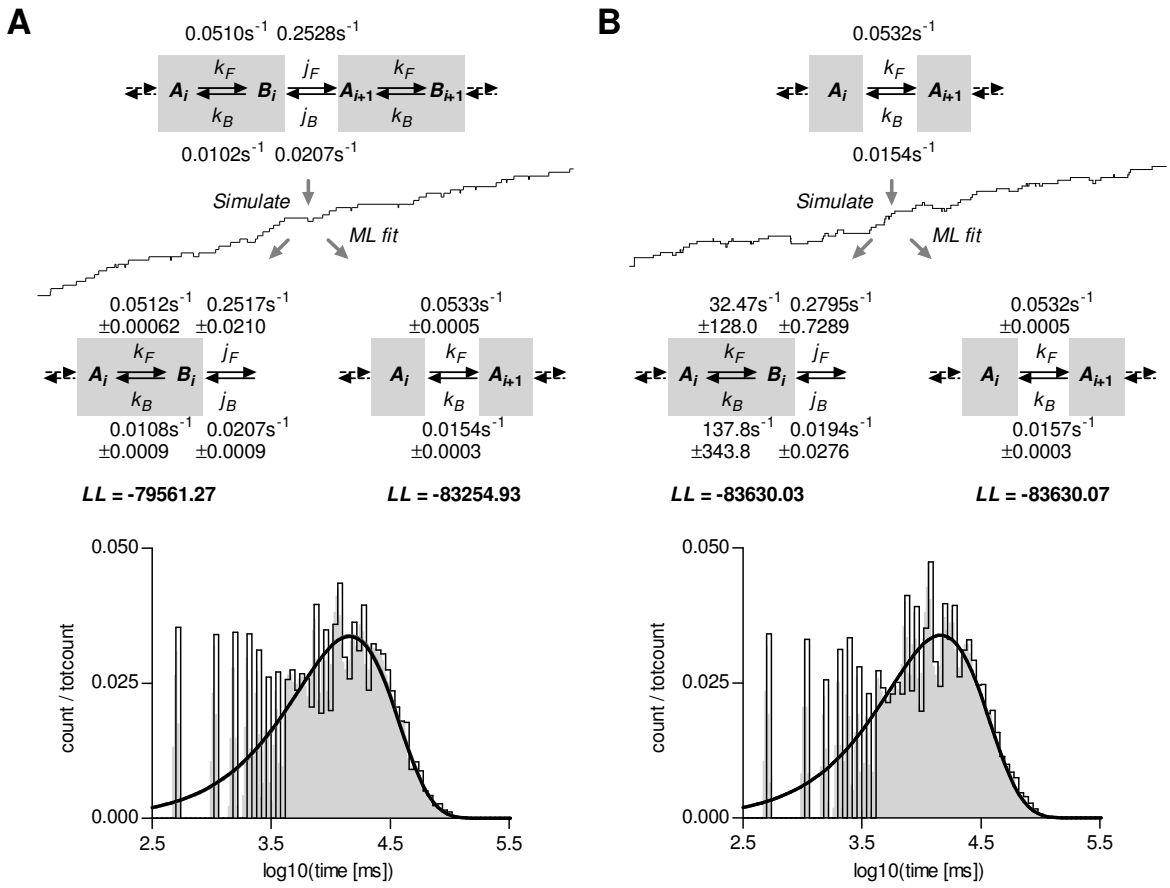
A
0.0510s-1 0.2528s-1
kF
Ai
Bi
kB
jF
jB
Ai+1
Bi+1
0.0102s-1 0.0207s-1
Simulate
ML fit
0.0512s-1 0.2517s-1
±0.00062 ±0.0210
0.0108s-1 0.0207s-1
±0.0009 ±0.0009
0.0533s-1
±0.0005
0.0154s-1
±0.0003
LL = -79561.27
LL = -83254.93
0.050
0.025
0.000
count / totcount
2.5
3.5
4.5
5.5
log10(time [ms])
B
0.0532s-1
0.0154s-1
32.47s-1 0.2795s-1
±128.0 ±0.7289
137.8s-1 0.0194s-1
±343.8 ±0.0276
0.0532s-1
±0.0005
0.0157s-1
±0.0003
LL = -83630.03
LL = -83630.07

Figure 4.

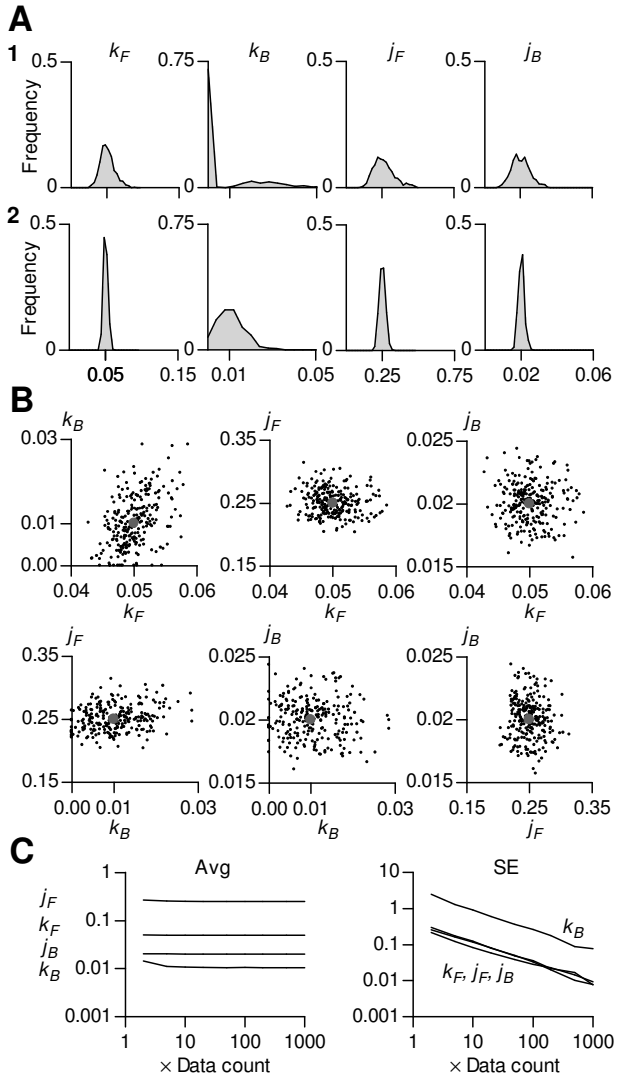

Figure 5.

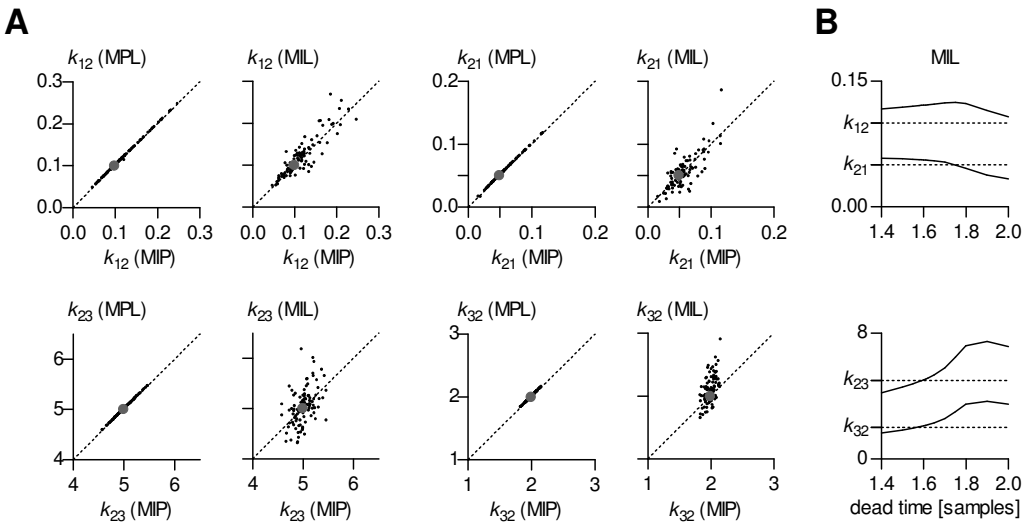
A
B
k_{12} (MPL)
k_{12} (MIL)
k_{21} (MPL)
k_{21} (MIL)
k_{12} (MIP)
k_{21} (MIP)
k_{23} (MPL)
k_{23} (MIL)
k_{32} (MPL)
k_{32} (MIL)
k_{23} (MIP)
k_{32} (MIP)
MIL
dead time [samples]

Figure 6.