Challenges in inferring directionality of active molecular processes from single-molecule FRET trajectories.

Aljaž Godec[1*] and Dmitrii E. Makarov[2,3,*]

[1]Mathematical bioPhysics Group, Max Planck Institute for Multidisciplinary Sciences, 37077 Göttingen, Germany

[2]Depatment of Chemistry and [3]Oden Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, TX 78712, USA

Abstract

We discuss some of the practical challenges that one faces in using stochastic thermodynamics to infer directionality of molecular machines from experimental single-molecule trajectories. Because of the limited spatio-temporal resolution of single-molecule experiments and because both forward and backward transitions between same pairs of states cannot always be detected, differentiating between the forward and backward directions of, e.g., an ATP-consuming molecular machine that operates periodically, turns out to be a nontrivial task. Using a simple extension of a Markov-state model that is commonly used to analyze single-molecule transition-path measurements, we illustrate how irreversibility can be hidden from such measurements, but in some cases can be uncovered when non-Markov effects in low-dimensional single-molecule trajectories are considered.

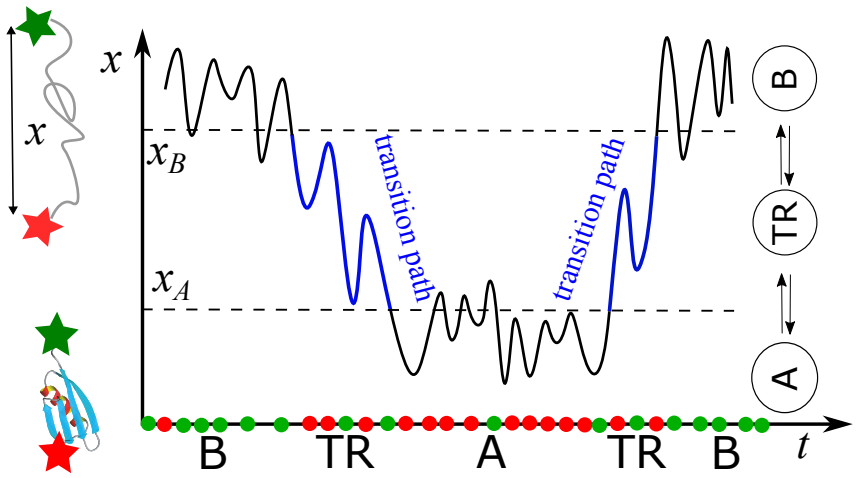
*Corresponding authors, agodec@mpinat.mpg.de, makarov@cm.utexas.edu

Authors' biographies:

Aljaž Godec obtained his PhD in theoretical physics in 2012 from the University of Ljubljana (Slovenia) and afterwards moved to a postdoc at the University of Potsdam (Germany) as an Alexander von Humboldt fellow. Since 2017 he is head of the research group "Mathematical bioPhysics" at the Max-Planck-Institute for Multidisciplinary Sciences in Göttingen (Germany). His research focuses on the statistical and mathematical physics of out-of-equilibrium systems, in particular on stochastic thermodynamics and emergent phenomena in soft matter and biophysics.

Dmitrii E Makarov received a PhD in theoretical physics from the Institute for Chemical Physics (Moscow, Russia) in 1992. Since 2001 he has been on the faculty at the Department of Chemistry at the University of Texas at Austin. He is also a core faculty at the Oden Institute for Computational Engineering & Sciences. His research interests are in the field of computational and theoretical chemical physics, with current research topics ranging from quantum reaction rate theory to molecular biophysics.

TOC graphic



1. Introduction

Nonequilibrium phenomena have direction or "arrow of time": that is, a movie of a nonequilibrium process played backwards will look statistically different from the original movie. The directionality or time-irreversibility of life is evident at the macroscopic scale. If, however, we zoom in on the microscopic motion of molecules that constitute a living organism, such directionality becomes far from evident. For example, the Maxwell-Boltzmann distribution of molecular velocities – a fingerprint of equilibrium and thus of time-reversible state of matter – not only enables us to discuss our body's temperature but also ensures the validity of the Arrhenius law that biochemists unabashedly use to describe the rates of elementary steps in nonequilibrium biochemical networks.

The question of whether specific components of biomolecular machinery undergo equilibrium or nonequilibrium dynamics has been the subject of some controversy. For example, while a molecular motor walking along its track clearly steps more often in one direction than the other, it has been argued that individual steps are time-reversible and follow the same (average) path in the forward and backward direction[1]. Yet we know that such microscopic reversibility must be violated at some length scale to give rise to directional persistence. For example, according to the "scallop theorem", a bacterium could not swim[2] if its movements were time-reversible (at least in incompressible fluids). Another obvious example is cell division.

The problem of detecting directionality at a molecular level is two-fold: first, adequate experimental tools are needed. Because molecular machines usually operate under nonequilibrium steady-state conditions (that is, concentrations of various molecules in the cell

do not change in time), single-molecule experiments are usually required to study their dynamics[3]. Despite single-molecule experiments becoming nearly routine tools of molecular biology, only a small number of studies[4, 5] have attempted to address the question of directionality of biochemical phenomena and/or employed nonequilibrium models[6] to describe single-molecule data except when directionality is self-evident (e.g. the rotation or walking of a molecular motor along its track). Second, proper data analysis is needed. There have been many recent efforts to identify and quantify non-equilibrium effects in living systems[7-18], yet, this has proven difficult even in mesoscopic systems[7-9], where the dynamics of the system can often be observed directly and quite accurately, e.g., with a fast camera. Inferring and quantifying irreversible behavior at smaller, molecular scales poses additional difficulties, as the time- and length-scales of the probes employed by single-molecule experiments often overlap with the time- and length-scales of the phenomena probed[19]. In this Perspective, we argue that those limitations are especially important in single-molecule studies that probe nonequilibrium steady-state phenomena, describe challenges faced when applying ideas of stochastic thermodynamics[20, 21] to single-molecule data, and outline promising theoretical and experimental approaches to overcoming these challenges.

2. Limitations of single-molecule measurements in application to active molecular processes.

2.1. Single-molecule measurements report on projected dynamics. Single-molecule experiments have the ability to probe conformational changes in individual molecules in real time. Such measurements, however, often suffer from limited spatial and temporal resolution, and they inevitably entail a coarse-grained view of the process: in fluorescence resonance

energy transfer (FRET) studies[22-24], for example, the ability to resolve a microscopic state of a molecule is limited by the number of individual photons that can be detected while the molecule resides in this state; moreover, experimentalists usually only have a limited number of distinct photon colors (usually two) available in order to differentiate among multiple states. Likewise, in single-molecule force spectroscopy studies the dynamics of the molecular coordinate of interest, such as the distance between two residues in a protein, have to be inferred from the displacement of a sluggish force probe[25-30].

These limitations are especially important when single-molecule techniques are applied to molecular machines[3] that operate away from equilibrium[31]. Sometimes the ensuing directionality (i.e., time irreversibility) in their motion is immediately evident from data: for example, a molecular motor can be observed walking in a particular direction[32]. But there are also many cases where this directionality is more subtle. For example, the dynamics of HSP90 is driven by ATP hydrolysis, yet the motion is periodic and observed along a one-dimensional reaction coordinate – thus its time irreversibility is not straightforward to discern[4, 33]. To illustrate the difficulties that arise from limited spatio-temporal resolution, consider a minimal hypothetic 3-state kinetic scheme of the type that has, for example, been used to describe the operation of the disaggregation machine ClpB[34] (Fig. 1).

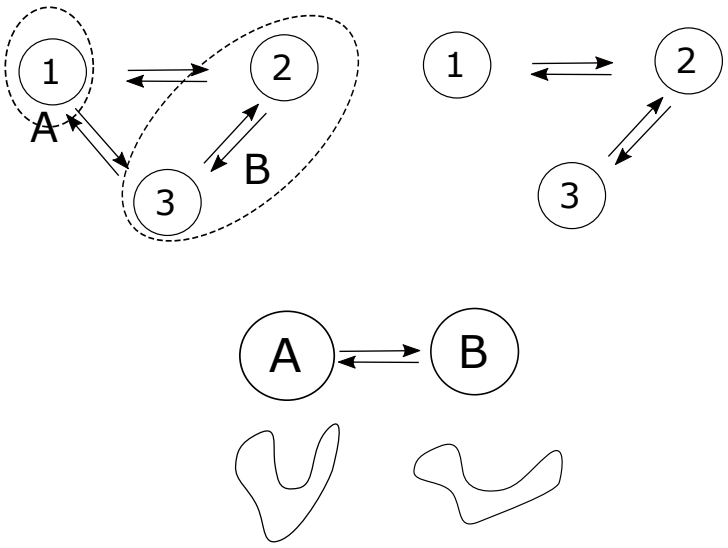


Figure 1. Transitions between, say, closed (A) and open (B) conformations of a protein complex (bottom) may be driven by a cycle $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$, yet if all of its states cannot be resolved experimentally then it may be difficult to differentiate between a nonequilibrium process (top, left) and an equilibrium one (top, right) even if hidden Markov analysis is used. Here we have assumed that states 2 and 3 correspond to a single observable state B.

The observed process is a conformational rearrangement between two mesoscopic states A and B, but the microscopic underlying process is described by a 3-state kinetic scheme driven by, e.g., coupling to ATP hydrolysis. According to Kolmogorov's cycle criterion[35], the dynamics of the system is time-reversible only if, for the cycle $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$, the product of the forward rate coefficients is the same as the product of the backward rate coefficients, i.e. when the ratio

$$e^{X} = \frac{k_{1\rightarrow 2}k_{2\rightarrow 3}k_{3\rightarrow 1}}{k_{2\rightarrow 1}k_{3\rightarrow 2}k_{1\rightarrow 3}} \quad (1)$$

is equal to 1. Here $k_{i\to j}$ denotes the (pseudo-)first-order rate coefficient for the transition from i to j. If this transition is coupled to, e.g., ATP hydrolysis then $k_{i\to j}$ would be proportional to the ATP concentration such that $e^X = 1$ only when the concentrations of molecules participating in the process assume their equilibrium values. The quantity X (which, in the language of stochastic thermodynamics, is sometimes called the affinity of the cycle $1 \to 2 \to 3 \to 1$) quantifies the thermodynamic force driving the process[31]. When $X > 0$, the steady state of the system entails overall motion in the clockwise direction. This directional motion can be further characterized by a nonzero flux J, which is equal to the total number of overall cycles completed per unit time.

The degree of irreversibility of the dynamics is usually quantified by the mean entropy production rate, which is a measure of both the heat dissipated to the environment and of how statistically different the forward process is from its time reverse[20, 21]. For Markovian dynamics, the entropy production rate is given by[20, 21]

$$\langle \dot{S} \rangle = \lim_{\tau\to\infty} \frac{1}{\tau} \left\langle \frac{P[x(t)]}{P[x(\tau-t)]} \right\rangle, \qquad (2)$$

where $P[x(t)]$ is the probability of a forward path $x(t)$, $t \in [0, \tau]$, $P[x(\tau - t)]$ the probability of its time reverse, Boltzmann's constant is set to 1, and the angular brackets indicate averaging over the ensemble of trajectories. For continuous-time Markov jump dynamics, the entropy production is generally given by[20]

$$\langle \dot{S} \rangle = \frac{1}{2} \sum_{i\neq j} \left(k_{i\to j} P_i - k_{j\to i} P_j \right) \ln \frac{k_{i\to j} P_i}{k_{j\to i} P_j}, \quad (3)$$

where P_i is the steady-state probability of finding the system in state i. In the case of a single cycle as in Fig. 1, this becomes

$\langle \dot{S} \rangle = JX$, (4)

We note that, more generally, Kolmogorov's cycle criterion demands that the product of the forward rates is equal to the product of the backward rates (and thus $X = 0$) for *any* cycle present in the system in order for its dynamics to be time-reversible.

We now suppose that the details about transitions between the three microscopic states are hidden from the observer, who only has access to two coarser states A and B. In Fig. 1, it is assumed that the observer cannot differentiate between states 2 and 3, which are therefore lumped to a state B. What can we say about the underlying process (and particularly about it's irreversible, driven nature) by observing transitions between A and B[36, 37]?

It is intuitively obvious that the direction of the process is much easier to discern from the sequence of the "microscopic" 3 states that it visits than from the corresponding coarse two-state sequence: for example the sequence 123123123123123… is obviously irreversible (i.e. distinct from its temporal reverse 321321321321…) while the corresponding sequence ABABABABAB… is reversible. For an arbitrary random sequence of states A and B, any transition from A to B is followed by a transition from B to A, and so the corresponding unidirectional fluxes (i.e., mean numbers of transitions per unit time), $J_{A\to B}$, and $J_{B\to A}$, are identical, with zero overall flux J. Using these fluxes, one can introduce the rate coefficients $k_{A\to B} = J_{A\to B}/P_A$, $k_{B\to A} = J_{B\to A}/P_B$, where $P_{A(B)}$ is the fraction of time spent in A(B) and model the two-state trajectory with jump rates $k_{A\to B}$ and $k_{B\to A}$ – at this level of description the process is manifestly reversible, and the measured entropy production rate is zero. Any information about the irreversible character of the original process is lost. This argument, however, does not imply that the irreversible character of the true dynamics cannot be

recovered from the coarse-grained two-state trajectory, only that the Markov description of this trajectory based on the average fluxes or average dwell times in each state does not suffice: non-Markov effects (i.e. memory) must be included in the model[36, 38, 39]. Given the experimental limitations further outlined below, however, experimental detection of memory effects is far from straightforward.

The fact that the Markovian kinetic scheme $A \rightleftarrows B$ is fundamentally time-reversible has further implications. Usually, the Markov approximation is justified for the observables that are slow, slower than all "hidden" degrees of freedom that equilibrate much faster. In application to the kinetic scheme in Fig. 1 (top, left), the states 2 and 3 that constitute the compound state B could be lumped together and the scheme could be reduced to the two-state scheme (Fig. 1, bottom) if the interconversion dynamics $2 \rightleftarrows 3$ within B is fast. But the detailed balance may still be violated by the full kinetic model, while it is not violated by the reduced two-state model – in this case irreversibility of the true dynamics is hidden from observation[40]. We will see another example of this kind in Section 3.2.

2.2. Only some transitions are observable. In interpreting single-molecule measurements, it is often beneficial to consider the combined dynamics of the molecule of interest and of the experimental probe. Here we illustrate this idea for a FRET experiment[41-43] (further discussed below), in which the internal state of the molecule (1, 2 or 3 in Fig. 2) is inferred indirectly from the colors and arrival times of photons emitted by two fluorescent labels called donor (D) and acceptor (A). Photon-by-photon analysis of such an experiment, which is especially important when there is no clear separation between photophysical and biochemical timescales, can be

accomplished by considering the combined states of the system specified by the molecule's state and the state of the donor and the acceptor[41], each of which can either be in the ground (D,A) or excited state (D*,A*). In the resulting kinetic network of combined photophysical and molecular states (Fig. 2), some transitions are observable directly while others are not. For example, transitions such as 1D*A→ 1DA or 3DA*→3DA are detected because the donor or acceptor emits a photon of a specific color. A FRET transition such as 1D*A→1DA*, in which excitation energy is transferred from the donor to the acceptor, is, however, unobservable directly: it can only be deduced from subsequent emission of a photon by the acceptor. Likewise, a transition in which the donor is excited by absorbing a photon from a light source is unobservable.

Detection of directionality and estimation of entropy production in such "partially observed" kinetic networks has been the subject of considerable recent interest in stochastic thermodynamics. A particularly promising approach to this problem[36, 44], focusing on the waiting times between *observed transitions* (here, photon emission events) rather than on dwell times in network states or first passage times to network states, can be used to infer quantities such as the entropy production. Unfortunately, the use of such an approach requires that, if transitions from one network state to another are observed, the reverse transitions are also observed[36], which, strictly speaking, is not true for FRET experiments: For example, while emission of a photon is observed, the reverse transition (i.e. photoexcitation) is not.

We therefore conclude that inferring directionality of molecular phenomena from single-molecule data poses both fundamental and practical open problems: on a fundamental side we need to understand how much information about directionality of the underlying

process is retained in projected/partially observed dynamics; on a practical side, we need workable tools for estimating the entropy production and related quantities from such partially observed dynamics. While progress toward these goals has been achieved in the field of stochastic thermodynamics, it is not clear whether such existing approaches are directly applicable to single-molecule FRET data given their limitations described above.

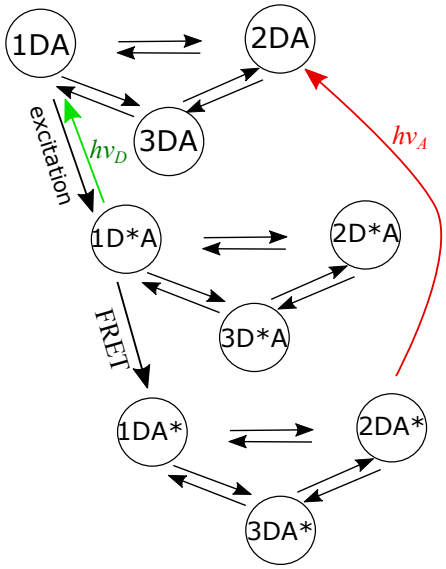


Figure 2. A combined network representing dynamics of a molecule (here with 3 internal states) along with the photophysics of the two fluorescent dyes (donor D and acceptor A) that report on the molecule's state. Some of the possible transitions within such a network are shown, including laser excitation, FRET (transfer of excitation from excited donor D* to acceptor), and emission by either donor or acceptor, are shown. Note that the arrow lengths do not necessarily correspond to the respective transition rates.

3. Directionality of molecular processes from transition paths.

***3.1. Transition paths: definition and importance*.** In what follows, we focus on the more specific problem of inferring directionality of molecular processes via measurements of transition paths. Transition paths have recently become a subject of considerable interest in single-molecule biophysics[23, 25]. A transition path is a short segment of a molecular trajectory that accomplishes a successful conformational transition (Fig. 3). More precisely, one considers the dynamics along a molecular reaction coordinate x of interest (such as the distance between two parts of a biopolymer labeled with two fluorescent dyes of different colors, Fig. 3) and associates the state A with all configurations with $x < x_A$, B with configurations with $x > x_B$, and a "transition region" with $x_A < x < x_B$. Here $x_{A,B}$ are the transition region boundaries that can (in principle) be chosen arbitrarily. A transition path enters the transition region through one boundary and exits through the other without escaping the transition region in between. The transition-path time thus does not include the time spent outside the transition region or the temporal duration of failed attempts to cross the transition region; it is the temporal duration of the successful transition itself. Importantly, while within the standard picture of chemical kinetics the transitions are deemed to be instantaneous, a kinetic description that accounts for transition path times is more general than a kinetic model with instantaneous jumps between states[38, 45, 46].

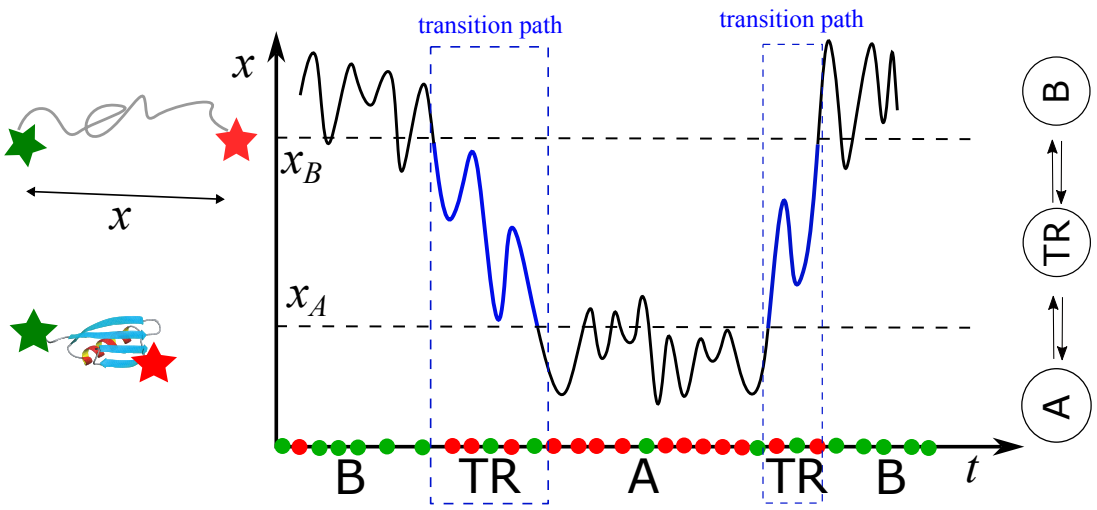


Figure 3. Transition paths are segments of a molecular trajectory that traverse a specified transition region staying continuously in this region (shown in blue). Experimentally, the transition region can be defined by requiring that the molecular reaction coordinate of interest x belongs to a specified interval (x_A, x_B). Such a coordinate, however, is often not directly observable. In 2-color FRET experiments, for example, the molecular trajectory $x(t)$ is deduced from the arrival times of photons of two colors; in general, there is no one-to-one correspondence between the current value of x and the color of the photon emitted, and extraction of transition path times from photon sequences usually involves maximum likelihood analysis using a discrete model in which the continuous trajectory is replaced by dynamics with only 3 states, A, B, and a "transition-region" intermediate TR.

Transition paths encapsulate the transition mechanisms that biochemists and biophysicists strive to glean from experimental data. For instance, transition path times contain information about whether or not the dynamics of the experimental observable is Markovian[38, 47, 48], whether transition involves multiple pathways that are unobservable directly[38, 47], or whether an on-pathway transient intermediate is visited by transition paths[49, 50]. For

nonequilibrium systems, examination of transition paths informs one about (ir)reversibility of the process in question[39, 51-54]. For example one may ask whether the pathway taken by a molecular motor during a forward step is different from that taken during a backward step, regardless of whether the motor is making more steps in one direction than the other thereby undergoing overall unidirectional motion[1]. At the same time, the relatively short temporal duration of transition paths illustrates the challenge encountered when there is no clear time-scale separation between the dynamics of the probes employed (e.g. photophysics of the FRET donor-acceptor pair) and the intrinsic biochemical dynamics that one desires to measure (see Section 2.2).

3.2. Experimental limitations in transition path measurements. The detailed molecular trajectory of interest, $x(t)$, is often inaccessible experimentally (at least with an unlimited spatio-temporal resolution) but, instead, is deduced from a different experimental observable. In 2-color FRET experiments, for example, one infers $x(t)$ from the arrival times of photos emitted by two fluorescent dyes. The time resolution is then obviously limited by the inter-photon lag times, but there is a further complication: in general, there is no one-to-one mapping between the color of the photon emitted at time t and the value $x(t)$, because the latter only determines the probabilities of detecting photons of different colors. Inferring whether the molecule is in states A, B, or is on a transition path, therefore, usually requires additional severe approximations and assumptions. So far, this problem has been tackled by replacing the continuous dynamics $x(t)$ with a discrete model that only includes 3 states: A, B, and the fictitious "transition region" state (TR), as shown in Fig.3[50, 55]. The model further

assumes that the dynamics of this 3-state model is Markovian, and thus hidden-Markov analysis of photon sequences allows one to map them onto to 3-state trajectories and to approximate the transition path time as the lifetime of the state TR[43, 56]. We note that, by construction, this model describes an equilibrium system satisfying detailed balance, as it lacks cycles. As argued above, nonequilibrium dynamics can be captured within such a model only if the Markov assumption is relaxed and memory effects are included.

We now illustrate how nonequilibrium effects are manifested in the observed transition-path dynamics using a simple nonequilibrium extension of the model used to describe experimental measurements[55], in which the "transition-region" state is comprised of two microscopic states, 1 and 2 (Fig. 4), which cannot be differentiated experimentally. The transition between the metastable states A and B can thus proceed via two parallel pathways. As a result of lumping two microscopic states into one coarser observable state TR, *the dynamics of the resulting 3- system that consists of states A, B, and TR, is no longer Markovian*, and the distributions of transition path times reflect this fact.

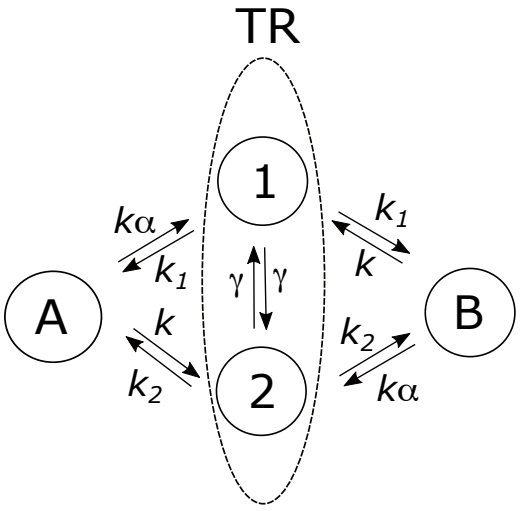


Figure 4. A two-pathway model, with the transition region (TR) comprised of two states, 1 and 2, predicts broad distributions of transition path times and a breakdown of forward/backward symmetry of transition paths when the system does not satisfy detailed balance.

The distributions of transition-path times from A to B and from B to A are generally given by[39]

$$p_{A\to B}(t) = \frac{2\alpha}{\alpha+1}[k_1 G_{11}(t) + k_2 G_{21}(t)] + \frac{2}{\alpha+1}[k_1 G_{12}(t) + k_2 G_{22}(t)]$$

$$p_{B\to A}(t) = \frac{2}{\alpha+1}[k_1 G_{11}(t) + k_2 G_{21}(t)] + \frac{2\alpha}{\alpha+1}[k_1 G_{12}(t) + k_2 G_{22}(t)] \quad (5)$$

where $k_{1(2)}$ is the rate of escape from the corresponding intermediate to either side (Fig. 4), and where

$$\begin{pmatrix} G_{11}(t) & G_{12}(t) \\ G_{21}(t) & G_{22}(t) \end{pmatrix} = \exp\left[\begin{pmatrix} -2k_1 - \gamma & \gamma \\ \gamma & -2k_2 - \gamma \end{pmatrix} t\right] \quad (6)$$

is the "transition region Green's function"[39]. We will first consider the case where there are no transitions within the transition region ($\gamma = 0$ in Fig. 4). For transition paths proceeding via intermediate 1 (2) the (conditional) distribution of the transition path time is exponential,

$$p_{1(2)}(t) = 2k_{1(2)}G_{11}(t) = 2k_{1(2)}e^{-2k_{1(2)}t}, \quad (7)$$

For the transitions from A or B into the transition region, the clockwise transitions take place with a rate $k\alpha$ and counterclockwise transitions with a rate k, the parameter α being a measure of the nonequilibrium driving force. As a result, for the cycle $A \to 1 \to B \to 2 \to A$ the rate of the forward rate coefficients differs from that of the backward ones, with an overall driving force,

$$e^X = \alpha^2, \quad (8)$$

and with equilibrium attained only when $\alpha = 1$.

The distributions of transition path times from A to B and B to A are linear combinations of the contributions from each pathway. Using the fact that the probability to accomplish a transition from A to B via the pathway 1 is $\alpha/(\alpha+1)$ etc., the distributions of transition path times in each direction are

$$p_{A\to B}(t) = \frac{\alpha p_1(t) + p_2(t)}{\alpha + 1}$$

$$p_{B\to A}(t) = \frac{p_1(t) + \alpha p_2(t)}{\alpha + 1},$$

(9)

a result that also follows directly from Eqs. 5-6. From Eqs. 7 and 9 several observations follow. First, unless one has $k_1 = k_2$, the distributions of transition path times are always broader than exponential. Specifically, the coefficients of variation $C_{A\to B}$ and $C_{B\to A}$, equal to the ratios of the distribution variances to their means, e.g.,

$$C_{A\to B} = \sqrt{\frac{\langle t^2_{A\to B}\rangle - \langle t_{A\to B}\rangle^2}{\langle t_{A\to B}\rangle^2}}, \langle t^n_{A\to B}\rangle = \int_0^\infty dt t^n p_{A\to B}(t), \quad (10)$$

are greater than 1, a value expected for a single-exponential distribution. As was shown recently[38, 47], such a broad distribution is impossible for any linear Markov kinetic scheme and, in particular, for the single-intermediate pathway model (Fig. 3) commonly used to infer experimental transition path times TR[50, 55]. Thus, the analysis of experimental transition path times with a broad distribution in terms of a single pathway mechanism would be internally inconsistent. On the other hand, we note that experimental observation of such a broad distribution would suggest that more than one pathway exists *even if the two intermediate states, 1 and 2, have the same FRET signature and thus cannot be distinguished directly*.

Second, the forward/backward symmetry is broken ($p_{A\to B}(t) \neq p_{B\to A}(t)$) when the system is

out of equilibrium, e.g. when $\alpha \neq 1$ (Fig. 4). Again, such a situation cannot be captured if the experimental distributions of transition path times $p_{A\to B}(t)$ and $p_{B\to A}(t)$ are interpreted in terms of a single-intermediate model or any linear Markov model such as the one shown in Fig. 3 right. Third, notice that increased driving (larger α) implies greater irreversibility (Fig. 4). In particular, in the limit of strong driving, $\alpha \gg 1$, the A-to-B transition preferentially proceeds via intermediate 1 and B-to-A transition proceeds via intermediate 2, with the corresponding distributions approaching $p_{A\to B}(t) \to p_1(t)$, $p_{B\to A}(t) \to p_2(t)$. As will be seen below, however, stronger driving does not necessarily imply greater asymmetry in transition path times.

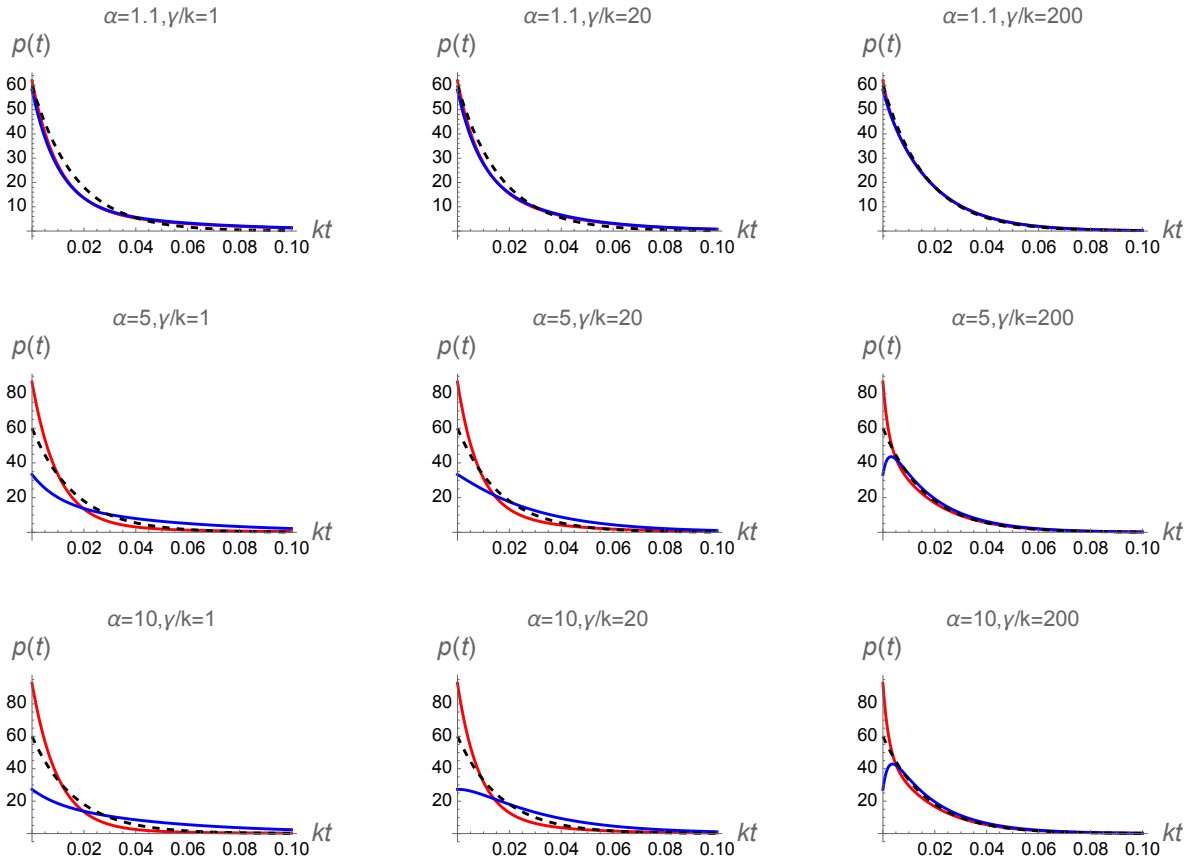


Figure 5. The distributions $p_{A\to B}(t)$ (red) and $p_{B\to A}(t)$ (blue) of the transition path times from left to right and from right to left in according to the kinetic scheme of Fig. 4. Here $k_1 = 50k$, $k_2 = 10k$, and the values of the parameters α and γ are indicated in each plot. The dashed black lines indicate the $\gamma \to \infty$ limit, Eq. 11.

Consider now the possibility that the system can switch between pathways 1 and 2 while in transit between A and B. This is captured by a nonzero rate γ of switching between the two intermediates (Fig. 4). Notice that even if $\alpha = 1$, nonzero switching in general breaks time-reversibility, as the Kolmogorov criterion is violated for the cycles $A \to 1 \to 2 \to A$ and $B \to 1 \to 2 \to B$. Would the nonequilibrium nature of the process be evident from observed transition paths? Interestingly, for $\alpha = 1$ the nonequilibrium character of the system does not lead to asymmetry between the forward and the backward transition paths: it is evident from Eq. 5 that the forward/backward symmetry is preserved, $p_{A\to B}(t) = p_{B\to A}(t)$, no matter how strong the driving is (i.e. regardless of the switching rate γ), as the contributions of the two pathways into A-to-B and B-to-A transitions are the same and independent of γ. However, the non-Markov character of the observed dynamics would still be discoverable because the distributions $p_{A\to B}(t)$ and $p_{B\to A}(t)$, while identical in this case, would be nonexponential (cf. Eq. 9 with $\alpha = 1$), with a coefficient of variation exceeding 1 and thus incompatible with a Markov model with linear topology[47].

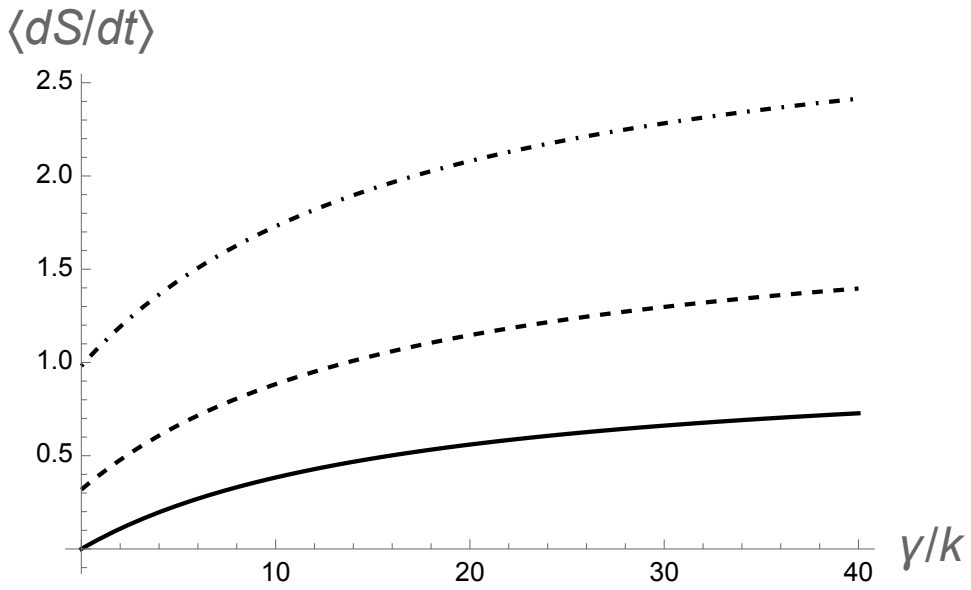


Figure 6. Mean entropy production rate (Eq. 3) for the 4-state system shown in Fig. 4 plotted as a function of the switching rate γ between two states 1 and 2 that form the transition region. Here $k_1 = 50k$, $k_2 = 10k$, and the values of the parameter α are $\alpha = 1$ (solid line), $\alpha = 2$ (dashed line), and $\alpha = 3$ (dot-dashed line).

Consider now the case where $\alpha \neq 1$. In the limit $\gamma \ll k, k\alpha, k_1, k_2$ the distributions of the transition path times approach the results of Eq. 9. The opposite limit where switching is the fastest timescale of the system, $\gamma \gg k, k\alpha, k_1, k_2$ is also easy to understand. In this case, many switches take place while the system remains in the transition region, and the transition region is effectively a single intermediate TR with apparent transition rates to each side equal to the average rate $(k_1 + k_2)/2$. Thus, the two-pathway scheme can be replaced with a single-pathway one (Fig. 3), with an exponential distribution of transition path times given by

$$p_{A\to B}(t) = p_{B\to A}(t) = (k_1 + k_2)e^{-(k_1+k_2)t}. \quad (11)$$

Forward/backward symmetry is therefore recovered for $\gamma \to \infty$ despite the nonequilibrium character of the process[57]. In this case the entropy production is increasing with the increasing exchange rate γ (Fig. 6), yet the observed distributions of the transition path time evolve from two distinguishable distributions, Eq. 9, to two identical distributions, Eq. 11 (see Figure 5). Moreover, the observed distribution approaches a single-exponential one, consistent with the assumption that the transition region consists of a single intermediate state. The system becomes increasingly indistinguishable from an inherently time-reversible, Markovian 3-state system with linear topology as long as the two states 1 and 2 are indistinguishable experimentally. Note, however, that even for high values of γ, *short-time behavior* of $p_{A\to B}(t)$ and $p_{B\to A}(t)$ is always different (Fig. 5). Indeed, it follows from Eqs. 5-6 that $p_{A\to B}(0)$ and $p_{B\to A}(0)$ are independent of γ. Therefore, given sufficiently high (and increasing with increasing γ) time resolution, it is still in principle possible to discern time-irreversible behavior of the system.

4. Summary and future directions

To summarize, here we argue that it may be surprisingly difficult to determine, from a trajectory observed at a single-molecule level, whether it represents an active (e.g., driven by ATP hydrolysis) process or is a manifestation of equilibrium fluctuations of the observed molecule. The reason is ultimately the low-dimensional character of most single-molecule observables, which often forces an analysis of data in terms of simple models with linear topology. When such models are assumed to obey Markovian dynamics, they, fundamentally, are equilibrium models that cannot capture active processes. Therefore, a successive fit of data

to a linear hidden-Markov model does not necessarily imply that the true dynamics is reversible.

A solution, on the experimental side, is a development of multidimensional techniques[4, 58-60]. On a data analysis side, irreversibility is often encoded in non-Markov effects. Such effects – fundamentally – cannot be neglected, and they can be detected with appropriate data analysis[48, 61]. If discovering and quantifying directionality of the observed dynamics is the objective, it may also be desirable to estimate entropy production, Eq.2, directly from raw experimental time series (e.g. the photon sequence in Fig. 3) rather than after post-processing the data using, e.g., hidden Markov models. In this regard, histogram entropy estimators and compression-algorithm-based estimators appear promising[11, 62, 63], although, to our knowledge, they have not yet been applied to photon sequences. On a theory side, there have been recent developments addressing the question of how measures of irreversibility such as the entropy production can be deduced from partial observations such as some but not all transitions[36, 44] – application of those ideas to experimental trajectories is a promising new direction.

Acknowledgments. AG was supported by German Research Foundation (DFG) through the Emmy Noether Program 'GO 2762/1-2'. DEM was supported by the Robert A. Welch Foundation (Grant No. F- 1514), the National Science Foundation (Grant No. CHE 1955552), and Erwin Neher fellowship of the Max Planck Institute for Multidisciplinary Sciences. DEM would also like to thank the Max Planck Institute for Multidisciplinary Sciences and particularly his co-author for the hospitality during his stay there. Discussions with Gilad Haran, Hagen Hofmann, Anatoly Kolomeisky, and Benjamin Schuler are greatly appreciated.

References

1. Astumian, R. D., Microscopic reversibility as the organizing principle of molecular machines. *Nat Nanotechnol* **2012,** *7* (11), 684-8.
2. Purcell, E. M., Life at Low Reynolds Number. *American Journal of Physics* **1977,** *45*, 3-11.
3. Mazal, H.; Haran, G., Single-molecule FRET methods to study the dynamics of proteins at work. *Curr Opin Biomed Eng* **2019,** *12*, 8-17.
4. Ratzke, C.; Hellenkamp, B.; Hugel, T., Four-colour FRET reveals directionality in the Hsp90 multicomponent machinery. *Nat Commun* **2014,** *5*, 4192.
5. Sanabria, H.; Rodnin, D.; Hemmen, K.; Peulen, T. O.; Felekyan, S.; Fleissner, M. R.; Dimura, M.; Koberling, F.; Kuhnemuth, R.; Hubbell, W.; Gohlke, H.; Seidel, C. A. M., Resolving dynamics and function of transient states in single enzyme molecules. *Nat Commun* **2020,** *11* (1), 1231.
6. Pelz, B.; Zoldak, G.; Zeller, F.; Zacharias, M.; Rief, M., Subnanometre enzyme mechanics probed by single-molecule force spectroscopy. *Nat Commun* **2016,** *7*, 10848.
7. Gnesotto, F. S.; Mura, F.; Gladrow, J.; Broedersz, C. P., Broken detailed balance and non-equilibrium dynamics in living systems: a review. *Rep Prog Phys* **2018,** *81* (6), 066601.
8. Gladrow, J.; Fakhri, N.; MacKintosh, F. C.; Schmidt, C. F.; Broedersz, C. P., Broken Detailed Balance of Filament Dynamics in Active Networks. *Phys Rev Lett* **2016,** *116* (24), 248301.
9. Battle, C.; Broedersz, C. P.; Fakhri, N.; Geyer, V. F.; Howard, J.; Schmidt, C. F.; MacKintosh, F. C., Broken detailed balance at mesoscopic scales in active biological systems. *Science* **2016,** *352* (6285), 604-7.
10. Li, J.; Horowitz, J. M.; Gingrich, T. R.; Fakhri, N., Quantifying dissipation using fluctuating currents. *Nat Commun* **2019,** *10* (1), 1666.
11. Roldan, E.; Parrondo, J. M., Estimating dissipation from single stationary trajectories. *Phys Rev Lett* **2010,** *105* (15), 150607.
12. Fodor, E.; Nardini, C.; Cates, M. E.; Tailleur, J.; Visco, P.; van Wijland, F., How Far from Equilibrium Is Active Matter? *Phys Rev Lett* **2016,** *117* (3), 038103.
13. Seara, D. S.; Machta, B. B.; Murrell, M. P., Irreversibility in dynamical phases and transitions. *Nat Commun* **2021,** *12* (1), 392.
14. Seifert, U.; Speck, T., Fluctuation-dissipation theorem in nonequilibrium steady states. *EPL (Europhysics Letters)* **2010,** *89* (1), 10007.
15. Cugliandolo, L. F.; Dean, D. S.; Kurchan, J., Fluctuation-Dissipation Theorems and Entropy Production in Relaxational Systems. *Physical Review Letters* **1997,** *79* (12), 2168-2171.
16. Mizuno, D.; Tardin, C.; Schmidt, C. F.; Mackintosh, F. C., Nonequilibrium mechanics of active cytoskeletal networks. *Science* **2007,** *315* (5810), 370-3.
17. Dieball, C.; Godec, A., Mathematical, Thermodynamical, and Experimental Necessity for Coarse Graining Empirical Densities and Currents in Continuous Space. *Physical Review Letters* **2022,** *129*, 140601.
18. Tu, Y., The nonequilibrium mechanism for ultrasensitivity in a biological switch: sensing by Maxwell's demons. *Proc Natl Acad Sci U S A* **2008,** *105* (33), 11737-41.
19. Makarov, D. E., *Single Molecule Science: Physical Principles and Models*. CRC Press, Taylor & Francis Group: Boca Raton, 2015; p 214.

20. Peliti, L.; Pigolotti, S., *Stochastic Thermodynamics*. Princeton University Press: 2021.
21. Seifert, U., Stochastic thermodynamics, fluctuation theorems and molecular machines. *Rep Prog Phys* **2012,** *75* (12), 126001.
22. Schuler, B.; Eaton, W. A., Protein folding studied by single-molecule FRET. *Curr Opin Struct Biol* **2008,** *18* (1), 16-26.
23. Chung, H. S.; Eaton, W. A., Protein folding transition path times from single molecule FRET. *Curr Opin Struct Biol* **2018,** *48*, 30-39.
24. Aviram, H. Y.; Pirchi, M.; Barak, Y.; Riven, I.; Haran, G., Two states or not two states: Single-molecule folding studies of protein L. *J Chem Phys* **2018,** *148* (12), 123303.
25. Hoffer, N. Q.; Woodside, M. T., Probing microscopic conformational dynamics in folding reactions by measuring transition paths. *Curr Opin Chem Biol* **2019,** *53*, 68-74.
26. Yu, H.; Siewny, M. G.; Edwards, D. T.; Sanders, A. W.; Perkins, T. T., Hidden dynamics in the unfolding of individual bacteriorhodopsin proteins. *Science* **2017,** *355* (6328), 945-950.
27. Makarov, D. E.; Berezhkovskii, A. M.; Haran, G.; Pollak, E., The effect of time resolution on apparent transition path time distributions observed in single-molecule studies of biomolecules. *J. Phys. Chem. B* **2022,** *126* (40), 7966–7974.
28. Cossio, P.; Hummer, G.; Szabo, A., On artifacts in single-molecule force spectroscopy. *Proc Natl Acad Sci U S A* **2015,** *112* (46), 14248-53.
29. Neupane, K.; Woodside, M. T., Quantifying Instrumental Artifacts in Folding Kinetics Measured by Single-Molecule Force Spectroscopy. *Biophys J* **2016,** *111* (2), 283-6.
30. Makarov, D. E., Communication: Does force spectroscopy of biomolecules probe their intrinsic dynamic properties? *J Chem Phys* **2014,** *141* (24), 241103.
31. Hill, T. L., *Free Energy Transduction and Biochemical Cycle Kinetics*. Springer-Verlag New York, Inc.: 1989.
32. Kolomeisky, A. B., *Motor Proteins and Molecular Motors*. CRC Press: 2015.
33. Barducci, A.; De Los Rios, P., Non-equilibrium conformational dynamics in the function of molecular chaperones. *Curr Opin Struct Biol* **2015,** *30*, 161-169.
34. Mazal, H.; Iljina, M.; Barak, Y.; Elad, N.; Rosenzweig, R.; Goloubinoff, P.; Riven, I.; Haran, G., Tunable microsecond dynamics of an allosteric switch regulate the activity of a AAA+ disaggregation machine. *Nat Commun* **2019,** *10* (1), 1438.
35. Kelly, F. P., *Reversibility and Stochastic Networks*. Wiley: Chichester, 1979.
36. Meer, J. v. d.; Ertel, B.; Seifert, U., Thermodynamic inference in partially accessible Markov networks: A unifying perspective from transition-based waiting time distributions. *Phys. Rev. X* **2022,** *12*, 031025.
37. Ertel, B.; van der Meer, J.; Seifert, U., Operationally accessible uncertainty relations for thermodynamically consistent semi-Markov processes. *Phys Rev E* **2022,** *105* (4-1), 044113.
38. Hartich, D.; Godec, A., Emergent memory and kinetic hysteresis in strongly driven networks. *Phys. Rev. X* **2021,** *11*, 041047.
39. Berezhkovskii, A. M.; Makarov, D. E., On the forward/backward symmetry of transition path time distributions in nonequilibrium systems. *Journal of Chemical Physics* **2019,** *151*, 065102.
40. Hartich, D.; Godec, A., Violation of Local Detailed Balance Despite a Clear Time-Scale Separation. *arXiv* **2022**, arXiv:2111.14734

41. Wang, Z.; Makarov, D. E., Nanosecond dynamics of single polypeptide molecules revealed by photoemission statistics of fluorescence resonance energy transfer: A theoretical study. *J. Phys. Chem. B* **2003,** *107*, 5617.
42. Nettels, D.; Gopich, I. V.; Hoffmann, A.; Schuler, B., Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc Natl Acad Sci U S A* **2007,** *104* (8), 2655-60.
43. Gopich, I. V.; Szabo, A., Decoding the pattern of photon colors in single-molecule FRET. *J Phys Chem B* **2009,** *113* (31), 10965-73.
44. Harunari, P. E.; Dutta, A.; Polettini, M.; Roldán, É., What to learn from few visible transitions' statistics? *Physical Review X* **2022,** *12*, 041026.
45. Elber, R., Perspective: Computer simulations of long time dynamics. *J Chem Phys* **2016,** *144* (6), 060901.
46. Elber, R.; Makarov, D. E.; Orland, H., *Molecular Kinetics in Condense Phases: Theory, Simulation, and Analysis*. Wiley and Sons: 2020.
47. Satija, R.; Berezhkovskii, A. M.; Makarov, D. E., Broad distributions of transition-path times are fingerprints of multidimensionality of the underlying free energy landscapes. *Proc Natl Acad Sci U S A* **2020,** *117* (44), 27116-27123.
48. Berezhkovskii, A. M.; Makarov, D. E., Single-Molecule Test for Markovianity of the Dynamics along a Reaction Coordinate. *J Phys Chem Lett* **2018,** *9* (9), 2190-2195.
49. Dutta, R.; Pollak, E., What can we learn from transition path time distributions for protein folding and unfolding? *Phys Chem Chem Phys* **2021,** *23* (41), 23787-23795.
50. Sturzenegger, F.; Zosel, F.; Holmstrom, E. D.; Buholzer, K. J.; Makarov, D. E.; Nettels, D.; Schuler, B., Transition path times of coupled folding and binding reveal the formation of an encounter complex. *Nat Commun* **2018,** *9* (1), 4708.
51. Gladrow, J.; Ribezzi-Crivellari, M.; Ritort, F.; Keyser, U. F., Experimental evidence of symmetry breaking of transition-path times. *Nat Commun* **2019,** *10* (1), 55.
52. Shin, J.; Berezhkovskii, A. M.; Kolomeisky, A. B., Crowding breaks the forward/backward symmetry of transition times in biased random walks. *J Chem Phys* **2021,** *154* (20), 204104.
53. Shin, J.; Kolomeisky, A. B., Asymmetry of forward/backward transition times as a non-equilibrium measure of complexity of microscopic mechanisms. *J Chem Phys* **2020,** *153* (12), 124103.
54. Berezhkovskii, A. M.; Makarov, D. E., From Nonequilibrium Single-Molecule Trajectories to Underlying Dynamics. *J Phys Chem Lett* **2020,** *11* (5), 1682-1688.
55. Chung, H. S.; Eaton, W. A., Single-molecule fluorescence probes dynamics of barrier crossing. *Nature* **2013,** *502* (7473), 685-8.
56. Chung, H. S.; Gopich, I. V., Fast single-molecule FRET spectroscopy: theory and experiment. *Phys Chem Chem Phys* **2014,** *16* (35), 18644-57.
57. Mehl, J.; Lander, B.; Bechinger, C.; Blickle, V.; Seifert, U., Role of hidden slow degrees of freedom in the fluctuation theorem. *Phys Rev Lett* **2012,** *108* (22), 220601.
58. Patrick, E. M.; Srinivasan, S.; Jankowsky, E.; Comstock, M. J., The RNA helicase Mtr4p is a duplex-sensing translocase. *Nature Chemical Biology* **2017,** *13* (1), 99-104.
59. Kim, J. Y.; Chung, H. S., Disordered proteins follow diverse transition paths as they fold and bind to a partner. *Science* **2020,** *368* (6496), 1253-1257.

60. Kolimi, N.; Pabbathi, A.; Saikia, N.; Ding, F.; Sanabria, H.; Alper, J., Out-of-Equilibrium Biophysical Chemistry: The Case for Multidimensional, Integrated Single-Molecule Approaches. *The Journal of Physical Chemistry B* **2021,** *125* (37), 10404-10418.
61. Lapolla, A.; Godec, A., Toolbox for quantifying memory in dynamics along reaction coordinates. *Physical Review Research* **2021,** *3* (2), L022018.
62. Roldan, E.; Parrondo, J. M., Entropy production and Kullback-Leibler divergence between stationary trajectories of discrete systems. *Phys Rev E Stat Nonlin Soft Matter Phys* **2012,** *85* (3 Pt 1), 031129.
63. Song, K.; Makarov, D. E.; Vouga, E., Memory effects and static disorder reduce information in single-molecule signals. *bioRxiv* **2022**, 2022.01.13.476256.