

## Toolbox for quantifying memory in dynamics along reaction coordinates

Alessio Lapolla<sup>1</sup> and Aljaž Godec<sup>1\*</sup>*Mathematical bioPhysics Group, Max Planck Institute for Biophysical Chemistry, Göttingen 37077, Germany*

(Received 11 January 2021; accepted 13 May 2021; published 28 May 2021)

Memory effects in time series of experimental observables are ubiquitous, have important consequences for the interpretation of kinetic data, and may even affect the function of biomolecular nanomachines such as enzymes. Here we propose a set of complementary methods for quantifying conclusively the magnitude and duration of memory in a time series of a reaction coordinate. The toolbox is general, robust, easy to use, and does not rely on any underlying microscopic model. As a proof of concept we apply it to the analysis of memory in the dynamics of the end-to-end distance of the analytically solvable Rouse-polymer model, an experimental time series of extensions of a single DNA hairpin measured by optical tweezers, and the fraction of native contacts in a small protein probed by atomistic molecular dynamics simulations.

DOI: [10.1103/PhysRevResearch.3.L022018](https://doi.org/10.1103/PhysRevResearch.3.L022018)

The dynamics of complex, high-dimensional physical systems such as complex biomolecules is frequently described by means of memoryless, Markovian diffusion along a low-dimensional reaction coordinate [1–10]. Such simplified models often accurately describe selected observations in experiments [11–15] and computer simulations [1,6,7]. However, as soon as latent, hidden degrees of freedom that become projected out do not relax instantaneously on the time scale we observe the reaction coordinate [16], or the reaction coordinate does not locally equilibrate in metastable mesostates [17], almost any projection of high-dimensional dynamics onto a lower dimensional coordinate introduces memory [16–25].

Memory effects can have intriguing manifestations in the evolution of both ensemble- [16,26–29] and time-averaged observables [16,30], and are often particularly well pronounced in observations that reflect, or couple to, intramolecular distances in conformationally flexible biomolecules [17,19,21–23,26,31–39]. Moreover, if the dynamics is ergodic in the sense that the system relaxes to a unique equilibrium probability density function from any initial condition (i.e., the reaction coordinate has a unique free energy landscape) then the memory is necessarily transient [16]. Whether or not memory is in fact relevant depends on how its extent compares to the relaxation time and whether or not the latter is reached in an experiment. If the extent of memory is comparable to, or longer than, the time scale on which biomolecules operate, such as, e.g., enzymes catalyzing chemical reactions [40,41], non-Markovian effects shape biological function.

It is therefore important to assess the presence and duration of memory effects in the dynamics along reaction coordinates. An elegant “test of Markovianity” of a reaction coordinate has recently been proposed by Berezhkovskii and Makarov, who considered the behavior of transition paths [42]. The authors provide a pair of inequalities whose violation conclusively reflects that the dynamics is non-Markovian. However, memory effects are typically transient [16], although their extent may exceed the duration of experimental observations [34]. There is thus a need to determine not only the presence of memory in a time series of a reaction coordinate but also its extent and attenuation on different time scales.

Here, we fill this gap by providing a toolbox for quantifying the magnitude and duration of memory in a time series of a reaction coordinate. We propose a set of model-free complementary methods that are easy to use and suited to treat reaction coordinates with arbitrary dimensionality. As a proof of concept we apply these methods to the analysis of an experimental time series of the extension of a DNA hairpin measured by optical tweezers, the fraction of native contacts in a protein probed by atomistic molecular dynamics (MD) simulations, and the exactly solvable Rouse model of polymer chain.

*Theory.* Our approach is twofold—(i) we quantify violations of the Chapman-Kolmogorov equation in a time series of the monitored *true dynamics* and (ii) we compare the true dynamics to a constructed *nominally memoryless diffusion* in the free energy and diffusion landscape of the true dynamics. This assumes all hidden degrees of freedom to be at equilibrium constrained by the instantaneous value of the observable.

Let  $q_t$  with  $0 \leq t \leq T$  denote the monitored time series of the reaction coordinate and  $q_t^M$  the constructed Markovian series. Without any loss of generality we assume that the reaction coordinate is one-dimensional—the generalization to multiple dimensions is straightforward. We assume  $q_t$  and  $q_t^M$  to be ergodic with an equilibrium probability density  $p_{\text{eq}}(q)$  that is by construction identical for both processes. Let  $G(q, t|q_0) = \langle \delta(q_t - q) \rangle_{q_0}$  denote the probability

\*agodec@mpibpc.mpg.de

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Open access publication funded by the Max Planck Society.

density that the reaction coordinate evolving from  $q_{t=0} = q_0$  is found at time  $t$  to have a value in an infinitesimal neighborhood of  $q$  and  $G^M(q, t|q_0) = \langle \delta(q_t^M - q) \rangle_{q_0}$  the Markovian counterpart, where  $\delta(x)$  denotes Dirac's delta function and the angular brackets  $\langle \cdot \rangle_{q_0}$  the average over all realizations of  $q_t$  evolving from  $q_0$ . We then have  $\lim_{t \rightarrow \infty} G(q, t|q_0) = \lim_{t \rightarrow \infty} G_M(q, t|q_0) = p_{\text{eq}}(q)$  as a result of ergodicity. In practice the limits are achieved as soon as  $t$  becomes sufficiently larger than the relaxation time  $t_{\text{rel}}$ , i.e.,  $t \gtrsim t_{\text{rel}}$ , which may or may not be reached in an experiment. Note that the relaxation times of the true and Markovian reference process are typically different [16,29].

We use two descriptors. The first is the Kullback-Leibler divergence between the transition probabilities of the true and a reference process defined as [43]

$$\mathcal{D}_{q_0}^a(t) \equiv \int dq G(q, t|q_0) \ln[G(q, t|q_0)/G^a(q, t|q_0)], \quad (1)$$

where  $a = \text{CK}, \text{M}$  denotes the particular kind of reference process that we detail below. By construction  $\mathcal{D}_{q_0}^a(t) \neq 0$  if and only if  $G(q, t|q_0) \neq G^a(q, t|q_0)$  and thus nonzero values of  $\mathcal{D}_{q_0}^a(t)$  reflect memory in the dynamics of  $q_t$ .

When  $q_t$  reaches equilibrium in the course of the experiment we also consider the normalized equilibrium autocorrelation function defined as

$$C_{\text{M}}(t) \equiv \frac{\langle q_t q_0 \rangle_{\text{M}} - \langle q \rangle_{\text{eq}}^2}{\langle q^2 \rangle_{\text{eq}} - \langle q \rangle_{\text{eq}}^2}, \quad (2)$$

where we have introduced

$$\begin{aligned} \langle q_t q_0 \rangle_{\text{M}} &\equiv \iint qq_0 G^{\text{M}}(q, t|q_0) p_{\text{eq}}(q_0) dq dq_0 \\ &\stackrel{T \gg t_{\text{rel}}}{\approx} (T-t)^{-1} \int_0^{T-t} q_{\tau+t} q_{\tau} d\tau, \quad t \ll T, \\ \langle q^n \rangle_{\text{eq}} &\equiv \int q^n p_{\text{eq}}(q) dq, \quad \text{for } n = 1, 2, \\ &\stackrel{T \gg t_{\text{rel}}}{\approx} T^{-1} \int_0^T q_{\tau}^n d\tau, \end{aligned} \quad (3)$$

where the definitions in terms of time averages hold when trajectories are much longer than the relaxation time, i.e.,  $T \gg t_{\text{rel}}$ . The absence of an index refers to the true process and M to the constructed Markovian counterpart.

We consider two distinct reference processes. The first one is a mathematical construction based on the Chapman-Kolmogorov equation (i.e.,  $a = \text{CK}$ ) that we may write as

$$G_{\tau}^{\text{CK}}(q, t|q_0) \equiv \int G(q, t - \tau|q') G(q', \tau|q_0) dq', \quad (4)$$

because the Green's function of a time-homogeneous Markov process is time-translation invariant,  $G(q, t - \tau|q') = G(q, t|q', \tau)$ , and  $G_{\tau}^{\text{CK}}(q, t|q_0) = G(q, t|q_0)$  independent of  $\tau$  [44]. The physical interpretation of Eq. (4), which is exact for Markov processes, is that we observe the true dynamics  $q_t$  until time  $\tau$  and then instantaneously reset the memory (if any) to zero.

If  $q_t$  is indeed memoryless we have  $G_{\tau}^{\text{CK}}(q, t|q_0) = G(q, t|q_0)$  for any  $\tau$  and thus  $\mathcal{D}_{\tau, q_0}^{\text{CK}}(t) = 0$  for any  $t$  and  $\tau$ . If  $G_{\tau}^{\text{CK}}(q, t|q_0) \neq G(q, t|q_0)$  for some  $t$  and  $\tau$ , then  $q_t$  is

conclusively non-Markovian and  $\mathcal{D}_{\tau, q_0}^{\text{CK}}(t) > 0$ , but the converse is not true. Namely, there exist non-Markovian processes that satisfy the Chapman-Kolmogorov equation [16,45]. Note that this method does not require  $q_t$  to reach equilibrium during an experiment and requires only  $G(q, t|q_0)$  that is straightforward to determine from a time series  $q_t$  given sufficient data. If equilibrium is reached,  $\mathcal{D}_{\tau, q_0}^{\text{CK}}(t \gg t_{\text{rel}}) \simeq 0$  for any  $q_0$ . By analyzing  $\mathcal{D}_{\tau, q_0}^{\text{CK}}(t)$  we can quantify the degree and range of memory as a function of  $\tau$  and  $q_0$ , which we demonstrate below.

In the second method we construct from  $q_t$  a Markovian time series  $q_t^{\text{M}}$  (i.e.,  $a = \text{M}$ ) evolving under the influence of the potential of mean force  $w(q) \equiv -k_{\text{B}}T \ln p_{\text{eq}}(q)$  according to the thermodynamically consistent anti-Itô (i.e., postpoint) [17] Langevin equation

$$\frac{d}{dt} q_t^{\text{M}} = D(q_t^{\text{M}}) f(q_t^{\text{M}}) / k_{\text{B}}T + \sqrt{2D(q_t^{\text{M}})} \otimes \xi_t, \quad (5)$$

where  $f(q_t^{\text{M}}) \equiv -k_{\text{B}}T \partial_q \ln p_{\text{eq}}(q)|_{q=q_t^{\text{M}}}$  and  $\xi_t$  denotes zero mean Gaussian white noise with covariance  $\langle \xi_t \xi_{t'} \rangle = \delta(t - t')$ ,  $D(q_t)$  is the diffusion landscape, and  $\otimes$  is the anti-Itô or Klimontovich product [46] [see Supplemental Material (SM) [47] for the discretized version of Eq. (5)]. This method assumes the ability to determine the equilibrium probability density  $p_{\text{eq}}(q)$  and thus requires  $q_t$  to reach equilibrium. In the simplest model the diffusion coefficient does not depend on  $q$  and we may interpret Eq. (5) according to Itô. However, this may not be the case (see below), and we note that the best possible Markovian approximation includes a positional dependence [48]. Efficient methods have been developed to infer  $D(q)$  [6,9,10].

On the level of the probability density function Eq. (5) corresponds to the Fokker-Planck equation

$$\partial_t G^{\text{M}}(q, t|q_0) = \partial_q D(q) [\partial_q - f(q) / k_{\text{B}}T] G^{\text{M}}(q, t|q_0), \quad (6)$$

with initial condition  $G^{\text{M}}(q, 0|q_0) = \delta(q - q_0)$  and natural boundary conditions imposed by the underlying physics. Depending on the specific problem  $G^{\text{M}}(q, t|q_0)$  can be found by a numerical integration of the Langevin equation and subsequent histogram analysis, i.e.,  $G^{\text{M}}(q, t|q_0) = \langle \delta(q_t^{\text{M}} - q) \rangle_{q_0}$ , or by projecting the full dynamics or directly solving Eq. (6) as done, e.g., for polymers [49], single-file models [16,27], and in the literature on persistence [50,51] in diffusive and critical dynamics [50–55]. Below we illustrate both approaches.

*End-to-end distance of a Rouse polymer.* As a first example we consider a Rouse polymer chain with  $N + 1$  beads ( $N$  bonds) in the absence of hydrodynamic interactions [56,57] and focus on the end-to-end distance as the reaction coordinate, i.e.,  $q_t \equiv |\mathbf{r}_1 - \mathbf{r}_{N+1}|$ , which is known to be non-Markovian. The model is exactly solvable and the explicit results for  $C(t)$ ,  $C_M(t)$ ,  $G(q, t|q_0)$ ,  $G^{\text{M}}(q, t|q_0)$ , and  $G_{\tau}^{\text{CK}}(q, t|q_0)$  are all given in [47]. We express time in units of  $t_{\text{Kuhn}}$ , the characteristic diffusion time of a Kuhn segment, i.e.,  $t_{\text{Kuhn}} = b^2/D$ , where  $b$  is the Kuhn length and  $D$  the diffusion coefficient of a bead.

A comparison of the autocorrelation function of the true dynamics and its Markovian approximation  $C_M(t)$  is shown in Fig. 1(a), with the inset depicting the corresponding

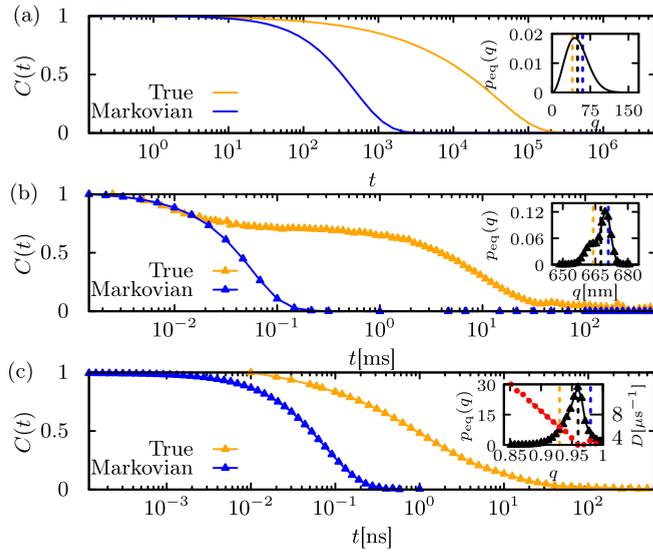


FIG. 1. Autocorrelation function of the true dynamics (orange) and its Markovian approximation (blue) for (a) a Rouse polymer with 1000 monomers with time expressed in units of the diffusion time of a Kuhn segment  $t_{\text{Kuhn}}$ , (b) the extension of a DNA hairpin, and (c) the fraction of native contacts in the WW domain of protein 2F21. The black line in the inset depicts the respective equilibrium probability density function  $p_{\text{eq}}(q)$  and the red one  $D(q)$ . The dashed lines depict the initial conditions we consider in Fig. 2.

equilibrium probability density  $p_{\text{eq}}(q)$ . Note that when the free energy landscape  $w(q)$  overestimates the confining effect

of hidden degrees of freedom on  $q_t$  the Markovian approximation overestimates the relaxation rate (e.g., Ref. [16]; see also [47]). Namely, the Markovian approximation assumes the hidden degrees of freedom to remain at equilibrium at all times, whereas the actual instantaneous, fluctuating restoring force on  $q_t$  is in this case smaller than the force arising from  $w(q)$ .

The Chapman-Kolmogorov construct for the Rouse polymer,  $G_{\tau}^{\text{CK}}(q, t|q_0)$  (given explicitly in [47]) differs from the true  $G(q, t|q_0)$  for all expected large values of  $t - \tau$ . A quantification of the discrepancy between the true and ‘‘Chapman-Kolmogorov’’ evolution of the end-to-end distance of the Rouse polymer in terms of the Kullback-Leibler divergence (1) is shown in Fig. 2(a). A typical time evolution of  $\mathcal{D}_{\tau, q_0}^{\text{CK}}(t)$  gradually increases from zero, reaches a maximum, and afterwards returns back to zero, which reflects the gradual buildup and attenuation of memory because  $q_t$  ‘‘remembers’’ the initial condition of the hidden degrees of freedom [16]. As a result, the Chapman-Kolmogorov Green’s function  $G_{\tau}^{\text{CK}}(q, t|q_0)$  fails to predict the true evolution of  $q_t$ , and  $\mathcal{D}_{\tau, q_0}^{\text{CK}}(t)$  constructed this way depends on both,  $\tau$  and initial condition  $q_0$ . For the Rouse polymer with 1000 beads  $\mathcal{D}_{\tau, q_0}^{\text{CK}}(t) \neq 0$  at least up to  $t \sim 10^4 \times t_{\text{Kuhn}}$ .

Next we examine  $\mathcal{D}_{q_0}^{\text{M}}(t)$ , the Kullback-Leibler divergence (1) between the true Green’s function  $G(q, t|q_0)$  and the Markovian approximation corresponding to the white-noise Markovian diffusion in the exact free energy landscape [i.e., Eq. (5)]. The results are shown in Fig. 2(d).

The qualitative features of the time dependence of  $\mathcal{D}_{q_0}^{\text{M}}(t)$  are similar to those observed in Fig. 2(a)—memory builds up

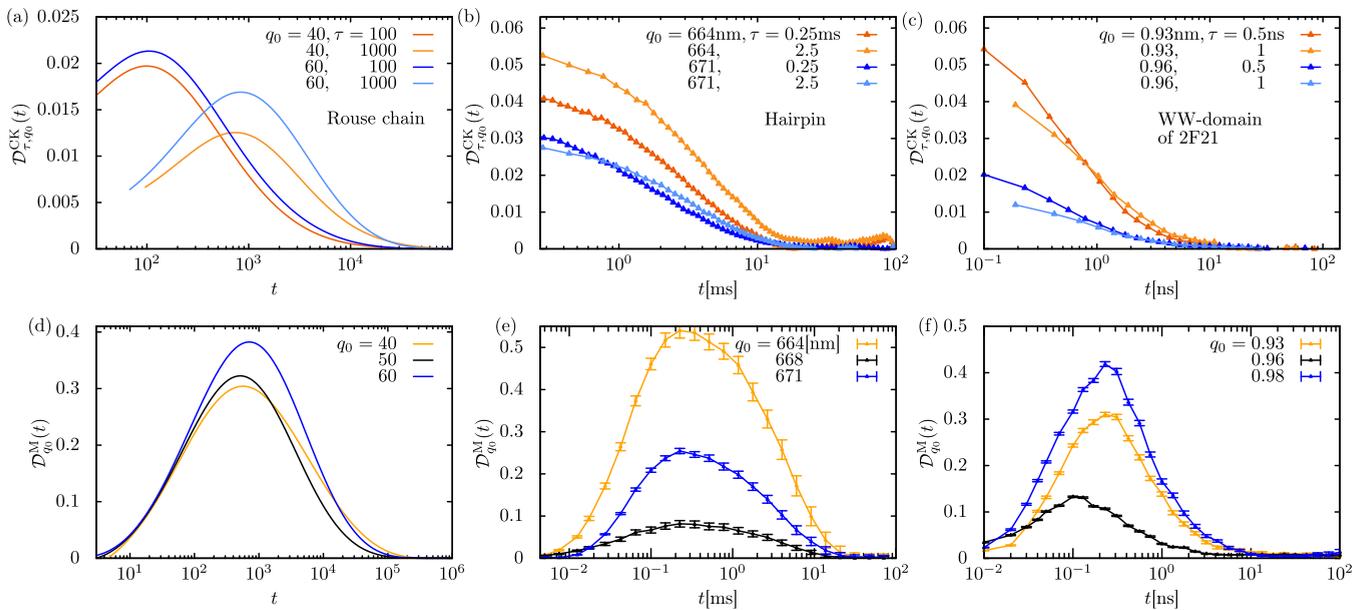


FIG. 2. Kullback-Leibler divergence  $\mathcal{D}_t^a$  in Eq. (1) between true Green’s function  $G(q, t|q_0)$  and (a)–(c) the Chapman-Kolmogorov Eq. (4) (i.e.,  $a = \text{CK}$ ), and (d)–(f) the Markovian approximation  $G^{\text{M}}(q, t|q_0)$  corresponding to the Langevin Eq. (5) (i.e.,  $a = \text{M}$ ) as a function of time  $t$  for (a) and (d) the Rouse polymer with 1000 beads evolving from several initial conditions  $q_0$ , (b) and (e) the extension of a DNA hairpin evolving from several initial conditions within a bin of thickness 1 nm centered at  $q_0$ , and (c) and (f) the fraction of native contacts in the WW domain of protein 2F21 for several  $q_0$ ; the error bars depict the standard deviation obtained by systematically neglecting  $\sim 20\%$  (in case of the hairpin) and  $\sim 40\%$  (in case of the protein) of the data. Due to the particular construction of Eq. (4) times shorter than depicted are not accessible due to numerical instability or poor statistics.

in a finite interval and smoothly returns back to zero from the attained maximum. The intuition behind this result is that it takes a finite time to allow for distinct evolutions of hidden degrees of freedom that introduce memory in the dynamics of the reaction coordinate  $q_t$ . At long times memory is progressively lost as a result of the gradual relaxation of the hidden degrees of freedom to their respective equilibrium that in turn renders the dynamics of the reaction coordinate effectively memoryless and correspondingly  $\mathcal{D}_{q_0}^M(t)$  vanishes.

*Single-molecule experiments on a DNA hairpin.* As a second example we consider a time series of the end-to-end distance of a single-strand DNA hairpin measured in an optical tweezers experiment performed by the Woodside group [58]. The data set contains 11 million measurements of the extension of the DNA hairpin 30R50T4 held in a pair of optical traps with stiffness 0.63 pN/nm and 1.1 pN/nm, respectively, sampled with a 2.5  $\mu$ s temporal resolution. It has been shown that this time series is non-Markovian [39]. The length of the time series is much larger than the relaxation time [see Fig. 1(b)] and therefore we slice it into several pieces that are statistically independent. More precisely, we use the time scale  $t_{\text{cut}}$ , where the autocorrelation function of the extension,  $C(t)$ , falls to  $\simeq 0.05$ . This ensures  $t_{\text{cut}} \gg t_{\text{rel}}$  and yields an ensemble of 50 statistically independent trajectories.

We determine the equilibrium probability density  $p_{\text{eq}}(q)$  [see inset of Fig. 1(b)] and two-point joint probability density  $p(q, t, q_0, 0) = p(q, t_0 + t, q_0, t_0)$  by performing a standard histogram analysis with a bin size of  $l_{\text{bin}} = 0.35$  nm, such that  $q$  refers to a bin of width  $l_{\text{bin}}$  centered at  $q$ . The Green's function is thereupon obtained by the law of conditional probability,  $G(q, t|q_0) = p(q, t, q_0, 0)/p_{\text{eq}}(q)$ , while  $C(t)$  in Eq. (2) is determined directly from the respective second lines of Eq. (3).

The Chapman-Kolmogorov construct is determined from  $G(q, t|q_0)$  by direct integration of Eq. (4) and is used to determine  $\mathcal{D}_{\tau, q_0}^{\text{CK}}(t)$ , while the corresponding fictitious Markovian process evolves as Markovian diffusion in a free energy landscape  $w(q)$  with a constant diffusion coefficient  $D$  that we determine according to standard methods as detailed in [47]. According to the results to a good approximation  $D$  is independent of  $q$ . The analysis yields  $D = 447 \pm 9$  nm<sup>2</sup>/ms that we use to generate the Markovian time series  $q_t^M$  by integrating the Itô Langevin equation (5) using the Euler-Mayurama scheme (for details see [47]), and determine  $\mathcal{D}_i^M(q_0)$  in Eq. (1) and  $C_M(t)$  in Eq. (2), respectively.

In contrast to the Rouse polymer the DNA hairpin exists in two characteristic conformational states—folded and unfolded. As a result, the equilibrium probability density function  $p_{\text{eq}}(q)$  is bimodal and the dynamics of  $q_t$  displays signatures of metastability [58]. However, since the two peaks corresponding to the two subpopulations are not separated [see inset of Fig. 1(b)] the potential of mean force  $w(q)$  is expected to underestimate the free energy barrier and therefore the Markovian evolution is likely to overestimate the relaxation rate. In complete agreement Fig. 1(b) displays an overestimation of the rate of decay of autocorrelations in the Markovian approximation by two orders of magnitude in time. Moreover, a long-lived plateau is observed in the true  $C(t)$  spanning more than an order of magnitude in time.

In order to assess whether the mismatch between true and Markovian time evolution is predominantly due to an underestimation of the free energy barrier between folded and unfolded states of the hairpin we inspect the Kullback-Leibler divergence (1) between the true and “Chapman-Kolmogorov evolution” shown in Fig. 2(b). The result clearly shows pronounced signatures of memory extending over more than  $\sim 10$  ms. Note that the “Chapman-Kolmogorov evolution” is exact until time  $t = \tau$ , whereupon memory is reset to zero. Therefore, a nonzero  $\mathcal{D}_{\tau, q_0}^{\text{CK}}(t)$  is a clear signature of memory arising from the dynamical coupling of  $q_t$  to hidden degrees of freedom. Similar to the Rouse polymer  $\mathcal{D}_{\tau, q_0}^{\text{CK}}(t)$  depends on the initial condition  $q_0$ .

A buildup and decay of memory similar to the Rouse polymer is also observed in the time evolution of  $\mathcal{D}_{q_0}^M(t)$ , the Kullback-Leibler divergence between the Green's function of the true evolution and the white-noise Markovian diffusion in the exact free energy landscape shown in Fig. 2(b). Notably, Fig. 2(b) and Fig. 2(e) display essentially the same extent of memory (though the peak is attained sooner in the white-noise Markovian diffusion), demonstrating that metastability does not necessarily destroy nor dominate memory in the evolution of reaction coordinates. Note that the presence of memory in metastable systems is not unusual (see, e.g., Refs. [21, 22] and [29]). In total, the analysis conclusively identifies extended memory in the dynamics of the extension of the hairpin.

It is important to note that the extent of memory (of the order of  $\sim 10$  ms) is clearly shorter than the relaxation time  $t_{\text{rel}}$  [compare Figs. 1(b) and 2(e)], and therefore the decay of memory does not coincide with  $t_{\text{rel}}$  and the corresponding “forgetting” of initial conditions of the coordinate itself. Instead the memory reflects correlations between  $q_t$  and the *initial conditions* of the hidden degrees of freedom [16]. The information encoded in  $C(t)$  and  $\mathcal{D}^{\text{M,CK}}(t)$  is therefore different— $\mathcal{D}^{\text{M,CK}}(t)$  is a genuine measure of the extent and duration of memory.

*MD simulation of WW domain of 2F21.* We analyzed 177 atomistic MD trajectories of the WW domain of the human Pin1 Fip (2F21) mutant [59] provided by the Grubmüller group, each 1  $\mu$ s long sampled every 10 ps. During this time the protein attains a pronounced local equilibrium in the folded state and does not unfold. The data set was produced in 15 days in “wall time.” We also analyzed two longer trajectories, 486 and 651  $\mu$ s long sampled every 200 ps, from [60] where the protein reversibly (un)folders several times but sampling of the unfolded state is limited (see [47]). The fraction of native contacts [61] was chosen as the reaction coordinate (see [47] for details). It reflects the displacement of the protein's structure from the native conformation. In contrast to the previous examples it is *not* known whether this coordinate displays memory. Technical details including the simulation parameters, estimation of  $D(q)$  (with error analysis), and corresponding results for the longer trajectories are shown in [47].

The results are qualitatively similar to the hairpin with one notable exception—the diffusion coefficient may not be considered to be constant. The equilibrium density  $p_{\text{eq}}(q)$  and diffusion landscape  $D(q)$  in the folded state are shown alongside  $C(t)$  in Fig. 1(c). As a first signature of memory the Markovian time series constructed according to Eq. (5)

overestimates the relaxation rate by almost two decades. The Kullback-Leibler divergence  $\mathcal{D}_{\tau, q_0}^{\text{CK}}(t)$  in Fig. 2(c) shows pronounced memory up to  $\sim 10$  ns, extending up to  $\sim 100$  ns when considering the longer trajectories that also capture the protein's dynamics in the unfolded state (see [47]). Occurring on time scales  $\simeq 20 \mu\text{s}$  [60], the (un)folding dynamics is thus memoryless. This example highlights that our method does *not* distinguish between local and global equilibrium in the case of a time-scale separation, such as the ns time-scale folded-state dynamics and  $\sim 20 \mu\text{s}$  time-scale (un)folding dynamics.

The constructed Markovian time series shows qualitatively similar signatures of memory as the hairpin [see Fig. 2(f)]. The extent of memory displayed by  $\mathcal{D}_{q_0}^{\text{M}}(t)$  matches that of  $\mathcal{D}_{\tau, q_0}^{\text{CK}}(t)$  and, similar to the Rouse polymer and hairpin, depends on the initial condition  $q_0$ . One may quite generally relate this dependence to the dynamics of hidden degrees of freedom with respect to how far  $q_0$  is displaced from the free energy minimum. When  $q_0$  is near the free energy minimum the dynamics of hidden degrees of freedom has a smaller effect.

*Remarks on feasibility.* The toolbox requires an ensemble of statistically independent or ergodically long trajectories. Most demanding is the Chapman-Kolmogorov analysis that requires sufficient sampling of the support of the integral in Eq. (4) at different times  $t, \tau$ . Constructing the Markovian time series requires accurate estimates of  $p_{\text{eq}}(q)$  and  $D(q)$ . The minimal data requirements depend on the system at hand, and may vary substantially. However, we propose a simple test of the reliability of the results—determining their uncertainty by a comparison with results obtained by omitting say  $\sim 10\%$ – $20\%$  of data as shown in Figs. 2(e) and 2(f). For a

reliable quantification of memory the statistical uncertainty should be substantially smaller than the value of the Kullback-Leibler divergence, as in the present case.

*Conclusion.* We presented a set of complementary methods to quantify conclusively the degree and duration of memory in a time series of a reaction coordinate  $q_t$ . The proposed toolbox does not assume any particular physical model. Instead it exploits the Chapman-Kolmogorov equation and constructs a fictitious Markovian diffusion process in the free energy landscape of  $q_t$ , and compares the artificially constructed transition probability density with the observed probability density. The analysis not only determines whether the dynamics of  $q_t$  has memory but also quantifies the magnitude and duration of memory and thus complements the recently proposed “test for Markovianity” based on transition paths [42]. Whereas in our examples we considered only one-dimensional coordinates, the toolbox generalizes straightforwardly to higher-dimensional reaction coordinates. The method is general, robust, and easy to use, and should be used before any attempt to describe a complex system with a low-dimensional Markovian reaction coordinate. We therefore hope that it will find numerous applications involving time series derived from experiments and computer simulations.

*Acknowledgments.* We thank K. Neupane and M. T. Woodside for providing access to their DNA-hairpin data, A. Volkhardt and H. Grubmüller for kindly providing unpublished MD trajectories, and the D.E. Shaw group for the two long MD trajectories published in [60]. The financial support from the German Research Foundation (DFG) through the Emmy Noether Program GO 2762/1-1 to A.G. is gratefully acknowledged.

- 
- [1] R. B. Best and G. Hummer, Reaction coordinates and rates from transition paths, *Proc. Natl. Acad. Sci. USA* **102**, 6732 (2005).
- [2] J. J. Portman, S. Takada, and P. G. Wolynes, Microscopic theory of protein folding rates. II. Local reaction coordinates and chain dynamics, *J. Chem. Phys.* **114**, 5082 (2001).
- [3] B. Peters, P. G. Bolhuis, R. G. Mullen, and J.-E. Shea, Reaction coordinates, one-dimensional Smoluchowski equations, and a test for dynamical self-consistency, *J. Chem. Phys.* **138**, 054106 (2013).
- [4] A. K. Faradjian and R. Elber, Computing time scales from reaction coordinates by milestoning, *J. Chem. Phys.* **120**, 10880 (2004).
- [5] A. Berezhkovskii and A. Szabo, One-dimensional reaction coordinates for diffusive activated rate processes in many dimensions, *J. Chem. Phys.* **122**, 014503 (2005).
- [6] G. Hummer, Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations, *New J. Phys.* **7**, 34 (2005).
- [7] R. B. Best and G. Hummer, Coordinate-dependent diffusion in protein folding, *Proc. Natl. Acad. Sci. USA* **107**, 1088 (2009).
- [8] W. Zhang, C. Hartmann, and C. Schütte, Effective dynamics along given reaction coordinates, and reaction rate theory, *Faraday Discuss.* **195**, 365 (2016).
- [9] A. M. Berezhkovskii and D. E. Makarov, Communication: Coordinate-dependent diffusivity from single molecule trajectories, *J. Chem. Phys.* **147**, 201102 (2017).
- [10] B. J. Berne, M. Borkovec, and J. E. Straub, Classical and modern methods in reaction rate theory, *J. Phys. Chem.* **92**, 3711 (1988).
- [11] O. K. Dudko, G. Hummer, and A. Szabo, Theory, analysis, and interpretation of single-molecule force spectroscopy experiments, *Proc. Natl. Acad. Sci. USA* **105**, 15755 (2008).
- [12] K. Neupane, A. P. Manuel, and M. T. Woodside, Protein folding trajectories can be described quantitatively by one-dimensional diffusion over measured energy landscapes, *Nat. Phys.* **12**, 700 (2016).
- [13] K. Neupane, D. A. N. Foster, D. R. Dee, H. Yu, F. Wang, and M. T. Woodside, Direct observation of transition paths during the folding of proteins and nucleic acids, *Science* **352**, 239 (2016).
- [14] J. Gladrow, M. Ribezzi-Crivellari, F. Ritort, and U. F. Keyser, Experimental evidence of symmetry breaking of transition-path times, *Nat. Commun.* **10**, 55 (2019).
- [15] A. L. Thorneywork, J. Gladrow, Y. Qing, M. Rico-Pasto, F. Ritort, H. Bayley, A. B. Kolomeisky, and U. F. Keyser, Direct detection of molecular intermediates from first-passage times, *Sci. Adv.* **6**, eaaz4642 (2020).

- [16] A. Lapolla and A. Godec, Manifestations of projection-induced memory: General theory and the tilted single file, *Front. Phys.* **7**, 182 (2019).
- [17] D. Hartich and A. Godec, Emergent memory and kinetic hysteresis in strongly driven networks, [arXiv:2011.04628](https://arxiv.org/abs/2011.04628).
- [18] N. van Kampen, Remarks on non-Markov processes, *Braz. J. Phys.* **28**, 90 (1998).
- [19] S. S. Plotkin and P. G. Wolynes, Non-Markovian Configurational Diffusion and Reaction Coordinates for Protein Folding, *Phys. Rev. Lett.* **80**, 5015 (1998).
- [20] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, Oxford, 2010).
- [21] D. E. Makarov, Interplay of non-Markov and internal friction effects in the barrier crossing kinetics of biopolymers: Insights from an analytically solvable model, *J. Chem. Phys.* **138**, 014102 (2013).
- [22] M. Ozmaian and D. E. Makarov, Transition path dynamics in the binding of intrinsically disordered proteins: A simulation study, *J. Chem. Phys.* **151**, 235101 (2019).
- [23] H. Meyer, P. Pelagejcev, and T. Schilling, Non-Markovian out-of-equilibrium dynamics: A general numerical procedure to construct time-dependent memory kernels for coarse-grained observables, *EPL (Europhys. Lett.)* **128**, 40001 (2020).
- [24] E. Herrera-Delgado, J. Briscoe, and P. Sollich, Tractable nonlinear memory functions as a tool to capture and explain dynamical behaviors, *Phys. Rev. Research* **2**, 043069 (2020).
- [25] F. Müller, U. Basu, P. Sollich, and M. Krüger, Coarse-grained second-order response theory, *Phys. Rev. Research* **2**, 043123 (2020).
- [26] W. Min, G. Luo, B. J. Cherayil, S. C. Kou, and X. S. Xie, Observation of a Power-Law Memory Kernel for Fluctuations within a Single Protein Molecule, *Phys. Rev. Lett.* **94**, 198302 (2005).
- [27] A. Lapolla and A. Godec, Faster Uphill Relaxation in Thermodynamically Equidistant Temperature Quenches, *Phys. Rev. Lett.* **125**, 110602 (2020).
- [28] A. Lapolla and A. Godec, Bethesf: Efficient computation of the exact tagged-particle propagator in single-file systems via the Bethe eigenspectrum, *Comput. Phys. Commun.* **258**, 107569 (2021).
- [29] A. Lapolla and A. Godec, Single-file diffusion in a bi-stable potential: Signatures of memory in the barrier-crossing of a tagged-particle, *J. Chem. Phys.* **153**, 194104 (2020).
- [30] A. Lapolla and A. Godec, Unfolding tagged particle histories in single-file diffusion: exact single- and two-tag local times beyond large deviation theory, *New J. Phys.* **20**, 113021 (2018).
- [31] S. C. Kou and X. S. Xie, Generalized Langevin Equation with Fractional Gaussian Noise: Subdiffusion within a Single Protein Molecule, *Phys. Rev. Lett.* **93**, 180603 (2004).
- [32] T. Neusius, I. Daidone, I. M. Sokolov, and J. C. Smith, Subdiffusion in Peptides Originates from the Fractal-Like Structure of Configuration Space, *Phys. Rev. Lett.* **100**, 188103 (2008).
- [33] S. Pressé, J. Peterson, J. Lee, P. Elms, J. L. MacCallum, S. Marqusee, C. Bustamante, and K. Dill, Single molecule conformational memory extraction: P5ab rna hairpin, *J. Phys. Chem. B* **118**, 6597 (2014).
- [34] X. Hu, L. Hong, M. Dean Smith, T. Neusius, X. Cheng, and J. Smith, The dynamics of single protein molecules is non-equilibrium and self-similar over thirteen decades in time, *Nat. Phys.* **12**, 171 (2015).
- [35] A. K. Sangha and T. Keyes, Proteins Fold by Subdiffusion of the Order Parameter, *J. Phys. Chem. B* **113**, 15886 (2009).
- [36] S. M. Avdoshenko, A. Das, R. Satija, G. A. Papoian, and D. E. Makarov, Theoretical and computational validation of the Kuhn barrier friction mechanism in unfolded proteins, *Sci. Rep.* **7**, 269 (2017).
- [37] Y. Cote, P. Senet, P. Delarue, G. G. Maisuradze, and H. A. Scheraga, Anomalous diffusion and dynamical correlation between the side chains and the main chain of proteins in their native state, *Proc. Natl. Acad. Sci. USA* **109**, 10346 (2012).
- [38] I. Grossman-Haham, G. Rosenblum, T. Namani, and H. Hofmann, Slow domain reconfiguration causes power-law kinetics in a two-state enzyme, *Proc. Natl. Acad. Sci. USA* **115**, 513 (2018).
- [39] A. G. T. Pyo and M. T. Woodside, Memory effects in single-molecule force spectroscopy measurements of biomolecular folding, *Phys. Chem. Chem. Phys.* **21**, 24527 (2019).
- [40] H. P. Lu, L. Xun, and X. S. Xie, Single-molecule enzymatic dynamics, *Science* **282**, 1877 (1998).
- [41] B. P. English, W. Min, A. M. van Oijen, K. T. Lee, G. Luo, H. Sun, B. J. Cherayil, S. C. Kou, and X. S. Xie, Ever-fluctuating single enzyme molecules: Michaelis-menten equation revisited, *Nat. Chem. Biol.* **2**, 87 (2005).
- [42] A. M. Berezhkovskii and D. E. Makarov, Single-molecule test for Markovianity of the dynamics along a reaction coordinate, *J. Phys. Chem. Lett.* **9**, 2190 (2018).
- [43] S. Kullback and R. Leibler, On information and sufficiency, *Ann. Math. Stat.* **22**, 79 (1951).
- [44] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and Natural Sciences*, 2nd ed. (Springer-Verlag, Berlin, 1985).
- [45] W. Feller, Non-Markovian processes with the semigroup property, *Ann. Math. Stat.* **30**, 1252 (1959).
- [46] Y. Klimontovich, Ito, stratonovich and kinetic forms of stochastic equations, *Physica A* **163**, 515 (1990).
- [47] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevResearch.3.L022018> for a discretization of the anti-Itô Langevin equation (5), exact results for the Rouse polymer, details about MD simulations, the fraction of native contacts, the estimation of the diffusion landscape  $D(q)$ , and a description of the uncertainty quantification.
- [48] A. Berezhkovskii and A. Szabo, Time scale separation leads to position-dependent diffusion along a slow coordinate, *J. Chem. Phys.* **135**, 074108 (2011).
- [49] S. Sunagawa and M. Doi, Theory of diffusion-controlled intrachain reactions of polymers, *Polymer J.* **7**, 604 (1975).
- [50] S. N. Majumdar, Persistence in nonequilibrium systems, *Curr. Sci.* **77**, 370 (1999).
- [51] A. J. Bray, S. N. Majumdar, and G. Schehr, Persistence and first-passage properties in nonequilibrium systems, *Adv. Phys.* **62**, 225 (2013).
- [52] A. J. Bray, B. Derrida, and C. Godrèche, Non-trivial algebraic decay in a soluble model of coarsening, *EPL (Europhys. Lett.)* **27**, 175 (1994).
- [53] B. Derrida, V. Hakim, and V. Pasquier, Exact First-Passage Exponents of 1D Domain Growth: Relation to a Reaction-Diffusion Model, *Phys. Rev. Lett.* **75**, 751 (1995).
- [54] S. N. Majumdar and A. J. Bray, Persistence with Partial Survival, *Phys. Rev. Lett.* **81**, 2626 (1998).

- [55] S. N. Majumdar, A. J. Bray, S. J. Cornell, and C. Sire, Global Persistence Exponent for Nonequilibrium Critical Dynamics, *Phys. Rev. Lett.* **77**, 3704 (1996).
- [56] P. E. Rouse, A Theory of the Linear Viscoelastic Properties of Dilute Solutions of Coiling Polymers, *J. Chem. Phys.* **21**, 1272 (1953).
- [57] K. H. Ahn, J. L. Schrag, and S. J. Lee, Bead-spring chain model for the dynamics of dilute polymer solutions, *J. Non-Newtonian Fluid Mech.* **50**, 349 (1993).
- [58] K. Neupane, A. P. Manuel, J. Lambert, and M. T. Woodside, Transition-path probability as a test of reaction-coordinate quality reveals DNA hairpin folding is a one-dimensional diffusive process, *J. Phys. Chem. Lett.* **6**, 1005 (2015).
- [59] M. Jager, Y. Zhang, J. Bieschke, H. Nguyen, M. Dendle, M. E. Bowman, J. P. Noel, M. Gruebele, and J. W. Kelly, Structure-function-folding relationship in a WW domain, *Proc. Natl. Acad. Sci. USA* **103**, 10648 (2006).
- [60] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, How fast-folding proteins fold, *Science* **334**, 517 (2011).
- [61] R. B. Best, G. Hummer, and W. A. Eaton, Native contacts determine protein folding mechanisms in atomistic simulations, *Proc. Natl. Acad. Sci. USA* **110**, 17874 (2013).