PNAS
www.pnas.org

Supplementary Information for

Non-Markovian Modeling of Protein Folding

Cihan Ayaz, Lucas Tepper, Florian N. Brünig, Julian Kappler, Jan O. Daldrop and Roland R. Netz (complete author list)

Roland R. Netz
E-mail: rnetz@physik.fu-berlin.de

This PDF file includes:

Contents

Supporting Information Text

1. Simulation Details

In the MD simulations, we first minimize the energy of the system using the steepest descent method. Then, we let the system equilibrate at 300 K in the NPT ensemble. During the equilibration, the polypeptide is positionally restrained to ensure that no conformational transitions occur at this stage. We do this by applying harmonic potentials with a force constant of $k = 1000\,\mathrm{kJ/mol/nm^2}$ to all $\mathrm{Ala_9}$ atoms. The equilibration time is 1 ns, i.e., 10^6 time steps. To speed up the equilibration process, we assign to each particle an initial velocity drawn from a Maxwell-Boltzmann distribution at the desired temperature. We perform production runs in the NVT ensemble using a modified Berendsen thermostat with a time constant of 0.1 ps (1). We constrain all bond lengths using the LINCS algorithm. The bond angles are unconstrained. The $\mathrm{Ala_9}$ polypeptide was build using the open-source molecular builder software Avogadro 1.2.0. (2). The N terminus is $\mathrm{NH_3^+}$ and the C terminus is $\mathrm{CO_2^-}$.

2. Numerical Extraction of the Memory Kernel

In this section, we derive an equation that allows us to compute the running integral of the memory kernel, $G(t) = \int_0^t \mathrm{d}s\,\Gamma(s)$, directly from time-series data. The final equation generalizes the method to extract $G(t)$ in a harmonic potential derived in (3) to an arbitrary potential U.

We multiply the GLE in Eq. (1) in the main text by the initial velocity $\dot{q}(0) = v(0)$ and ensemble average the result. This leads to

$$m\,\frac{\mathrm{d}}{\mathrm{d}t}C^{vv}(t) = -C^{\nabla U v}(t) - \int_0^t \mathrm{d}s\,\Gamma(s)C^{vv}(t-s), \qquad [1]$$

where $C^{AB}(t) = \langle A(t)B(0)\rangle$ denotes an equilibrium correlation function. In deriving Eq. (1), we used the orthogonality relation $\langle F_R(t)v(0)\rangle = 0$ which follows from the derivation of the GLE using orthogonal projection operators (4, 5). Time integration of Eq. (1) gives a continuous equation for the running integral $G(t)$

$$mC^{vv}(t) = C^{\nabla U q}(t) - C^{\nabla U q}(0) + mC^{vv}(0) - \int_0^t \mathrm{d}s'\,C^{vv}(s')G(t-s). \qquad [2]$$

By multiplying the GLE by the initial position $q(0)$ and ensemble averaging, we find

$$mC^{\dot{v}q}(t) = -C^{\nabla U q}(t) - \int_0^t \mathrm{d}s \Gamma(s)\, C^{vq}(t-s). \tag{3}$$

This follows from the fact that the GLE is the result of an orthogonal projection onto the subspace spanned by $\{q(0), v(0)\}$ and that the random force F_R lies completely in the orthogonal subspace to $\{q(0), v(0)\}$, i.e. $\langle F_R(t) q(0)\rangle = 0$ at all times t (6). Evaluating Eq. (3) at $t = 0$ and inserting $C^{\dot{v}q}(t) = -C^{vv}(t)$ gives

$$mC^{vv}(0) = C^{\nabla U q}(0). \tag{4}$$

Hence, Eq. (2) becomes

$$\frac{C^{vv}(t)}{C^{vv}(0)} C^{\nabla U q}(0) = C^{\nabla U q}(t) - \int_0^t \mathrm{d}s'\, C^{vv}(s') G(t-s). \tag{5}$$

Discretizing Eq. (5) by use of the trapezoidal rule gives

$$\frac{C_n^{vv}}{C_0^{vv}} C_0^{\nabla U q} = C_n^{\nabla U q} - \frac{\Delta t}{2} G_n C_0^{vv} - \Delta t \sum_{i=1}^{n-1} G_{n-i} C_i^{vv}, \tag{6}$$

where we used $G(0) = 0$. Solving Eq. (6) for $G_n = G(t)$ gives

$$G_n = \frac{2}{\Delta t\, C_0^{vv}} \left(C_n^{\nabla U q} - \frac{C_0^{\nabla U q}}{C_0^{vv}} C_n^{vv} - \Delta t \sum_{i=1}^{n-1} G_{n-i} C_i^{vv} \right), \tag{7}$$

where C_n^{AB} is the n-th value of the equilibrium correlation function $C(t) = \langle A(t), B(0)\rangle$.

3. Computation of the Effective Mass

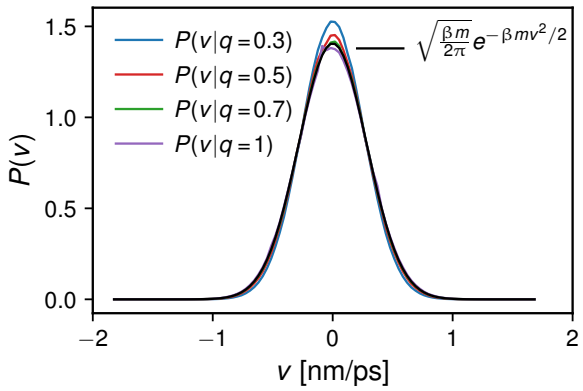

Fig. S1. From the MD data, we compute the velocity distribution of the hb4 reaction coordinate in a given interval $[q - \mathrm{d}q, q + \mathrm{d}q]$ with $\mathrm{d}q = 0.001$ nm. It can be seen that the distribution is independent of the position q.

In Fig. S1, the velocity distribution $P(v|q)$ conditioned on the position q is shown for different values of q. Fig. S1 shows that the velocity distribution is independent of the position q. This implies that the effective mass of the reaction coordinate is independent of q.

4. Markovian Embedding

The direct numerical integration of the GLE is not advisable since it would involve a nested loop. A more efficient way is to introduce additional degrees of freedom to map the GLE onto a system of linearly coupled standard Langevin equations (7). In the following, we demonstrate how this is done. We consider the system of coupled equations:

$$m\ddot{q}(t) = -\frac{\mathrm{d}U(q)}{\mathrm{d}q} - \sum_{n=1}^{N} k_n \left(q(t) - y_n(t)\right), \tag{8a}$$

$$\dot{y}_n = -\frac{k_n}{\gamma_n} \left(y_n(t) - q(t)\right) + \sqrt{\frac{k_B T}{\gamma_n}} \eta_n \quad \text{for } n = 1, 2, \ldots, N, \tag{8b}$$

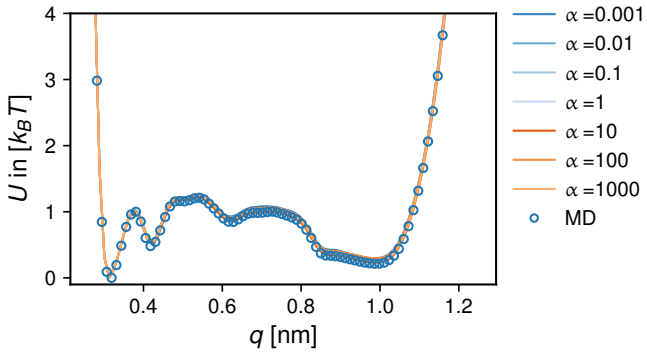

Fig. S2. Shown are free energies U computed from the GLE trajectories as a function of the hb4 reaction coordinate q, where the memory times are rescaled with a factor α as described in the main text. The GLE free energies are compared to the MD free energy.

where $U(q)$ is an external potential that only affects the reaction coordinate q. Each η_n is a stationary Gaussian process with $\langle \eta_n \rangle = 0$ and $\langle \eta_n(t), \eta_l(t') \rangle = 2\delta_{nl}\delta(t-t')$. Eq. (8) can be mapped onto a GLE for the reaction coordinate q. This is accomplished by solving the inhomogeneous first order differential equation for the y_n and by inserting the solution back into the equation for q. The solution for $y_n(t)$ reads

$$y_n(t) = (y_n(0) - q(0))\, e^{-t/\tau_n} + q(t) - \int_0^t \mathrm{d}s\, e^{-(t-s)/\tau_n} \dot{q}(s) + \int_0^t \mathrm{d}s\, e^{-(t-s)/\tau_n} \frac{1}{\gamma_n} \eta_n(s). \quad [9]$$

Inserting Eq. (9) into Eq. (8a) gives

$$m\ddot{q}(t) = -\frac{\mathrm{d}U(q)}{\mathrm{d}q} - \int_0^t \mathrm{d}s\, \Gamma(t-s)\dot{q}(s) + F(t), \quad [10]$$

with $F(t) = \sum_n \left[\frac{\gamma_n}{\tau_n}(q(0) - y_n(0))e^{-t/\tau_n} - \frac{1}{\tau_n}\int_0^t \mathrm{d}s\, e^{(t-s)/\tau_n}\eta_n(s) \right]$ and the kernel $\Gamma(t) = \sum_n \frac{\gamma_n}{\tau_n} e^{-t/\tau_n}$. If we assume that the system in Eq. (8) is initially in equilibrium, the initial conditions are distributed according to a Boltzmann distribution (7). Using this, it follows that $\langle F(t) \rangle = 0$ and $\langle F(t), F(0) \rangle = k_B T \Gamma(t)$. Hence, we obtain a GLE with a memory kernel consisting of a sum of exponentially decaying functions. One can generate a trajectory from the GLE in Eq. (10) by numerically integrating Eq. (8). In Fig. S2, we show free energies from trajectories generated using the GLE method described here. All free energies are independent of the memory rescaling factor α introduced in the main text and agree perfectly with the free energy extracted from the original MD simulation.

5. Error Estimation of Correlated Data

We consider a series of consecutive measurements

$$\mathbf{x} = (x_1, x_2, \ldots, x_N). \quad [11]$$

Each x_i is a fluctuating quantity distributed according to some probability distribution function $P(x)$. We assume that the system is equilibrated such that $P(x)$ will be the same for all x_i. In this case, the finite average $\bar{x} = 1/N \sum_{i=1}^N x_i \equiv \tau$ is also a fluctuating quantity.

In this work, $\mathbf{x}$ would represent the first passage times τ_{FP} "measured" using simulation data and $\bar{x} \equiv \tau$ the estimate of the mean first passage time. When one performs consecutive measurements of first passage times, the resulting data will in general be correlated. Therefore, the variance of $\bar{x}$ does not equal the variance of x. In the following, we present how to estimate the errors of averages of correlated data (8).

The variance of τ is given by

$$\sigma_\tau^2 = \langle \tau^2 \rangle - \langle \tau \rangle^2. \quad [12]$$

Inserting $\tau = 1/N \sum_{i=1}^N x_i$ into the above equation yields

$$\sigma_\tau^2 = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N \left(\langle x_i, x_j \rangle - \langle x_i \rangle \langle x_j \rangle \right). \quad [13]$$

We define $C_{i,j} = \langle (x_i - \langle x_i \rangle)\,(x_j - \langle x_j \rangle) \rangle = \langle x_i, x_j \rangle - \langle x_i \rangle \langle x_j \rangle$, the covariance matrix of x. Since we consider an equilibrated process, the covariance only depends on the difference $|i-j|$. Using this, we can rewrite the expression in Eq. (13) as

$$\sigma_\tau^2 = \frac{1}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} C_{|i-j|} = \frac{1}{N^2} \left(N C_0 + 2 \sum_{i=1}^{N-1} \sum_{j=1}^{N-i} C_j \right) \tag{14a}$$

$$= \frac{1}{N^2} \left[N C_0 + 2 \sum_{i=1}^{N-1} (N-i)\, C_i \right]. \tag{14b}$$

From Eq. (14b), the formula for averages of correlated data immediately follows

$$\Delta\tau_{MFP}^2 = \frac{1}{N} \left[C_0 + 2 \sum_{i=1}^{N-1} \left(1 - \frac{i}{N}\right) C_i \right]. \tag{15}$$

For our estimation of errors, we estimate the auto-correlation function by $C_i = \frac{1}{N-i} \sum_{k=1}^{N-i} (x_k - \bar{x})(x_{k+i} - \bar{x})$.

6. Mean First-Passage and Mean Transition-Path Times

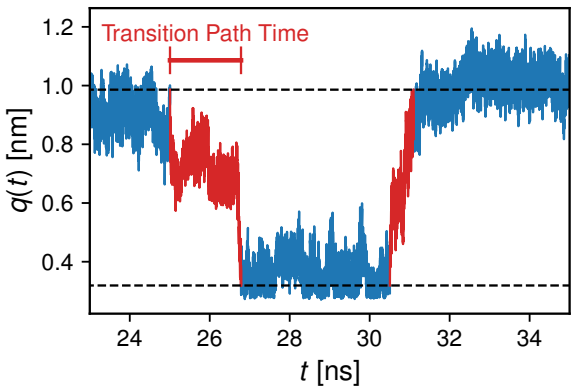

Fig. S3. Illustration of transition-path times using a hb4 trajectory from MD simulations. The red sections of the trajectory highlight two transition paths between the q values 0.32 nm and 0.99 nm.

In Fig. S3, we show exemplary transition paths between the two local minima in the free energy at $q = 0.32\,\text{nm}$ and at $q = 0.99\,\text{nm}$. A transition path between the two states $q_s \to q_f$ is a path that leaves q_s and reaches q_f for the first time without recrossing q_s or q_f. The mean transition path time $\tau_{\text{MTPT}}(q_s, q_f)$ is the average over all transition path times $q_s \to q_f$. In Fig. S4(a), we show $\tau_{\text{MTPT}}(q_s = 0.32\,\text{nm}, q_f)$ and $\tau_{\text{MTPT}}(q_s = 0.99\,\text{nm}, q_f)$ for various systems, i.e., MD, GLE, underdamped LE and overdamped LE. The curves for the underdamped LE and overdamped LE are computed using the friction coefficient $\bar{\gamma} = \int_0^\infty \Gamma(s)$ from the extracted memory kernel and the potential of mean force extracted from the MD system. We see a very good agreement between the GLE and MD simulations. Thus, the GLE with a time-dependent friction reproduces not only the mean first-passage times (MFPT) of the MD simulation, as shown in the main text, but also the mean transition-path times (MTPT). In Fig. S4(b), we show ratios of the mean transition-path times from MD and GLE simulations and from MD and overdamped LE simulations shown in Fig. S4(a). In Fig. S4(c), we show ratios of the MFPTs from MD and GLE simulations and from MD and overdamped LE simulations shown in Fig. 3(a) in the main text. Additionally, in Fig. S5 we show the first-passage time distributions between the positions $q_L = 0.32\,\text{nm}$, $q_B = 0.54\,\text{nm}$ and $q_R = 0.99\,\text{nm}$. No significant deviation can be discerned between the MD and the GLE system in the distributions shown in Fig. S5(a), (b) and (d). In Fig. S5(c), it can be seen that the MD system has higher probabilities towards larger first-passage times compared to the GLE system. The impact of this on the MFPT can be seen in Fig. 3A in the main text.

7. Non-Markovian Modeling for the End-to-end Distance

We repeat the analysis performed for the hb4 reaction coordinate in the main text for the end-to-end distance of Ala_9 (d_{e2e}). This reaction coordinate is of particular importance, since it is available from single-molecule Förster resonance energy transfer (FRET) experiments (9, 10). Here, d_{e2e} was computed from the same MD simulation data used in the main text to obtain the hb4 reaction coordinate. d_{e2e} is the distance between the center of masses of the first and the last residues of Ala_9, i.e.,

$$d_{e2e}(t) := \left\| \frac{1}{M_1^{\text{ala}}} \sum_{i \in I_{\text{resiude}_1}} m_i \vec{r}_i(t) - \frac{1}{M_9^{\text{ala}}} \sum_{i \in I_{\text{resiude}_9}} m_i \vec{r}_i(t) \right\|, \tag{16}$$

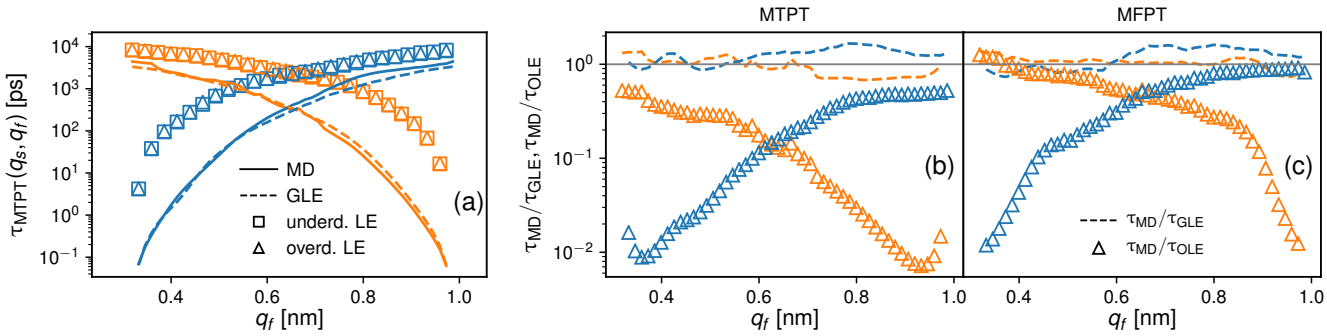

Fig. S4. (a) We show the mean transition path time (MTPT) between $q_s = \{0.32, 0.99\}$ nm (blue, orange) and q_f as a function of q_f. We compute the MTPT from the MD and GLE trajectories as well as trajectories generated using an underdamped (ULE) and overdamped Langevin equation (OLE) with the friction $\bar{\gamma} = \int_0^\infty \mathrm{d}s\, \Gamma(s)$ and the PMF $U(q)$ from the MD system shown in Fig. S2. In **(b)**, we compare the ratios MD/GLE and MD/OLE from the data shown in (a). In **(c)**, the same ratios are shown for the mean first passage times (MFPT) shown in the main text.

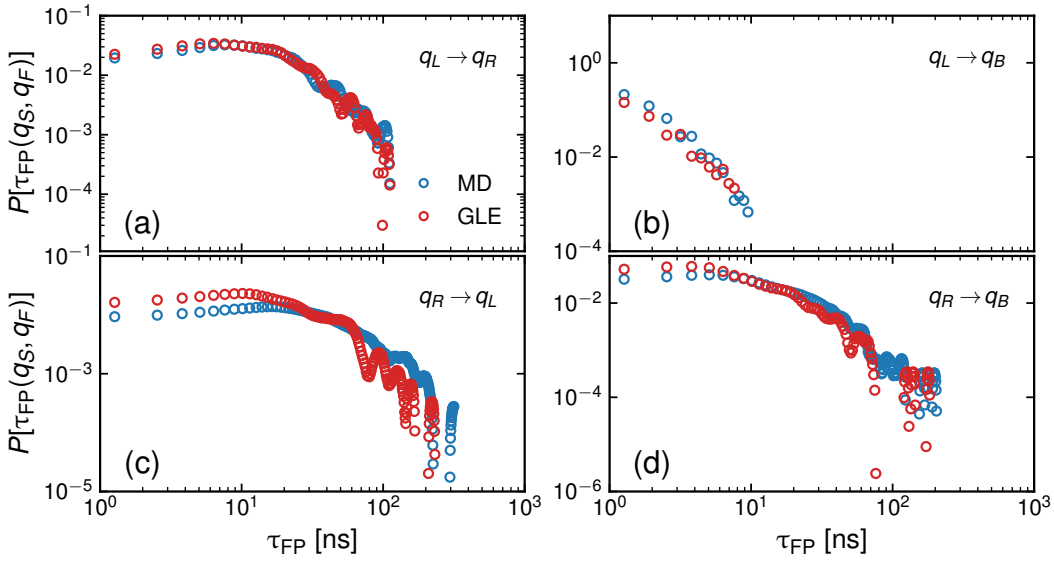

Fig. S5. First-passage time distributions between the positions $q_L = 0.32$ nm, $q_B = 0.54$ nm and $q_R = 0.99$ nm considered in the main text. We compare distributions computed from the MD system (blue) and the corresponding GLE system (red).

where m_i denotes the mass of the i-th atom in one of the terminal residues. In Eq. (16), the summation is performed over the index sets I_{residue_1} (all atom indieces of first residue) and I_{residue_9} (all atom indices of last residue). The vectors $\vec{r}_i(t)$ are the corresponding atom positions at time t and M_i^{ala} is the total mass of the i-th residue. The free energy along the d_{e2e}-coordinate shows a sharp minimum containing loop states and a very broad region made up of a diverse spectrum of states. The α-helix forms a shoulder on the barrier between this region and the loop state. The free energies $U(q)$ in Fig S6 show a good convergence for all but the sharp minimum. For this reason subsequent transition times were only calculated for the movement in the right minimum. As with hb4, the integral over the memory kernel was numerically extracted, together with the kernel itself, which is available as a numerical derivative of the integral. The kernel was fitted with five exponential functions using least-squares. The fit in Fig S7 a) agrees well with the numerical kernel, disregarding the oscillations. The integral over the kernel in Fig S7 b) extracted from the MD data follows an exponential regime up to approximately 7 ns, followed by a linear regime. The linear regime results in a kernel that is constant but just above zero. We assume that this is a numerical artifact, possibly as a result of the lack of convergence in the free energy. Since the velocity v and the mean force ∇U in Eq. (7) have vanishing means, their correlation functions should decay to zero at sufficiently large lag times such that the running integral G of the memory function should converge to a constant value. For this reason, we disregard the linear regime in the fitting procedure and obtain an estimation of 7501 u/ps for the total friction, indicated as a black, dashed line in Fig S7 a). The validity of the resulting kernel is verified by simulating a Markov-Embedding system according to Eq. (8). From the resulting trajectory, the MFPTs and the MSD of the GLE system and the MD data are compared in Fig S8. A nearly perfect agreement can be seen for both.

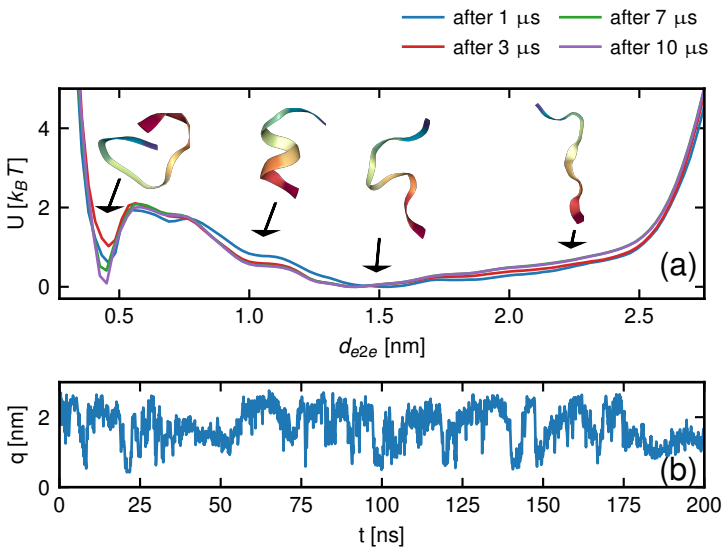

Fig. S6. (a) Free energy $U(q)$ for the end-to-end distance reaction coordinate of Ala_9 for different simulation lengths, representative snapshots of the polypeptide backbone are shown. **(b)** A 200 ns long segment of the trajectory is shown.

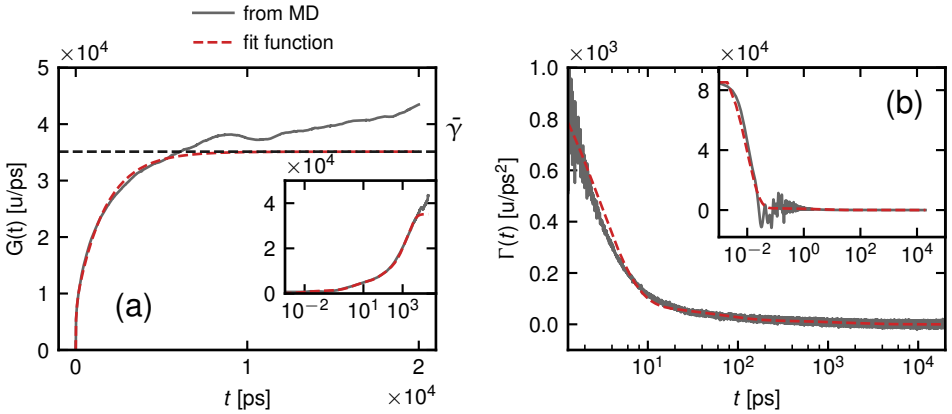

Fig. S7. (a) Running integral G(t) over the memory function, the inset shows a lin-log plot. The horizontal dashed line denotes the total friction coefficient $\bar{\gamma}$. **(b)** Memory function $\Gamma(t)$, the inset includes short times. Gray lines correspond to the numerical data, red lines correspond to the multi-exponential fit according to Eq. (3) in the main text.

8. Derivation of the Underdamped Langevin Equation from the GLE

The underdamped Langevin equation can be straightforwardly derived from the GLE given by

$$m\dot{v}(t) = -U'[q(t)] - \int_0^t \mathrm{d}s\, \Gamma(t-s)v(s) + F_R(t). \quad [17]$$

For this we expand the velocity $v(s)$ in the integral in Eq. (17) around $s = t$

$$m\dot{v}(t) = -U'[q(t)] - \sum_{n=0}^{\infty} \frac{v^{(n)}(t)}{n!} \int_0^t \mathrm{d}s\, (s-t)^n \Gamma(t-s) + F_R(t), \quad [18]$$

$$= -U'[q(t)] - \sum_{n=0}^{\infty} \frac{v^{(n)}(t)}{n!} \int_0^t \mathrm{d}s\, (-s)^n \Gamma(s) + F_R(t). \quad [19]$$

In the Markov limit, the memory kernel $\Gamma(s)$ is sharply peaked at $s = 0$. Thus, we can extend the upper integration limit to infinity without changing the value of the integral for $t > 0$. This results in

$$m\dot{v}(t) = -U'[q(t)] - \gamma_0 v(t) + \gamma_1 \dot{v}(t) - \gamma_2 \ddot{v}(t) + \cdots + F_R(t) \quad [20]$$

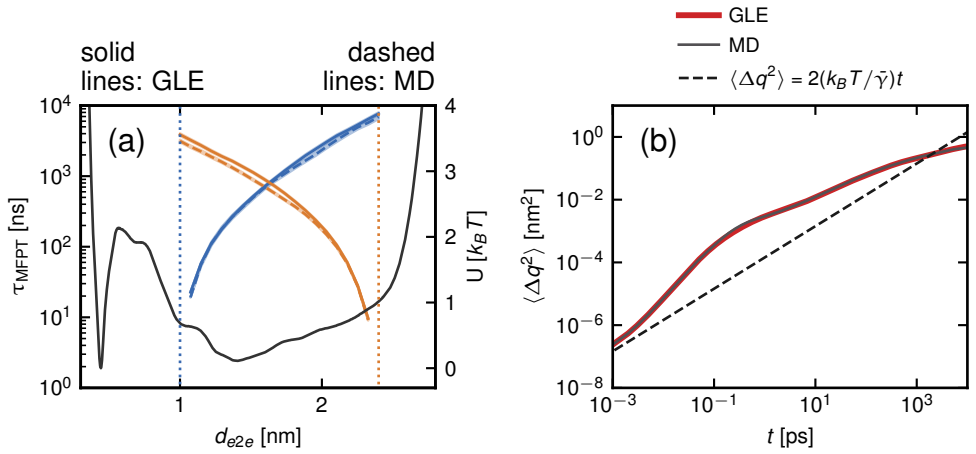

Fig. S8. (a) Comparison of the MFPTs from MD and GLE simulations as a function of the final position q_f for start positions $q_s = 1.0$ nm (blue) and $q_s = 2.4$ nm (orange). The vertical dotted lines indicate the starting positions. The gray curve shows the free energy $U(q)$. **(b)** Mean-squared displacement of the reaction coordinate, MD (gray line) and GLE (red line). Simulation results agree perfectly and exhibit superdiffusion for times up to 0.1 ps and subdiffusion up to 1 ns.

where $\gamma_n = \int_0^\infty \mathrm{d}s\, \frac{s^n}{n!}\Gamma(s)$ are moments of the memory function. For quickly decaying memory functions all higher moments vanish and we obtain the underdamped Langevin equation

$$m\dot{v}(t) = -U'[q(t)] - \gamma_0 v(t) + \eta(t), \qquad [21]$$

with $\langle\eta(t), \eta(t')\rangle = 2k_BT\gamma_0\delta(t-t')$.

9. Kramers Moyal Coefficients for Multidimensional Langevin Equations

A coupled system of N first-order Markovian Langevin equations for the variables $\mathbf{q}(t) = (q_1(t), q_2(t), \ldots, q_N(t))$ can be written as

$$\dot{q}_i(t) = h_i(\mathbf{q}(t)) + \sum_{j=1}^{N} g_{ij}(\mathbf{q}(t))\eta_j(t), \qquad [22]$$

where each η_j is a stationary Gaussian process that obeys

$$\langle\eta_i(t)\rangle = 0, \qquad \text{and} \qquad \langle\eta_i(t')\eta_j(t)\rangle = 2\delta_{ij}\delta(t'-t). \qquad [23]$$

To compute the Kramers Moyal Coefficients (KMCs) introduced in Sec. 11, one integrates Eq. (22) from t to $t+\Delta t$

$$q_i(t+\Delta t) - q_i(t) = \int_t^{t+\Delta t} \mathrm{d}s\, h_i(\mathbf{q}(s)) + \sum_{j=1}^{N}\int_t^{t+\Delta t} \mathrm{d}s\, g_{ij}(\mathbf{q}(s))\eta_j(s). \qquad [24]$$

We expand the functions h_i and g_{ij} around $q_i(s) = q_i(t)$ to first order, i.e.,

$$h_i(\mathbf{q}(s)) \approx h_i(\mathbf{q}(t)) + (\mathbf{q}(s) - \mathbf{q}(t)) \cdot \left.\frac{h_i(\mathbf{q}(s))}{\partial\mathbf{q}(s)}\right|_{\mathbf{q}(s)=\mathbf{q}(t)}, \qquad [25a]$$

$$g_{ij}(\mathbf{q}(s)) \approx g_{ij}(\mathbf{q}(t)) + (\mathbf{q}(s) - \mathbf{q}(t)) \cdot \left.\frac{g_{ij}(\mathbf{q}(s))}{\partial\mathbf{q}(s)}\right|_{\mathbf{q}(s)=\mathbf{q}(t)}. \qquad [25b]$$

For the differences $\mathbf{q}(s) - \mathbf{q}(t)$, one makes use of Eq. (24) to write

$$q_i(s) - q_i(t) = \int_t^s \mathrm{d}s'\, h_i(\mathbf{q}(s')) + \sum_{j=1}^{N}\int_t^s \mathrm{d}s'\, g_{ij}(\mathbf{q}(s'))\eta_j(s'). \qquad [26]$$

Eq. (25) in combination with Eq. (26) is inserted into Eq. (24) and the result is averaged over the random force. By making use of the relations in Eq. (23), one finds that the first two KMCs, i.e., the drift and diffusion coefficients, are uniquely determined by the functions h_i, g_{ij} and in the limit $\Delta t \to 0$ given by (7)

$$D_i^{(1)}(\mathbf{q}(t)) = h_i(\mathbf{q}(t)) + g_{kj}(\mathbf{q}(t))\frac{\partial}{\partial q_k} g_{ij}(\mathbf{q}(t)), \qquad [27a]$$

$$D_{ij}^{(2)}(\mathbf{q}(t)) = g_{ik}(\mathbf{q}(t))\, g_{jk}(\mathbf{q}(t)), \qquad [27b]$$

where the Einstein summation convention is used. All higher order KMCs vanish in the limit $\Delta t \to 0$. According to Pawula's Theorem (see Sec.11), it is sufficient to show that the fourth order KMCs vanish in the limit $\Delta t \to 0$, since from this it follows that all higher orders $n \geq 3$ must also vanish.

A. Underdamped One-Dimensional Langevin Equation. We consider the underdamped Langevin equation of the general form

$$\dot{q}(t) = v(t), \tag{28a}$$

$$\dot{v}(t) = \frac{1}{m}\left(h[q(t)] - \gamma[q(t)]v(t) + b[q(t)]\eta(t)\right) \tag{28b}$$

Defining $\mathbf{q}(t) = (q(t), v(t))$ and comparing Eq. (28) with Eq. (22) to determine the functions h_i and g_{ij}, we find

$$h_1(q(t), v(t)) = v(t), \qquad h_2(q(t), v(t)) = \frac{1}{m}\left(h[q(t)] - \gamma[q(t)]v(t)\right), \tag{29a}$$

$$g_{11}(q(t), v(t)) = g_{12}(q(t), v(t)) = g_{21}(q(t), v(t)) = 0, \qquad g_{22}(q(t), v(t)) = \frac{b[q(t)]}{m}. \tag{29b}$$

Thus, from Eq. (27) we obtain

$$D_q(q, v) = v, \qquad D_v(q, v) = \frac{h(q)}{m} - \frac{\gamma(q)}{m}v, \qquad D_{vv}(q, v) = D_{vv}(q) = \frac{b^2(q)}{m^2}, \tag{30}$$

and $D_{qv}(q, v) = D_{vq}(q, v) = D_{qq}(q, v) = 0$. The Fokker-Planck equation follows from Eq. (53) as

$$\dot{P}(q, v, t) = \left[-\frac{\partial}{\partial q}D_q(q, v) - \frac{\partial}{\partial v}D_v(q, v) + \frac{\partial^2}{\partial v^2}D_{vv}(q, v)\right] P(q, v, t). \tag{31}$$

The equilibrium distribution $P_{eq}(q, v) \simeq \exp(-\beta m v^2/2 - \beta U(q))$ must be a stationary solution of the FP equation, from which we straightforwardly obtain the conditions $h(q) = -U'(q)$ and $b^2(q) = m^2 D_{vv}(q) = k_B T \gamma(q)$. Thus the underdamped Langevin equation of a particle in a potential $U(q)$ reads

$$\dot{q}(t) = v(t), \tag{32a}$$

$$\dot{v}(t) = \frac{1}{m}\left(-U'[q(t)] - \gamma[q(t)]v(t) + \sqrt{k_B T \gamma[q(t)]}\eta(t)\right). \tag{32b}$$

B. Overdamped One-Dimensional Langevin Equation. We consider the overdamped Langevin equation of the general form

$$\dot{q}(t) = h[q(t)] + g[q(t)]\eta(t). \tag{33}$$

From Eq. (27) we obtain

$$D_{qq}(q) = g^2(q), \quad D_q(q) = h(q) + g(q)g'(q) = h(q) + D'_{qq}(q)/2. \tag{34}$$

The Fokker-Planck equation follows from Eq. (53) as

$$\dot{P}(q, t) = \left[-\frac{\partial}{\partial q}D_q(q) + \frac{\partial^2}{\partial q^2}D_{qq}(q)\right] P(q, t). \tag{35}$$

Again, the equilibrium distribution $P_{eq}(q) \simeq \exp(-\beta U(q))$ must be a stationary solution of the FP equation, from which we obtain the condition $h(q) = -D_{qq}(q)\beta U'(q) + D'_{qq}(q)/2$. Associating the function $D_{qq}(q)$ with the friction profile according to $D_{qq}(q) = 1/(\beta\gamma(q))$, the overdamped Langevin equation of a particle in a potential $U(q)$ finally reads

$$\dot{q}(t) = -\frac{U'(q)}{\gamma(q)} + \frac{k_B T}{2}\frac{\partial}{\partial q}\frac{1}{\gamma(q)} + \sqrt{\frac{k_B T}{\gamma(q)}}\eta(t). \tag{36}$$

In Eq. (36), the second term on the r.h.s. cancels a spurious drift term and ensures that the correct stationary state is obtained.

10. Diffusivity Profile from a Non-Linear Coordinate Transformation

Here, we demonstrate how a friction or diffusivity profile arises through a non-linear coordinate transformation. Consider the overdamped LE with the constant friction γ given by

$$\dot{q}(t) = -\frac{U'(q)}{\gamma} + \sqrt{\frac{k_B T}{\gamma}}\eta(t). \tag{37}$$

$U(q)$ denotes the potential of mean force and $\eta(t)$ is white noise. To use the results from section 9.B, we identify

$$h(q) = -\frac{U'(q)}{\gamma}, \qquad g(q) = \sqrt{\frac{k_B T}{\gamma}}. \tag{38}$$

Thus, we know that the drift coefficient is given by $D_1(q) = h(q) + g(q)g'(q) = -U'(q)/\gamma$ and the diffusion coefficient is obtained from $D_2(q) = g(q)^2 = k_BT/\gamma$ (see section 9.B). We now perform a coordinate transformation to a new variable $Q = \xi(q)$, where ξ is a non-linear function in q. Then Eq. (37) is transformed according to

$$\dot{Q} = \dot{\xi}(q) = \xi'(q)\,\dot{q} \tag{39a}$$

$$= \xi'(q)\left(-\frac{U'(q)}{\gamma} + \sqrt{\frac{k_BT}{\gamma}}\eta(t)\right) \tag{39b}$$

$$= \tilde{h}(Q) + \tilde{g}(Q)\,\eta(t), \tag{39c}$$

where we defined $\tilde{h}(Q) = -\xi'(q = \xi^{-1}(Q))U'(q = \xi^{-1}(Q))/\gamma$ and $\tilde{g}(Q) = \xi'(q = \xi^{-1}(Q))\sqrt{k_BT/\gamma}$. From the Langevin equation in Eq. (39c), it immediately follows that the transformed drift and diffusion coefficient in Q are given by

$$\tilde{D}_1(Q) = \tilde{h}(Q) + \tilde{g}(Q)\tilde{g}'(Q), \tag{40}$$

$$\tilde{D}_2(Q) = \tilde{g}(Q)^2. \tag{41}$$

By using the identity $\frac{\partial}{\partial q} = \xi'(q)\frac{\partial}{\partial Q}$, one finds

$$\tilde{D}_1(Q) = \frac{\tilde{D}_2(Q)}{k_BT}\frac{\partial}{\partial Q}\left(-U(Q) + k_BT\ln\xi'(Q),\right), \tag{42}$$

$$\tilde{D}_2(Q) = \xi'(Q)^2\frac{k_BT}{\gamma} \equiv \frac{k_BT}{\tilde{\gamma}(Q)}, \tag{43}$$

where ξ' denotes the derivative of ξ w.r.t. q. Hence, the overdamped LE in the coordinate Q has multiplicative noise and a friction profile given by $\tilde{\gamma}(Q) = \gamma/\xi'(Q)^2$.

11. Kramers-Moyal Expansion and Pawula Theorem

We here consider a general multidimensional stochastic process described by the trajectory $\mathbf{q}(t)$, in the main text we consider the special case of a two-dimensional process, i.e. $\mathbf{q} = (q, \dot{q}) \equiv (q, v)$. The Kramers-Moyal coefficients are derived from the transition probability $W(\mathbf{q}, t|\mathbf{q}', t')$, which is the probability that $\mathbf{q}(t) = \mathbf{q}$ given that $\mathbf{q}(t') = \mathbf{q}'$, for $t > t'$. The transition probability is a well-defined quantity for Markov and non-Markov processes. However, in contrast to Markov processes, for non-Markov processes $W(\mathbf{q}, t|\mathbf{q}', t')$ does not contain the full information of the dynamics, i.e., we can not describe the stochastic process completely based on the transition probabilities alone. For non-Markov processes, further information is required.

A. The Kramers-Moyal Expansion. To consider the multi-dimensional Kramers-Moyal expansion, it is useful to introduce the multi-index notation (11) which employs the multi-index $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_N) \in \mathbb{N}^N$. The convention is as follows: For some vector $\mathbf{r} \in \mathbb{R}^N$, the product of coefficients is defined as $\mathbf{r}^\alpha \equiv r_1^{\alpha_1} r_2^{\alpha_2} \cdots r_N^{\alpha_N}$ and we define $\alpha! \equiv \alpha_1!\alpha_2! \ldots \alpha_N!$. The absolute value of the multi-index is defined as $|\alpha| = \sum_{i=1}^N \alpha_i \equiv n$. Using the above described multi-index notation, the n-th multi-dimensional Kramers-Moyal coefficient with a finite lag time Δt can be defined as

$$D_\alpha^{(n)}(\mathbf{q}, t, \Delta t) = \frac{1}{n!\Delta t}\left\langle(\mathbf{q}(t + \Delta t) - \mathbf{q}(t))^\alpha\right\rangle_{\mathbf{q}(t)=\mathbf{q}}. \tag{44}$$

In equation Eq. (44), $\langle\ \ \rangle_{\mathbf{q}(t)=\mathbf{q}}$ denotes the conditional expectation value that at time t, $\mathbf{q}(t) = \mathbf{q}$. We start with the joint probability $P_2(\mathbf{q}, t + \Delta t; \mathbf{q}', t)$ of observing $\mathbf{q}$ at time $t + \Delta t$ and $\mathbf{q}'$ at time t with $\Delta t > 0$. For Markov and non-Markov processes, P_2 can be written as

$$P_2(\mathbf{q}, t + \Delta t; \mathbf{q}', t) = W(\mathbf{q}, t + \Delta t|\mathbf{q}', t)P_1(\mathbf{q}', t). \tag{45}$$

By integrating over $\mathbf{q}'$ we obtain

$$P_1(\mathbf{q}, t + \Delta t) = \int \mathrm{d}\mathbf{q}'\, W(\mathbf{q}, t + \Delta t|\mathbf{q}', t)P_1(\mathbf{q}', t), \tag{46}$$

where $\mathrm{d}\mathbf{q}' = \mathrm{d}q_1'\mathrm{d}q_2' \ldots \mathrm{d}q_N'$. Next we substitute $\Delta = \mathbf{q} - \mathbf{q}' \Rightarrow \mathrm{d}\Delta = -\mathrm{d}\mathbf{q}'$ and obtain

$$P_1(\mathbf{q}, t + \Delta t) = \int \mathrm{d}\Delta\, W(\mathbf{q}, t + \Delta t|\mathbf{q} - \Delta, t)P_1(\mathbf{q} - \Delta, t). \tag{47}$$

Now we expand the integrand in equation Eq. (47) in $\mathbf{q}$ around $\mathbf{q} + \Delta$

$$P_1(\mathbf{q}, t + \Delta t) = \int \mathrm{d}\Delta \sum_{|\alpha|\geq 0}\frac{\Delta^\alpha}{\alpha!}\left(-\frac{\partial}{\partial\mathbf{q}}\right)^\alpha W(\mathbf{q} + \Delta, t + \Delta t|\mathbf{q}, t)P_1(\mathbf{q}, t), \tag{48}$$

where the sum $\sum_{|\alpha|\geq 0}$ goes over all possible permutations and $|\alpha| = 0, 1, 2, \ldots, \infty$. Rearranging Eq. (48) gives

$$P_1(\mathbf{q}, t+\Delta t) = \sum_{|\alpha|\geq 0} \frac{n!}{\alpha!} \left(-\frac{\partial}{\partial \mathbf{q}}\right)^{\alpha} P_1(\mathbf{q}, t) \Delta t D_{\alpha}^{(n)}(\mathbf{q}, t, \Delta t), \tag{49}$$

where the equivalent definition of the finite-time Kramers-Moyal coefficients has been used

$$D_{\alpha}^{(n)}(\mathbf{q}, t, \Delta t) = \int \mathrm{d}\Delta \, \frac{1}{n! \Delta t} \Delta^{\alpha} W(\mathbf{q}+\Delta, t+\Delta t | \mathbf{q}, t). \tag{50}$$

From Eq. (50), it follows that $\Delta t D_{\alpha}^{(0)}(\mathbf{q}, t, \Delta t) = 1$. Therefore, we can write

$$P_1(\mathbf{q}, t+\Delta t) = P_1(\mathbf{q}, t) + \sum_{|\alpha|\geq 1} \frac{n!}{\alpha!} \left(-\frac{\partial}{\partial \mathbf{q}}\right)^{\alpha} P_1(\mathbf{q}, t) \Delta t D_{\alpha}^{(n)}(\mathbf{q}, t, \Delta t). \tag{51}$$

After rearranging we find

$$\frac{P_1(\mathbf{q}, t+\Delta t) - P_1(\mathbf{q}, t)}{\Delta t} = \sum_{|\alpha|\geq 1} \frac{n!}{\alpha!} \left(-\frac{\partial}{\partial \mathbf{q}}\right)^{\alpha} D_{\alpha}^{(n)}(\mathbf{q}, t, \Delta t) P_1(\mathbf{q}, t), \tag{52a}$$

$$= \sum_{n=1}^{\infty} \sum_{|\alpha|=n} \frac{n!}{\alpha!} \left(-\frac{\partial}{\partial \mathbf{q}}\right)^{\alpha} D_{\alpha}^{(n)}(\mathbf{q}, t, \Delta t) P_1(\mathbf{q}, t), \tag{52b}$$

where the sum $\sum_{|\alpha|=n}$ goes over all α with $|\alpha| = n$. From Eq. (50), it follows that $D_{\alpha}^{(n)}(\mathbf{q}, t, 0) = 0$ for $n > 0$. Therefore, taking the limit $\Delta t \to 0$ on both sides gives the standard Kramers-Moyal expansion of the Fokker-Planck equation

$$\frac{\partial P_1(\mathbf{q}, t)}{\partial t} = \sum_{n=1}^{\infty} \sum_{|\alpha|=n} \frac{n!}{\alpha!} \left(-\frac{\partial}{\partial \mathbf{q}}\right)^{\alpha} \lim_{\Delta t \to 0} D_{\alpha}^{(n)}(\mathbf{q}, t, \Delta t) P_1(\mathbf{q}, t) = L_{KM} P_1(\mathbf{q}, t), \tag{53}$$

where L_{KM} denotes the Kramers-Moyal operator.

B. Pawula Theorem. According to the Pawula Theorem the expansion in Eq. (53) either stops after the second term or it contains an infinite number of terms. This can be derived by using the generalized Schwarz inequality. For a non-negative function $P(\mathbf{q}) \geq 0$ and arbitrary functions $f(\mathbf{q})$ and $g(\mathbf{q})$, one finds

$$\left(\int \mathrm{d}\mathbf{q} \, P(\mathbf{q}) f(\mathbf{q}) g(\mathbf{q}) \right)^2 \leq \int \mathrm{d}\mathbf{q} \, P(\mathbf{q}) \, f(\mathbf{q})^2 \int \mathrm{d}\mathbf{q} \, P(\mathbf{q}) \, g(\mathbf{q})^2. \tag{54}$$

When $P(\mathbf{q})$ is the transition probability $W(\mathbf{q}+\Delta, t+\Delta t|\mathbf{q}, t)$, and $f(\mathbf{q}) \to \Delta^{\beta}$, $g(\mathbf{q}) \to \Delta^{\gamma}$, where the multi-indices β, γ are a partition of α, i.e., $|\alpha| = |\beta| + |\gamma| = n + (n+m)$ and $\alpha! = \beta! \gamma!$, the Schwarz inequality implies for the Kramers Moyal coefficients (7)

$$\left(D_{\alpha}^{(2n+m)}\right)^2 \leq \frac{(2n)! \, (2(n+m))!}{[(2n+m)!]^2} D_{\beta}^{(2n)} D_{\gamma}^{(2(n+m))}, \qquad \text{for } n, m \geq 1. \tag{55}$$

The factorial prefactors follow from the definition Eq. (44). If $D_{\beta}^{(2n)} = 0$, then from Eq. (55) we find that $D_{\alpha}^{(2n+1)} = D_{\alpha}^{(2n+2)} = \cdots = 0$, i.e.,

$$D_{\beta}^{(2n)} = 0 \;\Rightarrow\; D_{\alpha}^{(2n+m)} = 0, \qquad \text{for } m \geq 1. \tag{56}$$

From Eq. (55), it also follows that if $D_{\gamma}^{(2(n+m))} = D_{\gamma}^{(2r)} = 0$, we find that the coefficients $D_{\alpha}^{(2n+m)} = D_{\alpha}^{(2(n+m)-m)} = D_{\alpha}^{(2r-m)}$ for all combinations of n, m with $n + m = r$ vanish, meaning

$$D_{\gamma}^{(2r)} = 0 \;\Rightarrow\; D_{\alpha}^{(2r-m)} = 0 \qquad \text{for } m = 1, 2, \ldots, r-1. \tag{57}$$

Combining Eq. (56) with a repeated use of Eq. (57) gives in summary the following non-trivial properties:

1. The expansion stops after the first order, which means that there is no stochastic part in the corresponding Langevin equation. This become clear by consideration of Section 9: According to Eq. (27b), if all second order KMCs vanish, the amplitude of the random forces in the LE also vanish. The fact that the expansion may stop after the first order follows from Eq. (56) for $n = 1$. If the second order KMCs are zero, so are all higher orders.

2. The expansion stops after the second order, i.e, the Kramers Moyal expansion becomes the ordinary Fokker-Planck equation, which describes a diffusion process. This follows from Eq. (56) and Eq. (57). If we consider $n = 2$ in Eq. (56), it follows that if the fourth order KMCs are zero, then all orders higher than four are also zero. On the other hand, Eq. (57) says that if the fourth order KMCs are zero, the third order KMCs must also be zero. Hence, all orders larger than two must vanish if the fourth order KMCs turn out to be zero.

3. The expansion contains an infinite number of terms. This can be seen by setting the third order KMCs to be non-zero. Then, according to Eq. (57), the fourth order KMCs can not be zero. Furthermore, if the fourth order KMCs are non-zero, Eq. (57) again implies that then the sixth order KMCs must also be non-zero. In this way, the repeated use of Eq. (57) implies that the Kramers Moyal expansion contains an infinite number of terms.

12. Kramers Moyal Coefficients for the GLE

To obtain analytical approximations for the KMCs from the GLE, we use the general results for stationary Gaussian processes derived below in subsection A. To approximate $D_q(q, v)$, we integrate the relation $\dot{q}(t) = v(t)$ from t to $t + \Delta t$, which gives

$$q(t + \Delta t) - q(t) = \int_t^{t+\Delta t} \mathrm{d}s\, v(s). \tag{58}$$

For small Δt, the integral on the r.h.s. can be approximated by the trapezoidal rule as

$$q(t + \Delta t) - q(t) = \frac{v(t + \Delta t) + v(t)}{2}\,\Delta t + \mathcal{O}(\Delta t^3) = v(t)\Delta t + \frac{v(t + \Delta t) - v(t)}{2}\Delta t + \mathcal{O}(\Delta t^3). \tag{59}$$

By taking the conditional average on both sides and dividing by Δt, we can make use of the definition of D_q given by

$$D_q(q, v, \Delta t) = \frac{1}{\Delta t}\left\langle q(t + \Delta t) - q(t)\right\rangle_{\substack{q(t)=q \\ v(t)=v}}, \tag{60}$$

which is a special case of Eq. (44). In this way, we find D_q in terms of D_v for finite Δt

$$D_q(q, v, \Delta t) = v + \frac{\Delta t}{2} D_v(q, v, \Delta t) + \mathcal{O}(\Delta t^2). \tag{61}$$

To compute D_v in the presence of a potential, we divide the GLE in Eq. (17) by the mass m and integrate from t to $t + \Delta t$

$$v(t + \Delta t) - v(t) = \int_t^{t+\Delta t} \mathrm{d}s\, \left(-\frac{U'[q(s)]}{m} - \int_0^s \mathrm{d}s'\, \Gamma(s - s')v(s') + F_R(s) \right). \tag{62}$$

The function $U'(q(s))/m$ is expanded around $s = t$ and to leading order in Δt, we obtain

$$v(t + \Delta t) - v(t) = -\Delta t \frac{U'[q(t)]}{m} + \mathcal{O}(\Delta t^2) + \int_t^{t+\Delta t} \mathrm{d}s\, \left(-\int_0^s \mathrm{d}s'\, \Gamma(s - s')v(s') + F_R(s) \right). \tag{63}$$

Again, we take the conditional average on both sides and divide by Δt. The conditional average of the integral part on the r.h.s. is approximated by Eq. (79), i.e., by the result for D_v in the absence of a potential U. For this approximation to work, the velocity distribution should be independent of the position q, such that we can neglect the condition on q in the conditional averaging of the velocity in the integral. In Fig. S1, we show that the probability distribution of the velocity is indeed independent of q, i.e., $P(v|q) \approx P(v)$. Therefore, approximating conditional averages over q and v by conditional averages over v alone is a valid approximation. Also neglecting the impact of the potential U on the memory kernel we find

$$D_v(q, v, \Delta t) \approx -\frac{U'(q)}{m} + \widetilde{D}_v(v, \Delta t) + \mathcal{O}(\Delta t), \tag{64}$$

where $\widetilde{D}_v(v, \Delta t)$ denotes the drift coefficient in the absence of a potential U and is given in Eq. (79). To compute D_{qq}, we square Eq. (59) and find

$$(q(t + \Delta t) - q(t))^2 = \frac{(v(t + \Delta t) - v(t))^2}{4}\Delta t^2 + v(t)v(t + \Delta t)\Delta t^2 + \mathcal{O}(\Delta t^6). \tag{65}$$

Taking the conditional average and dividing by $2\Delta t$ and using the definitions of D_{qq} and D_{vv} gives

$$D_{qq}(q, v, \Delta t) = \frac{\Delta t^2}{4} D_{vv}(q, v, \Delta t) + \frac{\Delta t}{2}\langle v(t + \Delta t)v(t)\rangle_{q,v} + \mathcal{O}(\Delta t^5). \tag{66}$$

The conditional autocorrelation function on the r.h.s. is approximated by Eq. (81). This is again motivated by the results shown in Fig. S1, i.e., by the fact that $P(v|q) \approx P(v)$, which yields

$$D_{qq}(q, v, \Delta t) = \frac{\Delta t^2}{4} D_{vv}(q, v, \Delta t) + \frac{\Delta t}{2}\frac{C_{vv}(\Delta t)}{C_{vv}(0)} v^2 + \mathcal{O}(\Delta t^5). \tag{67}$$

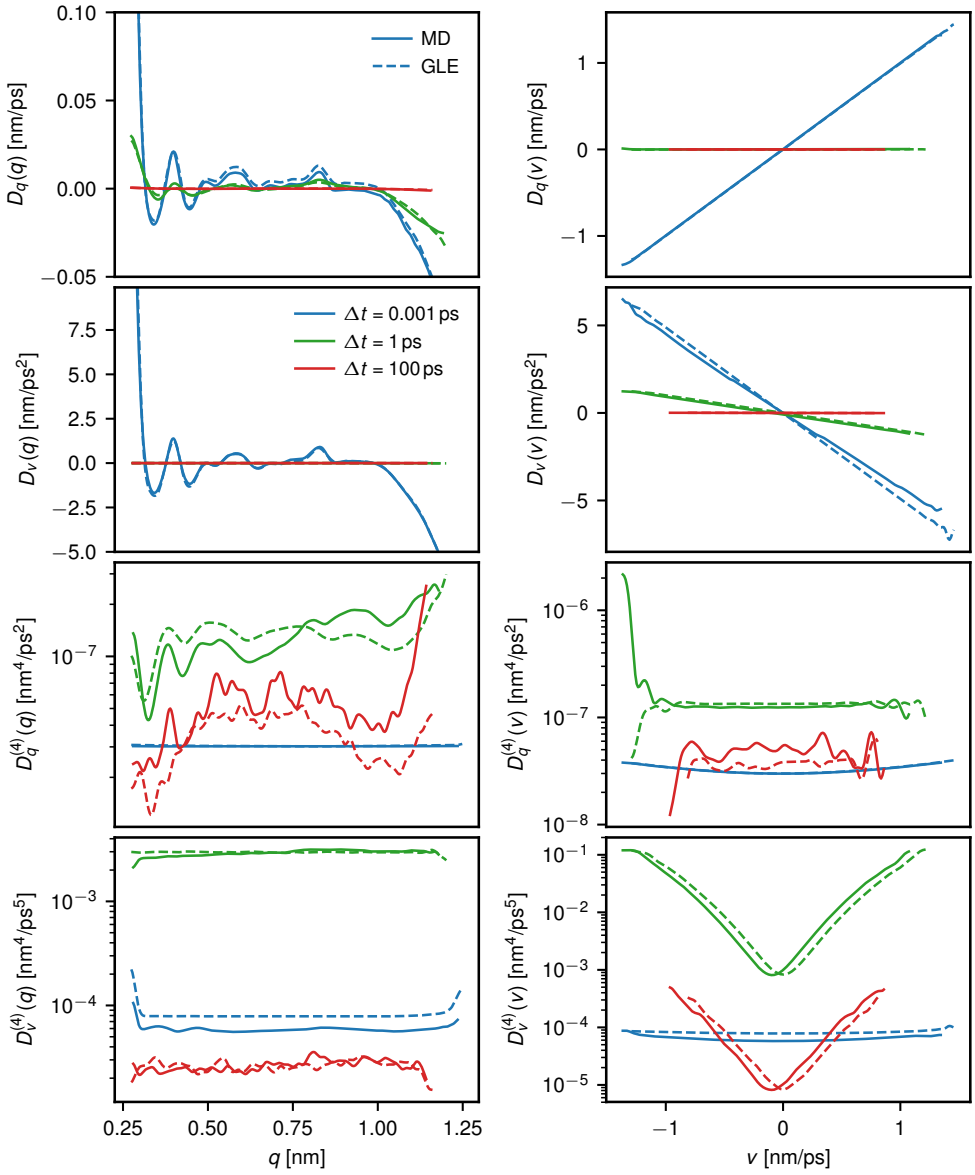

Fig. S9. Comparison of KMCs computed from MD and GLE trajectories for different lag times Δt.

In an analogous manner, $D_{vv}(q,v)$ is approximated by squaring Eq. (63) and using Eq. (83b), which gives

$$D_{vv}(q,v,\Delta t) = \frac{\Delta t}{2}\left(\frac{U'(q)}{m}\right)^2 + \widetilde{D}_{vv}(v,\Delta t) + \mathcal{O}(\Delta t^2). \quad [68]$$

If we take the limit $\Delta t \to 0$, only the first order KMCs do not vanish, as follows from Eq. (61), Eq. (64), Eq. (67) and Eq. (68)), so that we obtain

$$\lim_{\Delta t\to 0} D_q(q,v,\Delta t) = v, \qquad \lim_{\Delta t\to 0} D_v(q,v,\Delta t) = -\frac{U'(q)}{m}, \quad [69a]$$

$$\lim_{\Delta t\to 0} D_{qq}(q,v,\Delta t) = 0, \qquad \lim_{\Delta t\to 0} D_{vv}(q,v,\Delta t) = 0. \quad [69b]$$

Hence, in the limit of Δt going to zero, the KMCs of a GLE only reproduce the deterministic part of the dynamics and contain no information on the stochastic part.

By averaging over q with a weight function $\propto e^{-\beta U(q)}$ or over v with a weight function $\propto e^{-\beta m v^2/2}$, the results for finite Δt, Eq. (61), Eq. (64), Eq. (67) and Eq. (68)), read to leading order

$$D_q(q,\Delta t) = -\frac{\Delta t}{2}\frac{U'(q)}{m} + \mathcal{O}(\Delta t^2) \quad [70a]$$

$$D_q(v,\Delta t) = v + \mathcal{O}(\Delta t), \quad [70b]$$

$$D_{qq}(q,\Delta t) = \frac{\Delta t}{2}C_{vv}(\Delta t) - \frac{\Delta t^2}{4}\bar{C}(\Delta t)\Delta C(\Delta t) + \mathcal{O}(\Delta t^3), \quad [70c]$$

$$D_{qq}(v,\Delta t) = \frac{\Delta t}{2}\frac{C_{vv}(\Delta t)}{C_{vv}(0)}v^2 - \frac{\Delta t^2}{4}\bar{C}(\Delta t)\Delta C(\Delta t) + \mathcal{O}(\Delta t^3), \quad [70d]$$

$$D_v(q,\Delta t) = -\frac{U'(q)}{m} + \mathcal{O}(\Delta t), \quad [70e]$$

$$D_v(v,\Delta t) = v\Delta C(\Delta t) + \mathcal{O}(\Delta t), \quad [70f]$$

$$D_{vv}(q,\Delta t) = -\bar{C}(\Delta t)\Delta C(\Delta t) + \frac{\Delta t}{2}\left(\frac{U'(q)}{m}\right)^2 + \mathcal{O}(\Delta t^2), \quad [70g]$$

$$D_{vv}(v,\Delta t) = -\bar{C}(\Delta t)\Delta C(\Delta t) + \frac{\Delta t}{2}\left\langle\left(\frac{U'(q)}{m}\right)^2\right\rangle + \mathcal{O}(\Delta t^2). \quad [70h]$$

The quantities $\bar{C}(\Delta t) = (C_{vv}(\Delta t) + C_{vv}(0))/2$ and $\Delta C(\Delta t) = (C_{vv}(\Delta t) - C_{vv}(0))/(\Delta t\, C_{vv}(0))$ are defined in subsection A below. In Fig. S9 we compare the marginalized KMCs computed from MD and GLE data for various Δt and find very good agreement. The GLE model reproduces the KMCs from the MD trajectories over many orders of magnitude of the lag time Δt. In Fig. S10, we compare Eq. (70c) and Eq. (70g) with numerically computed KMCs from MD and GLE simulations as a function of the lag time Δt. For this, we average the numerical D_{vv} over v and evaluate it in the vicinity of the minimum at $q = 0.32\,\text{nm}$ of $U(q)$, i.e., where the mean force vanishes. The numerical D_{qq} is averaged over q and v. As can be seen in Fig. S10, the analytical approximations describe the KMCs very well for small Δt, as expected. This also confirms that the numerical computation of the KMCs is accurate.

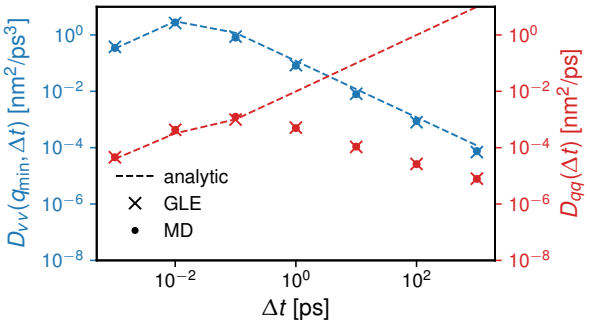

Fig. S10. The second-order KMCs $D_{qq}(\Delta t)$ and $D_{vv}(q_{\min},\Delta t)$ as a function of the lag time Δt. D_{vv} is averaged over v and is evaluated at the free energy minimum at $q_{\min} = 0.32\,\text{nm}$. D_{qq} is averaged over q and v. The data points show the KMCs computed from the MD (circles) and GLE (crosses) trajectories. The dashed lines show the analytic expressions in Eq. (70g) (blue) and Eq. (70c) (red).

A. Kramers Moyal Coefficients of a Stationary Gaussian Process. Here, we derive the KMCs for a stationary Gaussian process, i.e. for a GLE Eq. (17) in the absence of a potential U. The results of this subsection are used above to approximately compute the KMCs of a GLE. To proceed, we consider the velocity $v(\cdot)$ as an N dimensional stationary Gaussian process,

$$v(\cdot) \;\to\; \mathbf{v} = (v_1, v_2, \ldots, v_N)^T, \quad [71]$$

$$P(\mathbf{v}) = \frac{1}{\sqrt{(2\pi)^N \det\sigma}}\, e^{-\frac{1}{2}\mathbf{v}^T\cdot\sigma^{-1}\cdot\mathbf{v}}, \quad [72]$$

where $\sigma_{ij} = \langle v_i, v_j\rangle = \langle v_{|i-j|}, v_0\rangle$ denotes the covariance matrix of the N dimensional stationary Gaussian distribution $P(v)$. We can use Eq. (72) to compute the conditional probability distribution. For this, we partition $P(\mathbf{v})$ using

$$\mathbf{v}^T\cdot\sigma^{-1}\cdot\mathbf{v} = (\mathbf{v}_{N-1}\, v_N)\cdot\begin{pmatrix} M & O \\ O^T & N\end{pmatrix}\cdot\begin{pmatrix}\mathbf{v}_{N-1}\\ v_N\end{pmatrix}, \quad [73]$$

where we introduce the vectors $O \in \mathbb{R}^{N-1}$ and $\mathbf{v}_{N-1} \in \mathbb{R}^{N-1}$ and the matrix $M \in \mathbb{R}^{(N-1)\times(N-1)}$. $v_N = v$ denotes the value of the velocity at time t. The partitioning in Eq. (73) leads to

$$P(\mathbf{v}) = P_1(v_N)P_{N-1}(\mathbf{v}_{N-1}|v_N). \quad [74]$$

$P_{N-1}(\mathbf{v}_{N-1}|v_N)$ is the conditional probability distribution we must use when computing conditional averages. It has the form

$$P_{N-1}(\mathbf{v}_{N-1}|v_N) = \sqrt{\frac{\det M}{(2\pi)^N}}\, e^{-\frac{1}{2}\left(\mathbf{v}_{N-1}+M^{-1}Ov_N\right)^T\cdot M\cdot\left(\mathbf{v}_{N-1}+M^{-1}Ov_N\right)}. \quad [75]$$

As can be seen in Eq. (75), the mean of the distribution is shifted. We find

$$\langle v_i \rangle_{v_N = v} = -(M^{-1} O v_N)_i = \frac{\langle v_{|N-i|}, v_0 \rangle}{\langle v_0^2 \rangle} \, v. \tag{76}$$

In the continuous case, Eq. (76) becomes

$$\langle v(s) \rangle_{v(t)=v} = \frac{C_{vv}(|t-s|)}{C_{vv}(0)} \, v, \tag{77}$$

where C_{vv} denotes the velocity autocorrelation function.
We use Eq. (77) in the conditional expectation $\langle v(t+\Delta t) - v(t) \rangle_v = \langle v(t+\Delta t) \rangle_v - \langle v(t) \rangle_v$, and find

$$\langle v(t+\Delta t) - v(t) \rangle_v = \frac{C_{vv}(\Delta t) - C_{vv}(0)}{C_{vv}(0)} \, v. \tag{78}$$

Inserting the last result into Eq. (44), we obtain the first order KMC for a stationary Gaussian process

$$\widetilde{D}_v(v, \Delta t) = \frac{C_{vv}(\Delta t) - C_{vv}(0)}{\Delta t} \frac{v}{C_{vv}(0)} = \Delta C(\Delta t) \, v, \tag{79}$$

where $\widetilde{D}_v(v, \Delta t)$ denotes the KMC in the absence of a potential U and $\Delta C(\Delta t) \equiv (C_{vv}(\Delta t) - C_{vv}(0))/(\Delta t \, C_{vv}(0))$.
To compute D_{vv} we start from

$$\left\langle (v(t+\Delta t) - v(t))^2 \right\rangle_v = \left\langle v(t+\Delta t)^2 \right\rangle_v + \left\langle v(t)^2 \right\rangle_v - 2 \left\langle v(t+\Delta t) v(t) \right\rangle_v . \tag{80}$$

For the conditional velocity autocorrelation function, we obtain similarly as for the conditional average velocity in Eq. (77)

$$\langle v(s), v(s') \rangle_{v(t)=v} = C_{vv}(s-s') - \frac{C_{vv}(t-s) C_{vv}(t-s')}{C_{vv}(0)} + \frac{C_{vv}(t-s) C_{vv}(t-s')}{C_{vv}(0)^2} v^2. \tag{81}$$

Inserting this into Eq. (80) gives

$$\left\langle (v(t+\Delta t) - v(t))^2 \right\rangle_v = -\left(C_{vv}(\Delta t) + C_{vv}(0)\right) \frac{C_{vv}(\Delta t) - C_{vv}(0)}{C_{vv}(0)} + \left(\frac{C_{vv}(\Delta t) - C_{vv}(0)}{C_{vv}(0)} \, v \right)^2 . \tag{82}$$

From this the second order KMC immediately follows as

$$\widetilde{D}_{vv}(v, \Delta t) = -\frac{\left(C_{vv}(\Delta t) + C_{vv}(0)\right)}{2} \frac{\left(C_{vv}(\Delta t) - C_{vv}(0)\right)}{\Delta t} \frac{1}{C_{vv}(0)} + \frac{\Delta t}{2} \left(\frac{C_{vv}(\Delta t) - C_{vv}(0)}{\Delta t} \frac{v}{C_{vv}(0)} \right)^2 \tag{83a}$$

$$= -\bar{C}(\Delta t) \, \Delta C(\Delta t) + \Delta t \left(\Delta C(\Delta t) \, v\right)^2 / 2, \tag{83b}$$

where we defined $\bar{C}(\Delta t) \equiv \left(C_{vv}(\Delta t) + C_{vv}(0)\right)/2$ and $\Delta C(\Delta t)$ has been defined above.

13. Numerical Computation of KMCs: Kernel Density Estimators

To compute the KMCs numerically, we divide our parameter space (q, v) into a grid. In each cell of the grid, we must find an estimate for the conditional probability distribution $P(q, v, t+\Delta t | q', v', t)$, which is usually accomplished by histograms. When the statistics in a cell are insufficient, i.e., when the number of data points in a cell is too small, the estimate of the probability distribution via a histogram depends significantly on the choice of the boundaries of the cell. To avoid such ambiguities, we use kernel density estimators (12). Kernel density estimators have the advantage that each point of occurrence within a cell contributes to the estimate of the density at its point of occurrence and not in a range of a bin. This means that, for a given data set $\mathbf{q} \in \mathbb{R}^N$, a p.d.f. $P(q)$ is estimated by $\widehat{P}(q)$ via

$$\widehat{P}(q) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{q - q_i}{h} \right). \tag{84}$$

Here K denotes the kernel with bandwidth h. For example, one could use a Gaussian kernel

$$K\left(\frac{q - q_i}{h} \right) \propto e^{-\frac{1}{2}\left(\frac{q-q_i}{h}\right)^2}. \tag{85}$$

In this paper, we used an Epanechnikov kernel

$$K\left(\frac{q - q_i}{h} \right) \propto \max\left(1 - \left(\frac{q - q_i}{h} \right)^2, 0 \right), \tag{86}$$

where the optimal bandwidth h was estimated by the inverse of the grid size. For small time steps $\Delta t = 0.001$ ps, the grid size in q direction is 5000 and in v direction it is 300. Hence, the respective bandwidths are 0.0002 and 0.003. For larger time steps $\Delta t \geq 0.001$ ps, the grid size in q direction is decreased to 500 and therefore the bandwidth is 0.002.

References

1. G Bussi, D Donadio, M Parrinello, Canonical sampling through velocity rescaling. *The J. Chem. Phys.* **126**, 014101 (2007).
2. MD Hanwell, et al., Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminformatics* **4**, 17 (2012).
3. B Kowalik, et al., Memory-kernel extraction for different molecular solutes in solvents of varying viscosity in confinement. *Phys. Rev. E* **100**, 012126 (2019).
4. H Mori, Transport, Collective Motion, and Brownian Motion. *Prog. Theor. Phys.* **33**, 423–455 (1965).
5. R Zwanzig, Memory Effects in Irreversible Thermodynamics. *Phys. Rev.* **124**, 983–992 (1961).
6. R Zwanzig, *Nonequilibrium statistical mechanics*. (Oxford UnivPress, Oxford [u.a.]), (2001).
7. H Risken, Fokker-Planck Equation in *The Fokker-Planck Equation: Methods of Solution and Applications*, Springer Series in Synergetics, ed. H Risken. (Springer, Berlin, Heidelberg), pp. 63–95 (1996).
8. H Flyvbjerg, Error estimates on averages of correlated data in *Advances in Computer Simulation*, Lecture Notes in Physics, eds. J Kertész, I Kondor. (Springer, Berlin, Heidelberg), pp. 88–103 (1998).
9. A Möglich, K Joder, T Kiefhaber, End-to-end distance distributions and intrachain diffusion constants in unfolded polypeptide chains indicate intramolecular hydrogen bond formation. *Proc. Natl. Acad. Sci.* **103**, 12394–12399 (2006).
10. EV Kuzmenkina, CD Heyes, GU Nienhaus, Single-molecule förster resonance energy transfer study of protein dynamics under denaturing conditions. *Proc. Natl. Acad. Sci.* **102**, 15471–15476 (2005).
11. XS Raymond, *Elementary Introduction to the Theory of Pseudodifferential Operators*. (CRC Press), (1991) Google-Books-ID: kD5ZCJDIg4oC.
12. LR Gorjão, F Meirinhos, kramersmoyal: Kramers–Moyal coefficients for stochastic processes. *J. Open Source Softw.* **4**, 1693 (2019).