Journal of Physics A: Mathematical and Theoretical

**PAPER** • **OPEN ACCESS**

# Time- and ensemble-average statistical mechanics of the Gaussian network model

To cite this article: Alessio Lapolla *et al* 2021 *J. Phys. A: Math. Theor.* **54** 355601

OPEN ACCESS
IOP Publishing
Journal of Physics A: Mathematical and Theoretical
J. Phys. A: Math. Theor. **54** (2021) 355601 (30pp)
https://doi.org/10.1088/1751-8121/ac16e4

# Time- and ensemble-average statistical mechanics of the Gaussian network model

**Alessio Lapolla, Maximilian Vossel and Aljaž Godec***

Mathematical bioPhysics Group, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11, Göttingen 37077, Germany

E-mail: agodec@mpibpc.mpg.de

Received 19 February 2021, revised 7 June 2021
Accepted for publication 22 July 2021
Published 9 August 2021

**Abstract**

We present analytical results for a set of time- and ensemble-averaged physical observables in the non-Hookean Gaussian network model (GNM)—a generalization of the Rouse model to elastic networks with links that display a certain degree of extensional and rotational stiffness. We focus on a set of coarse-grained observables that may be of interest in the analysis of GNM in the context of internal motions in proteins and mechanical frames in contact with a heat bath. A C++ computer code is made available that implements all analytical results.

Keywords: conformation dynamics, projected physical observables, elastic network, network models of macromolecules, Fokker–Planck equation, occupation time, covariance matrix

(Some figures may appear in colour only in the online journal)

## 1. Introduction

Proteins utilize their unique dynamic character encoded in internal motions to execute a biological function [1]. These motions span fs to s time-scales and their study thus requires a multitude of experimental and/or computational methods [1]. The most detailed, atomically resolved information about these motions comes—with a grain of salt because of an underlying approximate, empirical potential energy function—from molecular dynamics (MD) simulations [2, 3]. However, even if the state-of-the-art hardware and highly parallel algorithms allow to reach ms time-scales [4] a substantial time-scale gap remains. In addition, the sheer amount of detail in such *tour de force* simulations [4] often poses a challenge if one aims at extracting

*Author to whom any correspondence should be addressed.

minimal, 'leading order' physical principles underlying protein internal motions. Moreover, physical or even topological properties alone may accurately predict selected features of protein dynamics [5, 6].

To describe internal motions in proteins on an effective, coarse-grained level disregarding chemical details Tirion introduced the so-called elastic network model (ENM) [7] akin to the seminal works of Rouse [8] and Flory [9] in polymer physics. The basic idea underlying ENMs is a tridimensional elastic network connecting those residues, more precisely the respective C$\alpha$ atoms, that lie within a cutoff distance typically chosen in the range 7–16 Å. Subsequent works considered various alternative models, e.g. so-called Gaussian network model (GNM) [10, 11] and the anisotropic network model (ANM) [12–14].

ENMs in various forms have been successfully applied (and extended) to refine NMR- [15] and x-ray crystallography-derived protein structures [16], derive NMR-structural order parameters [17], investigate structural correlations [18], function [19–22], conformational transitions [23], and allosteric effects [24] in proteins, and to identify and decompose protein domains [25]. Further applications involve improving MD simulations [26], the study of protein evolution [27], investigations of smart polymers [28, 29], viruses [30], membrane channels [31, 32], and nucleic acids [33], as well as the prediction of rupture points in single-molecule pulling experiments [34],

Most of these works rely on 'standard' normal mode analysis (NMA) [12, 35], i.e. on spectral characteristics of the underlying mechanical vibration spectrum. In the particular context of proteins NMA has been used predominantly to identify the large-scale collective motions encoded in the eigenvector corresponding to the principal eigenvalue of the Hessian. Notably, the low-frequency modes are quite insensitive to the precise value of the cutoff distance [36].

Here we go beyond and present analytical results for time- and ensemble-average characteristics of internal 'reaction coordinates' in GNM in contact with a heat bath at a finite temperature. More precisely, we consider the non-Markovian dynamics of internal distances at equilibrium. Our results may be relevant for interpreting single-molecule spectroscopy data or MD simulations.

## 2. The GNM

The Rouse model [8] is one of the earliest 'elastic network' models of flexible linear polymers (later on extended to more general network structures [9]). It neglects excluded volume effects and hydrodynamic interactions. Within this theoretical framework beads are connected by ideal, Hookean springs with vanishing resting length (i.e. at $T = 0$ the beads' positions would coincide). The strength of the springs is proportional to the temperature $T$ of the heat bath. The model does not accurately capture the features of molecules with a non-negligible internal rigidity.

ENMs [7] extend these core ideas by including a non-zero resting length, i.e. at $T = 0$ the residues are assumed to have distinct positions that are fixed in space. This idea is consistent with the results of NMR and x-ray crystallography that yield a set of positions $\mathbf{R}^0 = \{\mathbf{r}_i^0\}$ of the $N + 1$ residues to which we refer as 'the structure' of a protein (NMR experiments in fact yield an ensemble of such structures). These experiments are able to probe the positions of all atoms of a macromolecule at cryogenic temperatures, and are typically deposited in the protein data bank. We will assume that these experimentally determined positions form the set $\mathbf{R}^0$.

In GNMs a pair of residues $i$, $j$ within a cutoff distance (i.e. $|\mathbf{r}_i^0 - \mathbf{r}_i^0| \leqslant r_c$) are assumed to be connected by *identical* but *non-Hookean* springs with a constant $K$. The interaction energy

as a function of the particles' positions $\mathbf{R} = \{\mathbf{r}_i\}$ is written as

$$U_{\mathrm{GNM}}(\{\mathbf{r}_{ij}\}) = \frac{K}{2}\sum_{\langle i,j\rangle}(\mathbf{r}_{ij} - \mathbf{r}^0_{ij})^{\mathrm{T}}(\mathbf{r}_{ij} - \mathbf{r}^0_{ij}), \tag{1}$$

where the sum spans all connected pairs. We now introduce for convenience the *deviation from the equilibrium* '*structure*', $\Delta\mathbf{R} = \{\Delta\mathbf{r}_i \equiv \mathbf{r}_i - \mathbf{r}^0_i\}$. The main simplifying hypothesis of the GNM is that $\Delta\mathbf{R}$ at temperature $T$ corresponds to an isotropic Gaussian random vector, where each element $\mathbf{R}_i = \mathbf{r}_i - \mathbf{r}^0_i$ is a tridimensional vector describing the displacement of bead $i$ from its equilibrium position $\mathbf{r}^0_i$, i.e.

$$P(\Delta\mathbf{R}) = \left[(2\pi)^N \tilde{K}\ \det\Gamma^{-1}\right]^{-3/2} \exp\left(-\frac{\tilde{K}}{2}\Delta\mathbf{R}^{\mathrm{T}}\boldsymbol{\Gamma}\Delta\mathbf{R}\right), \tag{2}$$

where $\tilde{K} \equiv K/k_{\mathrm{B}}T$ is the dimensionless strength (in units of thermal energy $k_{\mathrm{B}}T$) and $\boldsymbol{\Gamma}$ is a $3(N+1)\times 3(N+1)$ block matrix in which each diagonal block is the positive semi-definite Laplacian (or Kirchhoff) matrix $\Gamma$ with elements

$$\Gamma_{ij} = \begin{cases} -1, & \text{if } i \neq j \text{ and } |\mathbf{r}^0_i - \mathbf{r}^0_j| \leqslant r_{\mathrm{c}} \\ 0, & \text{if } i \neq j \text{ and } |\mathbf{r}^0_i - \mathbf{r}^0_j| > r_{\mathrm{c}} \\ -\sum_{j;j\neq i}^{N+1}\Gamma_{ij}, & \text{if } i = j. \end{cases} \tag{3}$$

Two pictorial representations of the connectivity matrix obtained from a pair of protein structures are shown in figure 1. The dynamics of the beads' positions (i.e. deviations from the equilibrium 'structure') is assumed to follow the Itô equation

$$\mathrm{d}\Delta\mathbf{R}(t) = -\xi K\boldsymbol{\Gamma}\Delta\mathbf{R}(t)\mathrm{d}t + \sqrt{2D}\,\mathrm{d}\mathbf{W}(t), \tag{4}$$

where $D$ is the diffusion coefficient and $\xi \equiv D/k_{\mathrm{B}}T$ the mobility that are both assumed to be equal for all beads for the sake of simplicity, and $\mathrm{d}\mathbf{W}(t)$ is the increment of the multi-dimensional Wiener process (i.e. Gaussian white noise) with zero mean and covariance $\langle \mathrm{d}W_i(t)\mathrm{d}W_j(t')\rangle = \delta_{ij}\delta(t-t')$. Henceforth we measure energy in units of thermal energy $k_{\mathrm{B}}T$ (i.e. $U \to U/k_{\mathrm{B}}T$), distances in units of the cutoff distance $r_{\mathrm{c}}$ (i.e. $\Delta R_i \to \Delta R_i/r_{\mathrm{c}}$) and time in units of the *diffusion time*, $t_{\mathrm{D}} \equiv r_{\mathrm{c}}^2/D$—the time required for a bead with a diffusion coefficient $D$ to diffuse a distance $r_{\mathrm{c}}$ (i.e. $t \to t/t_{\mathrm{D}}$).

In order to generalize the GNM one may consider non-equal spring constants, i.e. $K \to K_{ij}$. In this case the matrix $\boldsymbol{\Gamma}$ loses the simple structure in equation (3) and corresponding symmetric spectral characteristics. However, it has been shown that such complications do not lead to any appreciable improvements [7, 12]. Alternatively, ANM in addition to an anisotropic stiffness $K_{ij}$ display the distinguishing characteristic that the interaction energy is given by [12]

$$U_{\mathrm{ANM}}(\{\mathbf{r}_{ij}\}) = \frac{1}{2}\sum_{\langle i,j\rangle}K_{ij}(|\mathbf{r}_{ij}| - |\mathbf{r}^0_{ij}|)^2, \tag{5}$$

such that $U_{\mathrm{ANM}}(\{\mathbf{r}_{ij}\})$ is invariant with respect to bond orientations in absence of changes of the bond lengths. We are not able to treat the ANM model analytically. Notably, it has been found that the predictions of the GNM often agree better with experimental data than the predictions of the ANM [11, 37]. It is conceivable that this is rooted in the fact that interactions between

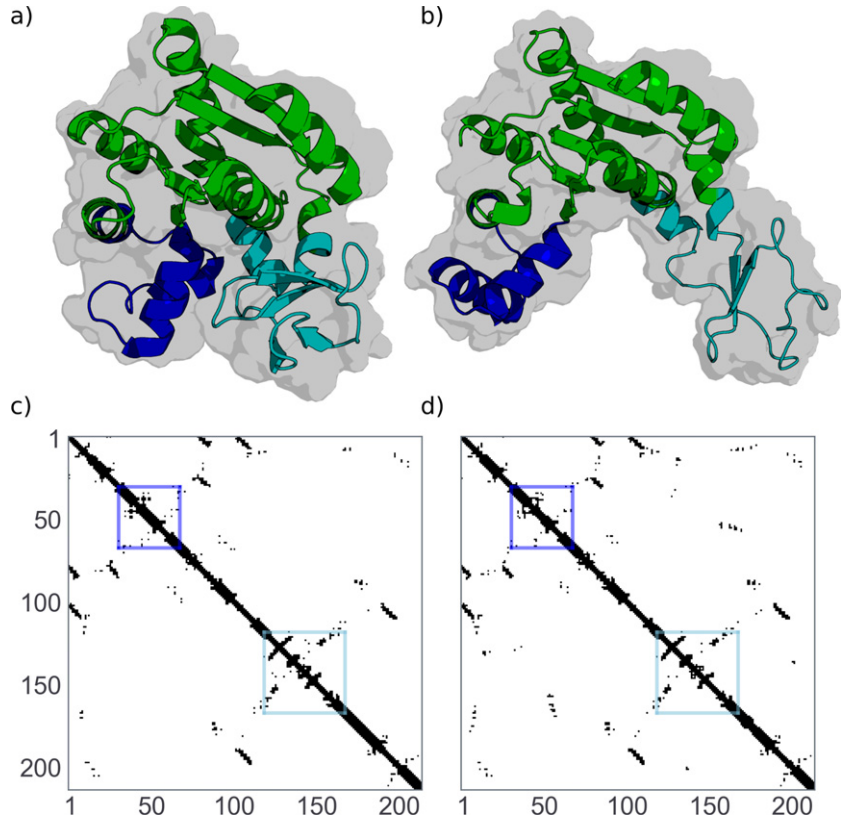


**Figure 1.** Panels (a) and (b) depict a cartoon and the molecular surface (gray) of the two protein structures, called (a) 'the closed' configuration 1AKE and (b) 'the open' configuration 4AKE. Panels (c) and (d) show the corresponding connectivity matrices for 1AKE and 4AKE, respectively. The blue and cyan square enclose, respectively, the NMP and LID residues. The cutoff distance used to obtain these matrices was 8 Å.

sets of chemical groups in general depend on the orientation [7] (e.g. the effective electrostatic interactions between groups of charges that represent individual beads depend on the center of mass distance between the groups *and* the respective mutual orientation, which follows trivially from a multipole expansion).

It is convenient to pass to the vector of normal tridimensional coordinates $\mathbf{Q} = \{\mathbf{q}_k\}$ that diagonalize $\mathbf{\Gamma}$, i.e. $\mathbf{Q}^{\mathrm{T}}\mathbf{\Gamma}\mathbf{Q} = \mathrm{diag}(\boldsymbol{\mu})$ with $(\mathbf{Q})_{ij} \equiv Q_{ij}\mathbb{1}$, $\mathbb{1}$ being the $3 \times 3$ identity matrix and where the matrix $Q$ diagonalizes the Kirchoff matrix, i.e. $Q^{\mathrm{T}}\Gamma Q = \mathrm{diag}(\mu_i)$, and therefore $\mathrm{diag}\,(\boldsymbol{\mu})_{ii} = \mu_i\mathbb{1}$. With a few exceptions that are, e.g. a highly symmetric and/or display a simple topology (like the rouse chain), or for very small networks ($N < 5$) the diagonalization of the Kirchhoff matrix cannot be performed analytically. This is the only step that may require a numerical evaluation. For convenience we let $k \in \{0, \ldots, N\}$ with $\mu_0 = 0$ and $Q_{i,0}$ referring to the center of mass motion, while $i \in \{1, \ldots, N+1\}$ such that

$$\Delta\mathbf{r}_i = \sum_{k=0}^{N} Q_{ik}\mathbf{q}_k, \quad \forall\, i\{1, \ldots, N+1\}. \tag{6}$$

In this notation the Itô equation corresponds to the Fokker–Planck equation describing $N$ independent isotropic three-dimensional Ornstein–Uhlenbeck processes [38]. Neglecting the center of mass motion we obtain the following equation for the Green's function (i.e. transition probability density function)

$$\partial_t G(\mathbf{Q}, t|\mathbf{Q}_0) = \sum_{k=1}^{N} \left[\partial^2_{\mathbf{q}_k} + \mu_k \partial_{\mathbf{q}_k} \mathbf{q}_k\right] G(\mathbf{Q}, t|\mathbf{Q}_0), \tag{7}$$

with localized initial condition $G(\mathbf{Q}, t=0|\mathbf{Q}_0) = \delta(\mathbf{Q}-\mathbf{Q}_0)$ and natural boundary conditions $\lim_{|\mathbf{Q}|\to\infty} G(\mathbf{Q}, t|\mathbf{Q}_0) = 0$. We solve equation (7) by means of an eigendecomposition [38] yielding

$$G(\mathbf{Q}, t|\mathbf{Q}_0) = \sum_{\mathbf{N}} \Psi^{\mathrm{R}}_{\mathbf{N}}(\mathbf{Q}) \Psi^{\mathrm{L}}_{\mathbf{N}}(\mathbf{Q_0}) \mathrm{e}^{-\Lambda_{\mathbf{N}} t}, \tag{8}$$

where $\Lambda_{\mathbf{N}}$ denote eigenvalues, $\mathbf{N}$ being a multiset of integer-triples $\{\mathbf{n}_1, \ldots, \mathbf{n}_N\}$ with $\mathbf{n}_i = \{n_{ix}, n_{iy}, n_{iz}\}$ such that

$$\Lambda_{\mathbf{N}} = \sum_{i=1}^{N} (n_{ix} + n_{iy} + n_{iz})\mu_i, \tag{9}$$

and $\Psi^{\mathrm{L}}_{\mathbf{N}}(\mathbf{Q})$ and $\Psi^{\mathrm{R}}_{\mathbf{N}}(\mathbf{Q})$ are the corresponding left and right eigenfunction given by

$$\Psi^{\mathrm{L}}_{\mathbf{N}}(\mathbf{Q}) = \prod_{i=1}^{N} \psi_{\mathbf{n}_i}(\mathbf{q}_i), \qquad \Psi^{\mathrm{R}}_{\mathbf{N}}(\mathbf{Q}) = P_{\mathrm{eq}}(\mathbf{Q}) \prod_{i=1}^{N} \psi_{\mathbf{n}_i}(\mathbf{q}_i), \tag{10}$$

where $P_{\mathrm{eq}}(\mathbf{Q}) \prod_{i=1}^{N} (\mu_i/2\pi)^{3/2} \mathrm{e}^{-\mu_i \mathbf{q}_i^2/2}$ is the equilibrium probability density function of normal coordinates and

$$\psi_{\mathbf{n}_i}(\mathbf{q}_i) = \frac{H_{n_{ix}}(\mu_i q_i^x/2) H_{n_{iy}}(\mu_i q_i^y/2) H_{n_{iz}}(\mu_i q_i^z/2)}{\sqrt{2^{n_{ix}+n_{iy}+n_{iz}} n_{ix}! n_{iy}! n_{iz}!}}, \tag{11}$$

where $q^x, q^y$ and $q^z$ are the components of the vector $\mathbf{q}$, and $H_n(x)$ denotes the $n$th 'physicist's' Hermite polynomial [39]. Using Mehler's formula [40]

$$\sum_{n=0}^{\infty} \frac{(y/2)^n}{n!} H_n(x) H_n(z) = \frac{1}{\sqrt{1-y^2}} \exp\left(-\frac{y^2[x-z]^2}{1-y^2}\right) \tag{12}$$

and recalling that $\mu_0 = 0$ we can also write equation (8) in a closed form [41]

$$G(\mathbf{Q}, t|\mathbf{Q}_0) = \prod_{i=1}^{N} \left(\frac{\mu_i}{2\pi(1-\mathrm{e}^{-2\mu_i t})}\right)^{3/2} \exp\left(-\frac{\mu_i(\mathbf{q}_i - \mathbf{q}_{i0}\,\mathrm{e}^{-\mu_i t})^2}{2(1-\mathrm{e}^{-2\mu_i t})}\right), \tag{13}$$

where the equilibrium probability density function corresponds

$$P_{\mathrm{eq}}(\mathbf{Q}) \equiv \lim_{t\to\infty} G(\mathbf{Q}, t|\mathbf{Q}_0). \tag{14}$$

In what follows we will use both forms of the Green's function, i.e. equations (8) and (13).

## 3. Conformational dynamics

Throughout we are interested in conformational motions encoded in the dynamics of some internal distance $d$, e.g. the distance between two beads $i$ and $j$, $l = |\mathbf{r}_i - \mathbf{r}_j|$ or the distance

between the center of masses of two sets of beads $\Omega_1, \Omega_2$ with $\Omega_1 \cap \Omega_2 = \{0\}$, $l_{\Omega_1,\Omega_2} = |\sum_{i\in\Omega_1}\mathbf{r}_i/\mathrm{card}(\Omega_1) - \sum_{i\in\Omega_2}\mathbf{r}_j/\mathrm{card}(\Omega_2)|$ where $\mathrm{card}(\Omega_i)$ is the cardinality the set $\Omega_i$. Without loss of generality we may thus focus on the distance between two arbitrary beads. Note that in absence of any dynamics in an equilibrium at $T = 0$ such a distance is constant and equal to $d_0$. Expressed in normal coordinates we in turn have

$$\mathbf{l} \equiv \mathbf{r}_i - \mathbf{r}_j = \sum_{k=1}^{N}(Q_{ik} - Q_{jk})\mathbf{q}_k + \mathbf{r}_i^0 - \mathbf{r}_j^0 \equiv \sum_{k=1}^{N} A_k\mathbf{q}_k + \mathbf{d}_0, \tag{15}$$

where in the second equality we have defined $A_k$ and $\mathbf{d}_0$ and omitted the labels $i, j$ to simplify the notation. Note, moreover, that $l \equiv |\mathbf{l}|$ and the generalization to $l_{\Omega_1,\Omega_2}$ follows by linear superposition.

We will focus on four types of observables. The first one is the (non-Markovian) conditional probability density of the time series of the coordinate, $l_t$, defined as

$$\mathcal{G}_{d_0}(l,t|l_0) \equiv \mathbb{P}(l_t \in l\,\mathrm{d}l|l_{t=0} \in l_0\,\mathrm{d}l) = \frac{\langle\delta(l(\mathbf{Q}_t) - l)\delta(l(\mathbf{Q}_0) - l_0)\rangle_{\mathbf{Q}_t}}{\langle\delta(l(\mathbf{Q}_0) - l_0)\rangle_{\mathrm{eq}}}, \tag{16}$$

with $\lim_{t\to\infty}\mathcal{G}_{d_0}(l,t|l_0) \equiv \mathcal{P}^{\mathrm{eq}}_{d_0}(l) = \langle\delta(l(\mathbf{Q}) - l)\rangle_{\mathrm{eq}}$, and where in the second equality we have used the law of conditional probability and introduced the expectation over all Markovian paths of the full system evolving from equilibrium $\langle\cdot\rangle_{\mathbf{Q}_t}$, i.e.

$$\langle\mathcal{B}\rangle_{\mathbf{Q}_t} \equiv \int \mathrm{d}\mathbf{Q}\int \mathrm{d}\mathbf{Q}_0\,\mathcal{B}(\mathbf{Q},\mathbf{Q}_0)G(\mathbf{Q},t|\mathbf{Q}_0)P_{\mathrm{eq}}(\mathbf{Q}_0). \tag{17}$$

and the expectation of any observable $\mathcal{B}(\mathbf{Q})$ over the equilibrium measure $\langle\cdot\rangle_{\mathrm{eq}}$ is $\langle\mathcal{B}\rangle_{\mathrm{eq}} \equiv \int \mathrm{d}\mathbf{Q}\,\mathcal{B}(\mathbf{Q})P_{\mathrm{eq}}(\mathbf{Q})$. The second observable is the normalized equilibrium autocorrelation function

$$\mathcal{C}_{d_0}(t) \equiv \frac{\langle l(t)l(0)\rangle - \langle l(t)\rangle\langle l(0)\rangle}{\langle l^2\rangle_{\mathrm{eq}} - \langle l\rangle^2_{\mathrm{eq}}}, \tag{18}$$

where we have introduced the expectations moments

$$\begin{aligned}
\langle l(t)l(0)\rangle &\equiv \langle l(\mathbf{Q}_t)l(\mathbf{Q}_0)\rangle_{\mathbf{Q}_t} = \int_0^\infty \mathrm{d}l\int_0^\infty \mathrm{d}l_0\, l l_0 \mathcal{G}_{d_0}(l,t|l_0)\mathcal{P}^{\mathrm{eq}}_{d_0}(l_0),\\
\langle l(t)\rangle &\equiv \frac{\langle l(\mathbf{Q}_t)\delta(l(\mathbf{Q}_0) - l_0)\rangle_{\mathbf{Q}_t}}{\langle\delta(l(\mathbf{Q}_0) - l_0)\rangle_{\mathrm{eq}}} = \int_0^\infty \mathrm{d}l\, l\,\mathcal{G}_{d_0}(l,t|l_0),\\
\langle l^n\rangle_{\mathrm{eq}} &\equiv \langle l^n(\mathbf{Q})\rangle_{\mathrm{eq}} = \int_0^\infty \mathrm{d}l l^n\,\mathcal{P}^{\mathrm{eq}}_{d_0}(l).
\end{aligned} \tag{19}$$

The third observable is the $3(N+1)\times 3(N+1)$ position-covariance matrix [42] whose elements are defined as

$$C^{ij}_{\alpha\beta}(t,t_0) = \langle(\mathbf{r}_{i,\alpha}(t+t_0) - \mathbf{r}^0_{i,\alpha})(\mathbf{r}_{j,\beta}(t_0) - \mathbf{r}^0_{j,\beta})\rangle_{\mathbf{Q}_t}, \tag{20}$$

where $\mathbf{r}_{i,\alpha}$ is the $\alpha = \{x, y, z\}$ component of the position vector of bead $i$, $\mathbf{r}_i$.

The fourth, time-average observable is a functional of the projected path $l_\tau$ evolving from $l_{\tau=0}$ called the *fraction of occupation time* or 'empirical density' [43]

$$\theta_{d_0}(l;t) \equiv t^{-1}\int_0^t \delta(l_\tau - l)\mathrm{d}\tau. \tag{21}$$

Note that all observables defined above are assumed to evolve from equilibrium. However, except for $C^{ij}_{\alpha\beta}(t,t_0)$, the initial distribution in fact corresponds to equilibrium constrained to a given value of the tagged distance $l_0$, i.e. from *all* those equilibrium configurations drawn from $P_{\rm eq}(\mathbf{Q})$ that are compatible with $l_0$. This introduces memory in the dynamics of $l_t$ [44].

### 3.1. Projected propagator

The non-Markovian projected propagator $\mathcal{G}_{d_0}(l,t|l_0)$ defined in equation (16) denotes the probability density that the distance between the two tagged beads is equal to $l$ at time $t$ given that it was initially equal to $l_0$. Introducing the auxiliary functions

$$\eta_t \equiv \sum_{k=1}^{N} \frac{A_k^2}{2\mu_k} \mathrm{e}^{-\mu_k t}, \quad \Xi_t(d_0,l,l') \equiv \mathrm{erfi}\left(\frac{d_0(\eta_0-\eta_t)+\eta_t(l+l')}{2\sqrt{\eta_t(\eta_0^2-\eta_t^2)}}\right) \tag{22}$$

we find (for details of the calculation see appendix A)

$$\mathcal{P}^{\rm eq}_{d_0}(l_0)\mathcal{G}_{d_0}(l,t|l_0) = \frac{l l_0 \,\exp\left(-\frac{(l^2+l_0^2)\eta_t+(\eta_0-\eta_t)d_0^2}{4\eta_t(\eta_0-\eta_t)}\right)}{8\sqrt{\pi\eta_t}d_0(\eta_0-\eta_t)}\left[\Xi_t(d_0,-l,-l_0)\right.$$
$$\left. -\,\Xi_t(d_0,-l,l_0)+\Xi_t(d_0,l,l_0)+\Xi_t(-d_0,-l,l_0)\right], \tag{23}$$

where $\mathrm{erfi}(x)$ is the imaginary error function [39] and

$$\mathcal{P}^{\rm eq}_{d_0}(l) = \frac{l}{d_0}\frac{\mathrm{e}^{-(l^2+d_0^2)/4\eta_0}}{\sqrt{\pi\eta_0}}\,\sinh\left(\frac{l d_0}{2\eta_0}\right). \tag{24}$$

We also derive the spectral expansion of $\mathcal{G}_{d_0}(l,t|l_0)$ that reads (see appendix B)

$$\mathcal{G}_{d_0}(l,t|l_0) = V_{\mathbf{00}}(l_0;d_0)^{-1}\sum_{\mathbf{N}} V_{\mathbf{0N}}(l;d_0)V_{\mathbf{N0}}(l_0;d_0)\mathrm{e}^{-\Lambda_N t}, \tag{25}$$

where the overlap elements $V_{\mathbf{0N}}$ and $V_{\mathbf{N0}}$ admit a closed-form expression that is, however, somewhat complicated and thus given in appendix B. Note that 'the ground state' element is simple and corresponds to $V_{\mathbf{00}}(l;d_0) = \mathcal{P}^{\rm eq}_{d_0}(l)$.

### 3.2. Equilibrium distance autocorrelation function

The (normalized) autocorrelation function defined in equation (18) is made explicit by means of the following results

$$\langle l\rangle_{\rm eq} = 2\sqrt{\frac{\eta_0}{\pi}}\,\mathrm{e}^{-d_0^2/4\eta_0} + \left(d_0+\frac{2\eta_0}{d_0}\right)\mathrm{erf}\left(\frac{d_0}{2\sqrt{\eta_0}}\right), \tag{26}$$

$$\langle l^2\rangle_{\rm eq} = d_0^2 + 6\eta_0, \tag{27}$$

where erf is the error function. Equation (27) follows from direct integration of the last line of equation (19) with the aid of equation (24). Conversely, an analytic computation of $\langle l(t)l(0)\rangle$ is possible only using the spectral expansion in equation (25) and yields,

$$\langle l(t)l(0)\rangle = \sum_{\mathbf{N}} \mathcal{V}^{d_0}_{\mathbf{0N}}\mathcal{V}^{d_0}_{\mathbf{N0}}\,\mathrm{e}^{-\Lambda_N t}, \tag{28}$$

$$\mathcal{V}^{d_0}_{\mathbf{0N}} = \int_0^\infty \mathrm{d}l\, l\, V_{\mathbf{0N}}(l; d_0), \qquad \mathcal{V}^{d_0}_{\mathbf{N0}} = \int_0^\infty \mathrm{d}l\, l\, V_{\mathbf{N0}}(l; d_0). \tag{29}$$

The analytic expression of the coefficients $\mathcal{V}^{d_0}_{\mathbf{0N}}$ is lengthy and can be found in appendix D. Plugging equations (29) and (27) into equation (18) delivers an exact analytical result for the equilibrium distance autocorrelation function $\mathcal{C}_{d_0}(t)$. Alternatively one may also evaluate $\mathcal{C}_{d_0}(t)$ by numerical integration of the first line of equation (17) using equation (23), which may in fact be numerically more convenient than implementing the analytical solution.

### 3.3. Position covariance matrix

In the analysis of atomistic MD simulations one often focuses on the position covariance matrix $C^{ij}_{\alpha\beta}(t, t_0)$ [42] and its eigendecomposition. The trajectory derived from an MD stimulation is then projected on the eigenvector (or principal component) corresponding the largest eigenvalue of the covariance matrix with the aim to identify the most important (potentially functionally relevant) motion in a protein [42]. To facilitate a comparison between the aforementioned analysis of MD simulation with GNM we compute $C^{ij}_{\alpha\beta}(t, t_0)$ analytically. Passing as before to normal coordinates we find

$$C^{ij}_{\alpha\beta}(t, t_0) = \Big\langle \sum_{k=1}^{N} Q_{ik} q_{k\alpha}(t + t_0) \sum_{l=1}^{N} Q_{jl} q_{l\beta}(t_0) \Big\rangle, \tag{30}$$

where the matrix elements $Q_{ij}$ do not depend on the spatial coordinate because the GNM is isotropic. Each process $\mathbf{q}_{k,\alpha}$ corresponds to an independent Ornstein–Uhlenbeck process, i.e. the solution of the Itô integral [45] (setting all constant to unity)

$$q_{k\alpha}(t) = \sqrt{2} \int_0^t \mathrm{e}^{-\mu_k (t-s)}\, \mathrm{d}W_{k\alpha}(s). \tag{31}$$

Since by construction (i.e. as a result of isotropy) only the elements of the same spatial coordinate for any given normal mode survive the averaging in equation (30), the elements of the covariance matrix read explicitly

$$C^{ij}_{\alpha\alpha}(t, t_0) = \sum_{k=1}^{N} \frac{Q_{ik} Q_{jk}}{\mu_k}\, \mathrm{e}^{-\mu_k |t - t_0|}. \tag{32}$$

Each entry of the covariance matrix $C^{ij}_{\alpha\alpha}(t, t_0)$ is stationary (i.e. depends only on the time difference, $C^{ij}_{\alpha\alpha}(t, t_0) = C^{ij}_{\alpha\alpha}(|t - t_0|)$).

### 3.4. Fluctuations of occupation time

Single molecule experiments typically probe time-averaged observables. For example, Förster resonance energy transfer [46] and plasmon ruler experiments [47] have been used to extract information about conformational motions of macro-molecules. A fundamental quantity that underlies this kind of observables is the fraction of occupation time, $\theta_{d_0}(l; t)$, defined in equation (21) [43, 48–52]—the random fraction of time a time-series (in our case an internal distance between two beads or between two center of masses) of length $t$ attains a given value of $l$.

In previous publications we have shown how to obtain the mean and the variance of $\theta_{d_0}(l; t)$ [43, 52]. Along these lines we here focus on the mean, $\langle \theta_{d_0}(l; t) \rangle$, and the variance,

$\sigma^2_{\theta;d_0}(l;t) \equiv \langle \theta^2_{d_0}(l;t)\rangle - \langle \theta_{d_0}(l;t)\rangle^2$, of the occupation time fraction at equilibrium that read, respectively (for a derivation see appendix F)

$$\langle \theta_{d_0}(l)\rangle = \mathcal{P}^{\mathrm{eq}}_{d_0}(l), \tag{33}$$

$$\sigma^2_{\theta;d_0}(l,t) = \frac{2}{t}\sum_{\mathbf{N}\neq\mathbf{0}} \frac{V_{\mathbf{0N}}(d;d_0)V_{\mathbf{N0}}(d;d_0)}{\Lambda_{\mathbf{N}}}\left(1 - \frac{1-\mathrm{e}^{-\Lambda_{\mathbf{N}}t}}{\Lambda_{\mathbf{N}}t}\right). \tag{34}$$

Note that $\langle \theta_{d_0}(l;t)\rangle$ corresponds to the equilibrium probability density for all times $t$ since we are considering an ergodic system evolving from equilibrium initial conditions. The variance of the occupation time fraction can equivalently be obtained from (see e.g. [53])

$$\sigma^2_{\theta;d_0}(l,t) = \frac{2}{t}\mathcal{P}^{\mathrm{eq}}_{d_0}(l)\left[\int_0^t (1-\tau/t)\mathcal{G}_{d_0}(l,\tau|l) - \mathcal{P}^{\mathrm{eq}}_{d_0}(l)\right]\mathrm{d}\tau. \tag{35}$$

The integral in equation (35) does not admit an explicit solution. However, it can easily be computed via numerical quadrature. Moreover, it is possible to expand $\mathcal{G}_{d_0}(l,\tau|l)$ for short times (details are given in appendix E) yielding the *small deviation limit*

$$\mathcal{G}_{d_0}(l,t|l) \overset{t\to 0}{=} 2\sqrt{\frac{1}{\pi}}\left(\frac{2}{\sqrt{\kappa t}} + \frac{\sqrt{\kappa t}}{l^2}\right) + \mathcal{O}(t^{3/2}), \tag{36}$$

where we have introduced the shorthand notation $\kappa = \sum_{k=1}^{N} A_k^2$. Plugging equation (36) into equation (35) and performing the integral in turn yields

$$\sigma^2_{\theta;d_0}(d,t) \overset{t\to 0}{\simeq} 2\mathcal{P}^{\mathrm{eq}}_{d_0}(l)\left(\frac{8}{3\sqrt{\kappa\pi t}} + \frac{4}{15l^2}\sqrt{\frac{\kappa t}{\pi}} - \mathcal{P}^{\mathrm{eq}}_{d_0}(l)\right). \tag{37}$$

Since the dynamics of every stable system at equilibrium can be 'linearized' for sufficiently small times $t$ the *small deviation* asymptotic in equations (37) and (36) is in fact a general result for the (large) fluctuations of $\theta_{d_0}(l;t)$ at sufficiently short times.

## 4. Examples

We now apply the result of the previous section to the analysis of a GNM of a protein called adenylate kinase (ADK) and the analysis of toy-model mechanical frames.

### 4.1. GNM of ADK

ADK is an enzyme catalyzing the reversible phosphorylation reaction that transforms adenosine monophosphate to adenosine triphosphate. The structure of ADK has been resolved using x-ray crystallography that uncovered two distinct conformations of the protein that are deposited in the protein data bank (PDB ID: 1AKE [54] and PDB ID: 4AKE [55]) and shown in figure 1.

ADK consists of 214 residues divided in three macro-domains called CORE (residues 1–29, 68–116, and 160–214), LID (residues 118–160), and NMP (residues 30–67). Distinct studies suggest the function to be coupled to open-closed transitions of both, LID and NMP domains with respect to the CORE domain [54, 55]. These transitions have been observed even in absence of nucleotides [56, 57]. However, there is a lively debate in the biophysical community about the precise mechanism and rate-limiting steps in the catalytic function of ADK [58].

**Table 1.** Distance between the center of masses of the three domains for both structures of ADK. All distances are expressed in units of the cutoff distance $r_c = 8$ Å.

| $d_0[r_c]$ | 1AKE | 4AKE |
|---|---|---|
| CORE–LID | 2.6 | 3.8 |
| CORE–NMP | 2.3 | 2.7 |
| LID–NMP | 2.6 | 4.5 |

Here we analyze the autocorrelation functions of distances between the center of mass of LID, NMP, and CORE using the results described in the previous sections. Note that each GNM describes only a single stable structure and therefore cannot capture transitions between the two structures. Nevertheless, the comparison between the two respective GNMs may highlight some differences of the dynamics around the two distinct stable minima.

We obtain the connectivity matrices (shown in figure 1) of the two GNMs using the Prody package [59] with a cutoff distance $r_c = 8$ Å. The static (zero-temperature) distances between the center of masses of the three domains in both structures are given in table 1.

Figure 2 shows the equilibrium probability density function $\mathcal{P}^{\mathrm{eq}}_{d_0}(l)$ (panels (a) and (b)) as well as the autocorrelation function $\mathcal{C}_{d_0}(t)$ (panels (c) and (d)) for all considered distances of the two GNMs representing the two conformational states of ADK. The structure 1AKE is evidently more compact than 4AKE and its corresponding autocorrelation functions consistently decay faster. Moreover, the CORE–NMP distance autocorrelation function decays faster compared to the other two distances whose autocorrelation functions are almost identical (see figure 2(d)). This difference in relaxation is a result of differences in the respective projection, i.e. whereas the eigenvalues of the underlying generator are identical (see equation (29)) the numerical coefficients $\mathcal{V}^{d_0}_{\mathbf{0N}}$ and $\mathcal{V}^{d_0}_{\mathbf{N0}}$ depend strongly on the particular type of projection and thus modify the relaxation rate substantially [60].

The lines in figures 2(c) and (d) have been obtained by means of a numerical integration of the first line of equation (19) using the Gauss–Kronrod quadrature [61]. Unfortunately the evaluation of the integrand is challenging for very short-times because it is a function sharply peaked along the diagonal of the $l, l_0$-plane. This feature prohibits us to obtain reliably (that is, due to numerical imprecision) the autocorrelation function for very short times.

Next we inspect the covariance matrix in equation (32) to identify the dominant, potentially functional important, motions in ADK. In order to reduce the information content while retaining the most essential physics about the extent of local fluctuations and how much the motion of each bead correlated to the motion of other beads we introduce the *covariance-time* defined as

$$\tau_{ij\alpha} = \int_0^\infty C^{ij}_{\alpha\alpha}(t)\mathrm{d}t, \tag{38}$$

which may be interpreted in a manner analogous to the correlation time [62–64], i.e. as a measure of how much the motion between the beads $i$ and $j$ is correlated over time. To measure how much the motion of a single bead is correlated with the rest of the system we consider the total covariance-time $\tau^{\mathrm{tot}}_{i,\alpha} \equiv \sum_{j\neq i}|\tau_{ij\alpha}|$. Conversely, the total variance-time is quantified directly by $\tau_{ii\alpha}$. Note that the model is isotropic and thus independent of $\alpha$. The results are shown in figure 3.

Notably, one can immediately observe that those residues that are involved in the large-scale open-closed motion (i.e. residues with a large $\tau_{ii\alpha}$) also participate in correlated motions denoted by large values $\tau^{\mathrm{tot}}_{i,\alpha}$. For the open structure 4AKE (see figures 3(a) and (b)) the two

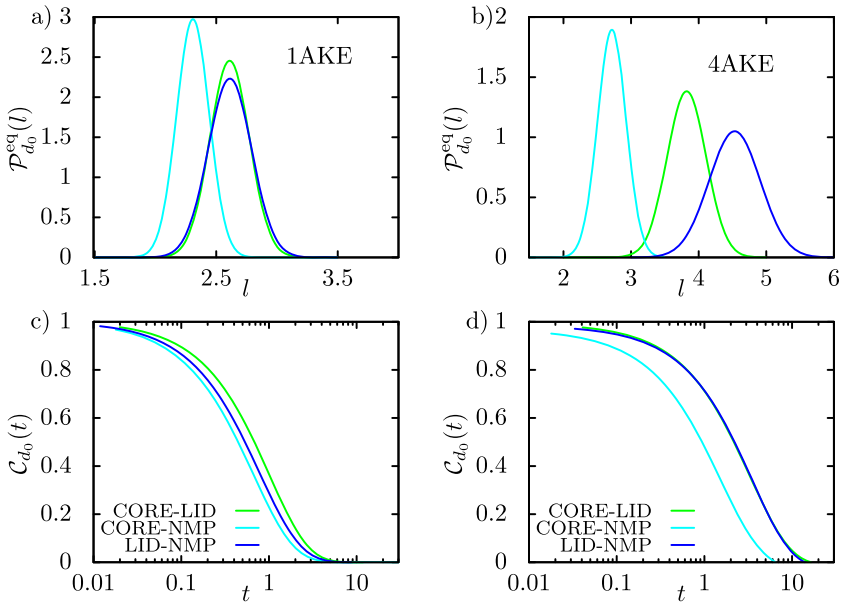


**Figure 2.** Panels (a) and (b) show the equilibrium probability density function for the three center-of-mass distances for both structures (see table 1 for the numerical values of $d_0$). Panels (c) and (d) depict the respective distance autocorrelation functions. Note that the first structure is more compact and its $\mathcal{C}_{d_0}(t)$ decorrelates faster. Moreover, panel (d) reveals that the CORE–NMP distance decorrelates faster than the remaining two distances.

ends of the LID and NMP domains move in a particularly correlated fashion. These residues are in fact those that move toward the core region in the functional open-closed motion of the protein [56, 57]. A remnant of this collective motion can also be seen in the closed structure 1AKE (figures 3(c) and (d)), where the same beads as in 4AKE have a larger $\tau^{\text{tot}}_{i,\alpha}$. This is likely a result of a higher local connectivity.

Note that such results may also be compared to MD simulations [65–68]. A rigorous comparison would in general require exceedingly long simulations [65], which is only feasible for small very small proteins [3, 4]. However, some may equilibrate locally on shorter time-scales (see e.g. [60]) thus potentially lifting the requirement on excessively long simulations. Ignoring potential issues with respect to the computational feasibility we may expect, according to previous results on GNM, no major qualitative differences between GNM and MD simulations.

The analytical results for the covariance matrix are simple, only require the numerical diagonalization of the Kirchhoff matrix and sums involving $N$ terms, and can therefore be rapidly evaluated in a matter of seconds or minutes on a laptop, in contrast to the large computer power required for MD simulations.

Only in the case of huge macromolecules composed of millions of residues the computations presented here could become challenging. However the Kirchhoff matrix of larger proteins is often sparse and/or a coarser representation may be justified [69]. In this case the evaluation of the low-lying $\mu_k$ is still computational efficient using iterative methods, yet would imply a truncation of the sums in equation (32). The same arguments may apply to the computation of the variance and correlation function of occupation time fractions of the intra-bead distance. In fact equation (23) can be easily computed as well (see appendix G for details), and the

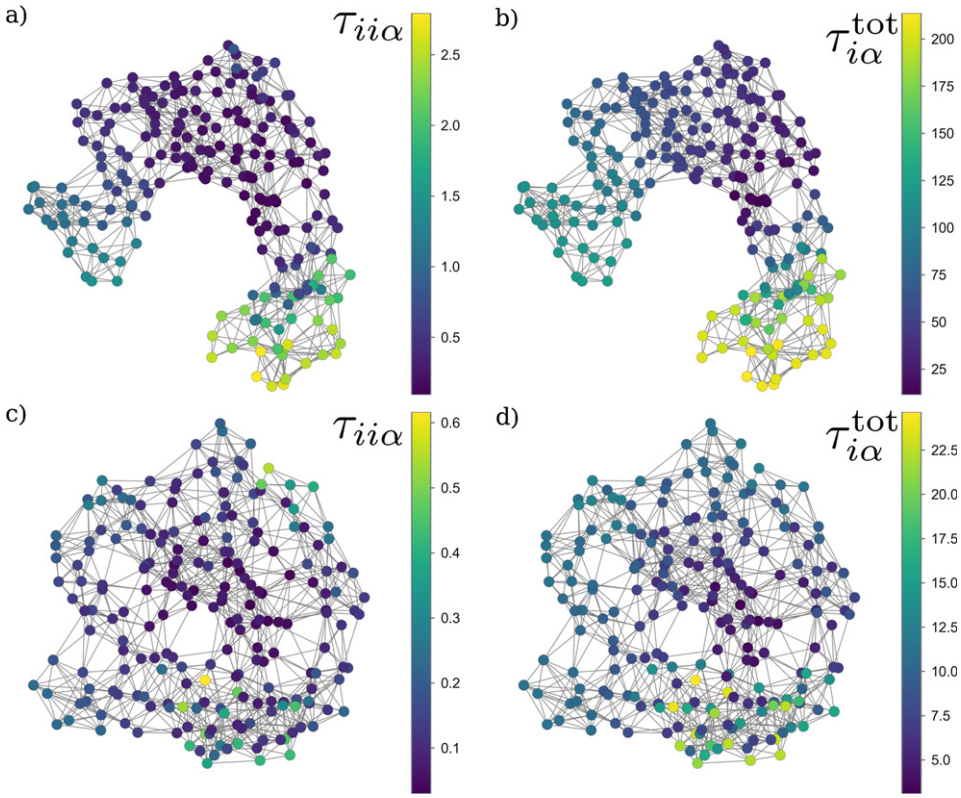


**Figure 3.** (a) and (c) depict $\tau_{ii\alpha}$ and (b) and (d) $\tau^{\mathrm{tot}}_{i,\alpha}$ for each bead in the 4AKE and 1AKE structures, respectively. Notably, the beads in the LID and NMP domains in the 4AKE structure display a particularly large covariance-time.

integrals in equations (35) and (19) can be evaluated numerically efficiently. Conversely, the implementation and computational complexity of the results based on the series expansion are more nuanced and we refer the reader to appendix G.

### *4.2. Simple mechanical frames*

Although GNMs were originally developed to describe proteins they can in fact be used to model any mechanical system in which some underlying network of links imposes constraints on the position of nodes while allowing small, Gaussian fluctuations driven by thermal noise. Examples may include nano-machines such as piezoelectric actuators that move probe-tips in atomic force microscopes [70, 71].

In the generic context of 'mechanical frames' the theory of structural rigidity deals with the question of whether frames are rigid or not [72]. A frame is said to be rigid if one cannot change the distance between pairs of nodes without simultaneously altering the length of at least one connection. A structure that is not rigid is in turn said to allow for inextensional mechanisms. These arise due to a too low number or a particular arrangement of links. In addition, in frames with redundant links there exist states of self-stress. Under given circumstances these states of self-stress impart stiffness to inextensional mechanisms [73].

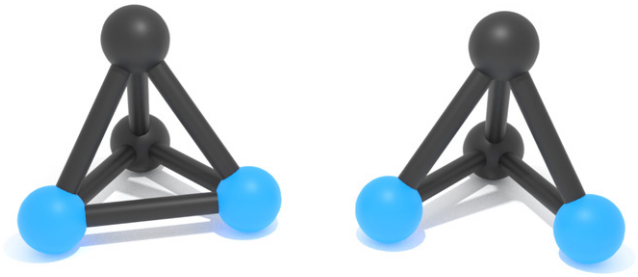

**Figure 4.** A schematic representation of a stable (left) and an unstable (right) frame we consider below.

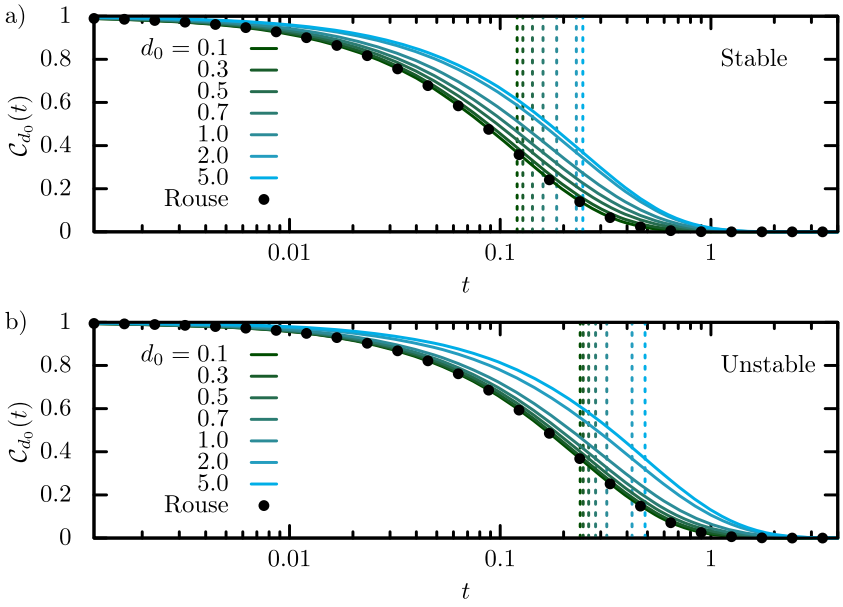


**Figure 5.** Distance autocorrelation function $\mathcal{C}_{d_0}(t)$ for various values of the rest length $d_0$ for the rigid (top panel) and non-rigid (bottom panel) frames depicted in figure 4. The black dots depict $\mathcal{C}_{d_0}(t)$ in the Rouse limit $d_0 = 0$ (see appendix D for details). The vertical dashed lines corresponds to the time $t_c$ at which $\mathcal{C}_{d_0}(t_c) = e^{-1}$. Note that the unstable structure relaxes slower.

As anticipated by Maxwell such a classification of mechanical frames is often non-trivial and may require more information than encoded in the topology of the network [74]. A complete analysis of the mechanisms of a given frame can be obtained by a 'singular value decomposition' of the respective equilibrium matrix $\mathbf{A}$ [75] that relates forces $\mathbf{f}$ on the nodes with tensions $\mathbf{t}$ in the links

$$\mathbf{At} = \mathbf{f}. \tag{39}$$

Singular value decomposition of $\mathbf{A}$ allows (among other things) to determine the rank $r$ of $\mathbf{A}$ and thereby the number of inextensional mechanisms $m$ and states of self-stress $s$ via $s = b - r$ and $m = 3j - 6 - r$, where $j$ is the number of joints and $b$ the number of links in the structure,

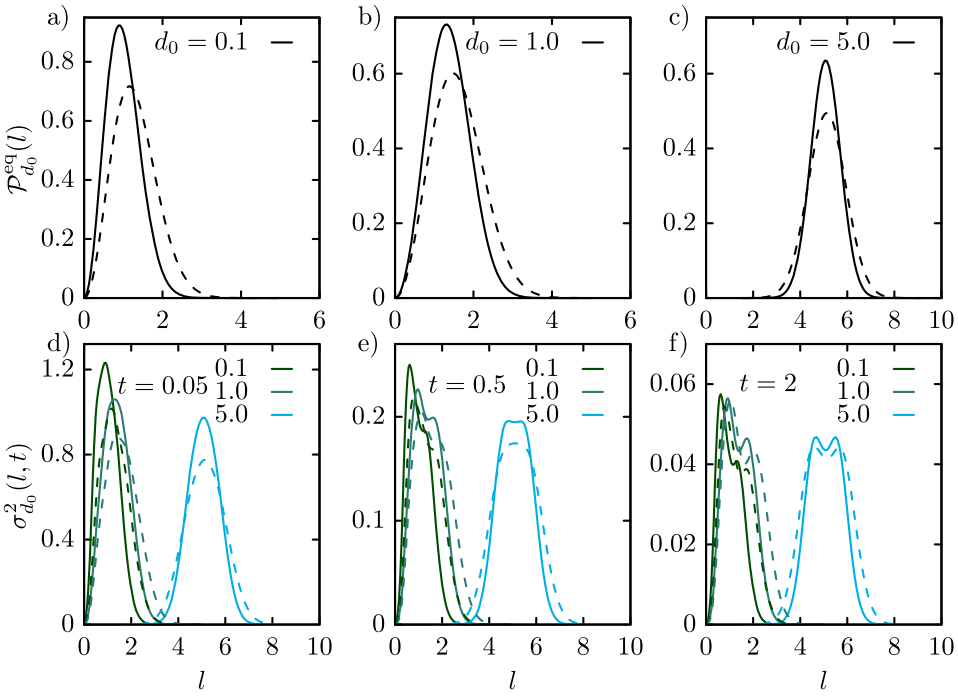


**Figure 6.** Panels (a)–(c) show the equilibrium probability density $\mathcal{P}^{\rm eq}_{d_0}(l)$ for the stable (full lines) and unstable (dashed lines) structure for several values of $d_0$. Panels (e)–(f) depict the variance of the occupation time $\sigma^2_{\theta;d_0}(l,t)$ for the stable (full lines) and unstable (dashed lines) structure, respectively, for different values of $d_0$. The length of the trajectory $t$ increases from (d) to (f).

and note that there are in general six rigid-body motions in three spatial dimensions. Maxwell's well-known formula $b = 3j - 6$ is then extended to:

$$b - 3j + 6 = s - m. \tag{40}$$

To illustrate the concept we consider two toy-model frames depicted in figure 4. Both have $j = 4$ nodes and $s = 0$ states of self-stress. The rigid structure with $b = 6$ links has no inextensional mechanism (i.e. 6–12–6 = 0–0) while the structure with $b = 5$ links has exactly $m = 1$ mechanism (i.e. 5–12–6 = 0–1).

To highlight the role of rigidity and to investigate the effect of a heat-bath we first analyze the autocorrelation function between the blue beads (see figure 4) as a function of the rest-length $d_0$. Notably, in a GNM such distance fluctuations depend on the equilibrium structure $\mathbf{R}^0$ *only* via the equilibrium distance between the tagged beads, $d_0 = |\mathbf{r}^0_i - \mathbf{r}^0_j|$. In turn there is a redundancy—many distinct equilibrium structures $\mathbf{R}^0$ may yield the same result that depends only on the connectivity matrix $\Gamma$ and $d_0$.

The (normalized) distance autocorrelation function $\mathcal{C}_{d_0}(t)$ (see equation (18)) for the two frames is shown figure 5. For $d_0 \lesssim 0.5$ (in dimensionless units) $\mathcal{C}_{d_0}(t)$ depends only very weakly on $d_0$. For larger values of $d_0$ the relaxation time (see dashed vertical lines in figure 5) increases. This observation may be explained by noticing that entropy dominates the motion for small $d_0$.

That is, in the limit of small $d_0$ the rest length may be neglected and the 'Rouse limit' suffices to explain the dynamics essentially quantitatively. Conversely, as $d_0$ increases a certain 'stiffness' emerges in the frame and the (random) oscillations become localized around the equilibrium value $d_0$. Note that the entropic contribution to $\mathcal{C}_{d_0}(t)$ is more important for the non-rigid frame (see right panel in figure 5) as we increase the value of $d_0$ (see figure 5(b)). Conversely, the departure from the Rouse limit toward the 'large stiffness' case is faster in the stable frame (see figure 5(a)). A larger $d_0$ leads to a slower decay of the autocorrelation function $\mathcal{C}_{d_0}(t)$.

Next we consider the fraction of occupation time $\theta_{d_0}(l;t)$ [43]. We assume that the initial condition evolves from equilibrium and therefore $\langle\theta_{d_0}(l;t)\rangle = \mathcal{P}^{\mathrm{eq}}_{d_0}(l)$ whereas $\sigma^2_{\theta;d_0}(l,t)$ depends on time (see equation (34) as well as [43, 52]). The aforementioned dominance of the entropic (heat bath) contribution at small values of $d_0$ is also noticeable the fluctuations of $\theta_{d_0}(l;t)$ as depicted in figure 6. Notably, as $d_0$ increases the support of $\sigma^2_{\theta;d_0}(l,t)$ progressively shifts toward larger $l$ and concentrates near $d_0$.

Notably, the variance of the occupation time fraction $\sigma^2_{\theta;d_0}(l,t)$ changes shape from unimodal, at short times $t$, $d$ to bimodal at long $t$. Such a behavior is characteristic for stochastic process in spatial confinement [43], i.e. fluctuations of $\theta_{d_0}(l;t)$ are larger in the vicinity of confining boundaries (even if these boundaries are 'soft').

Moreover as $d_0$ increases the shape of both, $\mathcal{P}^{\mathrm{eq}}_{d_0}(l)$ as well as $\sigma^2_{\theta;d_0}(l,t)$ becomes more symmetric. The reason seems to be that the effect of the confining boundary at $l = 0$ becomes irrelevant as the support of $\sigma^2_{\theta;d_0}(l,t)$ begins to concentrate near a substantial $d_0$. In other words although the projection of the dynamics of a link in three-dimensional space onto a (one-dimensional) distance destroys the Gaussian behavior, the latter becomes (partially) restored at large values of $d_0$.

## 5. Conclusions

We presented analytical results (up to a numerical diagonalization of a symmetric matrix) for a selection of relevant time- and ensemble-averaged physical observables in the GNM. One may think of GNM as certain generalization of the Rouse model to networks with links with a certain degree of extensional and rotational stiffness. We determined a set of coarse-grained observables—internal distances—that may be of interest in the analysis of GNM in the context of internal motions in proteins or mechanical frames in contact with a heat bath. We hope that our results will enable and motivate a more systematic analysis of GNM derived from proteins [59]. To this end a C++ computer code is provided in the supplementary material that implements all result (for more details about the implementation see appendix G).

## Acknowledgments

The authors thank David Hartich and Lars Bock for the useful discussions. The financial support from the German Research Foundation (DFG) through the Emmy Noether Program GO 2762/1-1 to AG is gratefully acknowledged.

## Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

## Appendix A. Derivation of the equilibrium probability density

The equilibrium probability density function of any link-vector $\mathbf{l}$ is defined by

$$\mathcal{P}^{\mathrm{eq}}_{\mathbf{d}_0}(\mathbf{l}) = V_{\mathbf{00}}(\mathbf{l};\mathbf{d}_0) \equiv \int \mathrm{d}\mathbf{Q}\Psi^{\mathrm{R}}_{\mathbf{0}}(\mathbf{Q})\delta\left(\sum_{k=1}^{N} A_k\mathbf{q}_k + \mathbf{d}_0 - \mathbf{l}\right)\Psi^{\mathrm{L}}_{\mathbf{0}}(\mathbf{Q}). \tag{A.1}$$

Applying the Fourier transform $\tilde{f}(s) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\mathrm{d}x\, f(x)\mathrm{e}^{-\mathrm{i}sx}$ component-wise to equation (A.1) (i.e. $\mathbf{l}\to\mathbf{s}$) yields

$$\frac{1}{(2\pi)^3}\int \mathrm{d}\mathbf{Q}\prod_{k=1}^{N}\left(\frac{\mu_k}{2\pi}\right)^{3/2}\exp\left(-\sum_{k=1}^{N}\frac{\mu_k}{2}\mathbf{q}_k^2 + i(A_k\mathbf{q}_k+\mathbf{d}_0)\mathbf{s}\right) = \tag{A.2}$$

$$\frac{1}{(2\pi)^3}\,\mathrm{e}^{-s^2\sum_{k=1}^{N}A_k^2/2\mu_k + i\mathbf{d}_0\mathbf{s}}. \tag{A.3}$$

Inverting the Fourier transform we obtain, defining $\eta_0 = \sum_{k=1}^{N}A_k^2/(2\mu_k)$,

$$\mathcal{P}^{\mathrm{eq}}_{\mathbf{d}_0}(\mathbf{l}) = V_{\mathbf{00}}(\mathbf{l};\mathbf{d}_0) = \frac{1}{(2\pi)^3}\left(\frac{\pi}{\eta_0}\right)^{3/2}\mathrm{e}^{-(\mathbf{l}-\mathbf{d}_0)^2/4\eta_0}. \tag{A.4}$$

Since we are only interested in the distance and not the direction we need to marginalize over angles, i.e.

$$\int_0^{\infty}\mathrm{d}x\, x^2\int_0^{2\pi}\mathrm{d}\phi\int_{-1}^{1}\mathrm{d}(\cos\,\theta)V_{\mathbf{00}}(\mathbf{x})\delta(|\mathbf{x}|-x), \tag{A.5}$$

where $\phi$ is the polar angle, $\theta$ is the azimuthal angle and without loss of generality we choose a frame of reference such that the vector $\mathbf{d}_0$ is parallel to the $z$-axis. The solution of this integral finally gives equation (24):

$$\mathcal{P}^{\mathrm{eq}}_{d_0}(l) = V_{\mathbf{00}}(l;d_0) = \frac{1}{\sqrt{\pi\eta_0}}\frac{l}{d_0}\,\mathrm{e}^{-(l^2+d_0^2)/4\eta_0}\,\sinh\left(\frac{ld_0}{2\eta_0}\right). \tag{A.6}$$

## Appendix B. Spectral solution for $\mathcal{G}_{d_0}$

In the spectral solution for the Green's function in equation (25) we have defined the elements $V_{\mathbf{0N}}(l;d_0)$, $V_{\mathbf{N0}}(l;d_0)$ which are derived as follows. Let

$$V_{\mathbf{0N}}(\mathbf{l};\mathbf{d}_0) = \int \mathrm{d}\mathbf{Q}\;\Psi^{\mathrm{R}}_{\mathbf{N}}(\mathbf{Q})\delta\left(\sum_{k=1}^{N}A_k\mathbf{q}_k + \mathbf{d}_0 - \mathbf{l}\right)\Psi^{\mathrm{L}}_{\mathbf{0}}(\mathbf{Q}), \tag{B.1}$$

$$V_{\mathbf{N0}}(\mathbf{l};\mathbf{d}_0) = \int \mathrm{d}\mathbf{Q}\;\Psi^{\mathrm{R}}_{\mathbf{0}}(\mathbf{Q})\delta\left(\sum_{k=1}^{N}A_k\mathbf{q}_k + \mathbf{d}_0 - \mathbf{l}\right)\Psi^{\mathrm{L}}_{\mathbf{N}}(\mathbf{Q}). \tag{B.2}$$

Fortunately, the above elements $V_{\mathbf{0N}}$ and $V_{\mathbf{N0}}$ are identical (cf equation (10)). Therefore what we need to solve for is

$$V_{\mathbf{N0}}(\mathbf{l};\mathbf{d_0}) = \prod_{k=1}^{N}\int \mathrm{d}\mathbf{q}_k \left(\frac{\mu_k}{2\pi}\right)^{\frac{3}{2}} \sqrt{\frac{1}{2^{n_{kx}+n_{ky}+n_{kz}} n_{kx}! n_{ky}! n_{kz}!}} \times H_{n_{kx}}\left(\sqrt{\frac{\mu_k}{2}} q_k^x\right) H_{n_{ky}}\left(\sqrt{\frac{\mu_k}{2}} q_k^y\right) H_{n_{kz}}\left(\sqrt{\frac{\mu_k}{2}} q_k^z\right) \times \mathrm{e}^{-\mu_k \mathbf{q}_k^2/2} \delta\left(\sum_{k=1}^{N} A_k \mathbf{q}_k + \mathbf{d}_0 - \mathbf{l}\right). \tag{B.3}$$

It is convenient to define the auxiliary variables $\{\mathbf{q}'_k\} \equiv \{q_k^x - d_0^x, q_k^y - d_0^y, q_k^z - d_0^z\}$, and then perform the Fourier transform $\mathbf{l} \to \mathbf{s}$ to obtain

$$\frac{1}{(2\pi)^3}\prod_{k=1}^{N}\int \mathrm{d}\mathbf{q}'_k \left(\frac{\mu_k}{2\pi}\right)^{\frac{3}{2}} \sqrt{\frac{1}{2^{n_{kx}+n_{ky}+n_{kz}} n_{kx}! n_{ky}! n_{kz}!}} \times H_{n_{kx}}\left(\sqrt{\frac{\mu_k}{2}} q_k'^x\right) H_{n_{ky}}\left(\sqrt{\frac{\mu_k}{2}} q_k'^x\right) H_{n_{kz}}\left(\sqrt{\frac{\mu_k}{2}} q_k'^z\right) \times \mathrm{e}^{-\mu_k \mathbf{q}_k'^2/2 - iA_k \mathbf{s}\cdot\mathbf{q}'_k}. \tag{B.4}$$

Factorizing in the three spatial dimensions, completing the square in the exponential, and changing the variable to $t_k^h = \sqrt{\mu_k} q_k'^h/\sqrt{2}$ (where the subscript $h$ denotes the respective spatial coordinate) we find

$$\frac{1}{(2\pi)^3}\prod_{k=1}^{N}\left(\frac{1}{\pi}\right)^{\frac{3}{2}}\prod_{h=1}^{3}\sqrt{\frac{1}{2^{n_{kh}} n_{kh}!}}\, \mathrm{e}^{-s_h^2 (A_k)^2/2\mu_k} \times \int_{-\infty}^{\infty} \mathrm{d}t_k^h\, H_{n_{kh}}(t_k^h) \exp\left(-\left[t_k^h - \left(-\frac{iA_k}{\sqrt{2\mu_k}} s_h\right)\right]^2\right) \tag{B.5}$$

whose solution is [76]

$$\frac{1}{(2\pi)^3}\prod_{k=1}^{N}\prod_{h=1}^{3}\sqrt{\frac{2^{n_{kh}}}{n_{kh}!}}\left(-\frac{iA_k}{\sqrt{2\mu_k}} s_h\right)^{n_{kh}} \mathrm{e}^{-s_h^2 (A_k)^2/2\mu_k}. \tag{B.6}$$

It turns out to be convenient to write equation (B.6) as

$$\frac{1}{(2\pi)^3}\left[\prod_{k=1}^{N}\prod_{h=1}^{3}\sqrt{\frac{2^{n_{kh}}}{n_{kh}!}}\left(-\frac{iA_k}{\sqrt{2\mu_k}}\right)^{n_{kh}}\right]\prod_{h=1}^{3} s_h^{\sum_{k=1}^{N} n_{kh}}\, \mathrm{e}^{-s_h^2 \sum_{k=1}^{N} (A_k)^2/2\mu_k}, \tag{B.7}$$

and to define[1]

$$M \equiv \prod_{k=1}^{N}\prod_{h=1}^{3}\sqrt{\frac{2^{n_{kh}}}{n_{kh}!}}\left(-\frac{iA_k}{2\mu_k}\right)^{n_{kh}}$$
$$= \sqrt{\frac{2^{\sum_{k=1}^{N}\sum_{h=1}^{3}n_{kh}}}{\prod_{k=1}^{N}\prod_{h=1}^{3}n_{kh}!}}(-i)^{\sum_{k=1}^{N}\sum_{h=1}^{3}n_{kh}}\prod_{k=1}^{N}\left(\frac{A_k}{\sqrt{2\mu_k}}\right)^{n_{kx}+n_{ky}+n_{kz}}. \tag{B.8}$$

We now invert the Fourier transform

$$\frac{M}{(2\pi)^3}\prod_{h=1}^{3}\int_{-\infty}^{\infty}\mathrm{d}s_h\, s_h^{\sum_{k=1}^{N}n_{kh}}\,\mathrm{e}^{-\eta_0 s_h^2+is_h d_h}. \tag{B.9}$$

Completing the square in the exponential and defining $t_h = \sqrt{\eta_0}s_h$, we can write:

$$\frac{M}{(2\pi)^3}\prod_{h=1}^{3}\frac{\mathrm{e}^{-d_h^2/4\eta_0}}{\sqrt{\eta_0}^{\sum_{k=1}^{N}n_{kh}+1}}\int_{-\infty}^{\infty}\mathrm{d}t_h\, t_h^{\sum_{k=1}^{N}n_{kh}}\,\mathrm{e}^{-(t_h-id_h/2\sqrt{\eta_0})^2}, \tag{B.10}$$

the integral in the previous equation can be solved analytically [76]

$$\frac{M}{(2\pi)^3}\prod_{h=1}^{3}\frac{\mathrm{e}^{-d_h^2/4\eta_0}}{\sqrt{\eta_0}^{\sum_{k=1}^{N}n_{kh}+1}}\sqrt{\pi}(2i)^{-\sum_{k=1}^{N}n_{kh}}(-1)^{\sum_{k=1}^{N}n_{kh}}H_{\sum_{k=1}^{N}n_{kh}}\left(\frac{l_h}{2\sqrt{\eta_0}}\right). \tag{B.11}$$

Using the definition of $M$ in equation (B.8), defining $N_h = \sum_{k=1}^{N}n_{kh}$ and $\mathcal{N} = N_x + N_y + N_z$, and going back to the original, non-shifted coordinates we arrive at the following form of equation (B.3)

$$V_{\mathbf{0N}}(\mathbf{l};\mathbf{d}_0) = \frac{1}{(2\sqrt{\pi})^3}\sqrt{\frac{1}{2^{\mathcal{N}}\prod_{k=1}^{N}n_{kx}!n_{ky}!n_{kz}!}}\prod_{k=1}^{N}\left(\frac{A_k}{\sqrt{2\mu_k}}\right)^{n_{kx}+n_{ky}+n_{kz}}$$
$$\times\frac{1}{\sqrt{\eta_0}^{\mathcal{N}+3}}H_{N_x}\left(\frac{l^x-d_0^x}{2\sqrt{\eta_0}}\right)H_{N_y}$$
$$\times\left(\frac{l^y-d_0^y}{2\sqrt{\eta_0}}\right)H_{N_z}\left(\frac{l^z-d_0^z}{2\sqrt{\eta_0}}\right)\mathrm{e}^{-(\mathbf{l}-\mathbf{d}_0)^2/4\eta_0}. \tag{B.12}$$

To integrate over the angular part it is convenient to use the following expansion of the Hermite polynomials [39]:

$$H_n(x) = n!\sum_{m=0}^{\lfloor\frac{n}{2}\rfloor}\frac{(-1)^m}{m!(n-2m)!}(2x)^{n-2m}. \tag{B.13}$$

[1] We use the convention $0^0 = 1$ for terms $\left(\frac{A_k}{\sqrt{2\mu_k}}\right)^{n_{kx}+n_{ky}+n_{kz}}$.

If we rotate our frame of reference such that $\hat{z}\|d_0^z$, (i.e.: $d_0^x = 0$, $d_0^y = 0$ and $d_0^z = d_0$) we find

$$\begin{aligned} V_{\mathbf{0N}}(\mathbf{l};\mathbf{d}_0) = & \frac{1}{(2\sqrt{\pi})^3}\sqrt{\frac{1}{2^{\mathcal{N}}\prod_{k=1}^{N} n_{kx}!n_{ky}!n_{kz}!}}\prod_{k=1}^{N}\left(\frac{A_k}{\sqrt{2\mu_k}}\right)^{n_{kx}+n_{ky}+n_{kz}} \\ & \times \frac{1}{\sqrt{\eta_0}^{\mathcal{N}+3}} N_x!N_y!N_z!\mathrm{e}^{-\frac{d^2+d_0^2}{4\eta_0}} \\ & \times \sum_{a=0}^{\lfloor N_x/2\rfloor}\sum_{b=0}^{\lfloor N_y/2\rfloor}\sum_{c=0}^{\lfloor N_z/2\rfloor}\frac{(-1)^{a+b+c}}{a!b!c!(N_x-2a)!(N_y-2b)!(N_z-2c)!} \\ & \times \left(\frac{l^x}{\sqrt{\eta_0}}\right)^{N_x-2a}\left(\frac{l^y}{\sqrt{\eta_0}}\right)^{N_y-2b}\left(\frac{l^z-d_0}{\sqrt{\eta_0}}\right)^{N_z-2c}\mathrm{e}^{\mathbf{l}\cdot\mathbf{d}_0/2\eta_0}. \end{aligned} \tag{B.14}$$

Using the binomial expansion of $(l^z - d_0)/\sqrt{\eta_0})^{N_z-2c}$ we obtain

$$\begin{aligned} V_{\mathbf{0N}}(\mathbf{l};\mathbf{d}_0) = & \frac{1}{(2\sqrt{\pi})^3}\sqrt{\frac{1}{2^{\mathcal{N}}\prod_{k=1}^{N} n_{kx}!n_{ky}!n_{kz}!}}\prod_{k=1}^{N}\left(\frac{A_k}{\sqrt{2\mu_k}}\right)^{n_{kx}+n_{ky}+n_{kz}} \\ & \times \frac{1}{\sqrt{\eta_0}^{\mathcal{N}+3}} N_x!N_y!N_z!\mathrm{e}^{-(l^2+d_0^2)/4\eta_0} \\ & \times \sum_{a=0}^{\lfloor N_x/2\rfloor}\sum_{b=0}^{\lfloor N_y/2\rfloor}\sum_{c=0}^{\lfloor N_z/2\rfloor}\frac{(-1)^{a+b+c}}{a!b!c!(N_x-2a)!(N_y-2b)!} \\ & \times \left(\frac{l^x}{\sqrt{\eta_0}}\right)^{N_x-2a}\left(\frac{l^y}{\sqrt{\eta_0}}\right)^{N_y-2b}\mathrm{e}^{\mathbf{l}\cdot\mathbf{d}_0/2\eta_0} \\ & \times \sum_{m=0}^{N_z-2c}\frac{1}{m!(N_z-2c-m)!}\left(\frac{l^z}{\sqrt{\eta_0}}\right)^{N_z-2c-m}\left(-\frac{d_0}{\sqrt{\eta_0}}\right)^{m}. \end{aligned} \tag{B.15}$$

At this point we can integrate over the angles of $\mathbf{l}$ fixing the length, hence

$$\begin{aligned} V_{\mathbf{0N}}(l,d_0) = & \frac{1}{(2\sqrt{\pi})^3}\sqrt{\frac{1}{2^{\mathcal{N}}\prod_{k=1}^{N} n_{kx}!n_{ky}!n_{kz}!}}\prod_{k=1}^{N}\left(\frac{A_k}{2_k}\right)^{n_{kx}+n_{ky}+n_{kz}} \\ & \times \frac{1}{\eta_0^{\mathcal{N}+1}} N_x!N_y!N_z!\mathrm{e}^{-(l^2+d_0^2)/4\eta_0^2} \\ & \times \sum_{a=0}^{\lfloor N_x/2\rfloor}\sum_{b=0}^{\lfloor N_y/2\rfloor}\sum_{c=0}^{\lfloor N_z/2\rfloor}\frac{(-1)^{a+b+c}}{a!b!c!(N_x-2a)!(N_y-2b)!} \\ & \times \left(\frac{l}{\eta_0}\right)^{N_x+N_y-2(a+b)+2} \\ & \times \int_0^{2\pi}\mathrm{d}\phi(\sin\,\phi)^{N_y-2b}(\cos\,\phi)^{N_x-2a} \end{aligned}$$

$$\times \sum_{m=0}^{N_z-2c} \frac{1}{m!(N_z-2c-m)!}\left(\frac{l}{\eta_0}\right)^{N_z-2c-m}\left(-\frac{d_0}{\eta_0}\right)^m$$
$$\times \int_0^\pi \mathrm{d}\theta(\sin\,\theta)^{N_x+N_y-2(a+b)+1}(\cos\,\theta)^{N_z-2c-m}\,\mathrm{e}^{\cos\,\theta l d_0/2\eta_0^2}. \quad \text{(B.16)}$$

The first integral is

$$\int_0^{2\pi} \mathrm{d}\phi \cos^n\phi \sin^m\phi = \frac{\pi n!m!}{2^{n+m-1}\left(\frac{n}{2}\right)!\left(\frac{m}{2}\right)!\left(\frac{n+m}{2}\right)!}, \quad \text{(B.17)}$$

and is non-zero only if $n$ and $m$ are even [77]. Therefore $N_x$ and $N_x$ must be even. While the second integral reads [77]

$$\begin{aligned}\int_0^\pi \mathrm{d}\theta(\sin\,\theta)^n(\cos\,\theta)^m\mathrm{e}^{k\,\cos\,\theta} = \frac{\sqrt{\pi}}{4}\gamma\left(\frac{1+n}{2}\right)\\ \times\left[2(1+(-1)^m)\gamma\left(\frac{1+m}{2}\right)\right.\\ \times {}_1\tilde{F}_2\left(\frac{1+m}{2};\frac{1}{2},\frac{2+m+n}{2};\frac{k^2}{4}\right)\\ -(-1+(-1)^m)k\gamma\left(1+\frac{m}{2}\right)\\ \left.\times {}_1\tilde{F}_2\left(\frac{2+m}{2};\frac{3}{2},\frac{3+m+n}{2};\frac{k^2}{4}\right)\right], \quad \text{(B.18)}\end{aligned}$$

where we have introduced the Euler's gamma function $\gamma(x)$ as well as the regularized hypergeometric function ${}_p\tilde{F}_q(a_1,\ldots,a_p;b_1,\ldots,b_q;x)$ [39]. Putting all together we finally arrive at

$$\begin{aligned}V_{\mathbf{0N}}(l;d_0) = \frac{1}{16}\sqrt{\frac{1}{2^{\mathcal{N}}\prod_{k=1}^N n_{kx}!n_{ky}!n_{kz}!}}\prod_{k=1}^N\left(\frac{A_k}{\sqrt{2\mu_k}}\right)^{n_{kx}+n_{ky}+n_{kz}}\\ \times\frac{1}{\sqrt{\eta_0}^{\mathcal{N}+1}}N_x!N_y!N_z!\mathrm{e}^{-(d^2+d_0^2)/4\eta_0}\\ \times\sum_{a=0}^{N_x/2}\sum_{b=0}^{N_y/2}\sum_{c=0}^{\lfloor N_z/2\rfloor}\frac{(-1)^{a+b+c}}{a!b!c!\left(\frac{N_x-2a}{2}\right)!\left(\frac{N_y-2b}{2}\right)!2^{N_x+N_y-2(a+b)}}\\ \times\sum_{m=0}^{N_z-2c}\frac{1}{m!(N_z-2c-m)!}\\ \times\left(\frac{l}{\sqrt{\eta_0}}\right)^{\mathcal{N}-2(a+b+c)-m+2}\left(-\frac{d_0}{\sqrt{\eta_0}}\right)^m\\ \times\left[2(1+(-1)^{N_z-2c-m})\gamma\left(\frac{1+N_z-2c-m}{2}\right)\right.\\ \left.\times {}_1\tilde{F}_2\left(\frac{1+N_z-2c-m}{2};\frac{1}{2},\frac{3+\mathcal{N}-2(a+b+c)-m}{2};\frac{l^2d_0^2}{16\eta_0^2}\right)\right.\end{aligned}$$

$$- (-1 + (-1)^{N_z-2c-m})\frac{ld_0}{2\eta_0}\gamma\left(1 + \frac{N_z - 2c - m}{2}\right)$$
$$\times {}_1\tilde{F}_2\left(\frac{2 + N_z - 2c - m}{2}; \frac{3}{2}, \frac{4 + \mathcal{N} - 2(a + b + c) - m}{2}; \frac{l^2 d_0^2}{16\eta_0^2}\right)\Bigg]; \quad \text{(B.19)}$$

which finally allows us to write down the non-Markovian Green's function expressed as an infinite series in equation (25). In addition, the series expansion allows us the compute the cross conditioned Green's function

$$\mathcal{G}_{d_0,d_0'}(l,t|l') = V_{\mathbf{00}}(l';d_0')^{-1}\sum_{\mathbf{N}} V_{\mathbf{0N}}(l;d_0)V_{\mathbf{N0}}(l';d_0')\mathrm{e}^{-\Lambda_N t} \quad \text{(B.20)}$$

that is the probability that the distance between the beads $i$ and $j$ is equal to $l$ at time $t$ conditioned to the fact that the distance between the beads $k$ and $l$ at time 0 was equal to $l'$, assuming that these two distances have rest lengths $d_0$ and $d_0'$, respectively. In particular in order to evaluate $V_{\mathbf{N0}}(l';d_0')$ we need to consider that the distance $\mathbf{l}'$ is expressed via the normal coordinates as

$$\mathbf{d}' = \mathbf{r}_k - \mathbf{r}_l = \sum_{i=1}^{N} B_i \mathbf{q}_i, \quad \text{(B.21)}$$

and we in turn use these coefficients to define $\zeta_t = \sum_{k=1}^{N} B_k^2/2\mu_k \mathrm{e}^{-\mu_k t}$ and $\zeta_0 = \sum_{k=1}^{N} B_k^2/2\mu_k$ instead of $\eta_t$ and $\eta_0$.

## Appendix C. Closed form solution for $\mathcal{G}_{d_0}$

In order to obtain the equivalent result in a closed form solution we should consider the following integral:

$$\mathcal{J}_{\mathbf{d}_0}(\mathbf{l},t;\mathbf{l}_1) = \int \mathrm{d}\mathbf{Q}\int \mathrm{d}\mathbf{Q}_1\, G(\mathbf{Q},t|\mathbf{Q}_1)P_{\mathrm{eq}}(\mathbf{Q}_1)$$
$$\times \delta\left(\sum_{k=1}^{N} A_k\mathbf{q}_{1k} + \mathbf{d}_0 - \mathbf{l}_1\right)\delta\left(\sum_{k=1}^{N} A_k\mathbf{q}_k + \mathbf{d}_0 - \mathbf{l}\right). \quad \text{(C.1)}$$

Performing the first Fourier transform, between $\mathbf{l}_1 \to \mathbf{u}$ the above integral becomes

$$\int \mathrm{d}\mathbf{Q}\delta\left(\sum_{k=1}^{N} A_k\mathbf{q}_k + \mathbf{d}_0 - \mathbf{l}\right)\mathrm{e}^{-\mathrm{i}\mathbf{d_0}\cdot\mathbf{u}}$$
$$\times \left(\frac{1}{2\pi}\right)^3 \int \mathrm{d}\mathbf{q}_{1k}\prod_{k=1}^{N}\left(\frac{\mu_k^2}{2\pi}\right)^{3/2}\left(\frac{\mu_k}{2\pi(1-\mathrm{e}^{-2\mu_k t})}\right)^{3/2}$$
$$\times \exp\left[-\frac{\mu_k}{2(1-\mathrm{e}^{-2\mu_k^2 t})}\left(\mathbf{q}_k^2 + \mathbf{q}_{1k}^2\mathrm{e}^{-2\mu_k t} - 2\mathbf{q}_k\cdot\mathbf{q}_{1k}\,\mathrm{e}^{-\mu_k t}\right)\right]$$
$$\times \mathrm{e}^{-\mathrm{i}\,A_k\mathbf{q}_{1k}\cdot\mathbf{u}}\,\mathrm{e}^{-\mu_k\mathbf{q}_{1k}^2/2}; \quad \text{(C.2)}$$

and the integration yields

$$\left(\frac{1}{2\pi}\right)^3 \int \mathrm{d}\mathbf{Q}\delta\left(\sum_{k=1}^{N} A_k\mathbf{q} + \mathbf{d}_0 - \mathbf{l}\right) \mathrm{e}^{-i\mathbf{d_0}\cdot\mathbf{u}} \prod_{k=1}^{N}\left(\frac{\mu_k^2}{2\pi}\right)^{3/2} \exp\left[-\frac{\mu_k}{2}\left(\mathbf{q}_k^2 + 2i\frac{A_k}{\mu_k}\,\mathrm{e}^{-\mu_k t}\mathbf{u}\cdot\mathbf{q}_k + \frac{A_k^2}{\mu_k^2}(1-\mathrm{e}^{-2\mu_k t})\mathbf{u}^2\right)\right]. \tag{C.3}$$

Performing the second Fourier transform $\mathbf{l}\to\mathbf{v}$ we find

$$\mathrm{e}^{-i\mathbf{d}_0\cdot(\mathbf{u}+\mathbf{v})}\left(\frac{1}{2\pi}\right)^6 \prod_{k=1}^{N}\int \mathrm{d}\mathbf{q}_k\left(\frac{\mu_k}{2\pi}\right)^{3/2} \exp\left[-\frac{\mu_k}{2}\left(\mathbf{q}_k^2 + 2i\frac{A_k}{\mu_k}(\mathrm{e}^{-\mu_k t}\,\mathbf{u}+\mathbf{v})\cdot\mathbf{q}_k + \frac{A_k^2}{\mu_k^2}(1-\mathrm{e}^{-2\mu_k t})\mathbf{u}^2\right)\right], \tag{C.4}$$

that reads

$$\mathrm{e}^{-i\mathbf{d}_0\cdot(\mathbf{u}+\mathbf{v})}\left(\frac{1}{2\pi}\right)^6 \prod_{k=1}^{N}\exp\left[-\frac{A_k^2}{2\mu_k}(\mathbf{u}^2+\mathbf{v}^2) + 2\frac{A_k^2}{2\mu_k}\,\mathrm{e}^{-\mu_k t}\,\mathbf{u}\cdot\mathbf{v}\right]. \tag{C.5}$$

It is convenient to define

$$\sum_{k=1}^{N}\frac{A_k^2}{2\mu_k}\,\mathrm{e}^{-\mu_k t} = \eta_t \to \sum_{k=1}^{N}\frac{A_k^2}{2\mu_k} = \eta_0 \tag{C.6}$$

so the Fourier transform of the joint-density is:

$$\tilde{\mathcal{J}}_{\mathbf{d}_0}(\mathbf{v},t;\mathbf{u}) = \frac{1}{(2\pi)^6}\,\exp\left(-\eta_0\mathbf{u}^2 - \eta_0\mathbf{v}^2 + 2\eta_t\mathbf{u}\cdot\mathbf{v} + i\mathbf{d}_0\cdot(\mathbf{u}+\mathbf{v})\right). \tag{C.7}$$

The inversion of the two Fourier transforms gives straightforwardly

$$\mathcal{J}_{\mathbf{d}_0}(\mathbf{l},t;\mathbf{l}_1) = \frac{1}{2^6\pi^3}\left(\frac{1}{\eta_0^2-\eta_t^2}\right)^{3/2} \times \exp\left[-\frac{\eta_0(\mathbf{l}-\mathbf{d_0})^2 + \eta_0(\mathbf{l}_1-\mathbf{d_0})^2 - 2\eta_t(\mathbf{l}-\mathbf{d_0})\cdot(\mathbf{l}_1-\mathbf{d_0})}{4(\eta_0^2-\eta_t^2)}\right]. \tag{C.8}$$

We now marginalize over the angles

$$\mathcal{J}_{d_0}(l,t;l_1) \equiv \int \mathrm{d}\mathbf{d}\int \mathrm{d}\mathbf{d}_1\int \mathrm{d}\mathbf{d}_0\,\delta(|\mathbf{d}_0|-d_0)\delta(|\mathbf{l}_1|-l_1)\delta(|\mathbf{l}|-l)\mathcal{J}_{\mathbf{d}_0}(\mathbf{l},t;\mathbf{l}_1), \tag{C.9}$$

by moving to a frame of reference where $\mathbf{d}_0$ is parallel to the $z$ axis, and express all the vectors in spherical coordinates. This removes all delta-functions and $d_0$ in the new frame of reference is just a scalar. By doing so we obtain

$$\mathcal{J}_{d_0}(l,t;l_1) = \frac{1}{2^6\pi^3}\left(\frac{1}{\eta_0^2-\eta_t^2}\right)^{3/2}\exp\left(-\frac{\eta_0 l^2 + \eta_0 l_1^2 + 2(\eta_0-\eta_t)d_0^2}{4(\eta_0^2-\eta_t^2)}\right) l^2 l_1^2 \times \int_0^{2\pi}\mathrm{d}\phi\int_0^{2\pi}\mathrm{d}\phi'\int_{-1}^{1}\mathrm{d}(\cos\,\theta)\int_0^{\pi}\mathrm{d}(\cos\,\theta')$$

$$\times \exp\left[\frac{(\eta_0-\eta_t)ld_0}{2(\eta_0^2-\eta_t^2)}\cos\theta+\frac{(\eta_0-\eta_t)l_1d_0}{2(\eta_0^2-\eta_t^2)}\cos\theta'+\frac{\eta_t ll_1}{2(\eta_0^2-\eta_t^2)}\right.$$
$$\times\big(\cos\phi\cos\phi'\sin\theta\sin\theta'$$
$$\left.+\sin\phi\sin\phi'\sin\theta\sin\theta'+\cos\theta\cos\theta'\big)\right]. \tag{C.10}$$

The two integrals over $\phi$ and $\phi'$ (keeping in mind that $\cos(\phi-\phi')=\cos\phi\cos\phi'+\sin\phi\sin\phi'$) give us

$$\frac{1}{16\pi}\left(\frac{1}{\eta_0^2-\eta_t^2}\right)^{3/2}\exp\left(-\frac{\eta_0 l^2+\eta_0 l_1^2+2(\eta_0-\eta_t)d_0^2}{4(\eta_0^2-\eta_t^2)}\right)l^2l_1^2$$
$$\times\int_{-1}^{1}\mathrm{d}(\cos\theta)\int_0^{\pi}\mathrm{d}(\cos\theta')\exp$$
$$\times\left[\frac{(\eta_0-\eta_t)ld_0}{2(\eta_0^2-\eta_t^2)}\cos\theta+\frac{(\eta_0-\eta_t)l_1d_0}{2(\eta_0^2-\eta_t^2)}\cos\theta'\right.$$
$$\left.+\frac{\eta_t ll_1}{2(\eta_0^2-\eta_t^2)}\cos\theta\cos\theta'\right]I_0$$
$$\times\left(\frac{\eta_t ll_1}{2(\eta_0^2-\eta_t^2)}\sqrt{1-\cos^2\theta}\sqrt{1-\cos^2\theta'}\right), \tag{C.11}$$

where $I_0(x)$ is the modified Bessel function of the first kind. The first integral in $\cos\theta'$ is solvable [76], and by changing the variable $\cos\theta\to x$ we are left with

$$\frac{1}{8\pi}\left(\frac{1}{\eta_0^2-\eta_t^2}\right)^{3/2}\exp\left(-\frac{\eta_0 l^2+\eta_0 l_1^2+2(\eta_0-\eta_t)d_0^2}{4(\eta_0^2-\eta_t^2)}\right)l^2l_1^2$$
$$\times\int_{-1}^{1}\mathrm{d}x\,\mathrm{e}^{\frac{(\eta_0-\eta_t)ld_0}{2(\eta_0^2-\eta_t^2)}x}\frac{\sinh\left(\sqrt{\frac{(\eta_0-\eta_t)^2l_1^2d_0^2+\eta_t^2l^2l_1^2+2\eta_t(\eta_0-\eta_t)ll_1^2d_0x}{4(\eta_0^2-\eta_t^2)^2}}\right)}{\sqrt{\frac{(\eta_0-\eta_t)^2l_1^2d_0^2+\eta_t^2l^2l_1^2+2\eta_t(\eta_0-\eta_t)ll_1^2d_0x}{4(\eta_0^2-\eta_t^2)^2}}}. \tag{C.12}$$

And the final integral yields [77]

$$\mathcal{J}_{d_0}(l,t;l_1)=\frac{1}{16\sqrt{\pi}}\left(\frac{1}{\eta_0^2-\eta_t^2}\right)^{3/2}\exp\left(-\frac{\eta_0 l^2+\eta_0 l_1^2+2(\eta_0-\eta_t)d_0^2}{4(\eta_0^2-\eta_t^2)}\right)$$
$$\times l^2l_1^2\frac{\mathrm{e}^{-ab/c-c/4a}}{\sqrt{ac}}\left[\mathrm{erfi}\left(\frac{2a\sqrt{b-c}-c}{2\sqrt{ac}}\right)\right.$$
$$-\mathrm{erfi}\left(\frac{2a\sqrt{b-c}+c}{2\sqrt{ac}}\right)$$
$$\left.+\mathrm{erfi}\left(\frac{c-2a\sqrt{b+c}}{2\sqrt{ac}}\right)+\mathrm{erfi}\left(\frac{c+2a\sqrt{b+c}}{2\sqrt{ac}}\right)\right] \tag{C.13}$$

having defined

$$a = \frac{(\eta_0 - \eta_t) l d_0}{2(\eta_0^2 - \eta_t^2)}, \tag{C.14}$$

$$b = \frac{(\eta_0 - \eta_t)^2 l_1^2 d_0^2 + \eta_t^2 l^2 l_1^2}{4(\eta_0^2 - \eta_t^2)^2}, \tag{C.15}$$

$$c = \frac{\eta_t(\eta_0 - \eta_t) l l_1^2 d_0}{2(\eta_0^2 - \eta_t^2)^2}; \tag{C.16}$$

the direct substitution of these auxiliary variables gives, upon division by $\mathcal{P}_{d_0}^{\mathrm{eq}}$ and some simplification, equation (23).

## Appendix D. Derivation of equilibrium autocorrelation function

In order to compute the autocorrelation function in equation (29) the following integrals must be evaluated

$$\mathcal{V}_{\mathbf{0N}}^{d_0} = \int_0^\infty \mathrm{d}x\, V_{\mathbf{0N}}(x, d_0) x, \qquad \mathcal{V}_{\mathbf{N0}}^{d_0} = \int_0^\infty \mathrm{d}x\, V_{\mathbf{N0}}(x, d_0) x. \tag{D.1}$$

These two integrals are identical and the integration yields [76]

$$\begin{aligned}
\mathcal{V}_{\mathbf{0N}}^{d_0} = {} & \frac{1}{16}\sqrt{\frac{1}{2^{\mathcal{N}}\prod_{k=1}^{N} n_{kx}! n_{ky}! n_{kz}!}}\prod_{k=1}^{N}\left(\frac{A_k}{\sqrt{2\mu_k}}\right)^{n_{kx}+n_{ky}+n_{kz}} \\
& \times \frac{1}{\sqrt{\eta_0}^{\mathcal{N}+1}} N_x! N_y! N_z! \mathrm{e}^{-d_0^2/4\eta_0} \\
& \times \sum_{a=0}^{N_x/2}\sum_{b=0}^{N_y/2}\sum_{c=0}^{\lfloor N_z/2\rfloor} \frac{(-1)^{a+b+c}}{a!b!c!\left(\frac{N_x-2a}{2}\right)!\left(\frac{N_y-2b}{2}\right)! 2^{N_x+N_y-2(a+b)}} \\
& \times \sum_{l=0}^{N_z-2c} \frac{1}{l!(N_z-2c-l)!}\left(-\frac{d_0}{\sqrt{\eta_0}}\right)^l \\
& \times \left[(1+(-1)^{N_z-2c-l})\gamma\left(\frac{1+N_z-2c-l}{2}\right) 2^{\mathcal{N}-2(a+b+c)-l+4}\eta_0\gamma\right. \\
& \times \left(\frac{\mathcal{N}-2(a+b+c)-l+4}{2}\right) \\
& \times {}_2\tilde{F}_2\left(\frac{1+N_z-2c-l}{2}, \frac{\mathcal{N}-2(a+b+c)-l+4}{2}; \frac{1}{2},\right. \\
& \left.\left. \times \frac{3+\mathcal{N}-2(a+b+c)-l}{2}; \frac{d_0^2}{4\eta_0}\right)\right. \\
& - (-1+(-1)^{N_z-2c-l}) d_0\sqrt{\eta_0}\gamma\left(1+\frac{N_z-2c-l}{2}\right) \\
& \times 2^{\mathcal{N}-2(a+b+c)-l+3}\gamma\left(\frac{\mathcal{N}-2(a+b+c)-l+5}{2}\right)
\end{aligned}$$

$$\times\, {}_2\tilde{F}_2\left(\frac{2+N_z-2c-l}{2}, \frac{\mathcal{N}-2(a+b+c)-l+5}{2}; \frac{3}{2}, \frac{4+\mathcal{N}-2(a+b+c)-l}{2}; \frac{d_0^2}{4\eta_0}\right)\Bigg] . \tag{D.2}$$

If we are instead interested in the cross-correlation the more general equation (B.20) must be used and the two integrals differ in therms of some constants, i.e. they are obtained by changing the following variables $d_0 \to d_0'$, $\{A_k\} \to \{B_k\}$ and $\eta_t \to \zeta_t$.

### *D.1. Rouse-limit autocorrelation function*

In figure 5 we showed how the autocorrelation for a GNM compares to the autocorrelation in the Rouse limit (i.e. $d_0 \to 0$). The latter can be obtained in a closed form [60]

$$\mathcal{C}(t) = \frac{\langle l(t)l(0)\rangle - \langle l\rangle^2}{\langle l^2\rangle - \langle l\rangle^2}; \tag{D.3}$$

$$\langle l(t)l(0)\rangle = \frac{4\left[3\eta_t\sqrt{\eta_0^2-\eta_t^2} + 2(\eta_0^2+\eta_t^2)\arctan(\eta_t/(\eta_0^2-\eta_t^2))\right]}{\pi\eta_t}, \tag{D.4}$$

$$\langle l\rangle = 4\sqrt{\eta_0/\pi}, \quad \langle l^2\rangle = 6\eta_0. \tag{D.5}$$

## Appendix E. Short-time expansion of $\mathcal{G}_{d_0}$

Introducing the auxiliary variable $\phi(t) = \eta_t/\eta_0$ in equation (23) we can write the return joint-density as and expanding to linear order in $t$ using

$$\phi(t) \overset{t\to 0}{\simeq} 1 - \frac{\sum_{k=1} A_k^2 t}{2\eta_0}\phi^2(t) \overset{t\to 0}{\simeq} 1 - \sum_{k=1} A_k^2 \frac{t}{\eta_0}; \tag{E.1}$$

we find the partial limits

$$\exp\left(-\frac{2d^2\phi(t) + (1-\phi(t))d_0^2}{4\eta_0\phi(t)(1-\phi(t))}\right) \overset{t\to 0}{\sim} \mathrm{e}^{-1/t} \to 0, \tag{E.2}$$

$$\mathrm{erfi}\left(\frac{\pm 2\mathrm{d}\phi(t) + d_0(1-\phi(t))}{2\sqrt{\eta_0\phi(t)(1-\phi(t)^2)}}\right) \overset{t\to 0}{\sim} \mathrm{erfi}(\pm t^{-1/2}) \to \pm\infty, \tag{E.3}$$

$$\mathrm{erfi}\left(\frac{d_0(1-\phi(t))}{2\sqrt{\eta_0\phi(t)(1-\phi(t)^2)}}\right) \overset{t\to 0}{\sim} \mathrm{erfi}(\sqrt{t}) \to 0; \tag{E.4}$$

where all the convergences are of exponential order. Therefore, while we can neglect the second erfi, we need to retain the product between the exponential and the two diverging erfis and only then plug them into in equation (E.1). Thus considering the expansion for large and real arguments of erfi [39]

$$\mathrm{erfi}(x) \overset{x\to\pm\infty, x\in\mathbb{R}}{\simeq} \mp i + \left(\frac{1}{x} + \frac{1}{2x^3} + O(x^{-5})\right)\frac{\mathrm{e}^{x^2}}{\sqrt{\pi}}, \tag{E.5}$$

and explicitly, multiplying by the remaining exponentials equation (23) becomes (note that $\mathcal{P}^{\rm eq}_{d_0}(l)\mathcal{G}_{d_0}(l,t|l) \equiv \mathcal{J}_{d_0}(l,t;l)$)

$$\begin{aligned}\mathcal{J}_{d_0}(d,t;d) \overset{t\to 0}{\simeq} & \frac{d^2}{8\pi d_0}\mathrm{e}^{-(l^2+d_0^2)/2\eta_0(1+\phi(t))} \\ & \times \left\{ \left[ \frac{2\sqrt{1+\phi(t)}}{\sqrt{1-\phi(t)}\eta_0(-2\,\mathrm{d}\phi(t)+d_0(1-\phi(t)))} \right.\right. \\ & \left. + 4\frac{\phi(t)\sqrt{1-\phi(t)}(1+\phi(t))^{3/2}}{(-2\,\mathrm{d}\phi(t)+d_0(1-\phi(t)))^3} \right] \mathrm{e}^{-ld_0/\eta_0(1+\phi(t))} \\ & + \left[ \frac{2\sqrt{1+\phi(t)}}{\sqrt{1-\phi(t)}\eta_0(2\,\mathrm{d}\phi(t)+d_0(1-\phi(t)))} \right. \\ & \left.\left. + 4\frac{\phi(t)\sqrt{1-\phi(t)}(1+\phi(t))^{3/2}}{(2\,\mathrm{d}\phi(t)+d_0(1-\phi(t)))^3} \right] \mathrm{e}^{ld_0/\eta_0(1+\phi(t))} \right] \right\}. \end{aligned} \tag{E.6}$$

Using equation (E.1) and expanding $t = 0$ and introducing $\kappa = \sum_{k=1}^{N} A_k^2$ we finally arrive at equation (36).

## Appendix F. Evaluation of the variance of the occupation time fraction

The direct implementation of equation (34) suffers from slow convergence issues. We suspect that this problem has his roots in the (well-known) slow convergence of series involving Hermite polynomials [78]. We therefore combine the analytical short-time asymptotics in equation (37) with the spectral solution. Defining a small cutoff time $t_{\rm s} \ll 1$ and rewriting equation (35) (using the linearity of integration) as

$$\begin{aligned}\sigma^2_{d_0}(l,t) = & \frac{2\mathcal{P}^{\rm eq}_{d_0}(l)}{t}\int_0^{t_{\rm s}} \mathrm{d}\tau(1-\tau/t)\mathcal{G}_{d_0}(l,\tau|l) + \frac{2\mathcal{P}^{\rm eq}_{d_0}(l)}{t} \\ & \times \int_{t_s}^{t} \mathrm{d}\tau(1-\tau/t)[\mathcal{G}_{d_0}(l,\tau|l) - \mathcal{P}^{\rm eq}_{d_0}(l)]. \end{aligned} \tag{F.1}$$

We can explicitly evaluate the first addend using equation (37) and evaluate the second term using the spectral expansion (25). Note that the first term in the series (with $\Lambda_{\mathbf{0}} = 0$) must be treated in a manner different than the rest. Therefore $\sigma^2_{d_0}(l,t)$ can be conveniently written (and implemented) in the form

$$\begin{aligned}\sigma^2_{d_0}(d,t) = & 2\mathcal{P}^{\rm eq}_{d_0}(l)\left(\frac{8}{3\sqrt{\kappa\pi t}} + \frac{4}{15l^2}\sqrt{\frac{\kappa t}{\pi}} - \mathcal{P}^{\rm eq}_{d_0}(l)\right) + \frac{2}{t^2}\sum_{\mathbf{N}\neq\mathbf{0}} V_{\mathbf{N0}}(l;d_0)V_{\mathbf{0N}}(l;d_0) \\ & \times \left[(t-t_{\rm s})\frac{\mathrm{e}^{-\Lambda_{\mathbf{N}}t_{\rm s}}}{\Lambda_{\mathbf{N}}} - \frac{\mathrm{e}^{-\Lambda_{\mathbf{N}}t_{\rm s}} - \mathrm{e}^{-\Lambda_{\mathbf{N}}t}}{\Lambda^2_{\mathbf{N}}}\right] + \mathcal{P}^{\rm eq}_{d_0}(l)^2\left(\frac{t_{\rm s}}{t}-2\right)\frac{t_{\rm s}}{t}. \end{aligned} \tag{F.2}$$

## Appendix G. Notes on the numerical implementation of the results

Accompanying this article there is a C++ implementation of all analytical results. The code allows the computation the Green's function $\mathcal{G}_{d_0}$, the mean $\langle\theta_t(l,d_0)\rangle$ and variance $\sigma^2_{d_0}(d,t)$ of the occupation time fraction, as well as the autocorrelation function $\mathcal{C}_{d_0}(t)$ for a generic Gaussian Network. The connectivity matrix of the network $\Gamma$ must be provided as a plain text file and is diagonalized using the Armadillo library [79, 80]. A closed-form expression of the joint density in equation (23) is implemented in the available C++ code. However, for numerical stability and speed of computation it is convenient to implement equation (C.12) and perform the final integral numerically using a Gauss-Kronrod quadrature routine [61].

The results based on the evaluation of both, equations (B.19) and (D.2) require the evaluation of the less common regularized hypergeometric functions ${}_p\tilde{F}_q$. A notable exception is the Arblib library [81], that implements several 'special' functions using arbitrary precision arithmetic. The reliable evaluation of such functions is challenging and often requires several different methods to cover the entire domain [82]. Unfortunately this higher reliability comes with a higher computational cost compared to machine precision arithmetic. However hypergeometric functions converge on the entire complex plane if $p \leqslant q$ [82]. In addition, we only need to evaluate them when all the parameters are positive real numbers. Therefore we implemented the series definitions of these function directly since in our case these converge reasonably fast to a desired accuracy as long as the parameters are not too large.

Many of our results, in particular the autocorrelation function and the variance of the fraction of occupation time, can only be expressed fully analytically using the eigendecomposition of the Fokker–Plank operator. Unfortunately the computational effort required in the generation of all necessary terms to achieve convergence is huge. In addition, this number scales non-polynomially with the number of beads in the network. Therefore the attached program should be used with care as it does not generate reliable results when the size of the network becomes too large.

### ORCID iDs

Alessio Lapolla https://orcid.org/0000-0002-3305-2611
Aljaž Godec https://orcid.org/0000-0003-1888-6666

### References

[1] Henzler-Wildman K and Kern D 2007 Dynamic personalities of proteins *Nature* **450** 964–72
[2] van Gunsteren W F and Berendsen H J C 1990 Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry *Angew. Chem., Int. Ed. Engl.* **29** 992–1023
[3] Klepeis J L, Lindorff-Larsen K, Dror R O and Shaw D E 2009 Long-timescale molecular dynamics simulations of protein structure and function *Curr. Opin. Struct. Biol.* **19** 120–7
[4] Shaw D E *et al* 2010 Atomic-level characterization of the structural dynamics of proteins *Science* **330** 341–6
[5] Baiesi M, Orlandini E, Seno F and Trovato A 2017 Exploring the correlation between the folding rates of proteins and the entanglement of their native states *J. Phys. A: Math. Theor.* **50** 504001
[6] Micheletti C, Banavar J R, Maritan A and Seno F 1999 Protein structures and optimal folding from a geometrical variational principle *Phys. Rev. Lett.* **82** 3372–5
[7] Tirion M M 1996 Large amplitude elastic motions in proteins from a single-parameter, atomic analysis *Phys. Rev. Lett.* **77** 1905–8

[8] Rouse P E 1953 A theory of the linear viscoelastic properties of dilute solutions of coiling polymers *J. Chem. Phys.* **21** 1272–80
[9] Flory P J 1976 Statistical thermodynamics of random networks *Proc. R. Soc.* A **351** 351–80
[10] Bahar I, Atilgan A R and Erman B 1997 Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential *Folding Des.* **2** 173–81
[11] Haliloglu T, Bahar I and Erman B 1997 Gaussian dynamics of folded proteins *Phys. Rev. Lett.* **79** 3090–3
[12] Bahar I, Lezon T R, Bakan A and Shrivastava I H 2010 Normal mode analysis of biomolecular structures: functional mechanisms of membrane proteins *Chem. Rev.* **110** 1463–97
[13] Doruker P, Atilgan A R and Bahar I 2000 Dynamics of proteins predicted by molecular dynamics simulations and analytical approaches: application to $\alpha$-amylase inhibitor *Proteins* **40** 512–24
[14] Atilgan A R, Durell S R, Jernigan R L, Demirel M C, Keskin O and Bahar I 2001 Anisotropy of fluctuation dynamics of proteins with an elastic network model *Biophys. J.* **80** 505–15
[15] Delarue M and Dumas P 2004 On the use of low-frequency normal modes to enforce collective movements in refining macromolecular structural models *Proc. Natl Acad. Sci.* **101** 6957–62
[16] Schröder G F, Brunger A T and Levitt M 2007 Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution *Structure* **15** 1630–41
[17] Ming D and Brüschweiler R 2006 Reorientational contact-weighted elastic network model for the prediction of protein dynamics: comparison with NMR relaxation *Biophys. J.* **90** 3382–8
[18] Tang Q-Y, Zhang Y-Y, Wang J, Wang W and Chialvo D R 2017 Critical fluctuations in the native state of proteins *Phys. Rev. Lett.* **118** 088102
[19] Hamacher K, Trylska J and McCammon J A 2006 Dependency map of proteins in the small ribosomal subunit *PLoS Comput. Biol.* **2** e10
[20] Rader A J, Anderson G, Isin B, Khorana H G, Bahar I and Klein-Seetharaman J 2004 Identification of core amino acids stabilizing rhodopsin *Proc. Natl Acad. Sci.* **101** 7246–51
[21] Keskin O, Durell S R, Bahar I, Jernigan R L and Covell D G 2002 Relating molecular flexibility to function: a case study of tubulin *Biophys. J.* **83** 663–80
[22] Weng J, Ma J, Fan K and Wang W 2008 The conformational coupling and translocation mechanism of vitamin $B_{12}$ ATP-binding cassette transporter BtuCD *Biophys. J.* **94** 612–21
[23] Putz I and Brock O 2017 Elastic network model of learned maintained contacts to predict protein motion *PLoS One* **12** e0183889
[24] Zheng W and Brooks B 2005 Identification of dynamical correlations within the myosin motor domain by the normal mode analysis of an elastic network model *J. Mol. Biol.* **346** 745–59
[25] Kundu S, Sorensen D C and Phillips G N 2004 Automatic domain decomposition of proteins by a Gaussian network model *Proteins* **57** 725–33
[26] Zhang Z, Shi Y and Liu H 2003 Molecular dynamics simulations of peptides and proteins with amplified collective motions *Biophys. J.* **84** 3583–93
[27] Hamacher K 2008 Relating sequence evolution of HIV1-protease to its underlying molecular mechanics *Gene* **422** 30–6
[28] Zhang J, Koo B, Liu Y, Zou J, Chattopadhyay A and Dai L 2015 A novel statistical spring-bead based network model for self-sensing smart polymer materials *Smart Mater. Struct.* **24** 085022
[29] Zhang J, Koo B, Subramanian N, Liu Y and Chattopadhyay A 2016 An optimized cross-linked network model to simulate the linear elastic material response of a smart polymer *J. Intell. Mater. Syst. Struct.* **27** 1461–75
[30] Tama F and Brooks C L 2002 The mechanism and pathway of pH induced swelling in cowpea chlorotic mottle virus *J. Mol. Biol.* **318** 733–47
[31] Shrivastava I H and Bahar I 2006 Common mechanism of pore opening shared by five different potassium channels *Biophys. J.* **90** 3929–40
[32] Bocquet N, Nury H, Baaden M, Le Poupon C, Changeux J-P, Delarue M and Corringer P-J 2009 X-ray structure of a pentameric ligand-gated ion channel in an apparently open conformation *Nature* **457** 111–4
[33] Pinamonti G, Bottaro S, Micheletti C and Bussi G 2015 Elastic network models for RNA: a comparative assessment with molecular dynamics and SHAPE experiments *Nucleic Acids Res.* **43** 7260–9
[34] Sułstrokkowska J I, Kloczkowski A, Sen T Z, Cieplak M and Jernigan R L 2008 Predicting the order in which contacts are broken during single molecule protein stretching experiments *Proteins* **71** 45–60

[35] Goldstein H, Poole C P and Safko J L 2002 *Classical Mechanics* 3rd edn (Reading, MA: Addison-Wesley Developers Press)
[36] Nicolay S and Sanejouand Y-H 2006 Functional modes of proteins are among the most robust *Phys. Rev. Lett.* **96** 078104
[37] Lezon T R, Shrivastava I H, Yang Z and Bahar I 2009 Elastic network models for biomolecular dynamics: theory and application to membrane proteins and viruses *Handbook on Biological Networks* (Hackensack, NJ: World Scientific) pp 129–58
[38] Pyun C W and Fixman M 1965 Intrinsic viscosity of polymer chains *J. Chem. Phys.* **42** 3838–44
[39] Abramowitz M and Stegun I A (ed) 2013 *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables* (*Dover Books on Mathematics*) 9th edn (New York: Dover)
[40] Foata D 1981 Some Hermite polynomial identities and their combinatorics *Adv. Appl. Math.* **2** 250–9
[41] Risken H and Frank T 1996 *The Fokker–Planck Equation: Methods of Solution and Applications* (*Springer Series in Synergetics*) 2nd edn (Berlin: Springer)
[42] Amadei A, Linssen A B M and Berendsen H J C 1993 Essential dynamics of proteins *Proteins* **17** 412–25
[43] Lapolla A, Hartich D and Godec A 2020 Spectral theory of fluctuations in time-average statistical mechanics of reversible and driven systems *Phys. Rev. Res.* **2** 043084
[44] Lapolla A and Godec A 2019 Manifestations of projection-induced memory: general theory and the tilted single file *Front. Phys.* **7** 182
[45] Gardiner C W 1985 *Handbook of Stochastic Methods for Physics, Chemistry and Natural Sciences* 2nd edn (Berlin: Springer)
[46] Truong K and Ikura M 2001 The use of FRET imaging microscopy to detect protein–protein interactions and protein conformational changes *in vivo Curr. Opin. Struct. Biol.* **11** 573–8
[47] Ye W *et al* 2018 Conformational dynamics of a single protein monitored for 24 h at video rate *Nano Lett.* **18** 6633–7
[48] Kac M 1949 On distributions of certain Wiener functionals *Trans. Am. Math. Soc.* **65** 1
[49] Yen J-Y and Yor M 2013 *Local Times and Excursion Theory for Brownian Motion: A Tale of Wiener and Itô Measures: 2088* (*Lecture Notes in Mathematics*) (Berlin: Springer)
[50] Majumdar S N and Comtet A 2002 Local and occupation time of a particle diffusing in a random medium *Phys. Rev. Lett.* **89** 060601
[51] Majumdar S N 2006 Brownian functionals in physics and computer science *The Legacy of Albert Einstein* (Singapore: World Scientific) pp 93–129
[52] Lapolla A and Godec A 2018 Unfolding tagged particle histories in single-file diffusion: exact single- and two-tag local times beyond large deviation theory *New J. Phys.* **20** 113021
[53] Gopich I V and Szabo A 2003 Single-macromolecule fluorescence resonance energy transfer and free-energy profiles *J. Phys. Chem* B **107** 5058–63
[54] Müller C W and Schulz G E 1992 Structure of the complex between adenylate kinase from *E. coli* and the inhibitor $Ap_5A$ refined at 1.9 Å resolution *J. Mol. Biol.* **224** 159–77
[55] Müller C, Schlauderer G, Reinstein J and Schulz G 1996 Adenylate kinase motions during catalysis: an energetic counterweight balancing substrate binding *Structure* **4** 147–56
[56] Henzler-Wildman K A *et al* 2007 Intrinsic motions along an enzymatic reaction trajectory *Nature* **450** 838–44
[57] Hanson J A, Duderstadt K, Watkins L P, Bhattacharyya S, Brokaw J, Chu J-W and Yang H 2007 Illuminating the mechanistic roles of enzyme conformational dynamics *Proc. Natl Acad. Sci.* **104** 18055–60
[58] Zheng Y and Cui Q 2018 Multiple pathways and time scales for conformational transitions in apo-adenylate kinase *J. Chem. Theory Comput.* **14** 1716–26
[59] Bakan A, Meireles L M and Bahar I 2011 ProDy: protein dynamics inferred from theory and experiments *Bioinformatics* **27** 1575–7
[60] Lapolla A and Godec A 2021 Toolbox for quantifying memory in dynamics along reaction coordinates *Phys. Rev. Res.* **3** L022018
[61] Agrawal Nikhar *et al* 2020 https://boost.org/doc/libs/1_73_0/libs/math/doc/html/math_toolkit/gauss_kronrod.html
[62] Lipari G and Szabo A 1982 Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1 Theory and range of validity *J. Am. Chem. Soc.* **104** 4546–59
[63] Perico A, Pratolongo R, Freed K F, Pastor R W and Szabo A 1993 Positional time correlation function for one-dimensional systems with barrier crossing: memory function corrections to the optimized Rouse–Zimm approximation *J. Chem. Phys.* **98** 564–73

[64] Lapolla A and Godec A 2020 Single-file diffusion in a bi-stable potential: signatures of memory in the barrier-crossing of a tagged-particle *J. Chem. Phys.* **153** 194104
[65] Balsera M A, Wriggers W, Oono Y and Schulten K 1996 Principal component analysis and long time protein dynamics *J. Phys. Chem.* **100** 2567–72
[66] Naritomi Y and Fuchigami S 2011 Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis: the case of domain motions *J. Chem. Phys.* **134** 065101
[67] Naritomi Y and Fuchigami S 2013 Slow dynamics of a protein backbone in molecular dynamics simulation revealed by time-structure based independent component analysis *J. Chem. Phys.* **139** 215102
[68] Sittel F and Stock G 2018 Perspective: identification of collective variables and metastable states of protein dynamics *J. Chem. Phys.* **149** 150901
[69] Gohlke H and Thorpe M F 2006 A natural coarse graining for simulating large biomolecular motion *Biophys. J.* **91** 2115–20
[70] Tian Y, Shirinzadeh B and Zhang D 2010 Design and dynamics of a 3-DOF flexure-based parallel mechanism for micro/nano manipulation *Microelectron. Eng.* **87** 230–41
[71] Sierra D P, Weir N A and Frank J 2005 *A Review of Research in the Field of Nanorobotics* Sandia National Laboratories
[72] Henry C 1979 Structural rigidity *Structural Topology Núm. 1* (Montréal: Université du Québec à Montréal)
[73] Calladine C R 1978 Buckminster Fuller's 'tensegrity' structures and Clerk Maxwell's rules for the construction of stiff frames *Int. J. Solids Struct.* **14** 161–72
[74] Clerk Maxwell J 1864 On the calculation of the equilibrium and stiffness of frames *London Edinburgh Dublin. Philos. Mag.* **27** 294–9
[75] Pellegrino S and Calladine C R 1986 Matrix analysis of statically and kinematically indeterminate frameworks *Int. J. Solids Struct.* **22** 409–28
[76] Gradshteyn I S and Ryzhik I M 2007 *Table of Integrals, Series, and Products* 7th edn (Amsterdam: Elsevier)
[77] Wolfram Research, Inc 2019 Mathematica *(Version 12.0)*
[78] Boyd J P 1980 The rate of convergence of Hermite function series *Math. Comp.* **35** 1309
[79] Sanderson C and Curtin R 2016 Armadillo: a template-based C++ library for linear algebra *J. Open Source Softw.* **1** 26 Publisher: The Open Journal
[80] Sanderson C and Curtin R 2018 A user-friendly hybrid sparse matrix class in C++ *Mathematical Software—ICMS 2018* ed J H Davenport, M Kauers, L George and J Urban (Berlin: Springer) pp 422–30
[81] Johansson F 2017 Arb: efficient arbitrary-precision midpoint-radius interval arithmetic *IEEE Trans. Comput.* **66** 1281–92
[82] Johansson F 2019 Computing hypergeometric functions rigorously *ACM Trans. Math. Softw.* **45** 30