Analysis of velocity autocorrelation functions from molecular dynamics simulations of a small peptide by the generalized Langevin equation with a power-law kernel

Daniel Abergel, Antonino Polimeno, Mirco Zerbetto

Analysis of velocity autocorrelation functions from molecular dynamics simulations of a small peptide by the generalized Langevin equation with a power-law kernel.

Daniel Abergel,*,† Antonino Polimeno,‡ and Mirco Zerbetto‡

†*Laboratoire des biomolécules, LBM, Département de chimie, Ecole normale supérieure, PSL University, Sorbonne Université, CNRS, 75005 Paris, France.*

‡*Dipartimento di Scienze Chimiche, Università degli Studi di Padova, via Marzolo, 1, I-35131, Padova, Italy.*

E-mail: daniel.abergel@ens.psl.eu

Abstract

Internal motions play an essential role in the biological functions of proteins and have been the study of numerous theoretical and spectroscopic studies. Such complex environments are associated to anomalous diffusion where, in contrast to the classical Brownian motion, the relevant correlation functions have power law decays with time. In this work, we investigate the presence of long memory stochastic processes through the analysis of atomic velocity auto-correlation functions. Analytical expressions of the VACF spectrum obtained through a Mori-Zwanzig projection approach were shown to be compatible with molecular dynamics simulations of a small helical peptide (8-polyalanine).

Introduction

Complex molecular systems are characterized by the existence of highly irregular energy landscapes that admit numerous local energy minima that originate from the presence of, potentially diverse, mutually interacting units. Such systems have been identified in various contexts, ranging from glasses, liquid crystals, to polymers and proteins,[1–3] in which the dynamics is described as ensembles of particles evolving in a rough potential. The existence of complex environments is associated to the phenomenon of anomalous diffusion (for a general review, see for instance[4]), where, in particular, relevant correlation functions have power law decays with time, in contrast to the classical Brownian motion. The interior of a protein provides an illustration of such an irregular medium, and the existence of subdiffusive processes has been indeed investigated in the context of protein dynamics using various techniques, such as fluorescence correlation spectroscopy,[3,5] as well as in the context of neutron scattering[6,7] or in an NMR perspective focusing on spin relaxation.[8–10]

Anomalous diffusion can be described theoretically by a variety of approaches (Continuous time random walk,[11–14] fractional Fokker-Planck equations,[4,15,16]), all of which have in common the existence of a stochastic process with long memory kernels. In the context of protein NMR relaxation, MD-computed orientation correlation functions of bond vectors were fitted to models based on Mittag-Leffler functions, which are solutions of fractional diffusion equations.[8–10] These correlation functions were shown to better fit the MD correlation functions based on mono-exponentials, without introducing any additional assumptions about local geometry and supporting the assumption of subdiffusive stochastic processes underlying the internal motions in proteins. Such models were used for backbone, side chain methyl groups as well as for chemical shift fluctuations. These studies support a fractional diffusion model, as suggested by the consistency of the subdiffusion parameter when motions are analyzed at different time scales.[10,17]

However, some ambiguity persists as a Mittag-Leffler function is itself a continuous

superposition of exponentials (albeit of a particular kind) and therefore may be compatible with a function decaying multi-exponentially, such as expected in a Markovian, memoryless process. This calls for a complementary, or more fundamental, characterization.

The projection approach in the phase space introduced by Mori[18] and Zwanzig[19] (MZ projection approach), and leading to the Generalized Langevin Equation (GLE), allows one to describe the evolution of a small subsystem of a larger one in order to focus on the dynamics of the former. The evolution of the complete system is partitioned according to the projection onto a set of phase variables, the evolution of which is governed by the GLE, where the random forces appear as an additive noise term, and friction arises as the convolution with a memory kernel. From the dynamical point of view, characterizing the statistical properties of the Langevin forces appears more fundamental than studying the decay of position correlation functions. However, stochastic forces are not directly inferred from MD simulations that give only access to the sum of the deterministic and random forces acting on the system. This can be circumvented by adopting an alternative projection proposed by Boon and Yip (BY projection approach)[20] and using the relationship between the memory kernels obtained by both approaches, as we show briefly in the next Section. We analyze then the existence of long memory processes by deriving exact expressions for the spectra of the velocity correlation functions in the case of a power-law kernel. The analytical expressions are then applied to the analysis of molecular dynamics trajectories of a small, 8-residue, polyalanine peptide (polyala8), considered as a simple case study.

Overview of MZ and BY projections The goal of a projection technique is to isolate the dynamics of a small *relevant* (e.g. directly observable) part (the "system") from the rest of the system (the "bath"). Consider a classical conservative system, characterized by the phase space of N positions and momenta $\mathcal{E} = (\mathbf{q}, \mathbf{p})$ where $\mathbf{q} = (q_1, \ldots, q_N)$, $\mathbf{p} = (p_1, \ldots, p_N)$. A vector of M phase variables $\mathbf{A}^{\mathrm{t}}(t) = [A_1(t), \ldots, A_M(t)]$, functions of the

positions and momenta, obeys the exact time evolution equations:

$$\frac{d}{dt}\mathbf{A}(t) = \dot{\mathbf{A}} = i\mathcal{L}\mathbf{A}(t) \tag{1}$$

where $\mathcal{L} = \{\cdot, \mathcal{H}\}$ is the Liouvillian of the system. following the MZ approach, the evolution of the phase variables $\mathbf{A}(t)$) of a larger system can be obtained by projecting Eq. 1 onto the phase variables $\mathbf{A}$ using the projector operator $\mathcal{P}f = \sum_{ij} A_i S_{ij} \langle A_j | f \rangle = \mathbf{A}^{\mathrm{t}} \mathbf{S} \langle \mathbf{A} | f \rangle$, where $\langle f | g \rangle$ is the Hermitian product and $\mathbf{S} = \mathbf{M}^{-1}$, having defined the overlap matrix $\mathbf{M} = \langle \mathbf{A} | \mathbf{A}^{\mathrm{t}} \rangle$. The GLE can now be recovered (cfr SI for details) in the form

$$\frac{d}{dt}\mathbf{A}^t(t) = i\mathbf{A}^t(t)\mathbf{\Omega} - \int_0^t \mathbf{A}^t(t-t')\mathbf{k}(t')dt' + \mathbf{F}(t) \tag{2}$$

where $i\mathbf{\Omega} = \mathbf{S}\langle \mathbf{A} | \dot{\mathbf{A}}^{\mathrm{t}} \rangle$, $\mathbf{F}$ is the Langevin random force and the memory kernel $\mathbf{k}(t)$ obeys the relation:

$$\mathbf{k}(t) = \mathbf{S}\mathbf{C}_F(t) \tag{3}$$

where the force correlation matrix $\mathbf{C}_F(t) = \langle \mathbf{F}(0) | \mathbf{F}(t)^{\mathrm{t}} \rangle$ is defined. Eq. 3 takes the role of a fluctuation-dissipation theorem, which relates the correlation function of the random force to the memory kernel.

Most physical measurable quantities in linear spectroscopy involve position or velocity correlation functions that are contained in the phase variable correlation matrix $\mathbf{\Psi}(t) = \mathbf{S}\langle \mathbf{A} | \mathbf{A}^{\mathrm{t}}(t) \rangle$. The latter can be derived from the Mori-Zwanzig GLE:

$$\frac{d}{dt}\mathbf{\Psi}(t) = \int_0^t dt' \mathbf{\Psi}(t-t')[i\mathbf{\Omega}\delta(t') - \mathbf{k}(t')] \tag{4}$$

Taking into account that $\mathbf{\Psi}(0) = \mathbf{1}$ one gets for the Laplace transform $\tilde{\mathbf{\Psi}}(s)$ of $\mathbf{\Psi}(t)$:

$$\tilde{\mathbf{\Psi}}(s) = \left[s - i\mathbf{\Omega} + \tilde{\mathbf{k}}(s) \right]^{-1}. \tag{5}$$

In principle, the GLE kernel contains all the necessary information to characterize the time evolution of relevant correlation functions. The direct evaluation of the random force via MD simulations, however, is difficult, since only the total forces, and not their stochastic components as defined within the MZ approach, are computed. An alternative approach developed by Boon and Yip[20] bridges this gap by making use of the projection operator directly on the evolution equation of $A_i(t)$. One gets an alternative expression for $\tilde{\mathbf{\Psi}}(s)$ (the derivation is detailed in the SI). The results is of the form:

$$\tilde{\mathbf{\Psi}}(s) = [s\mathbf{1} + \tilde{\mathbf{K}}(s)]^{-1}\mathbf{\Psi}(0), \tag{6}$$

where the memory kernel $\tilde{\mathbf{K}}(s)$ is given explicity in the SI. Equation 6 is the memory, or master equation:

$$\frac{d}{dt}\mathbf{\Psi}(t) = \int_0^t dt' \mathbf{K}(t')\mathbf{\Psi}(t - t') \tag{7}$$

The master equation is obtained using the same definition of the projection operator as in the Mori-Zwanzig approach that leads to the generalized Langevin equation. However, in the latter, the kernel $\tilde{\mathbf{k}}(s)$ is equal to the Laplace transform of the random force correlation function, to within a normalization factor. The connection between both projection approaches and their associated kernels can be deduced easily. By comparing Eqs. 5 and 6 after some passages we get

$$\tilde{\mathbf{K}}^{-1}(s) = \left[\tilde{\boldsymbol{\mathcal{F}}}(s) - i\mathbf{\Omega} \right]^{-1} - s^{-1}\mathbf{1} \tag{8}$$

linking the kernel of the BY approach directly to the total force and the streaming part of the MZ approach.

GLE with power law kernel As underlined previously, the rewriting of the Liouville equation yields the GLE with a memory kernel and a stochastic force $\mathbf{F}(t)$.We now proceed to apply such a description to atoms in a molecular structure (e.g. a peptide or a protein) in order to investigate the possible presence of long memory effects, i.e., of a power law Langevin kernels $k(t) \sim (t/\tau)^{-\alpha}$. For this purpose we compare the spectra of the normalized atomic velocity autocorrelation functions (VACFs) $\langle v_l(t)v_l(0)\rangle/\langle v^2\rangle$, $l = x, y, z$ computed from MD simulations, to the analytical solutions obtained from the GLE with long memory.

Consider the dynamics of a single component q of the vector position of a single atom with mass m, and its conjugate momentum p, with velocity $v = p/m$. It can be shown that the general form of the GLE given by Eq. 2 simplifies for the phase variables $\mathbf{A}_a^t = (q, p)^t$, to yield (see Eqs. SI 34-SI 37 in the SI):

$$\begin{cases} \dot{q}(t) = \dfrac{p(t)}{m} \\ \dot{p}(t) = -\kappa q(t) - \displaystyle\int_0^t k(t')p(t-t')dt' + F(t) \end{cases} \tag{9}$$

where the fluctuation-dissipation theorem is $\langle F(0)|F(t)\rangle = k(t)\langle p(0)|p(t)\rangle$. In the case of a local harmonic potential $\kappa q^2/2$, where κ is identified as a force constant, and $\kappa = k_BT/\langle q^2\rangle$. The (from now on) normalized momentum autocorrelation function is $C(t) = \langle p(t)p(0)\rangle/\langle p^2\rangle$ and its Laplace transform $\tilde{C}(s)$ can be obtained (see SI) from the autocorrelation matrix:

$$\tilde{C}(s) = \frac{m}{ms[s+\tilde{k}(s)]+\kappa} \begin{pmatrix} s+k(s) & \dfrac{1}{m} \\ -\kappa & s \end{pmatrix} \tag{10}$$

Consider the power law memory kernel:

$$k(t) = \frac{\langle |f_2|^2 \rangle}{\langle p^2 \rangle} \left(\frac{t}{\tau}\right)^{-\alpha} = \xi \left(\frac{t}{\tau}\right)^{-\alpha}, \ 0 < \alpha < 1 \tag{11}$$

Its Laplace transform is $\tilde{k}_{22}(s) = \xi\Gamma(1-\alpha)\tau(s\tau)^{\alpha-1} = \gamma_a(s\tau)^{\alpha-1}$, where $\gamma_a = \xi\Gamma(1-\alpha)\tau$ and $\Gamma(x)$ is the Gamma function, so that the normalized correlation function $\tilde{C}_{v_i}(s) = \tilde{\Psi}_{22}(s)/\langle p_i^2 \rangle$ writes:

$$\tilde{C}(s) = \frac{s}{\gamma_a \tau^{\alpha-1} s^\alpha + s^2 + \frac{\kappa}{m}} \tag{12}$$

In the case of free diffusion, i.e., when the driving term $\omega_0 = \kappa/m = 0$, the expression of $\tilde{C}(s)$ simplifies and one obtains:

$$\tilde{C}(s) = \frac{1}{s + \gamma_a (s\tau)^{\alpha-1}} \tag{13}$$

the inverse Laplace transform of which is:[21]

$$C(t) = E_{2-\alpha}(-\xi\Gamma(1-\alpha)\tau^2(t/\tau)^{2-\alpha}), \tag{14}$$

where:

$$E(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\alpha k + 1)} \tag{15}$$

is the Mittag-Leffler function.[22] Moreover, using the asymptotic property of the Mittag-Leffler functions:[21–26]

$$E_{\alpha,\beta}(-t) \sim \frac{1}{\Gamma(\beta-\alpha)} \frac{1}{t}, t \to +\infty \tag{16}$$

the following expression of $C(t)$ at long times is obtained:

$$C(t) \sim \frac{t^{\alpha-2}}{\Gamma(\alpha-1)\xi\Gamma(1-\alpha)\tau^{\alpha}} \tag{17}$$

It is clear that the projected dynamics of Eq. 9 alone, together with its associated VACF (Eq. 12), do not account for the typical VACF spectra computed from MD simulations that exhibit many vibrational resonant peaks. In contrast, the velocity correlation spectra can be fairly well reproduced based on a conventional Markovian (memoryless) Langevin or Fokker-Planck dynamics of atoms in the molecule.[27,28] The associated Langevin modes thus correspond in this case to vibration (underdamped) and diffusion (overdamped) modes of the velocity correlation function.[29] Importantly, the modes are all exponentially damped, owing to the Markovian and therefore memoryless nature of the Langevin equation. A remark is in order here. Power law memory kernels arise in analyses of subdiffusion motions, where the associated correlation functions are constructed based on stretched Mittag-Leffler functions $E_{\alpha}(-(t/\tau)^{\alpha})$. Subdiffusion processes are described as a continuous tile random walk in a "rugged" potential that represents the various conformational "substates" of the fractal potential energy surface. This eventually leads to long waiting times and time-fractal correlation functions (the streched Mittag-Leffler functions), where the parameter α characterizes the "rugosity" of the potential, that is, its self-similarity. Thus, by construction, α indicates the presence of a significant density of frequency components towards zero in the VACFs. However, this has to be distinguished from the time scales related to the different energy wells in a potential surface, as it can be shown that in this model of subdiffusion, each relaxation mode τ_i decays not as an exponential, but as a stretched Mittag-Leffler function $E_{\alpha}(-(t/\tau_i)^{\alpha})$. This expresses the fact that each mode is associated to the same degree of "rugosity" characterized by the parameter $0 \leq \alpha < 1$. This was discussed for instance in ref.[4] and illustrated in refs.[10,17] in the context of NMR spectroscopy. The obvious discrepancies noted between the model

of Eq. 9 and MD-computed velocity correlation functions are an expected consequence of the "linear" Mori-Zwanzig projection leading to a generalized Langevin equation that involves a single resonance for the selected dynamical variable $\{q\}$. Such a procedure actually transfers the nonlinearities to the Langevin forces and to the memory kernel.[30,31] However, we are mainly interested in the behaviour of the velocity correlation function for long lag times or, equivalently, in its frequency spectrum at extreme low frequencies, when $\omega \to 0^+$. Note that, by nature, a power law kernel is an asymptotic model, meaning that it is indeed expected to take the form of Eq. 11 only at long times, when $t \to \infty$, when the vibrational processes responsible for the resonance spectra of the vacfs have died out. Moreover, analyses based on neutron scattering experiments and MD simulations[32] suggest that in the low energy range ($\lesssim 10$ meV or $\lesssim 2.4$ ps^{-1}) protein vibrations propagate mainly along the backbone, and that the atoms in the molecule undergo a linear potential, in line with our assumption of a memory kernel free from absorbed nonlinearities in the extreme-low frequency regime.

A generalized Langevin equation can be obtained in several ways, the Mori-Zwanzig projection being only one of them, beside the linear response theory or a perturbation approach.[30,33,34] It has been proposed to analyze stochastic processes in various perspectives. In the present case, because we are ultimately interested in the motions of individual atoms in the molecule, we use the MZ projection approach that is directly derived from the hamiltonian equations in the phase space. This is consistent with the perspective of NMR spectroscopy that provides dynamical information on individual atoms, and where the characterization of the stochastic properties of the forces acting upon each atom is of primary interest. Note, however, that, by construction, the MZ approach yields a mean force term that is linear in the phase variable, which is the derivative of a potential only in special case where it is quadratic. Otherwise, the MZ projection rejects nonlinearities into the kernel, which may bias the analysis by lumping contributions from the potential with memory effects, potentially leading to multi-exponential behaviours of the VACFs that

are not related to memory effects. The use of the GLE in our work thus differs in this sense from other approaches recently introduced to study protein folding in MD simulations, where, in a completely different perspective, a phenomenological GLE was employed to study the evolution of some reaction coordinate in a complex energy landscape where the nonlinear potential is explicitly taken into account.[35]

These arguments therefore lead to a model where at short times the dynamics of the system is described along the conventional lines of an ensemble of particles governed by the Langevin equations, whereas at long times, each atom obeys a long memory diffusion process given by the generalized Langevin equation obtained by the Zwanzig-Mori projection approach. We implemented the proposed model using this simple strategy to the small polyala8 peptide.

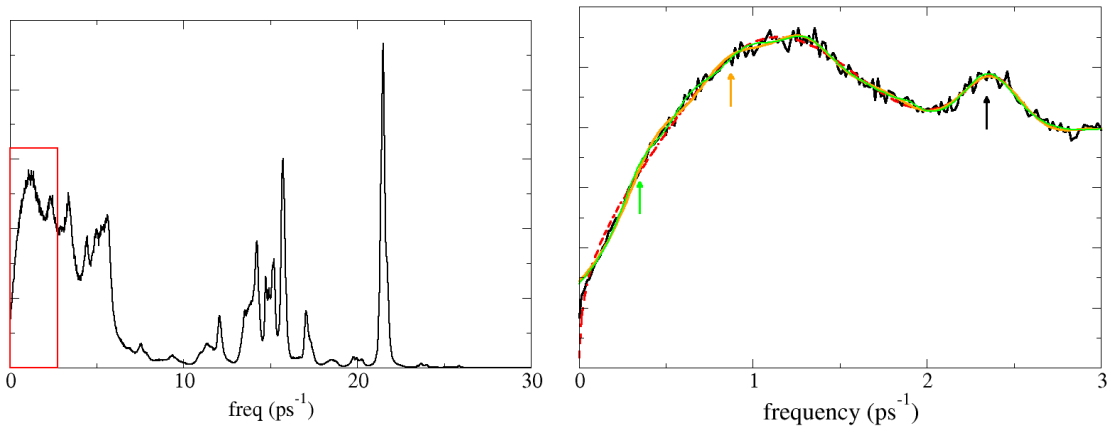
Because the Laplace transform of the VACF admits an analytical expression, it is natural to tackle the problem in the frequency domain. Thus, a frequency threshold ω_M is introduced that defines two different regimes. For frequencies above ω_M, the atomic motion is governed by a Markovian process that contributes a sum of exponentially damped cosine functions. The corresponding VACF spectrum is therefore composed of a series of Lorentzian lines centered at a number of frequencies ω_j. Alternatively, for frequencies below ω_M, the VACF is described by Eq. 12. The VACF spectrum is obtained for arguments $s = \epsilon + i\omega$ in the limit $\epsilon \to 0+$. Thus, the MD-computed VACF spectra were fitted as:

$$\begin{aligned} \tilde{C}(\omega) &= \sum_j \frac{\mathcal{A}_j}{w_J^2 + (\omega - \omega_j)^2},\ \mathcal{I}m(s) \geq \omega_M \\ \tilde{C}(\omega) &= \mathcal{R}e\left(\frac{i\omega}{\gamma_a \tau^{\alpha-1} e^{i\alpha\pi/2}\omega^\alpha - \omega^2 + \frac{\kappa}{m}}\right),\ \mathcal{I}m(s) \ll \omega_M \end{aligned} \tag{18}$$

where $\mathcal{A}_j$ and w_j are the amplitudes and widths of the Lorentzians components of the VACF.

Analysis of the polyala8 peptide In order to illustrate our approach, the above results were tested against MD simulations by using a simple model system poly-alanine peptide, which adopts a globally stable helix structure. The molecular geometry was first optimized via a Molecular Mechanics energy minimization, using the molecular modeling toolkit (MMTK) software package.[36] Minimization was carried out *in vacuo* using the Amber99 force field,[37] which is the default force field in MMTK. We carried on 200 ns-long MD simulations, using the MD package NAMD,[38] following a standard protocol consisting in energy minimization, heating, equilibration, and production. Parameters and conditions for the MD simulations are reported in Table 2 in the *Methods* section. Box dimensions have been chosen equal to 30 Å[39] have shown that 8-alanine is unaffected by the dimension of the simulation box ranging from 30 to 50 Å. It has also been shown that no particular differences are found in the backbone populations and that one configuration (PP_{II}) is relevant.[40] Such an observation is compatible with the description of the molecules (at least for relatively short polyalanines, like in our case) as flexible objects oscillating around a global minimum. The spectra of the VACFs were obtained from MD simulations for a selection of backbone atoms (amide N and C') as well as side chain methyl C_β carbons. In practice, fitting of the spectra was achieved by using a distribution of Lorentzian lines at frequencies down to a frequency threshold below which the spectra were fitted according to the second of Eq. 18. The choice of ω_M was guided by the fact that the model of Eqs.12 and 18 has a single maximum, so the additional maxima observed in the MD-computed VACF were fitted using Lorentztians. In all cases, the lower frequency boundary of the Lorentzian spectrum was taken $\omega_M \approx 1.5\ \mathrm{ps}^{-1}$. For each category of atoms, a single distribution of frequencies, albeit with different ω_M values, was used to define the Lorentzian part of the spectrum in the VACF analysis at different temperatures (278 K, 283 K, and 293 K). Finally, in order to facilitate the fitting process, and because the high frequency part of the spectrum does not affect the low frequency region, only the narrower spectral region $\lesssim 10\ \mathrm{ps}^{-1}$ of the VACF spectrum was considered in all

cases. Most of the VACF spectra were found to be satisfactorily reproduced by multiple Lorentzians, except for their extreme lower ends that could only be captured by the model with long memory. All spectra exhibited a convex curvature towards $\omega \to 0$, a feature well reproduced by the model of Eq. 18. On the contrary pure Lorentzian fitting showed, as expected, a concave curvature at low frequencies. This is illustrated in the example of Figure 1.



(a) (b)

Figure 1: Left: spectrum of the x component of the VACF computed from MD simulations for the Cβ atom of the Ala6 residue (T = 278 K). Right: the low frequency region of the spectrum (red box in (a)) is plotted in black, together with with model spectra obtained from a purely Lorentzian model (orange and green solid lines) or a mixed model that includes a long-memory kernel for low frequencies (red dashed line). The vertical black arrow indicates the lowest Lorentzian component of the spectrum for the long memory model. Two implementations of the purely Lorentzian model were computed, with lowest Lorentzian centered at 0.9 ps^{-1} (orange arrow and orange line) and 0.4 ps^{-1} (green arrow and green line). The latter two illustrate that such a model leads to a concave spectrum towards zero, whereas the long memory kernel is able to reproduce the concavity of the spectrum for $\omega \to 0$.

The VACF spectra for N, C′ and C$_\beta$ atoms in the peptide were fitted according to the model described above and results are depicted in Figures 2 and 3. In addition, the α parameter that characterizes the power law of the kernel was extracted for each atom of these selections. Results are shown in Fig. 4 and in Table 1. The fits all yielded α values that were significantly lower than 1 and that clustered in a relatively narrow range $\alpha \in [0.7 - 0.8]$. However, local variations were found to be more pronounced for amide

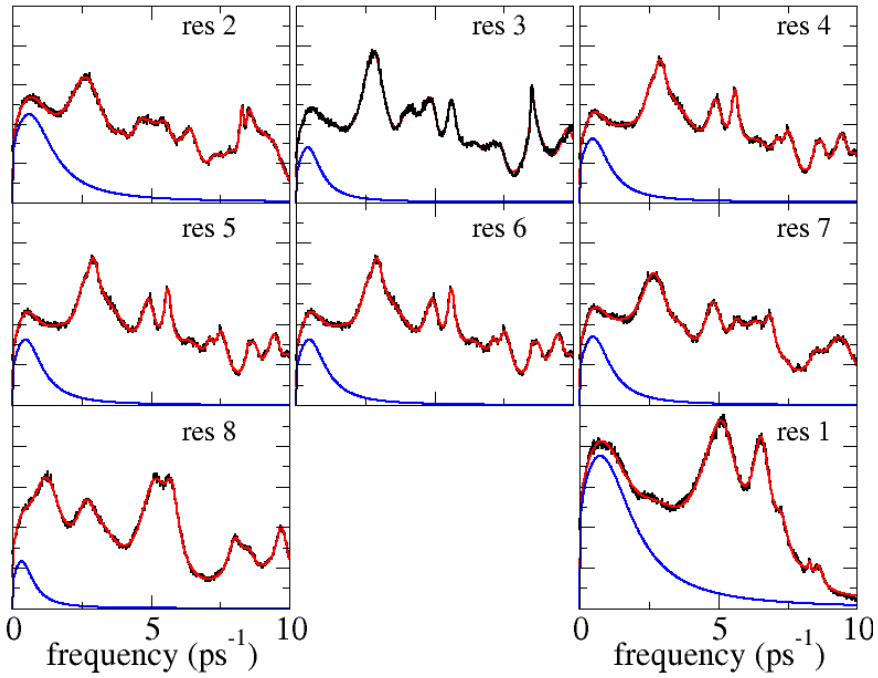


Figure 2: Spectra of the x component of the amide N VACF at $T = 298$ K. The model of Eq. 18 is shown in red, and the lower end contribution is depicted in blue.

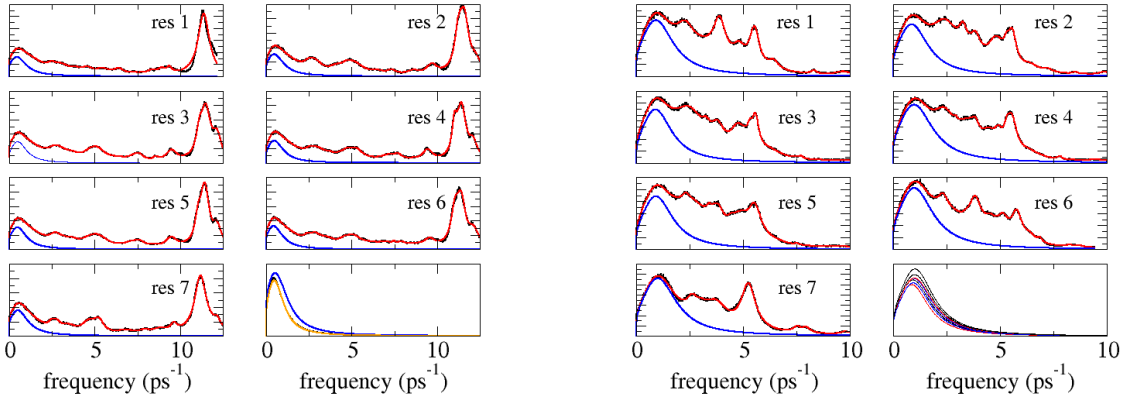


Figure 3: VACF spectra of the x component of the C′ (left) and C_β (right) velocity autocorrelation functions at T = 298 K. Color coding is the same as in Figure 2.

N atoms, whereas they appeared more uniform for C′ and C_β atoms. Nevertheless, all these values obviously indicate the departure from the exponential decay, and and therefore suggest an asymptotic power law decay of the relaxation functions at long times. This kind of behaviour has received several interpretations, in terms of the presence of a

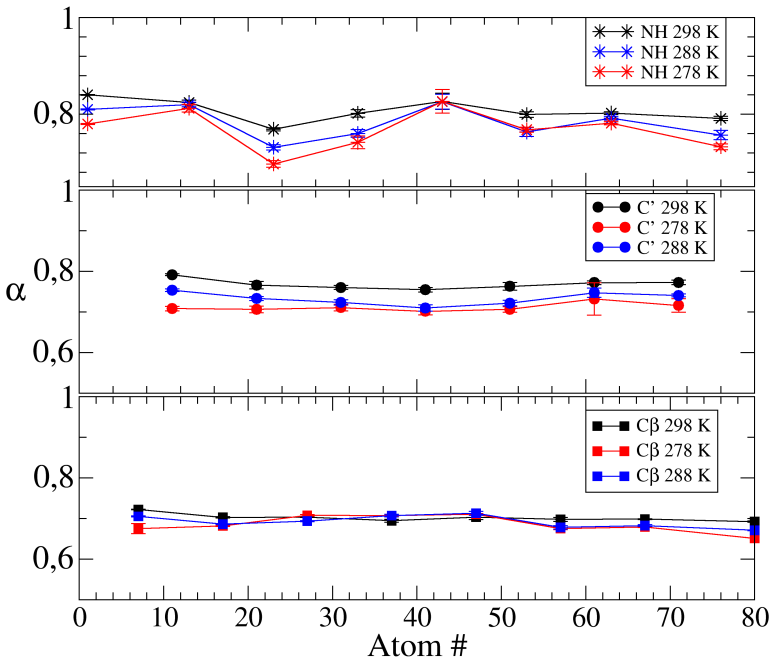


Figure 4: Parameter α of Equation 12 for amide, C' and C_β atoms of the poly-ALA_8 peptide.

Table 1: Average values and standard deviations of the parameter α, $\langle\alpha\rangle \pm \langle\sigma_\alpha\rangle$ computed for the three atom populations NH, C' and C_β of the polyAla8 peptide.

	NH	C'	C_β
278 K	0.77 ± 0.01	0.71 ± 0.01	0.69 ± 0.00
288 K	0.78 ± 0.01	0.73 ± 0.01	0.69 ± 0.00
298 K	0.81 ± 0.01	0.77 ± 0.01	0.70 ± 0.00

complex conformational energy landscape with a hierarchy of conformational substates organized in a self-similar manner, or to the fractal structure of the protein itself,[41] leading to fractal time processes and subdiffusion, although the origin of such arrangements may be several.[42] In the case at hand, the variations of the kernel parameter α suggest different local distributions of such substates across the backbone and for the different classes of backbone atoms (N, C', C_β). Interestingly, recent work based on the analysis of MD simulations of a similar 8-residue poly-Alanine peptide suggests the existence of a fractal topology of the energy landscape, where traps are physically correlated rather than random.[43,44]

Analysis of the MD simulations performed at the different temperatures (278 K, 288

K, and 298 K) for the chosen sets of atoms shows that the average value of α across the peptide sequence, $\langle\alpha\rangle$, tends to increase with temperature (see Table 1). This trend is more pronounced for the C′ atoms, and to a lesser degree for N amides, whereas C_β atoms seem to be less sensitive to temperature. In the perspective of rugged energy landscapes and correlated traps,[1,44] the observed trend can be interpreted as the progressive decrease of these correlations with temperature, therefore of the self-similarity of the conformational energy landscape. Interestingly, a similar increase of α with temperature was first noticed in ref.[1] There, oxygen binding to myoglobin was described as a reaction kinetic process between a ground and an excited state of the protein. The process was analyzed in terms of fractional diffusion to account for self-similarity of a continuous distribution of time scales.

Although a wide variety of approaches has been used to investigate memory effects in protein dynamics and kinetics,[1] including conformational dynamics or folding,[45] the present study is essentially driven by the perspective of NMR spectroscopy, where one is primarily interested in characterizing the motions of bond or atomic motions through the correlation functions attached to them. In this case, the interpretation of spin relaxation requires sound physical models in order to disentangle dynamics from the statistical properties, including memory effects, and to provide plausible models of dynamics that can be used for the analysis of relaxation data, whilst avoiding overfitting. This work focuses on the atomic VACF as the quantity connecting to the Langevin kernel because neither this or the Langevin random forces are directly accessible from MD simulations. However, the autocorrelation functions of the velocities and of the total forces (see Eq. SI 44 in the SI) involve the same (α, τ) characteristic parameters. The spectrum of the force autocorrelation function, which is numerically more demanding, can therefore be calculated. This approach, based on the long-time behaviour of the memory kernel leads to exact expressions that occur naturally in the frequency domain, and are easier to handle than time domain expressions. This therefore provides an effective tool to account for the

low frequency region of the spectra.

Methods

MD simulations

A 300 ns MD simulation of a poly-ALA peptide was performed and positions and velocities of the atoms were stored. Long MD simulations were performed, as the computation, of the ACFs required both high sampling rate (2 fs) of the trajectory in order to avoid aliasing (the highest expected frequencies are on the order of 250 ps^{-1}) and a long duration to achieve sufficient signal-to-noise ratio and good enough statistics in the low-frequency region of the velocity spectra. Details of the simulations are found in Table 2

Table 2: Setup conditions for the MD simulations of the polyala8 peptide

Parameter	Value
Cubic box side length / Å	30
Boundary conditions	periodic
Number of water molecules	845
Force field	CHARMM22 with CMAP correction for the peptides,[46] TIP3P for water
Ensemble	NpT
Thermostat	Langevin, $T = 298.15$ K
Barostat	Langevin, piston period 200 fs, piston decay 100 fs, $p = 1$ atm
Non-bonded interactions cut-off / Å	12.0
Electrostatics	particle mesh Ewald, order 6, tolerance 10^{-5}
Integration time step / fs[a]	2
Minimization steps	5000
Heating time / ps	72
Equilibration time / ns	2
Production time / ns	200

[a] All bonds with H atoms were constrained with the SHAKE algorithm.

Data processing and analysis

VACF computations and fittings were performed using homewritten scripts using the Scilab software.[47]

The trajectories of the velocities of individual atoms in the peptide were analyzed by splitting the trajectories in 16 k chunks. MD simulations were sampled every 2 fs, so that each of these spanned $\sim$ 32 ps. Velocity auto-correlation functions were computed for each of these extracts. Velocity ACFs are typically expected to decay very fast, so that these extracts of the MD trajectory are likely to sample processes that contain frequencies down to $\approx 30^{-1}$ ps^{-1} with satisfactory statistics.[48] The VACFs were then obtained by summing the functions obtained for each excerpt of the total trajectory. The associated velocity spectra were obtained for each atom through numerical Fast Fourier Transformation.

The VACF spectra were fitted based on the model described above using the differential evolution algorithm[49] implemented in the Scilab *Differential Evolution* Toolbox.

Supporting Information

Derivation of Eqs.1-8; derivation of the force autocorrelation function for a power law kernel.

Acknowledgement

Support from the Centre National de la Recherche scientifique is acknowledged (D.A.).

References

(1) Glöckle, W. G.; Nonnenmacher, T. F. A fractional calculus approach to self-similar protein dynamics. *Biophysical Journal* **1995**, *68*, 46–53.

(2) Lu, H.; Xun, L.; Xie, X. S. Single-Molecule Enzymatic Dynamics. *Science* **1998**, *282*, 1877.

(3) Kou, S. C.; Xie, X. S. Generalized Langevin Equation with Fractional Gaussian Noise: Subdiffusion within a Single Protein Molecule. *Physics Review Letters* **2004**, *93*, 180603(4).

(4) Metzler, R.; Klafter, J. The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Phys. Rep.* **2000**, *339*, 1–77.

(5) Min, W.; Luo, G.; Cherayil, B. J.; Kou, S. C.; Xie, X. S. Observation of a Power-Law Memory Kernel for Fluctuations within a Single Protein Molecule. *Physics Review Letters* **2005**, *94*, 198302.

(6) Hamon, V.; Calligari, P.; Hinsen, K.; Kneller, G. R. Simulation studies of structural changes and relaxation processes in lysozyme under pressure. *Journal of Non-Crystalline Solids* **2006**, *352*, 4417 – 4423.

(7) Calandrini, V.; Hamon, V.; Hinsen, K.; Calligari, P.; Bellissent-Funel, M.-C.; Kneller, G. *Chemical Physics* **2008**, *345*, 289 – 297.

(8) Calandrini, V.; Abergel, D.; Kneller, G. R. Fractional protein dynamics seen by nuclear magnetic resonance spectroscopy: Relating molecular dynamics simulation and experiment. *The Journal of Chemical Physics* **2010**, *133*, 145101.

(9) Calligari, P.; Calandrini, V.; Kneller, G. R.; Abergel, D. From NMR Relaxation to Fractional Brownian Dynamics in Proteins: Results from a Virtual Experiment. *The Journal of Physical Chemistry B* **2011**, *115*, 12370–12379.

(10) Calligari, P.; Abergel, D. Multiple scale Dynamics in Proteins Probed at Multiple Time Scales through Fluctuations of Chemical Shifts. *The Journal of Physical Chemistry B* **2014**, *118*, 3823–3831.

(11) Montroll, E. W. Random Walks in Multidimensional Spaces, Especially on Periodic Lattices. *SIAM Journal* **1953**, *4*, 241–260.

(12) Montroll, E. W.; Weiss, G. H. Random Walks on Lattices. II. *Journal of Mathematical Physics* **1965**, *6*, 167–181.

(13) Montroll, E. W. Random Walks on Lattices. III. Calculation of FirstPassage Times with Application to Exciton Trapping on Photosynthetic Units. *Journal of Mathematical Physics* **1969**, *10*, 753–765.

(14) Montroll, E. W.; Scher, H. Random Walks on Lattices. IV. Continuous-TimeWalks and Influence of Absorbing Boundaries. *Journal of Statistical Physics* **1973**, *9*, 101–135.

(15) Metzler, R.; Barkai, E.; Klafter, J. Anomalous diffusion and relaxation close to thermal equlibrium: A fractional Fokker-Planck equation approach. *Phys. Rev. Lett.* **1999**, *82*, 3563–3567.

(16) Metzler, R.; Barkai, E.; Klafter, J. Deriving fractional Fokker-Planck equations from a generalised master equation. *Europhysics Letters* **1999**, *46*, 431.

(17) Calligari, P.; Abergel, D. Toward the Characterization of Fractional Stochastic Processes Underlying Methyl Dynamics in Proteins. *The Journal of Physical Chemistry B* **2012**, *116*, 12955–12965, PMID: 23009081.

(18) Mori, H. Transport, Collective Motion, and Brownian Motion. *Progress of Theoretical Physics* **1965**, *33*, 423–455.

(19) Zwanzig, R. In *Statistical mechanics of irreversibility*; Brittin, W., Ed.; Lectures in Theoretical Physics; Wiley-Interscience: New York, 1961; pp 106–141.

(20) Boon, J.-P.; Yip, S. *Molecular Hydrodynamics*; McGraw-Hill, Inc., 1980.

(21) Haubold, H.; Mathai, A.; Saxena, R. Mittag-Leffler Functions and Their Applications. *Journal of Applied Mathematics* **2011**, 51 pages.

(22) *Higher Transcendantal Functions*; McGraw-Hill, 1955; Vol. 3; Chapter 18.

(23) Gorenflo, R.; Loutchko, J.; Luchko, Y. Computation of the Mittag-Leffler $E_{\alpha,\beta}$ Function and its Derivative. *Fractional Calculus and Applied Analysis* **2002**, *5*, 491–518.

(24) Hilfer, R.; Seybold, H. Computation of the generalized Mittag-Leffler function and its inverse in the complex plane. *Integral Transforms and Special Functions* **2006**, *17*, 637–652.

(25) Seybold, H.; Hilfer, R. Numerical Algorithm for the Computating the Generalized Mittag-Leffler Function. *SIAM J. Numer. Anal.* **2008**, *47*, 69–88.

(26) Mathai, A. M.; Haubold, H. J. *Special functions for applied scientists*; Springer, 2008; Vol. 4.

(27) Lamm, G.; Szabo, A. Langevin modes of macromolecules. *The Journal of Chemical Physics* **1986**, *85*, 7334–7348.

(28) Moritsugu, K.; Smith, J. C. Langevin Model of the Temperature and Hydration Dependence of Protein Vibrational Dynamics. *The Journal of Physical Chemistry B* **2005**, *109*, 12182–12194, PMID: 16852503.

(29) Hayward, S.; Kitao, A.; Hirata, F.; Go, N. Effect of Solvent on Collective Motions in Globular Protein. *Journal of Molecular Biology* **1993**, *234*, 1207–1217.

(30) Zwanzig, R. *Nonequilibrium Statistical Mechanics*; Oxford University Press, 2001.

(31) Lange, O. F.; Grubmüller, H. Collective Langevin dynamics of conformational motions in proteins. *The Journal of Chemical Physics* **2006**, *124*, 214903/1–18.

(32) Yu, M.; Tan, P.; Ye, Y.; Voneshen, D. J.; Xing, X.; Hong, L. One-dimensional nature of protein low-energy vibrations. *Phys. Rev. Res.* **2020**, *2*, 032050.

(33) Kubo, R. The fluctuation-dissipation theorem. *Reports on Progress in Physics* **1966**, *29*, 255.

(34) Attard, P. *Non-equilibrium Thermodynamics and Statistical Mechanics: Foundations and Applications*; Oxford University Press, 2012.

(35) Ayaz, C.; Tepper, L.; Brünig, F. N.; Kappler, J.; Daldrop, J. O.; Netz, R. R. Non-Markovian modeling of protein folding. *Proc Natl Acad Sci U S A.* **2021**, *118*, e2023856118.

(36) Hinsen, K. The Molecular Modeling Toolkit: A New Approach to Molecular Simulations. *J. Comput. Chem.* **2000**, *21*, 79–85.

(37) Case David, A.; Cheatham Thomas, E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J.Comput. Chem.* **2005**, *26*, 1668–1688.

(38) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Laxmikant, K.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

(39) Kasahara, K.; Sakuraba, S.; Fukuda, I. Enhanced Sampling of Molecular Dynamics Simulations of a Polyalanine Octapeptide: Effects of the Periodic Boundary Conditions on Peptide Conformation. *J. Phys. Chem. B* **2018**, *122*, 2495–2503.

(40) Graf, J.; Nguyen, P. H.; Stock, G.; Schwalbe, H. Structure and Dynamics of the Homologous Series of Alanine Peptides. A Joint Molecular Dynamics/NMR Study. *J. Am. Chem. Soc.* **2007**, *129*, 1179–1189.

(41) Granek, R.; Klafter, J. Fractons in Proteins: Can They Lead to Anomalously Decaying Time Autocorrelations? *Phys. Rev. Lett.* **2005**, *95*, 098106.

(42) Meroz, Y.; Ovchinnikov, V.; Karplus, M. Coexisting origins of subdiffusion in internal dynamics of proteins. *Phys. Rev. E* **2017**, *95*, 062403.

(43) Neusius, T.; Daidone, I.; Sokolov, I. M.; Smith, J. C. Subdiffusion in Peptides Originates from the Fractal-Like Structure of Configuration Space. *Phys. Rev. Lett.* **2008**, *100*, 188103.

(44) Xia, C.; He, X.; Wang, J.; Wang, W. Origin of subdiffusions in proteins: Insight from peptide systems. *Phys. Rev. E* **2020**, *102*, 062424.

(45) Dominic, A. J. I.; Cao, S.; Montoya-Castillo, A.; Huang, X. Memory Unlocks the Future of Biomolecular Dynamics: Transformative Tools to Uncover Physical Insights Accurately and Efficiently. *Journal of the American Chemical Society* **2023**, *145*, 9916–9927.

(46) Matthias, B.; Bouguet-Bonnet, S.; Pastor, R.; MacKerell, A. D. J. Importance of the CMAP Correction to the CHARMM22 Protein Force Field: Dynamics of Hen Lysozyme. *Biophys. J.* **2006**, *90*, 36–38.

(47) Scilab, a free scientific software package. Scilab Enterprises: www. scilab.org, 2011-2014.

(48) Zwanzig, R.; Ailawadi, N. K. Statistical Error Due to Finite Time Averaging in Computer Experiments. *Phys. Rev.* **1969**, *182*, 280–283.

(49) Storn, R.; Price, K. J. Differential Evolution - a Simple and Efficient Heuristic for Global Optimization over Continuous Spaces. *Journal of Global Optimization* **1997**, *11*.