



HAL
open science

Stochastic modeling of macromolecules in solution. I. Relaxation processes

Antonino Polimeno, Mirco Zerbetto, Daniel Abergel

► **To cite this version:**

Antonino Polimeno, Mirco Zerbetto, Daniel Abergel. Stochastic modeling of macromolecules in solution. I. Relaxation processes. *The Journal of Chemical Physics*, 2019, 150 (18), pp.184107. 10.1063/1.5077065 . hal-02414591

HAL Id: hal-02414591

<https://hal.science/hal-02414591>

Submitted on 16 Dec 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Stochastic modeling of macromolecules in solution. I. Relaxation processes

Antonino Polimeno,^{1, a)} Mirco Zerbetto,¹ and Daniel Abergel²

¹⁾*Dipartimento di Scienze Chimiche, Università degli Studi di Padova, I-35131, Padova, Italy*

²⁾*Laboratoire des biomolécules, LBM, Département de chimie, Ecole normale supérieure, PSL University, Sorbonne Université, CNRS, 75005 Paris, France*

(Dated: 16 December 2019)

A framework for the stochastic description of relaxation processes in flexible macromolecules, including dissipative effects, is introduced, from an atomistic point of view. Projection-operator techniques are employed to obtain multidimensional Fokker-Planck operators governing the relaxation of internal coordinates and global degrees of freedom, and depending upon parameters fully recoverable from classic force fields (energetics) and continuum models (friction tensors). A hierarchy of approaches of different complexity is proposed in this unified context, aimed primarily at the interpretation of magnetic resonance relaxation experiments. In particular, a model based on a harmonic internal Hamiltonian is discussed as a test case.

^{a)}Electronic mail: antonino.polimeno@unipd.it

I. INTRODUCTION

Internal and overall dynamics of proteins are involved in the determination and the regulation of their physical and chemical properties, biological functions and spectroscopic signatures. Examples of dynamic-controlled classes of processes are the allosteric effects in enzyme catalysis, the formation of non-specific transient encounter complexes in the protein-protein association¹ and the regulation of molecular recognition². Therefore, monitoring and describing protein dynamics is a fundamental area of investigation in modern physical chemistry. Internal and global motions in solution affect directly or indirectly most spectroscopic methods aimed at the characterization of proteins. This is the case, for instance, for Nuclear Magnetic Resonance (NMR) relaxation³⁻⁵, Fluorescence Anisotropy Decay⁶, Time Resolved X-Ray⁷, site-directed spin-labeled Electron Spin Resonance^{8,9}, Förster Fluorescence Resonance Energy Transfer¹⁰, Atomic Force Microscopy¹¹. Experimental observations need to be rationalized in order to provide a meaningful description of at least some of the many complex motions that combine to yield the observed relaxation processes. These include global reorientation and large amplitude motions of entire domains, as well as local readjustments and single-residue motions. In general, different spectroscopic techniques probe different physical observables and provide information on motions taking place at different time windows. It is convenient, when possible, to develop a computational interpretation adapted to a specific experimental approach, within a general theoretical framework. This concern is particularly relevant for the analysis of NMR relaxation experiments^{3,12} where the definition of such variables can be dictated either by ad hoc considerations, such as the model-free (MF) approach,^{13,14} or by explicit models of dynamics, as for instance in the Slowly Relaxing Local Structure (SRLS) model.^{15,16}

The description of the dynamics of a non-rigid molecule requires a precise definition of several frames of reference. This is necessary when one needs to account accurately for the tensorial nature of the magnetic interactions. To this aim, we shall use throughout this work the following frames, shown in Fig. 1: i) a laboratory frame (LF), i.e. a fixed frame, external to the molecule; ii) a molecular frame (MF), i.e. a frame attached to the molecule, which is usually model-dependent; iii) an interaction frame (μ F), i.e. a local frame linked to the MF where some specific second rank tensor defining the property μ observable spectroscopically is well-defined. Depending on the problem at hand, this could

be for instance the frame where the $^{15}\text{N} - ^1\text{H}$ dipolar or ^{15}N chemical shift (CSA) tensors are diagonal.^{17,18} The spectroscopic observable is written, usually, in terms of suitable time correlation functions or spectral densities (see below), and their fast and accurate evaluation is the main target of a well-defined dynamic modeling approach.

A. Modeling and monitoring molecular motions

Models with various degrees of simplification can be used to interpret experimental dynamical data in macromolecules, and proteins in particular. For instance, one can make the simplifying assumption that local motions are due, at least for semi-rigid systems, to a network of dynamically coupled neighbors (Network Model)^{19,20} or caused by partial diffusive reorientation within a local potential.¹⁶ One can also assume specific statistical characteristics (diffusive or Brownian dynamics, fractional Brownian dynamics²¹ etc.). In general, two alternative points of view, which are apparently not easily reconciled, are adopted. On one side, one can resort to full molecular dynamics simulations, trying to describe a priori the molecular object in its environment and attempting a direct prediction, from the collected trajectory, of the desired relaxation parameters. On the other side, coarse-grained modeling approaches, based on phenomenological hypotheses, can be employed. Both methods are familiar to computational chemists and are used often in combination to interpret NMR relaxation data.

Classical molecular dynamics (MD) simulations are nowadays the most powerful tool available to simulate relaxation processes in molecules. Recent computational developments now make it possible to carry on long simulations in the time scales of microseconds and even milliseconds,²²⁻²⁶ which is compatible with the study of slow processes such as large amplitude motions and folding processes. However, some considerations are in order. Although MD brute force approaches are increasingly a natural choice, even with the rapid development of computing power, the availability of sufficient computational resources to collect long trajectories remains limited. Besides, some issues have important consequences in the context of NMR relaxation. Firstly, sampling the MD trajectories can be tricky: correct statistics is based on ergodicity, which may be difficult to ensure in MD simulations,²⁷ leading to the need to calculate a number of trajectories, each of which starting from a chosen different initial configuration of the system. Such a choice, however, requires the knowledge

of the free energy in the whole phase space of atomic coordinates or internal coordinates in order to be done systematically. This information is usually not known, especially in complex systems. The limitations of available force-fields should always be taken into account, especially when long (hundred of nanoseconds or longer) trajectories are employed. Secondly, many of the commonly employed parameterizations for atomic interactions in biological systems do not perform well at long times (i.e., those required to access NMR relaxation data), although work is being carried out to correct the MD force-fields just using NMR data as reference.²⁷⁻²⁹ Finally, recent investigations show that hydrodynamic properties, typically affected by the accuracy of force fields and statistics can be simulated via MD with some difficulty.^{30,31} Phenomenological stochastic models offer an alternative for interpreting NMR relaxation in flexible molecules. Moreover, they provide a way to rationalize complex observables, allowing to establish a hierarchy of causes and effects, via the definition of a coarse-grained description of the systems investigated. The adoption of stochastic models in interpreting NMR relaxation data of complex systems dates back to 1982, with the introduction of the MF approach, first applied to globular proteins.^{13,14} In MF, analytic expressions for orientational correlation functions (from which spectral densities are obtained and consequently NMR relaxation data extracted) are introduced, describing the molecular relaxation as the combination of two motions: the global tumbling of the whole molecule and the effective, unspecified local motion of the $^{15}\text{N} - ^1\text{H}$ or $^{13}\text{C} - ^1\text{H}$ bond. Correlation functions take a bi-exponential time dependence, which is in many cases sufficient to fit NMR data. Since the first introduction of MF, different extensions of such an approach have been presented, each of which had the effect of increasing the number of exponential terms that contribute to the correlation function, by calling multiple timescale separation limits, thereby obscuring the physical meaning of the parameters outside the range of applicability of the approximations.

The SRLS model,^{32,33} developed in the early 90's, allows for a more accurate description of the geometry of the local motions. In this approach, the relevant dynamics of the molecule is still given by the combination of the global tumbling and the local $^{15}\text{N} - ^1\text{H} / ^{13}\text{C} - ^1\text{H}$ bond motion, but the assumption of timescale separation is not a prerequisite. Rather, decoupling occurs naturally if global and internal motions show timescale separation at least of 3 - 4 orders of magnitude, making the correlation function be equal to the MF one (given the same choice for the local geometry),³⁴ while the exact solution is obtained in all the other

cases. In the SRLS approach, molecular parameters (potential of mean force and diffusion tensor, see below) always maintain their clear physical meaning. Both the SRLS and MF approaches describe the relevant dynamics of the molecule as that of two diffusive stochastic rotors, one describing the rotational motion of the whole molecule, and the second collecting the local motions in which the probe (e.g., the amino-acid in a protein) is involved. The nature of the internal motion, which is not easily mapped on to a specific set of internal generalized coordinates, poses difficulties in the definition of a multi-scale approach aimed at the *ab initio* solution of the stochastic equation. Extensions to the SRLS-NMR have been introduced both for proteins³⁵ and small molecules³⁶ with some parameters, the most important of which is the diffusion tensor associated to the internal motion, evaluated via a multi-component fitting procedure.

A significant contribution toward the comprehension of the complex dynamics of flexible macromolecules is due to Perico and Guenza^{37,38}. By adopting a diffusive Gaussian model these authors calculate in an exact manner first and second rank orientation correlation functions, and demonstrate that a multiexponential decays is predicted unless local motion and global reorientational tumbling are separated in time. The model is especially relevant to discuss the crossover from flexible to semirigid chains, and it has been applied, for instance, to the analysis of the dynamics of polymers.³⁹

The Reorientational Eigenmode Dynamics (RED) approach proposed by Prompers and Brüschweiler makes use of (short) MD simulations to calculate NMR data.⁴⁰⁻⁴² The method is based on the factorization of the time correlation functions into global and internal, as in the MF method. However, the internal part of the correlation function takes into account the coupling among selected internal degrees of freedom, i.e. the orientations on the protein of the NMR probes. In particular, the rank 2 spherical harmonics, functions of the orientation of the NMR probes (e.g., N-H bonds) are used to build a variance-covariance matrix, which is then diagonalized, thereby providing a set of reorientational eigenmodes. The time series of the rank 2 spherical harmonics of each probe are then projected over the eigenmodes, and these projections are employed to access the proper correlation functions. The global tumbling is then added *ad hoc*, based on the overall rotational symmetry of the protein. Several other approaches rely on stochastic modeling of the dynamics. We recall here the Gaussian Network Model (GNM), the Anisotropic Network Model (ANM), the Normal Mode Analysis (NMA) and the LE4PD method. In the GNM the molecule is seen

as an ensemble of effective atoms (beads) each having a simple harmonic interaction with other beads within a chosen cutoff radius.⁴³ The GNM has been applied with some success in a number of instances, such as the interpretation of the Debye-Waller factors measured by X-ray crystallography of globular proteins (bovine pancreatic trypsin inhibitor⁴⁴). The GNM has also been adapted to the study of NMR relaxation data⁴⁵ by considering the motion of the N-H bond decoupled from the global tumbling of the molecule and assuming a MF approach. The ANM was introduced as an extension of the GNM to include anisotropy explicitly into the fluctuations, and is based on the diagonalization of the Hessian of the potential energy^{46,47} obtained from a molecular mechanics force field. The NMA describes the (internal) dynamics of the protein as a superposition of normal modes^{48,49} calculated theoretically from the Hessian matrix of the minimum energy conformation. The difference with respect to ANM resides in the shape of the potential. In the ANM approach the potential energy is explicitly approximated by a multidimensional parabola and the minimum energy configuration is assumed to be known. Alternatively, in the NMA analysis, Monte Carlo or MD simulations are performed beforehand in order to explore the potential energy surface and determine the most populated conformation⁵⁰ of the molecule. Then the Hessian matrix is calculated numerically, and the energy surface approximated to a multidimensional parabola around the reference configuration and the rest of the analysis is conducted as for ANM. The LE4PD^{51,52} method is also based on a Langevin equation, does not assume harmonic potentials, and takes into account the presence of energy barriers and more complex energy profiles, such as in crankshaft or local torsional motions. This approach can be linked formally to an initial Liouville equation, using a Mori-Zwanzig projection technique. Hydrodynamic interactions are accounted for by introducing site specific friction coefficients, which account for the screening of the hydrodynamic interaction inside the hydrophobic core of the protein, are derived from the Oseen tensor. The degree of exposure to the solvent of each amino acid, which enters the definition of the hydrodynamic radius, is calculated from the available surface area, which is measured at each step of the whole simulation trajectory under study.

B. Motivation and outline of the paper

We propose a description starting from the exact dynamics of a non-rigid protein interacting with its solvent. A projection technique, which averages out the degrees of freedom (dof) of the 'bath' (the irrelevant dof, when one is interested in a small number of protein atoms), leads to the reduced Fokker-Planck equation for the relevant atoms, naturally including hydrodynamic properties.

In order to combine the advantages of the semi-phenomenological methods described above, a general strategy for the treatment of the dynamics must include a systematic view of the system geometry, based on the molecular structure, as well as of the associated dynamical features. In this perspective, we propose to treat explicitly the molecule as a non-rigid body, initially described by a classical Liouville operator, and to obtain a Markovian stochastic differential or master equation, with a relevant set of degrees of freedom as dynamical variables. In order to do so, we proceed as follows. First, we set up the Liouville equation of motion⁵³ for a generic flexible body defined as a set of material points (atoms or extended atoms), in terms of roto-translational and natural internal coordinates. The general description of a macromolecule in solution is then carried out in terms of a collection of flexible bodies, to which a standard Nakajima-Zwanzig^{54,55} projection method is applied in order to eliminate the 'irrelevant', i.e. not directly observed, degrees of freedom. A generalized master equation is thus obtained from which alternative modeling options, based on different choices of the internal variables, of accuracy levels of the description of solvents effects etc., can be envisaged. In particular, in the case of fully averaged irrelevant coordinates (hydrodynamic solvent, Brownian model etc.) different description levels of the macromolecule are discussed. In such a limit, the partition of internal degrees of freedom into 'hard' fast relaxing constrained modes, and 'soft' ones - i.e. medium and large-amplitude motions - is discussed. We analyze the simplest case of a partially rigid Brownian system, i.e. with only fast relaxing internal modes.

The paper is organized as follows. Section II introduces the classic Liouville description of a flexible macromolecule for the probability density of all degrees of freedom. Section III derives, using a standard Zwanzig projection approach, a general master equation, which is simplified, under suitable approximation to the case of a semi-flexible rotator in Section IV. The proposed parametrization is tested in Section V on a series of homologous molecules, the

polyalanines (ALA-) $_n$, with $n = 2, \dots, 10$. A brief discussion is included in Section VI, while the actual computational treatment of the resulting model is presented in the accompanying paper.⁵⁶ We confine in the Appendices some of the more cumbersome derivations.

II. MODELING

A. Observables

In the case of nuclear magnetic resonance the experimental observables are related to the normalized autocorrelation function of 2-nd rank Wigner matrix elements for each interaction frame. In general we can write the normalized autocorrelation function as:

$$G_\mu(t) = \frac{\langle \mathcal{D}_{0,0}^2[\Omega_\mu(t)]^* \mathcal{D}_{0,0}^2[\Omega_\mu(0)] \rangle}{\langle \mathcal{D}_{0,0}^2(\Omega_\mu)^* \mathcal{D}_{0,0}^2(\Omega_\mu) \rangle}, \quad (1)$$

where Ω_μ is the set of Euler angles defining the orientations of the μ interaction frame μF with respect to the LF (cf Figure 1). The average $\langle \dots \rangle$ is defined with respect to the phase space of the system, in which the evolution of the phase density is dictated by the Markovian time evolution operator $\hat{\Gamma}$, associated to the equilibrium phase density ρ . The phase space includes all coordinates and momenta. $\hat{\Gamma}$ and ρ define the model and in principle they can be the complete Liouville operator (see next section) with respect to which ρ can also be defined (for instance as a Boltzmann distribution in a canonical ensemble). Therefore, assuming that the system is isotropic, and taking into account the geometry defined in Figure 1:

$$G_\mu(t) = 5 \langle \mathcal{D}_{0,0}^2(\Omega_\mu)^* | \exp(-\hat{\Gamma}t) | \mathcal{D}_{0,0}^2(\Omega_\mu) \rho \rangle \quad (2)$$

Here, Ω_μ is interpreted as a function of the phase space coordinates, and is deduced from the system geometry. The factor 5 arises from the normalization condition in an isotropic medium. It is useful to separate the contributions to the spectral density function that originate from molecular motions, on the one hand, from those that come from the geometrical features of the interactions involved, on the other hand. This is achieved using standard algebra⁵⁷ of Wigner matrices. Denoting Ω the Euler angles describing the orientation of the MF with respect to the LF (cf Figure 1), and $\bar{\Omega}_\mu$ the tilt of μF with respect to the MF, the

TABLE I. Table of main mathematical symbols

Symbol	Meaning
Ω_μ	Euler angles of the μ interaction frame μF with respect to the LF
Ω	Euler angles of the MF with respect to the LF
$\bar{\Omega}_\mu$	Euler angle of μF with respect to the MF
n	number of atoms
$\mathbf{r}_\alpha, \mathbf{v}_\alpha, \mathbf{p}_\alpha, M_\alpha$	Cartesian position, velocity, momentum and mass of α -th atom
H, K, U, L	Hamiltonian, kinetic energy, potential energy, Lagrangian
$\mathbf{R}, \mathbf{V}, \mathbf{P}, M,$	position, velocity, momentum of center of mass and total mass
\mathbf{L}	angular momentum
\mathbf{c}_α	position vectors relative to the center of mass (MF)
q^μ	μ -th internal variable
p_μ	μ -th conjugate momentum
\mathbf{I}	inertia tensor
\mathbf{A}_μ	gauge potential
$g_{\mu\nu}, g^{\mu\nu}$	metric tensor components
\mathcal{X}, \mathcal{P}	coordinate and momentum space
\mathcal{Q}	phase space \mathcal{X}, \mathcal{P}
$\mathcal{X}_I, \mathcal{P}_I$	coordinate space and momentum space, expressed in internal coordinates and momenta, for a set of external molecules
\mathcal{Q}_I	phase space $\mathcal{X}_I, \mathcal{P}_I$
$\rho(\mathcal{Q}, \mathcal{Q}_I, t)$	exact phase space density at time t
Γ_T	Liouville operator governing $\rho(\mathcal{Q}, \mathcal{Q}_I, t)$
$\bar{\rho}(\mathcal{Q}, t)$	average phase space density at time t after projection of \mathcal{Q}_I
Γ_{FP}	Fokker-Planck time evolution operator for $\bar{\rho}(\mathcal{Q}, t)$
$\bar{\rho}(\mathcal{Q})$	equilibrium distribution
ξ	friction tensor
F	free energy
$\mathbf{q} = (\mathbf{q}_s, \mathbf{q}_h), \mathbf{p} = (\mathbf{p}_s, \mathbf{p}_h)$	harmonic (h) and non-harmonic (s) internal coordinates and momenta
\mathbf{K}	curvature matrix
\mathbf{y}	non-orientation degrees of freedom ($\mathbf{q}, \mathbf{L}, \mathbf{p}$) in the SFB model (no non-harmonic modes)
\mathbf{k}	curvature matrix for the total Hamiltonian in the SFB model
\mathbf{x}	scaled and rotated coordinates $(k_B T)^{-1/2} \mathbf{S}(\mathbf{y} - \mathbf{y}^{(0)})$
Γ	SFB time evolution operator
$p(\mathbf{x})$	equilibrium distribution in the SFB model (non-orientation dof only)
ω_T	dissipation matrix of Γ
ω_K, ω_ξ	scaled curvature matrix and friction tensor
$\omega^{\text{io}}, \omega^{\text{int}}$	internal overall and interaction blocks of dissipation matrix ω_T

measurable spectral density for auto-correlated relaxation is given by:

$$G_\mu(t) = \sum_{l,\bar{l}=-2}^2 \mathcal{D}_{l,0}^2(\bar{\Omega}_\mu)^* \mathcal{D}_{\bar{l},0}^2(\bar{\Omega}_\mu) G_{l\bar{l}}(t) \quad (3)$$

where

$$G_{l\bar{l}}(t) = 5 \langle \mathcal{D}_{0,l}^2(\Omega)^* | \exp(-\hat{\Gamma}t) | \mathcal{D}_{0,\bar{l}}^2(\Omega) \rho \rangle \quad (4)$$

In practice, one is currently interested in spectral densities, i.e. the Fourier-Laplace transforms of $G_\mu(t)$ and $G_{l\bar{l}}(t)$. Our main objective will therefore be the evaluation of $J_\mu(\omega)$, $J_{l\bar{l}}(\omega)$, defined as:

$$J_\mu(\omega) = \sum_{l,\bar{l}=-2}^2 \mathcal{D}_{l,0}^2(\bar{\Omega}_\mu)^* \mathcal{D}_{\bar{l},0}^2(\bar{\Omega}_\mu) J_{l\bar{l}}(\omega) \quad (5)$$

$$J_{l\bar{l}}(\omega) = \int_0^\infty dt e^{-i\omega t} G_{l\bar{l}}(t) \quad (6)$$

Evaluating Eq. 6 constitutes the main object of our study, which requires to define a suitable model for $\hat{\Gamma}$ and ρ . In the remainder of this article, we analyze several possible models that we derive from various levels of approximation. Additionally, a specific model is also defined, which we discuss in full in the companion paper.⁵⁶ For sake of clarity, we summarize in Table I the meaning of some of the symbols defined in the following Sections.

B. Liouville equation

We begin by reviewing the derivation of the (classical) Liouville equation for a flexible body. The treatment of non-rigid molecules is a well-known subject for students of vibrational spectroscopy. The literature on the subject is vast, as for instance in⁵⁸⁻⁶³ to cite but a few, and based on the methods of classical mechanics.^{53,64} Here, we shall in particular follow closely the approach of Meyer,⁶⁵ Lauvergnat,⁶⁶ Littlejohn and Reinsch.⁶⁷ Although the procedure is known and well-established, we feel that a summary of the main points of the basic procedure can be helpful to the reader unfamiliar with these tools, and makes this report self-contained. The starting point is the description of the molecule (flexible body) as a collection of n material points, with Cartesian coordinates \mathbf{r}_α , velocities $\mathbf{v}_\alpha = \dot{\mathbf{r}}_\alpha$ and masses M_α with $\alpha = 1, \dots, n$. We define the Lagrangian function $L = K - U = \sum_\alpha M_\alpha v_\alpha^2/2 - U(\mathbf{r})$,

the momenta $\mathbf{p}_\alpha = \partial L / \partial \mathbf{v}_\alpha = M_\alpha \mathbf{v}_\alpha$ and the Hamiltonian function $H = K + U = \sum_\alpha p_\alpha^2 / 2M_\alpha + U(\mathbf{r})$, with the notation $\mathbf{r} = (\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_\alpha, \dots, \mathbf{r}_n) \equiv (r^1, r^2, \dots, r^i, \dots, r^{3n})$, where the latter notation makes clear the contravariant nature of the $3n$ -dimensional vector of all coordinates \mathbf{r} . Analogously $\mathbf{p} = (\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_\alpha, \dots, \mathbf{p}_n) \equiv (p_1, p_2, \dots, p_i, \dots, p_{3n})$; in the following we employ Einstein's summation convention of repeated indices whenever tensorial properties need to be evidenced, e.g. $u_i v^i \equiv \sum_i u_i v^i$, or a more compact matrix vector notation, as in $\mathbf{u}^{\text{tr}} \mathbf{v} \equiv \sum_i u_i v_i$, otherwise.

The Liouville equation for the density $\rho(\mathbf{r}, \mathbf{p}, t)$ is written as

$$\frac{\partial \rho}{\partial t} = \{H, \rho\} \quad (7)$$

where the Poisson brackets of two phase variables A and B is given by:

$$\{A, B\} = \frac{\partial A}{\partial r^i} \frac{\partial B}{\partial p_i} - \frac{\partial A}{\partial p_i} \frac{\partial B}{\partial r^i} \quad (8)$$

We now define a transformed set of variables, in analogy with similar treatments employed in vibrational spectroscopy. We first choose the six roto-translational degrees of freedom, namely the center of mass \mathbf{R} defined by $M\mathbf{R} = \sum_\alpha M_\alpha \mathbf{r}_{\text{LF}\alpha}$ where $M = \sum_\alpha M_\alpha$ is the total mass. The three Euler angles denoted by Ω describing the orientation of the MF with respect to the LF. Moreover, we choose $3n - 6$ *internal* coordinates q^μ ($\mu = 1, \dots, 3n - 6$). The specific definitions of Ω and \mathbf{q} as functions of the native Cartesian coordinates can be arbitrarily chosen, provided that proper invariance conditions are considered.⁶⁷ A possible choice can be based on the Casimir-Eckart conditions.⁶³ Here we shall only assume that these functions are known in terms of Cartesian coordinates and write:

$$\mathbf{r}_\alpha = \mathbf{R} + \mathbf{E}^{\text{tr}}(\Omega) \mathbf{c}_\alpha(\mathbf{q}) \quad (9)$$

where $\mathbf{E}^{\text{tr}}(\Omega)$ is the transpose of the Euler matrix⁵⁷, and rotates the components of a generic vector from the MF to the LF. In this expression, the vectors \mathbf{r}_α and \mathbf{R} are related through $\mathbf{c}_\alpha(\mathbf{q}) = \mathbf{E}(\Omega)(\mathbf{r}_\alpha - \mathbf{R})$, which represent the relative position vectors of the particles with

respect to the center of mass in the LF. We define $\mathbf{V} = \dot{\mathbf{R}}$, and

$$\mathbf{v}_\alpha = \mathbf{V} + \mathbf{E}^{\text{tr}} \left(\boldsymbol{\omega} \times \mathbf{c}_\alpha + \frac{\partial \mathbf{c}_\alpha}{\partial q^\mu} \dot{q}^\mu \right) \quad (10)$$

where the angular velocity is defined from the antisymmetric matrix $\boldsymbol{\omega}^\times = \overline{\mathbf{E}\dot{\mathbf{E}}^{\text{tr}}}$. Therefore, using the \mathbf{c}_α representation, and the property $\sum_{\alpha=1}^n M_\alpha \mathbf{c}_\alpha = 0$, the total angular momentum $\mathbf{L}_T = \sum_\alpha M_\alpha \mathbf{r}_\alpha \times \mathbf{v}_\alpha$ breaks into the sum of a center of mass, \mathbf{L}_{ext} , and about the center of mass, contributions:

$$\mathbf{L}_T = \mathbf{L}_{\text{ext}} + \mathbf{E}^{\text{tr}} \mathbf{L} = M \mathbf{R} \times \mathbf{V} + \mathbf{E}^{\text{tr}} \left[\sum_\alpha M_\alpha \mathbf{c}_\alpha \times \left(\boldsymbol{\omega} \times \mathbf{c}_\alpha + \frac{\partial \mathbf{c}_\alpha}{\partial q^\mu} \dot{q}^\mu \right) \right] \quad (11)$$

Defining the inertia tensor $\mathbf{I} = \sum_\alpha M_\alpha (c_\alpha^2 \mathbf{1}_3 - \mathbf{c}_\alpha \mathbf{c}_\alpha^{\text{tr}})$, and the gauge potential

$$\mathbf{A}_\mu = \mathbf{I}^{-1} \left(\sum_\alpha M_\alpha \mathbf{c}_\alpha \times \frac{\partial \mathbf{c}_\alpha}{\partial q^\mu} \right) \quad (12)$$

the angular momentum about the center of mass with component in the MF is:^{53,64,67}

$$\mathbf{L} = \mathbf{I}(\boldsymbol{\omega} + \mathbf{A}_\mu \dot{q}^\mu) \quad (13)$$

The Lagrangian is obtained in the form:^{53,64,67}

$$L = \frac{1}{2} M V^2 + \frac{1}{2} \mathbf{L}^{\text{tr}} \mathbf{I}^{-1} \mathbf{L} + \frac{1}{2} g_{\mu\nu} \dot{q}^\mu \dot{q}^\nu - U \quad (14)$$

where the internal energy is left here undefined: for an isolated molecule, one can assume it to depend only on the internal (shape) coordinates. The metric tensor is defined:⁶⁷

$$g_{\mu\nu} = \sum_\alpha M_\alpha \frac{\partial \mathbf{c}_\alpha}{\partial q^\mu} \cdot \frac{\partial \mathbf{c}_\alpha}{\partial q^\nu} - \mathbf{A}_\mu^{\text{tr}} \mathbf{I} \cdot \mathbf{A}_\nu \quad (15)$$

We now define the conjugate momenta $\mathbf{P} = \hat{\nabla}_{\mathbf{V}} L = M \mathbf{V}$, $\mathbf{L} = \hat{\nabla}_{\boldsymbol{\omega}} L$, $p_\mu = \partial L / \partial \dot{q}^\mu$. Together with \mathbf{R} , Ω and q^μ they form a complete set of phase space coordinates, with

respect to which the Liouville equation for $\rho(\mathbf{R}, \Omega, \mathbf{q}, \mathbf{P}, \mathbf{L}, \mathbf{p}, t)$ can now be written:⁶⁷

$$\frac{\partial \rho}{\partial t} = \{H, \rho\} \quad (16)$$

$$H = \frac{1}{2M} P^2 + \frac{1}{2} \mathbf{L}^{\text{tr}} \mathbf{I}^{-1} \mathbf{L} + \frac{1}{2} g^{\mu\nu} (p_\mu - \mathbf{L}^{\text{tr}} \mathbf{A}_\mu) (p_\nu - \mathbf{L}^{\text{tr}} \mathbf{A}_\nu) + U \quad (17)$$

With these definitions, the Poisson brackets take the form:

$$\begin{aligned} \{A, B\} = & (\hat{\nabla}_{\mathbf{R}} A)^{\text{tr}} \hat{\nabla}_{\mathbf{P}} B - (\hat{\nabla}_{\mathbf{P}} A)^{\text{tr}} \hat{\nabla}_{\mathbf{R}} B + (\hat{\mathbf{M}} A)^{\text{tr}} \hat{\nabla}_{\mathbf{L}} B \\ & - (\hat{\nabla}_{\mathbf{L}} A)^{\text{tr}} (\hat{\mathbf{M}} B) - \mathbf{L}^{\text{tr}} (\hat{\nabla}_{\mathbf{L}} A \times \hat{\nabla}_{\mathbf{L}} B) + \frac{\partial A}{\partial q^\mu} \frac{\partial B}{\partial p_\mu} - \frac{\partial A}{\partial p_\mu} \frac{\partial B}{\partial q^\mu} \end{aligned} \quad (18)$$

The contravariant tensor $g^{\mu\nu}$ ($g_{\mu\lambda} g^{\lambda\nu} = \delta_\mu^\nu$) is

$$g^{\mu\nu} = \sum_\alpha \frac{1}{M_\alpha} (\hat{\nabla}_{\mathbf{r}_\alpha} q^\mu)^{\text{tr}} \hat{\nabla}_{\mathbf{r}_\alpha} q^\nu \quad (19)$$

Finally $\hat{\mathbf{M}}$ is the generator of infinitesimal rotations, with components in the MF⁵⁷. Eq. 16 is completely defined once the following quantities are known, all functions of \mathbf{q} : \mathbf{I} , \mathbf{A}_μ , $g^{\mu\nu}$ (or $g_{\mu\nu}$), U . Eq. 16 can be written in a more compact form by taking into account the symplectic notation of Hamilton equations⁵³. Thus, introducing the notation $\mathcal{Q} = (\mathbf{R}, \Omega, \mathbf{q}, \mathbf{P}, \mathbf{L}, \mathbf{p}) \equiv (\mathcal{X}, \mathcal{P})$ (where the variables \mathcal{X} and \mathcal{P} account for all the configuration coordinates and momenta) we can write:

$$\frac{\partial \rho(\mathcal{Q}, t)}{\partial t} = -\hat{\Gamma} \rho(\mathcal{Q}, t) = (\hat{\nabla}_{\mathcal{Q}})^{\text{tr}} \mathbf{J} \rho(\mathcal{Q}) \hat{\nabla}_{\mathcal{Q}} \rho(\mathcal{Q})^{-1} \rho(\mathcal{Q}, t) \quad (20)$$

Here $\rho(\mathcal{Q}) = \exp(-H/k_B T) / \langle \exp(-H/k_B T) \rangle$ is the Boltzmann distribution and the notation $\langle f(\mathcal{Q}) \rangle = \int d\mathcal{Q} f(\mathcal{Q})$ is introduced. The generalized gradient and the $6n \times 6n$ matrix \mathbf{J} are defined as:

$$\hat{\nabla}_{\mathcal{Q}} = \begin{pmatrix} \hat{\nabla}_{\mathcal{X}} \\ \hat{\nabla}_{\mathcal{P}} \end{pmatrix} = \begin{pmatrix} \hat{\nabla}_{\mathbf{R}} \\ \hat{\mathbf{M}} \\ \hat{\nabla}_{\mathbf{q}} \\ \hat{\nabla}_{\mathbf{P}} \\ \hat{\nabla}_{\mathbf{L}} \\ \hat{\nabla}_{\mathbf{p}} \end{pmatrix} \quad \mathbf{J} = k_B T \begin{pmatrix} \mathbf{0} & -\mathbf{1} \\ \mathbf{1} & \mathbf{j} \end{pmatrix} \quad (21)$$

where the $3n \times 3n$ matrix \mathbf{j} is

$$\mathbf{j} = \begin{pmatrix} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{L}^\times & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix} \quad (22)$$

The 3×3 matrix \mathbf{L}^\times has elements defined in terms of the Levi-Civita tensor, $(\mathbf{L}^\times)_{\alpha\beta} = \sum_\gamma \epsilon_{\alpha\beta\gamma} L_\gamma$, i.e. for a generic vector \mathbf{a} , $\mathbf{L}^\times \mathbf{a} = \mathbf{L} \times \mathbf{a}$.

III. MASTER AND FOKKER-PLANCK EQUATIONS

In this section, we review the derivation of the master equation, as obtained by a Zwanzig projection, in the case where the flexible molecule is explicitly represented as an ensemble of interacting particles, as defined in the previous section. The corresponding Fokker-Planck⁶⁸ limit will be also introduced.

We first consider the case of a single flexible molecule (probe) interacting with a set of other N_I flexible molecules. Note that the set of additional molecules does not necessarily represent the 'solvent' or the 'bath', and as a matter of fact, some or all of the N_I objects may equally well be subsets of atoms of a large macromolecule or components of a supra-molecular arrangement. However, irrespective of their physical nature, they are considered coupled to the probe, which is assumed to be the only moiety that is targeted by the spectroscopic measurement, i.e. all or part of the molecule that is sensed by the spectroscopic apparatus. Therefore, they are irrelevant as far as the experimental detection is concerned (hence the suffix 'I'). Each element of the set is defined by its own MF, internal coordinates, total mass, moment of inertia, gauge potential, metric tensor and so on. To keep things reasonably tidy we define \mathcal{Q} as above, for the probe, and the corresponding set of coordinates and momenta for the other molecules $\mathcal{Q}_I = (\mathcal{Q}_{I1}, \dots, \mathcal{Q}_{IN_I}) = (\mathcal{X}_I, \mathcal{P}_I) = (\mathcal{X}_{I1}, \dots, \mathcal{X}_{IN_I}, \mathcal{P}_{I1}, \dots, \mathcal{P}_{IN_I})$, i.e. \mathcal{X}_I is the collection of all configuration coordinates of the irrelevant molecules and so on. The system Hamiltonian is written straightforwardly, in analogy with Eq. 17 as the sum of the kinetic energy, and the potential energy $U(\mathcal{X}, \mathcal{X}_I)$. The N_I terms of the irrelevant particles are collected in the irrelevant kinetic energy term $K_I(\mathcal{Q}_I)$, whilst the probe kinetic energy is denoted as $K(\mathcal{Q})$. Finally, the potential energy term is a function of all the

configuration coordinates. Thus, the total Hamiltonian H_T is partitioned as:

$$H_T(\mathcal{Q}, \mathcal{Q}_I) = K(\mathcal{Q}) + K_I(\mathcal{Q}_I) + U(\mathcal{X}, \mathcal{X}_I) \quad (23)$$

We do not for now further specify $U(\mathcal{X}, \mathcal{X}_I)$, which can be thought, as usual, as the sum of two-bodies terms that can be further partitioned between the probe internal energy, the internal energy of the set of irrelevant configuration variables \mathcal{X}_I , and the interaction potential energy between the probe and the irrelevant variables.

The total density $\rho_T(\mathcal{Q}, \mathcal{Q}_I, t)$ is now governed by expressions equivalent to Eq. 20. Defining the equilibrium density $\rho_T(\mathcal{Q}, \mathcal{Q}_I) = \exp(-H_T/k_B T) / \langle \exp(-H_T/k_B T) \rangle_T$, and the average of a phase variable $f(\mathcal{Q}, \mathcal{Q}_I)$ as: $\langle f(\mathcal{Q}, \mathcal{Q}_I) \rangle_T = \langle \langle f(\mathcal{Q}, \mathcal{Q}_I) \rangle_I \rangle = \int d\mathcal{Q} \int d\mathcal{Q}_I f(\mathcal{Q}, \mathcal{Q}_I)$, one has

$$\frac{\partial \rho_T(\mathcal{Q}, \mathcal{Q}_I, t)}{\partial t} = -\hat{\Gamma}_T \rho_T(\mathcal{Q}, \mathcal{Q}_I, t) \quad (24)$$

where

$$\begin{aligned} \hat{\Gamma}_T = \hat{\Gamma} + \hat{\Gamma}_I = \hat{\Gamma} + \sum_{\kappa=1}^{N_I} \hat{\Gamma}_{I\kappa} = & -(\hat{\nabla}_{\mathcal{Q}})^{\text{tr}} \mathbf{J} \rho_T(\mathcal{Q}, \mathcal{Q}_I) \hat{\nabla}_{\mathcal{Q}} \rho_T(\mathcal{Q}, \mathcal{Q}_I)^{-1} \\ & - \sum_{\kappa=1}^{N_I} (\hat{\nabla}_{\mathcal{Q}_{I\kappa}})^{\text{tr}} \mathbf{J}_{I\kappa} \rho_T(\mathcal{Q}, \mathcal{Q}_I) \cdot \hat{\nabla}_{\mathcal{Q}_{I\kappa}} \rho_T(\mathcal{Q}, \mathcal{Q}_I)^{-1} \end{aligned} \quad (25)$$

Quantities $\hat{\nabla}_{\mathcal{Q}_{I\kappa}}$ and $\mathbf{J}_{I\kappa}$ are defined by analogy with $\hat{\nabla}_{\mathcal{Q}}$ and \mathbf{J} .

We are now in a position to apply a standard projection approach with respect to the irrelevant coordinates \mathcal{Q}_I . The well-known procedure^{54,55} is summarized in Appendix A, in a form adapted to the case under investigation, so that only the main points are underlined in the following. First, we define the averaged equilibrium and time dependent probability $\bar{\rho}(\mathcal{Q})$, $\bar{\rho}(\mathcal{Q}, t)$ and the conditional equilibrium probability $\rho(\mathcal{Q}_I|\mathcal{Q})$:

$$\begin{aligned} \bar{\rho}(\mathcal{Q}) &= \langle \rho_T(\mathcal{Q}, \mathcal{Q}_I) \rangle_I \\ \bar{\rho}(\mathcal{Q}, t) &= \langle \rho_T(\mathcal{Q}, \mathcal{Q}_I, t) \rangle_I \\ \rho(\mathcal{Q}_I|\mathcal{Q}) &= \rho_T(\mathcal{Q}, \mathcal{Q}_I) / \bar{\rho}(\mathcal{Q}) \end{aligned} \quad (26)$$

and the projection operator \hat{P} acting on a phase variable $f(\mathcal{Q}, \mathcal{Q}_I)$ is:

$$\hat{P} \dots = \rho(\mathcal{Q}_I | \mathcal{Q}) \langle \dots \rangle_I \quad (27)$$

We assume, keeping in line with most analogous treatments^{54,55} - and in agreement, as we shall see below, with the requirements for calculating correlation functions of the type given in Eq. 6 - that the probe is initially statistically uncorrelated from the 'irrelevant' coordinates and momenta \mathcal{Q}_I , so that the initial conditions for the phase density are defined according to:

$$\rho_T(\mathcal{Q}, \mathcal{Q}_I, 0) = \bar{\rho}(\mathcal{Q}, 0) \rho(\mathcal{Q}_I | \mathcal{Q}) \quad (28)$$

with $\bar{\rho}(\mathcal{Q}, 0)$ to be specified later. A formal projection of the irrelevant coordinates can now be carried out, which leads to the derivation of the following master equation (ME), as shown in Appendix A, and equivalent to Eq. 24

$$\frac{\partial \bar{\rho}(\mathcal{Q}, t)}{\partial t} = - \int_0^t d\tau \hat{\Gamma}(\tau) \bar{\rho}(\mathcal{Q}, t - \tau) \quad (29)$$

where the time evolution operator is defined as $\hat{\Gamma}(\tau) = -(\hat{\nabla}_{\mathcal{Q}})^{\text{tr}} \mathbf{J} \bar{\rho}(\mathcal{Q}) \hat{\nabla}_{\mathcal{Q}} \bar{\rho}(\mathcal{Q})^{-1}$. The kernel matrix operator $\hat{\mathbf{J}}(\tau)$ is:

$$\hat{\mathbf{J}}(\tau) = k_B T \begin{pmatrix} \mathbf{0} & -\mathbf{1}\delta(\tau) \\ \mathbf{1}\delta(\tau) & \mathbf{j}\delta(\tau) + \hat{\boldsymbol{\xi}}(\tau) \end{pmatrix} \quad (30)$$

where the generalized friction tensor operator is written as

$$\hat{\boldsymbol{\xi}}(\tau) = \begin{pmatrix} \hat{\boldsymbol{\xi}}_{\text{TT}}(\tau) & \hat{\boldsymbol{\xi}}_{\text{TR}}(\tau) & \hat{\boldsymbol{\xi}}_{\text{TS}}(\tau) \\ \hat{\boldsymbol{\xi}}_{\text{RT}}(\tau) & \hat{\boldsymbol{\xi}}_{\text{RR}}(\tau) & \hat{\boldsymbol{\xi}}_{\text{RS}}(\tau) \\ \hat{\boldsymbol{\xi}}_{\text{ST}}(\tau) & \hat{\boldsymbol{\xi}}_{\text{SR}}(\tau) & \hat{\boldsymbol{\xi}}_{\text{SS}}(\tau) \end{pmatrix} \quad (31)$$

The generalized friction tensor blocks $\hat{\boldsymbol{\xi}}_{\text{AB}}(\tau)$, where the symbols A, B = T, R, S stand for translation, rotation, and shape coordinates, are defined as the correlation functions of the various forces and torques:

$$\hat{\boldsymbol{\xi}}_{\text{AB}}(\tau) = \langle \mathbf{f}_{\text{A}} \exp(-\hat{Q}\hat{\Gamma}\tau) \mathbf{f}_{\text{B}}^{\text{tr}} \rho(\mathcal{Q}_I | \mathcal{Q}) \rangle_I \quad (32)$$

here $\hat{Q} = 1 - \hat{P}$ and A, B = T, R, S, i.e. $\mathbf{f}_T = \hat{\nabla}_R U - \overline{\hat{\nabla}_R U}$, $\mathbf{f}_R = \hat{\mathbf{M}}U - \overline{\hat{\mathbf{M}}U}$, $\mathbf{f}_S = \hat{\nabla}_q U - \overline{\hat{\nabla}_q U}$. The free energy is $F = U_0(\mathcal{X}) + U_{\text{int}}(\mathcal{X}, \mathcal{X}_I)$, and $\bar{\rho}(\mathcal{Q})$ is the Boltzmann distribution for $H = K(\mathcal{Q}) + F(\mathcal{X})$. For a generic function $f(\mathcal{X})$ depending on the configuration coordinates only, one defines:

$$\bar{f} = \langle f \exp(-U_{\text{int}}/k_B T) \rangle_{\mathcal{X}_I} / \langle \exp(-U_{\text{int}}/k_B T) \rangle_{\mathcal{X}_I} \quad (33)$$

In the standard limit of fast relaxing irrelevant coordinates, the Fokker-Planck equation in \mathcal{Q} is obtained⁶⁹

$$\frac{\partial \bar{\rho}(\mathcal{Q}, t)}{\partial t} = -\hat{\Gamma}_{FP} \bar{\rho}(\mathcal{Q}, t) \quad (34)$$

where the time evolution operator is now

$$\hat{\Gamma}_{FP} = \hat{P}r - (\hat{\nabla}_{\mathcal{Q}})^{\text{tr}} \mathbf{J}_{FP} \bar{\rho}(\mathcal{Q}) \hat{\nabla}_{\mathcal{Q}} \bar{\rho}(\mathcal{Q})^{-1} \quad (35)$$

where

$$\mathbf{J}_{FP} = k_B T \begin{pmatrix} \mathbf{0} & -\mathbf{1} \\ \mathbf{1} & \boldsymbol{\xi} \end{pmatrix} \quad (36)$$

$\boldsymbol{\xi}$ is the $(3n - 6) \times (3n - 6)$ friction tensor $(3n - 6) \times (3n - 6)$, which in general depends upon \mathcal{X} :

$$\boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\xi}_{TT} & \boldsymbol{\xi}_{TR} & \boldsymbol{\xi}_{TS} \\ \boldsymbol{\xi}_{RT} & \boldsymbol{\xi}_{RR} & \boldsymbol{\xi}_{RS} \\ \boldsymbol{\xi}_{ST} & \boldsymbol{\xi}_{SR} & \boldsymbol{\xi}_{SS} \end{pmatrix} \quad (37)$$

For convenience we have collected in the term $\hat{P}r$ the precessional part (arising from the $-\mathbf{L}^\times$ block)

$$\hat{P}r = \mathbf{L}^{\text{tr}} \left(\hat{\nabla}_{\mathbf{L}} H \times \hat{\nabla}_{\mathbf{L}} \right) \quad (38)$$

The formal definitions of the friction tensor submatrices $\boldsymbol{\xi}_{AB}$ are:

$$\boldsymbol{\xi}_{AB} = \int_0^\infty d\tau \langle \mathbf{f}_A \exp(-\hat{\Gamma}_I \tau) \mathbf{f}_B^{\text{tr}} \rho(\mathcal{Q}_I | \mathcal{Q}) \rangle_I, \quad (39)$$

where $\hat{\Gamma}_I$ is given in Eq. 26. Hydrodynamic models^{70,71} can be employed to evaluate approximately $\boldsymbol{\xi}_{AB}$. A hydrodynamics approach for the evaluation of $\boldsymbol{\xi}$ is summarized in Appendix B, which generalizes the approach presented elsewhere^{70,71}, limited to torsional angles only,

to the case of generic internal coordinates.

IV. SEMI-FLEXIBLE BROWNIAN BODY

We now investigate possible approximations aimed at making the resulting model computationally manageable. First, we may safely discard the center-of-mass coordinates and momenta from our description. This is acceptable for a single flexible body in an isotropic medium, for which only rotation-dependent correlation functions need to be defined. Next we distinguish between two sets of internal coordinates \mathbf{q} , which we shall call simply \mathbf{q}_s (soft modes) and \mathbf{q}_h (harmonic modes): the latter set encompasses, as usual, more constrained and faster relaxing degrees of freedom. Similarly, \mathbf{p} is also split into \mathbf{p}_s and \mathbf{p}_h .

We perform a second order expansion about the \mathbf{q}_s -dependent values of \mathbf{q}_h with respect to which a minimum is reached, assuming that such a minimum is unique

$$F \approx U^{(0)}(\mathbf{q}_s) + [\mathbf{q}_h - \mathbf{q}_h^{(0)}(\mathbf{q}_s)]^{\text{tr}} \mathbf{K}(\mathbf{q}_s) [\mathbf{q}_h - \mathbf{q}_h^{(0)}(\mathbf{q}_s)] / 2. \quad (40)$$

Pushing this approximation a little further, we may consider that the dependence of $\mathbf{q}_h^{(0)}$ and \mathbf{K} on \mathbf{q}_s is negligible, i.e. we may assume that the hard internal coordinates are given in a fixed coordinate frame that is independent of soft variables \mathbf{q}_s . Similarly, we may consider zeroth-order expansions of all other tensorial properties with respect to \mathbf{q}_h , i.e. the gauge potential \mathbf{A} , the inertia tensor \mathbf{I} , the metric tensor \mathbf{g} , and the friction tensor $\boldsymbol{\xi}$ (reduced to its rotational and shape blocks). Finally, the FP equation takes the form:

$$\frac{\partial \bar{\rho}(\mathcal{Q}, t)}{\partial t} = -\hat{\Gamma} \bar{\rho}(\mathcal{Q}, t) \quad (41)$$

$$\hat{\Gamma} = \hat{\text{Pr}} - \hat{\nabla}_{\mathcal{Q}}^{\text{tr}} \mathbf{J}_{SFB} \bar{\rho}(\mathcal{Q}) \hat{\nabla}_{\mathcal{Q}} \bar{\rho}(\mathcal{Q})^{-1} \quad (42)$$

where $\mathcal{Q} = (\Omega, \mathbf{q}, \mathbf{L}, \mathbf{p})$, $\mathbf{q} = (\mathbf{q}_s, \mathbf{q}_h)$, $\mathbf{p} = (\mathbf{p}_s, \mathbf{p}_h)$, $\bar{\rho}(\mathcal{Q}) = \exp(-H/k_B T) / \langle -H/k_B T \rangle$, $\langle f(\mathcal{Q}) \rangle = \int d\mathcal{Q} f(\mathcal{Q})$, $\hat{\text{Pr}}$ defined by Eq. 38. Hamilton's function is:

$$H = \frac{1}{2} \mathbf{L}^{\text{tr}} \mathbf{I}^{-1}(\mathbf{q}_s) \mathbf{L} + \frac{1}{2} [\mathbf{p} - \mathbf{A}(\mathbf{q}_s) \mathbf{L}]^{\text{tr}} \mathbf{g}(\mathbf{q}_s) [\mathbf{p} - \mathbf{A}(\mathbf{q}_s) \mathbf{L}] + \frac{1}{2} [\mathbf{q}_h - \mathbf{q}_h^{(0)}(\mathbf{q}_s)]^{\text{tr}} \mathbf{K}(\mathbf{q}_s) [\mathbf{q}_h - \mathbf{q}_h^{(0)}(\mathbf{q}_s)] + U^{(0)}(\mathbf{q}_s) \quad (43)$$

and tensor \mathbf{J}_{SFB} is

$$\mathbf{J}_{SFB} = k_B T \begin{pmatrix} \mathbf{0} & \mathbf{0} & -\mathbf{1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & -\mathbf{1} \\ \mathbf{1} & \mathbf{0} & \boldsymbol{\xi}_{RR}(\mathbf{q}_s) & \boldsymbol{\xi}_{RS}(\mathbf{q}_s) \\ \mathbf{0} & \mathbf{1} & \boldsymbol{\xi}_{SR}(\mathbf{q}_s) & \boldsymbol{\xi}_{SS}(\mathbf{q}_s) \end{pmatrix} = k_B T \begin{pmatrix} \mathbf{0} & -\mathbf{1} \\ \mathbf{1} & \boldsymbol{\xi} \end{pmatrix} \quad (44)$$

with a similar definition for $\boldsymbol{\xi}$ as in Eq. 39.

A. Harmonic internal coordinates

Numerically solving the model described above is still a formidable task, although manageable in some specific circumstances. We address here the case of an empty \mathbf{q}_s set, i.e. only harmonic shape variables are left, subject to a free energy having a quadratic form: $\mathbf{q} \equiv \mathbf{q}_h$, $\mathbf{p} \equiv \mathbf{p}_h$.

All tensors and vectors \mathbf{I} , \mathbf{A} , \mathbf{g} , \mathbf{K} , $\boldsymbol{\xi}$, $\mathbf{q}^{(0)}$ are now considered approximately constant. In practice, we are neglecting activated torsional rearrangements and/or crankshaft motions^{72,73}. Instead, we concentrate on the description of internal motions adopting the common view of a harmonic or boson bath, but 1) directly recovered from the complete Liouville description, 2) retaining full coupling with external tumbling, 3) including dissipative/stochastic effects. The model is therefore formally similar to other approaches which have been proposed relying on a description of a macromolecule as a collection of harmonic modes (vide supra).⁴³⁻⁴⁹ Since the total energy is a quadratic form, the model is relatively manageable without neglecting neither inertial effects, nor internal-rotational coupling, at least for the purpose of calculating observables as in Eq. 6. Let $\mathbf{y} = (\mathbf{q}, \mathbf{L}, \mathbf{p})$, so that $\mathcal{Q} = (\Omega, \mathbf{y})$. The total energy is $H = (\mathbf{y} - \mathbf{y}^{(0)})^{\text{tr}} \mathbf{k} (\mathbf{y} - \mathbf{y}^{(0)})/2$, where $\mathbf{y}^{(0)} = (\mathbf{q}^{(0)}, \mathbf{0}, \mathbf{0})$. The time evolution equation is written exactly as in the previous cases, and it contains $\rho(\mathcal{Q})$, Boltzmann distribution defined on H (see below), the precessional term $\hat{\mathbf{P}}r$ is defined again by Eq. 38 and the general tensor \mathbf{J} , in which all terms are now supposed constant. The matrix \mathbf{k} has the form

$$\mathbf{k} = \begin{pmatrix} \mathbf{K} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I}^{-1} + \mathbf{A}^{\text{tr}} \mathbf{g} \mathbf{A} & -\mathbf{A}^{\text{tr}} \mathbf{g} \\ \mathbf{0} & -\mathbf{g} \mathbf{A} & \mathbf{g} \end{pmatrix} \quad (45)$$

It is convenient to introduce the scaled and rotated variables defined as $\mathbf{x} = (k_B T)^{-1/2} \mathbf{S}(\mathbf{y} - \mathbf{y}^{(0)})$, where $\mathbf{S} = \mathbf{\Lambda} \mathbf{U}^{\text{tr}}$. \mathbf{U} and $\mathbf{\Lambda}$ are defined as $\mathbf{U}^{\text{tr}} \mathbf{k} \mathbf{U} = \mathbf{\Lambda}^2$, with $\mathbf{\Lambda}$ diagonal. In terms of the new set of coordinates $\mathcal{Q} = (\Omega, \mathbf{x})$, we finally get

$$\frac{\partial \rho(\mathcal{Q}, t)}{\partial t} = -\hat{\Gamma} \rho(\mathcal{Q}, t) \quad (46)$$

$$\hat{\Gamma} = \hat{\text{Pr}} - (\hat{\nabla}_{\mathcal{Q}})^{\text{tr}} \boldsymbol{\omega}_T \rho(\mathcal{Q}) \hat{\nabla}_{\mathcal{Q}} \rho(\mathcal{Q})^{-1} \quad (47)$$

where $\hat{\nabla}_{\mathcal{Q}} = (\hat{\mathbf{M}}, \hat{\nabla}_{\mathbf{x}})$, $H = -x^2/2$, $\rho(\mathcal{Q}) = \rho(\mathbf{x}) = p(\mathbf{x})/8\pi^2 = \exp(-x^2/2)/(2\pi)^{(6n-9)/2} 8\pi^2$

$$\boldsymbol{\omega}_T = \begin{pmatrix} \mathbf{0} & -\boldsymbol{\omega}^{\text{int}} \\ (\boldsymbol{\omega}^{\text{int}})^{\text{tr}} & \boldsymbol{\omega}^{\text{io}} \end{pmatrix} \quad (48)$$

$\boldsymbol{\omega}^{\text{io}}$ (for the internal overall coordinates) and $\boldsymbol{\omega}^{\text{int}}$ (for interaction between rotation and internal coordinates) are non-symmetric matrices $(6n-9) \times (6n-9)$ and $3 \times (6n-9)$ respectively, with dimensions of frequencies: $\boldsymbol{\omega}^{\text{io}}$ is related to dissipative properties (friction tensors) and it has the form:

$$\boldsymbol{\omega}^{\text{io}} = \mathbf{S} \begin{pmatrix} \mathbf{0} & \mathbf{0} & -\mathbf{1} \\ \mathbf{0} & \boldsymbol{\xi}_{\text{RR}} & \boldsymbol{\xi}_{\text{RS}} \\ \mathbf{1} & \boldsymbol{\xi}_{\text{SR}} & \boldsymbol{\xi}_{\text{SS}} \end{pmatrix} \mathbf{S}^{\text{tr}} \quad (49)$$

while $\boldsymbol{\omega}^{\text{int}}$ is

$$\boldsymbol{\omega}^{\text{int}} = (k_B T)^{1/2} \mathbf{e} \mathbf{S}^{\text{tr}} \quad (50)$$

where $\mathbf{e} = (\mathbf{0} \ \mathbf{1} \ \mathbf{0})$. The precessional operator can be written in the general form $\hat{\text{Pr}} = \sum ijk \omega_{ijk}^P x_i x_j \frac{\partial}{\partial x_k}$, where coefficients ω_{ijk}^P can be found straightforwardly (from now on we neglect tensorial notation) as

$$\omega_{ijk}^P = (k_B T)^{1/2} \sum_{pqr} \epsilon_{pqr} (\mathbf{e} \mathbf{S}^{-1})_{ri} (\mathbf{e} \mathbf{S}^{\text{tr}})_{pj} (\mathbf{e} \mathbf{S}^{\text{tr}})_{qk} \quad (51)$$

A further simplifying approximation can be introduced by considering the order of magnitudes of the different contributions to the time evolution operator, as weighted by $\boldsymbol{\omega}^{\text{io}}$, $\boldsymbol{\omega}^{\text{int}}$, $\boldsymbol{\omega}^P$. The elements of $\boldsymbol{\omega}^{\text{io}}$ are of order $O(\Lambda \xi)$, while $\boldsymbol{\omega}^{\text{int}} \sim \boldsymbol{\omega}^P$ are of the order $O(k_B T)^{1/2} \Lambda$. In a normal solvent at room temperature $\boldsymbol{\omega}^P \sim \boldsymbol{\omega}^{\text{int}}$ elements can be considered negligible as

additive terms with respect to ω^{io} . This allows to neglect, at least as a first approximation, the direct contribution of the precession terms. We shall therefore write Eq. 47 in the form:

$$\hat{\Gamma} = \hat{\Gamma}_0 + \hat{\Gamma}_{\text{int}} = - \sum_{ij} \omega_{ij}^{\text{io}} \frac{\partial}{\partial x_i} p(\mathbf{x}) \frac{\partial}{\partial x_j} p(\mathbf{x})^{-1} + \sum_{ip} \omega_{ip}^{\text{int}} x_i \hat{M}_p \quad (52)$$

where $i = 1, \dots, 6n - 9$, $p = 1, 2, 3$.

Summarizing, this is the simplest description of the dynamic of a non-rigid body undergoing diffusive fluctuations, that is recoverable from an initially atomistic model for the Brownian probe, and fully retains inertial effects, dissipation and rotation/shape coupling. It describes the macromolecule as a rotator coupled to $6n - 9$ (i.e. $3n - 6$ internal coordinates, $3n - 6$ internal momenta and 3 components of the \mathbf{L} vector) harmonic degrees of freedom, in a fashion quite similar to standard spin-boson quantum mechanical approaches. Indeed, the similarity can be exploited and manageable expressions can be found for the evaluation of orientation correlation functions, at least in specific dynamic regimes. In the following, we address this approximated model as the semi-flexible body (SFB).

V. PARAMETRIZATION

In a following paper, II⁵⁶, we shall address the evaluation, based on the SFB model, of correlation functions and spectral densities of suitable observables expressed as functions of \mathcal{Q} . Here we analyze the structure of the time evolution operator given in Eq. 52 by looking at the parameters employed to evaluate ω_{ij}^{io} , $\omega_{i\alpha}^{\text{int}}$, which are in principle recoverable from classical force fields and/or standard molecular dynamics (MD) calculations and hydrodynamic estimates of friction tensors. We shall consider for a test trial of the model the series made from the first nine polyalanines (ALA-) $_n$, with $n = 2, \dots, 10$, as shown in Figure 2.

The molecular geometries shown in Figure 2 were optimized via a Molecular Mechanics (MM) energy minimization, using the molecular modeling toolkit (MMTK) software package.⁷⁴ Minimization was carried out *in vacuo* using the Amber99 force field,⁷⁵ which is the default force field in MMTK. These optimized structures were used as reference geometries for the calculation of tensors \mathbf{g} , \mathbf{A} , \mathbf{I} , and the friction tensor $\boldsymbol{\xi}$; the curvature matrix \mathbf{K} was evaluated first directly from the Hessian based on the same force field employed in the MM optimization and next via the evaluation of the variance-covariance matrix \mathbf{C}

(see below) for the internal coordinates based on a 200 ns-long MD simulation, carried out with NAMD,⁷⁶ following a standard protocol consisting in energy minimization, heating, equilibration, and production. Parameters and conditions for the MD simulations are reported in Table II. The friction tensor $\boldsymbol{\xi}$ was obtained following the algorithm discussed in Appendix B, based on a generalized approach for flexible molecules implemented in the software DiTe.^{70,71} Parameters for the calculation of the friction tensor were: temperature 298.15 K, viscosity 0.84 cP (water at 298.15 K), stick boundary conditions, and effective radius 0.5 Å (see Appendix B). Box dimensions have been chosen equal to 30 Å up to the case of 8-alanine, and tested for increased dimensions for 9 and 10-alanine. Kasahara et al.⁷⁷ 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 of 3-7 poly-alanines, and that one configuration (PP_{II}) is relevant.⁷⁸ Such an observation is compatible with the description of the molecules as flexible objects oscillating around a global minimum, therefore in agreement with the basic hypothesis of the SFB model. It should be stressed that poly-alanine molecules have been chosen as sand-box examples solely to discuss the method. As such, we are mainly concerned with testing the magnitude and behavior of the parameters derived in the previous sections, and we are aware that the applicability of the SFB model can be limited, at least in the case of the longer homologues. In particular the 10-alanine case requires additional investigations. In the literature, the folded-unfolded equilibrium has been studied both in vacuum and in water.^{79,80} In vacuum, the free energy profile has a single deep minimum centered on the folded structure, with very steep energy barriers growing up to 40 $k_B T$ (at room temperature) for the elongated structure. In water the free energy profile assumes a bi-stable form, showing a relative minimum of the elongated 10-alanine peptide at a energy 2 $k_B T$ higher with respect to the folded configuration, and a barrier of 8 $k_B T$.⁸⁰ To further explore this behavior we produced 80 ns MD trajectories of 10-alanine in cubic boxes of 40, 50, and 60 Å of side length (see below for further comments).

In the following, the MF is assumed to coincide with the principal frame reference for the inertia tensor. In Figure 3 the principal values of the inertia tensors are shown as functions of the number of alanines in the series. All molecules have, approximately, a spherical ($n \leq 5$) or spheroid prolate ($n > 5$) geometry.

In order to discuss the structure of the operator in Eq. 52, determined exclusively by

TABLE II. Setup conditions for the MD simulations of the 9 polyalanines (ALA-) $_n$, with $n = 2, \dots, 10$

Parameter	Value
Cubic box side length / Å	30 (60) ^a
Boundary conditions	periodic
Number of water molecules ^b	900, 892, 892, 885, 877, 866, 853, 845, 7150
Force field	CHARMM22 with CMAP correction for the peptides, ⁸¹ 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 ^c	12.0
Electrostatics	particle mesh Ewald, order 6, tolerance 10^{-5}
Integration time step / fs ^d	2
Minimization steps	5000
Heating time / ps	72
Equilibration time / ns	2
Production time / ns	200

^aBox dimension of 60 Å for (ALA-) $_{10}$.

^bFrom (ALA-) $_2$ to (ALA-) $_{10}$: number of molecules, chosen to reproduce the density of bulk water at 298.15 K.

^cSwitching at 10.0 Å; pairlist set equal to 13.5 Å.

^dAll bonds with H atoms were constrained with the SHAKE algorithm.

matrices ω^{io} , ω^{int} , we analyze first the parameters directly related to the energy profile, i.e. the curvature \mathbf{K} , and to energy dissipation due to internal coordinates i.e. ξ_{SS} . It is convenient to look at $\omega_K = \sqrt{\mathbf{gK}}$ and $\omega_\xi = \mathbf{g}\xi_{SS}$ which are, respectively, the intrinsic harmonic frequency matrix and the dissipation frequency matrix of the internal coordinates in the absence of coupling with overall rotation, i.e. the deterministic parameters of the set of damped coupled oscillators $\ddot{\mathbf{q}} = -\omega_\xi \dot{\mathbf{q}} - \omega_K^2 \mathbf{q}$.

In Figure 4 we show a map of $\log |(\omega_K)_{ij}|$, where ω_K is obtained through the Hessian of the internal energy. The matrix elements are organized in blocks related to residues; matrix elements are clustered around values of 10^{-2} fs $^{-1}$, as shown in the histograms of Figure 5, with significant tails at both higher and lower frequencies (down to 10^{-5} fs $^{-1}$). The evaluation of the curvature matrix \mathbf{K} based on the Hessian *in vacuo* is a crude, but fast shortcut to characterize the energetics of the internal coordinates. A more refined approach,

taking into account the presence of the surrounding medium (i.e. estimating a free energy instead of an internal energy), is to estimate \mathbf{K} from the variance-covariance matrix \mathbf{C} obtained from an MD trajectory, i.e. $\mathbf{K} = (k_B T)^{-1} \mathbf{C}$. This approach requires a significantly increased effort, and in practice requires the following steps: i) first evaluate from a MD trajectory covariance matrix expressed in Cartesian coordinates, then ii) convert it to the corresponding matrix in internal coordinates⁸² and finally invert it to get \mathbf{K} . At least for the test case of relatively rigid polyanines series, the results obtained are close to the simpler estimate based on the Hessian.

In Figure 6 we show the map of $\log |(\omega_K)_{ij}|$ obtained from the covariance calculated from MD (O-H bending and H-O-H stretching not constrained). The matrix structure is similar to the one observed in Figure 4, with a slightly broader distribution of values. This is confirmed in Figure 7, which also shows a slightly more extended tail at lower frequencies. All things considered, one can conclude that the Hessian-based evaluation, of negligible computational cost, of the curvature matrix is satisfactorily in agreement with the computationally intensive covariance-variance matrix method, which is based on a MD trajectory, and can therefore be employed safely, at least for the trial set of molecules considered here, with the exception of 10-alanine, for which we observe an increased internal mobility.

Next, we inspect the properties of the friction related matrix, ω_ξ . It is characterized by a similar structure, as shown in Figure 8, although with a significantly larger extension in the domain of higher frequencies, up to 10^{-5} fs^{-1} . This is related to the fast relaxation of conjugate momenta \mathbf{p} , which is to be expected in the diffusive regime (relatively high friction) of motion for molecular systems in water at room temperature. In the SFB operator we made the choice of putting together the internal coordinates, their conjugate momenta, and the angular momentum vector \mathbf{L} in the collective ensemble \mathbf{x} , since they relax significantly faster than the molecular rotation $\mathbf{\Omega}$. This can be finally verified, by looking at the coupling between internal dynamics and rotation, i.e. matrix ω^{int} , and the overall relaxation of \mathbf{x} , i.e. the matrix ω^{io} .

In Figure 10 we show the histogram graphs of $\log |(\omega^{\text{int}})_{ij}|$, while in Figure 11 the same information is given for $\log |(\omega^{\text{io}})_{ij}|$. The weak coupling is shown by the shift of the two sets of distributions centered around 10^{-7} fs^{-1} and a tail up to 10^{-5} fs^{-1} for ω^{int} , and with a bimodal shape for ω^{io} , with a peak around 10^{-3} fs^{-1} and one, much faster around 10^2 fs^{-1} . This time scale separation can be exploited to devise specific methods of solution for

evaluating orientation correlation functions as given in Eq. 6, to take into account 1) how the softness of the molecule influences the effective global rotational diffusion; and 2) that relaxation as defined in Eq. 6 involves a fast component in order to account for the local geometry. We discuss in the companion paper⁵⁶, some methods for a computationally convenient evaluation of Eq. 6 based on Eq. 52.

An additional analysis of the parameters obtained from the MD simulations is useful to discuss further the approximations introduced. Figure 12 compares the histograms of the ω_K frequencies for different box lengths, which are related to the configurational energy for the case of 10-alanine. The three distributions are similar to each other, but differ appreciably from the simulation in the 30 Å side length. The difference is in the fact that the distribution is narrowed to higher frequencies, which in turn means a floppier internal dynamics. This is confirmed by a comparison, shown in Figure 13, of map and histograms of $\log |\omega_K \times \text{fs}|$ for the 10-alanine peptide calculated via the Hessian (above) and the inversion of the covariance matrix (below).

The friction tensor has been considerably simplified in the formulation of the SFB model. The constant friction approximation is usually acceptable for small molecules and for molecules not showing large amplitude motions.⁸³⁻⁸⁵ Examples of trajectories for 2-alanine, 4-alanine, 6-alanine and 10-alanine of diagonal friction matrix elements $\log |(\omega_\xi)_{ii}|$ in time are shown in Figure 14. Fluctuations are observed which can be significant; however their overall effect is damped if one considers the average eigenvalues of ω_ξ and their standard deviation, Figure 15. For most eigenvalues, deviations from average values are below 1%. Few eigenvalues in the larger peptides show more important deviations, which in principle can be taken into account in the solution of the stochastic equation. This is, however, not always necessary as the friction tensor calculated in the absolute energy minimum can be sufficient to describe the most important features of the long-time dynamics of the system.⁸⁵ Since including the dependence on the configuration requires a larger computational effort, one should decide whether the constant-friction approximation is inappropriate based on the quality of the agreement with experimental data. For the goal of this and of the following paper,⁵⁶ which is to show how to build a stochastic model of a system from its all-atom description without resorting to phenomenological parameters, we have purposely chosen the simplest implementation. In any event, inclusion of fast fluctuations in the solution of the Fokker-Planck equation is possible and can be addressed with limited computational

effort.⁵⁶ The same rationale - limited computational effort for sake of testing - is behind other assumptions. The choices of the effective hydrodynamic radius value and neglecting hydrodynamic interactions, in particular, have been based on past experience of the authors on the hydrodynamic modeling of the friction/diffusion tensor. In ref.⁸⁶ it has been shown that when the generalized diffusion tensor is calculated omitting the hydrodynamic interactions between the beads (the heavy atoms of the molecule), the correct order of magnitude of the diffusion tensor is obtained using an effective hydrodynamic radius of the order of 0.5 Å. This value is smaller with respect to the usually employed value of 2.0 Å, which is the reference value when hydrodynamic interactions among beads are taken into account, e.g. via the Rotne-Prager model.^{71,87}

For the sake of completeness, the table below compares the principal values of the rotational diffusion tensor of 2-alanine calculated with and without hydrodynamic interactions using, respectively, effective radii of 0.5 Å and 2.0 Å. The principal values are of the same order

TABLE III. Principal values of the global rotational diffusion tensor of 2-alanine⁷¹ with and without hydrodynamic interactions among the beads.

	Hyd. coupling	No hyd. coupling
$D_{XX} / 10^9 \text{ Hz}$	3.97	2.50
$D_{YY} / 10^9 \text{ Hz}$	4.33	2.58
$D_{ZZ} / 10^9 \text{ Hz}$	9.65	4.33

of magnitude. Of course, the two models provide different values and anisotropy, and in the future, the more sophisticated choice of fully including hydrodynamics interactions can be followed. Here we are interested in defining the basis of the overall methodology and therefore we have dropped the implementation of the hydrodynamic interactions. Notice that within the obvious limitation of the SFB model, we show in the companion paper⁵⁶ that predicted correlation functions are nevertheless in relatively good agreement with correlation function obtained from molecular dynamics simulations. Relevant adjustments, such as adding fluctuating tensors, hydrodynamic interactions etc, will have to be considered for a fully quantitative comparison with experimental relaxation data, and to improve the model applicability, beyond the "sand-box" cases considered here.

VI. SUMMARY

A particularly interesting feature of stochastic approaches from a spectroscopist’s viewpoint, is that they allow to build an adjustable fine-to-coarse grained hierarchy of dynamical models of large molecular objects, depending on additional *a priori* extraneous information that may provide suitable approximation strategies. However, in so doing, the proper theoretical framework is often not made clearly explicit, making sometime difficult to assess the validity of the approximations used.

In this report, we propose a partial answer to this conundrum by establishing a systematic approach, summarized in Fig. 16, to describe the dynamics of a non rigid molecule, based on elaborations from fundamental classical and statistical mechanics, in the form of a family of multidimensional Fokker-Planck operators for the probability density of internal and external degrees of freedom, retaining inertial effects and dissipation. The approach developed provides a description of the dynamics of a non-rigid protein interacting with its solvent. The projection technique, which averages out the degrees of freedom (dof) of the ‘bath’ (the irrelevant dof, when one is interested in a small number of protein atoms) and leads to the reduced FPE of the relevant atoms, naturally includes hydrodynamic properties. These, by construction, include all elementary interactions between particles in the phase space.

The Fokker-Planck equation for rigid-body dynamics derived in this paper seems particularly relevant for the description of large molecular objects, such as proteins, which represent the main domain of application in the authors’ perspective. It is a versatile method that enables one to tackle virtually any classical non rigid dynamics, and can be seen as a complementary approach to full blown long MD simulations when the purpose is the evaluation and interpretation of correlations/spectral densities directly related to spectroscopic observables. The method provides a physically sound framework, which allows a clear introduction of approximations and integration from additional modeling sources, thus leading naturally to an integrated approach to the interpretation of relaxation properties of (large) molecular objects. Resulting models can be discussed at different level of complexity, and we have discussed the simplest illustration, the SFB model, which can be used to interpret magnetic resonance relaxation of a stable molecule in solution, not subject to crankshaft large amplitude torsional motions. Here, we have shown that the SFB model retains a wealth of

complex properties. The simplifying assumption of harmonic internal coordinates provided a convenient illustration of the potential of our approach and its links to more conventional approaches. As we show in the companion paper, the SFB model is amenable to practical numerical implementation, by retaining the description of all the internal degrees of freedom of the system at a modest computational price. Moreover, the SFB model provides a sound theoretical framework and paves the way to the study of more demanding problems, where soft degrees of freedom cannot be discarded and are explicitly considered to account for large conformational transitions, and friction tensors with memory may occur. The fine-tuning capability for the description of the dynamics of large molecular objects, for instance based on Eqs. 41-44, will be developed in further work, extending the computational treatment of the SFB model described in the companion paper.⁵⁶

Appendix A: Projection of the irrelevant coordinates

We review here the projection of an irrelevant set of coordinates. As in section III we have $\mathcal{Q} = (\mathcal{X}, \mathcal{P})$ and $\mathcal{Q}_I = (\mathcal{X}_I, \mathcal{P}_I)$; we start from the density of states $\rho_T(\mathcal{Q}, \mathcal{Q}_I, t)$ and the corresponding Boltzmann distribution $\rho_T(\mathcal{Q}, \mathcal{Q}_I) = \exp(-H_T/k_B T) / \langle \exp(-H_T/k_B T) \rangle_I$, $\langle f(\mathcal{Q}, \mathcal{Q}_I) \rangle_T = \langle \langle f(\mathcal{Q}, \mathcal{Q}_I) \rangle_I \rangle = \int d\mathcal{Q} \int d\mathcal{Q}_I f(\mathcal{Q}, \mathcal{Q}_I)$. The Hamiltonian is

$$H_T = K(\mathcal{Q}) + K_I(\mathcal{Q}_I) + U(\mathcal{Q}, \mathcal{Q}_I) \quad (\text{A1})$$

The time evolution equation is in general

$$\frac{\partial \rho_T(\mathcal{Q}, \mathcal{Q}_I, t)}{\partial t} = -\hat{\Gamma}_T \rho_T(\mathcal{Q}, \mathcal{Q}_I, t) \quad (\text{A2})$$

$$\begin{aligned} \hat{\Gamma}_T &= \hat{\Gamma} + \hat{\Gamma}_I = -(\hat{\nabla}_{\mathcal{Q}})^{\text{tr}} \mathbf{J} \rho_T(\mathcal{Q}, \mathcal{Q}_I) \hat{\nabla}_{\mathcal{Q}} \rho_T(\mathcal{Q}, \mathcal{Q}_I)^{-1} \\ &\quad - (\hat{\nabla}_{\mathcal{Q}_I})^{\text{tr}} \mathbf{J}_I \rho_T(\mathcal{Q}, \mathcal{Q}_I) \hat{\nabla}_{\mathcal{Q}_I} \rho_T(\mathcal{Q}, \mathcal{Q}_I)^{-1} \end{aligned} \quad (\text{A3})$$

we do not specify further $\partial/\partial\mathcal{Q}$ and $\partial/\partial\mathcal{Q}_I$, which are generalized gradients; we assume that \mathbf{J}, \mathbf{J}_I are functions of $\mathcal{Q}, \mathcal{Q}_I$ only, respectively, with the generic form

$$\mathbf{J} = k_B T \begin{pmatrix} \mathbf{0} & -\mathbf{1} \\ \mathbf{1} & \boldsymbol{\xi} \end{pmatrix}, \quad \mathbf{J}_I = k_B T \begin{pmatrix} \mathbf{0} & -\mathbf{1} \\ \mathbf{1} & \boldsymbol{\xi}_I \end{pmatrix} \quad (\text{A4})$$

and $\boldsymbol{\xi} = \boldsymbol{\xi}(\mathcal{Q})$, $\boldsymbol{\xi}_I = \boldsymbol{\xi}_I(\mathcal{Q}_I)$. We define the averaged equilibrium and time dependent probability $\bar{\rho}(\mathcal{Q})$, $\bar{\rho}(\mathcal{Q}, t)$ and the conditional equilibrium probability $\rho(\mathcal{Q}_I|\mathcal{Q})$:

$$\begin{aligned}\bar{\rho}(\mathcal{Q}) &= \langle \rho_T(\mathcal{Q}, \mathcal{Q}_I) \rangle_I \\ \bar{\rho}(\mathcal{Q}, t) &= \langle \rho_T(\mathcal{Q}, \mathcal{Q}_I, t) \rangle_I \\ \rho(\mathcal{Q}_I|\mathcal{Q}) &= \rho_T(\mathcal{Q}, \mathcal{Q}_I) / \bar{\rho}(\mathcal{Q})\end{aligned}\tag{A5}$$

and the projection operator

$$\hat{P} = \rho(\mathcal{Q}_I|\mathcal{Q}) \langle \dots \rangle_I\tag{A6}$$

Finally we assume the initial condition

$$\rho_T(\mathcal{Q}, \mathcal{Q}_I, 0) = \bar{\rho}(\mathcal{Q}, 0) \rho(\mathcal{Q}_I|\mathcal{Q})\tag{A7}$$

The following properties are easily verified: *i*) $\hat{P}^2 = \hat{P}$, *ii*) $\hat{P}^\dagger = \hat{P}$, *iii*) $\hat{P} \rho_T(\mathcal{Q}, \mathcal{Q}_I) = \rho_T(\mathcal{Q}, \mathcal{Q}_I)$, *iv*) $\hat{P} \hat{\rho}_T(\mathcal{Q}, \mathcal{Q}_I, t) = \rho(\mathcal{Q}_I|\mathcal{Q}) \bar{\rho}(\mathcal{Q}, t)$, *v*) $\hat{P} \hat{\Gamma}_I f(\mathcal{Q}, \mathcal{Q}_I) = 0$. After defining $\hat{Q} = 1 - \hat{P}$, the well-known formal solution is obtained

$$\begin{aligned}\frac{\partial \hat{P} \rho_T(\mathcal{Q}, \mathcal{Q}_I, t)}{\partial t} &= -\hat{P} \hat{\Gamma}_T \hat{P} \rho_T(\mathcal{Q}, \mathcal{Q}_I, t) + \int_0^t d\tau \hat{P} \hat{\Gamma}_T \exp(-\hat{Q} \hat{\Gamma}_T \tau) \hat{Q} \hat{\Gamma}_T \hat{P} \rho_T(\mathcal{Q}, \mathcal{Q}_I, t - \tau) \\ &\quad + \hat{P} \hat{\Gamma}_T \exp(-\hat{Q} \hat{\Gamma}_T t) \hat{Q} \rho_T(\mathcal{Q}, \mathcal{Q}_I, 0)\end{aligned}\tag{A8}$$

The last term goes to zero due to the initial conditions; the first terms is the averaged operator depending upon the average equilibrium distribution $\bar{\rho}(\mathcal{Q})$, the second and most complex term is defined in terms of a general kernel \hat{K} ; summarizing

$$\frac{\partial \bar{\rho}(\mathcal{Q}, t)}{\partial t} = (\hat{\nabla}_{\mathcal{Q}})^{\text{tr}} \mathbf{J} \bar{\rho}(\mathcal{Q}) \hat{\nabla}_{\mathcal{Q}} \bar{\rho}(\mathcal{Q})^{-1} \bar{\rho}(\mathcal{Q}, t) + \int_0^t \hat{\mathbf{K}}(\tau) \bar{\rho}(\mathcal{Q}, t - \tau)\tag{A9}$$

The kernel operator $\hat{\mathbf{K}}(\tau)$ can be evaluated after some passages. One obtains:

$$\begin{aligned}\hat{\mathbf{K}}(\tau) &= \langle \hat{\Gamma}_T \exp(-\hat{Q} \hat{\Gamma}_T \tau) \hat{Q} \hat{\Gamma}_T \rho(\mathcal{Q}_I|\mathcal{Q}) \rangle_I \\ &= \langle \hat{\Gamma} \exp(-\hat{Q} \hat{\Gamma}_T \tau) \hat{Q} \hat{\Gamma} \rho(\mathcal{Q}_I|\mathcal{Q}) \rangle_I \\ &= (k_B T)^2 (\hat{\nabla}_{\mathcal{Q}})^{\text{tr}} \mathbf{J}^{\text{tr}} \hat{\mathbf{k}}(\tau) \mathbf{J} \hat{\nabla}_{\mathcal{Q}} \bar{\rho}(\mathcal{Q})^{-1}\end{aligned}\tag{A10}$$

where

$$\begin{aligned}\hat{\mathbf{k}}(\tau) &= \langle \rho_T \left(\hat{\nabla}_{\mathcal{Q}} \right) \rho_T^{-1} \exp(-\hat{Q}\hat{\Gamma}_T\tau) \hat{Q} \left(\hat{\nabla}_{\mathcal{Q}} \right)^{\text{tr}} \rho(\mathcal{Q}_I|\mathcal{Q}) \rangle_I \\ &= \langle \rho_T \left(\hat{\nabla}_{\mathcal{Q}} \right) \rho_T^{-1} \exp(-\hat{Q}\hat{\Gamma}_T\tau) \hat{Q} \left(\hat{\nabla}_{\mathcal{Q}} \rho(\mathcal{Q}_I|\mathcal{Q}) \right)^{\text{tr}} \rangle_I\end{aligned}\quad (\text{A11})$$

where an integration by parts is used in the last passage. With some further work one gets

$$\hat{\mathbf{k}}(\tau) = -\frac{1}{(k_B T)^2} \langle \left(\hat{\nabla}_{\mathcal{Q}} H_T - \overline{\hat{\nabla}_{\mathcal{Q}} H_T} \right) \exp(-\hat{Q}\hat{\Gamma}_T\tau) \left(\hat{\nabla}_{\mathcal{Q}} H_T - \overline{\hat{\nabla}_{\mathcal{Q}} H_T} \right)^{\text{tr}} \rho(\mathcal{Q}_I|\mathcal{Q}) \rangle_I \quad (\text{A12})$$

where

$$\overline{\hat{\nabla}_{\mathcal{Q}} H_T} = \langle \hat{\nabla}_{\mathcal{Q}} H_T \rho(\mathcal{Q}_I|\mathcal{Q}) \rangle_I \quad (\text{A13})$$

Since $\hat{\nabla}_{\mathcal{P}} H_T = \hat{\nabla}_{\mathcal{P}} K$ which does not depend upon \mathcal{Q}_I , terms $\hat{\nabla}_{\mathcal{P}} H_T - \overline{\hat{\nabla}_{\mathcal{P}} H_T}$ go to zero, and (given the assumed form of \mathbf{J}) one is left with the generalized collision operator

$$\hat{\mathbf{K}}(\tau) = (\hat{\nabla}_{\mathcal{P}})^{\text{tr}} \hat{\xi}(\tau) \bar{\rho}(\mathcal{Q}) \hat{\nabla}_{\mathcal{P}} \bar{\rho}(\mathcal{Q})^{-1} \quad (\text{A14})$$

where

$$\hat{\xi}(\tau) = \left\langle \left(\frac{\partial U}{\partial \mathcal{X}} - \overline{\frac{\partial U}{\partial \mathcal{X}}} \right)^{\text{tr}} \exp(-\hat{Q}\hat{\Gamma}\tau) \left(\frac{\partial U}{\partial \mathcal{X}} - \overline{\frac{\partial U}{\partial \mathcal{X}}} \right) \rho(\mathcal{Q}_I|\mathcal{Q}) \right\rangle_I \quad (\text{A15})$$

It follows that

$$\frac{\partial \bar{\rho}(\mathcal{Q}, t)}{\partial t} = - \int_0^t d\tau \hat{\Gamma}(\tau) \bar{\rho}(\mathcal{Q}, t - \tau) \quad (\text{A16})$$

$$\hat{\Gamma}(\tau) = -(\hat{\nabla}_{\mathcal{Q}})^{\text{tr}} \hat{\mathbf{J}}(\tau) \bar{\rho}(\mathcal{Q}) \hat{\nabla}_{\mathcal{Q}} \bar{\rho}(\mathcal{Q})^{-1} \quad (\text{A17})$$

where

$$\hat{\mathbf{J}}(\tau) = \hat{\mathbf{J}}\delta(t) + k_B T \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \hat{\xi}(\tau) \end{pmatrix} \quad (\text{A18})$$

Appendix B: Generalized friction tensor

Following the hydrodynamics approach for flexible molecules^{70,71} we first calculate the matrix of constraints \mathbf{B} , that relates the generalized velocities of the molecule, $\dot{\mathbf{Q}} =$

($\mathbf{V}_{CM}, \boldsymbol{\omega}, \dot{\mathbf{q}}$), to the Cartesian velocities of the atoms, \mathbf{v} , i.e.

$$\mathbf{v} = \mathcal{B}\dot{\mathbf{Q}} \quad (\text{B1})$$

To this purpose, it is sufficient to write the explicit expression that gives the velocity of the j -th atom. We start by writing the relationship among the Cartesian coordinates of atom α (\mathbf{R}_α) in the laboratory frame (LF), and the generalized coordinates \mathbf{Q} as

$${}^L\mathbf{R}_\alpha = {}^L\mathbf{R}_{CM} + {}^L\mathbf{r}_\alpha = {}^L\mathbf{R}_{CM} + \mathbf{E}(\boldsymbol{\Omega})\mathbf{c}_\alpha(\mathbf{q}) \quad (\text{B2})$$

and taking the total time derivative

$$\begin{aligned} \mathbf{v}_\alpha &= \mathbf{V}_{CM} + \dot{\mathbf{E}}(\boldsymbol{\Omega})\mathbf{c}_\alpha(\mathbf{q}) + \mathbf{E}(\boldsymbol{\Omega})\dot{\mathbf{c}}_\alpha(\mathbf{q}) \\ &= \mathbf{V}_{CM} + [\mathbf{E}(\boldsymbol{\Omega})\mathbf{c}_\alpha(\mathbf{q})]^\times \boldsymbol{\omega} + \mathbf{E}(\boldsymbol{\Omega})\frac{\partial \mathbf{c}_\alpha}{\partial \mathbf{q}}\dot{\mathbf{q}} = \\ &= \mathcal{B}_\alpha^{(T)}\mathbf{V}_{CM} + \mathcal{B}_\alpha^{(R)}\boldsymbol{\omega} + \mathcal{B}_\alpha^{(I)}\dot{\mathbf{q}} = \\ &= \mathcal{B}_\alpha\dot{\mathbf{Q}} \end{aligned} \quad (\text{B3})$$

The derivatives of atoms internal Cartesian coordinates, \mathbf{c}_α with respect to the natural internal coordinates (\mathbf{q} , distances, bond angles and torsion angles) are carried out analytically. Collecting all of the atoms

$$\begin{pmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \\ \vdots \\ \mathbf{v}_N \end{pmatrix} = \begin{pmatrix} \mathcal{B}_1^{(T)} & \mathcal{B}_1^{(R)} & \mathcal{B}_1^{(I)} \\ \mathcal{B}_2^{(T)} & \mathcal{B}_2^{(R)} & \mathcal{B}_2^{(I)} \\ \vdots & \vdots & \vdots \\ \mathcal{B}_N^{(T)} & \mathcal{B}_N^{(R)} & \mathcal{B}_N^{(I)} \end{pmatrix} \begin{pmatrix} \mathbf{V}_{CM} \\ \boldsymbol{\omega} \\ \dot{\mathbf{q}} \end{pmatrix} \quad (\text{B4})$$

The \mathcal{B} matrix here defined can be used to calculate the generalized friction tensor as

$$\boldsymbol{\xi}_Q = \mathcal{B}^{tr}\boldsymbol{\xi}_x\mathcal{B} \quad (\text{B5})$$

where $\boldsymbol{\xi}_x$ is the model translational friction for a collection of spheres. If we neglect hydrodynamics interactions and assume that the same effective radius, R_e , is assigned to all the

atoms the model friction tensor for a number N of atoms reads

$$\boldsymbol{\xi}_x = \pi C R_e \eta \mathbf{1}_{3N} = \xi_0 \mathbf{1}_{3N} \quad (\text{B6})$$

and the generalized friction tensor is calculated as

$$\boldsymbol{\xi}_Q = \xi_0 \mathbf{B}^{tr} \mathbf{B} \quad (\text{B7})$$

ACKNOWLEDGMENTS

A.P. and M.Z. thank Fondazione della Cassa di Risparmio di Padova e Rovigo (Project: Modeling and Monitoring Motions in Proteins-M₃PC) and the Italian Ministry of Education, of University and of Research (Project: Prin 2012T9XHH7). Computational work has been carried out on the C3P (Computational Chemistry Community in Padua) HPC facility of the Department of Chemical Sciences of the University of Padua.

REFERENCES

- ¹Y. C. Kim, C. Tang, G. M. Clore, and G. Hummer, “Replica exchange simulations of transient encounter complexes in protein-protein association,” *P. Natl. Acad. Sci. USA* **105**, 12855–12860 (2008).
- ²T. Mittag, L. E. Kay, and J. D. Forman-Kay, “Protein dynamics and conformational disorder in molecular recognition,” *J. Mol. Recognit.* **23**, 105–116 (2010).
- ³J. Cavanagh, W. J. Fairbrother, I. A. G. Palmer, N. J. Skelton, and M. Rance, *Protein NMR Spectroscopy: Principles and Practice* (Elsevier Science, 2010).
- ⁴K. Loth, P. Pelulessy, and G. Bodenhausen, “Chemical shift anisotropy tensors of carbonyl, nitrogen, and amide proton nuclei in proteins through cross-correlated relaxation in nmr spectroscopy,” *J. Am. Chem. Soc.* **127**, 6062–6068 (2005).
- ⁵D. Sheppard, D.-W. Li, R. Godoy-Ruiz, R. Brueschweiler, and V. Tugarinov, “Variation in quadrupole couplings of a deuterons in ubiquitin suggests the presence of c-alpha-h-alpha center dot center dot center dot o=c hydrogen bonds,” *J. Am. Chem. Soc.* **132**, 7709–7719 (2010).

- ⁶E. Bucci and R. F. Steiner, “Anisotropy decay of fluorescence as an experimental approach to protein dynamics,” *Biophys. Chem.* **30**, 199–224 (1988).
- ⁷B. Vergani, M. Kintrup, W. Hillen, H. Lami, E. Piemont, E. Bombarda, P. Alberti, S. M. Doglia, and M. Chabbert, “Backbone dynamics of tet repressor alpha 8 boolean and alpha 9 loop,” *Biochemistry-US* **39**, 2759–2768 (2000).
- ⁸W. L. Hubbell, H. S. Mchaourab, C. Altenbach, and M. A. Lietzow, “Watching proteins move using site-directed spin labeling,” *Structure* **4**, 779–783 (1996).
- ⁹Z. Zhang, M. R. Fleissner, D. S. Tipikin, Z. Liang, J. K. Moscicki, K. A. Earle, W. L. Hubbell, and J. H. Freed, “Multifrequency electron spin resonance study of the dynamics of spin labeled t4 lysozyme,” *J. Phys. Chem. B* **114**, 5503–5521 (2010).
- ¹⁰H. Hofmann, F. Hillger, S. H. Pfeil, A. Hoffmann, D. Streich, D. Haenni, D. Nettels, E. A. Lipman, and B. Schuler, “Single-molecule spectroscopy of protein folding in a chaperonin cage,” *P. Natl. Acad. Sci. USA* **107**, 11793–11798 (2010).
- ¹¹L. Hamon, D. Pastre, P. Dupaigne, C. L. Breton, E. L. Cam, and O. Pietrement, “High-resolution afm imaging of single-stranded dna-binding (ssb) protein-dna complexes,” *Nucleic Acids Res.* **35**, e58 (2007).
- ¹²J. Kowalewski and L. Maler, *Nuclear Spin Relaxation in Liquids: Theory, Experiments, And Applications* (Taylor & Francis, 2006).
- ¹³G. Lipari and A. Szabo, “Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. theory and range of validity,” *J. Am. Chem. Soc.* **104**, 4546–4559 (1982).
- ¹⁴G. Lipari and A. Szabo, “Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. analysis of experimental results,” *J. Am. Chem. Soc.* **104**, 4559–4570 (1982).
- ¹⁵Y. E. Shapiro, E. Kahana, V. Tugarinov, Z. C. Liang, J. H. Freed, and E. Meirovitch, “Domain flexibility in ligand-free and inhibitor-bound escherichia coli adenylate kinase based on a mode-coupling analysis of n-15 spin relaxation,” *Biochemistry-US* **41**, 6271–6281 (2002).
- ¹⁶E. Meirovitch, Y. E. Shapiro, A. Polimeno, and J. H. Freed, “An improved picture of methyl dynamics in proteins from slowly relaxing local structure analysis of h-2 spin relaxation,” *J. Phys. Chem. B* **111**, 12865–12875 (2007).
- ¹⁷A. Abragam, *The Principles of Nuclear Magnetism* (Clarendon Press, 1961).

- ¹⁸J. W. Peng and G. Wagner (Academic Press, 1994) pp. 563–595.
- ¹⁹D. Abergel and G. Bodenhausen, “Predicting internal protein dynamics from structures using coupled networks of hindered rotators,” *J. Chem. Phys.* **123**, 204901 (2005).
- ²⁰A. Dhulesia, G. Bodenhausen, and D. Abergel, “Predicting conformational entropy of bond vectors in proteins by networks of coupled rotators,” *J. Chem. Phys.* **129**, 095107 (2008).
- ²¹V. Calandrini, D. Abergel, and G. R. Kneller, “Fractional protein dynamics seen by nuclear magnetic resonance spectroscopy: Relating molecular dynamics simulation and experiment,” *J. Chem. Phys.* **133**, 145101 (2010).
- ²²R. O. Dror, R. M. Dirks, J. P. Grossman, H. Xu, and D. E. Shaw, “Biomolecular Simulation: A Computational Microscope for Molecular Biology,” in *Annu. Rev. Biophys.*, Annual Review of Biophysics, Vol. 41, edited by Rees, DC (2012) pp. 429–452.
- ²³D. E. Shaw, M. M. Deneroff, R. O. Dror, J. S. Kuskin, R. H. Larson, J. K. Salmon, C. Young, B. Batson, K. J. Bowers, J. C. Chao, M. P. Eastwood, J. Gagliardo, J. P. Grossman, C. R. Ho, D. J. Ierardi, I. Kolossvary, J. L. Klepeis, T. Layman, C. Mcleavey, M. A. Moraes, R. Mueller, E. C. Priest, Y. Shan, J. Spengler, M. Theobald, B. Towles, and S. C. Wang, “Anton, a special-purpose machine for molecular dynamics simulation,” *Commun. ACM* **51**, 91–97 (2008).
- ²⁴K. A. Beauchamp, R. McGibbon, Y.-S. Lin, and V. S. Pande, “Simple few-state models reveal hidden complexity in protein folding,” *Proc. Natl. Acad. Sci. U.S.A.* **109**, 17807–17813 (2012).
- ²⁵H. Nguyen, J. Maier, H. Huang, V. Perrone, and C. Simmerling, “Folding Simulations for Proteins with Diverse Topologies Are Accessible in Days with a Physics-Based Force Field and Implicit Solvent,” *J. Am. Chem. Soc.* **136**, 13959–13962 (2014).
- ²⁶M. Gur, E. Zomot, and I. Bahar, “Global motions exhibited by proteins in micro- to milliseconds simulations concur with anisotropic network model predictions,” *J. Chem. Phys.* **139** (2013), 10.1063/1.4816375.
- ²⁷O. Fiset, P. Laguerre, S. Gagne, and S. Morin, “Synergistic Applications of MD and NMR for the Study of Biological Systems,” *J. Biomed. Biotechnol.* (2012), 10.1155/2012/254208.
- ²⁸T. Soares, X. Daura, C. Oostenbrink, L. Smith, and W. van Gunsteren, “Validation of the GROMOS force-field parameter set 45A3 against nuclear magnetic resonance data of hen egg lysozyme,” *J. Biomol. NMR* **30**, 407–422 (2004).

- ²⁹K. Lindorff-Larsen, P. Maragakis, S. Piana, M. P. Eastwood, R. O. Dror, and D. E. Shaw, “Systematic Validation of Protein Force Fields against Experimental Data,” *PLOS One* **7** (2012), 10.1371/journal.pone.0032131.
- ³⁰J. S. Anderson, G. Hernandez, and D. M. LeMaster, “Prediction of Bond Vector Autocorrelation Functions from Larmor Frequency-Selective Order Parameter Analysis of NMR Relaxation Data,” *J. Chem. Theory Comput.* **13**, 3276–3289 (2017).
- ³¹F. Hoffmann, F. A. A. Mulder, and L. V. Schaefer, “Accurate Methyl Group Dynamics in Protein Simulations with AMBER Force Fields,” *J. Phys. Chem. B* **122**, 5038–5048 (2018).
- ³²A. Polimeno and J. H. Freed, “Slow Motional ESR in Complex Fluids: The Slowly Relaxing Local Structure Model of Solvent Cage Effects,” *J. Phys. Chem.* **99**, 10995–11006 (1995).
- ³³E. Meirovitch, A. Polimeno, and J. H. Freed, *J. Phys. Chem. B* **110**, 20615 (2006).
- ³⁴M. Zerbetto, A. Polimeno, and E. Meirovitch, “General Theoretical/Computational Tool for Interpreting NMR Spin Relaxation in Proteins,” *J. Phys. Chem. B* **113**, 13613–13625 (2009).
- ³⁵M. Zerbetto, M. Buck, E. Meirovitch, and A. Polimeno, “Integrated Computational Approach to the Analysis of NMR Relaxation in Proteins: Application to ps-ns Main Chain $^{15}\text{n-}^1\text{h}$ and Global Dynamics of the Rho GTPase Binding Domain of Plexin B1,” *J. Phys. Chem. B* **115**, 376–388 (2011).
- ³⁶M. Zerbetto, D. Kotsyubynskyy, J. Kowalewski, G. Widmalm, and A. Polimeno, “Stochastic Modeling of Flexible Biomolecules Applied to NMR Relaxation. I. Internal Dynamics of Cyclodextrins: β -Cyclodextrin as a Case Study,” *J. Phys. Chem. B* **116**, 13159–13171 (2012).
- ³⁷A. Perico and M. Guenza, “Viscoelastic relaxation of segment orientation in dilute polymer solutions,” *J. Chem. Phys.* **83**, 3103–3109 (1985).
- ³⁸A. Perico and M. Guenza, “Viscoelastic relaxation of segment orientation in dilute polymer solutions. ii. stiffness dependence of fluorescence depolarization,” *J. Chem. Phys.* **84**, 510–516 (1985).
- ³⁹K. S. Kostov, K. F. Freed, E. B. Web, M. Mondello, and G. Greta, “Dynamics of linear and branched alkane melts: Molecular dynamics test of theory for long time dynamics,” *J. Chem. Phys.* **108**, 9155–9167 (1998).

- ⁴⁰J. J. Prompers and R. Brüschweiler, “Reorientational eigenmode dynamics: A combined md/nmr relaxation analysis method for flexible parts in globular proteins,” *J. Am. Chem. Soc.* **123**, 7305–7313 (2001).
- ⁴¹J. J. Prompers and R. Brüschweiler, “General framework for studying the dynamics of folded and nonfolded proteins by nmr relaxation spectroscopy and md simulation,” *J. Am. Chem. Soc.* **124**, 4522–4534 (2002).
- ⁴²Y. Gu, D.-W. Li, and R. Brüschweiler, “Nmr order parameter determination from long molecular dynamics trajectories for objective comparison with experiment,” *J. Chem. Theory Comput.* **10**, 2599–2607 (2014).
- ⁴³S. Miyazawa and R. L. Jernigan, “Estimation of effective interresidue contact energies from protein crystal-structures - quasi-chemical approximation,” *Macromolecules* **18**, 534–552 (1985).
- ⁴⁴B. Erman, “The gaussian network model: Precise prediction of residue fluctuations and application to binding problems,” *Biophys. J.* **91**, 3589–3599 (2006).
- ⁴⁵T. Haliloglu, I. Bahar, and B. Erman, “Gaussian dynamics of folded proteins,” *Phys. Rev. Lett.* **79**, 3090–3093 (1997).
- ⁴⁶A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar, “Anisotropy of fluctuation dynamics of proteins with an elastic network model,” *Biophys. J.* **80**, 505–515 (2001).
- ⁴⁷E. Eyal, L.-W. Yang, and I. Bahar, “Anisotropic network model: systematic evaluation and a new web interface,” *Bioinformatics* **22**, 2619–2627 (2006).
- ⁴⁸N. Go, T. Noguti, and T. Nishikawa, “Dynamics of a small globular protein in terms of low-frequency vibrational modes.” *P. Natl. Acad. Sci. USA* **80**, 3696–700 (1983).
- ⁴⁹B. Brooks and M. Karplus, “Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor.” *P. Natl. Acad. Sci. USA* **80**, 6571–5 (1983).
- ⁵⁰I. Bahar and A. J. Rader, “Coarse-grained normal mode analysis in structural biology,” *Curr. Opin. Struc. Biol.* **15**, 586–592 (2005).
- ⁵¹E. Caballero-Manrique, J. K. Bray, W. A. Deutschman, F. W. Dahlquist, and M. G. Guenza, “A theory of protein dynamics to predict nmr relaxation,” *Biophys. J.* **93**, 4128–4140 (2007).
- ⁵²J. Copperman and M. G. Guenza, “Predicting protein dynamics from structural ensembles,” *J. Chem. Phys.* **143**, 243131 (2015).

- ⁵³H. Goldstein, *Classical Mechanics* (Addison-Wesley, 1980).
- ⁵⁴R. Zwanzig, “On the identity of three generalized master equations,” *Physica* **30**, 1109–1123 (1964).
- ⁵⁵R. Zwanzig, “Nonlinear generalized langevin equations,” *J. Stat. Phys.* **9**, 215–220 (1973).
- ⁵⁶A. Polimeno, M. Zerbetto, and D. Abergel, “Stochastic modeling of macromolecules in solution. ii. spectral densities,” Submitted.
- ⁵⁷R. N. Zare, *Angular momentum: understanding spatial aspects in chemistry and physics* (Wiley, 1988).
- ⁵⁸J. K. Watson, “Simplification of the molecular vibration-rotation hamiltonian,” *Mol. Phys.* **15**, 479–490 (1968).
- ⁵⁹J. K. Watson, “The vibration-rotation of linear molecules,” *Mol. Phys.* **19**, 465–487 (1970).
- ⁶⁰J. Louck, “Derivation of the molecular vibration-rotation hamiltonian for the schrödinger equation for the molecular model,” *J. Mol. Spectrosc.* **61**, 107–137 (1976).
- ⁶¹E. B. Wilson, J. C. Decius, and P. C. Cross, *Molecular Vibrations: The Theory of Infrared and Raman Vibrational Spectra* (Dover Publications, 1955).
- ⁶²S. Califano, *Vibrational states* (Wiley, 1976).
- ⁶³W. D. Allen and A. G. Csaszar, “On the ab initio determination of higher-order force-constants at nonstationary reference geometries,” *J. Chem. Phys.* **98**, 2983–3015 (1993).
- ⁶⁴A. A. Shabana, *Dynamics of Multibody Systems* (Cambridge University Press, 2005).
- ⁶⁵H. Meyer, “The molecular hamiltonian,” *Annu. Rev. Phys. Chem.* **53**, 141–172 (2002).
- ⁶⁶D. Lauvergnat and A. Nauts, “Exact numerical computation of a kinetic energy operator in curvilinear coordinates,” *J. Chem. Phys.* **116**, 8560–8570 (2002).
- ⁶⁷R. G. Littlejohn and M. Reinsch, “Gauge fields in the separation of rotations and internal motions in the n -body problem,” *Rev. Mod. Phys.* **69**, 213–275 (1997).
- ⁶⁸H. Risken, *The Fokker-Planck Equation: Methods of Solution and Applications* (Springer-Verlag).
- ⁶⁹J. T. Hynes and J. Deutch, “Non-equilibrium problems - projection operator techniques,” in *Physical Chemistry - An Advanced Treatise*, Vol. XIB, edited by H. Eyring, D. Henderson, and W. Jost (Academic Press, 1975).
- ⁷⁰J. Rotne and S. Prager, “Variational treatment of hydrodynamic interaction in polymers,” *J. Chem. Phys.* **50**, 4831–4837 (1969).

- ⁷¹V. Barone, M. Zerbetto, and A. Polimeno, “Hydrodynamic modeling of diffusion tensor properties of flexible molecules,” *J. Comput. Chem.* **30**, 2–13 (2009).
- ⁷²G. J. Moro, “A stochastic model for crankshaft transitions,” *J. Phys. Chem.* **100**, 16419–16422 (1996).
- ⁷³B. Nigro and G. J. Moro, “A stochastic model for crankshaft transitions. ii. analysis of transition dynamics,” *J. Phys. Chem. B* **106**, 7365–7375 (2002).
- ⁷⁴K. Hinsen, “The molecular modeling toolkit: A new approach to molecular simulations,” *J. Comput. Chem.* **21**, 79–85 (2000).
- ⁷⁵A. Case David, E. Cheatham Thomas, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, “The amber biomolecular simulation programs,” *J. Comput. Chem.* **26**, 1668–1688 (2005).
- ⁷⁶J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, K. Laxmikant, and K. Schulten, “Scalable molecular dynamics with namd,” *J. Comput. Chem.* **26**, 1781–1802 (2005), <https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.20289>.
- ⁷⁷K. Kasahara, S. Sakuraba, and I. Fukuda, “Enhanced sampling of molecular dynamics simulations of a polyalanine octapeptide: Effects of the periodic boundary conditions on peptide conformation,” *J. Phys. Chem. B* **122**, 2495–2503 (2018).
- ⁷⁸J. Graf, P. H. Nguyen, G. Stock, and H. Schwalbe, “Structure and dynamics of the homologous series of alanine peptides. a joint molecular dynamics/nmr study,” *J. Am. Chem. Soc.* **129**, 1179–1189 (2007).
- ⁷⁹S. Park, F. Khalili-Araghi, E. Tajkhorshid, and K. Schulten, “Free energy calculation from steered molecular dynamics simulations using jarzynski equality,” *J. Chem. Phys.* **119**, 3559–3566 (2003).
- ⁸⁰A. Hazel, C. Chipot, and J. C. Gumbart, “Thermodynamics of deca-alanine folding in water,” *J. Chem. Theory Comput.* **10**, 2836–2844 (2014).
- ⁸¹B. Matthias, S. Bouguet-Bonnet, R. Pastor, and A. D. J. MacKerell, “Importance of the cmap correction to the charmm22 protein force field: Dynamics of hen lysozyme,” *Biophys. J.* **90**, 36–38 (2006).
- ⁸²J. Waser, “Dyadics and the variances and covariances of molecular parameters, including those of best planes,” *Acta Crystallogr. A* **29**, 621–631, <https://onlinelibrary.wiley.com/doi/pdf/10.1107/S0567739473001609>.

- ⁸³V. Barone, M. Brustolon, P. Cimino, A. Polimeno, M. Zerbetto, and A. Zoleo, “Development and validation of an integrated computational approach for the modeling of cw-esr spectra of free radicals in solution: p-(methylthio)phenyl nitronyl nitroxide in toluene as a case study,” *J. Am. Chem. Soc.* **128**, 15865–15873 (2006).
- ⁸⁴A. Piserchia, M. Zerbetto, M.-V. Salvia, G. Salassa, L. Gabrielli, F. Mancin, F. Rastrelli, and D. Frezzato, “Conformational mobility in monolayer-protected nanoparticles: From torsional free energy profiles to nmr relaxation,” *J. Phys. Chem. C* **119**, 20100–20110 (2015).
- ⁸⁵M. Zerbetto, T. Angles d’Ortoli, A. Polimeno, and G. Widmalm, “Differential dynamics at glycosidic linkages of an oligosaccharide as revealed by ¹³C nmr spin relaxation and stochastic modeling,” *J. Phys. Chem. B* **122**, 2287–2294 (2018).
- ⁸⁶M. Zerbetto, S. Carlotto, A. Polimeno, C. Corvaja, L. Franco, C. Toniolo, F. Formaggio, V. Barone, and P. Cimino, “Ab initio modeling of cw-esr spectra of the double spin labeled peptide fmoc-(aib-aib-toac)₂-aib-ome in acetonitrile,” *J. Phys. Chem. B* **111**, 2668–2674 (2007).
- ⁸⁷J. Campeggio, A. Polimeno, and M. Zerbetto, “Dite2: Calculating the diffusion tensor for flexible molecules,” *J. Comput. Chem.* **40**, 697–705 (2019).

FIG. 1. Example of reference frames for protein GB3: μF is chosen as the dipolar frame on the N-H group of the peptide bond between E24 and T25.

FIG. 2. Optimized molecular geometries of the sequence of polyalanines $(\text{ALA}-)_n$, with $n = 2, \dots, 10$

FIG. 3. Principal values of inertia tensors of the set of the series $(\text{ALA}-)_n$, with $n = 2, \dots, 10$: squares x axis values, circles y axis values, diamonds z axis values

FIG. 4. Color map of $\log |(\omega_K)_{ij}|$ for the series $(\text{ALA}-)_n$, with $n = 2, \dots, 10$; ω_K is calculated via the direct estimate of the Hessian of internal energy in the references geometries.

FIG. 5. Histograms of $\log |(\omega_K)_{ij}|$ for the series $(\text{ALA}-)_n$, with $n = 2, \dots, 10$; ω_K is calculated via the direct estimate of the Hessian of internal energy in the references geometries.

FIG. 6. Color map of $\log |(\omega_K)_{ij}|$ for the series $(\text{ALA}-)_n$, with $n = 2, \dots, 10$; ω_K is calculated via the inversion of the covariance matrix.

FIG. 7. Histograms of $\log |(\omega_K)_{ij}|$ for the series $(\text{ALA}-)_n$, with $n = 2, \dots, 10$; ω_K is calculated via the inversion of the covariance matrix.

FIG. 8. Color map of $\log |(\omega_\xi)_{ij}|$ for the series $(\text{ALA}-)_n$, with $n = 2, \dots, 10$; cfr. Appendix B for the evaluation of ξ .

FIG. 9. Histograms of $\log |(\omega_\xi)_{ij}|$ for the series $(\text{ALA}-)_n$, with $n = 2, \dots, 10$; cfr. Appendix B for the evaluation of ξ .

FIG. 10. Histograms of $\log |(\omega^{\text{int}})_{ij}|$ for the series $(\text{ALA}-)_n$, with $n = 2, \dots, 10$; the curvature matrix is calculated via the inversion of the covariance matrix.

FIG. 11. Histograms of $\log |\omega_{ij}^{\text{io}}|$ for the series $(\text{ALA}-)_n$, with $n = 2, \dots, 10$; the curvature matrix is calculated via the inversion of the covariance matrix.

FIG. 12. Histograms of $\log |\omega_K \times \mathbf{fs}|$ calculated via the inversion of the covariance matrix for the 10-alanine peptide simulated in cubic boxes of side length of 40 Å (left panel), 50 Å (middle panel), and 60 Å (right panel).

FIG. 13. Map and histograms of $\log |\omega_K \times \mathbf{fs}|$ for 10-alanine calculated via Hessian (a, b) and the inversion of the covariance matrix (c,d).

FIG. 14. Examples of trajectories for 2-alanine, 4-alanine, 6-alanine and 10-alanine of diagonal friction matrix elements $\log |(\omega_\xi)_{ii}|$ in time: $i = 4$ stretching degree of freedom in red, $i = 5$ bending degree of freedom in green, $i = 6$ torsional degree of freedom in green; trace of rotational part in black.

FIG. 15. Average eigenvalues of ω_ξ for 2-alanine, 4-alanine, 6-alanine and 10-alanine, with standard deviation (vertical bars).

FIG. 16. Summary view of levels of approximations: from the atomistic Liouville equation to the SFB model.

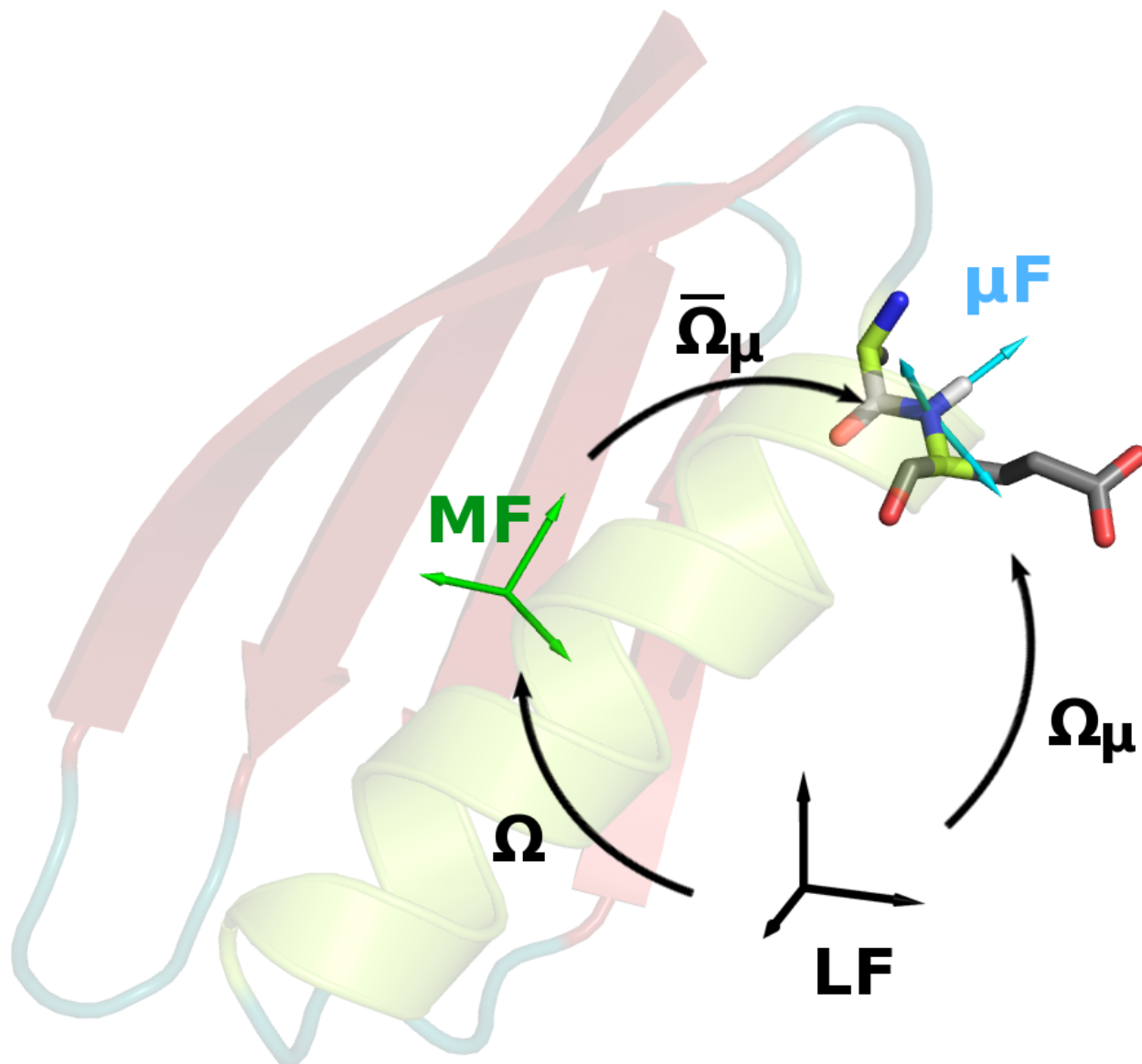


Figure 1

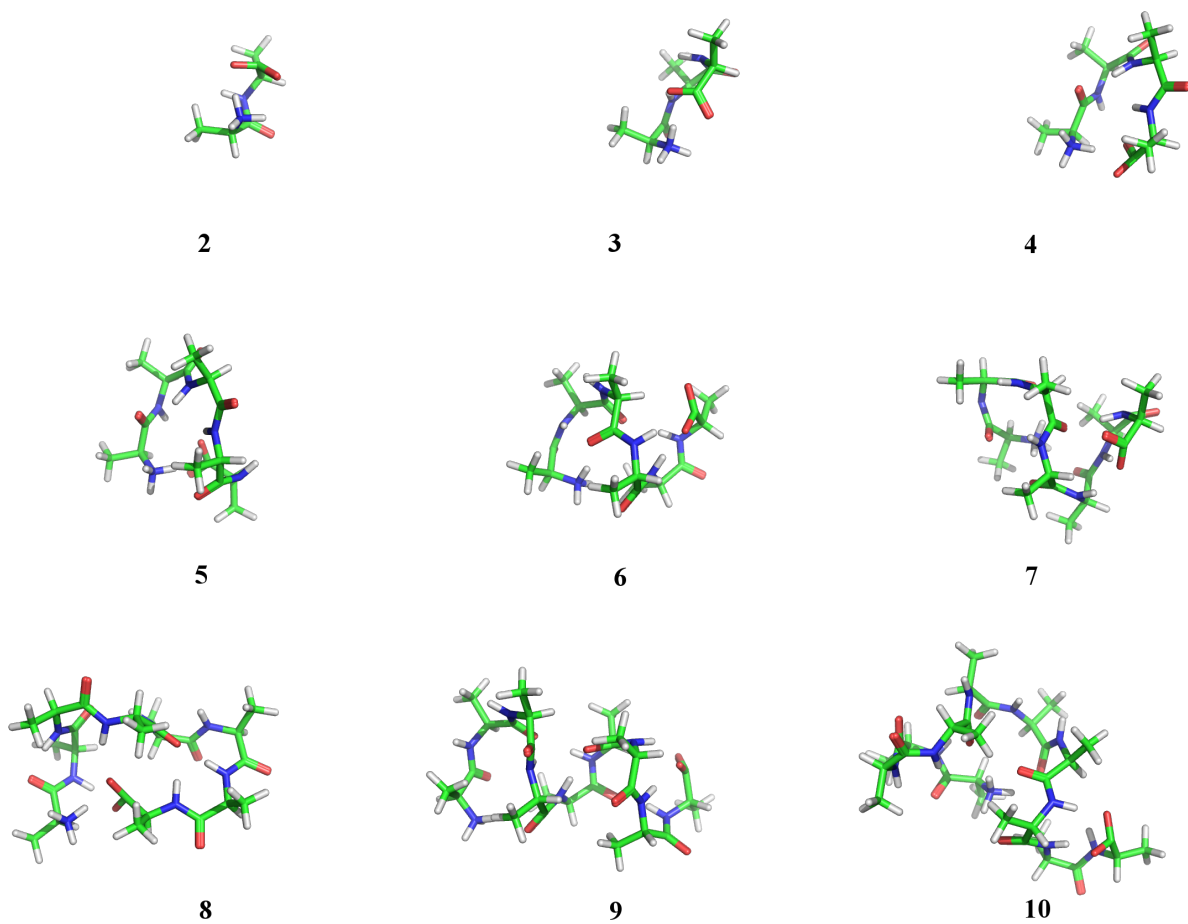


Figure 2

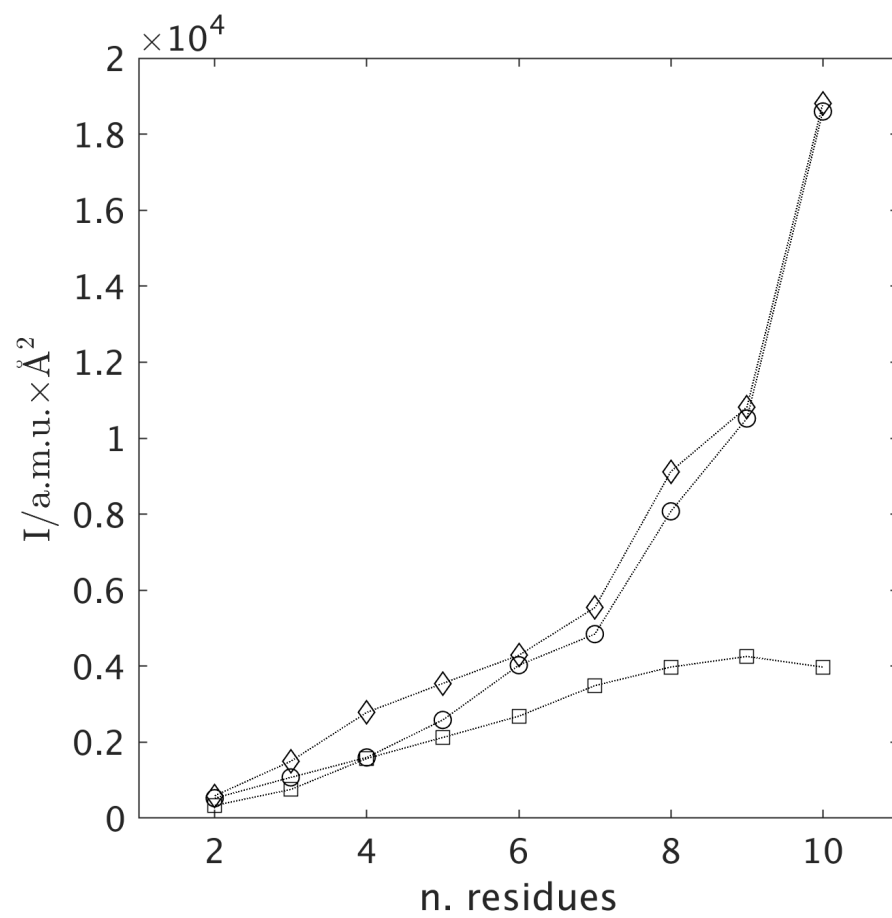


Figure 3

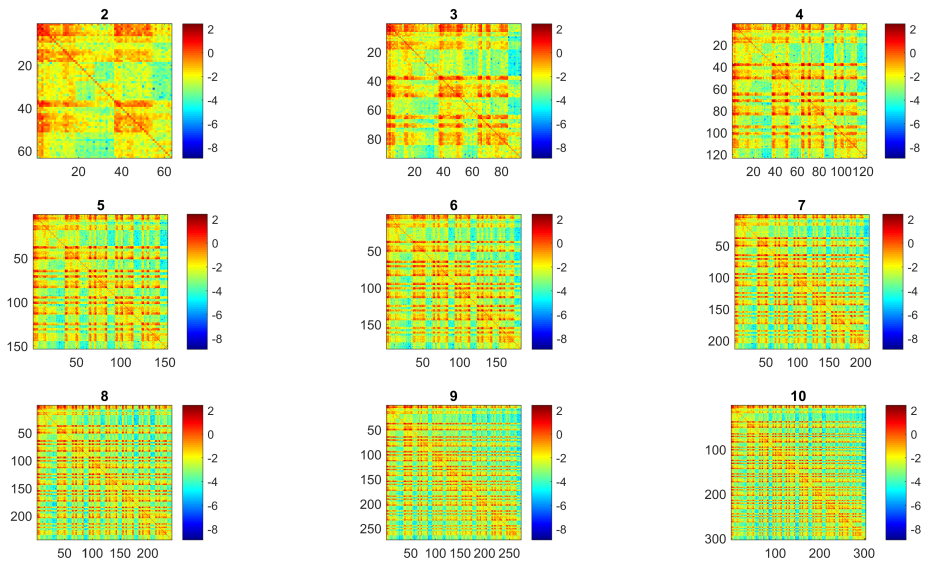


Figure 4

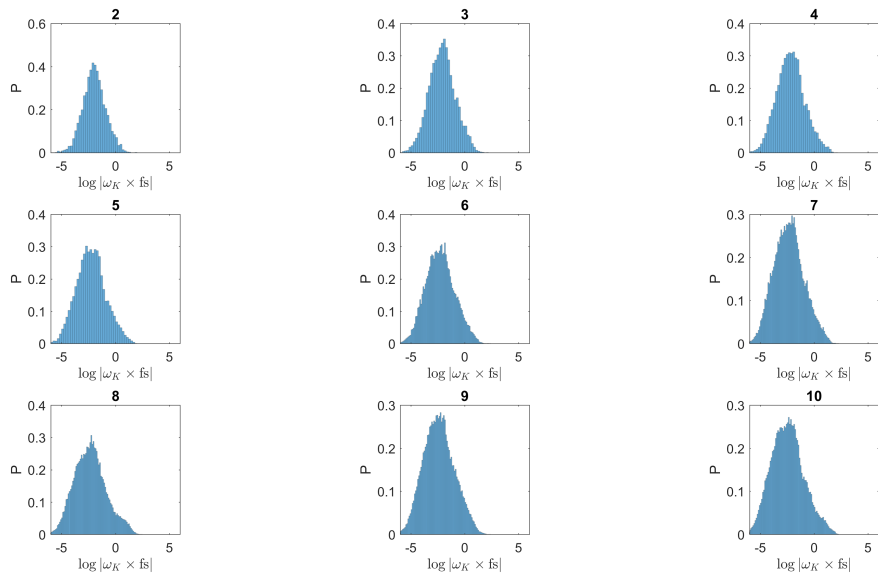


Figure 5

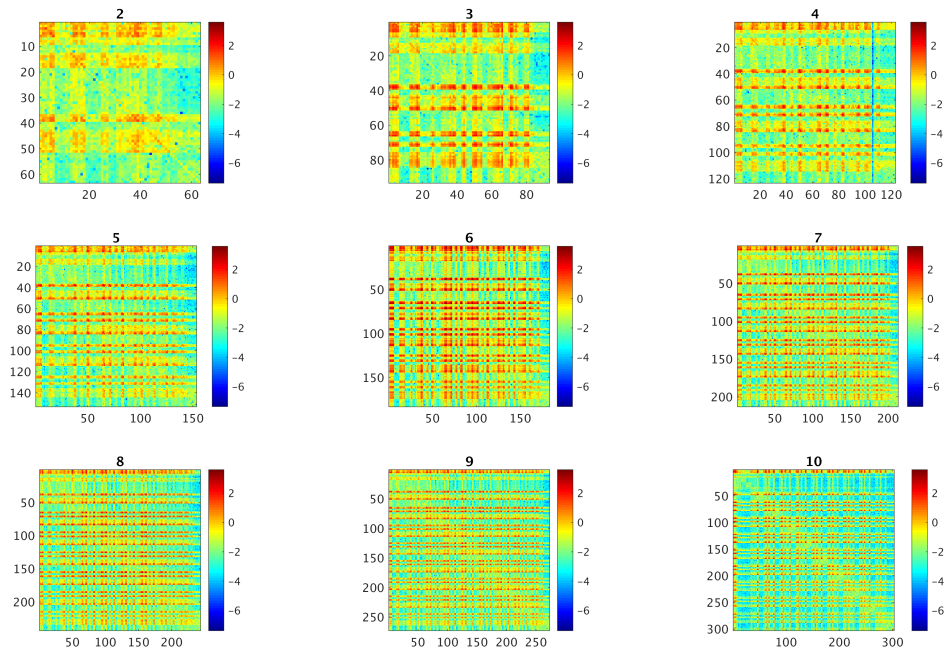


Figure 6

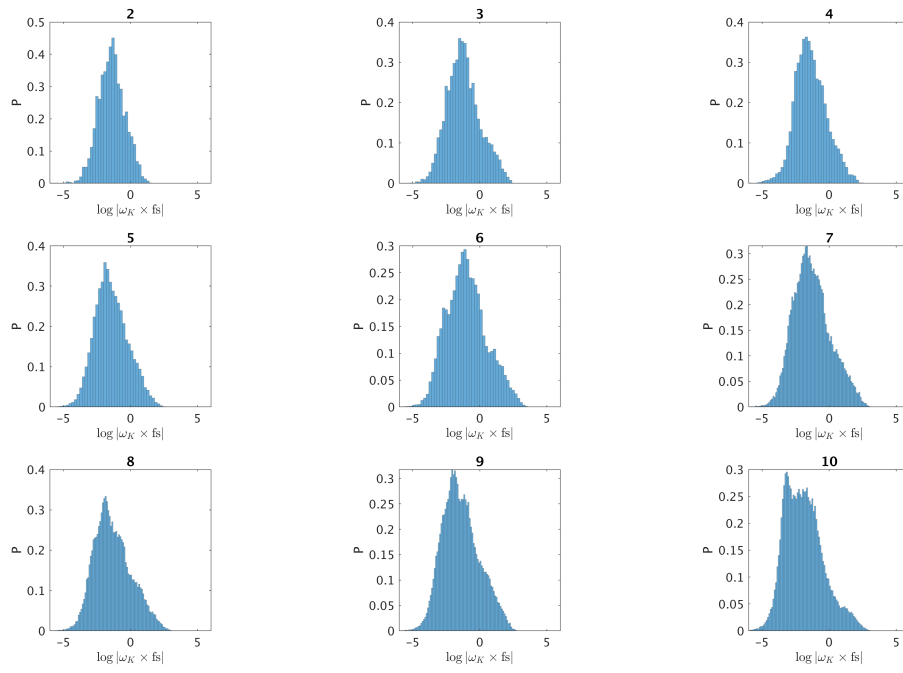


Figure 7

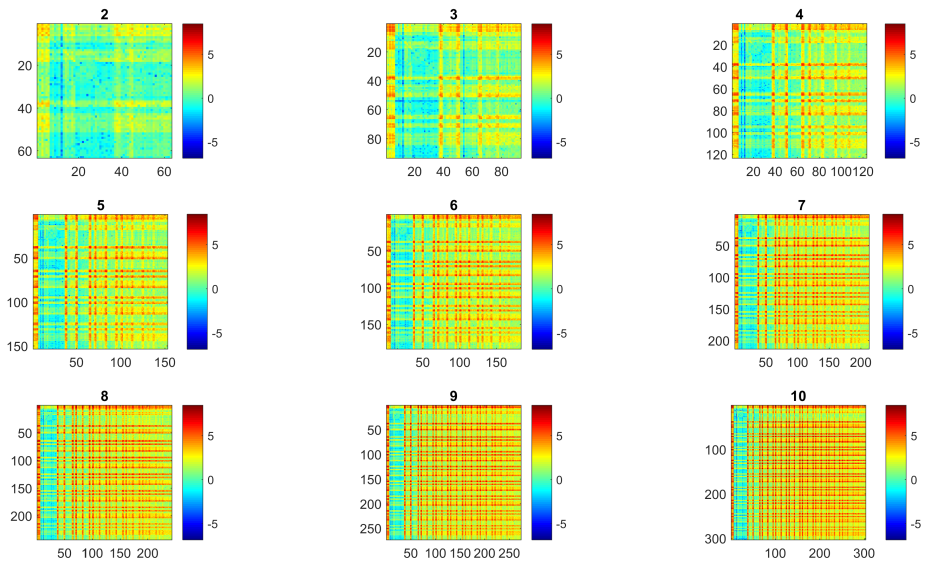


Figure 8

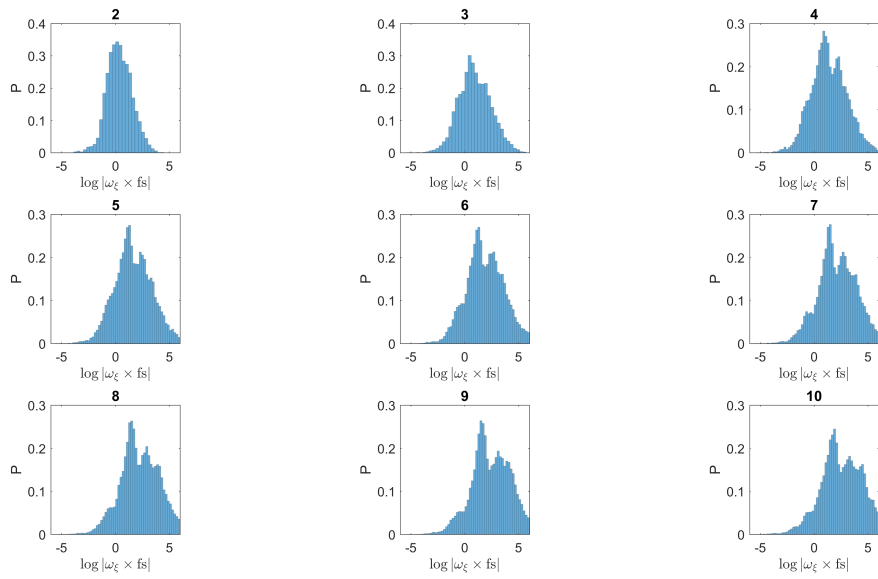


Figure 9

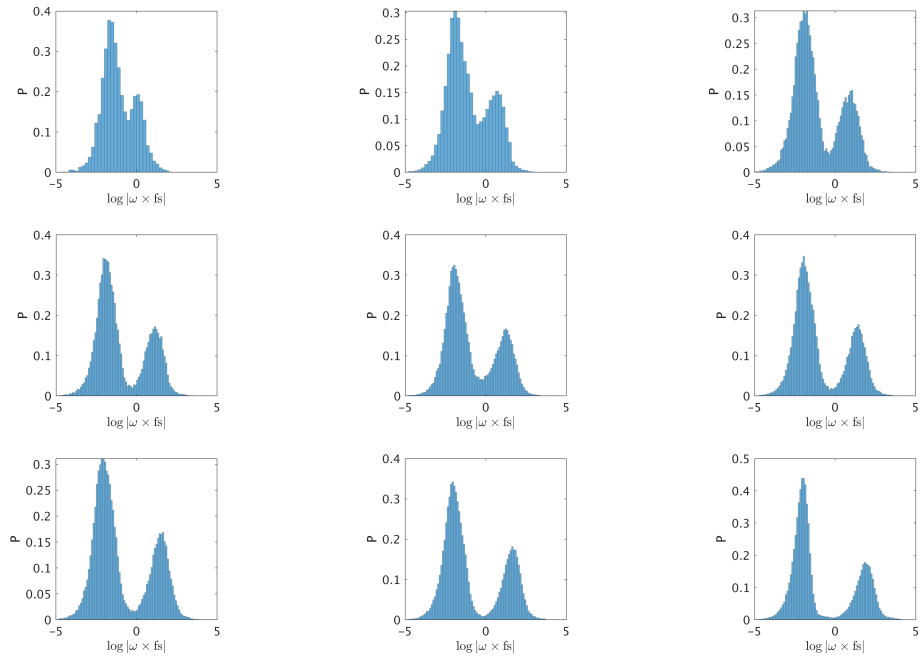


Figure 10

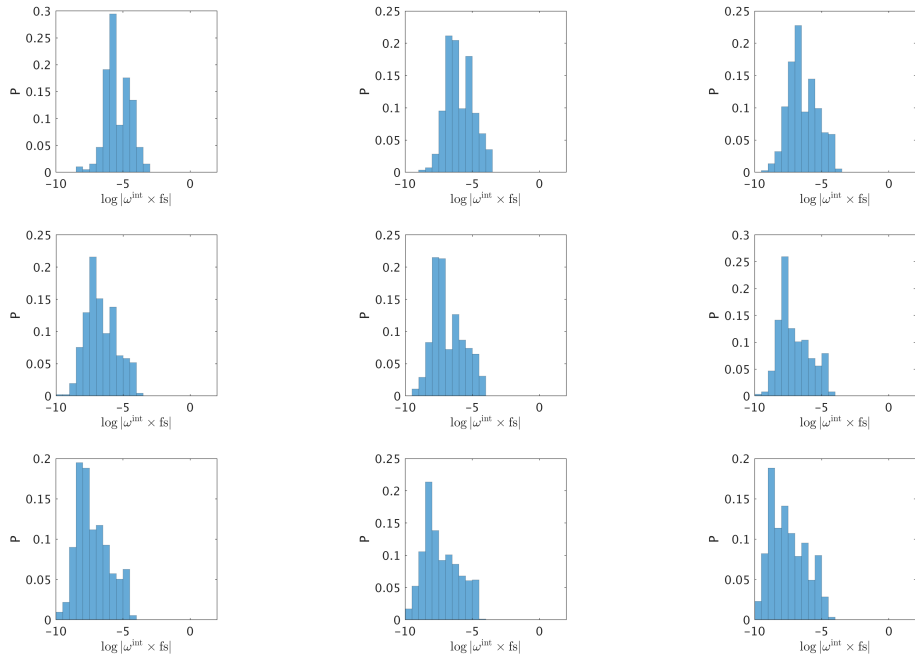


Figure 11

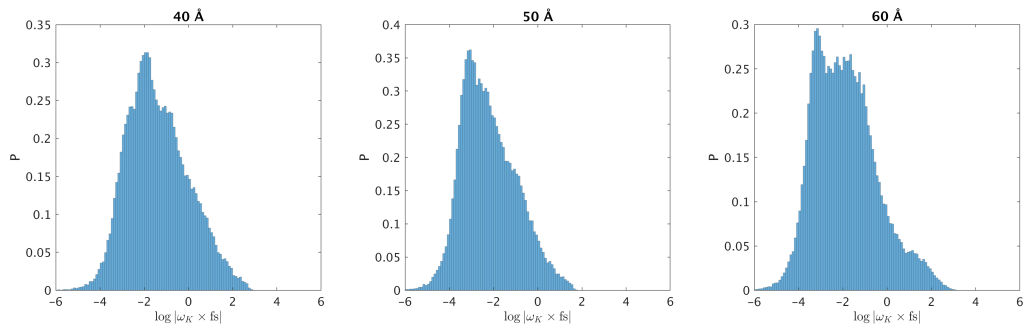
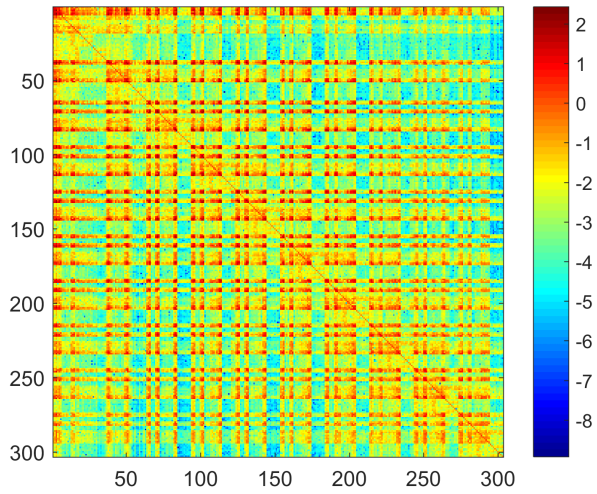
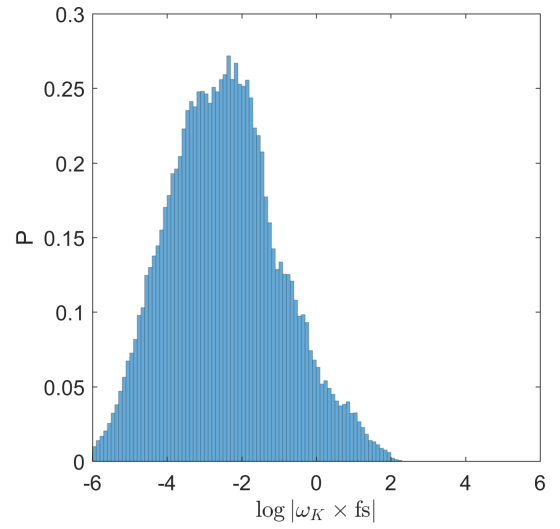


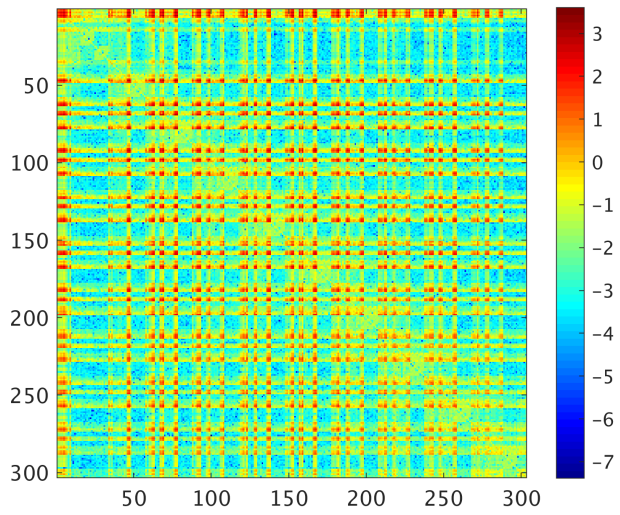
Figure 12



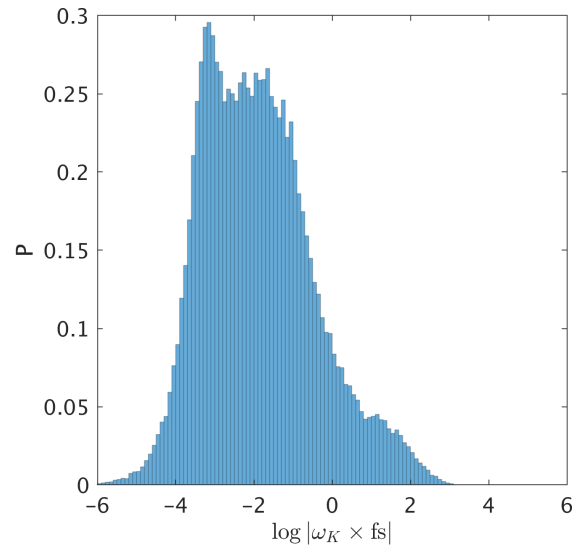
(a)



(b)



(c)



(d)

Figure 13

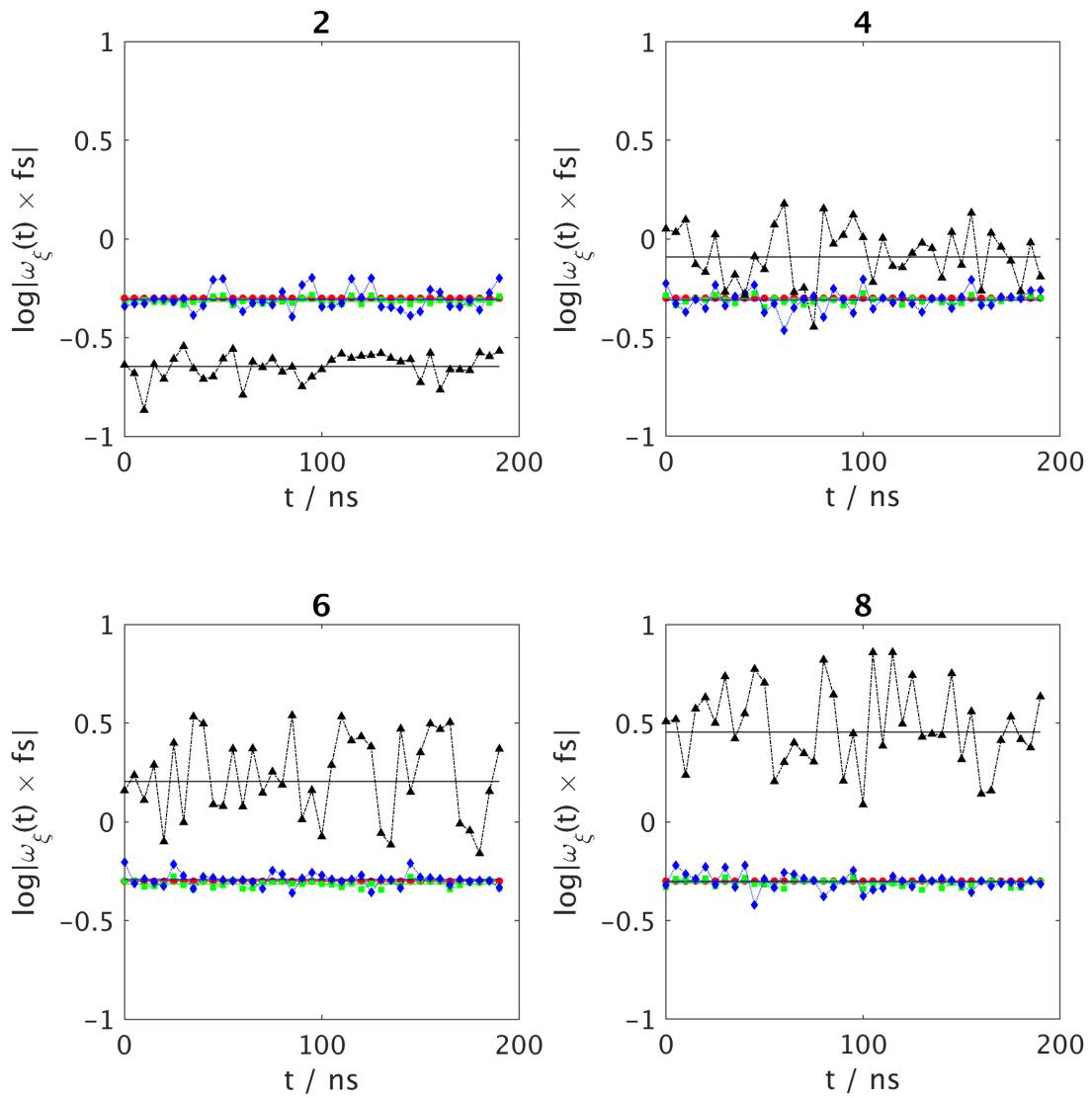


Figure 14

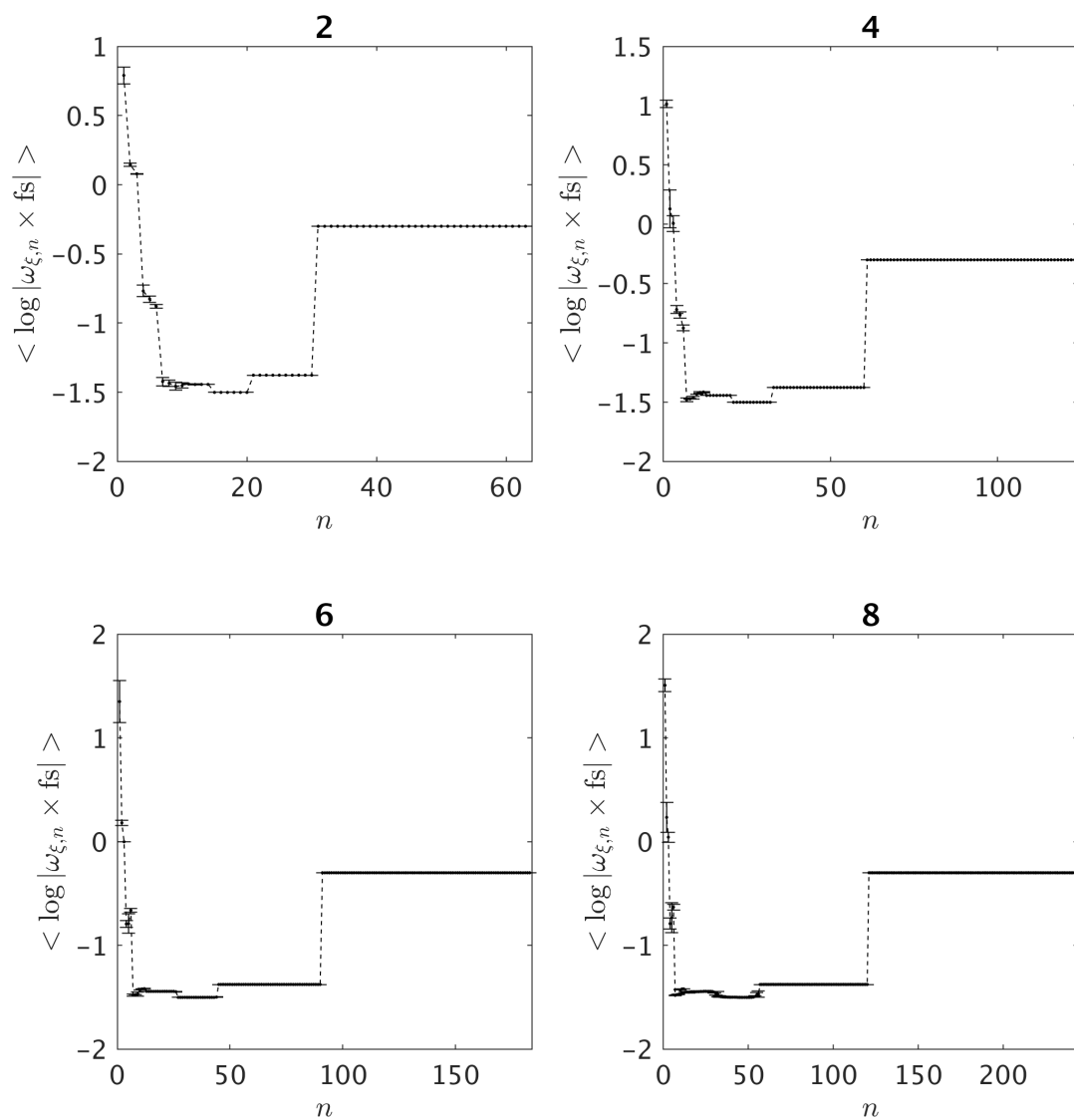


Figure 15

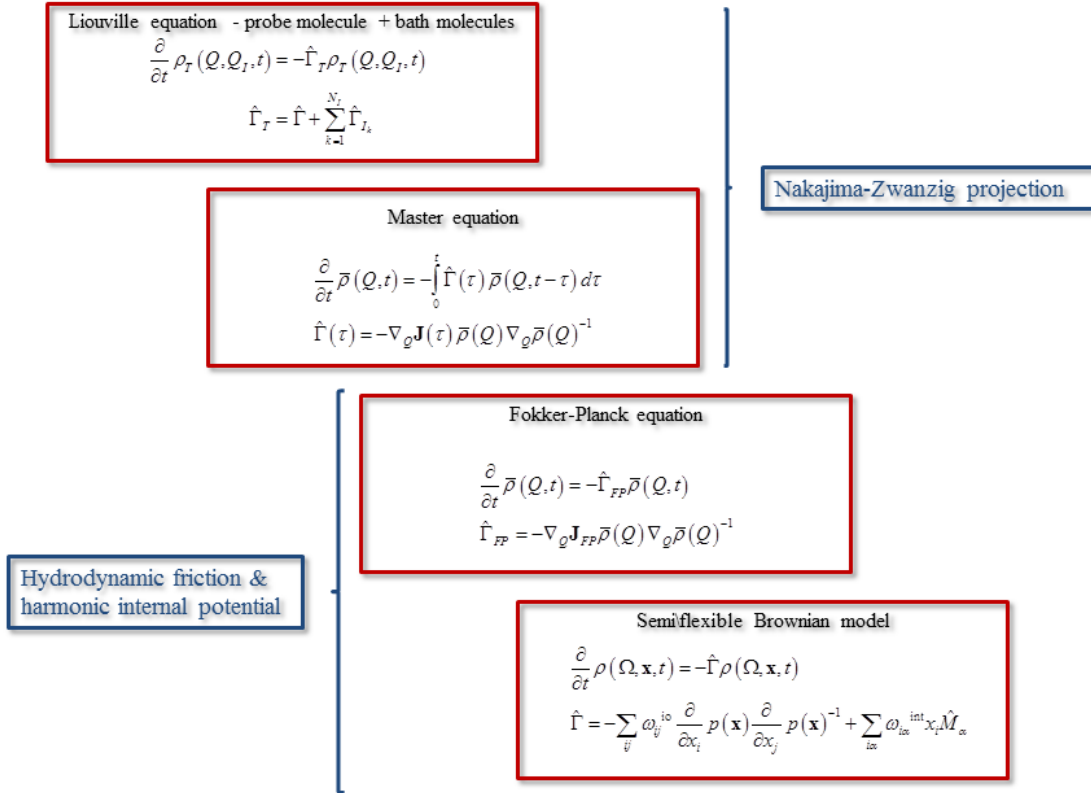


Figure 16