

Dynamic neutron scattering from conformational dynamics. II. Application using molecular dynamics simulation and Markov modeling

Zheng Yi, Benjamin Lindner, Jan-Hendrik Prinz, Frank Noé, and Jeremy C. Smith

Citation: *J. Chem. Phys.* **139**, 175102 (2013); doi: 10.1063/1.4824071

View online: <http://dx.doi.org/10.1063/1.4824071>

View Table of Contents: <http://jcp.aip.org/resource/1/JCPSA6/v139/i17>

Published by the [AIP Publishing LLC](#).

Additional information on *J. Chem. Phys.*

Journal Homepage: <http://jcp.aip.org/>

Journal Information: http://jcp.aip.org/about/about_the_journal

Top downloads: http://jcp.aip.org/features/most_downloaded

Information for Authors: <http://jcp.aip.org/authors>



Re-register for Table of Content Alerts

Create a profile.



Sign up today!



Dynamic neutron scattering from conformational dynamics. II. Application using molecular dynamics simulation and Markov modeling

Zheng Yi,¹ Benjamin Lindner,¹ Jan-Hendrik Prinz,² Frank Noé,² and Jeremy C. Smith¹

¹University of Tennessee/Oak Ridge National Laboratory Center for Molecular Biophysics, P.O. Box 2008, 1 Bethel Valley Road, Oak Ridge, Tennessee 37831, USA and Department of Biochemistry and Cellular and Molecular Biology, University of Tennessee, M407 Walters Life Sciences, 1414 Cumberland Avenue, Knoxville, Tennessee 37996, USA

²Department of Mathematics and Computer Science, FU Berlin, Arnimallee 6, 14159 Berlin, Germany

(Received 5 June 2013; accepted 12 September 2013; published online 1 November 2013)

Neutron scattering experiments directly probe the dynamics of complex molecules on the sub pico- to microsecond time scales. However, the assignment of the relaxations seen experimentally to specific structural rearrangements is difficult, since many of the underlying dynamical processes may exist on similar timescales. In an accompanying article, we present a theoretical approach to the analysis of molecular dynamics simulations with a Markov State Model (MSM) that permits the direct identification of structural transitions leading to each contributing relaxation process. Here, we demonstrate the use of the method by applying it to the configurational dynamics of the well-characterized alanine dipeptide. A practical procedure for deriving the MSM from an MD is introduced. The result is a 9-state MSM in the space of the backbone dihedral angles and the side-chain methyl group. The agreement between the quasielastic spectrum calculated directly from the atomic trajectories and that derived from the Markov state model is excellent. The dependence on the wavevector of the individual Markov processes is described. The procedure means that it is now practicable to interpret quasielastic scattering spectra in terms of well-defined intramolecular transitions with minimal *a priori* assumptions as to the nature of the dynamics taking place. © 2013 AIP Publishing LLC. [<http://dx.doi.org/10.1063/1.4824071>]

I. INTRODUCTION

Dynamic neutron scattering is a direct probe of the internal, conformational dynamics of complex molecules, such as polymers and biological macromolecules. Of particular interest in this regard is the “quasielastic” component of this scattering, which is a superposition of approximately Lorentzian functions in energy centred on the elastic line, and corresponds to nonoscillatory relaxation processes.¹ Given that the whole-molecule translational and rotational diffusion can be subtracted from the scattering signal, this scattering probes internal conformational transitions, and thus is particularly valuable information. In simple systems, such as methyl rotors, the quasielastic processes involved can be unambiguously identified without the aid of detailed simulation. However, in complex molecules with an internal energy landscape involving a large number of minima the direct assignment of quasielastic scattering to specific structural rearrangements is perilous, since many of the underlying dynamical processes may exist on similar timescales. A typical approach is to fit the time dependence using a phenomenological model, such as stretched exponential behavior, and, while this can provide important insight into the physics involved,^{2–4} it lacks the ability to form a clear picture of which atoms move where and when.

Molecular dynamics simulation is being used increasingly as an aid to interpreting dynamic neutron scattering experiments.⁵ Direct calculation of the dynamic structure factor, $S(\mathbf{q}, \omega)$, from the atomic trajectories can be

conveniently performed.⁶ In cases where simulation and experiment agree, the simulation can be decomposed to identify motions present. Early examples of this in protein biophysics include the identification of liquid-like rigid-body side-chain motions as dominating picosecond-timescale quasielastic scattering from myoglobin,^{7,8} and recently simulation was used to demonstrate that the neutron susceptibility of lysozyme can be decomposed into three types of motion.⁹ However, analyses such as these rely on first imagining a type of motion, then fitting it to the simulation results and comparing with experiment. A more general approach that does not rely on such intuition has been lacking.

In the accompanying article,¹⁰ we have derived a theory based on dynamical fingerprint analysis^{11,12} to compute neutron and x-ray scattering spectra from the conformational dynamics of the molecule, that provides a direct 1:1 assignment of experimentally observable relaxation timescales to structural transitions, each of which contributes to the scattering profile. In practice, this theory can be applied to Markov (state) models (MSMs) of conformational dynamics.^{13–17} In the Markov modeling approach, one discretizes the conformation space of the molecule into small sets (microstates), and uses MD simulation data to compute transition probabilities between microstates. An advantage of this approach is that slow kinetics can be computed with short trajectories and thus even systems with relaxation times longer than the affordable trajectory length are within reach.^{11,18–20} See Refs. 21–23 for a review of Markov models,²⁴ for an overview

of the methodology,^{25,26} and for software to construct Markov models from MD simulation data.

Note that MSMs are adequate for estimating the slow exponential relaxation processes in the data, but not necessarily for fast relaxations. A comment is in order to relate this restriction to the frequent observation of power-law or stretched exponential kinetics in experimental data. Although a power-law decay over infinite number of timescales cannot be reproduced by a sum of exponentials, a power law over a few orders of magnitude can very well be represented by a sum of rather few exponentials (e.g., see Ref. 11, supplementary Fig. 1 – this shows that an apparent power-law over three orders of timescales can be generated by as few as three appropriately spaced exponentials). Moreover, although much of the original mathematical theory of MSMs is based on the existence of a timescale separation, such a timescale separation does not seem to be necessary in practice, because many systems without a clear separation of timescales have been able to be usefully modeled by MSMs in the recent literature. In addition, recent mathematical theory^{27,28} suggests that, irrespective of whether a timescale separation exists or not, MSMs can approximate those processes with propagator eigenfunctions that are well discretized. A timescale separation helps in cases where this discretization is poor.

Here, we illustrate the practical application of dynamical neutron scattering from conformational dynamics by applying it to the well-characterized alanine dipeptide in solution. By using a Markov state model, the quasielastic region of the scattering intensity of neutron spectra is decomposed into individual components, each of which corresponds to a fully-characterized intramolecular structural transition with an associated exponential decay rate.

II. THEORY

The theory describing how neutron scattering functions can be reconstructed from MD simulation and MSM has been presented in detail in the accompanying previous study.¹⁰ Here, we provide a brief overview of the theory used to compute the neutron scattering quantities in the present case.

A. Markov state model

An MSM is based upon a discretization of the internal (molecular) conformational space into a set of substates $S = \{s_1, \dots, s_i, \dots, s_m\}$. A transition process between conformational substates can be described by a discrete-time Markov process using the transition matrix:

$$\mathbf{p}_{t+\tau}^T = \mathbf{p}_t^T \mathbf{T}(\tau), \quad (1)$$

where \mathbf{p}_t^T is an m -dimensional row vector containing the probabilities of finding the system in each of its m states at time t . The $m \times m$ transition matrix $\mathbf{T}(\tau)$ contains the conditional probabilities T_{ij} of finding the system in state j at time given that it was in state i at time t :

$$T_{ij} = \mathbb{P}(\mathbf{x}_{t+\tau} \in s_j \mid \mathbf{x}_t \in s_i) = \frac{1}{\pi_i} \int_{\mathbf{x}} d\mathbf{x} \mu(\mathbf{x}) \int_{\mathbf{y}} d\mathbf{y} p(\mathbf{x}, \mathbf{y}; \tau). \quad (2)$$

The transition matrix $\mathbf{T}(\tau)$ is dependent on the chosen *lag time* τ , the time step for building the MSM. The eigenvalue decomposition (EVD) of the transition matrix $\mathbf{T}(\tau)$ yields

$$\mathbf{T}(\tau) \hat{\boldsymbol{\psi}}_k = \hat{\lambda}_k(\tau) \hat{\boldsymbol{\psi}}_k,$$

$$\hat{\boldsymbol{\phi}}_k \mathbf{T}(\tau) = \hat{\boldsymbol{\phi}}_k \hat{\lambda}_k(\tau),$$

where $\hat{\boldsymbol{\phi}}_k = \mathbf{\Pi} \hat{\boldsymbol{\psi}}_k$ relate the left and right coarse-grained eigenvectors and $\mathbf{\Pi} = (\pi_1, \dots, \pi_m)$ is the equilibrium distribution of the system. $\hat{\boldsymbol{\phi}}_k$ and $\hat{\boldsymbol{\psi}}_k$ provide information on structural changes. The corresponding eigenvalues, $\hat{\lambda}_k(\tau)$, describe the estimated relaxation times and decay exponentially in τ , and thus each eigenpair is associated with a relaxation timescale \hat{t}_k , or a relaxation rate \hat{k}_k , via

$$\hat{\lambda}_k = e^{-\tau/\hat{t}_k} = e^{-\hat{k}_k \tau}. \quad (3)$$

Transitions with $\hat{\lambda}_k \ll 1$ correspond to fast processes, while transitions with $\hat{\lambda}_k \approx 1$ describe slow ones. Since $\mathbf{T}(\tau)$ is a stochastic matrix, the first process, $\hat{\boldsymbol{\phi}}_1$, has, by definition, an eigenvalue of $\hat{\lambda}_1 = 1$, and corresponds to the equilibrium distribution of the system $\mathbf{\Pi}$. The relaxation time is independent of the lag time τ when the jump process on the discrete state space is Markovian. Since this is only approximately true in reality, finding a suitable lag time requires an implied timescale analysis.²⁹

A time-correlation function between observables $f(\mathbf{x})$ and $g(\mathbf{x})$ is then approximated via the Markov model by

$$\begin{aligned} \langle f_t g_{t+\tau} \rangle_t &= \int d\mathbf{x} \int d\mathbf{y} f(\mathbf{x}) c(\mathbf{x}, \mathbf{y}; \tau) g(\mathbf{y}) \\ &\approx \sum_{i,j=1}^m \bar{f}_i \pi_i T_{ij}(\tau) \bar{g}_j \end{aligned} \quad (4)$$

$$= \sum_{k=1}^m \hat{\lambda}_k(\tau) \langle \bar{\mathbf{f}}, \hat{\boldsymbol{\phi}}_k \rangle \langle \bar{\mathbf{g}}, \hat{\boldsymbol{\phi}}_k \rangle, \quad (5)$$

where $c(\mathbf{x}, \mathbf{y}; \tau)$ is the probability of finding the molecule at position \mathbf{x} at time t and at position \mathbf{y} at time $t + \tau$, and $\bar{\mathbf{f}}$, $\bar{\mathbf{g}}$ are the observable values averaged over the Markov states, e.g., $\bar{f}_i = \int_{S_i} d\mathbf{x} \mu(\mathbf{x}) f(\mathbf{x})$. Clearly, these local averages can be efficiently computed by a direct average over the MD simulation frames that are located in set S_i .

B. Neutron scattering

Dynamic neutron scattering experiments measure the dynamic structure factor, $S(\mathbf{q}, \omega)$, where \mathbf{q} is the momentum exchange and $E = \hbar\omega$ is the energy transfer. $S(\mathbf{q}, \omega)$ contains a coherent part arising from self- and cross-correlations of atomic motions and an incoherent part describing only single atom motions:^{30,31}

$$S_{coh}(\mathbf{q}, \omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F_{coh}(\mathbf{q}, t) e^{-i\omega t} dt, \quad (6)$$

$$S_{inc}(\mathbf{q}, \omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} F_{inc}(\mathbf{q}, t) e^{-i\omega t} dt, \quad (7)$$

and

$$F_{coh}(\mathbf{q}, \tau) = \sum_{\alpha, \beta} b_{\alpha, coh} \cdot b_{\beta, coh}^* \cdot \langle e^{-i(\mathbf{q}, \mathbf{r}_\alpha(t))} e^{i(\mathbf{q}, \mathbf{r}_\beta(t+\tau))} \rangle_t, \quad (8)$$

$$F_{inc}(\mathbf{q}, \tau) = \sum_{\alpha} b_{\alpha, inc} \cdot b_{\alpha, inc}^* \cdot \langle e^{-i(\mathbf{q}, \mathbf{r}_\alpha(t))} e^{i(\mathbf{q}, \mathbf{r}_\alpha(t+\tau))} \rangle_t. \quad (9)$$

The coherent and incoherent dynamic structure factors are time Fourier transforms of the coherent and incoherent intermediate scattering functions, $F_{coh}(\mathbf{q}, \tau)$ and $F_{inc}(\mathbf{q}, \tau)$, respectively. α and β label individual atoms whose positions are specified by their time-dependent position vector operators, $\mathbf{r}_\alpha(t)$ and $\mathbf{r}_\beta(t)$, respectively. Each atom has a coherent scattering length $b_{\alpha, coh}$ and an incoherent scattering length $b_{\alpha, inc}$.³²

$S_{inc}(\mathbf{q}, \omega)$ can be expressed as the dynamic susceptibility $\chi''_{inc}(\mathbf{q}, \omega)$ for analyzing inelastic neutron scattering spectra:

$$\chi''_{inc}(\mathbf{q}, \omega) = \frac{S_{inc}(\mathbf{q}, \omega)}{n_B(\omega) + 1}, \quad (10)$$

where $n_B(\omega) = (e^{\hbar\omega/kT} - 1)^{-1}$ (for neutron energy loss) is the Bose temperature factor. The dynamic susceptibility allows the estimation of characteristic relaxation time of molecular structural processes from the peaks, ω_{max} , in the frequency spectrum³³ as

$$t_r = 1/\omega_{max}. \quad (11)$$

C. $F(\mathbf{q}, \tau)$, $S(\mathbf{q}, \omega)$, and $\chi''_{inc}(\mathbf{q}, \omega)$ calculated from a Markov model

In our companion paper,¹⁰ we demonstrated that both the incoherent and coherent intermediate scattering functions, $F(\mathbf{q}, \tau)$ assume the form:

$$F(\mathbf{q}, \tau) = \sum_k \lambda_k(\tau) \cdot A_k(\mathbf{q}) \quad (12)$$

$$= \sum_k e^{-\frac{\tau}{t_k}} \cdot A_k(\mathbf{q}) \quad (13)$$

$$= \sum_k e^{-\hat{\kappa}_k \tau} \cdot A_k(\mathbf{q}), \quad (14)$$

where $A_k(\mathbf{q})$ is the scattering amplitude, given by

$$A_k(\mathbf{q}) = A_k^{coh}(\mathbf{q}) + A_k^{inc}(\mathbf{q}),$$

$$A_k^{coh}(\mathbf{q}) = \sum_{\alpha, \beta} a_{k\alpha}(\mathbf{q}) a_{k\beta}^*(\mathbf{q}) \quad (15)$$

$$= \left(\sum_{\alpha} a_{k\alpha}(\mathbf{q}) \right) \left(\sum_{\beta} a_{k\beta}(\mathbf{q}) \right)^*, \quad (16)$$

$$A_k^{inc}(\mathbf{q}) = \sum_{\alpha} a_{k\alpha}(\mathbf{q}) a_{k\alpha}^*(\mathbf{q}), \quad (17)$$

where $a_{k\alpha}(\mathbf{q}) = \langle b_{\alpha} e^{-i(\mathbf{q}, \mathbf{r}_\alpha)} \cdot \hat{\phi}_k \rangle$ is the scattering intensity of atom α in the relaxation process k .

The dynamic structure factor, $S(\mathbf{q}, \omega)$, can then be expressed as

$$S(\mathbf{q}, \omega) = \frac{1}{2\pi} \left[A_1(\mathbf{q}) \sqrt{2\pi c} e^{-c^2 \omega^2 / 2} + \sum_{k=2}^m A_k(\mathbf{q}) L(\hat{\kappa}_k, \omega) \right], \quad (18)$$

where $L(\hat{\kappa}_k, \omega) \equiv \frac{2\hat{\kappa}_k}{\omega^2 + \hat{\kappa}_k^2}$. The first component, $A_1(\mathbf{q}) \sqrt{2\pi c} \exp(-c^2 \omega^2 / 2)$, corresponding to $\hat{\lambda}_1 = 1$, is the elastic structure factor, convoluted here with a Gaussian convolution function, and the other components, $\sum_{k=2}^m A_k(\mathbf{q})$, are associated with the relaxation processes in the MSM. We assume that the instrument is suitable to resolve all present relaxation processes well. Quantitatively, this means we assume $c \gg t_k$ for all terms $k > 1$, resulting in $R(\tau = t_k; c) = e^{-t_k^2/c^2} \approx 1$. Therefore, we only need to find the Fourier transform of a single-exponential relaxation for each term $k > 1$, which turns out to be a Lorentzian function.

Finally, the incoherent dynamic susceptibility, $\chi''_{inc}(\mathbf{q}, \omega)$ can also be directly represented in terms of the propagator eigenvalues and eigenfunctions:

$$\chi''_{inc}(\mathbf{q}, \omega) = \frac{1}{2\pi(n_B + 1)} A_1(\mathbf{q}) \sqrt{2\pi c} e^{-c^2 \omega^2 / 2} \quad (19)$$

$$+ \frac{1}{2\pi(n_B + 1)} \sum_{k=2}^m A_k(\mathbf{q}) L(\hat{\kappa}_k, \omega). \quad (20)$$

III. METHODS

A. MD simulation details

Equilibrium MD simulations were performed using NAMD³⁴ with the CHARMM22 all-atom force field^{35,36} for the alanine dipeptide and TIP3P for the explicitly modeled water molecules.³⁷ The alanine dipeptide was placed inside a cubic box with a distance of 7 Å to the closest edge, and was solvated by water molecules. Periodic boundary conditions were used and electrostatic interactions were calculated using the Particle Mesh Ewald (PME) method³⁸ with a grid spacing of 1 Å. Short range electrostatic and van der Waals interactions were switched to zero between 10 and 12 Å. Neighbor lists were updated every 10 steps and the non-bonded interactions were calculated every second step. Bond stretches in the water were constrained using the SHAKE algorithm.³⁹

The simulated system was first energy minimized for 10 000 steps using the conjugate gradient algorithm, followed by 1 ns MD equilibration and 1 μ s of production. During the equilibration the temperature was gradually increased from 0 to 300 K at a rate of 10 K/ps. The temperature was kept at 300 K using the Langevin thermostat with a 5 ps time constant

coupled to the heavy atoms. The pressure was maintained at 1 atm using the Nosé-Hoover Langevin piston barostat with a period of 100 fs, a decay time of 50 fs and a temperature of 300 K. The integration time step of the MD simulations was 1 fs and the recorded trajectory time step was 100 fs.

B. Markov state model for alanine dipeptide

The derivation of a MSM for the configurational dynamics of alanine dipeptide requires a number of steps, which are outlined in Fig. S1 of the supplementary material.⁴⁰ These steps comprise state space reduction, discretization, implied timescale analysis, and kinetic clustering. Each step builds upon the choices and results of the previous steps. Although the methodology for MSM construction is well established, the main points of the procedure are briefly summarized here and relevant information and construction parameters are provided. Further details are provided exhaustively in Ref. 24. The software package EMMA (<https://simtk.org/home/emma>)²⁵ was used for the discretization of the trajectory data, the implied timescale analysis, and the quality assessment of the MSM.

1. State space reduction

Markov models can be constructed using generalized metrics such as pair-wise minimal RMSD,⁴¹ that avoid the need to define reaction coordinates or order parameters *a priori*. However, whenever the number of state space variables can be reduced to those relevant to represent the slow kinetics, it is useful to do so as this will enhance the quality of the resulting state space discretization.

The choice of coordinates is often guided by existing knowledge of the accessible configurations and by the ability to distinguish between them. An important feature of the selected set of coordinates is that they must define a space that allows kinetically separated conformations to be distinguished. In the case of alanine dipeptide, the internal dihedral rotations are a natural choice for distinguishing between different peptide configurations.

In the present example, the set of rotational angles, ϕ , ψ , CI ($CAY - CY - N - CA$), NI ($CA - C - NT - CAT$), C -ter, N -ter, and χ' methyl were selected as candidates for selected coordinates (see Fig. 1), and their time series are illustrated in Fig. S2 of the supplementary material.⁴⁰ The values for ϕ , ψ , and χ' switch between distinguishable, long-lived (≥ 10 ps) plateaux, which means that their associated free energy profiles feature significant barriers with distinct minima. The values for the CI and NI peptide bonds fluctuate around $0^\circ \pm 15^\circ$ throughout the whole simulation of 1 μ s owing to their strong rigidity, while, in contrast, the C - and N -ter methyl angles have no preferred orientation and rotate freely, consistent with their low barriers (≤ 0.1 kcal/mol).⁴² To further illustrate the individual time evolution of the selected angles, the corresponding NMR based angle-angle self-correlation functions $C_{NMR}(\tau) = \langle P_2[\cos(\theta(t) - \theta(t + \tau))] \rangle$ ⁴³ were calculated and are plotted in Fig. S3 of the supplementary material.⁴⁰ P_2 is the second-order Legendre polynomial

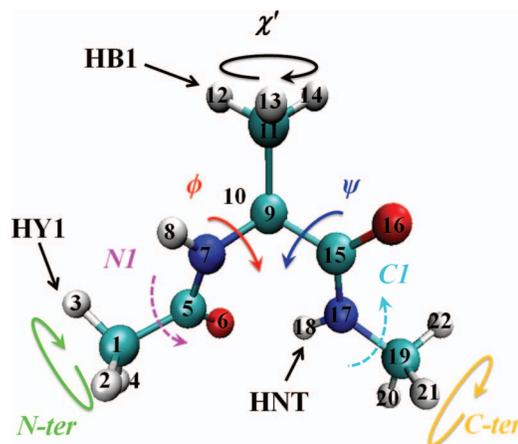


FIG. 1. Alanine dipeptide structure with atom indexes showing ϕ/ψ dihedral angles, CI/NI angles, C -ter/ N -ter/ χ' methyl angles, and representative hydrogen atoms.

given by $P_2(x) = \frac{1}{2}[3x^2 - 1]$. The estimated relaxation time for χ' is ~ 200 ps, for ϕ and ψ about ~ 50 ps, and for CI and NI and the C - and N -ter methyl angles ~ 0.1 ps and ~ 1 ps, respectively. Consequently, only ϕ , ψ , and χ' were selected as selected coordinates for the subsequent construction of Markov state model.

2. Discretization

The three selected coordinates, ϕ , ψ , and χ' span a continuous and fully periodic three-dimensional configurational space, each point representing a distinguishable configuration of the alanine dipeptide. To allow statistical analysis of the transitions within this state space, configurations which are geometrically close together must be clustered and assigned to the same microstate. Two common algorithms for this geometrical clustering are constant-space clustering and k -centers clustering.^{24,44-46} Constant-space clustering applies a regular grid to the complete state space and assigns distinct microstates to occupied grid elements, while the k -centers clustering algorithm traverses through the sampling data and generates new microstates for sampling points which do not fall within a certain geometrical distance of previous microstates.

Here, the k -centers clustering algorithm was used to derive a set of microstates for the alanine dipeptide data using the software EMMA with a target of $k = 1000$ cluster centers and the Euclidian metric for distance calculation. It has been shown that the clustering results converge for large k (e.g., $k \geq 1000$), independent of the order in which the trajectory is traversed.⁴⁴

After discretization, each configuration of the alanine dipeptide was assigned to one of a finite set of distinct microstates. The dynamics of the alanine dipeptide can then be described by a simple time series of these microstates. This serves as the basis for the estimation of the microstate state transition matrix, $T_{micro}(\tau)$. The element of the microstate transition matrix, T_{ij} , is the probability for finding the system in microstate j after a lag time of τ , given that it was in microstate i at time t .

3. Timescale analysis

The choice of the best lag time τ , with which to analyze the time series and for which the conditional probabilities in the MSM should be defined, requires some consideration. Processes operating on timescales significantly faster than the lag time cannot be accurately modeled, because they have already equilibrated at τ , precluding the garnering of any detailed time-dependent information. On the other hand, the lag time must be sufficiently long such that a Markov model on the discrete state space provides a good approximation to the kinetics of the real system.^{24,27}

Hence, the choice of the lag time is guided by an empirical approach in which the microstate transition matrix and its implied relaxation timescales, $\hat{t}_k^* = -\frac{\tau}{\ln|\lambda_k(\tau)|}$ ¹⁴ are estimated for a series for lag time τ . The implied relaxation times \hat{t}_k^* are then plotted against the lag time itself. For processes exhibiting the Markov property the implied relaxation timescales, \hat{t}_k^* , become independent of the lag time, τ , which is indicated by reaching a plateau. The ideal lag time is then given by the minimal value for which all implied timescales of interest have become independent of the value of the lag time.

Here, the implied relaxation timescale analysis for the derived microstate transition matrix of alanine dipeptide was performed for the lag times up to 100 ps, and is shown in Fig. S2 of the supplementary material⁴⁰ for the 20 largest eigenvalues. The solid line in Fig. 2 marks the onset for the resolution limit of the MSM which reflects the fact that fast processes cannot be reliably described by a MSM derived with a large lag time (lower triangle).

Here, the implied timescale analysis graph suggests two possible choices of metastable sets, which are also visible as timescale gaps between the estimated relaxation times.^{17,44} There is one gap at a timescale of 100 ps, separating two slow processes from the rest. At this timescale, three conformations are metastable. At a timescale of 10 ps, there is a second gap,

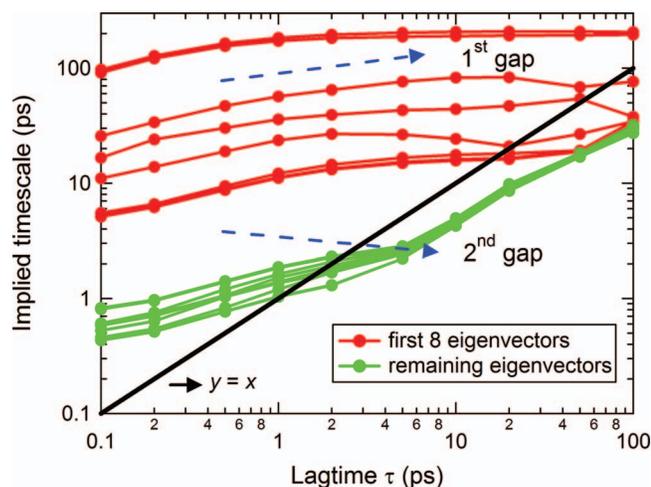


FIG. 2. Implied relaxation timescales of the processes associated with individual eigenvectors, depending on the lag time $n\tau$, computed as $\tau_k^*(n\tau) = n\tau / \ln|\lambda_k(n\tau)|$, where λ_k is the k th eigenvalue. The implied time scales of the first 8 eigenvectors become flat at around 10 ps, which is below the lifetimes of the metastable states, indicating that the interstate transitions are Markovian.

separating eight slow processes from the rest (9 metastable states).

4. Kinetic clustering

The implied timescale analysis provides a rationale for determining the number of relaxation processes reliably described by a Markov model. Since the number of Markov states is inherently equal to the number of relaxation processes (plus the equilibrium distribution), this also provides the number of distinguishable metastable states.

The microstate transition matrix contains explicit information about the kinetic connectivity between different groups of microstates, expressed as conditional probabilities for transitions. This can be exploited by clustering the microstates that are close kinetically, into macrostates. Here this is performed using the improved Perron-Cluster-Cluster-Analysis (PCCA) method^{47,48} implemented in EMMA. Note that in general it is not permitted to cluster the transition matrix into a transition matrix between metastable states, because a coarse-graining of state space is always associated with an increase of the discretization error of the associated Markov model.²⁷ Therefore, metastable states have in recent publications been used mostly for illustrative purposes.⁴⁹ However, in the present case the two timescale gaps are so pronounced that such a coarse-graining of the transition matrix is actually feasible without significant additional discretization error. This way two reduced macrostate transition matrices were obtained: one 3 state/100 ps matrix and the other a 9 state/10 ps matrix.

The micro- to macrostate assignment was visually inspected using a 3-dimensional plot of the cluster location and assignment in the $\phi - \psi - \chi'$ phase space. Figure S4(a) of the supplementary material⁴⁰ shows the assignment for the 3 state/100 ps matrix in which microstates with the same macrostate assignment have the same indicator, and Fig. S4(b) of the supplementary material⁴⁰ is a simplified schematic plot of the three macrostates: I ($\chi' \in [-120^\circ, 0^\circ]$), II ($\chi' \in [0^\circ, 120^\circ]$), and III ($\chi' \in [-180^\circ, -120^\circ] \cup \chi' \in [120^\circ, 180^\circ]$) in the $\phi - \psi - \chi'$ space. These plots indicate that the 3 state/100 ps matrix describes transitions between states with different values of the angle, which means that it exclusively describes the methyl group rotational dynamics. The higher resolution of the 9 state/10 ps matrix is required to resolve transitions between groups with different ϕ and ψ values, and these are shown in Fig. S5(a) of the supplementary material.⁴⁰ The Markov states are labeled according to their position in χ' (I, II, III) and their location within the $\phi - \psi$ plane (a, b, c). The projection of the distribution in $\phi - \psi - \chi'$ phase space onto the $\phi - \psi$ plane provides the Ramachandran plot (see Fig. S5(b) of the supplementary material⁴⁰), which leads to the identification of the Markov states with the Ramachandran states (a represents α_R , b represents β/C_5 , and c represents $C7_{ax}$). Figure 3 is a simplified schematic plot of these 9 macrostates.

The derived 9-state MSM, T_{macro} , is provided in Table I. The corresponding eigenvalues and eigenvectors are

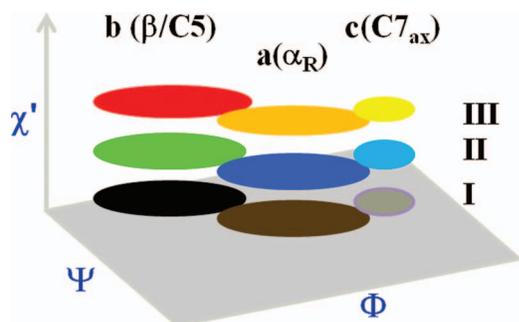


FIG. 3. The clustering leads to the identification of 9 metastable states, each occupying a region in the configurational space of ϕ , ψ , and χ' .

plotted in Fig. 4, and the relaxation times and frequencies are listed in Table II.

5. Markov model validation

The Markov model can be tested for its ability to reproduce certain time-dependent probability distributions derived directly from the underlying time series of Markov states. Here, the Chapman–Kolmogorov equation²⁴ was used to test the Markov property:

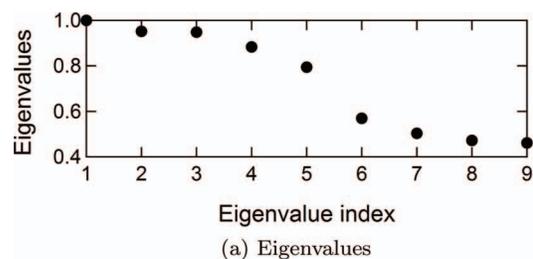
$$P(s(k\tau) = j | s(t=0) = i) = (T^k)_{ij}. \quad (21)$$

Equation (21) relates the conditional probability of an observed transition from state i to j after a lag time of $k\tau$ to the corresponding matrix element of the time developed transition matrix, T^k .

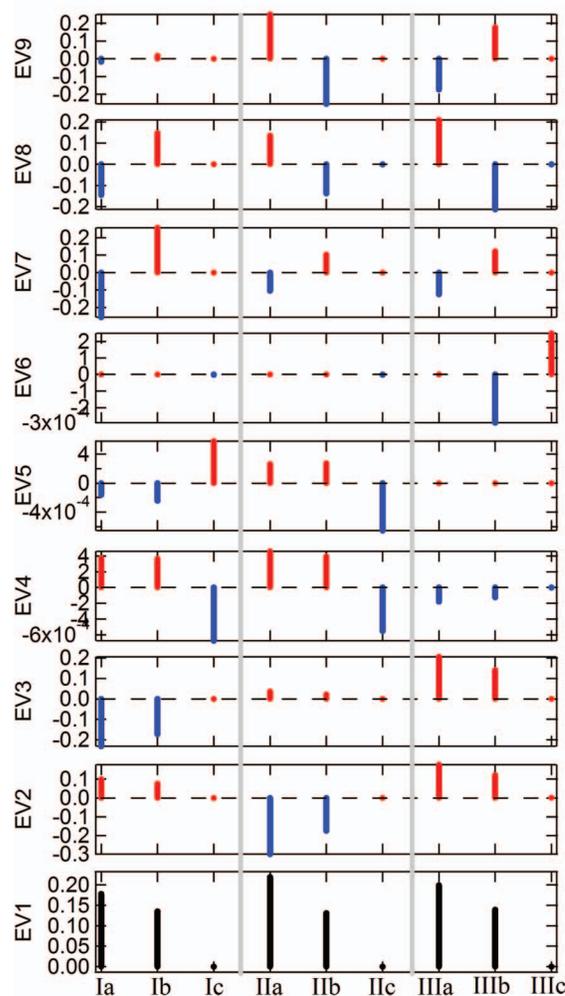
The projected probability distributions from the 9 state/10 ps transition matrix are shown in Fig. 5. The low populations of the Ic, IIc, and IIIc states lead to large uncertainties in the estimated probability distributions. However, the timescale of the time evolution of the probability distributions is well reconstituted.

C. Calculation of neutron scattering spectra

The neutron scattering spectra, i.e., the set of $F(q, \tau)$, $S(q, \omega)$ and $\chi''(q, \omega)$, were calculated from one 1 μ s MD trajectory using the publicly available software package Sassena,⁶ which directly computes the van-Hove correlation



(a) Eigenvalues



(b) Eigenvectors

FIG. 4. Eigenvalues and eigenvectors obtained from MSM using $m = 9$.

TABLE I. Row stochastic transition matrix $T^{(\chi', \phi, \psi)}$ for $m = 9$.

	\rightarrow Ia	\rightarrow Ib	\rightarrow Ic	\rightarrow IIa	\rightarrow IIb	\rightarrow IIc	\rightarrow IIIa	\rightarrow IIIb	\rightarrow IIIc
Ia	0.765	0.202	0.000	0.014	0.002	0.000	0.013	0.003	0.000
Ib	0.266	0.695	0.000	0.005	0.013	0.000	0.005	0.015	0.000
Ic	0.096	0.004	0.846	0.007	0.000	0.046	0.000	0.001	0.001
IIa	0.011	0.003	0.000	0.785	0.186	0.000	0.011	0.003	0.000
IIb	0.003	0.013	0.000	0.312	0.653	0.000	0.004	0.014	0.000
IIc	0.000	0.000	0.042	0.122	0.004	0.832	0.000	0.000	0.000
IIIa	0.012	0.004	0.000	0.012	0.003	0.000	0.767	0.202	0.000
IIIb	0.004	0.015	0.000	0.005	0.013	0.000	0.289	0.674	0.000
IIIc	0.000	0.000	0.002	0.001	0.001	0.002	0.304	0.029	0.662

TABLE II. Eigenvalues, relaxation times/frequencies, and A_k for $m = 9$ and $q = 1 \text{ \AA}^{-1}$.

Index	1	2	3	4	5	6	7	8	9
Eigenvalues	1	0.95	0.94	0.88	0.79	0.66	0.50	0.47	0.46
τ_k (ps)	∞	205.49	189.13	80.37	43.57	24.23	14.58	13.34	12.94
ω_k (GHz)	0	4.84	5.28	12.43	22.98	58.27	68.58	74.92	77.24
A_k^{inc}	0.71	3.4e-2	3.2e-2	1.4e-4	6.7e-5	1.4e-4	2.4e-2	5.5e-3	1.9e-3
A_k^{coh}	0.92	8.2e-6	2.6e-7	8.3e-5	7.9e-7	6.2e-5	3.5e-2	6.7e-3	5.4e-4

functions from the time-dependent MD position vectors given by Eqs. (6)–(10). The global translational and rotational motions were removed by aligning each simulation snapshot to the initial positions of the atoms C_α , CAY , and CAT . Thus, the calculated scattering functions contain information only about the internal motions of alanine dipeptide.

The neutron scattering functions were also obtained from the Markov state model by using Eqs. (12), (15), and (17)–(19). For the detailed analysis and reconstitution of $F_{inc}(q, \tau)$ from the Markov model, the MSM with $m = 9$ was used. Both in the direct calculation and for the MSM derived the orientational $F(q, \tau)$ averaging was performed over the same set of 50 random orientations of the molecule. Using 50 orientations was sufficient to get converged and representative estimates

for the relaxation times, as detailed in Figs. S6(a) and S6(b) of the supplementary material.⁴⁰

IV. RESULTS

A. Incoherent and coherent scattering functions

The total incoherent intermediate scattering function, $F_{inc}(q, \tau)$, for the whole molecule calculated from the Van-Hove correlation functions (Eq. (9)) and from the 9-state Markov model are (Eqs. (12) and (17)) compared in Fig. 6(a). The good agreement of the two methods indicates that the MSM reproduces $F_{inc}(q, \tau)$ very accurately, with small differences found only on fast time scales as would be expected (<10 ps). The full function $F_{inc}(q, \tau)$ from the 9-state MSM consists of the 9 components according to Eq. (12). The first component, $A_1^{inc}(q)$, corresponding to $\lambda_1 = 1$, is the elastic structure factor and the other components, $\sum_{k=2}^9 \exp(-\frac{\tau}{\tau_k}) A_k^{inc}(q)$, are associated with the relaxation processes in the MSM. The process-dependent scattering amplitudes, $A_k^{inc}(q)$, and the corresponding relaxation times, τ_k^{inc} , are provided in Table II.

The total $F_{inc}(q, \tau)$, which can be in principle determined in neutron scattering experiments, cannot be directly used to identify the set of structural relaxation processes that lead to the decay of $F_{inc}(q, \tau)$. Rather, in a typical experimental analysis, $F_{inc}(q, \tau)$ would be fitted with a set of simple exponential functions (assuming a small number of relaxation processes) or one single stretched exponential function (assuming a distribution of relaxation times³³). This would allow the relaxation timescales inherent to the system's dynamics to be probed, but would not provide any understanding of which structural rearrangements occur on these timescales. In contrast, the MSM explicitly provides the decomposition into atom-specific structural relaxation processes based on the molecular dynamics simulation and thus serves as a way to fully reconcile simulation and experiment.

The total coherent intermediate scattering function, $F_{coh}(q, \tau)$, calculated directly from the molecular dynamics trajectory using Eq. (8) and from the MSM using Eqs. (12) and (15) are plotted in Fig. 6(b). The agreement of the results illustrates that the MSM can also reliably reproduce the coherent scattering functions and was calculated for $q = 1-5 \text{ \AA}^{-1}$ with the results for $q = 2-5 \text{ \AA}^{-1}$ provided in Fig. S7 of the supplementary material.⁴⁰ The scattering amplitudes, $A_k^{inc}(q)$ and $A_k^{coh}(q)$, were calculated from Eqs. (15) and (17) and are listed in Table II for $q = 1.0 \text{ \AA}^{-1}$. $F_{coh}(q, \tau)$ decays more rapidly than $F_{inc}(q, \tau)$. The processes corresponding to the timescales ~ 200 ps ($\lambda_2 = 0.96$ and $\lambda_3 = 0.94$)

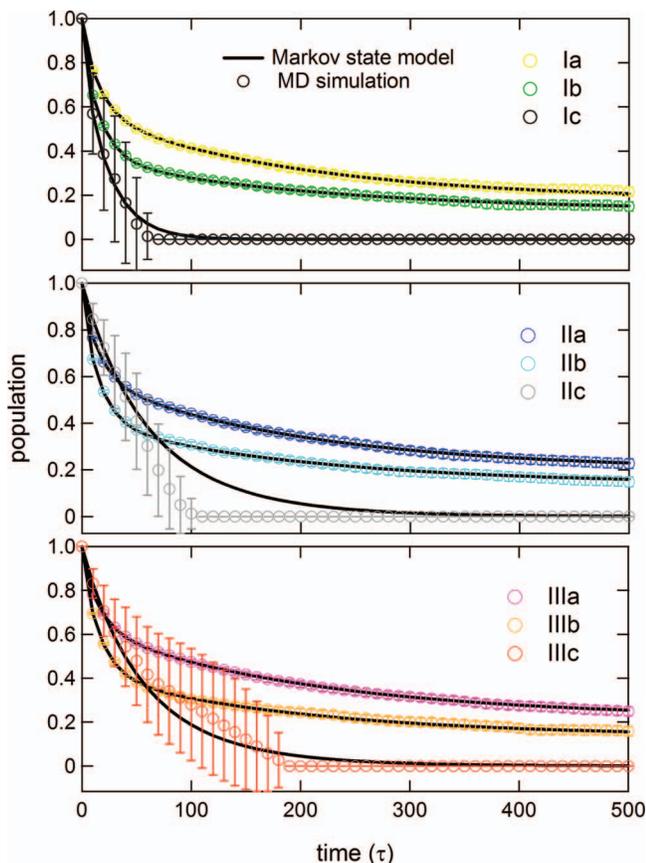


FIG. 5. Comparison of state population computed from MD simulations (open circles) and the Markov dynamics produced by the model network of macrostates (solid lines). The Markov dynamics is run C times, each time initializing a single state with population 1 and the remaining states with population 0. The evolution of the state population shows the relaxation of the selected state's population towards equilibrium.

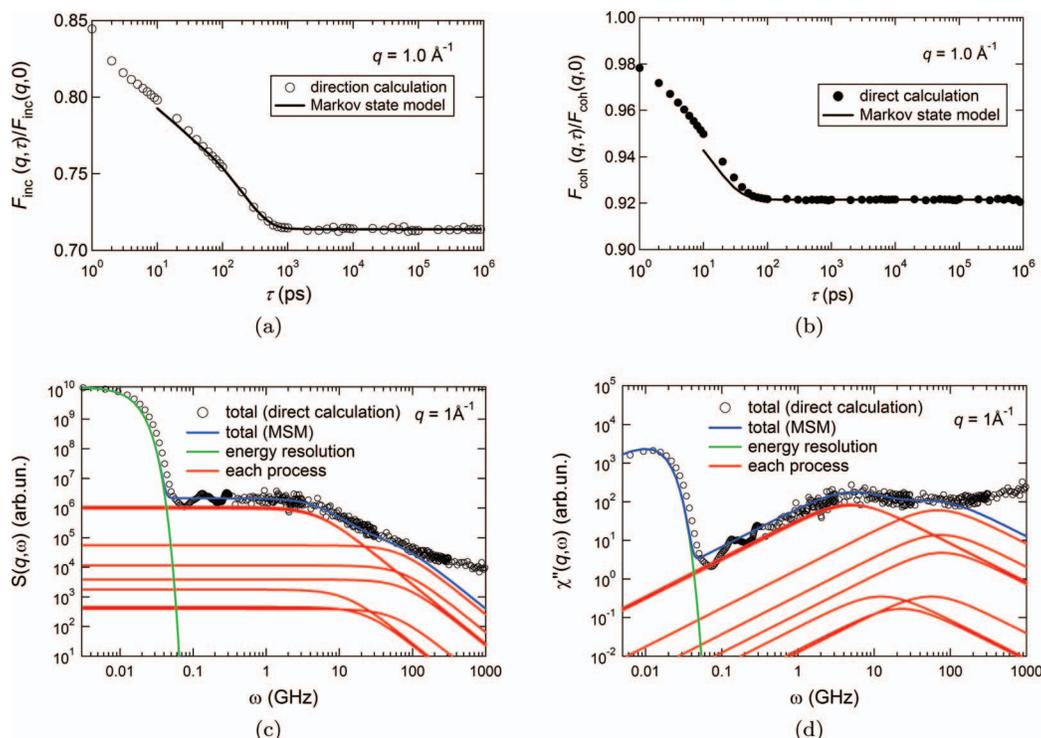


FIG. 6. Comparison of (a) $F_{inc}(q, \tau)$, (b) $F_{coh}(q, \tau)$, (c) $S_{inc}(q, \omega)$, and (d) $\chi''_{inc}(q, \omega)$ at $q = 1 \text{ \AA}^{-1}$ from Markov state model using Eqs. (12), (15), and (17) with the results directly calculated from MD trajectories using Eqs. (7)–(10).

have small coherent scattering amplitudes but relatively large incoherent scattering amplitudes. The tendency of the MSM to systematically underestimate the scattering intensity can be seen more pronounced in the case of the coherent intermediate scattering function, since the relevant relaxation processes are closer to the time resolution τ of the MSM and additional state space discretization is needed to resolve these accurately.

B. The incoherent dynamic structure factors and dynamic susceptibilities

The incoherent dynamic structure factors $S_{inc}(q, \omega)$ for the whole molecule directly calculated from the Van-Hove correlation functions (Eq. (7)) and from 9-state Markov transition matrix (Eq. (18)) are plotted in Fig. 6(c). The maximum at <0.1 GHz is elastic scattering, while quasielastic scattering ranges from 10^{-3} GHz to 10^4 GHz. $\chi''_{inc}(q, \omega)$ from the direct calculation and from the MSM are plotted in Fig. 6(d). $\chi''_{inc}(q, \omega)$ indicates the timescales of relaxation processes as peaks or maximums at frequencies corresponding to the associated relaxation times.¹⁰

In the frequency range up to 200 GHz the full spectra of $S_{inc}(q, \omega)$ and $\chi''_{inc}(q, \omega)$ from MSM (in blue) are in good agreement with the results directly calculated from the Van-Hove correlation functions (in black). By using Eq. (18), $S_{inc}(q, \omega)$ and $\chi''_{inc}(q, \omega)$ can now be fully decomposed into a set of individual structural relaxation processes: the elastic scattering component convoluted by the energy resolution function (in green), i.e., $A_1^{inc}(q)\sqrt{2\pi c} \exp(-c^2\omega^2/2)$, and the 8 quasielastic components (in red), i.e., $\sum_{k=2}^m A_k^{inc}(q) \frac{2/\hat{t}_k}{\omega^2 + (1/\hat{t}_k)^2}$, each associated

with a relaxation process in the 9-state MSM. The intensity of each relaxation process is dependent on the scattering amplitude $A_k^{inc}(q)$, and the frequency peak in $\chi''_{inc}(q, \omega)$ is located at the relaxation rate κ_k , or equivalently the inverse of the relaxation time, $1/\hat{t}_k$.

The MSM analysis allows all relevant relaxation processes to be distinguished even if they have similar timescales. The upper frequency limit for the MSM is determined by two factors. The slowest timescale neglected by 9-state MSM is approximated by the 10th relaxation timescale of the 1000-state MSM, which is here approximately 2 ps. The other limitation comes from the lag time on which the MSM is defined, which is $\tau = 10$ ps in the current case. Thus the direct calculation and the functions derived from MSM are expected to differ for $\omega > 200$ GHz.

C. Dependence on the scattering vector q

The dependence on the scattering vector q of the incoherent and coherent intermediate scattering functions, $F_{inc}(q, \tau)$ and $F_{coh}(q, \tau)$, is reflected in the scattering amplitudes, $A_k(q)$, and is shown for each relaxation process in Figs. 7(a) and 7(b). At low q (below 0.5 \AA^{-1}), the majority of the scattering is elastic and this results in $A_1(q)$ being significantly larger than all other components, $A_k(q)$, combined. As q increases, the fraction of inelastic scattering also rises, leading to a decrease in $A_1(q)$ and a respective increase for the inelastic components $A_k(q)$. The variation in the magnitude $A_k(q)$, at higher q (above 0.5 \AA^{-1}) may differ between processes. In the case of incoherent scattering, this leads to crossovers in some cases. For example, $A_2^{inc}(q)$ and $A_3^{inc}(q)$

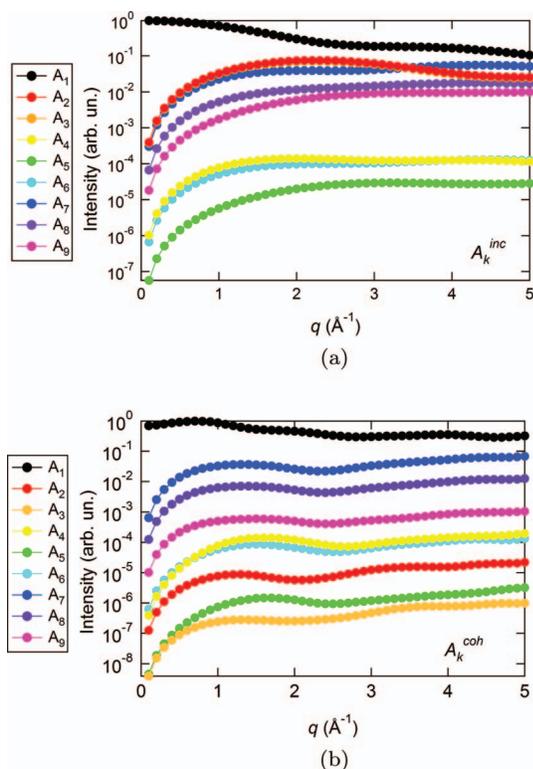


FIG. 7. The coherent and incoherent scattering functions are perfectly decomposable into a set of scattering amplitudes, each showing a distinct q vector length dependence. (a) A_k^{inc} shows significant intensity for processes $k = 2, 3,$ and 7 and (b) A_k^{coh} of for process $k = 7$. The elastic intensity is described by A_1 , respectively, while all other contributions describe dynamic processes.

overlap over the full q range and both fall below $A_7^{inc}(q)$ at $q > 3 \text{ \AA}^{-1}$. The difference in absolute magnitude of $A_k(q)$ between the processes provides direct information about which processes are experimentally relevant, these being processes $k = 2, 3, 7, 8,$ and 9 and $k = 7, 8,$ and 9 for incoherent and coherent scattering, respectively.

D. Decomposition

For incoherent neutron scattering, each component $A_k(q)$ is simply the sum of the individual contributions from single atoms. It is thus possible to directly identify the contribution of the scattering from individual atoms to a particular structural transition. Figure S8 of the supplementary material⁴⁰ shows the relative contribution of each atom to the amplitude of each of the relaxation processes at $q = 1 \text{ \AA}^{-1}$. A more complete picture of the decomposition of the components A_k into the contributions from the different hydrogen atoms as a function of the scattering vector length q is provided in Fig. 8, which illustrates that the $A_k(q)$ associated with processes $k = 2, 3, 5,$ and 9 are mainly determined by hydrogen atoms on the methyl group side chain, $k = 4$ is mainly determined by hydrogen HN, and $k = 6$ and 7 are strongly influenced by atoms HNT and HA. Interestingly, the dynamics of the atoms on the terminal methyl groups HY1/2/3 and HT1/2/3 do not contribute significantly to the components $A_k(q)$. The origin of this “invisibility” was clarified by in-

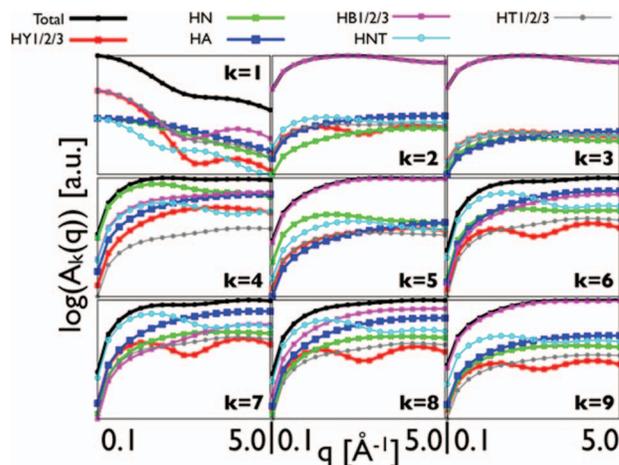


FIG. 8. Decomposition of the components A_k^{inc} into the contributions from distinct hydrogen atoms as a function of the scattering length q . This leads to an understanding how individual atom groups contribute to the observed q vector length dependence of the corresponding dynamic process k . Processes $k = 2, 3, 5,$ and 9 are dominated by atoms HB1/2/3. Atom HNT only plays a role in processes $k = 6, 7,$ and 8 at low values for the q vector length. Process 7 is dominated by movements of atoms HNT and HA, which are related to backbone angle rotations.

specting the spatial distributions of the individual atoms as a function of Markov state, which is shown in Fig. 9. The average location of HY1/2/3 and HT1/2/3 do not vary significantly between the states, which is a requirement for large amplitudes.

E. Eigenvector interpretation

While the MSM allows the construction of a transition network for Markov states, the connection to the underlying physical system is somewhat lost and has to be re-engineered with the help of the location of the Markov states within the

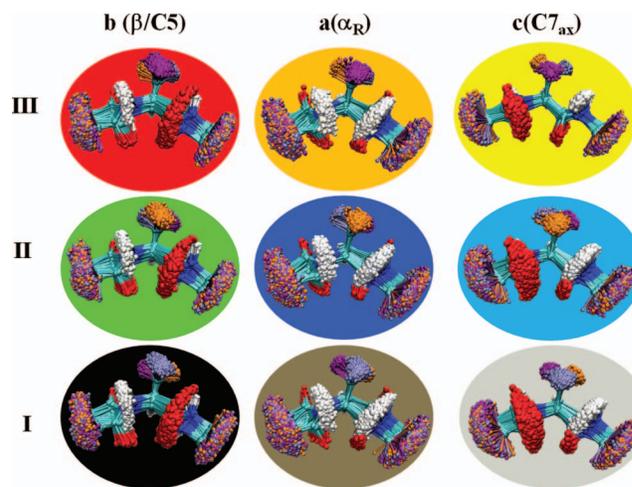


FIG. 9. Physical models of the Markov states. The configurational freedom within each Markov state causes the positions of individual atoms to transform into a probability density, represented with the help of individual atom colors. Since the terminal methyl groups were not used in the Markov decomposition, the corresponding atoms occupy all possible methyl group rotation angles. Representative pdb files are available as the supplementary material.⁴⁰

reduced state space ($\phi - \psi - \chi'$) and clues derived from the observables A_k and the spatial distribution of the atoms within each Markov state (see Fig. 9).

1. Process 2 and Process 3

The first two eigenvectors can be clearly identified as processes involving rotation of the side-chain methyl group: the eigenvector components (Fig. 4(b)) describe population changes from (II abc) to (I abc , III abc) for Process 2 and (I abc) to (III abc) for Process 3, where I, II, and III correspond to Markov states with distinct methyl group angles. The physical models for the Markov states (Fig. 9) also clearly show the rotation of the methyl group as the system transition between I, II, and III. Furthermore, the exclusive contribution of hydrogen atoms from the methyl group to A_2^{inc} and A_3^{inc} (Fig. 8) also indicates that Processes 2 and 3 describe the methyl group side chain rotation exclusively.

2. Process 4

The eigenvector components (Fig. 3(b)) associated with Process 4 indicate a transition $c \rightarrow (a, b)$. This corresponds to a change in the backbone dihedral ϕ , involving a displacement of the hydrogen HN (Fig. 1). This assignment is also supported by the decomposition A_k ($k = 4$) for incoherent scattering (Fig. 8), which contains a dominant contribution from hydrogen atom HN. Interestingly, the decomposition A_k ($k = 4$) in Fig. 8 shows that methyl group hydrogen atoms (HB1/2/3) and hydrogen HA contribute significantly at higher values of the scattering length q . The reason for this contribution can be understood with the help of the physical models of the Markov states (Fig. 9). When the dipeptide undergoes a backbone dihedral transition, the methyl group is slightly pushed away from the original location, resulting in a displacement of each hydrogen on the side chain. Consequently, the methyl group has a slightly different orientation in each of the substates a , b , and c , describing the backbone conformation. The same explanation holds true for atom HA, which adjusts its equilibrium position in response to the backbone dihedral changes.

3. Process 5

Summing the components of Process 5 yields a net flow between I abc and II abc , while the change in population of the substates a (I a +II a +III a), b (I b +II b +III b), and c (I c +II c +III c) is almost zero. The decomposition of A_k ($k = 5$) in Fig. 8 indicates that mainly atoms in the methyl group HB1/2/3 contribute to the scattering signal.

4. Process 6

The components of Process 6 clearly suggest a transition between substate b and c , corresponding to rotation of the backbone angle Ψ , which involves the HNT hydrogen atom. This conclusion is also supported by the decomposition of A_k ($k = 6$) in Fig. 8. As in the case of Process 4, HA and H1/2/3

contribute significantly at higher values of q , which indicates that the methyl group side chain and hydrogen HA adjust their position in response to changes in Ψ .

5. Process 7

The components of Process 7 immediately suggest that the eigenvector describes a transition between substates a and b , which corresponds to rotation of the backbone angle ψ . The strong influence of hydrogens HNT and HA in A_k ($k = 7$) in Fig. 8 can be understood by inspection of the physical models for the Markov States (Fig. 9). While hydrogen HNT clearly changes its orientation during ψ rotation, hydrogen HA changes from an “upward” to a “downward” position.

6. Process 8 and Process 9

The components of Process 8 and Process 9 mainly describe a transition between substates a and b . These transitions are accompanied by rotations of the methyl group side chain, which explains the strong contribution of HB1/2/3 to A_k ($k = 8$) and A_k ($k = 9$), where the contribution to Process 9 is from HB1/2/3 almost exclusively.

V. CONCLUSION

In the accompanying paper,¹⁰ a theoretical framework is constructed for interpreting quasielastic neutron scattering experiments on complex molecules with the aid of molecular simulation. The method overcomes the problem heretofore experienced that, even with the aid of molecular simulation, it has been difficult to unambiguously describe a quasielastic spectrum by specific sets of atoms transiting between free-energy minima on the energy landscape concerned. The approach presented is to reconstruct dynamic neutron scattering spectra from MD-based Markov state models. The method decomposes the quasielastic region of the scattering intensity of neutron spectra into components, each of which corresponds to a single structural transition with an associated exponential decay rate. The method, which is in principle applicable to both coherent and incoherent scattering, was illustrated here using conformational dynamics of the well-characterized alanine dipeptide in solution. The procedure is not completely without assumptions, as the state space in which the Markov state model is to be constructed must be defined *a priori*. However, the remainder of the procedure is automatic. Here, a 9-state MSM in the space of the backbone dihedral angles and the side-chain methyl group was found to describe the scattering: in the quasielastic region of the neutron scattering spectrum the agreement between the direct calculation and that derived from the Markov state model is excellent. The dependence of the individual scattering components on q and the underlying transition process (e.g., symmetry operations) connect the scattering tightly to the underlying molecular structure and dynamics. Simulation and neutron scattering have become common partners over the last 20 years. Therefore, we expect this methodology or variants of it to be broadly applied to the characterization of the dynamics of

complex molecules for which the potential energy surfaces possess many internal free-energy minima.

ACKNOWLEDGMENTS

We greatly thank Dr. Yinglong Miao for valuable discussion and help. We acknowledge the financial support provided by the National Science Foundation Award No. MCB-0842871. Frank Noé acknowledges the support provided by DFG Grant No. NO825/2, the ERC starting grant pcCell, and the DFG research center MATHEON.

- ¹M. Bee, *Quasielastic Neutron Scattering: Principles and Applications in Solid State Chemistry, Biology and Materials Science* (Adam Hilger, Bristol and Philadelphia, PA, 1988).
- ²M. Tehei, B. Franzetti, K. Wood, F. Gabel, E. Fabiani, M. Jasnin, M. Zamponi, D. Oesterhelt, G. Zaccai, M. Ginzburg, and B.-Z. Ginzburg, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 766 (2007).
- ³X.-Q. Chu, E. Mamontov, H. O'Neill, and Q. Zhang, *J. Phys. Chem. Lett.* **3**, 380 (2012).
- ⁴H. Jansson, F. Kargl, F. Fernandez-Alonso, and J. Swenson, *J. Chem. Phys.* **130**, 205101 (2009).
- ⁵J. Smith, *Q. Rev. Biophys.* **24**, 227 (1991).
- ⁶B. Lindner and J. Smith, *Comput. Phys. Commun.* **183**, 1491 (2012).
- ⁷G. R. Kneller and J. C. Smith, *J. Mol. Biol.* **242**, 181 (1994).
- ⁸S. Furois-Corbin, J. C. Smith, and G. R. Kneller, *Proteins* **16**, 141 (1993).
- ⁹L. Hong, N. Smolin, B. Lindner, A. P. Sokolov, and J. C. Smith, *Phys. Rev. Lett.* **107**, 148102 (2011).
- ¹⁰B. Lindner, Z. Yi, J.-H. Prinz, J. Smith, and F. Noé, *J. Chem. Phys.* **139**, 175101 (2013).
- ¹¹F. Noé, S. Doose, I. Daidone, M. Löllmann, J. Chodera, M. Sauer, and J. Smith, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 4822 (2011).
- ¹²B. Keller, J.-H. Prinz, and F. Noé, *Chem. Phys.* **396**, 92 (2012).
- ¹³C. Schütte, A. Fischer, W. Huisinga, and P. Deuffhard, *J. Comput. Phys.* **151**, 146 (1999).
- ¹⁴W. C. Swope, J. W. Pitera, and F. Suits, *J. Phys. Chem. B* **108**, 6571 (2004).
- ¹⁵N. Singhal and V. S. Pande, *J. Chem. Phys.* **123**, 204909 (2005).
- ¹⁶J. D. Chodera, K. A. Dill, N. Singhal, V. S. Pande, W. C. Swope, and J. W. Pitera, *J. Chem. Phys.* **126**, 155101 (2007).
- ¹⁷F. Noé, I. Horenko, C. Schütte, and J. C. Smith, *J. Chem. Phys.* **126**, 155102 (2007).
- ¹⁸V. Voelz, M. Jäger, L. Zhu, S. Yao, O. Bakajin, S. Weiss, L. Lapidus, and V. Pande, *Biophys. J.* **100**, 515a (2011).
- ¹⁹G. Bowman, V. Voelz, and V. Pande, *J. Am. Chem. Soc.* **133**, 664 (2011).
- ²⁰S. Sadiq, F. Noé, and G. de Fabritiis, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 20449 (2012).
- ²¹J.-H. Prinz, B. Keller, and F. Noé, *Phys. Chem. Chem. Phys.* **13**, 16912 (2011).
- ²²F. Noé and S. Fischer, *Curr. Opin. Struct. Biol.* **18**, 154 (2008).
- ²³V. Pande, K. Beauchamp, and G. Bowman, *Methods* **52**, 99 (2010).
- ²⁴J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. Chodera, C. Schütte, and F. Noé, *J. Chem. Phys.* **134**, 174105 (2011).
- ²⁵M. Senne, B. Trendelkamp-Schroer, A. Mey, C. Schütte, and F. Noé, *J. Chem. Theory Comput.* **8**, 2223 (2012).
- ²⁶K. Beauchamp, G. Bowman, T. Lane, L. Maibaum, I. Haque, and V. Pande, *J. Chem. Theory Comput.* **7**, 3412 (2011).
- ²⁷M. Sarich, F. Noé, and C. Schütte, *Multiscale Model. Simul.* **8**, 1154 (2010).
- ²⁸F. Noé and F. Nüske, *Multiscale Model. Simul.* **11**, 635 (2013).
- ²⁹B. Keller, P. Hünenberger, and W. van Gunsteren, *J. Chem. Theory Comput.* **7**, 1032 (2011).
- ³⁰L. Van Hove, *Phys. Rev.* **95**, 249 (1954).
- ³¹S. W. Lovesey, *Theory of Neutron Scattering from Condensed Matter* (Oxford University Press, New York, 1987).
- ³²V. Sears, *Neutron News* **3**, 26 (1992).
- ³³J. H. Roh, J. E. Curtis, S. Azzam, V. N. Novikov, I. Peral, Z. Chowdhuri, R. B. Gregory, and A. P. Sokolov, *Biophys. J.* **91**, 2573 (2006).
- ³⁴J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. V. Kalé, and K. Schulten, *J. Comput. Chem.* **26**, 1781 (2005).
- ³⁵A. D. Mackerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, D. T. Nguyen, T. Ngo, B. Prodhom, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, J. Wiorkiewiczkuczera, and M. Karplus, *FASEB J.* **6**, A143 (1992).
- ³⁶A. D. Mackerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewiczkuczera, D. Yin, and M. Karplus, *J. Phys. Chem. B* **102**, 3586 (1998).
- ³⁷W. Jorgensen, J. Chandrasekhar, J. Madura, R. Impey, and M. Klein, *J. Chem. Phys.* **79**, 926 (1983).
- ³⁸T. Darden, D. York, and L. Pedersen, *J. Chem. Phys.* **98**, 10089 (1993).
- ³⁹J.-P. Ryckaert, G. Ciccotti, and H. Berendsen, *J. Comput. Phys.* **23**, 327 (1977).
- ⁴⁰See supplementary material at <http://dx.doi.org/10.1063/1.4824071> for details regarding Markov modeling and the convergence and decomposition of the scattering intensity.
- ⁴¹W. Kabsch and C. Sander, *Biopolymers* **22**, 2577 (1983).
- ⁴²T. Kojima, E. Yano, K. Nakagawa, and S. Tsunekawa, *J. Mol. Spectrosc.* **112**, 494 (1985).
- ⁴³M. Krishnan, V. Kurkal-Siebert, and J. Smith, *J. Phys. Chem. B* **112**, 5522 (2008).
- ⁴⁴G. Bowman, X. Huang, and V. Pande, *Methods* **49**, 197 (2009).
- ⁴⁵S. Dasgupta and P. M. Long, *J. Comput. Syst. Sci.* **70**, 555 (2005).
- ⁴⁶G. Bowman, K. Beauchamp, G. Boxer, and V. Pande, *J. Chem. Phys.* **131**, 124101 (2009).
- ⁴⁷M. Weber, ZIB Report No. 03-04, 2003.
- ⁴⁸P. Deuffhard and M. Weber, ZIB Report No. 03-09, 2003.
- ⁴⁹F. Noé, C. Schütte, E. Vanden-Eijnden, L. Reich, and T. Weikl, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 19011 (2009).