

Frequency-Selective Anharmonic Mode Analysis of Thermally Excited Vibrations in Proteins

Michael A. Sauer and Matthias Heyden*

School of Molecular Sciences, Arizona State University, Tempe, AZ 85287, USA

E-mail: heyden@asu.edu

Abstract

Low-frequency molecular vibrations at far-infrared frequencies are thermally excited at room temperature. As a consequence, thermal fluctuations are not limited to the immediate vicinity of local minima of the potential energy surface and anharmonic properties cannot be ignored. The latter is particularly relevant in molecules with multiple conformations such as proteins and other biomolecules. However, existing theoretical and computational frameworks for the analysis of molecular vibrations have so far been limited by harmonic or quasi-harmonic approximations, which are ill-suited for the description of anharmonic low-frequency vibrations.

Here, we developed a fully anharmonic analysis of molecular vibrations based on a time correlation formalism that eliminates the need for harmonic or quasi-harmonic approximations. We use molecular dynamics simulations of a small protein to demonstrate that this new approach, in contrast to harmonic and quasi-harmonic normal modes, correctly identifies the collective degrees of freedom associated with molecular vibrations at any given frequency. This allows us to unambiguously characterize the anharmonic character of low-frequency vibrations in the far-infrared spectrum.

Introduction

The $3N - 6$ vibrations of a molecule with N atoms provide collective coordinates that can describe large and small changes in conformation. Computational procedures that identify collective degrees of freedom associated with internal vibrations of a molecular system are commonly based on harmonic^{1,2} or quasi-harmonic^{3,4} normal modes. Here, an Eigenvalue problem based on either a Hessian matrix of the potential energy or a co-variance matrix of coordinate displacements is solved. The Eigenvectors correspond to vibrational modes and their Eigenvalues are associated with force constants or vibrational frequencies, which can be used to estimate amplitudes of classical oscillations at a given temperature.¹

Principal component analysis (PCA) and time-lagged independent component analysis (TICA) are related methods, which extract collective degrees of freedom associated with large-amplitude motion or slow dynamics from molecular dynamics (MD) trajectories.⁵⁻⁸ PCA and TICA are powerful tools if extensive simulation trajectories are available for the system of interest that sample rare events such as protein conformational transitions. However, this is typically not the case in high throughput applications, which require predictions based on either structures or short simulations.

To predict conformational fluctuations in complex molecules, the vibrational modes associated with small force constants, low frequencies, and large amplitude fluctuations are of particular interest. For example, low-frequency harmonic normal modes have been used as collective degrees of freedom to predict potential conformational transitions in complex biomolecular systems.⁹⁻¹¹ The latter can then be used as collective variables in biased simulations to efficiently explore distinct conformational states.^{1,11,12}

However, the approximations associated with harmonic and quasi-harmonic normal modes are primarily applicable to high-frequency oscillations with large force constants and small fluctuations. Quantum harmonic oscillators (HO) with frequencies $\nu > 6$ THz (≈ 200 cm⁻¹) primarily populate the vibrational ground state at room temperature ($h\nu > k_B T$). This can be rationalized using the ratio of the quantum HO partition function in the canonical en-

semble, $\Omega_{\text{HO}} = \sum_{n=0}^{\infty} \exp \left[- \left(n + \frac{1}{2} \right) h\nu / (k_B T) \right]$, and the corresponding Boltzmann weight of the vibrational ground state, $p_0 = \exp \left[- \frac{1}{2} h\nu / (k_B T) \right]$, which simplifies to:

$$\frac{\Omega_{\text{HO}}}{p_0} = \frac{1}{1 - p_0^2} \quad (1)$$

This ratio describes the average number of populated states in the quantum HO at equilibrium, which evaluates approximately to 1.58 for $h\nu = k_B T$ and asymptotically decreases to 1 for higher frequencies.

Thus, harmonic vibrations at frequencies in the mid-infrared, including the fingerprint region (700-1200 cm^{-1}) and the spectrum of vibrations characteristic for functional groups up to $\sim 4000 \text{ cm}^{-1}$,¹³ are effectively restricted to their vibrational ground state. In molecular simulations with classical force fields and a fixed topology, such vibrations are restricted to small oscillations around the potential energy minimum, which justifies the use of harmonic approximations or even constraints to eliminate them from the equations of motion.¹⁴

In contrast, vibrations in the far-infrared spectrum with frequencies $\nu \leq 6 \text{ THz}$ are thermally excited at room temperature and populate increasing numbers of vibrational states with decreasing frequency. This includes, for example, vibrations of non-covalent hydrogen bonds in proteins.^{15–18} The significant population of excited vibrational states allows for classical approximations, e.g., in classical molecular dynamics simulations, and for the exploration of large portions of the potential energy surface. As a consequence, low-frequency vibrations are strongly affected by anharmonic properties of the potential energy surface. For example, vibrational modes that connect distinct potential energy minima and describe barrier crossings are anharmonic by definition. Thus low-frequency vibrations in complex molecules with multiple conformations are likely ill-described by theoretical frameworks based on harmonic or quasi-harmonic approximations. In this manuscript, we present a theoretical framework for the analysis of vibrational modes that eliminates the need for harmonic approximations and allows us to unambiguously identify and characterize anharmonic

vibrational modes in molecular systems.

Harmonic and quasi-harmonic normal modes aim to map $3N$ degrees of freedom on $3N$ orthogonal normal modes (translations, rotations and vibrations).^{18–21} In addition to the implied assumptions regarding the curvature of the potential energy, such a 1-to-1 mapping is only feasible if the system explores a single potential energy minimum, which is not applicable for any evolving system with thermally activated anharmonic dynamics and transitions between potential energy minima. In proteins, for example, the thermal activation of anharmonic dynamics can be observed as the protein dynamical transition using quasi-elastic neutron scattering (QENS),²² THz absorption spectroscopy,¹⁶ or MD simulations.^{23,24} In the analysis of molecular simulations, this can be addressed using instantaneous normal modes (INM)^{12,25} or normal mode ensemble analysis (NMEA)¹⁸ to generate multiple distinct sets of $3N$ normal modes for distinct conformations. However, both methods remain reliant on harmonic approximations to describe the local shape of the potential energy surface, which can limit the ability to compare to and interpret experimental observations.^{17,18}

In the approach presented here, the anharmonic features of the potential energy surface and the time-evolution of vibrational modes are fully acknowledged using a dedicated time-correlation formalism.²⁶ Instead of assigning a fixed number of vibrational modes to the system, the method uses input from MD simulations and assigns vibrational modes and their respective contributions to the ensemble-averaged vibrational density of states for any frequency sampled by the analyzed trajectories.

Theory

To derive a time-correlation formalism for a fully anharmonic analysis of vibrational modes from MD trajectories, we start with atomic velocities, $\mathbf{v}$, scaled by the square root of the atom mass.²⁷

$$\tilde{\mathbf{v}}_i(t) = \sqrt{m_i} \mathbf{v}_i(t) \tag{2}$$

Here, the index i describes the $3N$ degrees of freedom of an unconstrained N -atom system. The scaled velocities, $\tilde{v}$, provide a simple expression for the kinetic energy of the system, which is constant in the canonical ensemble.

$$E_{\text{kin}}(t) = 1/2 \sum_i^{3N} \tilde{v}_i(t)^2 \quad (3)$$

We then define time cross-correlations between the velocities of all pairs of degrees of freedom i and j , which include the velocity auto-correlations for $i = j$.

$$C_{\tilde{v},ij}(\tau) = \langle \tilde{v}_i(t) \tilde{v}_j(t + \tau) \rangle_t \quad (4)$$

Here, the brackets $\langle \dots \rangle_t$ indicate ensemble-averaging over the simulation time. All time auto- and cross-correlations can be considered to be elements of a velocity cross-correlation matrix, $\mathbf{C}_{\tilde{v}}(\tau)$, that depends on the correlation time τ . Each of the velocity auto- and cross-correlations are then Fourier-transformed into the frequency domain ($\omega = 2\pi\nu$), which provides the elements of a frequency-dependent velocity correlation matrix, $\mathbf{C}_{\tilde{v}}(\omega)$.²⁰

$$C_{\tilde{v},ij}(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(i\omega\tau) C_{\tilde{v},ij}(\tau) d\tau \quad (5)$$

The trace of the matrix, $\text{tr}[\mathbf{C}_{\tilde{v}}(\omega)] = \sum_i^{3N} C_{\tilde{v},ii}(\omega)$, now contains the mass-weighted sum of Fourier-transformed velocity auto-correlation functions, which is equivalent to standard expressions for the vibrational density of states (VDoS).^{15,27–30}

$$I_{\text{VDoS}}(\omega) = \frac{2}{k_B T} \text{tr}[\mathbf{C}_{\tilde{v}}(\omega)] \quad (6)$$

The VDoS defined in Eq. 6 describes how the kinetic energy in the $3N$ degrees of freedom is distributed over all frequencies. Due to the normalization by the average kinetic energy per degree of freedom, $k_B T/2$, the integral of the VDoS over positive frequencies yields the

total number of degrees of freedom in the system: $\int_0^\infty I_{\text{VDoS}}(\omega)d\omega = 3N$

It is often desirable to map the $3N$ degrees of freedom of a given system onto a set of $3N$ normal modes.²⁰ Previously, Mathias and Baer used the same frequency-dependent matrix $C_{\tilde{\mathbf{v}},ij}(\omega)$ to define so-called generalized normal modes. The latter are obtained as a solution of an optimization problem that searches for a unitary coordinate transformation that minimizes the off-diagonal elements of $C_{\tilde{\mathbf{v}},ij}(\omega)$ at all sampled frequencies simultaneously. However, the use of a single set of normal modes is only applicable if the system only explores the immediate vicinity of a single potential energy minimum. A detailed example of a simple anharmonic system for which normal modes fail to describe the vibrations is provided in the Supporting Information (SI).

In the approach described here, we let the dynamics of the system itself determine the number of modes needed to describe the vibrations at each frequency. Further, we do not require that an integer number of vibrational degrees of freedom contributes to the vibrations at each frequency, which is an implied assumption in harmonic approximations.

Instead, we compute Eigenvalues and Eigenvectors of $\mathbf{C}_{\tilde{\mathbf{v}}}(\omega)$ at selected frequencies sampled by the Fourier transform in Eq. 5 (e.g., for specific features observed in the VDoS or absorption spectrum, at regular frequency intervals, etc.). At first glance, this seems to generate a near-infinite number of modes, which would impede a meaningful analysis or assignments of spectral features. However, we show in the following that the Eigenvalues, obtained by this FREquency-SElective ANharmonic (FRESEAN) mode analysis report directly on the contributions of each vibrational mode to the VDoS at frequency ω_{sel} . Notably, the majority of these Eigenvalues are zero and thus indicate Eigenvectors that can be safely ignored. Only a small fraction of the Eigenvalues are non-zero and their associated Eigenvectors thus describe vibrational modes that contribute to the VDoS at frequency ω_{sel} .

The diagonal form of $\mathbf{C}_{\tilde{\mathbf{v}}}(\omega_{\text{sel}})$ is defined by a unitary coordinate transformation with the matrices $\mathbf{Q}(\omega_{\text{sel}})$ and $\mathbf{Q}^T(\omega_{\text{sel}})$, which contain the normalized orthogonal Eigenvectors of $\mathbf{C}_{\tilde{\mathbf{v}}}$

for frequency ω_{sel} as columns and rows, respectively.

$$\begin{pmatrix} \lambda_1(\omega_{\text{sel}}) & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_{3N}(\omega_{\text{sel}}) \end{pmatrix} = \mathbf{Q}^T(\omega_{\text{sel}}) \mathbf{C}_{\tilde{\mathbf{v}}}(\omega_{\text{sel}}) \mathbf{Q}(\omega_{\text{sel}}) \quad (7)$$

The trace of the matrix $\mathbf{C}_{\tilde{\mathbf{v}}}(\omega_{\text{sel}})$ is invariant under unitary transformations, which means that the Eigenvalue sum describes the VDoS at frequency ω_{sel} in analogy to Eq. 6.

$$I_{\text{VDoS}}(\omega_{\text{sel}}) = \frac{2}{k_B T} \sum_i^{3N} \lambda_i(\omega_{\text{sel}}) \quad (8)$$

In other words, each Eigenvalue $\lambda_i(\omega_{\text{sel}})$ provides a direct measure of the contribution to the VDoS at frequency ω_{sel} for vibrations along the corresponding Eigenvector $\mathbf{Q}_i(\omega_{\text{sel}})$.

Vibrational Dynamics for Individual Modes

In the following, we define an analysis of the fluctuations of a $3N$ -dimensional system along a single vibrational mode $\mathbf{q}$, which describes a normalized vector of displacements in Cartesian coordinates with $3N$ components q_i . This procedure is independent from the method used to define the vibrational mode $\mathbf{q}$, which can thus describe a harmonic or quasi-harmonic normal mode or a FRESEAN mode obtain for a given frequency ω_{sel} as introduced in the previous section. As a first step, we project the dynamics sampled in a MD trajectory onto the vibrational mode $\mathbf{q}$. Specifically, we project mass-weighted velocities in Cartesian coordinates on $\mathbf{q}$ to obtain the one-dimensional mass-weighted velocity along the vibrational mode $\dot{q}(t)$.

$$\dot{q}(t) = \sum_i^{3N} q_i \tilde{v}_i(t) \quad (9)$$

The projected mass-weighted velocity, $\dot{q}(t)$, describes the dynamics of the system along the one-dimensional collective coordinate $\mathbf{q}$. We use this to define a corresponding time auto-correlation function and its Fourier transform that describes the contribution of $\mathbf{q}$ to the VDoS at all frequencies ω , i.e., the one-dimensional (1D) VDoS along vibrational mode $\mathbf{q}$.

$$C_q(\tau) = \langle \dot{q}(t) \dot{q}(t + \tau) \rangle_t \quad (10)$$

$$C_q(\omega) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp(i\omega\tau) C_q(\tau) d\tau \quad (11)$$

$$I_{\text{VDoS}}^q(\omega) = \frac{2}{k_B T} C_q(\omega) \quad (12)$$

If $\mathbf{q}$ is an Eigenvector $\mathbf{Q}_i(\omega_{\text{sel}})$ obtained from Eq. 7 for a specific frequency ω_{sel} , $I_{\text{VDoS}}^q(\omega_{\text{sel}})$ for that frequency is equal to its Eigenvalue $\lambda_i(\omega_{\text{sel}})$. Further, the integral of $I_{\text{VDoS}}^q(\omega)$ over all frequencies ω is equal to 1 because it describes the VDoS contribution of a single collective degree of freedom.

Simulation & Analysis Protocol

To demonstrate the applicability of the FRESEAN mode analysis to biomolecular systems, we describe in the following its application to a small model protein. In this example, we include all atomistic details but note that a similar analysis can be performed using velocities of only a subset of the atoms, e.g., protein backbone C_α -atoms, to focus on the dynamics of secondary structure elements. Likewise, the approach described here can be adapted for the analysis of time derivatives of internal coordinates instead of atomic velocities in Cartesian coordinates.

We selected the small 20 amino acid protein Trp-cage (PDBID: 2job) for our fully anharmonic analysis of vibrational modes. The choice of a small system facilitates the interpretation of individual vibrational modes, even if all $3N$ degrees of freedom are taken into account and all sampled frequencies are analyzed. However, we note that our anharmonic

approach can be easily applied to significantly larger proteins, especially if the analysis is focused on low-frequencies and backbone vibrations.

The GROMACS 2018.1 software package was used for our simulations.³¹ The protein was placed in the center of a 40 Å by 40 Å by 40 Å simulation box and was subsequently solvated with 2073 water molecules. We used the AMBER99SB-ILDN force field³² and the TIP3P water model³³ to describe the potential energy of the system. No constraints were used for intramolecular bonds of the protein to sample the corresponding vibrations, but the SETTLE algorithm³⁴ was used to constrain the geometry of water molecules. To ensure proper sampling of any resulting mid-infrared vibrations, the simulation time step was set to 0.5 femtoseconds (fs) in all MD simulations. Short-ranged electrostatic and Lennard-Jones interactions were treated with a 10 Å real-space cutoff with energy and pressure corrections for dispersion interactions. Long-ranged electrostatic interactions were treated with the Particle Mesh Ewald algorithm³⁵ using a 1.2 Å grid and fourth order interpolation.

The potential energy of the system was initially minimized with a steepest descent algorithm until the maximum atomic force was ≤ 100 kJ/(mol Å). This was followed by an equilibration in the isobaric-isothermal (NPT) ensemble at 300 K and 1 bar for 100 picoseconds (ps). Berendsen thermostats³⁶ were applied separately to the protein and solvent with a time constant of 1.0 ps to control the temperature and a Berendsen barostat with a time constant of 2.0 ps and bulk water compressibility of 4.5×10^{-5} bar⁻¹ was used to control the pressure. This was followed by a production simulation of 1 nanosecond (ns) length in the NPT ensemble. For this simulation, Nosé-Hoover thermostats^{37,38} and the Parrinello-Rahman barostat³⁹ were applied with otherwise unchanged parameters. Coordinates and velocities from this simulation were stored every 4 fs and used for the subsequent analysis.

FRESEAN Mode Analysis

The vibrational analysis was performed in a reference frame aligned with the initial orientation of the protein in our simulation. Therefore, we pre-processed the coordinates and

velocities of all protein atoms in the simulation trajectory with a rotational matrix that compensates for the rotation of the protein during the simulation. The latter was determined by a minimization of the root mean squared deviations (RMSD) of protein atom coordinates relative to the initial structure. Time auto- and cross-correlation functions of weighted atomic velocities as defined in Eq. 4 were computed using a variant of the convolution theorem^{20,27} and transformed between the time and frequency domains as needed for processing. The maximum correlation time was set to 2.0 ps and both auto- and cross-correlation functions were symmetrized in time to enforce equilibrium ensemble properties. This allowed for a frequency resolution of 0.25 THz for the Fourier analysis in Eq. 5. For the Fourier transform in Eq. 5, we further employed a Gaussian window function with a 0.3 THz bandwidth.

Harmonic Normal Mode Analysis

To allow for comparisons with the FRESEAN mode analysis, we also performed harmonic normal mode analysis on the Trp-cage protein. For this purpose, 10 snapshots of the protein were extracted from the production simulation at regular intervals of 100 ps. The resulting protein configurations were then subjected to an energy minimization using the low-memory BFGS optimizer implemented in GROMACS^{40,41} (compiled with double precision) until all forces decreased to less than 0.01 kJ/(mol Å). We then computed the mass-weighted Hessian matrix, $\mathbf{H}$, of the protein potential energy, E_{pot} , with coordinates x_i and masses m_i for the $3N$ degrees of freedom of the protein, with the following matrix elements H_{ij} :

$$H_{ij} = \frac{1}{\sqrt{m_i m_j}} \frac{\partial^2 E_{\text{pot}}}{\partial x_i \partial x_j} \quad (13)$$

We then computed the Eigenvalues and Eigenvectors of $\mathbf{H}$. The Eigenvectors correspond to the harmonic vibrational modes, while the harmonic oscillator (HO) frequencies (in rad/s)

are described by the Eigenvalues λ^{H} .

$$\omega_i^{\text{HO}} = \sqrt{\lambda_i^{\text{H}}} \quad (14)$$

The first 6 Eigenvalues are equivalent to 0 and the corresponding Eigenvectors describe linear combinations of the 3 translational and 3 rotational degrees of freedom of the protein.

Quasi-Harmonic Normal Mode Analysis

To allow for additional comparisons, we also performed a quasi-harmonic normal mode analysis based on the simulation trajectory of the Trp-cage protein. For this purpose, we performed translational and rotational fitting of the coordinates in the Trp-cage protein trajectory to minimize the RMSD relative to the initial structure. We then computed the average coordinates $\langle x_i \rangle$ for all degrees of freedom and computed the mass-weighted co-variance matrix, $\mathbf{C}^{\text{covar}}$, of displacements from the average structure with the following matrix elements C_{ij}^{covar} :

$$C_{ij}^{\text{covar}} = \sqrt{m_i m_j} \langle x_i - \langle x_i \rangle \rangle \langle x_j - \langle x_j \rangle \rangle \quad (15)$$

We then computed the Eigenvalues and Eigenvectors of $\mathbf{C}^{\text{covar}}$. Assuming that the simulation trajectory describes thermal fluctuations around a single minimum of a harmonic potential, the Eigenvectors correspond to (quasi-)harmonic vibrational modes while the (quasi-)harmonic (QH) frequencies (in rad/s) are described by the Eigenvalues λ^{covar} .

$$\omega_i^{\text{QH}} = \sqrt{\frac{k_B T}{\lambda_i^{\text{covar}}}} \quad (16)$$

Results & Discussion

To illustrate the properties of thermally excited vibrations in proteins, we analyzed our classical MD simulation of the Trp-cage mini protein⁴² (see Fig. 1) in aqueous solution using

the theoretical framework described above. As mentioned earlier, a simpler two-dimensional model system is described in the SI to demonstrate the working mechanism of the FRESEAN mode analysis and to compare it to established methods.

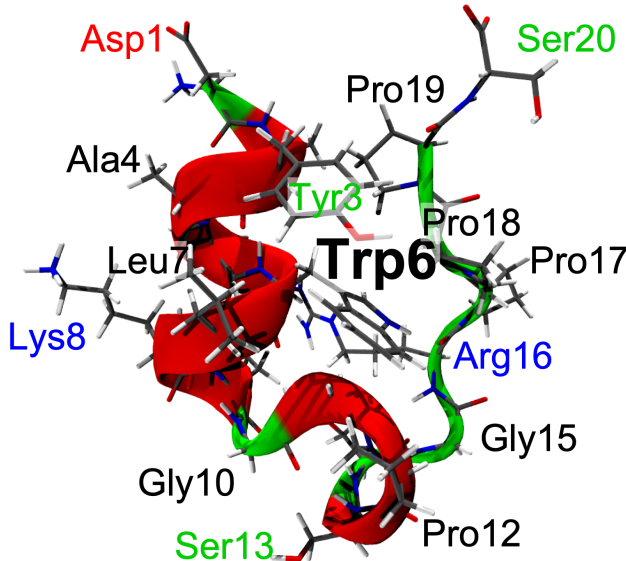


Figure 1: Structure of the Trp-cage mini protein. Secondary structure is highlighted as a red (α -helix) and green cartoon and chemical bonds are shown as CPK-colored sticks. Colored labels identify visible amino acid side-chains (blue: positively charged; red: negatively charged; green: polar; black: non-polar).

Vibrational Density of States

With 284 atoms, the Trp-cage protein has 852 Cartesian degrees of freedom and its vibrational density of states as defined in Eq. 6 features a continuous band of vibrational modes between 0 and 2000 cm^{-1} as shown in Fig. 2. Additional C-H, N-H and O-H vibrations are sampled at higher frequencies but are not shown for clarity. The background color at each frequency illustrates the logarithm of the number of thermally populated quantum HO states at 300 K (see Eq. 1), which highlights differences between vibrations at far- and mid-infrared frequencies.

The broad and continuous band in the VDoS indicates that each sampled frequency is associated with multiple vibrational modes. In harmonic and quasi-harmonic representations,

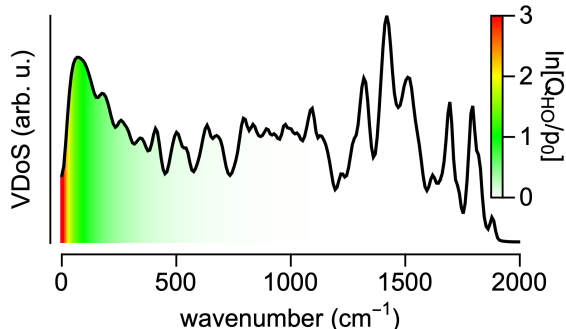


Figure 2: Vibrational density of states (VDoS) of the Trp-cage protein from 0-2000 cm^{-1} (frequencies $>2000 \text{ cm}^{-1}$ not shown for clarity). Color illustrates the logarithm of the number of thermally populated quantum HO states at 300 K for each frequency.

such a scenario results in degenerate representations of these modes since the Eigenvalues of the corresponding matrices in Eqs. 13 and 15 are define the vibrational frequencies.²¹ In contrast, the FRESEAN mode analysis allows us to identify unique vibrational modes based on their distinct contributions to the VDoS at any given frequency.

FRESEAN Eigenvalues

In Figure 3, we show sorted Eigenvalues obtained for selected frequencies in panel A and the resulting number of Eigenvalues required to describe (via summation according to Eq. 8) 25%, 50% and 75% of the total VDoS at each frequency in panel B. Our observations show that single vibrational modes are insufficient to describe the vibrations of the protein at any of the analyzed frequencies. However, only a fraction ($<10\%$) of the 852 orthogonal degrees of freedom with their corresponding Eigenvalues contribute to the vibrational spectrum at any frequency.

Projections on FRESEAN Modes

To confirm that the FRESEAN modes indeed describe vibrations/fluctuations at the selected frequencies, we utilize Eqs. 9 to 12 to project the simulation trajectory onto each mode and compute the corresponding 1D VDoS. The results are shown in Fig. 4 for the FRESEAN

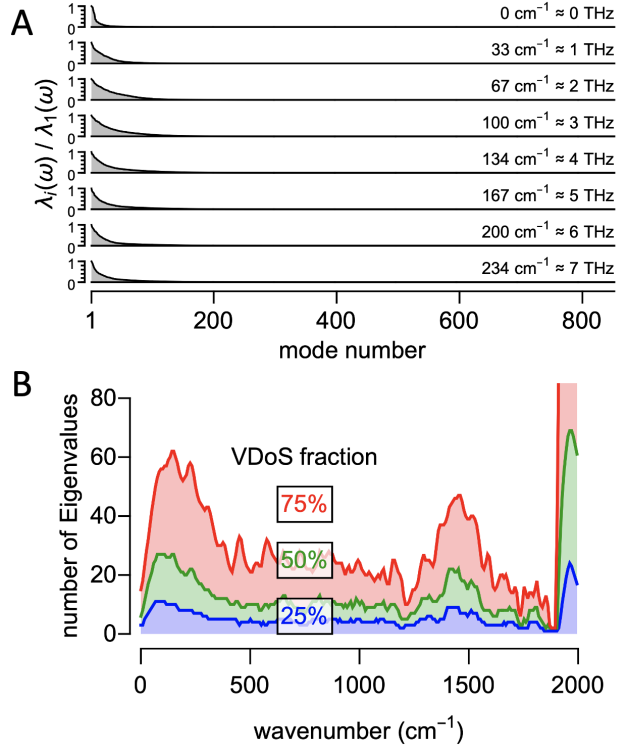


Figure 3: Analysis of frequency-dependent Eigenvalues. (A) Relative Eigenvalues (normalized by first Eigenvalue) of the frequency-dependent velocity cross-correlation matrix for selected frequencies between 0 and 7 THz (234 cm⁻¹). (B) Number of Eigenvalues (out of 852) needed to re-construct 25% (blue), 50% (green) and 75% (red) of the total VDoS (see Eq. 8).

modes (Eigenvectors) associated with the 10 largest VDoS contributions (Eigenvalues) at 0, 1 and 2 THz. For clarity, we use THz to identify frequencies selected for the FRESEAN mode analysis (ω_{sel}) and wavenumbers in cm^{-1} to define the frequency axis of vibrational spectra ($1 \text{ THz} \approx 33 \text{ cm}^{-1}$).

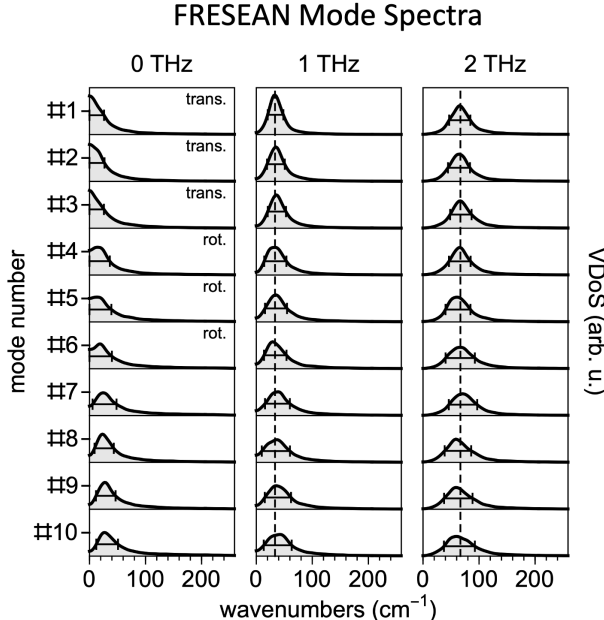


Figure 4: 1D VDoS for projections of the simulation trajectory on FRESEAN modes obtained at 0, 1 and 2 THz. At each selected frequency, spectra are shown for the modes with the 10 largest Eigenvalues. All spectra are normalized by their integral. For 1 and 2 THz, dashed vertical lines indicate the frequency for which the modes were selected. For each spectrum, black bars indicate the full width at half maximum (FWHM) of the peak.

For 0 THz, the FRESEAN mode analysis describes contributions of collective degrees of freedom to the VDoS at zero frequency, which reports on diffusive and relaxational dynamics. The first six modes at 0 THz correspond to translational and rotational diffusion of the protein as indicated in Fig. 4 and visualized in Fig. S4 of the SI. For the translations, the maximum of the VDoS is observed at 0 cm^{-1} , while the VDoS of the rotations peak at non-zero frequencies indicating librations of the protein in the solvation environment in addition to rotational diffusion. The FRESEAN modes #7-10 at 0 THz describe low-frequency vibrations of the protein with peak frequencies just below 1 THz but significant zero-frequency intensities. The latter can be attributed to anharmonic potentials and/or

damping, e.g., efficient exchange of vibrational energy with the surrounding solvent or other vibrational modes in the protein.

At 1 and 2 THz, the FRESEAN mode analysis identifies vibrational modes with well-defined peaks in the 1D VDoS at the frequency for which they were selected. The line widths characterized by the FWHM shown in Fig. 4 vary between 30 and 50 cm^{-1} . The 1D VDoS along the FRESEAN modes do not contain signals at any other frequencies, which demonstrates that the FRESEAN modes successfully isolate the collective degrees of freedom responsible for low-frequency vibrations. The following comparison to vibrational modes obtained from harmonic and quasi-harmonic normal mode analysis shows that both methods fail this test.

We computed harmonic and quasi-harmonic normal modes as described in a previous section. Harmonic normal modes were computed for 10 distinct snapshots of the simulation trajectory which resulted in 10 distinct sets of $3N$ harmonic normal modes. We observed that the qualitative results discussed below are reproducible between all 10 sets of harmonic normal modes so that we limit our analysis below to one of them. Quasi-harmonic normal modes were obtained as a single set of $3N$ Eigenvectors of the co-variance matrix computed as an average over the simulation trajectory.

The frequencies of harmonic and quasi-harmonic normal modes were determined based on Eqs. 14 and 16. For both, these frequencies were used to select 10 modes with predicted frequencies closest to: 1) 0 THz; 2) 1 THz; and 3) 2 THz. We then used Eqs. 9 to 12 to project the simulation trajectory on each of these modes and calculated the corresponding 1D VDoS. For far-infrared frequencies, the resulting 1D VDoS are shown in Fig. 5A for harmonic normal modes and in Fig. 5B for quasi-harmonic normal modes.

The harmonic normal modes #1-6 describe linear combinations of translations and rotations and thus result in 1D VDoS comparable to the FRESEAN modes #1-6 obtained at 0 THz in Fig. 4. However, the 1D VDoS of projections on harmonic normal modes #7-10 exhibit broad signals with FWHM's that exceed 100 cm^{-1} (e.g., mode #10). In addition,

the peak frequencies are significantly higher than the harmonic frequencies predicted by the Eigenvalues of the mass-weighted Hessian matrix (less than 10 cm^{-1}).

For harmonic normal modes with predicted frequencies close to 1 and 2 THz, the VDoS of the mode projections feature a main peak with a maximum intensity close to the predicted frequency (Eq. 14). However, the FWHM's are significantly larger than for the FRESEAN modes selected for the same frequencies and range from 50 to 75 cm^{-1} .

The broad bands of the 1D VDoS for vibrations along the harmonic normal modes can be easily misinterpreted as the result of anharmonic vibrations along a properly identified vibrational mode of the system. However, this interpretation is misleading because the FRESEAN modes show that vibrational modes with a significantly narrower spectrum exist at the corresponding frequency. Hence, the approximations implied by the definition of the harmonic normal modes amplify their perceived anharmonicity.

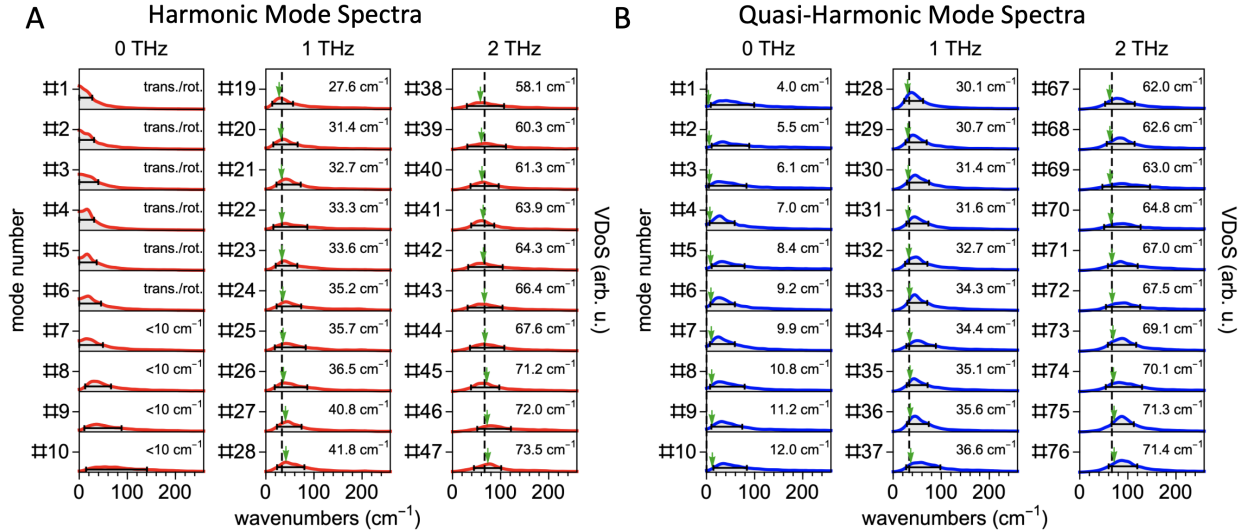


Figure 5: 1D VDoS for projections of the simulation trajectory on (A) harmonic (one representative set) and (B) quasi-harmonic normal modes. In each case, the spectra are shown for the first 10 modes (lowest predicted frequencies), and 10 modes with predicted frequencies closest to 1 THz and 2 THz, respectively (dashed vertical lines indicate 1 and 2 THz). All spectra are normalized by their integral and shown on the same scale as the VDoS for FRESEAN modes shown in Fig. 4. Green arrows indicate the predicted harmonic and quasi-harmonic frequencies (also given numerically as an inset). For each spectrum, black bars indicate the full width at half maximum (FWHM) of the peak.

Due to differences in the formalism, the first 6 quasi-harmonic normal modes do not

describe protein translations or rotations in contrast to harmonic normal modes. Instead, the quasi-harmonic normal modes with the 10 lowest predicted frequencies (largest Eigenvalues of co-variance matrix) describe distortions of the protein structure, which would correspond to vibrational modes in a harmonic system. However, we observe for all selected quasi-harmonic normal modes (the first 10 mode and the modes with predicted frequencies close to 1 and 2 THz) that the quasi-harmonic frequencies are consistently lower than the frequency of the main peak in the corresponding 1D VDoS. Further, the FWHM's of the main peaks range from 40 to 100 cm^{-1} and are thus again significantly larger than the spectra of vibrations along the FRESEAN modes in Fig. 4.

Similar to the 1D VDoS obtained for harmonic normal modes, it would be incorrect to use the line shape of vibrational bands to characterize the anharmonicity of low-frequency vibrations along quasi-harmonic normal modes. Harmonic and quasi-harmonic normal modes both fail to identify vibrational modes that isolate the low-frequency vibrations of the protein, which amplifies apparent anharmonicities. One indication is the integral of intensities in the far-infrared spectrum obtained for the harmonic and quasi-harmonic normal modes in Fig. 5. The intensity scale used in Figs. 4 and 5 is identical and the integral of each 1D VDoS is by definition equal to 1. Hence, missing intensities in the 1D VDoS for far-infrared frequencies indicate that there are additional contributions at higher frequencies that are not observed in Fig. 5. The latter becomes evident in a comparison of the 1D VDoS computed for FRESEAN, harmonic, and quasi-harmonic modes at 1 THz over the full range of sampled frequencies shown in Fig. 6.

In the discussion of Fig. 4, we indicated that the 1D VDoS obtained for FRESEAN modes feature only a single peak at the frequency that they were selected for. This is confirmed in Fig. 6, where the 1D VDoS of FRESEAN modes selected at 1 THz are shown with a 20-fold amplification for frequencies up to 4000 cm^{-1} . No additional peaks or signals are observed and even the magnified intensities of the high-frequency tail of the main peak are indistinguishable from zero for frequencies above 300 cm^{-1} . In contrast, the VDoS for

harmonic and quasi-harmonic normal modes feature high frequency tails that remain non-zero up to frequencies of 500 cm^{-1} . In addition, multiple resonance peaks are observed at mid-infrared frequencies. For the harmonic normal modes, such signals are primarily found around 1400 and 3000 cm^{-1} , which are associated with characteristic vibrations of individual functional groups. For the quasi-harmonic normal modes, the additional peaks are primarily restricted to the fingerprint region below 1200 cm^{-1} . Additional resonances are inconsistent with the harmonic oscillator model, which demonstrates that the harmonic and quasi-harmonic normal modes fail to isolate the low-frequency vibrations of the simulated protein. On the other hand, the absence of additional peaks in the VDoS of FRESEAN modes demonstrates that the latter successfully isolate collective degrees of freedom associated with low-frequency vibrations in the protein. We observed analogous results for the 1D VDoS along vibrational modes selected for diffusive dynamics (0 THz) and vibrations at 2 THz .

As a consequence of these findings, oscillations along the FRESEAN modes are actually in significantly better agreement with the harmonic oscillator model than oscillations along the harmonic and quasi-harmonic modes. Analyzing vibrations based on harmonic and quasi-harmonic normal modes thus risks to significantly overestimate the anharmonicity of the actual vibrations in a system, because the anharmonicity of the potential affects both the definition of vibrational modes and the fluctuations along them.

Anharmonicity of Low-Frequency Vibrations

To obtain more information on the potential energy surface associated with each FRESEAN mode, we analyze the probability distributions of displacements along each mode relative to the average structure:

$$q(t) = \sum_i^{3N} q_i \sqrt{m_i} [x_i(t) - \langle x_i \rangle] \quad (17)$$

The results are shown in Fig. 7. At 0 THz , we skipped the translational and rotational degrees of freedom and analyzed the displacement distributions for modes #7-16 instead.

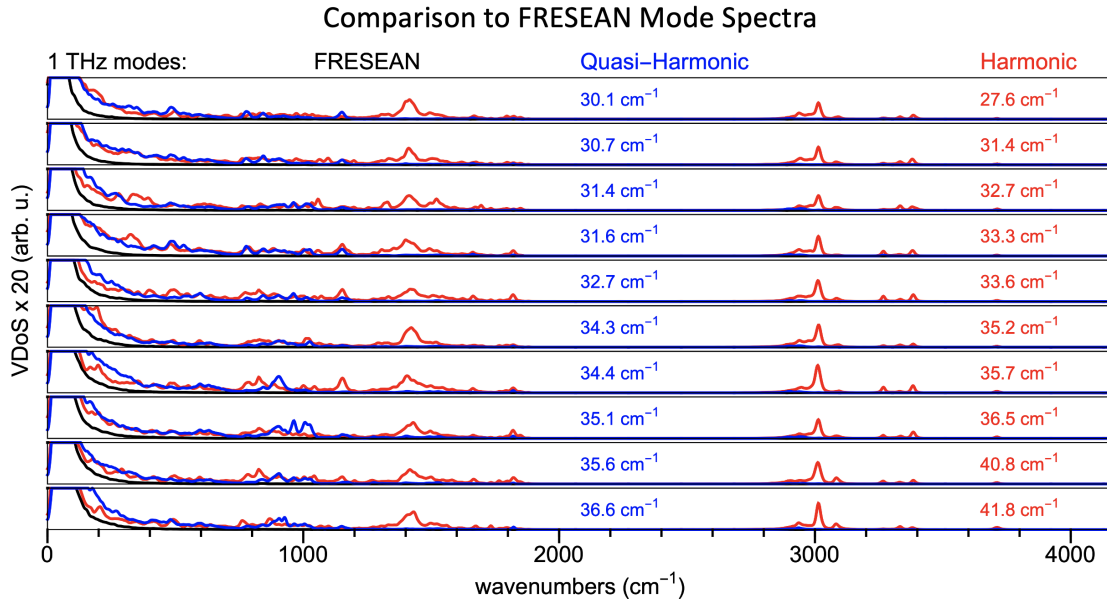


Figure 6: Comparison of 1D VDoS for projections of the simulation trajectory on 1 THz modes obtained with the FRESEAN mode analysis (black, modes #1-10), harmonic normal mode analysis (red, modes #19-28) and quasi-harmonic normal analysis (blue, modes #67-76). The 1D VDoS are identical to the 1 THz modes shown in Figs. 4 and 5, but the intensities are scaled by a factor of 20 and the shown frequency range is increased to 4000 cm^{-1} . The predicted frequencies used to select harmonic and quasi-harmonic normal modes for this comparison are given in red and blue, respectively.

For harmonic oscillators, such distributions would be described by Gaussians with a standard deviation σ_0 that is entirely defined by the harmonic oscillator frequency.

$$\sigma_0 = \sqrt{k_B T} / \omega_0 \quad (18)$$

In reality, i.e., for coupled anharmonic oscillators, the distributions are related to potentials of mean force that depend on time-averaged potentials and include entropic information due to correlations between distinct degrees of freedom.

To illustrate how displacement distributions along the FRESEAN modes deviate from the expected behavior of harmonic oscillators, we determined for each mode the position of the peak in the corresponding VDoS (see Fig. 4). Based on the peak frequency, we then plotted the corresponding Gaussian distribution expected for a harmonic oscillator at that frequency alongside the actual histogram sampled from the simulation in Fig. 7.

The histograms obtained from the simulations are unimodal with the exception of mode #15 at 0 THz, which is bimodal and thus describes barrier crossing events between two states already on the 1 ns time scale of the simulation. In addition, the plotted histograms feature varying degrees of asymmetry and, most of all, are significantly wider than expected for harmonic oscillators at the corresponding frequency. The increased width of the distributions is most pronounced for the FRESEAN modes at 0 THz, thus highlighting the anharmonicity of vibrations along these collective degrees of freedom. With increasing frequencies of the FRESEAN modes (1 and 2 THz) the sampled histograms become more similar to the Gaussians predicted for harmonic oscillations. However, at 2 THz the deviations from the expected behavior of harmonic oscillators are still noticeable.

Cross-Correlations

Another deviation from the harmonic oscillator model is the ability of vibrational modes to exchange energy with each other. The energy exchange between distinct modes can be

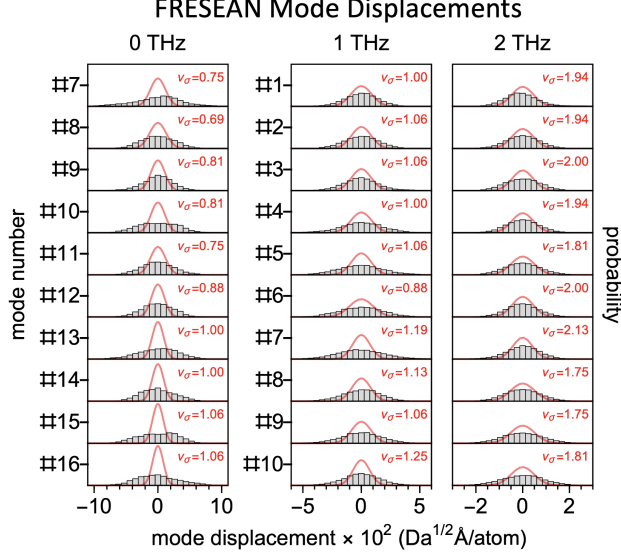


Figure 7: Histograms of displacement probability distributions along FRESEAN modes for (A) 0 THz, (B) 1 THz and (C) 2 THz. For 0 THz, translations and rotations were omitted and distributions are shown for vibrational modes #7-16. For 1 THz and 2 THz, distributions are shown for FRESEAN modes #1-10 with the largest Eigenvalues at the respective frequency. Each histogram is compared to the Gaussian distribution (red) of a harmonic oscillator with the frequency of the main peak in the 1D VDoS of each mode (indicated as $\nu_0 = \omega_0/(2\pi)$ in THz).

visualized via time cross-correlations of weighted velocities along distinct modes, which we analyze in Fig. 8.

These time auto- and cross-correlation functions can be obtained from projections of the simulated trajectory on each mode (see Eq. 9), or directly from the cross-correlation matrix $\mathbf{C}_{\tilde{\mathbf{v}}}(\omega)$. In the latter case, a unitary transformation with the Eigenvectors obtained for a selected frequency ω_{sel} is applied to $\mathbf{C}_{\tilde{\mathbf{v}}}(\omega)$ at all frequencies.

$$\mathbf{C}'_{\tilde{\mathbf{v}}}(\omega) = \mathbf{Q}^T(\omega_{\text{sel}}) \mathbf{C}_{\tilde{\mathbf{v}}}(\omega) \mathbf{Q}(\omega_{\text{sel}}) \quad (19)$$

The diagonal terms, $C'_{ii}(\omega)$, of the frequency-dependent matrix $\mathbf{C}'_{\tilde{\mathbf{v}}}(\omega)$ describe the 1D VDoS along each of the Eigenvectors, while the off-diagonal terms, $C'_{ij}(\omega)$, which are zero for $\omega = \omega_{\text{sel}}$, describe the spectra of cross-correlations. An inverse Fourier transform from the frequency into the time domain results in the corresponding time auto- and cross-correlation

functions for velocity fluctuations, $\dot{q}_i(t)$, along each of the FRESEAN modes, $\mathbf{q}_i$, generated for frequency ω_{sel} .

$$\langle \dot{q}_i(t) \dot{q}_j(t + \tau) \rangle_t = \int_{-\infty}^{\infty} \exp(-i\omega\tau) C'_{ij}(\omega) d\omega = C'_{ii}(\tau) \quad (20)$$

$C'_{ii}(\tau = 0)$ is directly related to the kinetic energy per degree of freedom and independent of the index i . With increasing delay time τ , the time auto-correlation functions shown in blue in Fig. 8 visualize the oscillations along each mode in the time domain. For each mode analyzed here, the resulting oscillations are reminiscent of damped oscillators.

Damping is caused by vibrational energy transfer between vibrational modes, both within the protein and the surrounding solvent. The vibrational energy transfer between FRESEAN modes of the protein can be observed in time cross-correlation functions shown in red in Fig. 8. We only analyzed cross-correlations between FRESEAN modes obtained at the same frequency ω_{sel} . As a consequence, the modes describe orthogonal Eigenvectors, which implies that cross-correlations are zero at $\tau = 0$ (apart from numerical noise). However, with increasing delay time τ , we observe the onset of correlated oscillations in pairs of distinct modes in Fig. 8, which decay back to zero on longer time scales.

The energy transfer between vibrational modes can be direct, i.e., within a single oscillation period at the corresponding frequency, or indirect via multiple modes for longer correlation times. From the cross-correlations in Fig. 8, it is apparent that the energy transfer between some pairs of vibrational modes, whether direct or indirect, is more efficient than for others. These results may provide new insights into energy transport pathways in proteins, which were previously analyzed in terms of harmonic normal modes.^{43,44}

Our analysis currently does not resolve the transfer of vibrational energy to the solvent, but we expect the latter to contribute significantly on timescales >1 ps. In previous work, we observed correlations between atomic velocities in proteins and solvating water molecules over length-scales of up to 25 Å.⁴⁵

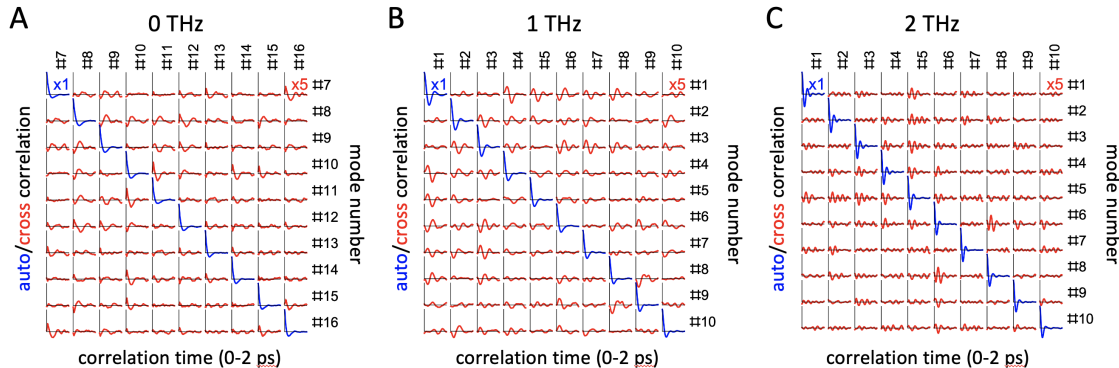


Figure 8: Time auto- and cross-correlation functions (blue and red, respectively) of mass-weighted velocities projected along FRESEAN modes at (A) 0 THz, (B) 1 THz and (C) 2 THz. For 0 THz, translations and rotations were omitted and correlations were analyzed for the vibrational modes #7-16. For 1 THz and 2 THz, correlations are shown for FRESEAN modes #1-10 with the 10 largest Eigenvalues at the respective frequency.

Conclusions

We developed the FRESEAN mode analysis of molecular vibrations, which allows for an unambiguous assignment of collective degrees of freedom associated with vibrations at any given frequency. Apart from the potential energy model used in molecular simulations, e.g., an empirical force field, no additional assumptions are required. Our methodology is free from harmonic and quasi-harmonic approximations and significantly improves our ability to characterize low-frequency vibrations of complex molecules at far-infrared frequencies. These vibrations are not restricted to their vibrational ground state at room temperature and thus tend to exhibit pronounced anharmonic properties. The collective degrees of freedom associated with these vibrations are critical for our understanding of thermal fluctuations in proteins and other complex molecules^{17,18}

A key feature of our anharmonic analysis is that it is entirely based on a time correlation formalism that uses information from MD trajectories. Until now, time correlation formalisms were only capable to evaluate the overall vibrational spectrum of a system, e.g., the vibrational density of states or the absorption spectrum.^{15,26,27} Assignments of collective degrees of freedom to vibrational frequencies primarily relied on normal modes and harmonic approximations.^{1,2,18} However, fluctuations along harmonic and quasi-harmonic

normal modes are contaminated by oscillations at a multitude of distinct frequencies, which is inconsistent with the harmonic oscillator model. Characterizing anharmonic properties of low-frequency vibrations based on such modes can thus be misleading because deviations from harmonic behavior result from both the definition of the modes and the actual vibrations of the system. Instead, the FRESEAN mode analysis successfully isolates the collective degrees of freedom associated with low-frequency vibrations. This allows for a detailed analysis of anharmonic properties, e.g., anharmonic fluctuations and vibrational energy exchange between vibrational modes.

Vibrational modes and the frequency spectrum of an anharmonic system evolve continuously as a function of time, especially if the system transitions between distinct minima of the potential energy surface. This presents a challenge for standard methods for the characterization of molecular vibrations, which led to the development of instantaneous normal modes and normal mode ensemble analysis.^{12,18} In the FRESEAN mode analysis, the time evolution of vibrational properties is naturally accounted for by non-integer populations of each vibrational mode that describe the respective contributions to the ensemble-averaged vibrational density of states.

Here, we implemented the FRESEAN mode analysis for Cartesian velocities of the simulated Trp-cage mini protein using an all-atom representation. We note that the approach is straightforward to adapt for other representations of a simulated protein. For example, the analysis can be performed for a subset of atoms, e.g., backbone C_α -atoms, other coarse-grained representations such as residue center-of-mass velocities, and time-derivatives of internal coordinates.

At the lowest frequencies, we anticipate FRESEAN mode analysis to be a powerful tool for the high throughput prediction of collective degrees of freedom associated with large amplitude motion and conformational transitions in proteins and other complex molecules. Further, an analysis of molecular vibrations free from harmonic approximations may yield novel approaches to estimate thermodynamic properties such as conformational entropy.

In addition, we expect the FRESEAN mode analysis to be a critical tool for the correct assignment of far-infrared vibrations observed experimentally, e.g., in anisotropic terahertz microspectroscopy^{46,47} or optical Kerr effect spectroscopy.¹⁷

Acknowledgement

This work is supported by the National Science Foundation (CHE-2154834) and the National Institute of General Medical Sciences (1R01GM148622-01). The authors acknowledge Research Computing at Arizona State University for providing high performance computing resources that have contributed to the research results reported within this work.

References

- (1) Brooks, B.; Karplus, M. Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. U.S.A* **1983**, *80*, 6571–6575.
- (2) Brooks, B. R.; Janežič, D.; Karplus, M. Harmonic analysis of large systems. I. Methodology. *J. Comput. Chem.* **1995**, *16*, 1522–1542.
- (3) Levy, R.; Srinivasan, A.; Olson, W.; McCammon, J. Quasi-harmonic method for studying very low frequency modes in proteins. *Biopolymers* **1984**, *23*, 1099–1112.
- (4) Levy, R. M.; De la Luz Rojas, O.; Friesner, R. A. Quasi-harmonic method for calculating vibrational spectra from classical simulations on multi-dimensional anharmonic potential surfaces. *J. Phys. Chem.* **1984**, *88*, 4233–4238.
- (5) Kitao, A.; Hirata, F.; Gō, N. The effects of solvent on the conformation and the collective motions of protein: normal mode analysis and molecular dynamics simulations of melittin in water and in vacuum. *Chem. Phys.* **1991**, *158*, 447–472.

- (6) García, A. E. Large-amplitude nonlinear motions in proteins. *Phys. Rev. Lett.* **1992**, *68*, 2696.
- (7) Amadei, A.; Linssen, A. B.; Berendsen, H. J. Essential dynamics of proteins. *Proteins Struct. Funct. Bioinf.* **1993**, *17*, 412–425.
- (8) Noé, F.; Clementi, C. Kinetic distance and kinetic maps from molecular dynamics simulation. *J. Chem. Theory Comput.* **2015**, *11*, 5002–5011.
- (9) Chennubhotla, C.; Rader, A.; Yang, L.-W.; Bahar, I. Elastic network models for understanding biomolecular machinery: from enzymes to supramolecular assemblies. *Phys. Biol.* **2005**, *2*, S173.
- (10) Yang, L.; Song, G.; Jernigan, R. L. How well can we understand large-scale protein motions using normal modes of elastic network models? *Biophys. J.* **2007**, *93*, 920–929.
- (11) Mahajan, S.; Sanejouand, Y.-H. Jumping between protein conformers using normal modes. *J. Comput. Chem.* **2017**, *38*, 1622–1630.
- (12) Peng, C.; Zhang, L.; Head-Gordon, T. Instantaneous normal modes as an unforced reaction coordinate for protein conformational transitions. *Biophys. J.* **2010**, *98*, 2356–2364.
- (13) Stuart, B. H. *Infrared spectroscopy: fundamentals and applications*; John Wiley & Sons, 2004.
- (14) Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (15) Chakraborty, S.; Sinha, S.-K.; Bandyopadhyay, S. Low-frequency vibrational spectrum of water in the hydration layer of a protein: A molecular dynamics simulation study. *J. Phys. Chem. B* **2007**, *111*, 13626–13631.

- (16) He, Y.; Ku, P. I.; Knab, J.; Chen, J.; Markelz, A. Protein dynamical transition does not require protein structure. *Phys. Rev. Lett.* **2008**, *101*, 178103.
- (17) Turton, D. A.; Senn, H. M.; Harwood, T.; Lapthorn, A. J.; Ellis, E. M.; Wynne, K. Terahertz underdamped vibrational motion governs protein-ligand binding in solution. *Nat. Commun.* **2014**, *5*, 1–6.
- (18) Romo, T. D.; Grossfield, A.; Markelz, A. G. Persistent protein motions in a rugged energy landscape revealed by normal mode ensemble analysis. *J. Chem. Inf. Model.* **2020**, *60*, 6419–6426.
- (19) Yu, X.; Leitner, D. M. Anomalous diffusion of vibrational energy in proteins. *J. Chem. Phys.* **2003**, *119*, 12673–12679.
- (20) Mathias, G.; Baer, M. D. Generalized normal coordinates for the vibrational analysis of molecular dynamics simulations. *J. Chem. Theory Comput.* **2011**, *7*, 2028–2039.
- (21) Na, H.; Song, G. The effective degeneracy of protein normal modes. *Phys. Biol.* **2016**, *13*, 036002.
- (22) Wood, K.; Frölich, A.; Paciaroni, A.; Moulin, M.; Härtlein, M.; Zaccai, G.; Tobias, D. J.; Weik, M. Coincidence of dynamical transitions in a soluble protein and its hydration water: direct measurements by neutron scattering and MD simulations. *J. Am. Chem. Soc.* **2008**, *130*, 4586–4587.
- (23) Tarek, M.; Tobias, D. Role of protein-water hydrogen bond dynamics in the protein dynamical transition. *Phys. Rev. Lett.* **2002**, *88*, 138101.
- (24) Tournier, A. L.; Smith, J. C. Principal components of the protein dynamical transition. *Phys. Rev. Lett.* **2003**, *91*, 208106.
- (25) Straub, J. E.; Thirumalai, D. Exploring the energy landscape in proteins. *Proc. Natl. Acad. Sci. U.S.A* **1993**, *90*, 809–813.

- (26) McQuarrie, D.-A. *Statistical mechanics*; University Science Books: Sausalito, CA 94965, USA, 2000.
- (27) Heyden, M.; Sun, J.; Funkner, S.; Mathias, G.; Forbert, H.; Havenith, M.; Marx, D. Dissecting the THz spectrum of liquid water from first principles via correlations in time and space. *Proc. Natl. Acad. Sci. U.S.A* **2010**, *107*, 12068–12073.
- (28) Lin, S.-T.; Blanco, M.; Goddard III, W. A. The two-phase model for calculating thermodynamic properties of liquids from molecular dynamics: Validation for the phase diagram of Lennard-Jones fluids. *J. Chem. Phys.* **2003**, *119*, 11792.
- (29) Heyden, M.; Sun, J.; Forbert, H.; Mathias, G.; Havenith, M.; Marx, D. Understanding the origins of dipolar couplings and correlated motion in the vibrational spectrum of water. *J. Phys. Chem. Lett.* **2012**, *3*, 2135–2140.
- (30) Heyden, M. Resolving anisotropic distributions of correlated vibrational motion in protein hydration water. *J. Chem. Phys.* **2014**, *141*, 22D509.
- (31) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25.
- (32) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins* **2010**, *78*, 1950–1958.
- (33) Jorgensen, W.-L.; Chandrasekhar, J.; Madura, J.-D.; Impey, R.-W.; Klein, M.-L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (34) Miyamoto, S.; Kollman, P.-A. SETTLE: An analytical version of the SHAKE and RATTLE algorithms for rigid water models. *J. Comput. Chem.* **1992**, *13*, 952–962.

- (35) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (36) Berendsen, H.-J.-C.; Postma, J.-P.-M.; van Gunsteren, W.-F.; DiNola, A.; Haak, J.-R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
- (37) Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984**, *52*, 255–268.
- (38) Hoover, W.-G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.
- (39) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (40) Byrd, R. H.; Lu, P.; Nocedal, J.; Zhu, C. A limited memory algorithm for bound constrained optimization. *SIAM J. Sci. Comput.* **1995**, *16*, 1190–1208.
- (41) Zhu, C.; Byrd, R. H.; Lu, P.; Nocedal, J. Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM T. Math. Software* **1997**, *23*, 550–560.
- (42) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. Designing a 20-residue protein. *Nat. Struct. Biol.* **2002**, *9*, 425–430.
- (43) Yu, X.; Leitner, D. M. Vibrational energy transfer and heat conduction in a protein. *J. Phys. Chem. B* **2003**, *107*, 1698–1707.
- (44) Yu, X.; Leitner, D. M. Heat flow in proteins: computation of thermal transport coefficients. *J. Chem. Phys.* **2005**, *122*, 054902.
- (45) Päs lack, C.; Schäfer, L. V.; Heyden, M. Atomistic characterization of collective protein–water–membrane dynamics. *Phys. Chem. Chem. Phys.* **2019**, *21*, 15958–15965.

- (46) Acbas, G.; Niessen, K. A.; Snell, E. H.; Markelz, A. Optical measurements of long-range protein vibrations. *Nat. Commun.* **2014**, *5*, 3076.
- (47) Singh, R.; George, D. K.; Benedict, J. B.; Korter, T. M.; Markelz, A. G. Improved mode assignment for molecular crystals through anisotropic terahertz spectroscopy. *J. Phys. Chem. A* **2012**, *116*, 10359–10364.

TOC Graphic

