

Examining the origins of observed terahertz modes from an optically pumped atomistic model protein in aqueous solution

Khatereh Azizi ^{a,b}, Matteo Gori ^b, Uriel Morzan ^a, Ali Hassanali ^a and Philip Kurian ^{b,*}

^aThe Abdus Salam International Centre for Theoretical Physics, Strada Costiera 11, 34151 Trieste, Italy

^bQuantum Biology Laboratory, Howard University, Washington, DC 20060, USA

*To whom correspondence should be addressed: Email: pkurian@howard.edu

Edited By: P. Fratzl

Abstract

The microscopic origins of terahertz (THz) vibrational modes in biological systems are an active and open area of current research. Recent experiments [Phys Rev X, 8, 031061 (2018)] have revealed the presence of a pronounced mode at ~ 0.3 THz in fluorophore-decorated bovine serum albumin (BSA) protein in aqueous solution under nonequilibrium conditions induced by optical pumping. This result was heuristically interpreted as a collective elastic fluctuation originating from the activation of a low-frequency phonon mode. In this work, we show that the sub-THz spectroscopic response emerges in a statistically significant manner ($>2\sigma$) from such collective behavior, illustrating how photoexcitation can alter specific THz vibrational modes. We revisit the theoretical analysis with proof-of-concept molecular dynamics that introduce optical excitations into the simulations. Using information theory techniques, we show that these excitations can give rise to a multiscale response involving two optically excited chromophores (tryptophans), other amino acids in the protein, ions, and water. Our results motivate new experiments and fully nonequilibrium simulations to probe these phenomena, as well as the refinement of atomistic models of Fröhlich condensates that are fundamentally determined by nonlinear interactions in biology.

Keywords: terahertz spectra, Fröhlich condensates, bovine serum albumin (BSA) protein, molecular dynamics (MD) simulation, optomechanical transduction

Significance Statement

Recent spectroscopic experiments suggest that photoexcited biomolecules can behave like optomechanical transducers in the THz domain. Photoexcitation may activate an energy downconversion cascade in the biomolecule, resulting in energy “funneling” to a specific (and collective) elastic mode. However, an atomistic picture of this activation mechanism is still lacking. Here, through MD simulations, we explore how the mechanical properties of a solvated model protein at thermodynamic equilibrium are affected by the photoexcitation of intrinsic tryptophan chromophores. In particular, we show how optical excitations and consequent charge reorganization alter these THz collective vibrational modes. Moreover, we show how photoexcitation can give rise to a multiscale response involving not only the amino acids in the protein but also the surrounding counterions and water.

Introduction

In the last decade, there has been an increased interest in studying the terahertz (THz) regime of biological systems (1). Both proteins and the aqueous solvent have a distinct absorption in the THz regime (2–6), extending from the sub-THz to several tens of THz. Advances in THz Fourier-transform infrared (FTIR), time-domain (TDS), and continuous-wave (CW) near-field spectroscopies have illuminated the variety of vibrational modes and timescales associated with the protein-water interface (7–11), which has revealed correlated dynamics occurring at much larger distances than

previously thought possible in a range of proteins (6). However, the molecular origins of these vibrational modes still remain elusive.

In a recent work bridging theory, computation, and experiment (12), the model protein bovine serum albumin (BSA) was decorated with Alexa488 fluorophores and pumped far from equilibrium in an aqueous solution (see Fig. 1), by energy injection from both a blue 488-nm laser and a broadband ultraviolet light-emitting diode (UV LED) (13). While the blue laser served to excite the Alexa488 fluorophores, the UV LED centered at 255 nm can be used to excite the tryptophan amino acid intrinsic to the protein.

Competing Interest: The authors declare no competing interest.

Received: December 28, 2022. **Revised:** July 14, 2023. **Accepted:** July 26, 2023

© The Author(s) 2023. Published by Oxford University Press on behalf of National Academy of Sciences. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

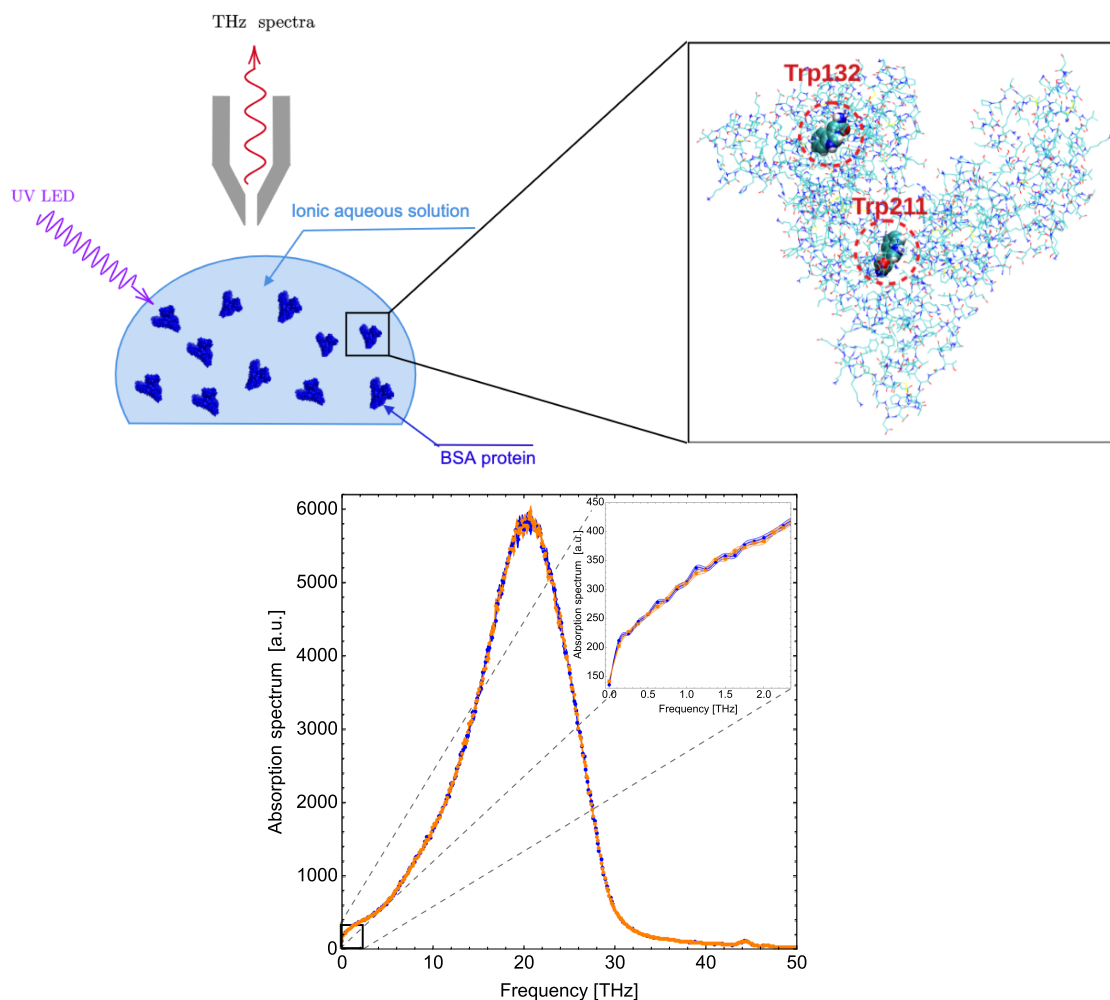


Fig. 1. (Left upper panel) A schematic presentation of the experimental setup motivating our proof-of-concept atomistic simulations studying the terahertz spectra of an aqueous protein solution. Right upper panel (zoomed-in picture): Atomic structure of the protein bovine serum albumin (BSA), with tryptophan residues highlighted. Water and ions are omitted for clarity. (Bottom panel) Simulated absorption coefficient $\alpha_{GS(EX)}(\nu)$ (in arbitrary units, a.u.) of the whole system vs. frequency ν for the BSA aqueous solution, with tryptophans in the ground state (blue) and the photoexcited state (orange). In the inset, the detail of the absorption spectrum of the protein-ions-water system is shown, highlighting the low-frequency range (0–2.35 THz).

Two experimental groups independently found an absorption peak at ~ 0.314 THz under these nonequilibrium (NEQ) conditions.¹ This absorption peak vanished if the fluorophores were removed or if the external pumping ceased. The feature could also be recovered by applying the energy injection again for a sufficiently long time. Heuristic calculations (14) modeling BSA as an elastic sphere predict the lowest-frequency breathing mode at 0.308 THz with a precision of $\sim 2\%$ with respect to the observed absorption peak. These results have been interpreted in terms of the emergence of a pronounced collective phonon mode in the aqueous BSA system due to the activation of an optomechanical energy downconversion cascade from photoexcited chromophores to the specific elastic vibration. Such unexpected protein behavior has been observed in a warm and wet environment, demanding a deeper study of its origins and the sub-THz response in these conditions, which remain poorly understood.

The possibility of activating giant coherent dipole oscillations in biomolecules is a long-standing proposal (12, 15–21), one which

may be realized by the channeling of a large part of the vibrational energy of a biomolecule into its lowest-frequency normal modes. Such giant dipole oscillations in the sub- to few-THz range could have a major impact on the description of biomolecular dynamics. It has already been theoretically proposed (22) and very recently experimentally realized (23) that resonance among these oscillating dipoles is responsible for (selective) long-range electrodynamic interactions, which could play a relevant role in recruiting cognate partners involved in a range of biochemical reactions. Thermal activation of particular “hot” residues by solvent collisions (24), and charting the potentially rate-promoting dissipation channels through protein systems, is also a lively area of current research (25, 26).

A schematic presentation of the experimental setup motivating our atomistic simulations is shown in the upper left panel of Fig. 1 requires UV excitation. It is well known that one of the most important intrinsic chromophores in proteins is tryptophan, which absorbs broadly in the UV and fluoresces generally near 350 nm in an aqueous solution due to its large absorption-emission Stokes shift (27). The two tryptophans in BSA play a critical role in stimulating the observed THz modes after energy downconversion through the protein. In order to take a first step at using hybrid classical

¹ The solution was thermostatted, and therefore the NEQ state of each protein molecule is a stationary one, loosely speaking, as in the case of a lasing medium where pumping and dissipation coexist.

atomistic models with excited-state charge configurations to understand the origins of Fröhlich-like condensates, modeling the naturally occurring tryptophan chromophores is more computationally tractable than including Alexa488 fluorophores, and more widely applicable to other protein systems.

The interpretation of these experiments within the framework of a Fröhlich-like condensate considers the tryptophans as essential to optical pumping, initiating the downconversion process from photon excitation to phonon vibration through the protein, but such a framework provides only a coarse-grained perspective on the water, ions, and protein response. Tryptophan plays dual roles in this phenomenon, as (1) highly efficient absorbers of energetic UV photons, which can change whole-system response properties in the electronic degrees of freedom at thermal equilibrium; and as (2) centers of energy downconversion through nonlinear interactions with the protein and aqueous environment, which when pumped far from thermal equilibrium can populate protein vibrational modes in such a manner that violates energy equipartition. In this article, we consider only the first of these roles, which is an essential prerequisite to analyzing the second from an atomistic point of view.

It should be stressed that the goal of this work is not to reproduce particular experimental conditions but instead to provide a proof-of-concept of how charge reorganization on intrinsic chromophores like tryptophan can alter the low-frequency terahertz response of solvated proteins. Indeed, the probability of a UV lamp or LED simultaneously exciting both tryptophans in BSA with independent photons, on the fluorescence timescale for tryptophan (2–3 ns), is essentially negligible. However, it is worth noting that with an appropriate choice of laser, photoexcitation of both tryptophans—or other attached fluorophores—can be achieved. It is important to remember that in the experiments (12), several protein-attached fluorophores (Alexa488) with a blue laser were used in addition to UV-excited tryptophan to excite the system. So clearly, the specific experimental conditions in that work are far more complicated and beyond the scope of this work. We have thus decided to simplify the problem to gain general insight, and instead focus on constructing an atomistic minimal model where photoexcitation is effectively included in a non-steady charge state of both tryptophans in BSA, which is a conservative underestimate of the total optical excitation energy input per BSA molecule in the experiment. Such a minimal modeling approach has been used effectively in previous theoretical studies to examine the absorption-emission Stokes shift in diverse biomolecular systems (28–30).

In this work, we revisit the analysis of the optically pumped BSA experiments using classical atomistic molecular dynamics in thermal equilibrium, taking the first steps in providing molecular-level information that can be used to refine existing theory on the microscopic origins of the phenomenon. Indeed, at the heart of the Fröhlich rate equations is the nature of the crucial nonlinear couplings with the external pumping field and with the thermal bath, which are not generally modeled with classical molecular dynamics. Nevertheless, as an essential milestone en route to NEQ simulations with such nonlinear interactions included, insights can be gleaned from the standard approach. Herein, we use a simplified, computationally tractable atomistic model to investigate the mechanical perturbations that can occur when electronically excited tryptophan residues in the BSA protein are simulated with classical molecular dynamics simulations in thermal equilibrium. We find that UV photoexcitation of tryptophan introduces large structural and dynamical changes in BSA involving regions far away from the two excited residues, suggesting the importance of a potential nonlinear response in the

system. Furthermore, we also observe preliminary signatures of nonlinear response in the water and surrounding ions. By computing the THz spectra from the ground- and photoexcited-state BSA simulations, we illustrate the possibility of exciting specific THz modes emerging from a collective multiscale response of the protein, ions, and water.

Computational methods

Classical molecular dynamics

All molecular dynamics simulations were performed using the GROMACS code (31). The initial crystal structure was obtained from the Protein Data Bank (PDB, with ID 3V03) (32). Following energy minimization, we subsequently performed NVT (isothermal-isochoric) and NPT (isothermal-isobaric) equilibrations each for 500 ps. For these runs, the temperature was set to 300 K and a pressure of 1 bar was used in the NPT ensembles. After these steps, the simulations were run within the NVT ensemble for approximately 0.5 μ s in either the ground or optically excited state, as will be explained in detail later. As will be shown later in the paper, we found that the initial equilibration of 1 ns was not sufficient. For both the ground- and excited-state simulations, we required an additional 100 ns in order for properties such as the root mean square deviation (RMSD) to equilibrate properly. For this reason, in all our ensuing analyses, the first 101 ns was not included.

The GROMOS96 54a7 (33) force field was used for the simulation together with the SPC/E (34) water model. Bond lengths were constrained using LINCS (35). The protonation states of the amino acids were chosen to be those relevant to neutral pH, which leads to a net negative charge of $-16e$ in the protein. The system is thus neutralized by adding 16 sodium ions. Additional sodium and chloride ions were added, leading to a buffer solution of concentration 0.15 M. Long-range electrostatics were calculated using the particle mesh Ewald (PME) method (36). The cutoff for both pairwise van der Waals (vdW) and short-range electrostatic interactions was set at 1.2 nm (37). Analytical corrections for pairwise interactions beyond the vdW cutoff, as implemented in GROMACS, were used to correct for the total energy and pressure of the system. The simulations were done in a cubic box with a side length of ~ 13.3 nm, with periodic boundary conditions, using a timestep of two femtoseconds. The total number of water molecules in the system was 75,414.

When computing dipole moments from the simulation, either for examining equilibrium properties such as the dipole distributions or in extracting time-correlation functions for the THz spectra, one has to carefully take into account the periodic boundary conditions of the simulation box, which can cause potential artifacts in the dipole estimation. For the dipole moment of the protein, this is simplified by ensuring that the trajectory is centered with the protein in the middle of the box. For computing the THz spectra, as shown below, we use the time-derivative of the dipole instead of the dipole moment itself, which essentially employs the velocities to create a charge-weighted correlation function that is always continuous. It can be shown that the infrared intensities extracted from the dipole and dipole-derivative differ only by a factor of ω^2 , where ω is the angular velocity.

Photoexcitation model

Simulations of BSA were performed both in the ground (Gr) and excited (Ex) state. The excited state of BSA was modeled by electronically exciting its two tryptophan (Trp) residues (with residue numbers 132 and 211). The two lowest-lying transition dipole

states for tryptophan are usually referred to as the L_a and L_b states, which demonstrate polarization along the short and long (orthogonal) axes in the $\pi - \pi^*$ orbital, respectively (see Fig. S1). L_a is generally the lower-lying fluorescent state in Trp in solvated proteins, and in water, there is ultrafast internal conversion (on the order of tens of femtoseconds) from L_b to L_a upon absorption, which is not modeled. We focus instead on the BSA protein relaxation and response due to both tryptophans being in the excited L_a state.

The ground-state charges are those of the GROMOS96 54a7 forcefield (33). The photoexcitation is modeled by modifying the point charges of the indole-group atoms in both Trp residues in the protein. The charge differences between the ground and excited states are based on multi-reference quantum chemistry (CASSCF/CASPT2) electronic structure calculations performed by Sobolewski and coworkers (38). This procedure has been applied in several previous studies in our group and has been found to give good agreement with experiments (28, 29).

Terahertz spectra analysis

To compute the terahertz spectra we employed a time-correlation formalism based on dipole fluctuations sampled in molecular dynamics simulations (39, 40). This entails extracting the Fourier transform of the dipole-dipole time-correlation function. The total dipole moment of the system is given by contributions coming from the protein ($\vec{M}_p(t)$), water ($\vec{M}_w(t)$), and ions ($\vec{M}_i(t)$). Specifically, the absorption coefficient per unit length is given by

$$\alpha(\nu) = F(\nu) \int_{-\infty}^{\infty} dt e^{-i2\pi\nu t} \left\langle \frac{d}{dt} \vec{M}(0) \cdot \frac{d}{dt} \vec{M}(t) \right\rangle, \quad (1)$$

$$\vec{M} = \vec{M}_p + \vec{M}_w + \vec{M}_i, \quad (2)$$

where $F(\nu) = (1/4\pi\epsilon_0)(2\pi/(k_B T))(1/(3Vcn(\nu)))$, with frequency ν , refractive index $n(\nu)$, volume V , and temperature T .

Note that the terahertz spectra include contributions coming from both autocorrelations and cross-correlations involving the protein, water, and ions. This will be discussed in more detail later in the manuscript. In order to compute the terahertz spectra, a total of 211 configurations were sampled with interval spacings of 2 ns between each initial condition, from the long equilibrated simulations in both ground and excited states as described earlier. From each of these initial configurations, microcanonical dynamics (holding particle number, volume, and energy constant) were performed for 100 ps. The terahertz spectra were then averaged over all these trajectories. This protocol has been previously applied with success to both protein and water systems (41).

Experimental evidence (12) of collective phononic modes in bio-systems driven far from thermal equilibrium is observed in sub-terahertz difference spectra between the photoexcited and ground states of BSA protein in aqueous solution. To ascertain the statistical relevance of the observed signals in molecular dynamics simulations, we assume the difference in the absorption coefficient of Equation (1) between the photoexcited state and the ground state follows a Gaussian distribution at each frequency.

In order to identify differences in the various terahertz spectral contributions from different modes upon photoexcitation, we quantify statistically significant changes as those exceeding two times the standard deviation,

$$\sigma(\nu) = \sqrt{\sigma_{\alpha_{Ex}}^2(\nu) + \sigma_{\alpha_{Gr}}^2(\nu)}, \quad (3)$$

where $\sigma_{\alpha_{Ex(Gr)}}(\nu)$ is the statistical error over the absorption coefficients α in the photoexcited (ground) state. Signals in $\Delta\alpha(\nu) = \alpha_{Ex}(\nu) - \alpha_{Gr}(\nu)$ have been considered statistically significant if $|\Delta\alpha(\nu)|$ exceeds $2\sigma(\nu)$, such that random fluctuations would not account for these outlier values in at least 95% of cases.

Centrality analysis

In order to characterize the dynamical changes arising from photoexcitation, we performed a mutual information (MI)-based eigenvector centrality and degree centrality analysis. These centrality metrics are widely employed to capture the fundamental properties of networks. Negre et al. have recently demonstrated the capability of the eigenvector centrality measure based on the mutual information to disentangle the complex interplay of amino acid interactions giving rise to allosteric signaling (42). This measure enables pinpointing of the key residues involved in the main collective motions of the system. To perform the centrality analysis we map the protein MD trajectory onto a graph composed of nodes interconnected by edges. Each node represents a single atom in the system, and the edges connecting pairs of these nodes are defined as the dynamical cross-correlations experienced by the corresponding atoms along an MD trajectory. The latter can be quantified by employing the generalized correlation coefficients, based on the MI between two atoms G_{ij} :

$$G_{ij} = \left[1 - \exp\left(-\frac{2}{3} I[\mathbf{x}_i, \mathbf{x}_j]\right) \right]^{\frac{1}{2}}, \quad (4)$$

where \mathbf{x}_i are the zero-mean atomic displacement vectors extracted from the MD simulations. The MI between the two atoms, $I[\mathbf{x}_i, \mathbf{x}_j]$, is defined in the following way:

$$I[\mathbf{x}_i, \mathbf{x}_j] = H_i + H_j - H_{ij}, \quad (5)$$

where

$$H_k = \int p(\mathbf{x}_k) \ln p(\mathbf{x}_k) d\mathbf{x}_k \quad (6)$$

$$H_{ij} = \iint p(\mathbf{x}_i, \mathbf{x}_j) \ln p(\mathbf{x}_i, \mathbf{x}_j) d\mathbf{x}_i d\mathbf{x}_j \quad (7)$$

are the marginal and joint Shannon entropies, respectively, extracted from MD simulations of the system at equilibrium. The generalized correlation coefficient G_{ij} ranges from zero to one for uncorrelated to fully correlated variables, respectively. We have recently shown that the compression of the generalized correlation matrix into a centrality vector enables the subtle characterization of the role played by every atom in the major collective modes of the system.

The eigenvector centrality metric E_c was obtained employing the linear approximation of the mutual information introduced by Lange et al. (43) The linear mutual information (LMI) relies on a Gaussian approximation of the probability density, which can be expressed as

$$p(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{(2\pi)^3 \det(\text{Cov})} \exp\left[-\frac{1}{2} (\mathbf{x}_i, \mathbf{x}_j) \text{Cov}^{-1} (\mathbf{x}_i, \mathbf{x}_j)^T\right], \quad (8)$$

where $\text{Cov} = \langle (\mathbf{x}_i, \mathbf{x}_j)^T \otimes (\mathbf{x}_i, \mathbf{x}_j) \rangle$ is the covariance matrix, and $\mathbf{x}_i, \mathbf{x}_j$, and $(\mathbf{x}_i, \mathbf{x}_j)$ are row vectors, with $(\mathbf{x}_i, \mathbf{x}_j)$ defined as a concatenation of vectors. Marginal probabilities are computed from the marginal covariance matrices $\text{Cov}_i = \langle \mathbf{x}_i^T \otimes \mathbf{x}_i \rangle$. Hence the entropies H are obtained analytically from the Gaussian density approximations as

$$H = \frac{1}{2} [q + \ln(\det(\text{Cov}))], \quad (9)$$

where $q = 2d(1 + \ln(2\pi)) = 17.03$, and $d = 3$ is the dimension of each individual coordinate vector. The LMI is then defined as

$$\text{LMI}_{ij} = \frac{1}{2} [(\ln(\det(\text{Cov}_i))) + (\ln(\det(\text{Cov}_j))) - (\ln(\det(\text{Cov})))] \quad (10)$$

As discussed above, the MI-based E_c distribution is then obtained from the leading eigenvector of the G matrix (Equation (4)), where $I[\mathbf{x}_i, \mathbf{x}_j]$ is approximated by LMI_{ij} , which by virtue of the Perron-Frobenius theorem is unique (i.e. non-degenerate):

$$E_c(i) = \frac{1}{\lambda} \sum_j G_{ij} E_c(j) \quad (11)$$

We would like to highlight that the covariant matrices Cov and Cov_i are based on time (and statistical) averages of correlations among the displacements \mathbf{x}_i . From this follows that the correlation matrices are based on the zero-frequency ($\omega = 0$) correlations in the frequency domain. It is possible, in principle, to construct a centrality score based on frequency-filtered correlations among atomic displacements $\text{Cov}(\omega) = \langle (\mathbf{x}_i, \mathbf{x}_j)^T \otimes (\mathbf{x}_i, \mathbf{x}_j) e^{i\omega t} \rangle$ and $\text{Cov}_i(\omega) = \langle \mathbf{x}_i^T \otimes \mathbf{x}_i e^{i\omega t} \rangle$. Such an analysis is beyond the scope of the present paper and will be developed in further works.

Results

Here we elucidate the mechanisms by which the photoexcitation of the BSA protein can alter the collective response involving various modes in the solvated protein system. In particular, we will begin by identifying specific changes in the sub-terahertz regime involving the protein, water, and ions. We will then subsequently illustrate structural and dynamical changes that involve these various contributions.

Origins of sub- and few-terahertz modes

The bottom panel of Fig. 1 shows the total absorption spectra of the system consisting of all the auto- and cross-correlations of the protein, ions, and water in the range between ~ 0 –30 THz. Since the specific focus of our work is to examine more deeply the region between ~ 0 –1 THz, we show a zoomed inset. Indeed, there is a significant contribution of intensity in this region as well.

As alluded to earlier, the terahertz signal can arise from a complex coupling of various modes in the system. There have been numerous theoretical studies in the literature illustrating how low-frequency terahertz spectra (see inset of Fig. 1 bottom panel) can arise from the interplay of protein, ion, and water fluctuations. Here we find that the photoexcitation of the two tryptophan chromophores in BSA gives rise to a multiscale response involving various different contributions.

In equilibrium measurements of THz spectra in the low-frequency regime, it is common to observe broad linewidth absorptions, rather than sharp peaks at specific frequencies. However, the BSA experiments reporting an absorption peak at ~ 0.3 THz are conducted in nonequilibrium settings. Since we cannot make a direct comparison at the moment in our simulations, we focus in the next section on reporting the differences in the bare intensities of the spectra between ground and excited states, without applying any smoothing procedures to the spectra. Significant peaks are seen at points throughout the low-frequency range, which contributes to the observed elastic mode of the whole BSA protein in an ionic solution. This elastic mode is constrained by the intrinsic vibrational properties (Young's modulus, hydrodynamic radius, specific volume) of the protein (12).

Fig. 2 shows the differences in the autocorrelation contributions to the terahertz absorption differences between excited and ground states, from various components of the protein solvated with ions and water as described earlier in the manuscript. In particular, the top row of panels in Fig. 2 displays the differences in the autocorrelations arising from the whole system, from water, and from ions. Though compared to the Trp132 and Trp211 autocorrelations there are far fewer data points across the entire low-frequency THz range that are significantly different from zero ($|\Delta a(\nu)| > 2\sigma(\nu)$), we observe that for the whole system, for the water, for the ions, and for the protein, the region between 0 and 2.35 THz is characterized by some rather large autocorrelation changes upon photoexcitation (significant at the 2σ -level, twice the usually reported error, at 95% or greater confidence interval). Particularly notable are the statistically significant autocorrelation changes in the sub-THz region for the protein, for the water, and for the whole system, suggesting pathways for sub-THz response of the whole system that has been observed experimentally under NEQ conditions.

We observe in the bottom row of panels in Fig. 2 that both excited tryptophan chromophores (Trp132 and Trp211) undergo clearly significant enhancements across the entire low-frequency THz range. This likely reflects subtle changes in the local environment of each tryptophan on the picosecond timescale consistent with previous studies (29). The changes in the THz spectra associated with the autocorrelations of the two tryptophans upon photoexcitation are over a factor of 10 larger than what one would expect from random fluctuations. Based on the current simulation data, if relevant to the NEQ observation, these large absorption differences are prime candidates for energy funneling to stimulate the BSA collective vibrational mode. Indeed, without a UV LED exciting these tryptophans in the NEQ experiment, no collective phononic mode is observed in the BSA solution (13).

In addition, as shown in Supplementary Material Fig. S2, there are statistically significant cross-correlation changes in the few-THz regime for the entire protein with ions and with water, and for both tryptophans with water. In recent years, there has been much interest in the relevance of low-frequency THz modes to ionic solutions (44). Our results suggest the possibility that photoexcitation or other local charge reorganization processes may alter global changes in the low-frequency THz modes of protein, ions, and water. It is also important to note that the differences in the THz modes we are quantifying in Fig. 2 and Fig. S2 range between 0 and 0.43 ps. The conformational ensembles we are using emerge from molecular dynamics simulations that are run for 500 ns and therefore, the statistical tests and large separation of timescales (500 ns compared to less than one picosecond) give us adequate confidence in our inferences.

Still, the physical relevance of the “oscillations” observed in the autocorrelation spectra—especially for Trp132-AC and Trp211-AC in Fig. 2—remains unclear at the moment. On the one hand, it may arise from the treatment of the linewidth analysis (focusing on the bare intensity differences at the frequency resolution applied), or on the other hand, it may be a signature of charge-phonon correlations that are beyond the scope of this paper. Further simulations and experiments are thus warranted to determine the sensitivity and importance of these thermodynamic averaging procedures in NEQ settings.

It is also essential to include that the THz spectra we compute from our simulations correspond to frequencies that are infrared-active. In other words, the y-axes in Fig. 2 and Fig. S2 are physically interpretable infrared intensities. Although one can compute also the vibrational density of states (VDOS) from the simulations, those active frequencies may not necessarily be

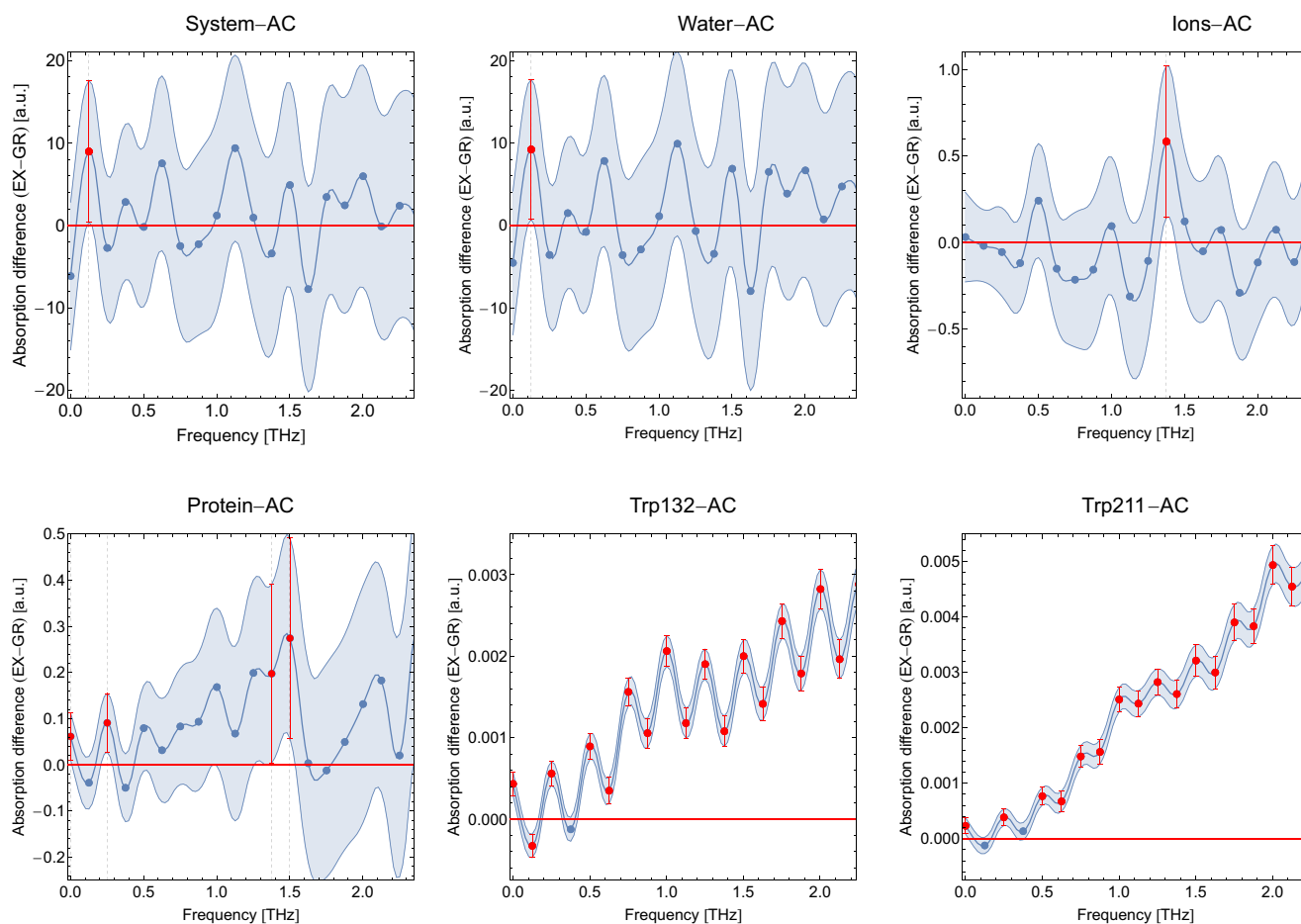


Fig. 2. Autocorrelations (AC) of electric dipole fluctuations from different inclusions of the system (top row: whole system, water, ions; bottom row: protein, tryptophan 132, tryptophan 211) contributing to the simulated terahertz absorption spectra in the range 0–2.35 THz (Fig. 1) upon photoexcitation of bovine serum albumin (BSA) protein, presented as differences in the absorption between photoexcited (Ex) and ground (Gr) state simulations ($\Delta\alpha(\nu) = \alpha_{\text{Ex}}(\nu) - \alpha_{\text{Gr}}(\nu)$). The shaded band above and below the blue data points is an interpolated second-order polynomial and has a width at each data point of four times the standard deviation, $\Delta\alpha(\nu) \pm 2\sigma(\nu)$. The black vertical dashed lines extending from the bold red error bars correspond to the frequencies where the absorption difference exceeds 2σ -statistical significance (i.e. when the mean value $\Delta\alpha$ is separated from zero (x-axis, in bold red) by more than 2σ). Vertical dashed lines have not been reported for the two tryptophan AC spectra for the sake of readability. The simulated terahertz absorption spectra from which these differences were derived have each been obtained by averaging over 211 NVE trajectories holding particle number, volume, and energy constant.

infrared-active. In the Supplementary Material Fig. S3 we show the VDOS obtained for various components of the system, where it is clear that the coupling of the charge geometry and mechanical distortions simultaneously is needed to understand the active THz modes and how they change upon photoexcitation.

These 2σ -significant values of different autocorrelations and cross-correlations from Fig. 2 and Fig. S2 are summarized in Table S1. Near the experimentally observed reference frequency ($\nu \approx 0.3$ THz), besides autocorrelations from protein, water, and the whole system, cross-correlations between Trp132 and water, and between Trp211 and ions, appear integral to mediating the multiscale response of the whole system to photoexcitation. All in all, the emerging picture from this analysis suggests that the few-THz signal involves a complex and multiscale response involving two photoexcited tryptophan amino acids, the entire BSA protein, ions, and water. Although the spectra are constructed from equilibrium simulations, the synopsis in Table S1 indicates the possibility of injecting energy into a wide variety of different modes involving water, ions, and protein, with the individual components and collective interactions supporting the maintenance of a coherent elastic vibration constrained by the

protein bulk mechanical properties. Since the few-THz modes involve low-frequency collective vibrations, in the remainder of the manuscript we will quantify the molecular reverberations associated with changes in the protein, ions, and solvent that emerge from UV photoexcitation of the two tryptophans in BSA.

Collective structural reorganization of BSA protein

The BSA protein simulated in this work has two tryptophans (Trp132 and Trp211), which are both substantially exposed to the solvent. The upper right panel of Fig. 1 shows a snapshot of the protein taken from our simulations, where the two tryptophans are highlighted with appropriate labels.

In order to examine how the photoexcitation perturbs the mechanical behaviors of the protein, we determined the root-mean-square deviation (RMSD), radius of gyration (R_G), and the total dipole moment (M) of the protein, in both the ground and excited state. The left column of panels in Fig. 3 shows the time-series associated with these three quantities, while the right column shows the corresponding probability distributions. We observe that for both the ground- and excited-state trajectories it takes about 100 ns to equilibrate the system from the initial

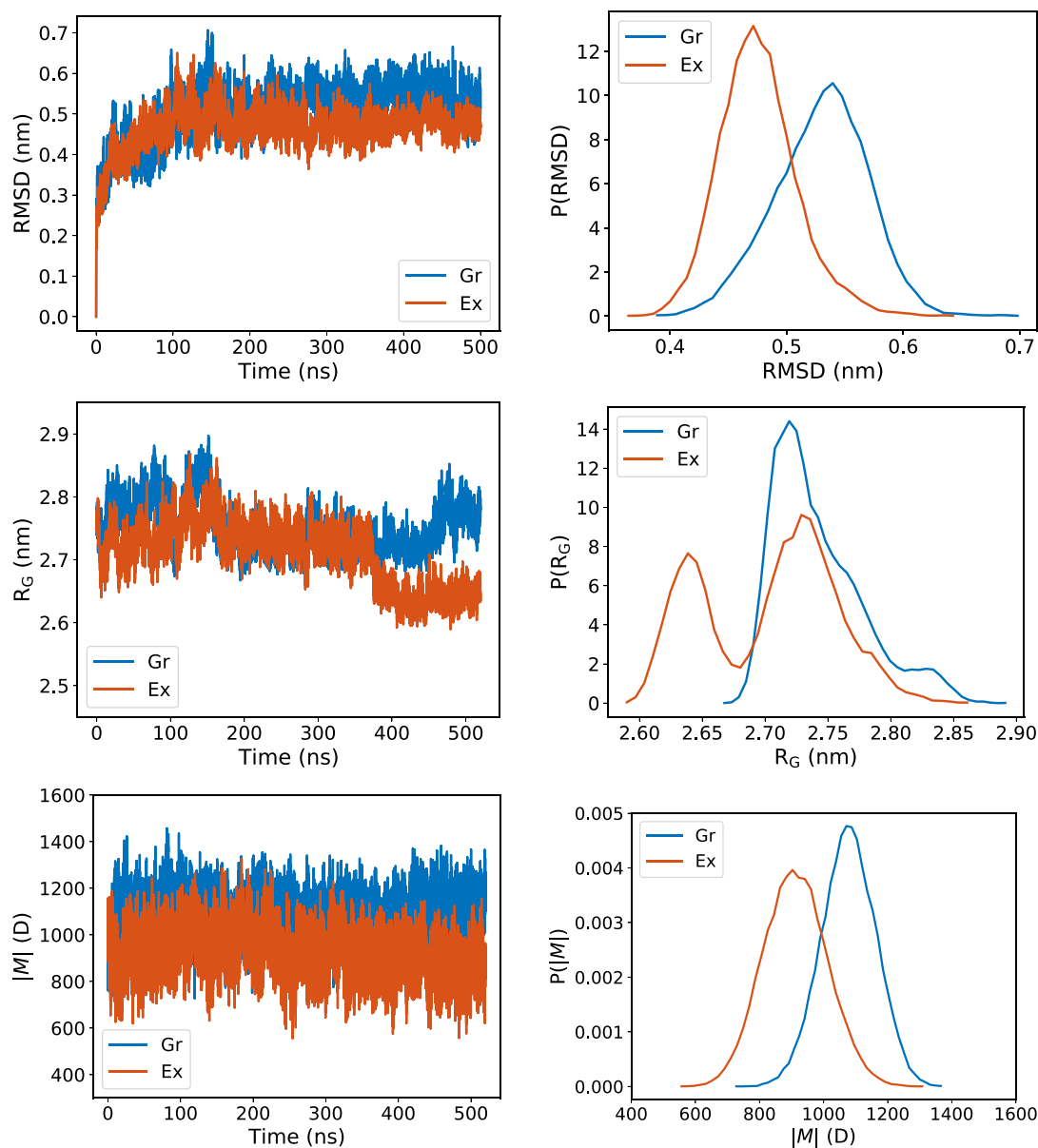


Fig. 3. Time evolution and distributions of different properties of BSA protein in aqueous solution: root-mean-square deviation (RMSD), radius of gyration (R_G), and the total dipole moment ($|M|$), in the ground (Gr) and excited (Ex) state. The distributions are shown for equilibrated time-series data after 101 ns. The equilibrium time is chosen based on the RMSD plot, which can also be confirmed by its distribution. The R_G plot shows a conformational change in the excited state at about 375 ns, which results in a bimodal behavior in the distribution and may be indicative of protein oscillations occurring on microsecond timescales. Total dipole moments of the protein—not including ionic contributions from the aqueous environment—are quite robust during the whole simulation time, with a lower mean value but larger distribution width for the excited state due to the larger dipole fluctuations caused by photoexcitation.

configuration. Thus in order to construct the histograms we do not include the first 100 ns.

Interestingly, we observe that the photoexcitation results in an overall rigidification of the protein as indicated by changes in the RMSD and R_G distributions. The time-series of the R_G shows that there are protein conformational changes occurring on a rather long time scale. In particular, we see a transition occurring at about 375 ns in the excited-state trajectory that results in a reduction of the R_G by approximately 1 Å. This change in the R_G leads to the bimodal structure of the probability distribution in the excited state. Furthermore, the distribution of the RMSD in the excited state is slightly more narrow than that in the ground state.

It is important to stress that the bi-modality observed along R_G in the excited state simulation arises from only a single

conformational change in our 0.5- μ s simulation. Our goal in this work is not to converge the distributions along the order parameters such as R_G , as this would take significantly more sampling, but we note that the THz spectra calculated from trajectories sampled equivalently in the periods before and after the conformational change remain essentially the same. We recall at this juncture a restatement of Levinthal's paradox: If we assume each amino acid of BSA can take one of two states (a gross underestimate), then we have a conformational space of at least $2^{500} \approx 10^{150}$ states, about 70 orders of magnitude larger than the number of atoms in the observable universe. If switching between the two amino acid states occurs on the order of 1 ps, then our 0.5- μ s simulation can at best sample only about 10^6 of these configurations, which is hopelessly inadequate to sample the full conformational

space. In fact, we would need at least 10^{140} such simulations, requiring far beyond 10^{100} times the age of the universe, to even approach the entire conformational space. Therefore, the best we can do is give error estimates for a given number of trajectories, determining the statistical significance of our results from there. We have thus calculated values for the statistical significance of our results in Equation (3) from 211 short trajectories (100 ps) each run independently, with 211 initial conformations sampled from throughout the equilibrated portion of the 0.5- μ s MD simulation. Indeed, we acknowledge this is but a small sample of the entire protein conformational space, but the 211 trajectories are sufficient for 95% confidence (statistically significant at the 2σ -level) that our autocorrelation results reported in Fig. 2 cannot be accounted for by random fluctuations.

The bottom row of panels in Fig. 3 shows the time-series and corresponding distributions of the total dipole moment for the protein itself. Note that for a charged system, as in our case, the dipole moment $\vec{M} = \sum_i q_i \vec{r}_i$ is not translationally invariant. It is possible, however, to define the dipole relative to the geometric centroid, $\vec{M} = \sum_i (q_i - Q/N) \vec{r}_i = \sum_i q_i \vec{r}_i - Q(\sum_i \vec{r}_i)/N$, where $Q = \sum_i q_i$ is the total charge of the system of N atoms. This quantity is *invariant* to net charge and net position shifts, as explained in Ref. (45). We found that the dipole moment distributions computed using the centroid versus the center of mass are essentially the same (see Supplementary Material Fig. S4). Consistent with the findings of the RMSD and R_C , the UV photoexcitation of BSA results in a significant reduction of the dipole moment of the protein (by about 20%). It is interesting to note that the changes in R_C are not manifested in $|M|$, which in the excited state retains a unimodal character. It is clear, however, that these types of atomistic details would not be captured by heuristic models and therefore are complemented by experiments (23) examining the source of long-range dipole interactions as electrodynamic fluctuations of proteins dressed by their local ionic environment.

The changes in the RMSD, R_C , and $|M|$ upon photoexcitation suggest that there is a rather significant change in the structure of the protein. However, these quantities only probe the global structure of the protein. In order to provide a more refined view of the molecular reorganization and collective changes in BSA upon photoexcitation, we determined the difference in the so-called distance contact map between all pairs of amino acids in the protein. The contact map quantifies the local environment around each amino acid in terms of the neighbors in close proximity. Specifically, we sampled frames from the molecular dynamics time-series every 10 ps after equilibration for 420 ns (42,000 frames) and determined the fraction of frames in which amino acids i and j , for all amino acid pairs, are separated by less than 5 Å. This is done for both the excited- and ground-state simulations, after which the difference between the two fractions is taken yielding the difference contact map. The choice of this cutoff length has been shown to provide a good balance of probing short-to-medium range correlations in protein structure in previous studies (42, 46, 47).

Fig. 4a illustrates the difference contact map between excited- and ground-state simulations. As indicated above, the first 100 ns of equilibration were not included in this analysis. For clarity in the left panel, the dashed horizontal (vertical) lines colored in black indicate the amino acid indices of the two tryptophans. The presence of a green spot along one of these lines indicates those amino acids that move further away from the tryptophan on average over the sampled frames in the excited state (fewer contacts) relative to those in the ground state, while a purple spot corresponds to amino acids that move closer to the tryptophan on average over the sampled frames in the excited state

(more contacts) relative to those in the ground state. It is worth noting that we observe changes in the rest of the protein as observed by the presence of green and purple spots across many amino acids in the system.

In order to quantify the net change in the total number of contacts, we sum over all the changes for each amino acid. To visualize these changes, in Fig. 4b, amino acids undergoing an increase in the number of contacts are colored in red while those that decrease in the number of contacts are shown in blue. For the purposes of visualization, we focus on only those contacts changing by more than 70%. Interestingly, we observe that these regions are spread all over the protein. The distance of the tryptophans to these regions of the protein ranges from between one to over six nanometers and shows that there is a rather long-range structural reorganization upon photoexcitation.

The contact map analysis presented above provides insights into the difference in fluctuations of the BSA configuration between photoexcited and ground states, but the contact map analysis introduces an arbitrary length scale to define the contact frequency. To obtain complementary information on the global BSA configuration altered by the tryptophan photoexcitation, we examine the difference $\Delta d(IJ) = d_{\text{Ex}}(IJ) - d_{\text{Gr}}(IJ)$ of the residue distance matrices between two configurations representative of the photoexcited and ground state. The distance $d_{\text{Ex(Gr)}}(IJ)$ between residues have been defined as the pairwise distance between the geometric barycenters of each residue in a configuration averaged over the equilibrated NVT time-series data presented in Fig. 3, after the 100-ns equilibration for the excited (ground) state. Such a choice seems natural to encode in a single configuration, a representative whole of information on the thermodynamic equilibrium properties both in the ground and in the excited state, at the price of neglecting higher-order moments of the distribution and, as a result, losing information on dynamical properties.

The results have been reported in Fig. 4c, which affirms how the photoexcitation of the tryptophans deeply affects the global structure of the protein, with a rearrangement of mutual residue distances even for residue pairs not involving the tryptophans themselves. Moreover, it is interesting to notice that darkened regions, corresponding to the most important conformational changes, form a “checkerboard” pattern parallel to the dashed Trp132 and Trp211 lines. Such darkened patterns correspond to flexible and mobile secondary-structure motifs (coils and turns), and their presence suggests the propagation of a structural rearrangement effect from the tryptophan centers to the rest of the protein. Fig. 4d shows a possible encoding of the information contained in the difference matrix of Fig. 4d into a residue-dependent scalar field $\phi(I) = \sum_j \Delta d(IJ)$ over the entire protein structure, analogous to Fig. 4b. The red spots show the residues for which $\phi(I)$ is large and positive: we observe how the photoexcitation of tryptophans leads to a configuration rearrangement of residue barycenters in sites both near and also far from the tryptophans.

In order to quantify this behavior more precisely, a correlation plot has been reported in Fig. 4e between the magnitude of the difference in mutual residue distances $|\Delta d(IJ)|$ and the distance of the residue in each pair closest to a tryptophan in the ground state defined as

$$d_{\text{Trp}}(IJ) = \min_{K \in \{\text{Trp131}, \text{Trp211}\}} (d_{\text{Gr}}(IK), d_{\text{Gr}}(JK)). \quad (12)$$

With such a definition, the residue pairs with $d_{\text{Trp}}(IJ) = 0$ are the ones containing at least one tryptophan. The residue pairs have been divided into three groups: the pairs whose residues both

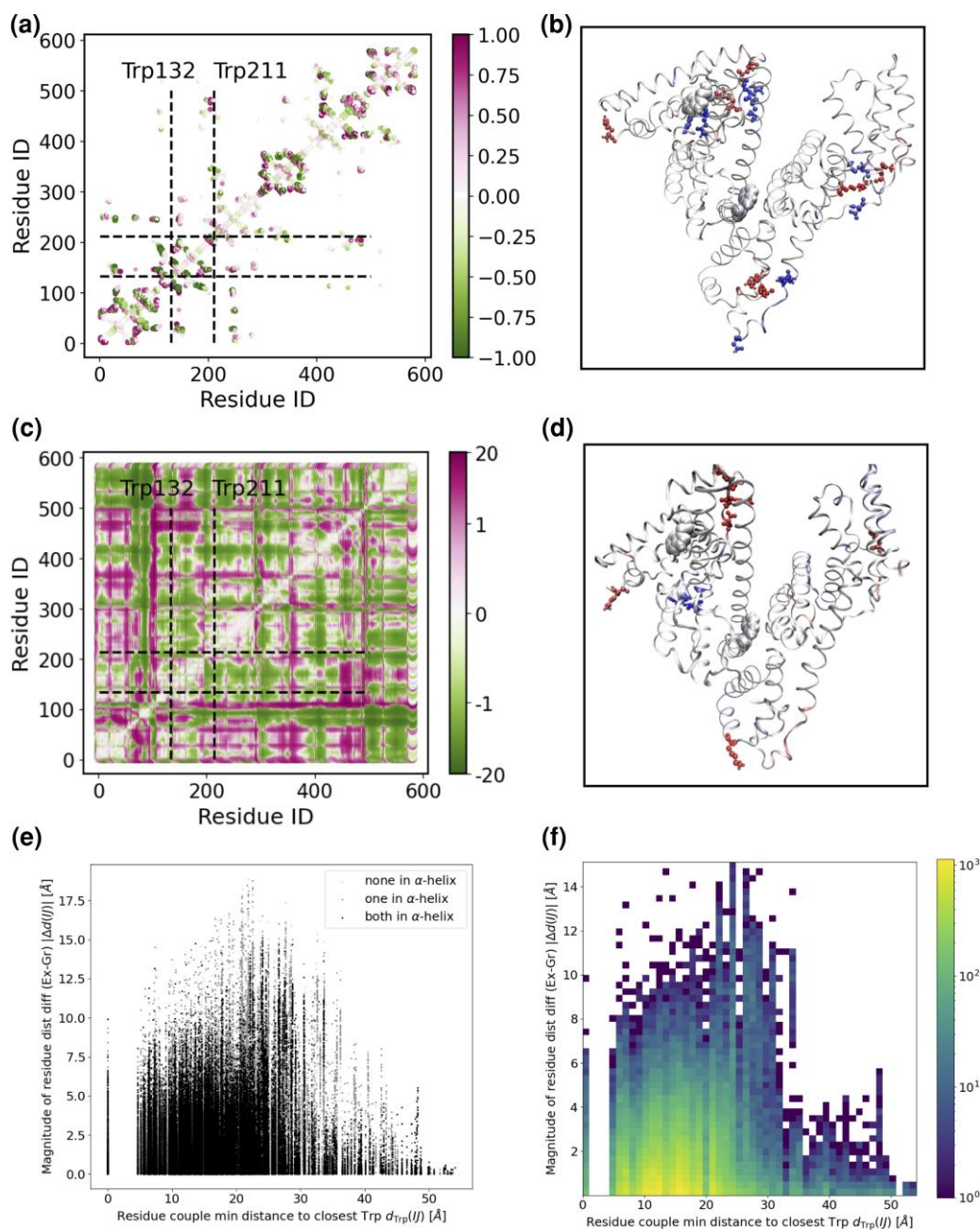


Fig. 4. a,b) The left panel shows the difference in the contact map between the excited- and ground-state simulations of the BSA aqueous solution. The presence of green and purple spots located across many residues, and not only along the vertical black lines (corresponding to the two tryptophans), indicates there are changes in the local geometries of amino acids throughout the protein. The right panel visually depicts this global dynamic restructuring. c,d) Variation of BSA average conformations between excited and ground states, giving a complementary static picture to the dynamic behaviors shown in the upper panels. e,f) The left panel displays a correlogram between the minimum distance $d_{\text{Trp}}(IJ)$ of a residue pair to the closest tryptophan and the magnitude of residue distance difference $|\Delta d(IJ)|$ between representative configurations for the excited and ground states of the BSA solution. The right panel displays a two-dimensional histogram representing in color scale the number of unique residue pairs with a given $|\Delta d(IJ)|$ and $d_{\text{Trp}}(IJ)$, whose residues both belong to a BSA α -helix. See the main text for further details.

belong to an α -helix (black points), the pairs where only one residue belongs to an α -helix (dark gray points), and the pairs whose residues do not belong to any α -helix (light gray points). Such a classification has been introduced because the α -helices are generally more rigid than other secondary-structure motifs (48, 49), especially more so than the flexible coil and linker segments in BSA. Still, the correlogram analysis reveals quite high values for the magnitude of residue distance difference ($|\Delta d(IJ)| > 10$ Å) for some pairs belonging to α -helices. This can be interpreted as a

manifestation of the effects of the photoexcitation on the BSA global configuration, including more rigid structures such as the α -helices. In fact, we can see from Fig. 4f that a significant fraction ($\sim 4\%$) of all pairs whose residues both belong to an α -helix ($N = 409$, out of 581 BSA residues, resulting in $N(N-1)/2 = 83,436$ unique α -helical residue pairs) and having $d_{\text{Trp}}(IJ) \leq 20$ Å show a magnitude of residue distance difference $|\Delta d(IJ)| > 4$ Å (the smallest length of an amino acid (50)) due to the photoexcitation of the tryptophans.

Numerous previous studies have shown, using similar models, that the optical excitation of tryptophan can produce a wide variety of protein structural changes, such as the movement of positively charged lysine groups due to the change in the strength of pi-cation interactions (28) as well as perturbations to the protein secondary structure (29). The preceding analysis in Fig. 4 provides a simpler way to pinpoint these structural perturbations. Moreover, these collective structural changes are likely to be important for tuning the THz modes discussed earlier.

In order to understand better how the fluctuations of each amino acid in the protein respond to optical excitations, we determined the difference in the root-mean-square fluctuations (RMSF) between the ground- and excited-state simulations (see Fig. S5). Interestingly, the response of the protein is characterized by some amino acids incurring an enhancement in their fluctuations, while others become more restricted. The right panel in Fig. S5 provides a visual depiction of the changes in RMSF. We see clearly that, contrary to some expectations, the amino acids undergoing large changes in RMSF do not necessarily involve those that are proximal to the optically excited chromophores.

Photoinduced reorganization of ions and water

Besides the collective structural changes in the protein, the terahertz spectra also suggest that there are important contributions coming from the reorganization of both ions and water molecules. Several previous studies have shown with time-dependent fluorescence Stokes shift experiments that the optical response involves protein, ionic, and aqueous contributions (7, 30, 51, 52). In addition, there have been several experimental measurements showing the importance of low-frequency THz modes that are sensitive to fluctuations of ion-related modes (53–55).

In Fig. S6 we show the pair correlation functions between the two tryptophan residues and the sodium/chloride ions for both ground and excited states. The left panel shows the distribution functions for Trp132, and the right panel is for Trp 211. For Trp 132, there appear to be subtle changes occurring in the ion density upon photoexcitation, which occur over several nanometers around the excited chromophore. In the case of Trp211, the changes appear to be much more drastic, especially for the chloride ions within one nanometer. Visual examination of the environment of Trp211 shows that there is a cluster of positively charged arginine residues in close proximity, which attract more anions. Upon photoexcitation, the structural reorganization that occurs throughout the protein reduces the accessibility of these positively charged residues to the chloride ions.

The perturbations in the ion densities around the two tryptophans prompted us to examine whether there are any changes in the solvent exposure of the tryptophans. Toward this end, we determined the solvent-accessible surface area (SASA) of all the amino acids following photoexcitation. Specifically, the distribution of the change in the SASA is shown in the left panel of Fig. S7. While it is peaked around zero, there are amino acids whose SASA increases or decreases upon photoexcitation, particularly those at the solvent boundary. A visual illustration of the change in the SASA is shown in the right panel of Fig. S7, which shows that there are regions of the protein close to and far away from the two excited tryptophans that undergo changes in their solvent exposure.

It is important to note that if ion distributions change significantly near the protein, as suggested in Fig. S6, there could be substantial contributions to the effective dipole moment arising from this recruitment. Indeed, notwithstanding the probability distributions for $|M|$ in Fig. 3 that only consider protein contributions

without surrounding ions, further investigations are warranted to verify if the recruitment of ions around the rigidified protein in the excited state could be the source of a “giant dipole” enhancement upon photoexcitation.

Photoinduced dynamic correlations via eigenvector centrality analysis

In the last decade, there has been significant development in the confluence of ideas from both network and information theory in order to understand the nature of fluctuations in biological systems (42, 43, 46, 47, 56, 57). Within the context of protein dynamics, the eigenvector centrality provides a manner to quantify how dynamically correlated an amino acid is with other components of its networked environment. Batista and coworkers recently employed the centrality measure to examine the dynamical changes that occur during effector-enzyme binding (42). Details of how the eigenvector centrality is computed can be found in the Methods section.

The upper panel of Fig. S8 presents a visual depiction of the change in the centrality of BSA upon tryptophan photoexcitation. Specifically, we employ the centrality defined in Equation (11), where the E_c vector has a unit norm. Areas colored in red correspond to regions of the protein that become more correlated with the major collective fluctuation modes of BSA, while those in blue are regions that lose correlation with the principal fluctuation modes of the protein. Interestingly, and in agreement with the structural behavior, the dynamical changes highlighted by this analysis are not limited to the proximity of the excited tryptophans but rather delocalized, indicating that the global dynamics of BSA is modified upon tryptophan photoexcitation. Fig. S8 shows that the main photoexcitation-induced dynamical perturbation of BSA occurs at the nexus of its three globular domains. This suggests that tryptophan photoexcitation will strongly impact the inter-domain vibrational modes of the protein, and the changes in centrality around this nexus would be highly correlated to spheroidal breathing motions of the whole protein that have been predicted (14) and observed (12, 25) at ~ 0.3 THz. Further analysis tracking which residues specifically contribute to such vibrational modes—projecting dynamical correlations into that basis—would be the subject of future work.

Discussion and conclusions

To derive the Fröhlich rate equations from simple idealized models of quantum (or classical) harmonic oscillators, many assumptions are considered that often do not hold in real experiments. For instance, it is usually assumed that the characteristic timescale of the dynamics of the occupation number of the collective modes of the protein is much larger than the inverse of their frequencies, so that they are completely decoupled. However, it has been found that the experimentally measured quality (Q) factor of the absorption peak associated with the nonequilibrium (NEQ) oscillations is about 50–90 (58). This means that the coherence time of such activated collective modes is estimated to be $\tau = Q/(\pi\nu) \approx 16$ – 30 ps (23). Comparing the full width at half-maximum of the absorption peak observed in photoexcited BSA to the unexcited one, we can estimate $Q \approx 30$. This means that the characteristic timescale of the observed coherent oscillations is very close to the timescale of the characteristic BSA breathing mode that is conjectured to be responsible for the observed signal.

Our THz and structural/dynamical analyses presented in this paper have illuminated some of the main features and “levers”

of the multiscale optomechanical transduction process in an atomistic model of solvated BSA. We have shown that in the critical sub-THz regime where the NEQ observation at ~ 0.3 THz was made, a variety of statistically significant modes arise dynamically from auto- and cross-correlations of interacting system components (tryptophan residues, entire protein, ions, water, whole system), upon optical excitation of the two BSA tryptophan residues during equilibrium MD simulation. The autocorrelation frequency of the entire protein, and the cross-correlations for Trp211-ions and for Trp132-water, are the closest in frequency to the experimental value, suggesting that a truly collective and strongly coupled interaction between protein molecules and their ionic solvent produces the observed difference in THz absorption between excited and ground states.

To explain the structural reverberations emerging from the optical excitation, we have shown the global restructuring of residue contacts throughout the protein, even more than six nanometers away from the photoexcited tryptophans. Commensurately, we observe stark changes in the radius of gyration of the BSA protein, which exhibits a pronounced conformational change in the excited state near the end of the $0.5\text{-}\mu\text{s}$ simulation, suggesting that protein oscillations might manifest on longer timescales. We have also demonstrated that significant reorganizations take place in the ions and water around the tryptophan residues and at the protein-solvent boundary, by tracking radial distribution functions for sodium and chloride ions and solvent-accessible surface areas for all the amino acids. Such rearrangements could have a considerable impact on the total effective dipole moment of the protein in an ionic solvent. Taken together, these data suggest drastic deviations in our simulated system from the standard expectations of linear response theory.

This work is intended to be the first step toward investigating the origins of the experimentally observed spectroscopic THz features using atomistic simulations. Atomistic simulations of electromagnetic radiation at different frequencies interacting with biological materials, in our case a solution of BSA protein, present many challenges, including realization of the precise binding sites and orientations for the Alexa488 fluorophores, the details of vibrational relaxation, and downconversion of energy from electronically excited Alexa488 and tryptophan to the protein vibrational and solvent degrees of freedom. Here we specifically focused on the problem of how photoexcitation of the two tryptophans in BSA protein can result in a collective response to the environment. The aim is twofold: From one side we have investigated how photoexcitation of tryptophans of native BSA (not marked with Alexa488) in an ionic solution affects the mechanical and electronic response properties of the protein and its environment, and on the other side we have studied how local changes in the biomolecule could have global effects on the atomistic dynamics, for instance in the generation of new collective normal modes and the centrality evolution of components in the network.

It should be noted that all aromatic residues, and certain other complexes in proteins, absorb in the UV range, which can affect biomolecular photophysics and resulting downstream responses. Tyrosine and phenylalanine, which generally have smaller transition dipole moments than tryptophan in the UV, frequently show up in proteins, including BSA, as a larger proportion of the total number of amino acids. There is some additional evidence that a visible fluorophore in close proximity to a lysine residue (namely, the experimental scenario with Alexa488 attached to BSA) absorbs significantly in the UV upon Schiff-base deprotonation (59, 60).

Despite this progress in our understanding of the emergence of collective optomechanical vibrational phenomena, many aspects remain to be clarified in open systems pumped far from thermal equilibrium. Our current molecular dynamics simulations of photoexcited BSA solution are done at equilibrium. Indeed, there have been several experimental and theoretical studies that have examined the role of NEQ effects within the context of deviations from linear response theory. This nonlinear response—and its role in light-driven, mechanical-structural phase transitions—originates from the strong coupling of electronic and vibrational degrees of freedom. This effect has also been described for Fröhlich polarons (61), where the impulsive movement of an electron in the highly nonlinear regime induces coherent phonons that persist for several hundred femtoseconds.

Such nonlinearities in protein vibrational dynamics have also been well described by Davydov and others in the theory of solitonic propagation along α -helices (62). This theory poses significant challenges for MD simulations of nonlinear dissipative systems (63) that are ubiquitous in biology, and which reflect the flow of energy under NEQ conditions. Indeed, the overdamping of sidechain rotovibrational modes in fully equilibrated MD simulations is related to the suppression of transient states that may mask the emergence of pronounced conformational changes occurring on much longer timescales. Our results suggest that these protein conformational changes in MD simulations can occur, in spite of overdamping, after several hundred nanoseconds (see Fig. 3, radius of gyration), and that residues in α -helices are significantly affected (see Fig. 4, bottom correlograms). These changes would likely take on greater prominence should additional nonlinear effects be included. Short of this, we interpolate in this work between strictly equilibrium simulations and NEQ conditions by the inclusion of electronically excited tryptophan residues in our thermodynamically equilibrated BSA simulations. In this way, our results motivate future NEQ simulations with nonlinear inclusions.

The breakdown of linear response theory may be closely related to the deviations observed from ergodic behavior in systems of biological relevance. UV fluorescence up-conversion experiments combined with NEQ MD simulations have observed deviations from the linear response approximation for the relaxation dynamics of photoexcited tryptophan in water (27). It has also been demonstrated that metabolic activities drive the biological milieu toward non-ergodicity far from thermodynamic equilibrium, resulting in increased cytoplasmic fluidization that allows larger components to escape their local environment anomalously and explore larger regions of the cellular compartment (64). The limitations of equilibrium MD simulations allude to the need for interfacing quantum excited states in biomolecules with an atomistic bath that includes a nonlinear dissipative term describing the flow of energy between the two. Indeed, pioneering work (24, 65, 66) has demonstrated the role of such thermal conduits in the kinetic activation of catalytically efficient states in enzyme dynamics. These perspectives and our results will motivate future, more realistic NEQ MD simulations and additional experiments to characterize the extent, limits of control, and applications of Fröhlich-like phenomena and optomechanical transduction in biology.

Acknowledgments

This research used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility supported under contract DE-AC02-06CH11357. This research also used

resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725. We thank the Howard University RCMI program for supporting publication charges to make this article open access. Finally, we would like to acknowledge conversations with Matthias Heyden and Marco Pettini in advancing this work.

Supplementary material

Supplementary material is available at PNAS Nexus online.

Funding

This work has been supported in part by the Defense Advanced Research Projects Agency.

Author contributions

M.G., A.H., and P.K. designed the research. All authors performed the research. K.A. performed the MD simulations. K.A., M.G., and U.M. analyzed the data. All authors contributed to write the paper.

Preprints

A preprint of this article is published at arXiv:2210.14912.

Data availability

All data is included in the manuscript and/or supporting information.

References

- He Y, Chen J-Y, Knab JR, Zheng W, Markelz AG. 2011 Feb. Evidence of protein collective motions on the picosecond time-scale. *Biophys J*. 100(4):1058–1065.
- Nibali VC, Havenith M. 2014 Sep. New insights into the role of water in biological function: studying solvated biomolecules using terahertz absorption spectroscopy in conjunction with molecular dynamics simulations. *J Am Chem Soc*. 136(37):12800–12807.
- Heyden M, et al. 2010 Jul. Dissecting the THz spectrum of liquid water from first principles via correlations in time and space. *Proc Natl Acad Sci USA*. 107(27):12068–12073.
- Falconer RJ, Markelz AG. 2012 Oct. Terahertz spectroscopic analysis of peptides and proteins. *J Infrared Millim Terahertz Waves*. 33(10):973–988.
- Bye JW, et al. 2014. Analysis of the hydration water around bovine serum albumin using terahertz coherent synchrotron radiation. *J Phys Chem A*. 118(1):83–88.
- Sushko O, Dubrovka R, Donnan RS. 2015. Sub-terahertz spectroscopy reveals that proteins influence the properties of water at greater distances than previously detected. *J Chem Phys*. 142(5):055101.
- Bellissent-Funel M-C, et al. 2016 Jul. Water determines the structure and dynamics of proteins. *Chem Rev*. 116(13):7673–7697.
- Ebbinghaus S, et al. 2007 Dec. An extended dynamical hydration shell around proteins. *Proc Natl Acad Sci USA*. 104(52):20749–20752.
- Born B, Kim SJ, Ebbinghaus S, Gruebele M, Havenith M. 2008 Nov. The terahertz dance of water with the proteins: the effect of protein flexibility on the dynamical hydration shell of ubiquitin. *Faraday Discuss*. 141:161–173.
- Born B, Kim SJ, Ebbinghaus S, Gruebele M, Havenith M. 2009. The terahertz dance of water with the proteins: the effect of protein flexibility on the dynamical hydration shell of ubiquitin. *Faraday Discuss*. 141:161–173.
- Yu B, et al. 2004 Mar. Torsional vibrational modes of tryptophan studied by terahertz time-domain spectroscopy. *Biophys J*. 86(3):1649–1654.
- Nardecchia I, et al. 2018. Out-of-equilibrium collective oscillation as phonon condensation in a model protein. *Phys Rev X*. 8(3):031061.
- Pettini M. 2022. Private communication.
- Bastrukov SI. 1994 Apr. Low-frequency elastic response of a spherical particle. *Phys Rev E*. 49(4):3166–3170.
- Fröhlich H. 1968. Long-range coherence and energy storage in biological systems. *Int J Quantum Chem*. 2(5):641–649.
- Wu TM, Austin S. 1978. Bose-Einstein condensation in biological systems. *J Theor Biol*. 71(2):209–214.
- Pokorný J, Wu T-M. 1998. *Biophysical aspects of coherence and biological order*. Prague: Springer.
- Pokorný J, Fiala J. 1992. Heat bath coupling effects in coherent vibration systems. *EPL*. 19(8):729–734.
- Turcu I. 1997. A generic model for the Fröhlich rate equations. *Phys Lett A*. 234(3):181–186.
- Preto J. 2017. Semi-classical statistical description of Fröhlich condensation. *J Biol Phys*. 43(2):167–184.
- Zhang Z, Agarwal GS, Scully MO. 2019. Quantum fluctuations in the Fröhlich condensate of molecular vibrations driven far from equilibrium. *Phys Rev Lett*. 122(15):158101.
- Preto J, Pettini M, Tuszynski JA. 2015. Possible role of electrodynamic interactions in long-distance biomolecular recognition. *Phys Rev E*. 91(5):052710.
- Lechelon M, et al. 2022. Experimental evidence for long-distance electrodynamic intermolecular forces. *Sci Adv*. 8(7):eabl5855.
- Zaragoza JPT. 2019 Oct. Detecting and characterizing the kinetic activation of thermal networks in proteins: thermal transfer from a distal, solvent-exposed loop to the active site in soybean lipoxigenase. *J Phys Chem B*. 123(41):8662–8674.
- Cheatum CM. 2020 Apr. Low-frequency protein motions coupled to catalytic sites. *Annu Rev Phys Chem*. 71(1):267–288.
- Faraji E, Franzosi R, Mancini S, Pettini M. 2021 Dec. Energy transfer to the phonons of a macromolecule through light pumping. *Sci Rep*. 11(1):6591.
- Bram O, et al. 2010 Sep. Relaxation dynamics of tryptophan in water: a UV fluorescence up-conversion and molecular dynamics study. *J Phys Chem A*. 114(34):9034–9042.
- Hassanali AA, Li T, Zhong D, Singer SJ. 2006. A molecular dynamics study of Lys-Trp-Lys: structure and dynamics in solution following photoexcitation. *J Phys Chem B*. 110(21):10497–10508.
- Li T, Hassanali AA, Kao Y-T, Zhong D, Singer SJ. 2007. Hydration dynamics and time scales of coupled water-protein fluctuations. *J Am Chem Soc*. 129(11):3376–3382.
- Furse KE, Corcelli SA. 2008. The dynamics of water at DNA interfaces: computational studies of Hoechst 33258 bound to DNA. *Journal of the American Chemical Society* *J Am Chem Soc*. 130(39):13103–13109.
- Abraham MJ, et al. 2015. Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 1–2:19–25.
- Majorek KA, et al. 2012. Structural and immunologic characterization of bovine, horse, and rabbit serum albumins. *Mol Immunol*. 52(3-4):174–182.
- Schmid N, et al. 2011 Apr. Definition and testing of the gromos force-field versions 54a7 and 54b7. *Eur Biophys J*. 40(7):843–856.

- 34 Chatterjee S, Debenedetti PG, Stillinger FH, Lynden-Bell RM. 2008. A computational investigation of thermodynamics, structure, dynamics and solvation behavior in modified water models. *J Chem Phys.* 128(12):124511.
- 35 Hess B, Bekker H, Berendsen HJC, Fraaije JGEM. 1997. LINCS: a linear constraint solver for molecular simulations. *J Comput Chem.* 18(12):1463–1472.
- 36 Darden T, York D, Pedersen L. 1993. Particle mesh Ewald: an $N \cdot \log(n)$ method for Ewald sums in large systems. *J Chem Phys.* 98(12):10089–10092.
- 37 Piana S, et al. 2012. Evaluating the effects of cutoffs and treatment of long-range electrostatics in protein folding simulations. *PLoS One.* 7(6):e39918–e39918.
- 38 Sobolewski AL, Domcke W. 1999. Ab initio investigations on the photophysics of indole. *Chem Phys Lett.* 315(3–4):293–298.
- 39 McQuarrie DA. 1975. *Statistical mechanics.* New York: Harper & Row.
- 40 Bornhauser P, Bougeard D. 2001. Intensities of the vibrational spectra of siliceous zeolites by molecular dynamics calculations. I. Infrared spectra. *J Phys Chem B.* 105(1):36–41.
- 41 Heyden M, Tobias DJ. 2013. Spatial dependence of protein-water collective hydrogen-bond dynamics. *Phys Rev Lett.* 111(21):218101.
- 42 Negre CFA, 2018. Eigenvector centrality for characterization of protein allosteric pathways. *Proc Natl Acad Sci USA.* 115(52):E12201–E12208.
- 43 Lange OF, Grubmüller H. 2006. Generalized correlation for biomolecular dynamics. *Proteins.* 62(4):1053–1061.
- 44 Funkner S, et al. 2012. Watching the low-frequency motions in aqueous salt solutions: the terahertz vibrational signatures of hydrated ions. *J Am Chem Soc.* 134(2):1030–1035.
- 45 Veit M, Wilkins DM, Yang Y, DiStasio RA, Ceriotti M. 2020. Predicting molecular dipole moments by combining atomic partial charges and atomic dipoles. *J Chem Phys.* 153(2):024113.
- 46 Rivalta I, et al. 2012. Allosteric pathways in imidazole glycerol phosphate synthase. *Proc Natl Acad Sci USA.* 109(22):E1428–E1436.
- 47 East KW, et al. 2020. Allosteric motions of the CRISPR–Cas9 HNH nuclease probed by NMR and molecular dynamics. *J Am Chem Soc.* 142(3):1348–1358. PMID: 31885264
- 48 Emberly EG, Wingreen NS, Tang C. 2002 Aug. Designability of α -helical proteins. *Proc Natl Acad Sci USA.* 99(17):11163–11168.
- 49 Emberly EG, Mukhopadhyay R, Wingreen NS, Tang C. 2003 Mar. Flexibility of α -helices: results of a statistical analysis of database protein structures. *J Mol Biol.* 327(1):229–237.
- 50 Ching CB, Hidajat K, Uddin MS. 1989. Evaluation of equilibrium and kinetic parameters of smaller molecular size amino acids on KX zeolite crystals via liquid chromatographic techniques. *Sep Sci Technol.* 24(7–8):581–597.
- 51 Grisanti L, Pinotsi D, Gebauer R, Schierle GSK, Hassanali A. 2017. A computational study on how structure influences the optical properties in model crystal structures of amyloid fibrils. *Phys Chem Chem Phys.* 19(5):4030–4040.
- 52 Muiño PL, Callis PR. 1994. Hybrid simulations of solvation effects on electronic spectra: indoles in water. *J Chem Phys.* 100(6):4093–4109.
- 53 Schmidt DA, et al. 2009. Rattling in the cage: ions as probes of sub-picosecond water network dynamics. *J Am Chem Soc.* 131(51):18512–18517.
- 54 Funkner S, et al. 2012. Watching the low-frequency motions in aqueous salt solutions: the terahertz vibrational signatures of hydrated ions. *J Am Chem Soc.* 134(2):1030–1035.
- 55 Schwaab G, Sebastiani F, Havenith M. 2019. Ion hydration and ion pairing as probed by THz spectroscopy. *Angew Chem Int Ed.* 58(10):3000–3013.
- 56 Nierzwicki L, et al. 2021 Dec. Enhanced specificity mutations perturb allosteric signaling in CRISPR–Cas9. *eLife.* 10:e73601.
- 57 VanWart AT, Eargle J, Luthey-Schulten Z, Amaro RE. 2012 Aug. Exploring residue component contributions to dynamical network models of allostery. *J Chem Theory Comput.* 8(8):2949–2961.
- 58 Gori M, Floriani E, Marco P. 2021 Jun. Theoretical proposals for the experimental detection of electrodynamic interactions between biomolecules. Working paper or preprint.
- 59 Kolodner P, Lukashov EP, Ching YC, Rousseau DL. 1996 Oct. Electric-field-induced Schiff-base deprotonation in D85N mutant bacteriorhodopsin. *Proc Natl Acad Sci USA.* 93(21):11618–11621.
- 60 Salcedo E, Zheng L, Phistry M, Bagg EE, Britt SG. 2003 Nov. Molecular basis for ultraviolet vision in invertebrates. *J Neurosci.* 23(34):10873–10878.
- 61 Gaal P, et al. 2007 Dec. Internal motions of a quasiparticle governing its ultrafast nonlinear response. *Nature.* 450(7173):1210–1213.
- 62 Biswas A, Moran A, Milovic D, Majid F, Biswas KC. 2010. An exact solution for the modified nonlinear Schrödinger’s equation for Davydov solitons in α -helix proteins. *Math Biosci.* 227(1):68–71.
- 63 Kondepudi D, Petrosky T, Pojman JA. 2017. Dissipative structures and irreversibility in nature: celebrating 100th birth anniversary of Ilya Prigogine (1917–2003).
- 64 Parry BR, et al. 2014. The bacterial cytoplasm has glass-like properties and is fluidized by metabolic activity. *Cell.* 156(1–2):183–194.
- 65 Kohen A, Cannio R, Bartolucci S, Klinman JP, Klinman JP. 1999 Jun. Enzyme dynamics and hydrogen tunnelling in a thermophilic alcohol dehydrogenase. *Nature.* 399(6735):496–499.
- 66 Zhang J, Klinman JP. 2011 Nov. Enzymatic methyl transfer: role of an active site residue in generating active site compaction that correlates with catalytic efficiency. *J Am Chem Soc.* 133(43):17134–17137.