

# Persistent protein motions in a rugged energy landscape revealed by Normal Mode Ensemble Analysis

Tod D. Romo†, Alan Grossfield†, and Andrea G. Markelz‡

†Department of Biochemistry and Biophysics, University of Rochester Medical Center, Rochester, NY, 14642

‡Department of Physics, University at Buffalo, SUNY, Buffalo, NY, 14260

---

**ABSTRACT:** Proteins are allosteric machines that couple motions at distinct, often distant, sites to control biological function. Low-frequency structural vibrations are a mechanism of this long-distance connection and are often used computationally to predict correlations, but experimentally identifying the vibrations associated with specific motions has proved challenging. Spectroscopy is an ideal tool to explore these excitations, but measurements have been largely unable to identify important frequency bands. The result is at odds with some previous calculations and raises the question what methods could successfully characterize protein structural vibrations. Here we show the lack of spectral structure arises in part from the variations in protein structure as the protein samples the energy landscape. However, by averaging over the energy landscape as sampled using an aggregate 18.5 $\mu$ s of all-atom molecular dynamics simulation of hen egg white lysozyme and normal mode analyses, we find vibrations with large overlap with functional displacements are surprisingly concentrated in narrow frequency bands. These bands are not apparent in either the ensemble averaged vibrational density of states or isotropic absorption. However, in the case of the ensemble averaged anisotropic absorption, there is persistent spectral structure and overlap between this structure and the functional displacement frequency bands. We systematically lay out heuristics for calculating the spectra robustly, including the need for statistical sampling of the protein and inclusion of adequate water in the spectral calculation. The results show the congested spectrum of these complex molecules obscures important frequency bands associated with function, and reveal a method to overcome this congestion by combining structurally sensitive spectroscopy with robust normal mode ensemble analysis.

---

## ■ INTRODUCTION

It has been speculated that long range protein vibrations can serve to actuate allosteric control by perturbing the configurations sampled by the vibrational atomic displacements<sup>1-4</sup>. Elastic network models and all-atom calculations have shown that the structural changes necessary for protein function can be reproduced using the lowest frequency vibrations, suggesting these motions are essential to biological outcomes<sup>5-7</sup>. This possibility is currently being explored as a mechanism for allosteric control where the agonist or inhibitor is designed to perturb the long-range structural vibrations<sup>8-9</sup>. While some all-atom calculations have suggested that mutations change the directionality of motion and therefore enzyme promiscuity<sup>10</sup>, there has been no direct experimental confirmation of the control via these vibrations. In particular, standard optical techniques fail to even isolate specific structural vibrations, a first step towards identifying those vibrations with biochemical impact.

The lack of spectroscopic structure in optical measurements has been confounding, as some computational investigations have suggested that narrow resonances should be present<sup>11-13</sup>, allowing the monitoring of the vibrations with mutation and/or inhibitor binding to establish dynamical hot spots and reveal the design strategies that lead to robust and efficient biochemistry. Often calculations predicting spectroscopic fingerprints in isotropic absorption measurements is the use of a single starting structure and normal mode analysis. Here we show that this ap-

proach is insufficient and prone to significant error because it does not account for the time and ensemble averaging as the macromolecule samples the energy landscape. To address this we developed Normal Mode Ensemble Analysis (NMEA). After establishing a protocol which includes sufficient water and vibrations to accurately calculate an individual spectrum, we use NMEA to calculate experimentally accessible spectra: vibrational density of states; isotropic absorption; and anisotropic absorption. For these studies we use the benchmark protein chicken egg white lysozyme (CEWL), for which there are extensive measurements of picosecond dynamics using a wide variety of techniques and whose results are representative of other proteins. Neutron scattering, Raman, Optical Kerr and isotropic terahertz spectroscopy have found broad vibrational density of states and isotropic absorbance, with only slight changes in this absorbance with functional state<sup>14-15</sup>. However, recent terahertz anisotropic absorption studies on CEWL find narrow band resonances that are sensitive to inhibitor binding<sup>16-17</sup>. By rigorously averaging spectra calculated from structures sampled from microsecond-scale all-atom trajectories, we find that indeed the averaged isotropic spectra do not have narrow resonances, and resemble the smooth absorbance seen experimentally. On the other hand, resonant bands are present in the average anisotropic absorbance spectra. We compare the spectra of CEWL with and without inhibitor to examine spectral sensitivity and find that only the anisotropic absorption shows a statistically significant change with inhibitor binding. We then examine if spectral changes can

be assigned to specific structural motions, by projecting the massive data set onto known functional motions. The analysis of the full dynamical ensemble identifies several low frequency narrow spectral bands that overlap the regions of change in the anisotropic absorption.

## ■ MATERIALS AND METHODS

Figure 1 shows a schematic of the overall computational strategy. Multiple replica trajectories are simulated. NMA starting structures are sampled from these trajectories, and the VDOS, isotropic absorption and anisotropic absorption are then calculated and averaged; this averaging is essential to identify robust features that could in principle be observed experimentally. A projection analysis of the full data set is then used to perform a simple distillation of the large data set for the prevalence of vibrations overlapping displacements for substrate binding.

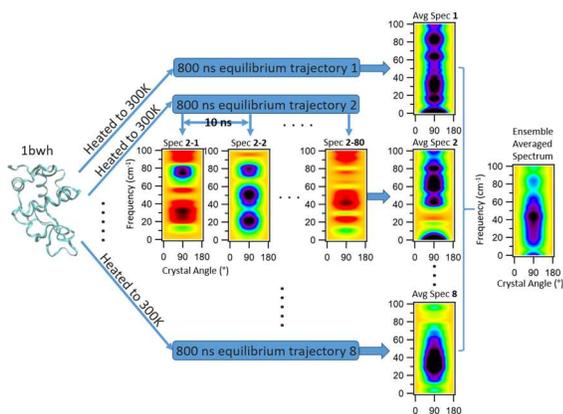


Figure 1. Schematic of the computational approach for Normal Mode Ensemble Analysis. Spectra are computed for individual structures to produce an average anisotropic spectrum.

### Molecular dynamics simulations

The apo and inhibitor-bound simulations at 300K were constructed from their respective X-ray structures (Apo: 1BWH, Inhibitor: 1HEW, containing 1 4C1 beta-D-glucose molecule), then inserted into a thermalized cubic box containing 8682 waters with 8 chlorides to produce an electrically neutral system using the OptimalMembraneGenerator solvate tool from LOOS<sup>18-19</sup>; 10 replicas were built, using the same protein coordinates but different placements of water and ions. We then performed 1000 steps of minimization with the alpha carbons fixed in space, followed by 1000 steps with the alpha carbons harmonically restrained using a force constant of 20 kcal/mol-Å<sup>2</sup>. The temperature was gradually increased from 30K to 300K by running 20 ps simulations followed by 30K temperature increments. We then ran 200 ps of unrestrained dynamics before beginning production simulations.

All simulations were performed using NAMD<sup>20</sup> version 2.11 on the Blue Gene/Q supercomputer at the University of Rochester's Center for Integrated Research Computing. Simulations were performed using a 2 fs timestep, with bonds to hydrogen constrained to their equilibrium length using SETTLE and the tolerance set to 10<sup>-10</sup>. A target temperature of 300K was maintained by using stochastic dynamics with collision frequency set to 2 ps<sup>-1</sup>, and no collisions applied to the hydrogens. The pressure was set to 1

atm via the Langevin-Piston algorithm<sup>21</sup>, with piston period set to 200 ps, a piston temperature of 300K, and a piston decay of 100 ps. Repulsion-dispersion interactions were cut off using force-switching between 9 and 10 Å, and long-range electrostatics computed using Ewald summation<sup>22</sup> on a 72 x 72 x 72 grid (roughly 1 Å /grid point). The CHARMM36 force field<sup>23</sup> was used, in combination with the CHARMM version of TIP3P water. Parameters for tri acetyl glucosamine (3NAG) were obtained from Niessen, et al.<sup>16</sup>

### Computing spectra from simulations

The vibrational density of states (VDOS) for a given structure is computed using normal mode analysis (NMA), as implemented in CHARMM v39. We compute the first 6000 vibrational modes rather than the full spectrum, to reduce the cost of computation and storage required by CHARMM, and to reduce time required to post-process the resulting modes into absorbance spectra (see supplemental information). The same normal mode analysis is used to compute the absolute isotropic absorbance:

$$Abs_{iso}(\omega) \approx \sum_{i=7}^{3N} \frac{\gamma^2/\omega_i}{(\omega - \omega_i)^2 + \gamma^2} \left[ \left| \frac{\partial \vec{p}}{\partial q_i} \right| \right]^2 \quad (1)$$

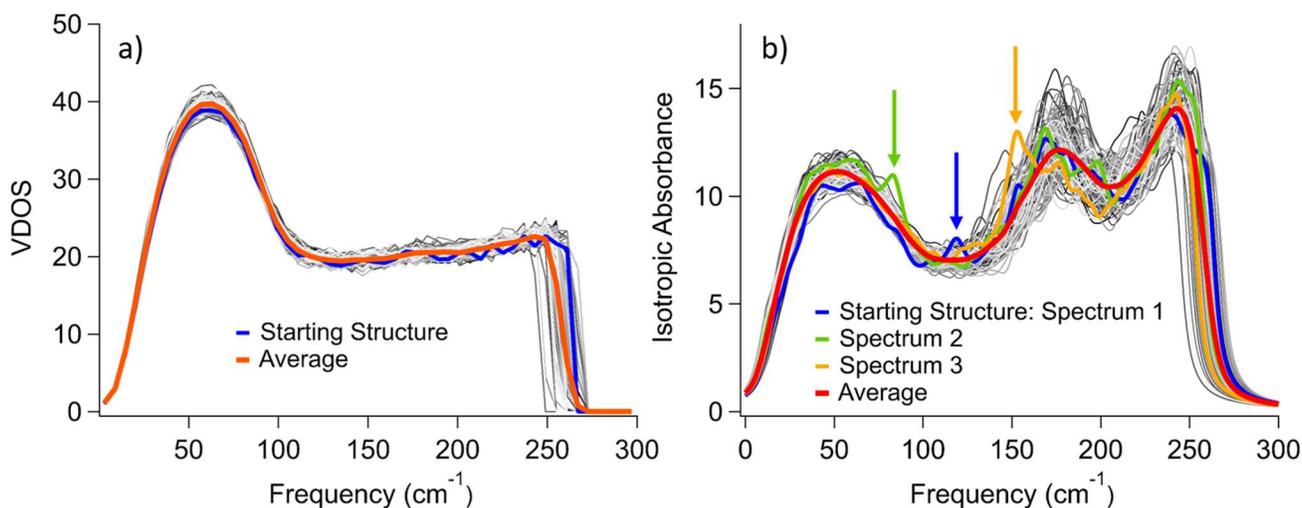
Here,  $\omega$  is the output frequency,  $\omega_i$  is the frequency of the  $i$ th mode,  $\frac{\partial \vec{p}}{\partial q_i}$  is the dipole derivative associated with the  $i$ th mode, and  $\gamma = 4 \text{ cm}^{-1}$  is the linewidth.  $N$  is the number of atoms included in the normal mode calculation, generally the protein plus a number of surrounding water molecules, so  $3N$  is the resulting number of modes. The 6 lowest frequency modes are discarded as they are due to rigid body translation and rotation.

The anisotropic absorbance measures the absorption dependence of an aligned sample on the light polarization direction given by the angle of the polarization with respect to a reference direction. It is computed in an analogous manner:

$$Abs_{ani}(\omega, \theta) \approx \sum_{j=xt} \sum_{\substack{\text{rotation} \\ \text{matrices}}} \sum_{i=7}^{3N} \frac{\gamma^2/\omega_i}{(\omega - \omega_i)^2 + \gamma^2} \left[ \vec{R}_j \frac{\partial \vec{p}}{\partial q_i} \cdot \hat{\lambda} \right]^2 \quad (2)$$

Here,  $\vec{R}_j$  is the rotation matrix for the crystal and  $\hat{\lambda}$  is the light polarization vector associated with the rotation angle  $\theta$ , and other terms are the same as in Equation 1. In this work, we computed spectra appropriate for a tetragonal crystal with 8 protein chains in the unit cell, oriented as was found experimentally<sup>24</sup>. The equivalent spectra for other crystal forms could be computed from the same simulations by using  $\vec{R}_j$  matrices appropriate to the contents of the unit cell and space group. To emphasize the variation with orientation, we subtract the spectrum at  $\theta = 0$  from all values and this is referred to as  $\Delta Abs_{ani}(\omega, \theta)$ . The polarization vector is  $\hat{\lambda} = (\frac{1}{\sqrt{2}} \sin\theta, \frac{1}{\sqrt{2}} \sin\theta, \frac{1}{\sqrt{2}} \cos\theta)$ , which lays in the plane of the (1,1,0) facet of the crystal.

As part of this work, we systematically determined how much water must be included in the normal mode calculations to correctly represent the protein dynamics, as well



**Figure 2** Sensitivity of calculated VDOS and Isotropic Absorbance to specific structure. The gray curves are spectra calculated by sampling structures from a representative trajectory. Panel a shows the VDOS, while Panel b shows the isotropic absorbance. The average and starting structure spectra are shown in red and blue, respectively. The green and orange curves in Panel b were generated using two other random structures from the ensemble, to illustrate that finer features in the spectra from individual structures average away when considering the whole ensemble.

as the number of modes one must calculate to produce reliable spectra over the typical experimental frequency range. In the supplemental information, we show that the spectral calculations require a water shell of at least 8-10 Å to converge for the spectral frequency 0 - 100  $\text{cm}^{-1}$ . With this water thickness, at least 6000 modes are needed for a single structure NMA to adequately cover the range of 0-100  $\text{cm}^{-1}$ . The mode number cutoff depends on the size of the system, so it is necessary to tune the number of modes included based on the specific system contents and the frequency range of interest. We did not systematically test the relative value of running a small number of long trajectories vs. an ensemble of relatively short ones. This is an active area of research,<sup>25-26</sup> and the optimal bang for the buck is almost certainly dependent on the specifics of the system (including the temperature), as well as the computational resources available.

**Table 1. Summary of simulations performed.** “Duration/traj” denotes the range of trajectory lengths, rounded off. “Aggregate” denotes the total simulation time of all simulations.

Ligand	Replicas	Duration/traj (ns)	Aggregate (ns)
Apo	10	940	9426
Inhibitor	10	755-985	9608

## DATA SHARING AND REPRODUCIBILITY

The scripts to perform the MD simulations, generate the modes in CHARMM, and compute the isotropic and anisotropic spectra are available from GitHub at <https://github.com/GrossfieldLab/NMEA-paper>.

## RESULTS

Figure 2 shows the calculated vibrational density of states (VDOS) and isotropic absorbance for the apo protein at

300 K, using one of the replicate simulations. Each gray curve represents the spectra calculated from single snapshots, while the thicker red lines represent the average; the blue lines represent the spectrum computed using the starting structure from the simulation. In panel a), the VDOS shows a single broad peak centered at ca. 63  $\text{cm}^{-1}$ , followed by a broad plateau. The steep drop in amplitude after  $\sim 250 \text{ cm}^{-1}$  is due to the truncation of the eigendecomposition at 6000 modes (see Supplemental Information for further discussion). While there is some snapshot-to-snapshot variation, overall the VDOS is not sensitive to starting structure. The peak and plateau are consistent with Raman and neutron scattering measurements however the measured peak is ca. 80  $\text{cm}^{-1}$ <sup>27-28</sup>. Velocity autocorrelation calculations of the CEWL VDOS from Lerbret et al. are in better agreement with the frequency of the peak, and this may be due to the SPC/E water model used. The NMEA calculations capture the overall VDOS structure and can be used to assign structural motions to spectral peaks, which is not possible with autocorrelation functions.

Figure 2b) shows the isotropic absorption spectrum from the same trajectory calculated using equation 1. In the contrast to the VDOS, the spectra from individual frames vary significantly; each individual spectrum has rapid variations in the absorbance with frequency, often with a few prominent narrow band resonances highlighted for a few of the spectra by arrows in the figure. However, these features do not persist with averaging over the ensemble and only three broad bands remain at 52  $\text{cm}^{-1}$ , 177  $\text{cm}^{-1}$  and 253  $\text{cm}^{-1}$  (thick red line). As before, the signal drop beyond  $\sim 250 \text{ cm}^{-1}$  is an artifact due to truncation of the eigendecomposition rather than any physical factors. CEWL measurements have found even less structure, with only strong broad peaks at ca. 160  $\text{cm}^{-1}$  and ca. 320  $\text{cm}^{-1}$ <sup>29-30</sup>. We note that the spectrum computed using just the starting structure (blue curve) is fairly close to the ensemble average, but appears to contain fine structure not present in the average. Those features are not numerical noise, but rather

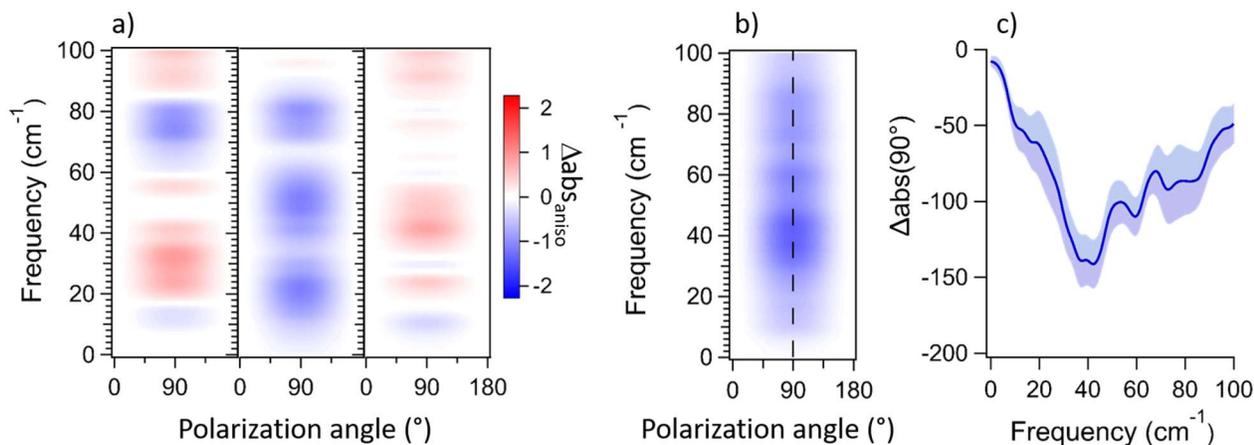


Figure 3 Sensitivity of anisotropic absorption to specific structure. Panel a shows the anisotropic spectra for three individual structures. Panel b shows the average over 8000 spectra. Panel c shows a slice of the average spectrum with polarization = 90°; the shaded region is the standard error.

are due to structural details that vary as the protein moves; we believe they would be visible if it were possible to do a time-resolved single-molecule variant of the experiment. **This demonstrates that spectra intended to be compared to experiment cannot be computed reliably from a single structure**; at a minimum it is necessary to run a reasonable-length MD simulation.

NMEA anisotropic absorption spectra for the same configuration in those measurements for several different starting structures is shown in Fig. 3a). The heat maps of the  $\Delta Abs_{aniso}(\theta)$  vary dramatically for the different structures with no obvious common features among the spectra. It would be reasonable to expect that averaging in this case will lead to a completely featureless spectrum, however as the data in Fig. 3 b) shows, this is not the case. Distinct and relatively narrow spectral bands emerge as the  $\Delta Abs_{aniso}(\theta)$  is averaged over the landscape. The specific bands that emerge are not obvious from any of the individual spectra, (see Supplemental information movie illustrating the emergence of the spectrum), indicating that it is the ensemble sampling of motions that has a net bias. The spectra are symmetric about 90° due to the symmetry of the tetragonal crystal used; other crystal forms will show more angular dependence. The symmetry of the tetragonal spectra makes it easier to understand the spectrum in 1-dimensional form, by considering only the curve at a polarization of 90 degrees, shown with its standard error in Fig. 3c). There is substantial structure in the averaged spectrum beyond the standard error, consistent with recent experimental results.

To examine if the different spectral methods can be used to follow changes in structural dynamics with functional significance, we computed spectra for the inhibitor-bound structure of lysozyme, in addition to the apo results discussed above. Figures 4a) and 4b) show that the VDOS and isotropic spectra are largely insensitive to the changes in structure and dynamics upon ligand binding; the amplitude of the change is less than 1.5% of the total signal, significantly smaller the statistical uncertainty. It is perhaps notable that the NMEA decrease in the VDOS with inhibitor binding for frequencies less than 60  $\text{cm}^{-1}$  is consistent with neutron scattering measurements<sup>16</sup>.

By contrast there are clear differences between the NMEA anisotropic absorbance spectra for the apo and inhibitor-bound ensembles, as shown in Figure 4c). The difference between the  $\Delta Abs_{ani}(\omega, \theta)$  for the apo and inhibitor-bound states,  $\Delta \Delta Abs_{ani}(\omega, \theta)$ , which we refer to as the double difference spectrum, has several sharp features, most notably the peak at 9  $\text{cm}^{-1}$ . This feature is particularly intriguing in light of the projection results to be discussed next.

To understand the relevance of  $\Delta Abs_{ani}(\omega, \theta)$  and  $\Delta \Delta Abs_{ani}(\omega, \theta)$  to biological function requires connecting the spectral features to structural motions. This is a question that is uniquely suited to the atomic resolution of the normal mode approach. We can analyze the large data set of vibrations robustly sampled over the landscape to see if specific displacements dominate, and further if they dominate in distinct frequency ranges. We can do this by projecting the eigenvectors for the full data set onto the displacement vector of interest. Previously this projection strategy was used to look for commonality in motions within a single data set<sup>31</sup>. Here we examine persistent motions throughout the energy landscape.

To demonstrate this strategy, we consider the normalized C- $\alpha$  displacement vector defined by the difference between the apo and inhibitor bound structures,  $\Delta \vec{x}_{binding}$ . For each of the 6000 modes computed for each structure (totaling > 1.8 million vibrational modes in aggregate), we take the absolute dot product between the normalized eigenvector and  $\Delta \vec{x}_{binding}$ . An overlap of 1 would indicate that the atomic displacements for a vibration are identical to  $\Delta \vec{x}_{binding}$ , while zero indicates orthogonality. We plot the results with the amplitude of this projection along the y-axis, and the frequency as the x-axis. The results are shown as Figure 5.

Figure 5a) shows a grayscale heat map of the relative density of projections at a given frequency and projection value, with the average projection value indicated by the blue curve. There are two sharp peaks at 6  $\text{cm}^{-1}$  and 8.5  $\text{cm}^{-1}$  and a lesser peak at 15  $\text{cm}^{-1}$ . The amplitudes of the first two peaks are both above 0.3; to put this in perspective, the overlap distribution for random unit vectors of the same dimension drops below  $10^{-9}$  at a value of roughly 0.18 (see Supplemental Figure 6), making an average overlap of 0.3

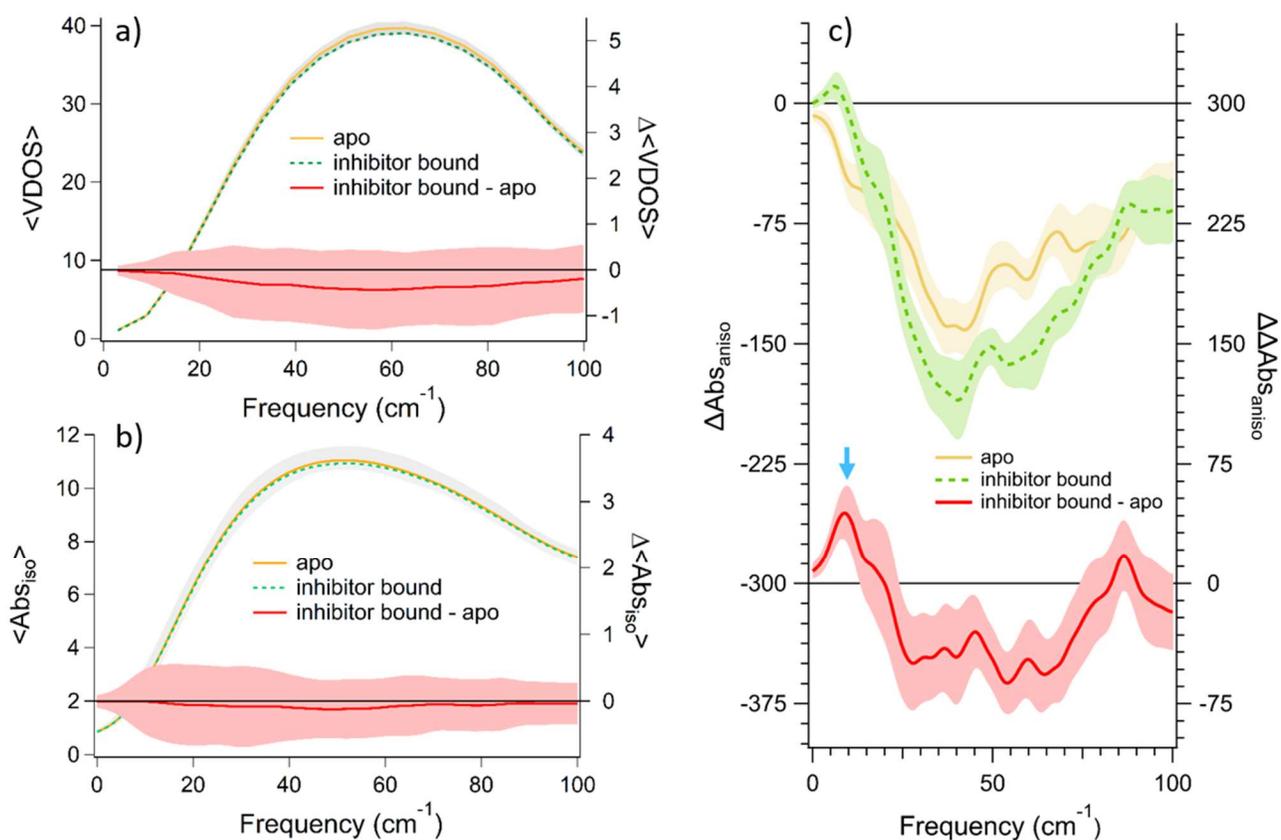


Figure 4. Anisotropic spectra detect signal due to inhibitor binding. a) VDOS. b) Isotropic absorbance and difference spectra. c) Anisotropic absorption at polarization= $90^\circ$ . In each panel, the left y-axis shows the average spectra for the apo (yellow) and inhibitor-bound (green) states, while the right y-axis applies to the difference spectrum (red). The shaded regions are the standard error, computed treating each trajectory as a single measurement.

for a particular band exceedingly unlikely to occur at random. Conventional wisdom would dismiss the idea that structural vibrations would have any bias towards functional long-range motions. However, with the present data set where the vibrations are robustly sampled, we find that in fact there is a predominance of vibrational displacements corresponding to functional motions.

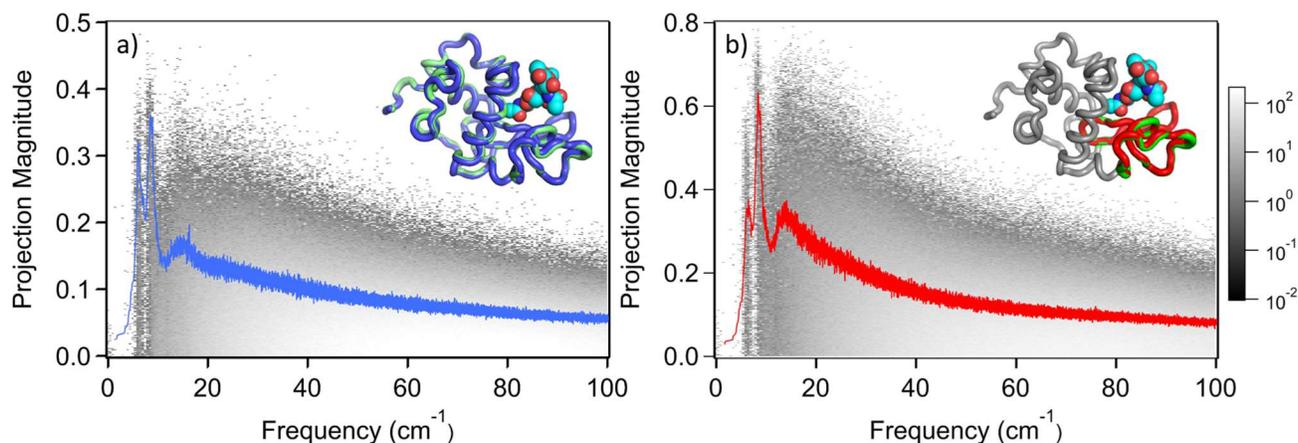
These peaks in Figure 5b) are even more striking; here we focus the atomic displacements for the loop with highest displacements with substrate binding, residues 42-79,  $\Delta\vec{x}_{loop}$ . These peaks are at  $6\text{ cm}^{-1}$ ,  $8\text{ cm}^{-1}$  and  $14\text{ cm}^{-1}$ , with the  $8\text{ cm}^{-1}$  peak dominating. The average overlap for the  $8\text{ cm}^{-1}$  is an astoundingly high 0.6, indicating some real specificity of displacements at this frequency; Supplemental Figure 6 shows that the probability distribution of the overlap for random vectors with this dimensionality drops below  $10^{-9}$  by about 0.3; given that there are roughly 2 million computed modes in in Figure 5, this suggests we wouldn't expect to see even a single mode with that high an overlap at random, let alone a strong peak. This sharp band (and the  $8.5\text{ cm}^{-1}$  band from the full protein spectrum) overlaps the  $9\text{ cm}^{-1}$  peak in the  $\Delta\Delta\text{Abs}_{ani}(\omega, \theta)$  shown in Figure 4 c), allowing us for the first time to assign physical significance to a specific band in a terahertz absorbance spectrum.

It is striking that there are identifiable peaks in the functional projection spectrum. If proteins are optimized for function, it is not obvious why there should be any sharp

features in the projection spectrum, especially in the case of lysozyme, where there is no apparent need for resonant energy exchange between enzyme and substrate. We speculate that it is possible that the strong energy-dependence for functionally important motions is a general feature that may be highly useful for those systems where energy exchange is crucial, such as photo-active proteins.

## DISCUSSION

In this work, we examine if different spectral measurements can isolate correlated structural displacements towards understanding how such motions impact function. The normal mode approach provides both the vibrational energies and their displacement vectors. It should be noted that the present normal mode method is distinct from approaches based on principal component analysis or quasiharmonic analysis. The modes obtained from principal component analysis capture the directions of the fluctuations in the protein's ensemble. By contrast, each set of normal modes is computed from a single structure, and reflects the instantaneous vibrational spectrum of the molecule at that moment. Given the extreme roughness of a protein's energy landscape, these instantaneous modes do not necessarily converge to those from principal component analysis, even after computing their ensemble average. From a classical perspective, the interaction of a photon with the protein occurs instantaneously, so only the normal mode approach is appropriate to model absorption



*Figure 5. Vibrational displacement overlap with functional motions. The projection of each computed mode onto the difference between the apo and inhibited structures is shown, using either the full protein (Panel a) or just the most-effected loop (Panel b). The grayscale heat map shows the relative density of projections at a given frequency and projection value. The solid lines show the binned average projection value as a function of frequency. The inset in a) shows the apo CEWL structure in green and the inhibitor-bound in blue, with the inhibitor rendered in space-filling mode. The inset in b) shows the apo CEWL in gray. The green (red) region corresponds to residues 42-79 for the apo (bound) states.*

spectra. The ensemble averaging we discuss above is essentially an average over the incident photons, each of which picks out 1 snapshot from within the ensemble; the experiment is an average over a large number of photon absorption events.

A common objection to the normal mode approach is that it effectively assumes zero temperature. The present approach addresses that by explicitly sampling at finite temperature, such that the structures included in the calculation are representative of the ensemble under ambient conditions. We would argue that the weakness is relatively minor, considering that normal modes allow structural interpretations not available from other arguably more rigorous approaches, such as those based the dipole autocorrelations 32-34. We note that NMEA is distinct from instantaneous normal modes which was first introduced for modeling solvents and is focused on the rotational and translational motions of the individual solvent molecules 35-36. Instantaneous normal modes have been applied to protein dynamics 37 to investigate energy relaxation as opposed to describing the distinct vibrational spectra sampled discussed here.

One modest limitation of the current approach is the difference between the simulated system and that measured experimentally. The anisotropic measurements are performed in a crystalline environment 16-17, 38, while the simulations contained a single protein molecule in solution. While it is in principle possible to simulate a protein crystal 39-41, these calculations are extremely challenging. Most space groups are not readily compatible with the Ewald-based electrostatics methods considered to be molecular dynamics best-practice, and beyond that the presence of the full unit cell would greatly increase the system size and thus the computational cost of the simulation. Moreover, protein crystals contain a relatively large amount of water, and other than specific crystal contacts the dynamics are expected to at least grossly resemble a solvated protein.

The anisotropic absorbance has finer structure because the orientation variable selects motion in specific directions; essentially, the orientation vector gives another dimension to the signal, uncovering underlying structure in a manner analogous to 2D spectroscopic techniques. One must then ask: Is there a way to know why those frequencies are different? Experimentally, it is possible through repeated experiments with mutations or chemical constraints that eliminate specific motion; the effect on those spectral features would begin to assign the motions. This approach is labor intensive whereas simulations provide an effective and efficient avenue to identify structure significance of the frequency features. Our present projection approach reveals that functional motions are concentrated in specific energy bands. In principle, we can extend this analysis to more general motional types, such as net torsional motions about the binding site.

## ■ CONCLUSION

We present a new computational technique where normal mode analysis is applied to a protein ensemble to extract the VDOS, isotropic and anisotropic spectra, and for the first time to assign a frequency band to a specific change in structure associated with inhibitor binding. We find the ensemble cannot be ignored for accurately calculating the measurements and for revealing underlying vibrational sampling of functional motions. NMEA finds a smooth VDOS with little sensitivity to the binding state. Isotropic spectra computed from individual structures often have fine features, but these average away over a thermal distribution in agreement with the majority of measurements. Further we find large structure to structure variations in the anisotropic absorption spectra, however persistent bands emerge with averaging of these spectra, and these bands change with inhibitor binding. A projection analysis of the thermal distribution of vibrations shows eigenvectors corresponding to binding displacements are concentrated in specific frequency bands and at least one of these bands overlaps the spectral changes seen in the anisotropic absorption. These bands would not have been obvious

from any of spectral measurements and only emerge when analyzing the massive computed data set. The congested vibrational spectrum obscures these underlying resonances for standard spectroscopy. To measure the vibrations that may impact biology, techniques must isolate those vibrations based on structural motions, such as transition dipole direction. All-atom molecular dynamics can provide an unprecedented capability to identify functional motions that is unavailable using prior techniques.

Testing the premise of evolutionarily optimized protein dynamics has remained an experimental challenge. Most measurements fail to isolate specific structural motions. Our simulations show that the structural variation of a single protein in time results in variation in the vibrations leading to the observed broad and featureless optical absorption. However, when the thermal population of a protein's configurations are considered, vibrations with functional displacements are concentrated in specific frequency bands. These emergent dynamics are apparent in anisotropic optical absorbance, indicating an experimental avenue for measuring these motions and their impact on biological function.

## ASSOCIATED CONTENT

**Supporting Information.** There are 6 supplemental figures, plus a movie of the average anisotropic spectrum showing the emergence of bands as spectra are averaged. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Author

\*[amarkelz@buffalo.edu](mailto:amarkelz@buffalo.edu)

### Author Contributions

T. Romo performed calculations. The manuscript was written through contributions of all authors.

### Funding Sources

National Science Foundation (MCB 1616529); U.S. Department of Energy (DE-SC0016317)

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENT

We thank the Center for Integrative Research Computing at the University of Rochester for access to their high-performance computing facility. We also thank the Center for Computational Research at the University at Buffalo for use of their computational facilities (<http://hdl.handle.net/10477/79221>).

## REFERENCES

1. Bahar, I.; Chennubhotla, C.; Tobi, D., Intrinsic dynamics of enzymes in the unbound state and, relation to allosteric regulation. *Curr. Opin. Struct. Biol.* **2007**, *17* (6), 633-640.
2. Tehver, R.; Chen, J.; Thirumalai, D., Allostery Wiring Diagrams in the Transitions that Drive the GroEL Reaction Cycle. *J. Mol. Biol.* **2009**, *387* (2), 390-406.
3. Gerek, Z. N.; Ozkan, S. B., Change in Allosteric Network Affects Binding Affinities of PDZ Domains: Analysis through Perturbation Response Scanning. *Plos Computational Biology* **2011**, *7* (10).
4. Katebi, A. R.; Jernigan, R. L., The critical role of the loops of triosephosphate isomerase for its oligomerization, dynamics, and functionality. *Protein Sci* **2014**, *23* (2), 213-28.
5. Whitmire, S. E.; Wolpert, D.; Markelz, A. G.; Hillebrecht, J. R.; Galan, J.; Birge, R. R., Protein flexibility and conformational state: A comparison of collective vibrational modes of wild-type and D96N bacteriorhodopsin. *Biophys. J.* **2003**, *85* (2), 1269-1277.
6. Yu, X.; Leitner, D. M., Heat flow in proteins: Computation of thermal transport coefficients. *J. Chem. Phys.* **2005**, *122* (5).
7. Go, N.; Noguti, T.; Nishikawa, T., Dynamics of a Small Globular Protein in Terms of Low-Frequency Vibrational-Modes. *Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences* **1983**, *80* (12), 3696-3700.
8. Porter, J. R.; Moeder, K. E.; Sibbald, C. A.; Zimmerman, M. I.; Hart, K. M.; Greenberg, M. J.; Bowman, G. R., Cooperative Changes in Solvent Exposure Identify Cryptic Pockets, Switches, and Allosteric Coupling. *Biophys. J.* **2019**, *116* (5), 818-830.
9. Zorba, A.; Nguyen, V.; Koide, A.; Hoemberger, M.; Zheng, Y. J.; Kutter, S.; Kim, C.; Koide, S.; Kern, D., Allosteric modulation of a human protein kinase with monobodies. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (28), 13937-13942.
10. Miller, D. W.; Agard, D. A., Enzyme specificity under dynamic control: A normal mode analysis of alpha-lytic protease. *J. Mol. Biol.* **1999**, *286* (1), 267-278.
11. Thirumuruganandham, S. P.; Urbassek, H. M., Low-frequency vibrational modes and infrared absorbance of red, blue and green opsin. *Journal of Molecular Modeling* **2009**, *15* (8), 959-969.
12. Zhang, H. L.; Zukowski, E.; Balu, R.; Gregurick, S. K., A dynamics study of the A-chain of ricin by terahertz vibrational calculation and normal modes analysis. *J. Mol. Graph.* **2009**, *27* (5), 655-663.
13. Castro-Camus, E.; Johnston, M. B., Conformational changes of photoactive yellow protein monitored by terahertz spectroscopy. *Chem. Phys. Lett.* **2008**, *455* (4-6), 289-292.
14. Falconer, R. J.; Markelz, A. G., Terahertz Spectroscopic Analysis of Peptides and Proteins. *Journal of Infrared Millimeter and Terahertz Waves* **2012**, *33* (10), 973-988.
15. Cote, Y.; Nomine, Y.; Ramirez, J.; Hellwig, P.; Stote, R. H., Peptide-Protein Binding Investigated by Far-IR Spectroscopy and Molecular Dynamics Simulations. *Biophys. J.* **2017**, *112* (12), 2575-2588.
16. Niessen, K. A.; Xu, M. Y.; Paciaroni, A.; Orecchini, A.; Snell, E. H.; Markelz, A. G., Moving in the Right Direction: Protein Vibrational Steering Function. *Biophys. J.* **2017**, *112* (5), 933-942.
17. Niessen, K. A.; Xu, M.; George, D. K.; Chen, M. C.; Ferre-D'Amare, A. R.; Snell, E. H.; Cody, V.; Pace, J.; Schmidt, M.; Markelz, A. G., Protein and RNA dynamical fingerprinting. *Nat Commun* **2019**, *10* (1), 1026.
18. Romo, T. D.; Grossfield, A., LOOS: an extensible platform for the structural analysis of simulations. *Conf Proc IEEE Eng Med Biol Soc* **2009**, *2009*, 2332-5.
19. Romo, T. D.; Leioatts, N.; Grossfield, A., Lightweight object oriented structure analysis: tools for building tools to analyze molecular dynamics simulations. *J Comput Chem* **2014**, *35* (32), 2305-18.
20. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K., Scalable molecular dynamics with NAMD. *J Comput Chem* **2005**, *26* (16), 1781-1802.
21. Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R., Constant-Pressure Molecular-Dynamics Simulation - the Langevin Piston Method. *J. Chem. Phys.* **1995**, *103* (11), 4613-4621.
22. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103* (19), 8577-8593.

23. Huang, J.; MacKerell, A., CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **2013**, *34* (25), 2135-2145.
24. Dong, J.; Boggon, T. J.; Chayen, N. E.; Raftery, J.; Bi, R. C.; Helliwell, J. R., Bound-solvent structures for microgravity-, ground control-, gel- and microbatch-grown hen egg-white lysozyme crystals at 1.8 Å resolution. *Acta Crystallogr D* **1999**, *55*, 745-752.
25. Grossfield, A.; Patrone, P. N.; Roe, D. R.; Schultz, A. J.; Siderius, D. W.; Zuckerman, D. M., Best Practices for Quantification of Uncertainty and Sampling Quality in Molecular Simulations [Article v1.0]. *Living J Comput Mol Sci* **2018**, *1* (1).
26. Grossfield, A.; Zuckerman, D. M., Quantifying uncertainty and sampling quality in biomolecular simulations. *Annu Rep Comput Chem* **2009**, *5*, 23-48.
27. Sassi, P.; Perticaroli, S.; Comez, L.; Giugliarelli, A.; Paolantoni, M.; Fioretto, D.; Morresi, A., Volume properties and spectroscopy: A terahertz Raman investigation of hen egg white lysozyme. *J. Chem. Phys.* **2013**, *139* (22).
28. Lerbret, A.; Affouard, F.; Bordat, P.; Hedoux, A.; Guinet, Y.; Descamps, M., Low-frequency vibrational properties of lysozyme in sugar aqueous solutions: A Raman scattering and molecular dynamics simulation study. *J. Chem. Phys.* **2009**, *131* (24).
29. Zakaria, H. A.; Fischer, B. M.; Bradley, A. P.; Jones, I.; Abbott, D.; Middelberg, A. P. J.; Falconer, R. J., Low-Frequency Spectroscopic Analysis of Monomeric and Fibrillar Lysozyme. *Appl. Spectrosc.* **2011**, *65* (3), 260-264.
30. Stehle, C. U.; Abuillan, W.; Gompf, B.; Dressel, M., Far-infrared spectroscopy on free-standing protein films under defined temperature and hydration control. *J. Chem. Phys.* **2012**, *136* (7).
31. Horiuchi, T.; Go, N., Projection of Monte-Carlo and Molecular-Dynamics Trajectories onto the Normal Mode Axes - Human Lysozyme. *Proteins-Structure Function and Genetics* **1991**, *10* (2), 106-116.
32. Xu, D.; Phillips, J. C.; Schulten, K., Protein Response to External Electric Fields: Relaxation, Hysteresis, and Echo. *J. Phys. Chem.* **1996**, *100*, 12108-12121.
33. He, Y. F.; Chen, J. Y.; Knab, J. R.; Zheng, W. J.; Markelz, A. G., Evidence of Protein Collective Motions on the Picosecond Timescale. *Biophys. J.* **2011**, *100* (4), 1058-1065.
34. Choi, J. H.; Cho, M., Terahertz Chiroptical Spectroscopy of an alpha-Helical Polypeptide: A Molecular Dynamics Simulation Study. *J. Phys. Chem. B* **2014**, *118* (45), 12837-12843.
35. Cho, M.; Fleming, G. R.; Saito, S.; Ohmine, I.; Stratt, R. M., Instantaneous Normal-Mode Analysis of Liquid Water. *J. Chem. Phys.* **1994**, *100* (9), 6672-6683.
36. Kindt, J. T.; Schmuttenmaer, C. A., Far-infrared absorption spectra of water, ammonia, and chloroform calculated from instantaneous normal mode theory. *J. Chem. Phys.* **1997**, *106* (11), 4389-4400.
37. Straub, J. E.; Choi, J. K., Extracting the Energy Barrier Distribution of a Disordered System from the Instantaneous Normal-Mode Density-Of-States - Applications to Peptides and Proteins. *Journal of Physical Chemistry* **1994**, *98* (42), 10978-10987.
38. Acbas, G.; Niessen, K. A.; Snell, E. H.; Markelz, A. G. Optical measurements of long-range protein vibrations *Nat Comm* [Online], 2014. <http://dx.doi.org/10.1038/ncomms4076> (accessed 2014).
39. Cerutti, D. S.; Le Trong, I.; Stenkamp, R. E.; Lybrand, T. P., Simulations of a Protein Crystal: Explicit Treatment of Crystallization Conditions Links Theory and Experiment in the Streptavidin-Biotin Complex. *Biochemistry* **2008**, *47* (46), 12065-12077.
40. Cerutti, D. S.; Case, D. A., Molecular dynamics simulations of macromolecular crystals. *Wiley Interdisciplinary Reviews-Computational Molecular Science* **2019**, *9* (4).
41. Meinhold, L.; Merzel, F.; Smith, J. C., Lattice dynamics of a protein crystal. *Phys. Rev. Lett.* **2007**, *99* (13).

---

Authors are required to submit a graphic entry for the Table of Contents (TOC) that, in conjunction with the manuscript title, should give the reader a representative idea of one of the following: A key structure, reaction, equation, concept, or theorem, etc., that is discussed in the manuscript. Consult the journal's Instructions for Authors for TOC graphic specifications.

Insert Table of Contents artwork here

