Site-selective dynamics of ligand-free and ligand-bound azidolysozyme

 Seyedeh Maryam Salehi and Markus Meuwly

COLLECTIONS

Paper published as part of the special topic on Time-resolved Vibrational Spectroscopy

View Online   Export Citation   CrossMark

ARTICLES YOU MAY BE INTERESTED IN

Site-selective dynamics of ligand-free and ligand-bound azidolysozyme

View Online   Export Citation   CrossMark

Seyedeh Maryam Salehi [ID] and Markus Meuwly[a] [ID]

AFFILIATIONS

Department of Chemistry, University of Basel, Klingelbergstrasse 80, CH-4056 Basel, Switzerland

Note: This paper is part of the JCP Special Topic on Time-Resolved Vibrational Spectroscopy.
[a]Author to whom correspondence should be addressed: m.meuwly@unibas.ch

ABSTRACT

Azido-modified alanine residues (AlaN$_3$) are environment-sensitive, minimally invasive infrared probes for the site-specific investigation of protein structure and dynamics. Here, the capability of the label is investigated to query whether or not a ligand is bound to the active site of lysozyme and how the spectroscopy and dynamics change upon ligand binding. The results demonstrate specific differences for center frequencies of the asymmetric azide stretch vibration, the longtime decay, and the static offset of the frequency fluctuation correlation function (FFCF)—all of which are experimental observables—between the ligand-free and the ligand-bound N$_3$-labeled protein. The center-frequency shifts range from 1 to 8 cm^{-1}, which is detectable from state-of-the art experiments. Similarly, the nonvanishing static component Δ_0 of the FFCF between ligand-free and ligand-bound protein can differ by up to a factor of 2.5. This makes the azide label a versatile and structurally sensitive probe to report on the dynamics of proteins in a variety of environments and for a range of different applications. Ligand-induced differences in the dynamics are also mapped onto changes in the local and through-space coupling between residues by virtue of dynamical cross correlation maps. This demonstrates that the position where the label is placed also influences the local and global protein motions.

INTRODUCTION

Proteins are essential for function and sustaining the life of organisms. Experimentation and computational studies have clarified that protein function involves both structure *and* dynamics.[1–3] However, characterizing structural and functional dynamics of proteins at the same time under physiological conditions in the condensed phase, which is prerequisite for understanding cellular processes at a molecular level, remains challenging.[1] Vibrational spectroscopy, in particular, two-dimensional infrared (2D IR) spectroscopy, has been shown to be a powerful tool for studying the structural dynamics of various biological systems.[4] One of the particular challenges is to obtain structural and environmental information in a site-specific manner. To address this, significant effort has been focused on the development and application of various infrared (IR) reporters[5,6] that absorb in the frequency range of 1700–2800 cm^{-1} to discriminate the signal from the strong protein background.[7,8] Such IR probes have provided valuable information about the structure and dynamics of complex systems. For example, nitrile probes have helped to clarify the

role of electrostatic fields in enzymatic reactions[9,10] or to elucidate the mode of drug binding to proteins.[11,12] Isotope edited carbonyl spectroscopy was used to characterize the mechanism of protein folding and amyloid formation[13,14] or the structure and function of membrane proteins.[15,16] Additional molecular groups, such as thiocyanate,[17] cyanamide,[18] sulfhydryl vibrations of cysteines,[19] deuterated carbons,[20] carbonyl vibrations of metal-carbonyls,[21–23] cyanophenylalanine,[24] and azidohomoalanine (Aha),[25] have also been explored.

In this work, AlaN$_3$, an analog of azidohomoalanine that has been shown to sensitively report on local structural changes while still being minimally invasive,[26,27] is used as the probe. This modification can be incorporated into proteins through established expression techniques.[28] The asymmetric stretch frequency of –N$_3$ is at ~2100 cm^{-1} and has a comparatively high extinction coefficient of 300–400 M^{-1} cm^{-1}, which makes it a suitable spectroscopic reporter.[25] Aha has been used for biomolecular recognition after incorporation into the peptide directly[25,26] or in the vicinity of the binding area of a PDZ2 domain[29] to detect the water-specific response of azide vibrations when attached to small

organic molecules[30] or to probe the frequency shift and fluctuation due to its sensitivity to the local electrostatic environments and dynamics.[27,31] Such studies confirm that AlaN$_3$ and/or Aha are environment-sensitive IR probes and suitable modifications for site-specific investigations of protein dynamics. It should be noted that actual experiments are carried out in D$_2$O rather than in H$_2$O to minimize overlap with the bend–libration combination band of H$_2$O that absorbs in the vicinity of the azide asymmetric stretch vibration.

With its picosecond time resolution, IR spectroscopy provides direct information about the structural dynamics around a probe molecule with high temporal resolution.[4,32] Moreover, introducing IR probes with isolated vibrational frequencies overcomes the problem of spectral congestion that complicates discrimination and analysis of desired vibrational bands. With that, the inter- and intramolecular coupling between degrees of freedom or the local structure or dynamics of biological systems can be specifically probed and characterized. Such an approach relies on the sensitivity of the probe to report on changes in the vibrational frequencies induced by alterations in the local electrostatic interactions in the vicinity of the probe.[24]

IR spectroscopy is a potentially advantageous technique to characterize ligand binding to proteins.[33,34] Its success depends, in part, on the notion that when a ligand binds to a protein, the frequency of an infrared active vibration shifts due to the different electric field in solution—often water—and in the protein binding site due to the Stark effect. Such an approach often requires the ligand to be modified, e.g., through the addition of a suitable label, such as –CN as in benzonitrile. The relationship between shifting the IR response of the label and the influence on the binding affinity has been explicitly assessed through computer simulations for benzonitrile in the active site of wild type (WT) and mutant lysozyme.[34]

Alternatively, the protein can be selectively modified by attaching a spectroscopic label at strategic positions so that the binding process and functional dynamics can be interrogated with the functional and unmodified ligand. This has the potential advantage that interactions between the ligand and the surrounding protein are unaltered. These interactions contribute the majority of the enthalpic part to the binding free energy and, therefore, directly affect the affinity of the ligand and its rate of unbinding. The dynamics of WT lysozyme without and with labeled alanine (AlaN$_3$) has been recently found to provide position-specific information about the spectroscopy and dynamics of the modification site.[27] The structure of the protein with the labeled Ala residues is shown in Fig. 1 together with the binding site lined by residues Leu84, Val87, Leu91, Leu99, Met102, Val111, Ala112, Phe114, Ser117, Leu118, Leu121, Leu133, and Phe153. Following Ref. 34, cyano-benzene (PhCN) was used as the ligand to allow direct comparison. The WT structure was used here (a) to compare directly with earlier results[27] and (b) because PhCN has a comparatively small binding free energy toward the WT protein ($\Delta G_{\text{bind}} = -0.5$ kcal/mol), which suggests that the interaction between the ligand and the protein is weak.[34] For the L99A mutant protein, $\Delta G_{\text{bind}} = -3.9$ kcal/mol for PhCN,[34] which compares with an experimentally determined value of ∼ − 3.5 kcal/mol for iodobenzene from isothermal titration calorimetry.[35]

In this work, changes in 1D- and 2D-IR signatures of the azido group attached to all alanine (Ala) residues of lysozyme upon
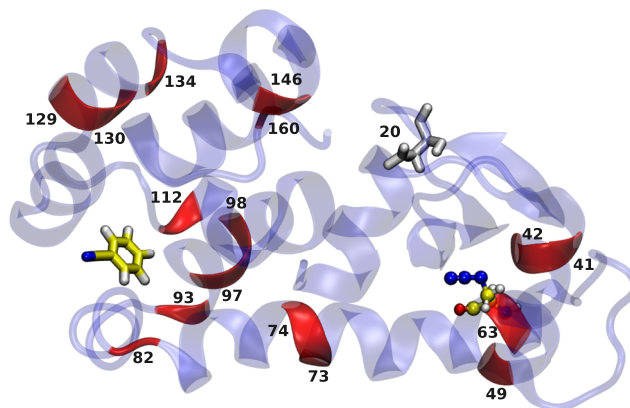

FIG. 1. The structure of lysozyme, including PhCN (licorice) in the active site and Ala63N$_3$ (CPK) as an example of attaching the –N$_3$ label. The Ala residues at positions 41, 42, 49, 63, 73, 74, 82, 93, 97, 98, 112, 129, 130, 134, 146, and 160 (red) are labeled one at a time. The rest of the protein is shown as blue, except for residue Asp20 that is in white licorice.

binding of cyano-benzene (PhCN) are determined. In addition, the changes of the environmental dynamics around all AlaN$_3$ are quantified for ligand-free vs ligand-bound lysozyme. Such differences are experimentally observable and yield valuable insight into the energetics and dynamics of protein–ligand binding. This paper is structured as follows: First, the methods are presented. This is followed by results and discussion of the structural dynamics and the spectroscopy and frequency fluctuation correlation functions (FFCFs). Finally, conclusions are drawn.

METHODS

Molecular dynamics simulations

MD simulations for the WT and all azide-labeled (AlaN$_3$) modified lysozymes were carried out using an adapted version of the Chemistry at Harvard Molecular Mechanics (CHARMM) program[36] with an interface to perform simulations with the reproducing kernel Hilbert space (RKHS) potential energy surface (PES).[37] The initial structure is that of T4 lysozyme with benzene in the cavity (PDB 1L83).[38] Benzene was replaced by cyano-benzene (PhCN) maintaining carbon atom positions.[34] The protein was solvated in explicit TIP3P water[39] using a cubic box of size $(78)^3$ Å^3. First, all systems were minimized, followed by heating and equilibration. Next, 2 ns NVE production simulations were carried out with and without the PhCN ligand present in the active site for all 16 protein variants with Ala replaced by AlaN$_3$. Bonds involving H-atoms were constrained using the SHAKE[40] algorithm, and all nonbonded interactions were evaluated with shifted interactions using a cutoff of 14 Å and switched at 10 Å.[41] Snapshots for analysis were recorded every 5 fs.

The energy function

The total energy function for the system consisted of the CHARMM force field[42] for the protein, the TIP3P[39] parameterization for water, and a reproducing kernel (RKHS)[43,44] based representation for the spectroscopic probe (–N$_3$).[27] For the –N$_3$ label, a

full-dimensional, accurate potential energy surface (PES) calculated at the pair natural orbital based coupled cluster [PNO-LCCSD(T)-F12/aVTZ][45,46] level is available. The RKHS-representation exactly reproduces the reference energies and provides analytical derivatives, which allows us to carry out high-accuracy, energy-conserving MD simulations.[37,44] This is of particular relevance for applications to vibrational spectroscopy. Details of the representation and validation of the full-dimensional PES are provided in Ref. 27. For the PhCN ligand, parameterization is based on Swissparam,[47] while charges are fitted to the electrostatic potential MP2/aug-cc-pVDZ level of theory.[34]

RESULTS AND DISCUSSION

Structural dynamics

The effect of ligand binding on the overall flexibility of the modified protein can be assessed from considering the root mean squared fluctuation (RMSF) of the C_α atoms. Depending on the position at which the $-N_3$ label is located, the changes in RMSF range from insignificant (Ala82N$_3$ or Ala160N$_3$) to major (Ala73N$_3$ or Ala112N$_3$) (see Figs. 2 and S1). Changes in flexibility can either be local (Ala82N$_3$ or Ala160N$_3$) or global (Ala73N$_3$ or Ala112N$_3$) similar to the RMSF variation. In addition, the average RMSF for the ligand-bound protein can be larger or smaller than that for the ligand-free protein. This suggests that insertion of the ligand can either rigidify the protein or increase its flexibility.

It is also of interest to compare all ligand-free $-N_3$-labeled RMSFs (blue traces in Figs. 2 and S1), which provides information about the influence of attaching the spectroscopic label on the

protein dynamics. As an example, flexibility at position Asp20 is, in general, high. The same applies to the region Ile100 to Met120. In general, attaching $-N_3$ to alanine residues along lysozyme leads to differential changes in the flexibility depending on the location of the modification site. Hence, although the azide is considered a "minimally invasive" probe, depending on the modification site the local or even global dynamics can still be affected.

For a more global assessment of the changes in dynamics and potential couplings between different parts of the protein induced by attaching the $-N_3$ label and/or introducing the PhCN ligand, dynamical cross-correlation maps[48,49] (DCCMs) were calculated from the trajectories using the Bio3D package.[50] Dynamic cross-correlation matrices are based on the expression

$$C_{ij} = \langle \Delta r_i \Delta r_j \rangle / \left(\langle \Delta r_i^2 \langle \Delta r_j^2 \rangle \rangle \right)^{1/2}, \tag{1}$$

where r_i and r_j are the C_α atom positions of the respective ith and jth amino acids and Δr_i corresponds to the displacement of the ith C_α atom from its trajectory-averaged position. DCCMs report on the correlated and anticorrelated motions within a protein, and difference DCCMs (ΔDCCM) provide a global view of the positionally resolved differences in the dynamics. In the following, only absolute values for C_{ij} and differences between them that are larger than 0.5 are reported. The DCCMs are symmetrical about the diagonal, and for clarity, positive correlations (for DCCM) or positive differences in C_{ij} (for ΔDCCM) are displayed in the lower right triangle and negative values or differences in C_{ij} are displayed in the upper left triangle.
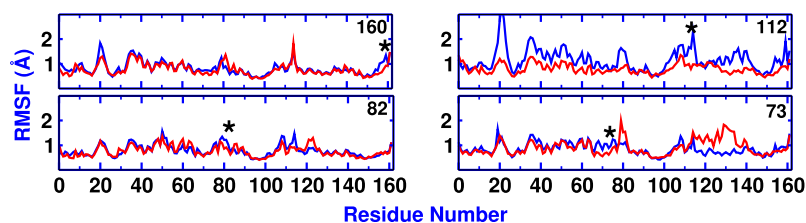
FIG. 2. RMSFs for the C_α atoms for ligand-free (blue) and ligand-bound (red) lysozyme with N_3^- attached to Ala73, Ala82, Ala112, and Ala160 residues. The label in each panel refers to the Ala residue number that carries the azide label, and the corresponding position of the residue is indicated with an asterisk above the RMSF trace.

FIG. 3. DCCM for ligand-free [panel (a)], ligand-bound [panel (b)], and ΔDCCM [panel (c)] between ligand-free and ligand-bound for Ala129N$_3$-PhCN. Positive correlations are in the lower right triangle, and negative correlations in the upper left triangle. Only correlation coefficients and differences between them (for ΔDCCM) with an absolute value greater than 0.5 are displayed.

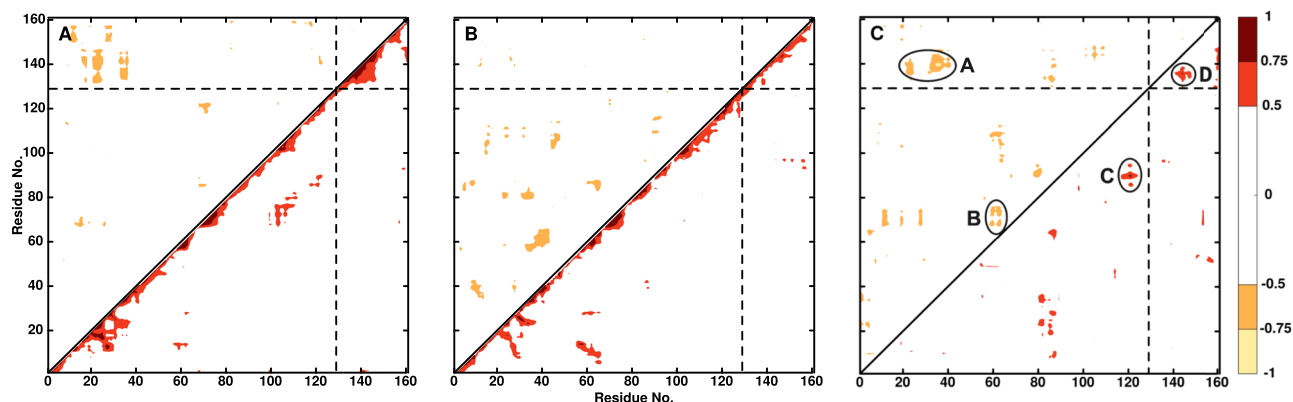The DCCM for lysozyme with Ala129N$_3$ ligand-free, ligand-bound, and the difference between the two is shown in Fig. 3. These maps reveal ligand-induced differences in the correlated and anti-correlated motions with appreciable amplitudes [see features A–D in Fig. 3(c)]. For ligand-free lysozyme, there are pronounced couplings between residues [130, 147] and [20, 25]/[32, 37] for anticorrelated motions and residues [68, 80] and [103, 112] for correlated motions. As demonstrated in Fig. 3(b), upon binding the ligand to lysozyme, the DCCM shows different coupled residues compared to the ligand-free protein. As an example, residues [35, 45] and [55, 68] are affected more for anticorrelated motions, whereas for correlated motions the coupling is between residues [5, 15] and [55, 65]. Note that these effects may not be visible in the ΔDCCM as the magnitude of the difference between the two systems may be below the threshold of 0.5 in the ΔC_{ij}.

In the difference map [Fig. 3(c)], feature A indicates coupling between residues [135, 145] and [20, 25]/[30, 42], whereas feature B refers to coupled residues [65, 75] and [58, 65]. Furthermore, feature C demonstrates prominent variations between residues [84, 95] and [117, 125], whereas for feature D residues [129, 140] and [140, 147] are strongly correlated. These findings suggest that residues couple both locally (feature B/D) and through space (feature A/C). It should also be pointed out that residues involved in features A–C are among those with higher RMSF (see Fig. S1). Interestingly, the region around residue Ala146 with larger differences ΔC_{ij} displays correlated dynamics with spatially close residues around residue Asp20 (white licorice in Fig. 1). On the other hand, the pronounced differences in the RMSF of Ala129N$_3$ (see Fig. S1) for residues [42, 57] do not show up in ΔDCCM because their C_{ij} coefficients are below the threshold of 0.5.

Difference DCCMs (ΔDCCM) between WT and ligand-bound (WT-PhCN) lysozyme are shown in Fig. 4(a) together with the difference map between WT and azido modified lysozyme at position 134 (Ala134N$_3$) [see Fig. 4(b)]. With the PhCN ligand bound to the protein, the ΔDCCM compared with that for the ligand-free protein is sparsely populated [see Fig. 4(a)]. This indicates that the conformational dynamics of the two systems is similar. Contrary to that,

a larger number of differences in the dynamics between WT and Ala134N$_3$ arise, as shown in Fig. 4(b). Finally, the ΔDCCM between the ligand-free (Ala134N$_3$) and ligand-bound (Ala134N$_3$-PhCN) labeled lysozyme at Ala134 shown in Fig. 4(c) demonstrates that the "contrast" further increases. The major difference in the conformational dynamics between the ligand-free and ligand-bound protein arises for coupled residues [57, 65] with [105, 120] (feature A), [103, 112] with [112, 122] (feature B), [60, 68] with [47, 52]/[55, 62] (feature C), and [57, 65] with [17, 25]/[32, 42] (feature D). Interestingly, as mentioned before for residue 129, residues involved in features A and B are also among those with higher RMSF (see Fig. S1). However, the DCCM and ΔDCCM provide considerably more information about changes in the average dynamics. Additional difference DCCMs (ΔDCCM) for the remaining AlaN$_3$ residues are shown in Figs. S2–S15.

Correlated motions in hen egg white lysozyme (HEWL) labeled with ruthenium dicarbonyl attached to His15 were also investigated from a normal mode analysis of the C$_\alpha$ atoms and based on an elastic network model.[21] This analysis also reported both correlated and anticorrelated motions for low-frequency normal modes as is found in the present case. Furthermore, such analyses help us to more quantitatively characterize the slaving motion of the protein to the surrounding solvent.[21]

Vibrational spectroscopy and frequency correlation functions

Using instantaneous normal mode (INM) analysis,[27,37,51] the frequency trajectory $\omega(t)$ of the asymmetric stretch vibration of the −N$_3$ label was determined. Based on this, the 1D infrared spectra corresponding to the azide asymmetric stretch vibration for each of the 16 AlaN$_3$ residues was computed for the ligand-free and ligand-bound protein [see Figs. 5(a), 5(b), and S16]. A direct comparison of the maximum position of the infrared line shape in Fig. 5(c) shows that for three N$_3$-modified alanine residues (Ala41, Ala98, and Ala130), the difference in the absorption frequency is insignificant. For positions Ala63, Ala73, and Ala160, the differences are 8, 3, and
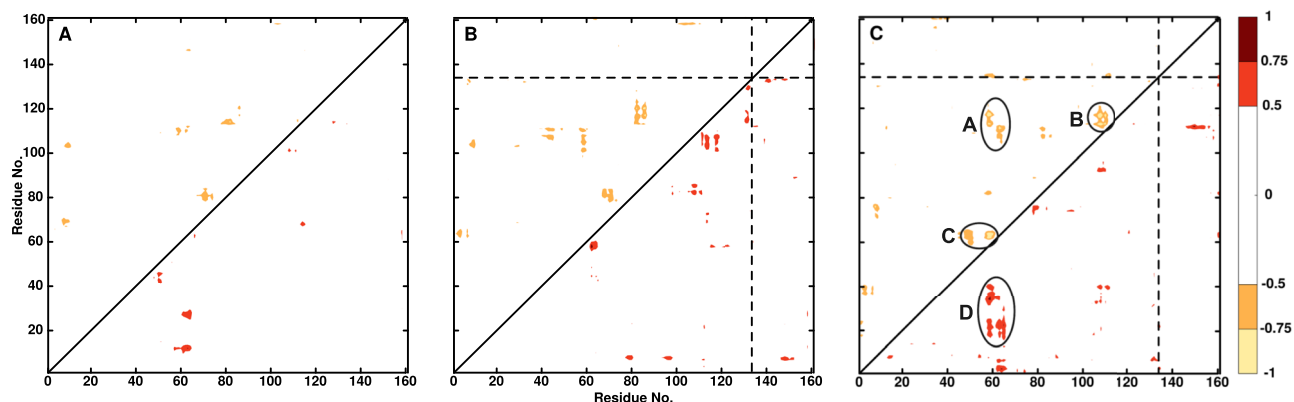

FIG. 4. Difference DCCMs (ΔDCCM) between WT and WT-PhCN [panel (a)], WT and Ala134N$_3$ [panel (b)], and Ala134N$_3$ and Ala134N$_3$-PhCN [panel (c)]. Positive correlations are in the lower right triangle, and negative correlations are in the upper left triangle. Only differences in correlation coefficients with an absolute value greater than 0.5 are displayed.
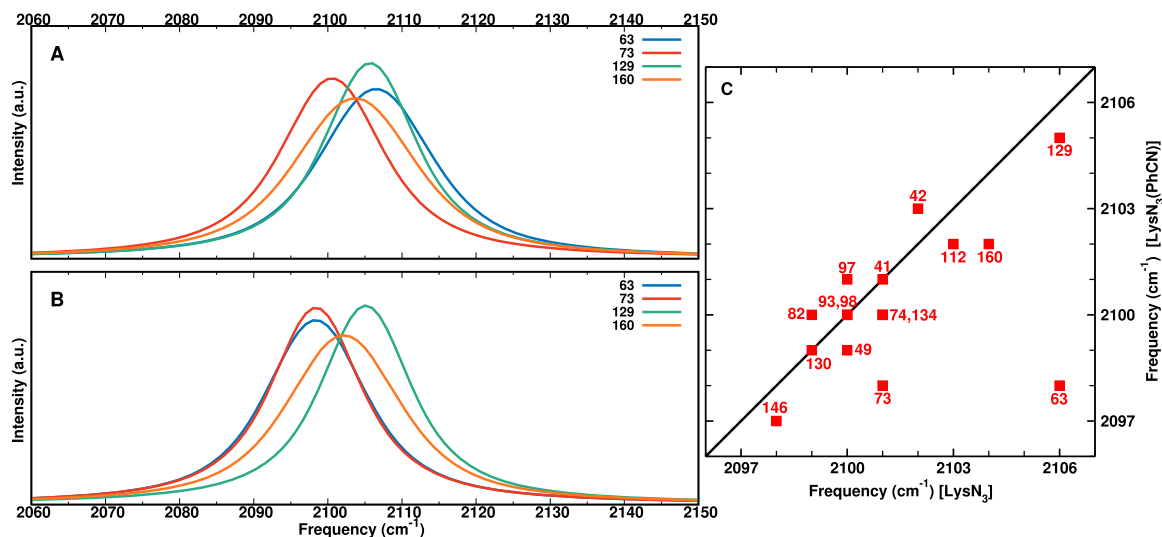
FIG. 5. 1D IR spectra from INM for four AlaN$_3$ residues (Ala63, Ala73, Ala129, and Ala160) for ligand-free [panel (a)] and ligand-bound [panel (b)] lysozyme. Panel (c) compares the maximum frequency of the 1D IR spectra for all modified Ala residues for ligand-free (along the x-axis) and ligand-bound (along the y−axis) N$_3$-labeled lysozyme.

2 cm^{-1}, respectively, whereas for the other residues the change is within 1 cm^{-1}.

Such frequency changes can be measured with state-of-the art experiments,[24] and their magnitude is also consistent with previous simulations of the vibrational Stark effect for the −CN probe in PhCN with red shifts of up to 3.5 cm^{-1} in going from the WT to the L99A and L99G mutants of T4-lysozyme.[34] Similarly, the 1D and 2D infrared spectroscopy of −CO as the label for the insulin monomer and dimer found[52] that the relative shifts of the spectroscopic response was correctly described, whereas the absolute frequencies may differ by some 10 cm^{-1}. In a recent study, such an approach found a splitting of 13 cm^{-1}, compared with 25 cm^{-1} from experiment, for the outer and central −CO labels

in cationic trialanine in water.[53] Hence, MD simulations together with instantaneous normal modes are a meaningful approach to determine relative frequency shifts, whereas capturing absolute frequencies in such simulations requires slight reparameterization of the underlying force field, e.g., through morphing techniques.[54,55]

The magnitude of frequency shifts found from the present simulations is also comparable with −3 cm^{-1} reported from experiments of the nitrile stretch in ligand IDD743 bound to WT vs V47N mutant hALR2[33] or a +6 cm^{-1} blue shift of the −CO vibrational frequency due to the binding of 19-NT to the Asp40Asn mutant of the protein ketosteroid isomerase compared to the WT.[56] Thus, differences of ~1 cm^{-1} for the frequency of the reporter in different chemical environments can be experimentally detected.[24]
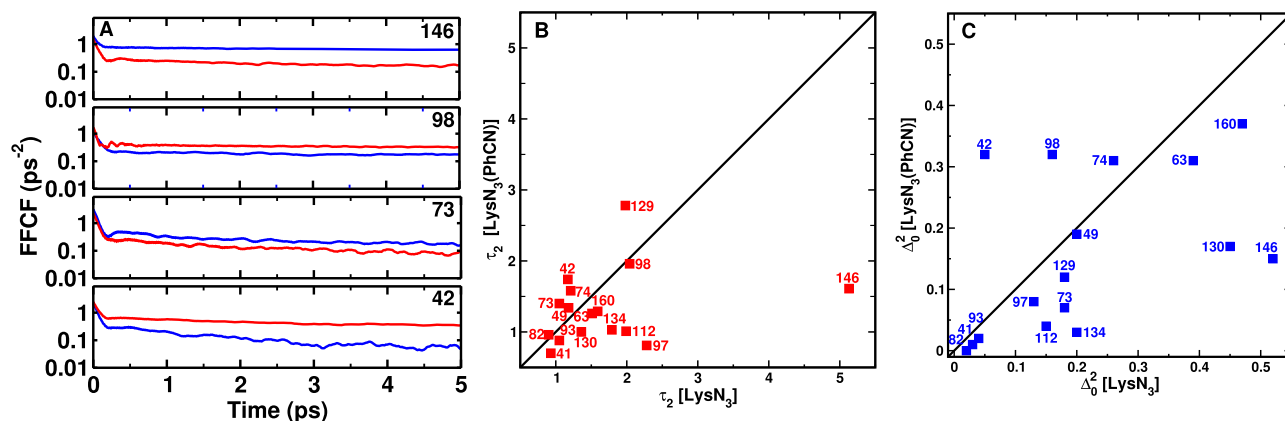
FIG. 6. [Panel (a)] FFCFs with pronounced differences from correlating the INM frequencies for ligand-free and ligand-bound Ala73, Ala146, Ala42, and Ala98 in lysozyme. The labels in each panel refer to the Ala residue that carries the azide label. Blue (ligand-free) and red (ligand-bound) traces are the fits to Eq. (2). The y−axis is logarithmic. Panels (b) and (c) compare τ_2 and Δ_0^2 for ligand-bound and ligand-free lysozyme, respectively.

From the frequency trajectories, the frequency fluctuation correlation function (FFCF) can be determined, which contains valuable information on relaxation time scales corresponding to the solvent dynamics around the solute. The FFCFs are fit to an empirical expression

$$\langle \delta\omega(t)\delta\omega(0)\rangle = a_1 \cos(\gamma t)e^{-t/\tau_1} + \sum_{i=2}^{n} a_i e^{-t/\tau_i} + \Delta_0^2, \quad (2)$$

which allows analytical integration to obtain the line shape function[57] using an automated curve fitting tool from the SciPy library.[58] As was found for the RMSFs and 1D IR spectra, the FFCFs from the simulations with and without the ligand bound to the protein can be very similar or differ appreciably [see Fig. 6(a)]. The slow decay time, τ_2, of the $-N_3$ asymmetric stretch mode of the

TABLE I. Parameters obtained from fitting the FFCF to Eq. (2) for INM frequencies for all different AlaN$_3$ residues in lysozyme. The average frequency $\langle\omega\rangle$ of the asymmetric stretch in cm^{-1}, the amplitudes a_1-a_3 in ps^{-2}, the decay times $\tau_1-\tau_3$ in ps, the parameter γ in ps^{-1}, and the static term Δ_0^2 in ps^{-2}.

Res.	$\langle\omega\rangle$	a_1	γ	τ_1	a_2	τ_2	Δ_0^2
			LysN$_3$				
41	2099.69	1.86	11.23	0.068	0.32	0.93	0.03
42	2100.53	1.76	10.21	0.069	0.33	1.17	0.05
49	2099.02	2.00	10.67	0.069	0.36	1.18	0.20
63	2105.49	2.05	11.71	0.069	0.37	1.51	0.39
73	2099.56	2.49	13.04	0.078	0.40	1.05	0.18
74	2099.59	1.94	12.80	0.072	0.52	1.21	0.26
82	2098.36	1.76	9.15	0.062	0.30	0.90	0.02
93	2099.00	1.92	8.88	0.065	0.27	1.05	0.04
97	2098.86	1.70	7.96	0.067	0.39	2.28	0.13
98	2099.47	1.15	0.0	0.057	0.06	2.04	0.16
112	2101.62	2.40	9.41	0.072	0.44	1.99	0.15
129	2104.69	2.83	18.08	0.066	0.09	1.98	0.18
130	2097.57	2.19	12.36	0.080	0.28	1.36	0.45
134	2100.17	1.87	11.55	0.074	0.29	1.79	0.20
146	2096.84	1.16	6.15	0.057	0.24	5.13	0.52
160	2102.67	2.64	10.20	0.065	0.45	1.59	0.47
			LysN$_3$–PhCN				
41	2099.90	1.77	11.78	0.068	0.38	0.70	0.01
42	2102.22	1.66	10.28	0.074	0.40	1.74	0.32
49	2098.31	1.56	10.52	0.072	0.23	1.34	0.19
63	2097.35	1.99	13.90	0.094	0.16	1.26	0.31
73	2097.38	1.94	8.92	0.067	0.22	1.40	0.07
74	2098.61	2.12	10.80	0.064	0.44	1.58	0.31
82	2098.62	1.76	9.10	0.063	0.28	0.96	0.02
93	2099.00	1.80	10.00	0.067	0.27	0.88	0.02
97	2099.65	1.46	10.88	0.064	0.31	0.81	0.08
98	2099.44	1.23	18.07	0.054	0.09	1.96	0.32
112	2100.91	1.79	11.01	0.074	0.32	1.01	0.04
129	2104.11	2.55	17.75	0.066	0.14	2.78	0.12
130	2098.13	1.64	13.09	0.090	0.13	1.00	0.17
134	2099.16	1.76	8.28	0.063	0.29	1.03	0.03
146	2096.41	1.17	10.61	0.068	0.15	1.61	0.15
160	2101.14	2.08	11.65	0.069	0.42	1.29	0.37

label is typically shorter for the ligand-bound protein compared to that without PhCN [see Fig. 6(b)], although exceptions exist. For Ala97N$_3$, Ala112N$_3$, and Ala134N$_3$, the slow relaxation time τ_2 is faster by 75% up to a factor of ~2.5, and for Ala146N$_3$, the slow time scale, τ_2, differs by a factor of ~3 between the ligand-free (τ_2 = 5.13 ps) and ligand-bound (τ_2 = 1.61 ps) lysozyme. For the other alanine residues, the τ_2 times between ligand-free and ligand-bound lysozyme are similar. As an exception, for Ala129N$_3$, the decay is slowed down by ~50% for PhCN-bound lysozyme. All FFCFs without (Fig. S17) and with (Fig. S18) the ligand are given in the supplementary material and the fitting parameters are reported in Table I.

As a last feature of the FFCF, it is found that the static component Δ_0 can differ appreciably between ligand-free and ligand-bound lysozyme [see Fig. 6(c)]. The static offset Δ_0 is an experimental observable and characterizes the structural heterogeneity around the modification site. There are only four alanine residues for which the static offset is similar (Ala41, Ala49, Ala82, and Ala93) for ligand-bound and ligand-free lysozyme. For all others, the differences range from 15% to a factor of ~3. As an example, for Ala73N$_3$ the difference for Δ_0^2 between bound and ligand-free lysozyme is a factor of ~2.5 (Δ_0^2 = 0.18 vs 0.07 ps^{-2} or Δ_0 = 0.42 vs Δ_0 = 0.26 ps^{-1}), and for Ala146N$_3$, they differ by a factor of ~3.5 (Δ_0^2 = 0.52 vs 0.15 ps^{-2}, i.e., Δ_0 = 0.72 vs Δ_0 = 0.39 ps^{-1}). Thus, the environmental dynamics around the spectroscopic label can be sufficiently perturbed by binding of a ligand in the protein active site to be reported directly as an experimentally accessible quantity with typical errors[24] between 0.1 and 0.3 cm^{-1} (~0.05 ps^{-1}). Hence, the differences found from the simulations are well outside the expected error bars from experiment.

Nonvanishing static components of the FFCF were also reported from experiments. For trialanine (Ala)$_3$, a value of Δ_0 = 5 cm^{-1} was reported,[59] compared with Δ_0 = 4.6 cm^{-1} from MD simulations (0.94 ps^{-1} vs 0.86 ps^{-1}) with multipolar force fields.[53] Similarly, CN$^-$ in water features a nonvanishing tilt angle by τ = 10 ps[60] with $\Delta_0 \sim 0.1$ ps^{-1} ~ 0.5 cm^{-1}.[61] Finally, 2D IR experiments for p-cyanophenylalanine bound to six distinct sites in a Src homology 3 domain reported static components ranging from Δ_0 = 1.0 to 3.7 cm^{-1} (corresponding to 0.19–0.70 ps^{-1}).[24]

Structural dynamics on longer time scales

Up to this point, the structural dynamics was characterized on the time scale required to converge the infrared spectroscopy within a given conformational substate. It was previously shown that IR spectra and FFCFs converge on the 1–2 ns time scale.[52] During this time, the average C$_\alpha$ root mean squared deviation (RMSD) for the modified lysozymes compared with the WT reference structure is 1.5 Å with a minimal value of 0.85 Å (for Ala63N$_3$) and a maximum value of 2.6 Å (for Ala42N$_3$). However, attaching an azide label to alanine residues may incur structural changes of the protein on longer time scales. For example, it is known that lysozyme samples open and closed conformations,[62] which occurs on the microsecond time scale.[63]

In order to probe and quantitatively assess dynamics on longer time scales, additional simulations were carried out for ligand-free WT and modified lysozyme with buried and solvent exposed labels.

The radial distribution functions $g(r)$ for water around the middle nitrogen atom (NB; see Fig. S19) of the azide label show that the attached N_3 is either solvent exposed (residues 41, 42, 49, 63, 73, 74, 82, 93, 97, 112, 134, and 160) or buried (residues 98, 129, 130, and 146). First, two independent 10 ns simulations were run for Ala82N_3 (solvent exposed), Ala98N_3 (buried), and the WT protein (for comparison) using the RKHS representation for the $-N_3$ label [see Fig. S19 for $g(r)$]. Because the RKHS PES is computationally more expensive than evaluating an empirical energy function, a second set of simulations was run with an empirical parameterization for the N–N bond and the N–N–N angle. The N–N bond potential was harmonic with $r_e = 1.14$ Å and $k_e = 877.4$ kcal/mol, while for the N–N–N angular potential the parameters were $\theta_e = 180°$ and $k_e = 46.7$ kcal mol^{-1} rad^{-2}; otherwise, the energy function remained unchanged. For these simulations, the OpenMM[64] implementation of CHARMM was used. With this setup, one 100 ns simulation was carried out for each of the three systems.

For all these simulations, the root mean squared deviation (RMSD) and the RMSF with respect to the initial structure after heating were determined (see Figs. S20 and S21). The top panel in Fig. S20 reports the RMSF for WT (black), Ala82N_3 (blue), and Ala98N_3 (red) from the 10 ns production run with the N_3-label described by the RKHS PES. The results show that the residues around [70–80] and [105–116] have higher fluctuation for Ala98N_3 compared to the other two systems. Moreover, Ala82N_3 generally has lower RMSFs for all residues, except for Thr142. The bottom panel of Fig. S20 shows the corresponding RMSDs. Ala82N_3 stabilizes at ~1.2 Å, whereas for Ala98N_3 it increases to ~2.8 Å and WT lysozyme at ~ 2.2 Å. This suggests that Ala82N_3 is structurally more stable than the other two systems on the 10 ns time scale.

Similarly, for the 100 ns simulations using the empirical energy function throughout, the bottom panel of Fig. S21 shows that the RMSDs for WT and Ala98N_3 stabilize at ~5.0 and ~3.5 Å, respectively, whereas Ala82N_3 stabilizes at ~2.5 Å. The RMSF results for 100 ns confirm that Ala82N_3 still fluctuates least compared with the other two systems. For some of the residues, such as [32–45], [65–83], [102–130], and [139–143], the WT lysozyme has larger fluctuations compared to the other two systems. For comparison, it is of interest to note that the RMSD from a 50 μs simulation for the M6I mutant of lysozyme[63] reached between 4 and 5 Å, and the experimentally reported[62] RMSD between the open and closed structure for the I3P mutant is ~5 Å, all of which is consistent with the present findings.

Furthermore, the DCCMs for the WT and the two modified lysozymes were determined for the two sets of simulations (see Figs. S22–S24). For the WT protein, the DCCM from 10 to 100 ns simulations are comparable with some difference between residues [30–42] and [17–25] together with a few additional smaller features (see Fig. S22). For Ala82N_3, the DCCM from the 10 ns simulation using the RKHS PES for the label shows reduced anticorrelations but is overall similar to that from the 100 ns simulation using the conventional FF. On the other hand, for Ala98N_3, a pronounced coupling between residues [61–75] and [23–26] emerges (see Fig. S25), which was not found from simulations on the 10 ns time scale. In addition, the anticorrelations for some of the residues are less pronounced on the 100 ns time scale.

Finally, it is also of interest to determine the number of water oxygen atoms as a function of simulation time within a given cutoff (here 4 Å was chosen) of the middle nitrogen atom NB of the azide label for Ala82N_3 and Ala98N_3 (see Fig. S26). This confirms that throughout the 2 ns, the azide label is hydrated for Ala82N_3 and often "dry" for Ala98N_3. This is also consistent with the radial distribution functions (see Fig. S19). The findings from the simulations with the empirical force field for the entire system on the 100 ns time scale are similar: NB of Ala82N_3 is hydrated throughout the simulation, whereas for Ala98N_3 the middle nitrogen atom of the label is usually "dry" or is occasionally partially hydrated (see Fig. S27).

CONCLUSIONS

In summary, this work demonstrates that the 1D and 2D IR spectroscopy of the azide bound to alanine residues within one conformational substate for WT lysozyme provides valuable site-specific and temporal information about ligand binding of PhCN to the active site of WT lysozyme. The static component Δ_0 of the FFCF, which is an experimentally accessible observable, shows pronounced differences between the ligand-bound and ligand-free protein and can serve as a useful indicator for ligand binding. Changes in the maximum of the infrared absorbance are of the order of one to several cm^{-1}, which can be detected with state-of-the-art experiments.[24] The contrast between ligand-free and ligand-bound lysozyme increases when the azido-label is present, as demonstrated for Ala134N_3. The structural dynamics of the modified lysozymes depends on the position of the alanine residue along the chain. Given that even within the same conformational substate the spectroscopy and dynamics depend on the position of the modification, it is expected that when sampling a wider distribution of conformations, differences in the dynamics and spectroscopy persist. However, quantifying this is outside the scope of this work. Finally, this work also lays the foundations to investigate the binding of biologically relevant substrates, such as N-acetyl-D-glucosamine (NAG) that is hydrolyzed at the β-(1,4)-glycosidic bond by hen egg white lysozyme in Gram-positive bacteria.[65]

SUPPLEMENTARY MATERIAL

See the supplementary material for Figs. S1–S27.

ACKNOWLEDGMENTS

AUTHOR DECLARATIONS

Conflict of Interest

The authors declare no conflict of interest.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] J. M. Plitzko, B. Schuler, and P. Selenko, "Structural Biology outside the box—Inside the cell," Curr. Opin. Struct. Biol. **46**, 110–121 (2017).

[2] J. Guo and H.-X. Zhou, "Protein allostery and conformational dynamics," Chem. Rev. **116**, 6503–6515 (2016).

[3] S. Lu, X. He, D. Ni, and J. Zhang, "Allosteric modulator discovery: From serendipity to structure-based design," J. Med. Chem. **62**, 6405–6421 (2019).

[4] P. Hamm and M. Zanni, *Concepts and Methods of 2D Infrared Spectroscopy* (Cambridge University Press, New York, 2011).

[5] J. Ma, I. M. Pazos, W. Zhang, R. M. Culik, and F. Gai, "Site-specific infrared probes of proteins," Annu. Rev. Phys. Chem. **66**, 357–377 (2015).

[6] H. Kim and M. Cho, "Infrared probes for studying the structure and dynamics of biomolecules," Chem. Rev. **113**, 5817–5847 (2013).

[7] M. M. Waegele, R. M. Culik, and F. Gai, "Site-specific spectroscopic reporters of the local electric field, hydration, structure, and dynamics of biomolecules," J. Phys. Chem. Lett. **2**, 2598–2609 (2011).

[8] K. L. Koziol, P. J. M. Johnson, B. Stucki-Buchli, S. A. Waldauer, and P. Hamm, "Fast infrared spectroscopy of protein dynamics: Advancing sensitivity and selectivity," Curr. Opin. Struct. Biol. **34**, 1–6 (2015).

[9] N. M. Levinson and S. G. Boxer, "A conserved water-mediated hydrogen bond network defines bosutinib's kinase selectivity," Nat. Chem. Biol. **10**, 127–132 (2014).

[10] J. P. Layfield and S. Hammes-Schiffer, "Calculation of vibrational shifts of nitrile probes in the active site of ketosteroid isomerase upon ligand binding," J. Am. Chem. Soc. **135**, 717–725 (2013).

[11] D. G. Kuroda, J. D. Bauman, J. R. Challa, D. Patel, T. Troxler, K. Das, E. Arnold, and R. M. Hochstrasser, "Snapshot of the equilibrium dynamics of a drug bound to HIV-1 reverse transcriptase," Nat. Chem. **5**, 174–181 (2013).

[12] J. Liu, J. Strzalka, A. Tronin, J. S. Johansson, and J. K. Blasie, "Mechanism of interaction between the general anesthetic halothane and a model ion channel protein, II: Fluorescence and vibrational spectroscopy using a cyanophenylalanine probe," Biophys. J. **96**, 4176–4187 (2009).

[13] S. H. Brewer, B. Song, D. P. Raleigh, and R. B. Dyer, "Residue specific resolution of protein folding dynamics using isotope-edited infrared temperature jump spectroscopy," Biochemistry **46**, 3279–3285 (2007).

[14] C. T. Middleton, P. Marek, P. Cao, C.-c. Chiu, S. Singh, A. M. Woys, J. J. de Pablo, D. P. Raleigh, and M. T. Zanni, "Two-dimensional infrared spectroscopy reveals the complex behaviour of an amyloid fibril inhibitor," Nat. Chem. **4**, 355–360 (2012).

[15] A. Ghosh, J. Qiu, W. F. DeGrado, and R. M. Hochstrasser, "Tidal surge in the M2 proton channel, sensed by 2D IR spectroscopy," Proc. Natl. Acad. Sci. U. S. A. **108**, 6115–6120 (2011).

[16] A. Remorino, I. V. Korendovych, Y. Wu, W. F. DeGrado, and R. M. Hochstrasser, "Residue-specific vibrational echoes yield 3D structures of a transmembrane helix dimer," Science **332**, 1206–1209 (2011).

[17] L. J. G. W. van Wilderen, D. Kern-Michler, H. M. Müller-Werkmeister, and J. Bredenbeck, "Vibrational dynamics and solvatochromism of the label SCN in various solvents and hemoglobin by time dependent IR and 2D-IR spectroscopy," Phys. Chem. Chem. Phys. **16**, 19643–19653 (2014).

[18] G. Lee, D. Kossowska, J. Lim, S. Kim, H. Han, K. Kwak, and M. Cho, "Cyanamide as an infrared reporter: Comparison of vibrational properties between nitriles bonded to N and C atoms," J. Phys. Chem. B **122**, 4035–4044 (2018).

[19] M. Kozinski, S. Garrett-Roe, and P. Hamm, "2D-IR spectroscopy of the sulfhydryl band of cysteines in the hydrophobic core of proteins," J. Phys. Chem. B **112**, 7645–7650 (2008).

[20] J. Zimmermann, M. C. Thielges, W. Yu, P. E. Dawson, and F. E. Romesberg, "Carbon–deuterium bonds as site-specific and nonperturbative probes for time-resolved studies of protein dynamics and folding," J. Phys. Chem. Lett. **2**, 412–416 (2011).

[21] J. T. King and K. J. Kubarych, "Site-specific coupling of hydration water and protein flexibility studied in solution with ultrafast 2D-IR spectroscopy," J. Am. Chem. Soc. **134**, 18705–18712 (2012).

[22] J. T. King, E. J. Arthur, C. L. Brooks III, and K. J. Kubarych, "Site-specific hydration dynamics of globular proteins and the role of constrained water in solvent exchange with amphiphilic cosolvents," J. Phys. Chem. B **116**, 5604–5611 (2012).

[23] J. T. King, E. J. Arthur, C. L. Brooks III, and K. J. Kubarych, "Crowding induced collective hydration of biological macromolecules over extended distances," J. Am. Chem. Soc. **136**, 188–194 (2014).

[24] S. Ramos, R. E. Horness, J. A. Collins, D. Haak, and M. C. Thielges, "Site-specific 2D IR spectroscopy: A general approach for the characterization of protein dynamics with high spatial and temporal resolution," Phys. Chem. Chem. Phys. **21**, 780–788 (2019).

[25] R. Bloem, K. Koziol, S. A. Waldauer, B. Buchli, R. Walser, B. Samatanga, I. Jelesarov, and P. Hamm, "Ligand binding studied by 2D IR spectroscopy using the azidohomoalanine label," J. Phys. Chem. B **116**, 13705–13712 (2012).

[26] P. J. M. Johnson, K. L. Koziol, and P. Hamm, "Quantifying biomolecular recognition with site-specific 2D infrared probes," J. Phys. Chem. Lett. **8**, 2280–2284 (2017).

[27] S. M. Salehi and M. Meuwly, "Site-selective dynamics of azidolysozyme," J. Chem. Phys. **154**, 165101 (2021).

[28] K. L. Kiick, E. Saxon, D. A. Tirrell, and C. R. Bertozzi, "Incorporation of azides into recombinant proteins for chemoselective modification by the Staudinger ligation," Proc. Natl. Acad. Sci. U. S. A. **99**, 19–24 (2002).

[29] C. Zanobini, O. Bozovic, B. Jankovic, K. L. Koziol, P. J. M. Johnson, P. Hamm, A. Gulzar, S. Wolf, and G. Stock, "Azidohomoalanine: A minimally invasive, versatile, and sensitive infrared label in proteins to study ligand binding," J. Phys. Chem. B **122**, 10118–10125 (2018).

[30] M. P. Wolfshorndl, R. Baskin, I. Dhawan, and C. H. Londergan, "Covalently bound azido groups are very specific water sensors, even in hydrogen-bonding environments," J. Phys. Chem. B **116**, 1172–1179 (2012).

[31] K.-I. Oh, J.-H. Lee, C. Joo, H. Han, and M. Cho, "β-azidoalanine as an IR probe: Application to amyloid Aβ(16-22) aggregation," J. Phys. Chem. B **112**, 10352–10357 (2008).

[32] A. L. Le Sueur, R. N. Schaugaard, M.-H. Baik, and M. C. Thielges, "Methionine ligand interaction in a blue copper protein characterized by site-selective infrared spectroscopy," J. Am. Chem. Soc. **138**, 7187–7193 (2016).

[33] I. T. Suydam, C. D. Snow, V. S. Pande, and S. G. Boxer, "Electric fields at the active site of an enzyme: Direct comparison of experiment with theory," Science **313**, 200–204 (2006).

[34] P. Mondal and M. Meuwly, "Vibrational Stark spectroscopy for assessing ligand-binding strengths in a protein," Phys. Chem. Chem. Phys. **19**, 16131–16143 (2017).

[35] L. Liu, W. A. Baase, and B. W. Matthews, "Halogenated benzenes bound within a non-polar cavity in T4 lysozyme provide examples of I···S and I···Se halogen-bonding," J. Mol. Biol. **385**, 595–605 (2009).

[36] B. R. Brooks, C. L. Brooks III, A. D. MacKerell, Jr., L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch *et al.*, "CHARMM: The biomolecular simulation program," J. Comput. Chem. **30**, 1545–1614 (2009).

[37] S. M. Salehi, D. Koner, and M. Meuwly, "Vibrational spectroscopy of N_3^- in the gas and condensed phase," J. Phys. Chem. B **123**, 3282–3290 (2019).

[38] A. E. Eriksson, W. A. Baase, J. A. Wozniak, and B. W. Matthews, "A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene," Nature **355**, 371–373 (1992).

[39] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," J. Chem. Phys. **79**, 926–935 (1983).

[40] W. F. van Gunsteren and H. J. C. Berendsen, "Algorithms for macromolecular dynamics and constraint dynamics," Mol. Phys. **34**, 1311–1327 (1977).

[41] P. J. Steinbach and B. R. Brooks, "New spherical-cutoff methods for long-range forces in macromolecular simulation," J. Comput. Chem. **15**, 667–683 (1994).

[42] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha *et al.*, "All-atom empirical potential for molecular modeling and dynamics studies of proteins," J. Phys. Chem. B **102**, 3586–3616 (1998).

[43] T. S. Ho and H. Rabitz, "A general method for constructing multidimensional molecular potential energy surfaces from *ab initio* calculations," J. Chem. Phys. **104**, 2584–2597 (1996).

[44] O. T. Unke and M. Meuwly, "Toolkit for the construction of reproducing kernel-based representations of data: Application to multidimensional potential energy surfaces," J. Chem. Inf. Model. **57**, 1923–1931 (2017).

[45] M. Schwilk, Q. Ma, C. Köppl, and H.-J. Werner, "Scalable electron correlation methods. 3. Efficient and accurate parallel local coupled cluster with pair natural orbitals (PNO-LCCSD)," J. Chem. Theory Comput. **13**, 3650–3675 (2017).

[46] Q. Ma, M. Schwilk, C. Köppl, and H.-J. Werner, "Correction to scalable electron correlation methods. 4. Parallel explicitly correlated local coupled cluster with pair natural orbitals (PNO-LCCSD-F12)," J. Chem. Theory Comput. **14**, 6750 (2018).

[47] V. Zoete, M. A. Cuendet, A. Grosdidier, and O. Michielin, "SwissParam: A fast force field generation tool for small organic molecules," J. Comput. Chem. **32**, 2359–2368 (2011).

[48] T. Ichiye and M. Karplus, "Collective motions in proteins: A covariance analysis of atomic fluctuation in molecular-dynamics and normal mode simulations," Proteins: Struct., Funct., Genet. **11**, 205–217 (1991).

[49] G. E. Arnold and R. L. Ornstein, "Molecular dynamics study of time-correlated protein domain motions and molecular flexibility: Cytochrome P450BM-3," Biophys. J. **73**, 1147–1159 (1997).

[50] B. J. Grant, A. P. C. Rodrigues, K. M. ElSawy, J. A. McCammon, and L. S. D. Caves, "Bio3d: An R package for the comparative analysis of protein structures," Bioinformatics **22**, 2695–2696 (2006).

[51] M. Cho, G. R. Fleming, S. Saito, I. Ohmine, and R. M. Stratt, "Instantaneous normal mode analysis of liquid water," J. Chem. Phys. **100**, 6672–6683 (1994).

[52] S. M. Salehi, D. Koner, and M. Meuwly, "Dynamics and infrared spectroscopy of monomeric and dimeric wild type and mutant insulin," J. Phys. Chem. B **124**, 11882–11894 (2020).

[53] P. Mondal, P.-A. Cazade, A. K. Das, T. Bereau, and M. Meuwly, "Multipolar force fields for amide-I spectroscopy from conformational dynamics of the alanine trimer," J. Phys. Chem. B **125**, 10928–10938 (2021).

[54] M. Meuwly and J. M. Hutson, "Morphing *ab initio* potentials: A systematic study of Ne–HF," J. Chem. Phys. **110**, 8338–8347 (1999).

[55] J. M. Bowman and B. Gazdy, "A simple method to adjust potential energy surfaces: Application to HCO," J. Chem. Phys. **94**, 816–817 (1991).

[56] S. D. Fried, S. Bagchi, and S. G. Boxer, "Extreme electric fields power catalysis in the active site of ketosteroid isomerase," Science **346**, 1510–1514 (2014).

[57] K. Moller, R. Rey, and J. Hynes, "Hydrogen bond dynamics in water and ultrafast infrared spectroscopy: A theoretical study," J. Phys. Chem. A **108**, 1275–1289 (2004).

[58] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright *et al.*, "SciPy 1.0: Fundamental algorithms for scientific computing in Python," Nat. Methods **17**, 261–272 (2020).

[59] S. Woutersen, R. Pfister, P. Hamm, Y. Mu, D. S. Kosov, and G. Stock, "Peptide conformational heterogeneity revealed from nonlinear vibrational spectroscopy and molecular-dynamics simulations," J. Chem. Phys. **117**, 6833–6840 (2002).

[60] M. Koziński, S. Garrett-Roe, and P. Hamm, "Vibrational spectral diffusion of CN^- in water," Chem. Phys. **341**, 5–10 (2007).

[61] M. W. Lee, J. K. Carr, M. Göllner, P. Hamm, and M. Meuwly, "2D IR spectra of cyanide in water investigated by molecular dynamics simulations," J. Chem. Phys. **139**, 054506 (2013).

[62] M. M. Dixon, H. Nicholson, L. Shewchuk, W. A. Baase, and B. W. Matthews, "Structure of a hinge-bending bacteriophage T4 lysozyme mutant, Ile3 → Pro," J. Mol. Biol. **227**, 917–933 (1992).

[63] M. Ernst, S. Wolf, and G. Stock, "Identification and validation of reaction coordinates describing protein functional motion: Hierarchical dynamics of T4 lysozyme," J. Chem. Theory Comput. **13**, 5076–5088 (2017).

[64] P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L. P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern *et al.*, "OpenMM 7: Rapid development of high performance algorithms for molecular dynamics," PLoS Comput. Biol. **13**, e1005659 (2017).

[65] C. C. Blake, L. N. Johnson, G. A. Mair, A. C. North, D. C. Phillips, and V. R. Sarma, "Crystallographic studies of the activity of hen egg-white lysozyme," Proc. R. Soc. London, Ser. B **167**, 378–388 (1967).