Quantifying the Effects of Hydrogen Bonding on Nitrile Frequencies in GFP: Beyond Solvent Exposure

Jeremy T. First, Joshua D. Slocum, and Lauren J. Webb*

Department of Chemistry, Texas Materials Institute, and Institute for Cell and Molecular Biology, The University of Texas at Austin, 10SE 24th Street, STOP A5300, Austin, Texas 78712-1224, United States

ABSTRACT: Vibrational spectroscopy is a powerful tool for characterizing the complex noncovalent interactions that arise in biological systems. The nitrile stretching frequency has proven to be a particularly convenient biological probe, but the interpretation of nitrile spectroscopy is complicated by its sensitivity to local hydrogen bonding interactions. This often inhibits the straightforward interpretation of nitrile spectra by requiring knowledge of the molecular-level details of the local environment surrounding the probe. While the effect of hydrogen bonds on nitrile frequencies has been well-documented for small molecules in solution, there have been relatively few studies of these effects in a complex protein system. To address this, we introduced a nitrile probe at nine locations throughout green fluorescent protein (GFP) and compared the mean vibrational frequency of each probe to the specific hydrogen bonding geometries observed in molecular dynamics (MD) simulations. We show that a continuum of hydrogen bonding angles is found depending on the particular location of each nitrile, and that the differences in these angles account for the differences in the measured nitrile frequency. We further observed that the temperature dependence of the nitrile frequencies (measured as a frequency temperature line slope, FTLS) was a good indicator of the hydrogen bonding interactions of the probe, even in scenarios where the nitrile was involved in complex and restricted hydrogen bonds, both from protein and from water. While constant offsets to the nitrile frequency to all hydrogen bonding environments have been applied before to interpret shifts in nitrile frequency, we show that this is insufficient in systems where the hydrogen bonds may be restricted by the surrounding medium. However, the strength of the observed correlation between nitrile frequency and hydrogen bonding angle suggests that it may be possible to disentangle electrostatic effects and effects of the orientation of hydrogen bonding on the nitrile stretching frequency. Meanwhile, the experimental measurement of the FTLS of the nitrile is an excellent tool to identify changes in the hydrogen bonding interactions for a particular probe.

INTRODUCTION

The complex networks of noncovalent interactions that arise from the arrangement of thousands of partially charged atoms in biological molecules are responsible for many important biomolecular functions, including protein folding, protein–protein interactions, membrane fluidity and permeation, ligand binding, signaling pathways, and catalysis. Characterizing and understanding the molecular details of the environments that give rise to biological function is of critical importance to modern biophysics. The inherent spatial and temporal resolution of molecular vibrations has been exploited by many researchers to directly probe biological systems. Specifically, the nitrile stretching frequency has been widely used as a probe of site-specific dynamics,\(^5\,^6\) hydration,\(^7\,^8\) and electric fields\(^9\,^10\) in a variety of biological systems due to its environmental sensitivity. Nitrile vibrational probes absorb in a relatively transparent region of the protein’s infrared spectrum, have reasonable oscillator strengths,\(^11\) and can be easily incorporated through a variety of synthetic and biosynthetic methods\(^12\,^13\) to interrogate DNA, proteins, and membranes.

Despite the widespread use of nitriles as biomolecular probes and the relative ease with which their absorption energies can be measured in a variety of applications, it is not trivial to meaningfully interpret a measured nitrile frequency in a complex biological system. This is largely due to the convolution of the competing effects of Coulombic frequency shifts, commonly referred to as the vibrational Stark effect (VSE), and frequency shifts due to specific interactions such as hydrogen bonding. While the effects of solvent electric field and specific hydrogen bonding interactions on nitrile frequencies have been well-characterized for small molecules in solution,\(^14\,^15\) the theoretical basis for these convoluted effects does not always successfully translate to the interpretation of nitrile frequencies in more complex systems like a solvated protein. Indeed, the ability to interpret nitrile frequencies in terms of electric fields via a VSE analysis when the nitrile experiences hydrogen bonds has been under...
question for quite some time.6,19–21 While it is clear that VSE analyses of nitrile frequencies can be used to quantify relative electric fields between nitriles that are engaged in similar hydrogen bonding interactions, there is still debate on how to define criteria that must be met in order for two nitriles to be classified within this regime. This is further complicated by a lack of experimental methodologies to specifically test in situ if these hydrogen bonding criteria are met. Conversely, the use of nitrile frequencies as site-specific probes of solvent exposure inherently ignores any electrostatic contribution to the frequency by equating blue-shifted spectra with solvent exposed locations and red-shifted spectra with buried locations. The possibility that electric field effects would red-shift the spectrum of a solvent exposed nitrile or blue-shift the spectrum of a buried nitrile is usually not considered.

Significant theoretical effort has been spent toward developing quantitative models that can capture the competing effects of hydrogen bonding and Coulombic forces on nitrile frequencies. Early theories focused on relating bulk properties and those that can disentangle the contributions of hydrogen bonding from other interactions that are engaged in similar hydrogen bonding geometries, rates of diffusion, and interaction strength. Each of these different hydrogen bonding environments could contribute to a frequency shift of differing magnitude in a way not described by the VSE. Here, we use a combination of FTLS measurements and MD simulations to show that the FTLS is related to small differences in hydrogen bonding profiles and is therefore useful beyond just the Boolean classification of a nitrile as either engaging in hydrogen bonding or free from hydrogen bonding. Instead, the FTLS can be used to differentiate a range of solvent exposure and even microscopic differences in specific hydrogen bonding angles. This makes the FTLS a powerful diagnostic tool for interpreting any nitrile spectrum, as it allows for the rapid comparison of hydrogen bonding interactions of different states.

To evaluate the nitrile FTLS as a diagnostic tool for a hydrogen bonding environment, we biosynthetically incorporated p-cyanophenylalanine (CNF) at nine locations in GFP. The locations were chosen on the basis of the PDB crystal structure 2b3p to sample a broad range of solvent accessibility and hydrogen bonding environments. Using the crystal structure, we classified a priori the locations as solvent exposed or buried on the basis of the crude assessment of whether or not the side chain was pointed into the interior of the β-barrel. The nine locations chosen are shown in Figure 1, where the colored spheres indicate a residue that was independently mutated to CNF. We collected temperature-
amino acid code of the residue that was replaced by CNF, denoted by X. The first letter indicates the one-letter amino acid code of the residue that was replaced by CNF, denoted by X.

**EXPERIMENTAL SECTION**

**Protein Expression and Purification.** A pBAD vector containing the gene for 6xHis-tagged superfolder GFP (Addgene plasmid 85482) and a pDule vector containing a tRNA synthetase/tRNA pair evolved for CNF (Addgene plasmid 85494) were kindly provided by the Mehl group.31,32 Amber codons were introduced into the superfolder GFP (hereafter, simply GFP) gene at the desired locations, and the amber-coded pBAD vector was cotransformed alongside the pDule vector into DH10β cells. Single colonies from plates containing ampicillin and tetracycline were used to seed the growth of 1 L cultures in an autoinduction media described by Mehl et al.33 After 24–30 h of expression, cells were harvested and either purified immediately or stored at −80 °C until further use. 6xHis-tagged GFP mutants containing CNF were purified by immobilized metal affinity chromatography, and the 6xHis tags were cleaved with trypsin proteolysis. Cleaved GFP mutants were buffer exchanged into degassed phosphate buffered saline (PBS) at pH 7.4 and used immediately or stored at −80 °C.

**FTIR Spectroscopy.** Nitrile-containing GFP mutants were concentrated to ~2 mM using Amicon Ultra 10 000 molecular weight cutoff filters according to the manufacturer's protocol. Following concentration, samples were spun for 5 min at 15 000 rpm to remove any air bubbles. Samples were injected into a temperature-controlled cell at 5 °C between two sapphire windows separated by two 100 μm Teflon spacers. Spectra were recorded at 5, 15, 25, and 35 °C, in that order, averaging over 600 scans at a resolution of 0.5 cm⁻¹, allowing the sample to equilibrate for at least 10 min before scanning at each temperature. At the end of the 35 °C measurement, the intact cell was cleaned with water and ethanol, and then PBS was added to the 35 °C equilibrated cell for background spectra, which were collected in the reverse order of temperature to avoid the formation of air bubbles during the scans.

**Molecular Dynamics Simulations.** A template protein was modeled by homology from the 2b3p crystal structure30 using a procedure described elsewhere.33 Using this template, nine separate mutations from the WT residue to phenylalanine were made using the mutagenesis wizard tool in PyMol34 at each of the CNF locations. Using the Avogadro molecular editing package,35 the mutated phenylalanine residue was further modified to CNF and independently minimized. Parameters for the internal GFP fluorophore were kindly provided by Riccardo Nifosi,36 and parameters for the CNF chromophore have been described previously.34 All further minimizations and MD simulations were performed using the Gromacs 5.0.4 molecular dynamics simulation package,36,37 and the Amber03 force field (Amber03).38,39

Each of the nine protein systems was energy minimized in vacuum using the steepest descent algorithm and then solvated in a dodecahedron box of TIP3P water,40 with a minimum distance of protein to the edge of the box of 1.5 nm. The solvated systems were then energy minimized using a steepest descent approach, and then heated under the NVT ensemble at 300 K, and equilibrated under the NPT ensemble at 1 atm. Production MD was run on each equilibrated protein system for 50 ns, using the same simulation parameters as described previously,34 and a snapshot was saved every 4 ps. The first 10 ns was discarded as equilibration time, and only the last 40 ns was used in the analyses that follow.

**RESULTS**

**Vibrational Absorption Measurements.** We measured the nitrile stretching frequencies of biosynthetically incorpo- rated CNF probes at nine positions in GFP, as well as CNF in water and p-tolunitrile in THF. The resulting spectra at 25 °C are shown in Figure 2, where the number to the right of each spectrum indicates the position of the probe according to Figure 1. Each spectrum was fit to a Gaussian curve using an in-house fitting program described elsewhere.41 The mean vibrational frequency and full-width at half maximum (fwhm) of each spectrum are found in the Supporting Information (Table S1). Most of the probe locations (114, 117, 143, 198, and 212) had mean vibrational frequencies that were centered...
on the value for CNF in water (\(\sim 2238 \text{ cm}^{-1}\)). These five probe locations also had peak widths (measured as fwhm) similar to that of CNF in water (\(\sim 10 \text{ cm}^{-1}\)). Two probe locations (165 and 218) had mean vibrational frequencies that were shifted to lower frequencies (\(\sim 2234 \text{ and } 2231 \text{ cm}^{-1}\), respectively) but were still blue-shifted relative to \(p\)-tolunitrile in THF. These two locations had broader peak widths (\(\sim 8 \text{ and } 7 \text{ cm}^{-1}\), respectively). One probe location (145) had an even lower mean vibrational frequency of \(\sim 2225 \text{ cm}^{-1}\) than \(p\)-tolunitrile in THF and had the narrowest peak width (\(\sim 5 \text{ cm}^{-1}\)). The final probe location (92) had a mean vibrational frequency higher (2243 cm\(^{-1}\)) than that of CNF in water and the largest peak width of 12 cm\(^{-1}\). The range of measured frequencies of the nitrile in these locations is larger than the range for the small CNF analogues in THF (2228 cm\(^{-1}\)) and water (2238 cm\(^{-1}\)), which are the two solvents commonly used to represent the extremes of potential solvent hydrogen bonding for CNF. Clearly, the broad range of hydrogen bonding environments available within a complex protein system is not adequately recapitulated by the solvatochromism of small nitrile derivatives.

Next, we measured the temperature dependence of the nitrile stretching frequencies for each of the nine positions in GFP, as well as CNF in water and \(p\)-tolunitrile in THF, over the range 5–35 °C. Over this temperature range, GFP has been shown to remain stably folded.\(^{29,30}\) The mean vibrational frequency and fwhm of each spectrum at each temperature are also found in the Supporting Information (Table S1). Figure 3 shows the change in nitrile center frequency (relative to the frequency at 5 °C) as a function of temperature. Similar to Adhikary et al.,\(^{29}\) we measured a linear response over this temperature range for each of the probe locations, as well as for the small molecules. For the aprotic solvent THF, we did not measure any change in the nitrile stretching frequency over this temperature range (Figure 3, black). Conversely, for CNF in water we measured a large change in frequency as a function of temperature (Figure 3, light blue). Interestingly, each nitrile location in our nine systems had a distinct FTLS, showing that our nine protein systems sampled a broad range of different hydrogen bonding interactions. Most of the CNF probes in GFP gave temperature responses that were in between the extremes of THF and water, which suggests that these probes all experience varying amounts of hydrogen bonding that are not as strong as the well-solvated CNF in water. However, for the nitriles at positions 212 and 92, we measured a FTLS greater than that for CNF in water, possibly suggesting that these two probe locations experience either more frequent, or stronger, hydrogen bonds than CNF in water. This is not surprising for position 212, which is on a solvent exposed loop in GFP (Figure 1). However, position 92 is pointed into the...
GFP barrel, where we do not expect much interaction between the nitrile and water. To interpret the differences in the FTLS of the nitrile in different locations, we turned to MD simulations to investigate the microscopic hydrogen bonding interactions around each probe, including to the nitrile itself.

**SASA Calculations.** In principle, the hydrogen bonding environment of a probe at any location could be investigated in a straightforward way through atomistic simulations. One easy and popular choice for quantifying solvent exposure to specific residues within a protein is with the SASA of the amino acid. However, SASA calculations suffer from the major disadvantage that they lack molecular details about solvent interactions. Specifically, they do not include information about geometric orientations of hydrogen bonding, and they do not capture interactions that arise when a residue is not exposed to bulk solvent, such as those from confined waters or from the protein itself. Nevertheless, SASA serves as a good first approximation to a measurement of hydrogen bonding interactions to the nitrile probe.

We calculated the SASA of several subsets of the atoms within the CNF residue of each mutant (including the amide backbone atoms) over the course of the 40 ns of production MD simulation for each GFP mutant using the gmx sasa tool. Specifically, we calculated the SASA of the nitrile alone; the nitrile plus the Cβ two Hα, and the phenyl ring; and the entire CNF residue including the backbone atoms. These are illustrated in the Supporting Information (Figure S1), and the results are shown in Figure S2. We found that the SASA of just the nitrile with its low total surface area was very sensitive to small changes in the surrounding protein structure, giving large fluctuations in SASA over the course of the simulation. The SASA of the entire residue, shown in Figure 4, gave the broadest range of values with the smallest standard deviations. We have therefore chosen to focus on these values for the analysis that follows. Across each of the nine CNF locations, there was a continuum of SASA results ranging from completely solvent exposed to completely inaccessible. SASA calculations revealed that two probes (positions 117 and 212) were highly solvent exposed and had SASAs that approached the calculated maximum of CNF by itself in solvent (∼2.5 nm²). Three probes, at positions 114, 143, and 198, had intermediate SASAs (∼0.5–1.5 nm²) relative to that of free CNF. Two probes (positions 145 and 165) had a low SASA (∼0.25 nm² to 0.5 nm²) relative to that of free CNF. Finally, two probes (positions 92 and 218) showed nearly zero SASA throughout the simulation, though these two had quite different FTLS values, shown in Figure 3. This discrepancy between the SASA and the measured FTLS suggests that the nitriles in these nine positions experienced differences in local environment more complex than could be captured by SASA calculations.

**Hydrogen Bonding Calculations.** The MD simulations provide molecular-level information about the hydrogen bonds around each nitrile probe that is significantly more detailed than the SASA. For each of the nine systems, we analyzed the MD trajectories for the presence and geometry of hydrogen bonding using an in-house code to characterize hydrogen bonds to a nitrile probe in each frame of the simulation. All nearby O, N, S, and Cα atoms were considered to be potential donors. A hydrogen bond donor was determined to be hydrogen bonding if three conditions were satisfied: (1) the C–N–H angle was greater than 99° (θ₁ in Figure 5); (2) the N–H–R donor angle was greater than 120°, where R donor represents any hydrogen bond donor. dNH is the distance between the hydrogen and the acceptor nitrogen. In this analysis, any S, N, O, or Cα atom was considered to be a potential hydrogen bond donor. To be considered a hydrogen bond, θ₁ must be greater than 99°, θ₂ must be greater than 120°, and dNH must be less than 2.45 Å. N–H–R donor angle was greater than 120° (θ₂ in Figure 5); and (3) the N–H distance was less than 2.45 Å (dNH in Figure 5). These cutoff values were chosen to accept hydrogen bonds that fall within 2 standard deviations of a typical hydrogen bond to nitriles, as defined by Le Questel et al. The code was implemented in C++ for use with the Gromacs package.

**Hydrogen Bonds Donated from Water.** Hydrogen bond geometries between the nitrile and water were calculated for each snapshot for each of the nine MD simulations of CNF containing variants. For each hydrogen bond, the geometry was calculated, and the different geometries were binned by θ₁ and dNH. These results are plotted as heat maps in Figure 6a.

![Figure 4. SASA of the CNF residue as a function of simulation time in 50 ns of production MD simulation for the nine GFP mutants containing the nitrile group.](image-url)
Hydrogen bonding geometries were also binned according to $\theta_2$ and $d_{NH}$. However, the $\theta_2$ (data not shown) and $d_{NH}$ distributions (Figure 6) did not vary greatly across the nine systems and therefore contained less information than shown in Figure 6. For the rest of these analyses, we will focus primarily on $\theta_1$ as the determinant angle. Figure 6a shows that several different hydrogen bonding geometries between the nitrile probes and water were present in our MD simulations. The nitriles at positions 114, 117, 198, and 212 experienced broad distributions of hydrogen bonding angles, with each distribution covering the entire range of allowed angles ($99^\circ$–$180^\circ$). This type of hydrogen bonding profile is consistent with a large degree of solvent exposure, where the hydrogen bonds are free to take on many orientations. A representative snapshot of CNF in a solvent exposed position is shown in Figure 7a. In contrast, Figure 6a shows the nitriles at positions 145 and 218 experienced a much narrower distribution of hydrogen bonds, with hydrogen bonding angles localized between $110^\circ$ and $150^\circ$. A representative snapshot of a hydrogen bond between water and CNF 145 is shown in Figure 7b, and a representative snapshot of a hydrogen bond between water and CNF 218 is shown in Figure 7c. These specific hydrogen bond geometries will be discussed in more detail below. The nitriles at positions 92 and 143 experienced very few hydrogen bonds from water throughout the simulation (Figure 6a). The nitrile at position 165 experienced slightly more hydrogen bonding from water than at positions 92 and 143, with two distinct geometries, one centered near $150^\circ$ and one centered near $100^\circ$ (Figure 6a). Even using only three geometric parameters and considering only hydrogen bonds from water, we observed that CNF at these nine locations experienced a wide variety of possible hydrogen bonding interactions within these constraints. Clearly, a Boolean classification of hydrogen bonding is insufficient to describe the individual positions of each nitrile vibrational probe used in this work.

**Figure 6.** Heat maps of the hydrogen bonding geometries to CNF at each probe location. The titles are colored to correspond to the color codes used elsewhere in the paper. The hydrogen bonding geometries to CNF were calculated at each snapshot (every 4 ps) of the last 40 ns of a 50 ns MD simulation. Contour lines are drawn to guide the eye. $d_{NH}$ (x-axis) and $\theta_1$ (y-axis) are defined in Figure 5. (A) Hydrogen bonds that are donated specifically from water, where $R_{\text{donor}}$ is the oxygen atom of a water molecule. (B) Hydrogen bonds that are donated from protein, where $R_{\text{donor}}$ may be any S, N, O, C$_\alpha$ atom on the protein.

**Figure 7.** Representative snapshots of select hydrogen bonding configurations observed in simulation. (A) The nitrile at position 117 was solvent exposed and involved in multiple simultaneous hydrogen bonds that exchanged frequently throughout the simulation. (B) The nitrile at position 145 was involved with a $\pi$-hydrogen bond with a confined water molecule, where $\theta_1$ was approximately $120^\circ$. (C) The nitrile at position 218 was involved in a hydrogen bond with a confined water molecule, which is also hydrogen bound to the OH group of Ser205. The Connolly surface computed with gmx sasa is shown in pink lines to emphasize how a small available SASA can still have significant hydrogen bonding. (D) The nitrile at position 92 was involved in a linear hydrogen bond with the amide hydrogen of Gln69, where $\theta_1$ was approximately $180^\circ$.

**Hydrogen Bonds Donated from Protein.** In addition to hydrogen bonds from water, we also investigated possible hydrogen bonds formed between the nitrile and donors from protein moieties. Specifically, any O, N, S, or C$_\alpha$ atom in the
protein was considered to be a possible hydrogen bond donor. For each hydrogen bond between the nitrile and a protein donor, the geometry of the interaction was calculated and binned, and the results are shown as a heat map in Figure 6b. Most nitrile locations (114, 117, 145, 198, 212, and 218) did not experience significant hydrogen bonding to donors from the protein. However, the nitriles at positions 143 and 165 experienced some hydrogen bonding from the protein, and the nitrile at position 92 experienced extensive hydrogen bonding from a protein donor. To investigate these protein-donated hydrogen bonding populations, we inspected the trajectories of each nitrile location. For the nitrile at location 165, we found that the nitrile accepted transient hydrogen bonds from nitrogen atoms on two different nearby side chains (Arg96 and Gln183), leading to the two distinct populations shown in Figure 6b. For the nitrile at position 143, we found that the nitrile accepted hydrogen bonds from a nitrogen on nearby Lys207, and interestingly, the nitrile also accepted transient hydrogen bonds from the Cα of nearby Met218. Carbon-donated hydrogen bonds, especially from Cα atoms to nearby oxygen atoms, are common in proteins and have been under investigation for some time.46−48 but this is the first evidence to our knowledge of a carbon–nitride hydrogen bond. Finally, at position 92, we found that the nitrile had a high probability of accepting a linear hydrogen bond from the amide hydrogen of Gln69. A representative snapshot of this interaction is shown in Figure 7d and will be discussed in more detail below. The observation that several of these CNF probes in the interior of GFP were engaged in hydrogen bonds with protein donors leads us to believe that because hydrogen bonds to the nitrile are so energetically favorable, the nitrile will accept weak hydrogen bonds from unconventional donors rather than remain isolated.

**DISCUSSION**

**Relationship between FTLS and SASA.** To assess if the FTLS is a good measure of hydrogen bonding interactions, we plotted the measured FTLS values against the ensemble averaged SASA of the CNF residues from simulation in Figure 8a. Our results show that there is a strong correlation between FTLS and SASA ($r = -0.920$), with two outliers (positions 92 and 218) that are not included in the regression fit. We also performed this analysis using the SASA calculations of the aromatic side chain atoms and the nitrile alone (from Figure S2). The resulting comparison can be found in Figure S3 in the Supporting Information. While the trends in Figure 8a and Figure S3 are consistent, because the relative standard deviations of the SASA increase dramatically as the region of the calculation is reduced, we have chosen to focus on the SASA values calculated for all atoms of the CNF residue. This correlation between FTLS and SASA indicates that the CNF probes in locations that have higher SASA, and therefore higher potential hydrogen bonding interactions with water, show a greater change in the nitrile absorption frequency as temperature is changed. The two outliers of positions 92 and 218 experienced hydrogen bonds that are not well-described by SASA calculations and will be discussed in more detail below. The SASA calculations for positions 117 and 198 have larger standard deviations than the rest of the data. Examining snapshots from the MD simulations revealed that CNF 117 alternated between two states with different SASAs through a torsional rotation about $\chi_1$ of CNF. Similarly, CNF 198 drifted from a state of high SASA to a state of lower SASA over the course of the simulation. These two examples of complex dynamic behavior of the CNF side chain indicate that enhanced MD sampling of CNF and its environment could lead to an even more quantitative relationship between SASA and FTLS; we are currently investigating this hypothesis. It is important to note that not all hydrogen bonding environments are captured by a calculation of the SASA. For the two outliers discussed previously, CNF 92 and 218, we saw that the nitriles in these two locations experienced hydrogen bonding interactions that were not well-described by SASA. The data in Figure 6 clearly demonstrate that CNF 92 was involved in a strong linear hydrogen bond with the amide hydrogen of Gln69, a representative snapshot of which is shown in Figure 7d. Since this interaction was not due to solvent, it was not captured by the SASA. In this case, relying only on the SASA calculation would have resulted in the incorrect conclusion that this nitrile was in an environment free from hydrogen bonding. The FTLS, however, was the largest of any of the probe locations. This leads us to believe that nitriles in buried locations as predicted by the calculation of SASA can still be involved in significant hydrogen bonding,
and that SASA should not solely be relied upon for evaluating the hydrogen bonding potential of a probe location. Further, we believe that this shows that in such cases, the FTLS still provides a good experimental measure of hydrogen bonding interactions of the nitrile. Because the FTLS was large over the temperature range measured in this particular case, we hypothesize that at increased temperatures, the thermal fluctuations of the protein backbone will increase and break the linear hydrogen bond between CNF 92 and the protein backbone, resulting in a large FTLS.

Additionally, CNF 218 had a very low SASA, but the data in Figure 6a clearly show extensive hydrogen bonding to the solvent. While the entire residue was surrounded by protein, leading to a low SASA, the nitrile group itself pointed directly to a single confined water stabilized by a hydrogen bond to Ser205, a representative snapshot of which is shown in Figure 7c. This water formed hydrogen bonds with CNF 218 and could not easily diffuse away from the nitrile when the hydrogen bond was broken. While the water molecule exchanged throughout the simulation, it did so rarely. In this case, the calculated SASA did not capture the hydrogen bonding interactions, and relying solely upon that result would have led to the incorrect conclusion that this nitrile location was free from hydrogen bonding; however, the measured FTLS still captured the hydrogen bonding interactions. We interpret the FTLS of CNF 218 to indicate that this water is more likely to exchange with increased temperature, weakening the hydrogen bonding interaction to the nitrile and shifting the frequency. We are currently investigating both of these hypothesized interpretations of the FTLS with additional simulations at varying temperatures.

These two observations from the MD trajectories of CNF 92 and 218 provide clear and direct explanations as to why they are outliers of the otherwise strong correlation between SASA and FTLS shown in Figure 8a. When the collection of nitrile locations is taken together, it is apparent that the temperature dependence of the nitrile frequencies is directly related to their respective hydrogen bonding environments in a quantitative manner. These measurements demonstrate that the FTLS may be used to assess whether or not a particular nitrile location experiences a change in specific hydrogen bonding interactions as a result of some perturbation, such as an amino acid mutation or ligand docking event. This would therefore allow the change in nitrile frequency to be attributed to that particular perturbation, and not to a change in specific hydrogen bonding environment.

Nitrile Frequency Interpretations. With the MD simulations of CNF-containing GFP mutants and the predicted effects of hydrogen bond angle on the nitrile frequency reported by Choi et al., we now have a straightforward way to interpret the relative frequency of our nitrile absorption spectra that would not have otherwise been possible. The solvent exposed probes at positions 114, 117, 198, and 212 all had very similar FTIR absorption spectra, with frequencies centered on ~2238 cm⁻¹ and peak widths of ~10 cm⁻¹. Accordingly, the MD simulations showed that all four had similar water hydrogen bonding geometry profiles (Figure 6a) and the four largest SASA values (Figure 4).

To see whether the nitrile frequency predictions reported by Choi et al. could be used in other cases, we extended this analysis to the CNF probe locations that experienced different hydrogen bonding geometries. For example, the mean vibrational frequency of the nitrile at position 145 was the most red-shifted of any of the spectra and had the narrowest fwhm. In simulation, CNF 145 was involved frequently in a hydrogen bond to water at ~120° (Figure 7b). In agreement with our experimental observations, Choi et al. predicted that these π-hydrogen bonds to nitriles red-shift the stretching frequency. Additionally, hydrogen bonds to CNF 145 were more constrained with a smaller angle distribution than any other probe, corresponding to the narrow fwhm of the measured FTIR spectrum.

Similarly, we found that the nitrile at position 218 was also involved in π-hydrogen bonding (Figure 7c) but with a larger maximum angle probability than that of CNF 145 by ~10°. Consistent with the predictions from Choi et al. for π-hydrogen bonding, the frequency of the nitrile at position 218 was blue-shifted relative to the nitrile at position 145, but also red-shifted relative to all the other probe locations (Figure 2). This spectrum was again quite narrow, consistent with the MD simulation information of a constricted hydrogen bond.

The two nitrile locations (143 and 165) that experienced the fewest hydrogen bonds of any of the locations are more difficult to interpret in this manner. Nevertheless, the few hydrogen bonds experienced by the nitrile at position 143 are of similar θ₁ (centered around 145°) to the solvent exposed residues. Experimentally, its mean vibrational frequency was also similar to the solvent exposed residues. The nitrile at position 165 experienced a small population of hydrogen bonds from protein at sharp angles (~100°, Figure 6b), and while this angle should red-shift the frequency, the effect was small due to the size of the population. Correspondingly, the absorbance frequency of the nitrile at position 165 is red-shifted relative to the solvent exposed residues, but not as red-shifted as positions 92 and 218.

Finally, we found a direct interpretation of the far blue-shifted spectrum of the nitrile at position 92. In the absence of any molecular-level information, the interpretation of this spectrum would be very difficult, because this location is in the interior of the β-barrel, removed from potential bulk water hydrogen bonds. However, the simulations revealed that the nitrile at this location was involved in a strong linear hydrogen bond (Figure 7d) to the amide backbone of Gln69. Such σ-hydrogen bonds are predicted by Choi et al. to blue-shift the spectra by withdrawing electron density from the antibonding molecular orbital of the nitrile, increasing the overall bond order of the group and shortening the bond, leading to higher frequency. Additionally, we found that there were two distinct hydrogen bonding populations for the nitrile at this location. While the dominant population was linear (~160–180°) and blue-shifting, there was a smaller population at a smaller value of θ₁ (~100°). This smaller population caused a lowering of the nitrile frequency relative to the linear hydrogen bonded population and resulted in a broadening of the experimental FTIR spectrum. Indeed, the nitrile at position 92 had the largest fwhm and was the only mutant from the MD trajectories that had two significant hydrogen bonding populations that occurred at sharply different angles.

To quantify the effect of hydrogen bonding geometry on the measured nitrile frequencies, we binned all of the hydrogen bonding geometries observed for each probe by angle and fit a polynomial to smooth the noise of the histogram (Figure S4a). Because of the geometric considerations of the hydrogen bonding probability, we calculated the probability density for each angle of θ₁ by normalizing by the volume element of each angular bin, calculated via eq 1:
Here, $V$ is the volume element of the bin, $\theta_{l}'$ and $\theta_{l}''$ are the bin edges of $\theta_{l}$, and $d_{NH}$ is the maximum hydrogen bond distance. The normalized probability densities are shown in Figure S4b. We then took the weighted average of all the hydrogen bonds by probability (black vertical line in Figure S4b) and plotted this average angle against the mean vibrational frequency of each nitrile in Figure 8b.

For all nine probe locations investigated, we observed a strong correlation ($r = 0.944$) between the hydrogen bonding angle and the mean vibrational frequency, as Choi et al. predicted in their quantum mechanical calculations. The strength of the correlation is somewhat surprising given the well-known inaccuracies of hydrogen bonding geometries in the Amber03 force field, and that we have not accounted for electrochromic effects contributing to the measured frequencies or for exchange-repulsion interaction-induced frequency shifts. These two effects are expected to also influence the measured vibrational frequency.

To account for electrochromic effects that may contribute to the measured frequency, we estimated the extent of the Coulombic frequency shift by calculating the Coulombic electric field at the midpoint of the nitrile using a procedure described elsewhere. The calculated electric field is plotted against the mean vibrational frequency in the Supporting Information (Figure S5). The calculated electric fields are poorly correlated to the mean vibrational frequency, suggesting that electrochromic effects are a minor contribution to the measured frequency. We separately investigated the extent of the exchange-repulsion interaction-induced frequency shift in these systems by analyzing the hydrogen bond length parameter, $d_{NH}$. The histograms and normalized probability densities for each hydrogen bonding distance bin are shown in the Supporting Information (Figure S6), and the average distance is plotted against the mean vibrational frequency in Figure S7. However, presumably because each of these distance interactions are governed by the same Lennard-Jones potential from the Amber03 force field, the probability distributions of the lengths of hydrogen bonds to water are remarkably similar. Thus, the average hydrogen bond distances are not well-correlated to the experimental mean vibrational frequencies. Because of the strong correlation of the hydrogen bonding angle, the lack of correlation of the calculated Coulombic field, and the lack of correlation of the average hydrogen bonding distance, we believe that the hydrogen bonding angle is the dominant contribution to the frequency in these cases.

In order to contribute to the line shape of the nitrile spectrum, a given hydrogen bond would need to have a lifetime longer than the vibrational lifetime of the nitrile probe. To determine whether or not that was the case for the hydrogen bond populations in Figure 6, we tracked each hydrogen bond in all nine trajectories. As a function of simulation time, we plotted the number of hydrogen bonds that persisted for at least that length of time. These plots were fit to an exponential decay function and are shown in the Supporting Information (Figure S8). The lifetimes were calculated from the decay rate and are reported in Table S2. All but one hydrogen bonding population had a lifetime approximately equal to or longer than the vibrational lifetime of CNF, which is estimated to be on the order of a picosecond or less, suggesting that these populations contribute to the line shape of the nitrile absorption spectrum.

It is interesting to note that out of all of our nine protein systems, not a single nitrile location was free from hydrogen bonding. While this certainly may be an artifact of GFP with its unique water filled $\beta$-barrel structure, it appears that hydrogen bonding to the nitrile is so favorable that even when placed in a location with no solvent accessibility, the nitrile will find a donor other than water. For example, when we put the nitrile at position 92, a location with virtually no solvent exposure, we observed that the nitrile was involved in significant hydrogen bonding with protein atoms. Additionally, when we put the nitrile at position 143, a location that was void of any conventional hydrogen bond donors, the nitrile experienced transient hydrogen bonds from an electron poor carbon atom. The observation of these weak hydrogen bonds from carbon suggests that even if there are no strong hydrogen bond donors nearby, the nitrile will still engage in interactions to fulfill its hydrogen bond accepting capabilities. If these observations of nitrile-labeled GFP extend to other systems, this would certainly complicate the interpretation of nitrile spectra in biological systems. Our results suggest that to properly interpret nitrile spectra, one must always consider hydrogen bonding effects. Because of the consistent presence of hydrogen bonds and the strong correlation between the hydrogen bonding angle and the mean vibrational frequency, it is necessary to determine the geometry of hydrogen bonds to the nitrile in all reference and perturbed states to ensure that the measured frequency shifts are not due to a change in hydrogen bond geometry. Specifically, when measuring a particular perturbation via vibrational spectroscopy of a nitrile, the FTLS of the reference and perturbed state may be used to confirm that specific hydrogen bonding interactions are not changed substantially by the perturbation.

Certainly, we have shown that a Boolean classification of the existence of hydrogen bonding is insufficient when the nitrile is placed in complex protein environments. There is a broad range of hydrogen bonding interactions available to a nitrile probe, and each circumstance contributes differently to the nitrile frequency. This means that there is not a constant change in frequency that can universally be applied to all hydrogen bonding interactions. However, the ability of nitriles to accept hydrogen bonds should not be interpreted as a disadvantage for its use as a vibrational probe. The strength of the correlation in Figure 8b demonstrates that it is possible to decouple the hydrogen bonding shift from other spectroscopic effects, such as the VSE, when there is sufficient molecular information about the hydrogen bonding geometry and FTLS.

### CONCLUSION

We have experimentally measured the temperature-dependent frequencies of CNF in nine different locations across the GFP protein and demonstrated the broad range of hydrogen bonding environments present in a complex protein system. We showed that the FTLS is directly related to molecular-level details of the specific hydrogen bonding interactions to the nitrile probe. The FTLS is strongly correlated to the SASA of the CNF probe, except in cases where constrained molecular geometry resulted in significant hydrogen bonding that was not captured by SASA. Nevertheless, it appears that the FTLS of a nitrile probe can be used to rapidly assess the degree of hydrogen bonding of the nitrile, even in the dynamic and complex environment found in and around a protein. These
molecular-level details are essential for properly interpreting the vibrational spectrum of the nitrile probe in any experiment where the specific hydrogen bonding interactions may be changing. This is especially necessary to ensure that a VSE experiment is actually reporting on a change in electric field instead of reflecting sensitivity to subtle changes in hydrogen bonding geometries.

We have further demonstrated the significant effect of hydrogen bonding on the nitrile spectra by showing that in our nine CNF containing variants, the mean nitrile vibrational frequency was dominated by the hydrogen bonding geometry, and that the measured frequencies matched what was expected from a theoretical prediction of the effect of hydrogen bonding geometry on nitrile frequency. In every one of our systems, the CNF probe experienced some type of hydrogen bonding. For all observed variants, the average hydrogen bonding angle was well-correlated with the mean vibrational frequency observed in experiment. Additionally, the fwhm of the experimental spectra were qualitatively consistent with the distribution of the hydrogen bonding geometries from simulation. Finally, we have shown that FTLS measurements may be used to assess the hydrogen bonding interactions of a nitrile so that they may be properly accounted for in the interpretation of the spectra.

## ASSOCIATED CONTENT

2 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpbc.8b03907.

Vibrational frequencies, SASA values, histograms of hydrogen bond angles and lengths, Coulombic field calculations, and hydrogen bond lifetimes (PDF)

## AUTHOR INFORMATION

Corresponding Author
*E-mail: lwebb@cm.utexas.edu.

ORCID

Lauren J. Webb: 0000-0001-9999-5500

Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This work was supported by the Welch Foundation (F-1722). We gratefully acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing high performance computing resources that have contributed to the results reported within this paper. We would also like to thank Riccardo Nifosi for providing the Amber parameters for the GFP fluorophore and Ryan Mehl for providing the pDule and pBAD plasmids.

## REFERENCES

The Journal of Physical Chemistry B


