Site-Specific Spectroscopic Reporters of the Local Electric Field, Hydration, Structure, and Dynamics of Biomolecules

Matthias M. Waegele, Robert M. Culik, and Feng Gai*  
Department of Chemistry, University of Pennsylvania, Philadelphia, PA 19104

Abstract

Elucidating the underlying molecular mechanisms of protein folding and function is a very exciting and active research area, but poses significant challenges. This is due in part to the fact that existing experimental techniques are incapable of capturing snapshots along the ‘reaction coordinate’ in question with both sufficient spatial and temporal resolutions. In this regard, recent years have seen increased interests and efforts in development and employment of site-specific probes to enhance the structural sensitivity of spectroscopic techniques in conformational and dynamical studies of biological molecules. In particular, the spectroscopic and chemical properties of nitriles, thiocyanates, and azides render these groups attractive for the interrogation of complex biochemical constructs and processes. Here, we review their signatures in vibrational, fluorescence and NMR spectra and their utility in the context of elucidating chemical structure and dynamics of protein and DNA molecules.

Keywords

Infrared probe; unnatural amino acid; nitrile stretching vibration; azide stretching vibration; spectral modeling

The free energy landscape of biomolecules, such as proteins, is a function of a large number of degrees of freedom. Thus, elucidating the structural dynamics of any protein with high spatial and temporal resolution is a difficult enterprise. In this regard, recent years have seen an increased interest in the utilization of various extrinsic molecular probes that can be incorporated into biomolecules in a site-specific manner to monitor, for example, the local electrostatics and dynamics of proteins and DNA. Herein, we offer our perspectives on this fast-growing field, with a focus on the recent development of unnatural amino acids as infrared (IR) environmental and/or structural reporters of proteins, especially those containing a nitrile, thiocyanate, or azide group. While other site-specific IR probes, such as those based on isotopic labeling, are also very useful in protein folding and conformational studies, they have been discussed and reviewed elsewhere.1-7 One of the advantages of using unnatural amino acid spectroscopic probes is that they often induce minimal perturbation to the protein structure in question and can be incorporated into the polypeptide sequence of interest using either in vitro or in vivo methods.8,9 Although the development of incorporation methods itself is a very active field, an extensive review of the recent advances in selectively and site-specifically incorporating unnatural amino acids into proteins is beyond the scope of this Perspective. Briefly, besides standard solid phase peptide synthesis or semi-synthesis methods,10,11 other in vitro methods include those via post-translational modification of a specific amino acid sidechain, such as cysteine,12-15 arginine,16 lysine,17 and tyrosine,18 whereas in vivo methods generally involve direct

*To whom correspondence should be addressed; gai@sas.upenn.edu; Phone: 215-573-6256; Fax: 215-573-2112.
subcloning (e.g., green fluorescent protein fusions), suppressor tRNA modifications, and amino acid replacement techniques.

Site-Specific Vibrational Probes

Vibrational transitions in condensed phases typically occur on ultrafast (fs−ps) timescales. Therefore, time-resolved vibrational spectroscopy is capable of offering high temporal resolution in protein conformational studies. However, without introduction of a suitable and site-specific vibrational reporter, this technique is often limited to low structural resolutions. To overcome this limitation, many strategies have been developed to achieve site-specificity. One such strategy, which recently is becoming more widely used, is to employ unnatural amino acids that contain an environmentally (and hence structurally) sensitive and relatively strong vibrator (compared to the peak extinction coefficient of the amide I transition, 720 M \(^{-1}\) cm\(^{-1}\)). As pointed out by Getahun \textit{et al.}, for a vibrational mode to be useful as a monitor of the local environment of proteins it should be a simple transition, largely decoupled from the rest of the molecule, and it should also have a relatively intense, narrow absorption band that is separated from other absorption bands of the molecule. An additional requirement, perhaps a more important one, is that the extrinsic vibrational probe should minimally perturb the conformation of the protein in question. Nitrile (C\(\equiv\)N) and azide (N\(_3\)) moieties largely meet these criteria as their stretching vibrations not only have relatively high extinction coefficients, but are also well separated from other vibrational modes of proteins. In addition, these groups can serve as an atomic substitution within an amino acid sidechain (Table 1), thus minimizing any perturbative effects. For example, Getahun \textit{et al.} have shown that nitrile-derivatized amino acids could be used as local IR environmental probes of proteins (Figures 1 and 2). Below, we will first discuss experimental investigations and theoretical modeling of the vibrational transitions of these probes, and then review specific examples of their applications in several representative biological and biophysical studies.

Properties of the C\(\equiv\)N stretching mode of nitriles

The nitrile stretching band of a variety of nitrile compounds have been extensively studied in the gas\(^{27-31}\) and condensed\(^{32-63}\) phases. In particular, the solvent,\(^{41,43-45,47,48,50-52,59,60,62,63}\) temperature,\(^{44,45,49}\) pressure,\(^{32,51}\) electrostatic field,\(^{55-57}\) or complexation\(^{33,34,36-40,46,64,65}\) induced vibrational frequency shifts from a reference value and intensity changes of this normal mode have been the focus of many previous investigations. For example, the nitrile stretching band of acetonitrile (CH\(_3\)CN), which is partly affected by a Fermi resonance\(^{50,54}\) between this mode and the combination band of the CC stretch and the symmetric CH\(_3\) bend, as well as a hot band (\(\nu_2 + \nu_8 - \nu_8\)),\(^{27-31,45,54}\) has been studied in the neat liquid,\(^{42,44,45,49,51,54}\) under dilute conditions in a large number of solvents,\(^{47,48,50,52,54}\) at air/liquid interfaces,\(^{53,58}\) as well as in solid matrices.\(^{55-57}\) While a theoretical model that can quantitatively predict the frequency shift of the C\(\equiv\)N stretching vibration for any ‘solvent’ condition is still lacking,\(^{54}\) the corresponding frequency-solvent relationship is well understood at a semi-quantitative level. When dissolved in most solvents, the peak position of the C\(\equiv\)N stretching band of acetonitrile is found to red-shift from its gas phase value of 2266.5 cm\(^{-1}\)\(^{27-31}\) except when dissolved in fluorinated alcohols, for which small blue-shifts are observed.\(^{48,52,54}\) Specifically, the solvent acceptor number, as a quantitative measure of Lewis acidity, was found to be a qualitative reporter of the magnitude of the solvent-induced red-shift with a relation that a larger solvent acceptor number corresponds to a smaller red-shift.\(^{50,52}\) This observation is attributed to an increasingly stronger coordination bond between the nitrogen of the nitrile group and a stronger Lewis acid, as stronger coordination was found to increase the C\(\equiv\)N bond strength.\(^{54,65}\) However, it has been shown that Lewis acidity alone is incapable of quantifying all solvent induced frequency shifts.\(^{48,52,54}\) This observation was interpreted as
a complex interplay between non-specific and specific interactions between acetonitrile and the solvent. Generally, the solvatochromatic shift of the C≡N stretching band, ν, may be expressed as

\[ \Delta \nu = \Delta \nu_{NS} + \Delta \nu_{S} = \Delta \nu_{C} + \Delta \nu_{R} + \Delta \nu_{A} \]  

(1)

Here, \( \Delta \nu_{NS} \) and \( \Delta \nu_{S} \) refer to the solvent-induced shifts by non-specific and specific interactions, respectively. The non-specific contribution component can be further decomposed into three independent terms: Upon solvation, the centrifugal distortion of the acetonitrile molecule due to free rotation in the gas phase is reduced, resulting in a red-shift of \( \Delta \nu_{C} = -0.5 \text{ cm}^{-1} \), whereas the short-range repulsive forces, which arise from collisions of acetonitrile with solvent molecules, were estimated to give rise to a blue-shift of \( \Delta \nu_{R} = +4.2 \text{ cm}^{-1} \). Thus, the observed red-shifts arise mainly from attractive forces, i.e. dispersive and dipole-dipole interactions between the solvent and solute molecules, namely,

\[ \Delta \nu_{S} = D \alpha + P \alpha, \]  

(2)

where \( D \) and \( P \) are the dielectric strength and polarizability of the solvent, respectively.

Reimers and Hall\(^{54} \) estimated the relative contributions of the terms in Eq. (1) by setting \( \Delta \nu_{C} \) and \( \Delta \nu_{R} \) to the aforementioned values and by fitting the non-specific component of Eq. (1), i.e. \( \Delta \nu_{NS} = \Delta \nu_{C} + \Delta \nu_{R} + \Delta \nu_{A} \), to the frequency shifts of C≡N stretching band of acetonitrile obtained in 33 solvents using \( A_{\varepsilon} \) and \( A_{\alpha} \) as fitting parameters. They found that whereas \( PA_{\alpha} \) falls within the range of \(-8.2 \) to \(-15.1 \text{ cm}^{-1} \), making the largest contribution to \( \Delta \nu_{A} \), the \( PA_{\varepsilon} \) term, which was found to be between \(-1.5 \) to \(-6.5 \text{ cm}^{-1} \), also contributes significantly to the observed red-shifts. The difference between this fit and the experimentally observed shift for a given solvent was then taken to be the specific contribution of this solvent to \( \Delta \nu \) (i.e., \( \Delta \nu_{S} \)). Their findings indicate that the specific contribution term can vary greatly from solvent to solvent. For example, the \( \Delta \nu_{S} \) was found to be \(-0.8 \text{ cm}^{-1} \) for benzene, but \(+6.7 \text{ cm}^{-1} \) for water and \(+15.4 \text{ cm}^{-1} \) for trifluoroethanol.

The oscillator strength of the C≡N vibrational transition was also found to be strongly dependent on solvent\(^{47,48} \) or coordination partner.\(^{33,34,36,37,39} \) In particular, previous studies have shown that the C≡N band of acetonitrile is significantly smaller in aprotic solvents than in protic solvents,\(^{47,48} \) due to an increased polarization of the nitrile group upon interaction of the nitrogen-lone pair with a Lewis acid. This can be understood by the following simple theoretical line of argument. If the dipole moment of the C≡N group is thought of as arising from two partial atomic charges \( \pm \delta q \), then, upon displacement \( \delta r \), Thus, an increase in the from the equilibrium C≡N bond length \( r_{e} \), the dipole is \( \mu = r_{e} \delta q + \delta r \delta q \), magnitude of the partial charges \( \pm \delta q \) results in a larger vibrational transition dipole moment, i.e.

\[ \mu_{0} = \int \left( \frac{d \mu}{d \delta r} \right) \langle v_0 | \sigma | v_1 \rangle = \delta q \langle v_0 \sigma | v_1 \rangle, \]  

(3)

where \( v_0 \) and \( v_1 \) are the wavefunctions of the vibrational levels involved in the transition.

Moreover, it was found that aryl nitriles typically have larger vibrational transition dipole strengths than alkyl nitriles, although the exact value depends on the structure of the aromatic ring (Figure 3).\(^{35,36,55,63} \) For example, whereas the transition dipole of \( p \)-tolunitrile was found to be approximately a factor of 1.6 times larger than that of benzonitrile, the oscillator strengths of \( o \)-nitrobenzonitrile and \( m \)-nitrobenzonitrile were found to be about 6...
and 3 times smaller, respectively.\textsuperscript{36} Similarly, Andrews and Boxer\textsuperscript{55} found a correlation between the Hammett number and the transition dipole strength of substituted benzonitriles. Cho and co-workers\textsuperscript{63} found that the extinction coefficient of 4-cyanophenol is smaller than that of 4-cyanophenoxide by a factor of two. These observations can be understood in terms of Eq. (3) as a greater electron donation group to the aromatic ring is expected to result in a larger partial charge on the nitrile moiety.

Besides alkyl and aryl nitriles, thiocyanates (R-S-C≡N) and cyanates (R-O-C≡N) have also been studied in detail. For example, Londergan and co-workers\textsuperscript{67} have studied the shape, peak position and extinction coefficient of the nitrile stretching band of methyl thiocyanate in nine solvents. Their results show that the line shape of the C≡N stretching band is symmetric in all solvents studied, except in fluorinated alcohols, which was attributed to the existence of two or more distinct populations of hydrogen-bonded complexes formed between the solvent and solute. Organic cyanates (R-O-C≡N) are also interesting candidates to serve as vibrational probes due to the comparatively large extinction coefficient of the O-C≡N asymmetric stretching mode.\textsuperscript{68,69} A recent study on phenyl cyanate found that the IR spectrum of this compound exhibits a complex band shape in the O-C≡N stretching region due to the involvement of this vibrational mode in at least two Fermi resonances with overtones,\textsuperscript{69} rendering derivatives of phenyl cyanate less likely to be useful as vibrational probes in conventional IR studies.

Finally, for experiments in which the chemical dynamics of a system are of interest, the vibrational lifetime of the probe is of great importance as it determines the time window in which the dynamics can be followed. The vibrational lifetimes of a few nitriles have been measured and they typically range from ~1 to 5 ps in water.\textsuperscript{62,63,70,71}

\textbf{Properties of the asymmetric stretching mode of organic azides (R-N$_3$)}

The asymmetric stretching vibration of alkyl and aryl azides (R-N$_3$) are also attractive candidates as vibrational probes of biomolecules as their vibrational transition dipole moments are typically stronger than that of C≡N stretching modes (Figure 4).

Unfortunately, contrary to the IR spectra of most nitrile compounds, which typically show a single, sharp band ($\Delta \nu = 6$–15 cm$^{-1}$) in the C≡N stretching region, the band of the N$_3$ asymmetric stretch of organic azides is much broader ($\Delta \nu = 25$–40 cm$^{-1}$) and typically convolved with one or more bands in the same region. The origin of this complexity has been the topic of various studies.\textsuperscript{72-75} Early studies tentatively attributed the band structure to mixing between a combination tone of low frequency vibrations and the N$_3$ asymmetric stretch due to anharmonic coupling.\textsuperscript{72,74} For example, in a systematic study of the vibrational spectra of various aromatic acid azides (R-CON$_3$), Lieber and coworkers\textsuperscript{72} found that the IR spectra of all aryl acid azides have two peaks in the N$_3$ asymmetric stretching region (around 2200 cm$^{-1}$), except those whose CON$_3$ group is separated from the aromatic phenyl ring by one or more methylene groups. They interpreted this splitting as the result of a Fermi resonance between a combination band and the N$_3$ asymmetric stretching band. The combination band was thought to arise from the combination of the N$_3$ symmetric stretch (~1200 cm$^{-1}$) and a band at ~1000 cm$^{-1}$, which was tentatively attributed to one of the aromatic ring modes. Similarly, the study of Dyall and Kemp\textsuperscript{74} on a series of aryl azides also suggested that the complex band structure is more likely to arise from a Fermi resonance than from hot bands. In a more recent experiment, Cheatum and coworkers\textsuperscript{75} found direct evidence of a Fermi resonance interaction between the N$_3$ asymmetric stretch and a combination or overtone state in the form of cross-peaks in the 2D-IR spectrum of 3-azidopyridine at early delay times.
The presence of a Fermi resonance will certainly complicate the interpretation of the experimental results as those vibrations involved in the resonance are expected to be affected differently by the corresponding environmental changes. In several instances, however, Brewer and coworkers have demonstrated that such Fermi resonances can be eliminated by isotope editing of the azide group. In addition, in support of the interpretations provided in earlier studies, Hochstrasser and coworkers found that the IR spectrum of the non-aromatic compound 2′-azido-2′-deoxyuridine in the N₃ asymmetric stretch region exhibits a single, sharp peak.

**Theoretical modeling of vibrational lineshapes of nitriles, thiocyanates, and azides**

In order to facilitate a quantitative, structurally-based interpretation of any experimentally determined vibrational spectra, it is highly desirable to model such spectra via molecular simulations. To this end, many theoretical and computational studies have been carried out. In particular, for the calculation of nitrile stretching frequencies and vibrational lineshapes, two empirical approaches have gained popularity. The first approach originated in the groups of Cho and Skinner, while the second was recently introduced by Corcelli and coworkers. Both methods rely on a robust classical molecular mechanical (MM) or quantum mechanical/molecular mechanical (QM/MM) simulation of the system of interest. In the first approach, QM harmonic or anharmonic frequency analyses of the vibrational transition of interest are performed on a large number (tens to thousands) of solute-solvent clusters, which are randomly chosen from the trajectory, resulting in a set of distinct ab initio frequencies to which the following equation is fit:

\[ \omega = \omega_{\text{gas}} + \sum_{i=1}^{j} \alpha_i \phi_i + \sum_{j=N,Y} \beta_i E_{i,j} \]  

where \( \phi_i \) and \( E_{i,j} \) are the electrostatic potential and Cartesian electric field component \( j \) at position \( i \), arising from the solvent environment, and the constants of proportionality \( \alpha \) and \( \beta \) are determined by fitting the model to the ab initio determined vibrational frequencies of the probe-solvent clusters, assuming the appropriate gas phase frequency \( \omega_{\text{gas}} \) of the probe. Once the parameters \( \alpha \) and \( \beta \) are determined, the fluctuation of the frequency along the molecular dynamics (MM or QM/MM) trajectory may be easily calculated via Eq. (4) by summing over all solvent partial charges. Using a related method, we recently showed that the broadening of the nitrile band of 5-cyanoindole upon hydration is partly due to interactions between water and the aromatic indole ring and does not only arise from direct interactions between water and the nitrile moiety.

The second approach, developed by Corcelli and coworkers, is based on the PM3 semi-empirical method, which was optimized to reproduce the C≡N vibrational frequencies of acetonitrile-water clusters. Using the single set of parameters obtained via this optimization procedure, they found quantitative agreement between the experimental and the calculated C≡N vibrational bandwidths of acetonitrile in both water and tetrahydrofuran (THF). Furthermore, they correctly predicted the solvent-induced frequency shift when going from THF to water, indicating that the method is “solvent transferable”. As the purpose of using a site-specific spectroscopic reporter is, in many cases, to probe the heterogeneous environment found in biomolecular systems, solvent transferability of a method for lineshape calculation is, therefore, of paramount importance. Though the former method based on Eq. (4) typically provides good agreement between experimental and calculated lineshapes, it is not entirely clear to what extent the resulting frequency-frequency map is solvent transferable. However, some evidence for transferability can be found in the literature.
Force field parameters are currently available for a range of model compounds of the aforementioned probes, \(^{81,87,89,93-96}\) but not all of those may be entirely compatible with commonly used force fields in molecular dynamics simulations, as they were derived by procedures that are different from those typically employed for a particular force field. To ensure the integrity of a molecular dynamics simulation of a more complex system containing a specific vibrational probe, it would be beneficial to have parameters available that were derived using the standard procedures for a given force field.

### Specific Examples

Taken together, results emerging from previous studies on simple nitrile, thiocyanate, cyanate and azide compounds indicated that the stretching vibration of some diatomic and/or triatomic moieties such as C≡N and N\(_3\) are sensitive vibrational probes of their local environment. However, to be applicable in biological studies, such probes need to be first incorporated into the biological molecules of interest. The most common practice is to incorporate these probes into polypeptides or nucleic acid strands in the format of unnatural amino acids or nucleic acids. As summarized in Table 1, a large number of unnatural amino acids and nucleic acids have been tested as site-specific spectroscopic probes of the local environment and/or conformation of biological systems. Below we highlight a few specific examples, showing how such unnatural vibrational reporters could be used to provide new insight into biological and/or biophysical problems of interest. Unfortunately, the entire body of deserving studies cannot be reviewed in this section and the interested reader may find other very interesting applications of these probes in the references cited.\(^{97-113}\)

### Probing local electrostatics

Because of the relatively large Stark tuning rate of the nitrile stretching vibration, one of its applications is to measure the local electric field in chemical and biological systems. For example, by incorporating β-Ala\(_{SCN}\) into the active site pocket of the enzyme ketosteroid isomerase (KSI) and using the nitrile stretching mode as a vibrational Stark probe, Boxer, Herschlag, and coworkers\(^ {114}\) studied the changes in the local electrostatic field associated with substrate binding and steroid isomerization induced charge localization. Based on the estimate that the steroid will displace \(8-10\) water molecules, they found a surprisingly small change in the electrostatic character of the active site of \(-0.6 \text{ MV/cm}\) along the C≡N axis upon binding. On the contrary, they found that binding of a transition-state analogue, equilenin, whose negative charge is localized on the carbonyl, lead to a more significant change in the local electrostatic field \((-3 \text{ MV/cm})\). In a similar study using β-Ala\(_{SCN}\) as an IR electrostatic reporter, Webb and coworkers\(^ {115}\) investigated how the local electrostatic field at the interface of the Ras binding domain of the protein Ral guanine nucleotide dissociation stimulator (RalGDS) changes upon binding of the protein Rap or Ras. Their results showed that the stretching frequencies of the S-C≡N reporter obtained with eleven RalGDS mutants are weakly correlated with the solvent accessible surface areas (SASAs) of the mutational sites. This observation suggests that the stretching frequency of the S-C≡N group provides a direct, site-specific experimental measure of the SASA. In addition, their results indicated that this IR reporter could be used to identify residues that are responsible for binding. In a more recent study, Webb and coworkers\(^ {116}\) employed Stark spectroscopy and p-Phe\(_{CN}\) as a local electrostatic probe to directly assess the dipole electric field in bicelle model membranes by inserting a p-Phe\(_{CN}\)-labeled trans-membrane helix into the bilayer. Their results provided an estimate of this electric field, which extends from the intermediate region near the lipid headgroup to the center of the lipid bilayer and is about \(-6 \text{ MV/cm}\).

As discussed above, the vibrational frequency of a nitrile or azide can be affected by both electric field and hydrogen bonding interactions. Thus, it is imperative to separately determine these two effects when using such vibrational reporters as local electrostatic
probes of biological molecules. Recently, Boxer and coworkers\textsuperscript{117} proposed a method for decomposing the apparent vibrational frequency shifts into hydrogen bonding- and electrostatic field-induced components.

**Characterization of amyloid formation and intrinsically disordered proteins**

Recently, Raleigh, Zanni, and coworkers\textsuperscript{118} and the groups of Decatur and Kirschner\textsuperscript{119} demonstrated that the nitrile stretching vibration is a useful probe of the kinetics and mechanism of amyloid formation. For example, by using \textit{p\textendash}Phe\textsubscript{CN} as a site-specific IR and fluorescence probe Marek \textit{et al.}\textsuperscript{118} were able to characterize the transient species that are populated in the lag-phase of amyloid formation of human islet amyloid polypeptide (IAPP). By combining fluorescence and IR measurements on \textit{p\textendash}Phe\textsubscript{CN} labeled fibrils, they further demonstrated that the aromatic residues are exposed to different extents in the fibrils of IAPP. In a different application, utilizing the S-C≡N reporter, Londergan and coworkers\textsuperscript{120} mapped the disorder-to-order transition of the intrinsically disordered C-terminal domain of the measles virus nucleoprotein (N\textsubscript{TAIL}) upon binding to another protein, which triggers structure formation in N\textsubscript{TAIL}. They demonstrated that the observed vibrational shifts report on both the local gain in secondary structure and on changes in the hydrophobicity of the microenvironment in the vicinity of the probe upon transition from the disordered to the ordered state.

**Probing the structure and orientation of membrane-bound peptides**

Using a series of \textit{p\textendash}Phe\textsubscript{CN}-derivatized mutants of a membrane-binding peptide (i.e., mastoparan \texttimes) in conjunction with IR polarization measurements, Tucker \textit{et al.}\textsuperscript{121} demonstrated that the nitrile stretching band could be used to report on the location and structure of membrane-bound helical peptides. The successful implementation of this application is based on the notion that in an \textalpha\textendash{}helix the orientations of individual sidechains follow a specific pattern and thus, by measuring the linear dichroic ratio of the C≡N stretching band of representative peptide mutants, it is possible to provide a highly quantitative structural description of the membrane-bound peptide of interest. Similarly, using \textit{p\textendash}Phe\textsubscript{CN} as an IR environmental reporter Mukherjee \textit{et al.}\textsuperscript{122} studied how mastoparan \texttimes interacts with another model membrane system, AOT reverse micelles (RMs). Their results indicated that in low water loading RMs mastoparan \texttimes forms an \textalpha\textendash{}helix with both the backbone and hydrophobic sidechains mostly dehydrated, and more interestingly, that the \textalpha\textendash{}helix orients itself in such a manner that its hydrophobic face is directed towards the water pool.

**Probing light-induced conformational changes in membrane proteins**

Using the azido stretching vibration as an IR probe, Ye \textit{et al.}\textsuperscript{123} studied the conformational changes of rhodopsin induced by retinal photoisomerization. This process is known to trigger a cascade of protein conformational events from the dark state involving discrete and spectroscopically distinguishable states, including the inactive states Lumi and Meta I, as well as the active state Meta II (for rhodopsin). They found that the azido stretching frequency of \textit{p\textendash}Phe\textsubscript{N3}, which was incorporated at position 250 on helix 6 (H6), does not show any detectable changes between the dark and Lumi states, but shows a red-shift when the Meta I state is populated. Combined with results obtained with other azido-derivatized versions of the protein, they concluded that H6 in the Meta I is rotated from its original alignment in the dark state, which brings the azido group from a hydrophilic to a hydrophobic environment. This study highlights the usefulness of nitrile and azide based probes in the study and understanding of the structure-dynamics-function relationship of proteins.
Other Spectroscopic Applications

Some of the aforementioned vibrational probes also find interesting applications in other forms of spectroscopy, an added advantage that most probes do not offer. Below we summarize the relevant fluorescence and NMR applications.

Fluorescence and FRET applications

Besides its application as a vibrational probe, \( p\text{-PheCN} \) can also be used as a fluorescent probe due to its appreciable quantum yield of \( \sim 0.11 \) in water.\textsuperscript{124,125} Tucker \textit{et al.}\textsuperscript{124} found that the fluorescence intensity of \( p\text{-PheCN} \) is very sensitive to solvent environment and can be selectively excited in the presence of other aromatic amino acids. For example, they found that the fluorescence intensity of \( p\text{-PheCN} \) is about ten times lower in acetonitrile than in water and hence could be used to determine the binding constant of peptide-protein or protein-protein complexes. Recently, the photophysics of \( p\text{-PheCN} \) was investigated in more detail by Serrano \textit{et al.}\textsuperscript{125} as well as Raleigh and coworkers.\textsuperscript{126,127} Utilizing \( p\text{-PheCN} \) fluorescence, Tang \textit{et al.}\textsuperscript{128} probed the folding dynamics of the membrane binding peptide \( \text{anti-\( \alpha \)IIb} \). Following rapid mixing of \( \text{anti-\( \alpha \)IIb} \) with model membranes using stopped-flow, they found three distinct phases in the observed fluorescence signal, which were attributed to the kinetic steps of binding, insertion, and dimerization of the peptide in the membrane. Interestingly, they found that the helix-helix association process occurs on the order of a few seconds, which is substantially slower compared to the folding of the coiled-coil motif in solution as well as orders of magnitudes slower than that expected for a diffusion limited reaction, suggesting that the helix-helix association rate is determined by the actual process itself, which requires proper backbone orientations and well-defined intermolecular side chain-side chain interactions.

Moreover, Tucker \textit{et al.}\textsuperscript{129} demonstrated that \( p\text{-PheCN} \) can be used as a fluorescence resonance energy transfer (FRET) donor to tryptophan (Trp). With a Förster radius of approximately 16.0 Å, this FRET pair is ideally suited to probe distances within small proteins and peptides and has been used to probe the conformational distribution of unstructured peptides as well as protein unfolding.\textsuperscript{130,131} More recently, Rogers \textit{et al.}\textsuperscript{131} expanded the FRET application of \( p\text{-PheCN} \) by including FRET acceptors consisting of other unnatural aromatic amino acids, such as 5-hydroxytryptophan and 7-azatryptophan, showing the possibility of performing more sophisticated (e.g., three-step) FRET measurements using unnatural amino acid FRET pairs. The studies of Taskent-Sezgin \textit{et al.} and Waegle \textit{et al.} indicated that tyrosine\textsuperscript{126} and 5-Trp\textsubscript{CN}\textsuperscript{132} are also effective quenchers of the fluorescence of \( p\text{-PheCN} \), further expanding the potential applications of this versatile spectroscopic probe.

NMR applications

The chemical shift of the carbon atom \( (^{13}\text{C}) \) in nitriles is located in an uncongested region of the NMR spectrum of proteins and is sensitive to its environment.\textsuperscript{133} Similarly, the chemical shift of the nitrile nitrogen \( (^{15}\text{N}) \) is found to be sensitive to solvent with the relation that the chemical shift decreases with increasing solvent acceptor number. In addition, Brewer and coworkers\textsuperscript{134} have recently shown, using \( ^{15}\text{N}\)-labeled 2'-azido-2'-deoxyuridine as an example, that the chemical shift of azide is also sensitive to its environment. Taken together, these previous studies indicate that both nitrile and azide groups could serve as site-specific NMR probes of the local environment of biomolecules. For example, Waldman \textit{et al.}\textsuperscript{12} and Doherty \textit{et al.}\textsuperscript{13} have both used \( ^{13}\text{C} \) cyanlated cysteines to qualitatively examine how ligand binding affects the environment of protein binding sites, while Boxer and workers\textsuperscript{117} have used nitrile NMR chemical shifts to facilitate a quantitative assessment of changes in the electrostatic field of the binding pocket of ketosteroid isomerase upon ligand association.
Potential Structural and Energetic Perturbations

Despite their small sizes, incorporation of nitrile, thiocyanate, or azide probes into proteins will lead to unavoidable perturbations of the native properties of the system of interest, such as the structure, stability and enzymatic activity. To address this concern, several studies have been carried out to characterize the degree of structural and energetic perturbations of various nitrile and/or azide probes in detail.\textsuperscript{127,135-137} For example, Raleigh and coworkers\textsuperscript{127} investigated how azidohomoalanine affects the folding stability of the protein NTL9 and found that its destabilizing effect (20–40\%) is dependent on the position of the probe. Similarly, Romesberg and coworkers\textsuperscript{137} investigated the energetic and structural perturbations of $p$-Phe$_{\text{CN}}$ on cytochrome c, while Boxer and coworkers\textsuperscript{136} showed that incorporation of $\beta$-Ala$_{\text{SCN}}$ into RNase results in only small decreases in both $K_{\text{m}}$ and $k_{\text{cat}}$, as well as minor perturbations to the global and local structures of the protein as judged by crystallographic data. Taken together, these previous studies indicate that incorporation of a nitrile or azide probe into proteins, when in the form of an unnatural amino acid, causes context dependent structural and energetic perturbations that could be minimized.

Summary and Outlook

Nitriles, thiocyanates and azides represent powerful spectroscopic probes for the investigation of structure and dynamics of a range of biological and biochemical systems. Each of the presented probes has intrinsic advantages and disadvantages and the best choice for a specific probe depends on the particular system of interest and application. In addition, previous works have mostly focused on testing the applicability of these probes, rather than elucidating specific questions. Therefore, we expect an increase in application of these probes to a wide range of biological and biophysical questions. Below we offer some potential future areas of interest in this rapidly expanding field.

Using two unnatural IR probes to infer distance

The study of Zanni and coworkers\textsuperscript{102} suggested that coupling between vibrational dipoles of two nearby nitriles might be used to probe short distances in biomolecules. However, further nonlinear experiments that can directly probe vibrational couplings\textsuperscript{138} of nitriles are required to substantiate the utility of this method. Recently, using model compounds, Rubtsov and coworkers\textsuperscript{61} have shown that vibrational energy transfer via through-bond relaxation could be used to infer distances between vibrators whose separation is beyond the length scale for which appreciable direct anharmonic coupling occurs. Thus, another direction worth exploring is to determine the feasibility of this technique for protein conformational studies using, for example, peptide systems that contain either two nitriles (Figure 3) or one nitrile and one azide group (Figures 4).

Probing local dynamics and dynamic heterogeneity of proteins

Several recent time-resolved studies\textsuperscript{137,139-141} have shown that nitrile and azide probes are extremely useful in revealing the heterogeneity of local protein dynamics or the dynamics of interfacial water molecules. Thus, it is expected that application of these probes in conjunction with various linear and nonlinear IR methods will continue to provide new insights into the relation between structure, dynamics and function of proteins. In this context, however, new challenges arise. Obtaining reliable spectroscopic signals from some of the probes presented here in photon echo experiments and also time-resolved linear vibrational spectroscopies has proven to be difficult owing to their comparatively low extinction coefficients. Utilization of azide reporters with larger extinction coefficients should mitigate this problem (Table 1). However, a larger extinction coefficient comes at the expense of a larger dipole moment (Eq. 3), which in turn may give rise to a larger
perturbation of the system. Thus, a trade-off exists between signal strength and energetic and structural perturbation. We anticipate that these considerations will become increasingly more important as the field is moving towards the investigation of more complex systems.

Developing new probes and applications

While the aforementioned probes are useful in many applications, they cannot possibly meet all needs. Thus, we expect that the pool of site-specific spectroscopic probes\textsuperscript{142} will continue to expand. For example, it would be very beneficial to develop a site-specific IR reporter of protein backbones. Among the potential strategies, amide to ester backbone mutations\textsuperscript{143} are especially worth pursuing because the CO stretching vibration of esters has a large molar absorption coefficient and is centered in the range of 1700–1800 cm\textsuperscript{-1} where other functional groups do not absorb. Another interesting direction we believe is worth exploring is to quantitatively examine the factors that control the signature of the Fermi resonance of aryl azides. While in many spectroscopic applications Fermi resonances often pose a nuisance, their signatures could be a very useful probe of protein structure and environment. Furthermore, Spiro and coworkers\textsuperscript{144} have shown that \textit{p}-PheCN also exhibits strong Raman scattering, suggesting that the vibrational probes discussed above are expected to find important Raman applications, especially in cases where IR measurements are not feasible (e.g., in solids or strongly scattering samples and surfaces).

Developing new computational and theoretical methods

We can also expect that continued effort will be made to refine existing computational methods and/or develop more advanced methods for easy and accurate calculation of the vibrational property of the probe of interest, which would facilitate a more detailed interpretation of experiments.\textsuperscript{92} Further exploration of the mechanistic origins\textsuperscript{64,65,145} of vibrational frequency shifts will also be helpful in both refining computational methods and interpreting experimental results.

**QUOTES**

- One of the advantages of using unnatural amino acid spectroscopic probes is that they often induce minimal perturbation to the protein structure in question and can be incorporated into the polypeptide sequence of interest using either \textit{in vitro} or \textit{in vivo} methods.

- Nitriles, thiocyanates and azides represent powerful spectroscopic probes for the investigation of structure and dynamics of a range of biological and biochemical systems. Each of the presented probes has intrinsic advantages and disadvantages and the best choice for a specific probe depends on the particular system of interest and application.

- In addition, previous works have mostly focused on testing the applicability of these probes, rather than elucidating specific questions. Therefore, we expect an increase in application of these probes to a wide range of biological and biophysical questions.

Acknowledgments

We thank the National Institutes of Health (GM-065978, RR01348 and GM-008275) for funding. R.M.C. is a Structural Biology Training Grant Fellow.
Biography

Matthias M. Waegele obtained his Ph.D. degree in Physical Chemistry from the University of Pennsylvania in 2011. He is currently a postdoctoral fellow at Lawrence Berkeley National Laboratory, where he investigates the mechanism of the water oxidation reaction on photo-driven transition metal oxide catalyst surfaces.

Robert M. Culik is a Ph.D. candidate in the graduate group of Biochemistry and Molecular Biophysics at the University of Pennsylvania. He graduated with a B.S. in Molecular Biology from the University of Pittsburgh in 2008. His research interests include developing new methods for studying complex folding dynamics.

Feng Gai is a Professor of Chemistry at the University of Pennsylvania. His current research interests focus on protein folding, structure-dynamic-function relationship, as well as the development of new spectroscopic probes and methods. http://gailab4.chem.upenn.edu/

REFERENCES


Figure 1.
FTIR spectra (in the C≡N stretching region) of β-Ala$_{CN}$ and p-Phe$_{CN}$ in H$_2$O and Fmoc-β-
Ala$_{CN}$ and Fmoc-p-Phe$_{CN}$ in THF, as indicated (reproduced from Ref. 26 with permission).
Figure 2.
The C≡N stretching bands of a calmodulin (CaM) binding peptide (MLCK$_{3CN}$) in the presence and absence of CaM, showing the sensitivity of this vibration to its environment (reproduced from Ref. 26 with permission).
Figure 3.
FTIR spectrum of the peptide Phe$_{\text{CN}}$-Ala-Trp$_{\text{CN}}$ in THF at 20 °C, showing that the intensity of the C≡N stretching band of Trp$_{\text{CN}}$ is larger than that of Phe$_{\text{CN}}$. 
Figure 4.
FTIR spectrum of the peptide Phe$_{\text{CN}}$-Ala-Phe$_{\text{N3}}$ in water at 20 °C, showing the relative oscillator strengths of nitrile and azide. The peptide concentration was ~2 mM and the optical pathlength was ~70 μm.
Table 1

Overview of nitrile and azido probes.

<table>
<thead>
<tr>
<th>Name (Abbreviation)</th>
<th>Structure</th>
<th>Band Position (cm⁻¹) (Band Width)</th>
<th>Intensity (M⁻¹ cm⁻¹) (Solvent, Ref.)</th>
<th>Application (Ref.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Para-cyanophenylalanine (p-PheCN)</td>
<td><img src="image" alt="Para-cyanophenylalanine" /></td>
<td>2237.2 (9.8)</td>
<td>2228.5b (5.0)</td>
<td>~220 (THF, 26)</td>
</tr>
<tr>
<td>Para-cyanobenzylcysteine (p-CysB-CN)</td>
<td><img src="image" alt="Para-cyanobenzylcysteine" /></td>
<td>2236.6 (11.8)</td>
<td>2228.5 (7.4)</td>
<td>210 ± 60 (H₂O, 15)</td>
</tr>
<tr>
<td>Para-cyanophenylcysteine (p-CysP-CN)</td>
<td><img src="image" alt="Para-cyanophenylcysteine" /></td>
<td>2233.7 (11.2)</td>
<td>2226.8 (6.1)</td>
<td>240 ± 60 (H₂O, 15)</td>
</tr>
<tr>
<td>5'-cyanotryptophan (5-Trp-CN)</td>
<td><img src="image" alt="5'-cyanotryptophan" /></td>
<td>2224.6c (15.6)</td>
<td>2220.4 (9.3)</td>
<td>~450 (2-MeTHF at 74 K, 57)d</td>
</tr>
<tr>
<td>Beta-cyanoalanine (β-Ala-CN)</td>
<td><img src="image" alt="Beta-cyanoalanine" /></td>
<td>2261.9 (18.2)</td>
<td>2252.1b (11.0)</td>
<td>32.2 (H₂O, 26)</td>
</tr>
<tr>
<td>Beta-Thiocyanatoalanine (β-AlaSCN)</td>
<td><img src="image" alt="Beta-Thiocyanatoalanine" /></td>
<td>2162.0 (12.8)</td>
<td>~2158</td>
<td>153.8 (H₂O, 67)</td>
</tr>
<tr>
<td>5'-Cyano-2' deoxyuridine (dUCN)</td>
<td><img src="image" alt="5'-Cyano-2' deoxyuridine" /></td>
<td>2241.9 (~11.5)</td>
<td>2232.7 (8)</td>
<td>332 (2-MeTHF at 74 K, 150)</td>
</tr>
<tr>
<td>Name (Abbreviation)</td>
<td>Structure</td>
<td>Band Position (cm(^{-1})) (Band Width)</td>
<td>Intensity (M(^{-1}) cm(^{-1}))(^a) (Solvent, Ref.)</td>
<td>Application (Ref.)</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------</td>
<td>------------------------------------------</td>
<td>-------------------------------------------------</td>
<td>-------------------</td>
</tr>
<tr>
<td>N2-nitri-2’-deoxyguanosine (dG-CN)</td>
<td><img src="image" alt="Structure" /></td>
<td>n.a.</td>
<td>2170 (29)(^e)</td>
<td>412 (2-MeTHF at 74 K, 150)</td>
</tr>
<tr>
<td>2’-Azido-2’-deoxuridine (N(_3)-dU)</td>
<td><img src="image" alt="Structure" /></td>
<td>2124 (22)</td>
<td>2110 (18)</td>
<td>1160 ± 150 (H(_2)O, 77)</td>
</tr>
<tr>
<td>Para-Azidophenylalanine (p-Phe(_{N3}))</td>
<td><img src="image" alt="Structure" /></td>
<td>~2128.6</td>
<td>~2115.5(^f)</td>
<td>n.a.</td>
</tr>
<tr>
<td>Beta-Azidoalanine (β-Ala(_{N3}))</td>
<td><img src="image" alt="Structure" /></td>
<td>2118.4 (27.8)</td>
<td>2104.6 (25.6)(^g)</td>
<td>414 (H(_2)O, 152)</td>
</tr>
<tr>
<td>Azidohomoalanine (Aha)</td>
<td><img src="image" alt="Structure" /></td>
<td>2115 (~38)</td>
<td>2098 (~38)(^h)</td>
<td>1570 (H(_2)O, 127)</td>
</tr>
<tr>
<td>Azido-NAD(^+)</td>
<td><img src="image" alt="Structure" /></td>
<td>2140 (33)</td>
<td>2138 (26)(^i)</td>
<td>250 (H(_2)O, 135)</td>
</tr>
</tbody>
</table>

\(^{a}\) Peak extinction coefficient was estimated based on reported values of concentration and optical pathlength. The reader might find it useful to compare the coefficients with the one for the amide I vibration of 720 M\(^{-1}\) cm\(^{-1}\) (Ref. 152)

\(^{b}\) measurements in THF were performed with the corresponding Fmoc-versions of the amino acids

\(^{c}\) measurements in water were performed in 95:5 mixture of water/methanol on 5-cyanoindole
5-cyanoindole

e measured in 2-MeTHF

f measured in isopropanol

g measured in DMSO

h measured in DMF

i measured on FDH-azido-NAD⁺ complex in H₂O.