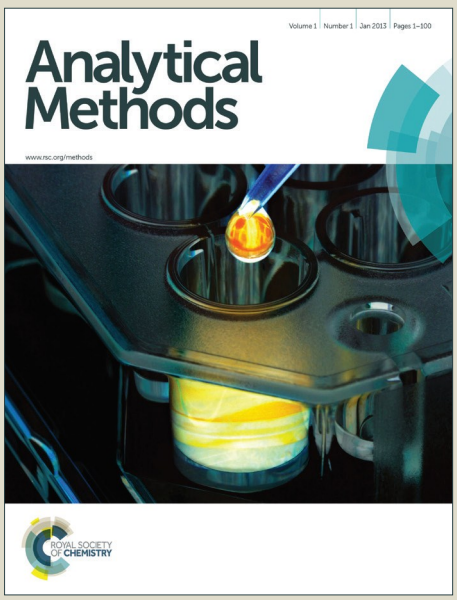
Analytical Methods

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the Information for Authors.

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard Terms & Conditions and the Ethical guidelines still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/methods

Site-selective Characterization of Src Homology 3 Domain Molecular Recognition with Cyanophenylalanine Infrared Probes.

Rachel E. Horness, Edward J. Basom, and Megan C. Thielges

Local heterogeneity of microenvironments in proteins is important in biological function, but difficult to characterize experimentally. One approach is the combination of infrared (IR) spectroscopy and site-selective incorporation of probe moieties with spectrally resolved IR absorptions that enable characterization within inherently congested protein IR spectra. We employed this method to study molecular recognition of a Src homology 3 (SH3) domain from the yeast protein Sho1 for a peptide containing the proline-rich recognition sequence of its physiological binding partner Pbs2. Nitrile IR probes were introduced at four distinct sites in the protein by selective incorporation of *p*-cyanophenylalanine via the amber codon suppressor method and characterized by IR spectroscopy. Variation among the IR absorption bands reports on heterogeneity in local residue environments dictated by the protein structure, as well as on residue-dependent changes upon peptide binding. The study informs on the molecular recognition of $SH3^{Sho1}$ and illustrates the speed and simplicity of this approach for characterization of select microenvironments within proteins.

1. Introduction

The structures of proteins create highly heterogeneous microenvironments. The complex spatial variation in local electrostatic fields and interactions throughout a protein presents challenges in understanding their folding, catalysis, protein-protein interactions, and other aspects of function,[1] and has motivated efforts toward the site-specific characterization of the local environments

throughout proteins.[2-18] Infrared (IR) spectroscopy provides a route to the direct characterization of the structural degrees of freedom of a molecule with bond-specific spatial resolution and an inherent temporal resolution that ensures detection of very rapidly exchanging species. Unfortunately, the spectral congestion inherent to such large macromolecules hinders the application of IR spectroscopy to proteins.

To overcome the spectral congestion, researchers have taken advantage of extrinsic IR probes that have absorptions around 1900-2500 cm^{-1}, a region of the IR spectrum free of intrinsic protein absorptions. The first such studies utilized small molecule probes, such as carbon monoxide, azide, and cyanide, which spontaneously bind heme-containing proteins.[3-5] Side chain labeling with carbon-deuterium bonds was later performed to incorporate site-specific IR probes in proteins.[6] Although strictly non-perturbative, the absorptions of carbon-deuterium bonds are weak and so are challenging to characterize. In contrast, azide-derivatized side chains provide very intense absorption bands, but suffer from spectral complexity due to Fermi resonances.[7,19] In practice, nitrile-functionalized amino acids provide a useful compromise as IR probes. Compared to carbon-deuterium bonds, their absorption bands are relatively intense; but compared to azides, their small size is minimally perturbative, and spectral interpretation of their IR absorption bands is relatively straightforward.[8,20-22]

Several routes exist for the site-selective incorporation of nitriles into proteins. A cyano-derivatized amino acid may be directly incorporated during total synthesis or incorporated into a peptide and built into a larger protein via semisynthesis employing native chemical or expressed protein ligation.[23,24] Although in principle any unnatural amino acid might be incorporated in this manner, for even moderately sized proteins this route can be time consuming and expensive. Alternately, chemical modification of cysteine residues can generate thiocyanates, as first

described by the Boxer group.[9,10] Although a relatively easy route to nitrile incorporation, many proteins contain multiple cysteine residues, and so require extensive mutagenesis to create a unique cysteine to achieve selective labeling at individual sites.

Site-selective introduction of nitriles into proteins is also possible by incorporation of *p*-cyanophenylalanine (*CN*Phe) via the amber codon suppressor method, developed by Schultz and coworkers.[25-27] In general, this route relies on amber suppressor transfer RNA ($tRNA^{CUA}$)-tRNA synthetase pairs, where the $tRNA^{CUA}$ is selectively charged with an unnatural amino acid, which it then incorporates into a protein during ribosomal synthesis in response to an amber codon (TAG). The amber codon can be introduced at any desired position in the gene of interest by site-directed mutagenesis to enable site-selective incorporation of *CN*Phe virtually anywhere in a protein. Via this method, a particular residue can be targeted for labeling without introducing additional mutations, and the labeled proteins are otherwise produced as in a typical recombinant expression, which is relatively rapid, cheap, and high yielding. Of the available nitrile-derivatized amino acids, the aromatic *CN*Phe also provides the most intense absorption signals. It has been successfully utilized as an IR probe in a number of studies, for example, of protein folding, ligand binding, and membrane insertion.[8,11-16]

In this study, we explored *CN*Phe as a vibrational probe of the Src homology 3 domain from the Sho1 protein of *Saccharomyces cerevisiae* ($SH3^{Sho1}$) that recognizes the protein Pbs2 as part of the yeast osmotic stress pathway (Fig. 1).[28] SH3 domains recognize proline-rich sequences to mediate protein-protein interactions that underlie myriad cellular processes including signaling, cytoskeletal remodeling, and development.[29] As one of the most prevalent domains in human and other eukaryotic proteomes, SH3 domains have emerged as archetypal models for the study of biological molecular recognition. We employed the amber codon

suppression methodology to specifically incorporate *CN*Phe as an IR probe of SH3^{Sho1} at four locations: Tyr5, Tyr11, Phe25 and Tyr57. IR spectroscopy was then used to generate a site-specific picture of the changes in SH3^{Sho1} at these sites upon binding a peptide consisting of the Pbs2 proline-rich recognition sequence (Pbs2 peptide). Because *CN*Phe is also a potential fluorescent probe,[11,12] we additionally characterized the SH3^{Sho1} variants with fluorescence spectroscopy. The study suggests that *CN*Phe is an excellent IR probe of local side chain hydration, and demonstrates that a combination of the amber codon labeling approach with IR spectroscopy provides a relatively quick and easy approach to the study of local environments in proteins.

2. Experimental

2.1 SH3^{Sho1} Expression and Purification

DNA encoding the wild-type SH3^{Sho1} domain was generously supplied by the laboratory of Alan Davidson (University of Toronto).[30] The gene of interest had been ligated into the pet21d+ vector (Novagen) between the Nco1 and Xho1 restriction sites such that the protein domain would be expressed with a C-terminal hexahistidine (His_6) tag. A Phusion Site-Directed Mutagenesis Kit (Thermo Scientific) was used to incorporate a thrombin cleavage recognition sequence between the C-terminus of the SH3 domain and the terminal His_6 tag. For SH3^{Sho1} variants, plasmids were prepared with appropriate mutations to TAG codons at the Tyr5, Tyr11, Phe25, and Tyr57 sites on the wild-type SH3^{Sho1} template DNA with standard site-directed mutagenesis methods and a Stratagene Site-Directed Mutagenesis kit (Agilent). The presence of all desired and lack of undesired mutations was confirmed by sequencing. The pUltraCNF

plasmid which encodes the orthogonal tRNA synthetase and tRNA for amber codon incorporation of *CN*Phe was generously provided by Peter Schultz (Scripps Research Institute).[27]

For wild-type expression, the plasmid was first transformed into BL21 (DE3) *E. coli*. A single colony was then cultured for 12 hours in 4 mL of Luria-Bertani medium (LB) at 37 °C in the presence of 100 μg/mL ampicillin. Subsequently, 100 μL of culture were added to 250 mL of LB supplemented with the same concentration of antibiotics and grown overnight at 37 °C. Five mL of overnight culture were added to 1 L of LB supplemented with ampicillin and grown to OD_{600} of 0.6-0.8. Expression was induced with 1mM IPTG for 4 to 6 hours. Cells were isolated by centrifugation, resuspended in 50 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, pH 8.0, and lysed with lysozyme treatment and by sonication. The lysate was clarified by centrifugation and added to NiNTA affinity media (Qiagen) and the slurry was rocked gently on ice. Column media was washed with three volumes of 50 mM sodium phosphate, 300 mM NaCl, 20 mM imidazole, pH 8.0, followed by an equal volume of the same buffer containing 250 mM imidazole, to elute the bound SH3. Eluted fractions were dialyzed into phosphate-buffered saline (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, pH 7.4) before cleavage with thrombin (Novagen). Cleavage was allowed to proceed for 16 hours at room temperature at a concentration of 1U thrombin/mg of protein. The uncleaved protein was removed by passage over NiNTA media, as described above, followed by size exclusion chromatography.

The expression of each $SH3^{Sho1}$ variant was similar to the wild-type procedure with a few modifications. The pUltraCNF plasmid was co-expressed with those containing the SH3 gene with each TAG mutation. All variants were expressed in Terrific Broth supplemented with ampicillin (100 μg/mL) and streptomycin (60 μg/mL) in the presence of 1 mM *CN*Phe

(ChemPep Inc.), which was added at OD_{600} of 0.1-0.2. Expression was induced at OD_{600} of 0.6-0.8 with 0.2 mM IPTG for 13 (*CN*Phe25) to 24 hours (*CN*Phe11 and *CN*Phe57) or 1 mM IPTG for 20 hours (*CN*Phe5). Isolation of cells and subsequent purification proceeded as described above.

2.2 Pbs2 Peptide Synthesis

The Pbs2 peptide binding partner of $SH3^{Sho1}$ was synthesized with the sequence acetyl-VNKPLPPLPVA-NH_2 on an Applied Biosystems 433A peptide synthesizer using standard Fmoc solid-phase peptide synthesis. Amino acids and other materials were generously provided by the laboratory of Richard DiMarchi (Indiana University). The product was purified on a Luna C18 reversed-phase HPLC column (Phenomenex) and the identity of the peptide was confirmed by mass spectrometry. To determine Pbs2 peptide concentration, one stock of dissolved peptide was analyzed by both absorbance spectroscopy and amino acid analysis (AstraOmics). Knowledge of both the absorbance and concentration allowed determination of the extinction coefficient at 205 nm (ε = 43,318.56 $M^{-1}cm^{-1}$) through Beer's law. This extinction coefficient was used in all further measurements of Pbs2 peptide concentration.

2.3 $SH3^{Sho1}$ Characterization

Circular dichroism spectra of the variants were obtained to assess potential perturbation to secondary structure that may have resulted from *CN*Phe incorporation. Spectra were acquired at a concentration of 0.01 mM protein in 10 mM sodium phosphate pH 7.0 in a 1 mm cuvette. All results are an average of 5 scans with subtraction of a buffer spectrum (Supplementary Information).

Additionally, intrinsic fluorescence assays were performed to assess perturbation of the binding affinity for Pbs2. All assays were performed with 3 μM protein solutions. Pbs2 peptide was added at concentrations of 0, 1, 2, 5, 10, 25, and 50 μM, and samples were equilibrated overnight at 4 °C before acquiring fluorescence spectra. Samples were excited at 280 nm and the fluorescence emission spectra were recorded from 290 to 470 nm in triplicate. The fluorescence intensity at 350 nm for the samples of varying Pbs2 concentration was fit assuming a single-site binding model (equation 1):

$$\frac{\Delta F}{\Delta F_{max}} = \frac{PL}{P_t} = \frac{(L_t+P_t+K_d)-\sqrt{(L_t+P_t+K_d)^2-(4*P_t*L_t)}}{2*P_t} \qquad (1)$$

where ΔF is the change in fluorescence at 350 nm with respect to 0 μM peptide, ΔF_{max} is the maximum observed change in fluorescence intensity, PL is the concentration of bound complex, P_t is the total protein concentration, L_t is the total peptide concentration, and K_d is the dissociation constant. Fluorescence spectra were also obtained with 240 nm excitation for the 0 and 50 μM Pbs2 peptide samples. Both emission and excitation slit widths were held at 5 nm for all data collection.

2.4 IR Spectroscopy

Protein samples for Fourier transform (FT) IR spectroscopy were concentrated to 1.5-2.0 mM and loaded between CaF_2 windows with a 38.1 μm Teflon spacer. FT IR spectra were recorded at 4 cm^{-1} resolution with an Agilent Cary 670 FT IR spectrometer and $N_{2(l)}$-cooled MCT detector. The instrument was purged with dry N_2 for 30 minutes before any data were collected. A 4-term Blackman Harris apodization function and zero-filling by a factor of 8 were applied to process all interferograms. Background spectra of the wild-type protein were subtracted from the spectra of *CN*Phe-labeled samples. A polynomial fit to the baseline was subtracted from

each of the spectra to correct for the slowly varying residual background absorbance. Baseline-corrected spectra were fit with a single Gaussian function (Supplementary Information). Spectra were acquired of at least three independent samples and all spectra were averaged over 10,000 scans. FT IR spectra of Pbs2 peptide-bound samples were taken in an identical manner with protein and peptide concentrations of 1.5 and 1.53 mM, respectively.

3. Results and Discussion

We introduced *CN*Phe as a vibrational probe at four individual sites in $SH3^{Sho1}$ via the *in vivo* amber codon suppression method. *CN*Phe11 and *CN*Phe57 are located along the binding interface with the Pbs2 peptide, *CN*Phe25 lies in the hydrophobic core, and *CN*Phe5 is distant from the peptide-binding surface and was chosen as a control (Fig. 1, PDB ID: 2VKN). All *CN*Phe-labeled proteins were obtained in relatively high yields: 12 mg/L (*CN*Phe5, *CN*Phe11), 2.5 mg/L (*CN*Phe57), and 34 mg/L (*CN*Phe25), compared to 40 mg/L for wild-type. Complete labeling at the desired sites except *CN*Phe25, and the absence of unwanted mutations, was confirmed by mass spectrometry of tryptic digests (Supplementary Information). The mass spectra for the *CN*Phe25 sample, however, do indicate the presence of $SH3^{Sho1}$ with the wild-type sequence. This variant was also obtained in the highest yield. The results suggest that the recognition specificity of the synthetase for *CN*Phe over Phe is likely lower compared to Tyr, resulting in charging the tRNA with Phe instead of *CN*Phe. Nonetheless, because the wild-type protein does not absorb in the IR spectral region of interest, its presence only results in an apparent decrease in the *CN*Phe25 absorption band intensity. Comparing the band intensity of the *CN*Phe25 sample with those from the other variants, we estimate ~60-70% *CN*Phe incorporation at Phe25. The circular dichroism spectra of the labeled $SH3^{Sho1}$ show no

significant differences from wild-type, indicating that incorporation of *CN*Phe does not highly perturb the secondary structure (Supplementary Information).

Fluorescence spectroscopy

In addition to an IR probe, *CN*Phe has been utilized as a fluorescent probe of environments in proteins.[11,12,31] Thus we also characterized the wild-type and *CN*Phe-labeled variants in both unligated SH3^{Sho1} and the Pbs2 peptide-bound states with fluorescence spectroscopy using 280 nm or 240 nm excitation to preferentially excite the native Trp or *CN*Phe residues, respectively (Fig. 2). In all cases the fluorescence intensity at 350 nm increases upon binding, and no significant binding-induced differences are observed among the variants. With 240 nm excitation, fluorescence intensity at 295 nm characteristic of *CN*Phe is observed for all but the *CN*Phe11 variant. The unique quenching of *CN*Phe11 fluorescence cannot be explained by its location at the protein surface, where *CN*Phe57 and *CN*Phe5 are also located, nor by energy transfer to Trp residues, as *CN*Phe11 is the most distant of the labeled residues (Fig. S2) Interpretation of the fluorescence spectra of *CN*Phe when incorporated into proteins does not appear straightforward. (See Supplementary Information for a more detailed discussion.)

Fluorescence-based binding assays were performed to assess perturbation of the affinity of SH3^{Sho1} for the Pbs2 peptide ligand (Fig. 3). Incorporation of *CN*Phe11 and *CN*Phe57, but not *CN*Phe25 or *CN*Phe5, slightly decreases the binding affinity by ~1.5-fold (in addition, the change at *CN*Phe57 is within error). The indifference to *CN*Phe5 incorporation is expected, as the residue is distant from the binding interface. Similarly, *CN*Phe25 lies in the hydrophobic core of SH3^{Sho1} away from the binding surface and was not anticipated to hinder binding. In contrast, *CN*Phe11 and *CN*Phe57 are at the binding interface. Tyr57 serves as a hydrogen bond donor to

the peptide backbone carbonyl in the structure of the SH3^{Sho1}-Pbs2 peptide complex (PDB ID: 2VKN). Substitution of the hydroxyl group with a nitrile removes the donor ability and as such the slight reduction in binding affinity is not surprising. Tyr11 also lies along the binding interface and participates in van der Waals interactions involving packing between the aromatic ring and the side chains of Pro9' and Leu8' of the Pbs2 peptide. The hydroxyl group of Tyr11 does not appear to participate in any hydrogen bonding interactions with the peptide in the crystal structure, so the mild perturbation of SH3^{Sho1}-Pbs2 peptide binding due to incorporation of *CN*Phe11 likely arises from altered packing, rather than hydrogen bonding interactions. Overall, our data show that incorporation of *CN*Phe is at most only mildly perturbative to binding and thus did not preclude characterization of the bound complex.

IR spectroscopy

FT IR spectroscopy was then applied to characterize the microenvironment at each incorporated *CN*Phe. The spectra of all variants show relatively intense absorption bands around 2235 cm^{-1}, characteristic of the nitrile stretching frequency of *CN*Phe (Fig. 4).[8,12-18] Although the line shapes are best described as Voigt due to contributions from both homogeneous and inhomogeneous line broadening mechanisms, the bands are reasonably well fit with a single Gaussian function (Supplementary Information). All absorption bands also appear symmetrical about the center frequency. These observations suggest that a single conformational state is experienced by the side chain at each labeled residue. The Gaussian fit yielded the center frequencies and line widths of the *CN*Phe absorption bands (Table 1). Although the spectral differences among the variants might appear small, the intense bands provided by the *CN*Phe probe enable very high precision in the determination of the center frequencies and line widths. (An average standard deviation of 0.07 cm^{-1} in center frequency was obtained, while the standard

deviation of 0.5 cm^{-1} in line width was slightly greater, primarily due to uncertainty introduced during baseline correction.)

In general, vibrational frequency shifts arise from a change in the electric field at the probe, or in localized, specific interactions with other chemical functional groups in its vicinity.[2] The absorption bands of *CN*Phe in different environments from proteins and/or solvents generally show higher (lower) frequencies when in more nonpolar (polar) environments.[20,22,32] From the crystal structure of the SH3^{Sho1}-Pbs2 peptide complex, we expect that *CN*Phe25 is buried whereas all other *CN*Phe residues are solvent-exposed in the unligated protein. As all residues show red-shifted frequencies compared to *CN*Phe in aqueous solution, where the environment should be more polar than any in the protein, our spectral data are inconsistent with the previously observed electrochromism.[22,32]

Nitriles also are known to be highly prone to hydrogen bond formation with water solvent or functional groups within a protein. Moreover, it has been well demonstrated that their vibrational frequencies are highly sensitive to the hydrogen bonding interactions.[8,22] The formation of a hydrogen bond to water involving the nitrogen lone pair electrons leads to a blue-shift in frequency from electron withdrawal out of σ* antibonding orbitals.[33,34] Computational studies also find that hydrogen bonding involving the nitrile π electrons can occur and lead to an opposite red-shift in frequency, and so cautious interpretation is required.[33] However, previous experimental studies of *CN*Phe in different solvents and protein environments generally show an increase in frequency associated with an increase in the hydrogen bonding potential of the environment and so suggest that the σ-type interaction dominates.[8,16,21,22] Our data also is consistent with this interpretation. Although all residues except *CN*Phe25 are at least partially solvent exposed, it is likely that none of the nitrile groups can form as optimal hydrogen bonding

interactions with water as does the free amino acid due to hindrance by neighboring protein side chains. Thus, the red-shift in the frequencies of all absorption bands of *CN*Phe residues in SH3^{Sho1} is consistent with reduced hydration and/or suboptimal σ-type hydrogen bonding upon incorporation into the protein.

When incorporated at the different positions of unligated SH3^{Sho1}, the absorption of *CN*Phe shows a variation of 2.3 cm^{-1} in center frequency (Table 1). Although these differences might appear small, they are highly significant given the small uncertainty in their determination, and are also similar in magnitude to those typically observed in studies of other proteins employing nitrile probes.[8,12-18] The relatively high frequency found for *CN*Phe57 suggests that, of the labeled residues in unligated SH3^{Sho1}, *CN*Phe57 can form the most optimal hydrogen bonding interactions with water. Thus *CN*Phe57 is likely the most highly hydrated side chain and the least secluded from water by other parts of the protein. The similar frequencies found for *CN*Phe5, the solvent-exposed "control" residue distant from the peptide binding site, and *CN*Phe11, the other residue on the binding surface, suggest similar levels of hydration. Compared to *CN*Phe57, the frequencies of *CN*Phe5 and *CN*Phe11 were red-shifted by 1.7 and 1.6 cm^{-1}, respectively, whereas that of *CN*Phe25 was red-shifted by 2.3 cm^{-1}. The highly red-shifted frequency of *CN*Phe25 likely reflects its lack of participation in hydrogen bonding interactions due to its buried location within the protein.

In addition to differences in their frequencies, the absorption bands of the *CN*Phe residues show variation in line width. Heterogeneity in the environment of *CN*Phe at a site will lead to a distribution in the frequencies of the nitrile vibration and broaden the line width. The line widths for *CN*Phe5 and *CN*Phe25 are greater than those of *CN*Phe11 and *CN*Phe57 by ~2 cm^{-1} (Table 1). Thus, the IR spectra suggest that the environmental heterogeneity is greater at *CN*Phe5 and

*CN*Phe25 than at *CN*Phe11 and *CN*Phe57, which are similar in line width to the free amino acid in aqueous solution. Although *CN*Phe5 is a solvent-exposed residue, the nitrile apparently experiences variation in its environment due to the protein surface. *CN*Phe25 is the only residue buried within the protein. In previous studies employing nitrile probes, narrow line widths of 5-7 cm^{-1} have been observed when the nitrile experiences a homogeneously nonpolar environment within a protein core or lipid bilayer.[17,18] Contrary to this, the broader line width found for *CN*Phe25 suggests that the nitrile group experiences a relatively heterogeneous environment. The side chain might adjust from its buried position in the crystal structure of the native protein to point the nitrile group toward the solvent surface or a more polar group of the protein and create a more varied environment. *CN*Phe25 is also located on the RT loop of $SH3^{Sho1}$, which is believed to be relatively mobile due to a lack of a hydrogen bonding network apparent elsewhere in the domain,[35] which might enable the side chain to sample a broader distribution of states. It should be noted, however, that both homogeneous and inhomogeneous line broadening mechanisms can contribute to line width changes, so an interpretation of a linear spectral line width in terms of heterogeneity is necessarily an approximation. 2D IR studies of the thermal unfolding of *CN*Phe-labeled HP35, however, find that inhomogenous broadening dominates the line width changes associated with the transition from a hydrophobic to a solvent-exposed environment; the homogeneous line width appears to vary more strongly with temperature rather than the extent of solvent exposure.[36] Efforts are underway to employ 2D IR spectroscopy to more rigorously define the observed differences in line widths among the $SH3^{Sho1}$ variants.

We next used FT IR spectroscopy to characterize the changes at each introduced *CN*Phe upon binding the Pbs2 peptide. As expected, the frequency of *CN*Phe5, the control residue located distant from the binding surface, shows only a 0.1 cm^{-1} binding-induced change, a value

similar to the error in the measurement. In contrast, for *CN*Phe57 and *CN*Phe11, binding-induced shifts in frequency of similar magnitude, but opposite direction, were observed. A red-shift of 0.5 cm^{-1} for *CN*Phe57 suggests that the nitrile forms less optimal hydrogen bonding interactions in the peptide-bound state. This is consistent with the crystal structure of the $SH3^{Sho1}$-Pbs2 peptide complex, which shows that the Tyr57 side chain points toward the bound peptide where it should be partially shielded from solvent. In contrast, a blue-shift of 0.7 cm^{-1} for *CN*Phe11 suggests that the residue becomes better hydrated upon binding the Pbs2 peptide. The structure shows that *CN*Phe11 interacts with the peptide primarily via van der Waals interactions involving the aromatic ring, and the nitrile is expected to point out toward solvent. The observation of opposite binding-induced frequency shifts for nitriles introduced at the two locations along the binding surface indicates that the spectral changes depend on the specific interaction of the *CN*Phe residue with the peptide. Finally, the absorption of residue *CN*Phe25 shows the greatest binding-induced change in frequency, blue-shifting by 1.3 cm^{-1}. This is interesting because *CN*Phe25 is not expected to directly contact the peptide ligand, and so the frequency change must be a secondary effect transmitted through changes in other parts of $SH3^{Sho1}$ upon Pbs2 peptide binding.

As with the frequency, no significant changes in line width were observed for the control residue *CN*Phe5 upon binding the Pbs2 peptide. A small increase in the line width for *CN*Phe25 was found, but the difference was not outside of error. Thus, the environmental heterogeneity at both *CN*Phe5 and *CN*Phe25 appears largely unaffected by the Pbs2 peptide binding. In contrast, the line widths *CN*Phe11 and *CN*Phe57 increase by ~1.5 cm^{-1} upon binding to the Pbs2 peptide to become similar to those of the other two *CN*Phe residues. *CN*Phe11 and *CN*Phe57 are the two residues that contact the Pbs2 peptide, and thus it is reasonable that they uniquely show binding-

induced changes in line width. The increase suggests greater variation in the environments experienced by *CN*Phe11 and *CN*Phe57 in the bound complex. This might reflect that they experience similar heterogeneity as in aqueous solvent in the unligated states, and that Pbs2 peptide binding generates local variation in electrostatic fields or hydrogen bonding arrangements. The greater heterogeneity might also reflect a distribution of bound conformations, which suggests an induced-fit rather than a lock-and-key or conformational selection mechanism of binding, as the latter mechanisms would predict either no change or conformational restriction upon binding. The increased bound state heterogeneity is consistent with a previous NMR study which suggested that SH3-peptide complexes can exist in a dynamic equilibrium between fully engaged and partially engaged conformations.[37]

4. Conclusions.

We combined site-selective labeling with *CN*Phe and FT IR spectroscopy to characterize the microenvironments for distinct sites and their binding-induced changes throughout SH3^{Sho1}. The *CN*Phe IR probes report greater variation in local environment due to their location in unligated SH3^{Sho1} than result from Pbs2 peptide binding. The protein structure thus highly dictates the local environment of an amino acid side chain. The relatively moderate frequency changes induced upon Pbs2 binding might simply reflect the small area of the interaction surface with the Pbs2 peptide, and consequently the small potential for binding-induced variation in environment. The small size of the ligand might also present limited structural hindrance for the *CN*Phe side chains to conformationally adjust to orient the nitrile group toward solvent and satisfy its hydrogen bonding propensity, resulting in minimal binding-induced changes in side chain hydration. This is consistent with the observation of the largest binding-induced spectral changes at interior residue *CN*Phe25, which presumably has a more constricted side chain than

do the solvent exposed residues. We are currently working toward investigation of the same sites via introduction of ring-d_2-tyrosine, which, albeit producing much less intense absorption bands, would provide characterization from the perspective of a different part of the same residues and, furthermore, eliminate issues concerning perturbation. Nonetheless, the current study finds that the *CN*Phe absorption is sensitive to ligand binding when placed at all sites except for the control residue, as expected, illustrating its general utility as an IR probe that provides intense IR bands for characterization of microenvironments in proteins.

*CN*Phe provides a spectrally resolved IR probe of its microenvironment regardless of a protein's amino acid composition. In contrast, our fluorescence characterization of the variants, however, finds *CN*Phe of limited utility as a fluorescent probe for SH3^{Sho1} and likewise other proteins containing multiple Trp or other fluorescent amino acids. Although the interpretation of the IR frequencies of nitriles can be complicated by the potential effects of both hydrogen bonding and electrostatics, our results are consistent with previous studies that suggest the former dominate. More extensive analysis of the spectral changes can be achieved via protein modeling.[18] In addition, the combination of IR and NMR spectroscopy to correlate IR frequencies and chemical shifts can clarify spectral interpretation,[21] but requires ^{13}C-labeled nitriles (^{13}*CN*Phe unfortunately is not currently commercially available). Nonetheless, this study demonstrates how the combination of site-selective *CN*Phe labeling using the amber codon approach with FT IR spectroscopy provides a fast, simple, minimally-perturbative, and reproducible approach to characterize the variation in side chain hydration throughout proteins and potential changes associated with protein function.

Acknowledgments

The authors thank John Mayer and Richard DiMarchi at Indiana University for assistance in preparation of the Pbs2 peptide. We also thank Peter Schultz at the Scripps Research Institute for providing plasmid pUltraCNF and Alan Davidson at the University of Toronto for providing an expression plasmid for $SH3^{Sho1}$. This work was supported by Indiana University. R.H. and E.B. were also supported by the Graduate Training Program in Quantitative and Chemical Biology (T32 GM109825).

5. References

1. B. Honig and A. Nicholls, Classical Electrostatics in Biology and Chemistry, *Science*, 1995, **268**, 1144-1149.
2. H. Kim and M. Cho, Infrared probes for studying the structure and dynamics of biomolecules, *Chem. Rev.*, 2013, **113**, 5817-5847.
3. G. N. Phillips Jr., M. L. Teodoro, T. Li, B. Smith and J. S. Olson, Bound CO Is A Molecular Probe of Electrostatic Potential in the Distal Pocket of Myoglobin, *J. Phys. Chem. B*, 1999, **103**, 8817-8829.
4. T. G. Spiro and I. H. Wasbotten, CO as a vibrational probe of heme protein active sites, *J. Inorg. Biochem.*, 2005, **99**, 34-44.
5. S. Yoshikawa, D. H. O'Keeffe and W. S. Caughey, Investigations of cyanide as an infrared probe of hemeprotein ligand binding sites, *J. Biol. Chem.*, 1985, **260**, 3518-3528.
6. J. K. Chin, R. Jimmenez and F. E. Romesberg, Direct Observation of Protein Vibrations by Selective Incorporations of Carbon-Deuterium Bonds in Cytochrome *c*, *J. Am. Chem. Soc.*, 2001, **123**, 2426-2427.
7. I. T. Suydam and S. G. Boxer, Vibrational Stark Effects Calibrate the Sensitivity of Vibrational Probes for Electric Fields in Proteins, *Biochemistry*, 2003, **42**, 12050-12055.
8. J. Zimmermann, M. C. Thielges, Y. J. Seo, P. E. Dawson and F. E. Romesberg, Cyano Groups as Probes of Protein Microenvironments and Dynamics, *Angew. Chem. Int. Ed. Engl.*, 2011, **50**, 8333-8337.
9. A. T. Fafarman, L. J. Webb, J. I. Chuang and S. G. Boxer, Site-Specific Conversion of Cysteine Thiols into Thiocyanate Creates an IR Probe for Electric Fields in Proteins, *J. Am. Chem. Soc.*, 2006, **128**, 13356-13357.
10. K. N. Alfieri, A. R. Vienneau and C. H. Londergan, Using infrared spectroscopy of cyanylated cysteine to map the membrane binding structure and orientation of the hybrid antimicrobial peptide CM15, *Biochemistry*, 2011, **50**, 11097-11108.
11. M. J. Tucker, R. Oyola and F. Gai, A novel fluorescent probe for protein binding and folding studies: p-cyano-phenylalanine, *Biopolymers*, 2006, **83**, 571-576.

12. H. Taskent-Sezgin, J. Chung, V. Patsalo, S. J. Miyake-Stoner, A. M. Miller, S. H. Brewer, R. A. Mehl, D. F. Green, D. P. Raleigh and I. Carrico, Interpretation of p-cyanophenylalanine fluorescence in proteins in terms of solvent exposure and contribution of side-chain quenchers: a combined fluorescence, IR and molecular dynamics study, *Biochemistry*, 2009, **48**, 9040-9046.
13. A. T. Fafarman and S. G. Boxer, Nitrile Bonds as Infrared Probes of Electrostatics in Ribonuclease S, *J. Phys. Chem. B*, 2010, **114**, 13536–13544.
14. P. Marek, S. Mukherjee, M. T. Zanni and D. P. Raleigh, Residue-specific, real-time characterization of lag-phase species and fibril growth during amyloid formation: a combined fluorescence and IR study of p-cyanophenylalanine analogs of islet amyloid polypeptide, *J. Mol. Biol.*, 2010, **400**, 878-888.
15. S. Bagchi, S. G. Boxer and M. D. Fayer, Ribonuclease S dynamics measured using a nitrile label with 2D IR vibrational echo spectroscopy, *J. Phys. Chem. B*, 2012, **116**, 4034-4042.
16. R. Adhikary, J. Zimmermann, J. Liu, P. E. Dawson and F. E. Romesberg, Experimental characterization of electrostatic and conformational heterogeneity in an SH3 domain, *J. Phys. Chem. B*, 2013, **117**, 13082-13089.
17. W. Hu and L. J. Webb, Direct Measurement of the Membrane Dipole Field in Bicelles Using Vibrational Stark Effect Spectroscopy, *J. Phys. Chem. Lett.*, 2011, **2**, 1925-1930.
18. R. Shrestha, A. E. Cardenas, R. Elber and L. J. Webb, Measurement of the Membrane Dipole Electric Field in DMPC Vesicles Using Vibrational Shifts of p-Cyanophenylalanine and Molecular Dynamics Simulations, *J. Phys. Chem. B*, 2015, **119**, 2869-2876.
19. M. J. Tucker, X. S. Gai, E. E. Fenlon, S. H. Brewer and R. M. Hochstrasser, 2D IR photon echo of azido-probes for biomolecular dynamics, *Phys. Chem. Chem. Phys.*, 2011, **13**, 2237-2241.
20. N. M. Levinson, S. D. Fried and S. G. Boxer, Solvent-Induced Infrared Frequency Shifts in Aromatic Nitriles Are Quantitatively Described by the Vibrational Stark Effect, *J. Phys. Chem. B*, 2012, **116**, 10470-10476.
21. A. T. Fafarman, P. A. Sigala, D. Herschlag and S. G. Boxer, Decomposition of Vibrational Shifts of Nitriles into Electrostatic and Hydrogen-Bonding Effects, *J. Am. Chem. Soc.*, 2010, **132**, 12811-12813.
22. S. Bagchi, S. D. Fried and S. G. Boxer, A solvatochromic model calibrates nitriles' vibrational frequencies to electrostatic fields, *J. Am. Chem. Soc.*, 2012, **134**, 10373-10376.
23. P. E. Dawson, M. C. Fitzgerald, T. W. Muir and S. B. H. Kent, Methods for the Chemical Synthesis and Readout of Self-Encoded Arrays of Polypeptide Analogues, *J. Am. Chem. Soc.*, 1997, **119**, 7917-17927.
24. P. E. Dawson and S. B. H. Kent, Synthesis of Native Proteins by Chemical Ligation, *Annu. Rev. Biochem.*, 2000, **69**, 923-960.
25. K. C. Schultz, S. Supekova, Y. Ryu, J. Xie, R. Perera and P. G. Schultz, A Genetically Encoded Infrared Probe, *J. Am. Chem. Soc.*, 2006, **128**, 13984-13985.
26. C. C. Liu and P. G. Schultz, Adding new chemistries to the genetic code, *Annu. Rev. Biochem.*, 2010, **79**, 413-444.

27. A. Chatterjee, S. B. Sun, J. L. Furman, H. Xiao and P. G. Schultz, A Versatile Platform for Single- and Multiple-Unnatural Amino Acid Mutagenesis in Escherichia coli, *Biochemistry*, 2013, **52**, 1828-1837.
28. A. Zarrinpar, S.-H. Park and W. A. Lim, Optimization of specificity in a cellular protein interaction network by negative selection, *Nature*, 2003, **426**, 676-680.
29. B. J. Mayer, SH3 domains: complexity in moderation, *J. Cell Sci.*, 2001, **114**, 1253-1263.
30. J. A. Marles, S. Dahesh, J. Haynes, B. J. Andrews and A. R. Davidson, Protein-protein interaction affinity plays a crucial role in controlling the Sho1p-mediated signal transduction pathway in yeast, *Molec. Cell*, 2004, **14**, 813-823.
31. R. F. Wissner, S. Batjargal, C. M. Fadzen and E. J. Petersson, Labeling proteins with fluorophore/thioamide Forster resonant energy transfer pairs by combining unnatural amino acid mutagenesis and native chemical ligation, *J. Am. Chem. Soc.*, 2013, **135**, 6529-6540.
32. Z. Getahun, C.-Y. Huang, T. Wang, B. De Leon, W. F. DeGrado and F. Gai, Using Nitrile-Derivatized Amino Acids as Infrared Probes of Local Environment, *J. Am. Chem. Soc.*, 2003, **125**, 405-411.
33. J. H. Choi, K. I. Oh, H. Lee, C. Lee and M. Cho, Nitrile and thiocyanate IR probes: quantum chemistry calculation studies and multivariate least-square fitting analysis, *J. Chem. Phys.*, 2008, **128**, 134506.
34. B. A. Lindquist, K. E. Furse and S. A. Corcelli, Nitrile groups as vibrational probes of biomolecular structure and dynamics: an overview, *Phys. Chem. Chem. Phys.*, 2009, **11**, 8119-8132.
35. S. Arold, R. O'Brien, P. Franken, M.-P. Strub, F. Hoh, C. Dumas and J. E. Ladbury, RT Loop Flexibility Enhances the Specificity of Src Family SH3 Domains for HIV-1 Nef, *Biochemistry*, 1998, **37**, 14683-14691.
36. J. K. Chung, M. C. Thielges and M. D. Fayer, Dynamics of the folded and unfolded villin headpiece (HP35) measured with ultrafast 2D IR vibrational echo spectroscopy, *Proc. Natl. Acad. Sci.*, 2011, **108**, 3578–3583.
37. E. J. Stollar, H. Lin, A. R. Davidson and J. D. Forman-Kay, Differential Dynamic Engagement within 24 SH3 Domain: Peptide Complexes Revealed by Co-Linear Chemical Shift Perturbation Analysis, *PLoS ONE*, 2012, **7**, e51282.

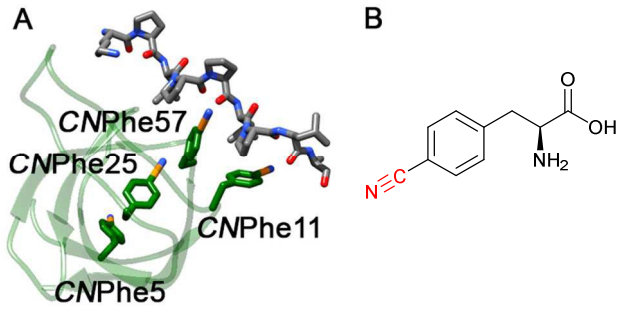


Figure 1. (A) Structure of SH3^{Sho1} showing labeled residues and Pbs2 ligand (2vkn). *CN*Phe11 and *CN*Phe57 are located in the binding interface and are shown interacting with the Pbs2 peptide ligand. *CN*Phe25 is located in the core of SH3^{Sho1} and *CN*Phe5 is located most distal to the binding groove. Image generated using UCSF Chimera. (B) *p*-cyanophenylalanine

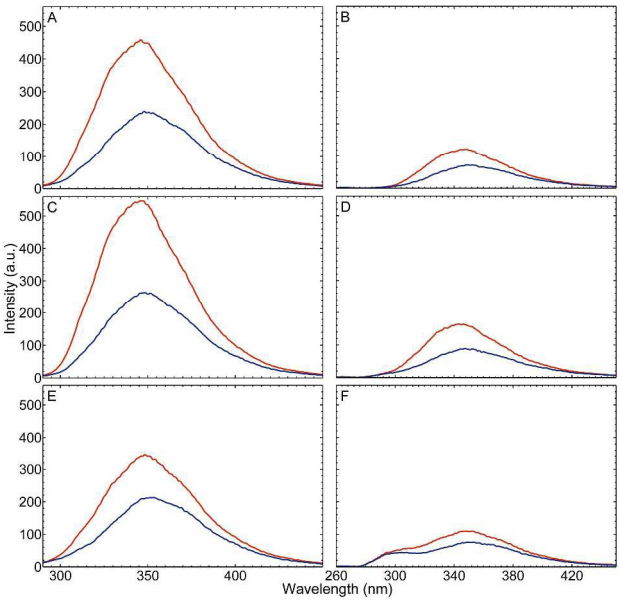


Figure 2. Fluorescence spectra of unligated (blue) and Pbs2 peptide–bound (red) states of wild-type (A and B), *CN*Phe11 (C and D), and *CN*Phe25 (E and F) SH3^{Sho1}at excitation wavelengths of 280 nm (left) and 240 nm (right).

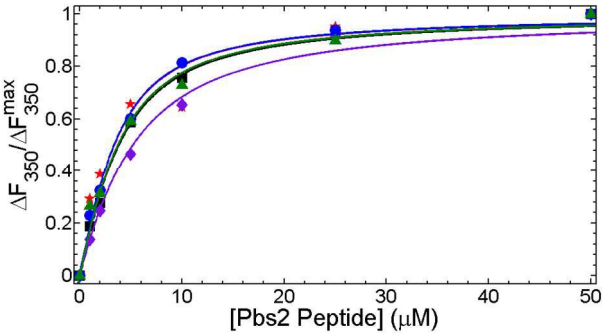


Figure 3. Fluorescence binding data (symbols) and fits (lines) for the SH3^{Sho1} variants with Pbs2 peptide: wild-type (black squares), *CN*Phe5 (red stars), *CN*Phe11 (purple diamonds), *CN*Phe25 (blue circles), and *CN*Phe57 (green triangles).

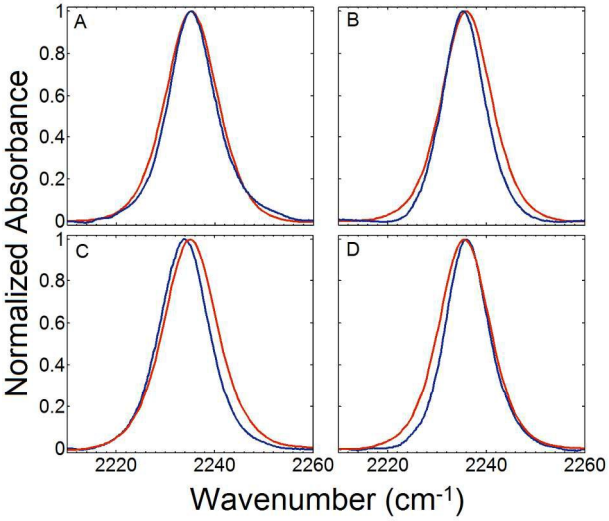


Figure 4. FT IR spectra of unligated (blue) and Pbs2 peptide-bound (red) states of SH3^{Sho1} variants. A) *CN*Phe5, B) *CN*Phe11, C) *CN*Phe25, D) *CN*Phe57.

Table 1. IR Spectral Fit Parameters.

	unligated		Pbs2 peptide-bound	
	ν (cm^{-1})	**fwhm (cm^{-1})**	**ν (cm^{-1})**	**fwhm (cm^{-1})**
*CN*Phe5	2235.3 ± 0.2	12.9 ± 1.7	2235.2 ± 0.01	12.5 ± 0.2
*CN*Phe11	2235.4 ± 0.01	10.4 ± 0.2	2236.1 ± 0.1	12.1 ± 0.2
*CN*Phe25	2233.8 ± 0.1	12.2 ± 0.7	2235.1 ± 0.04	13.1 ± 0.7
*CN*Phe57	2236.1 ± 0.1	10.5 ± 0.4	2235.6 ± 0.01	11.9 ± 0.1

fwhm = full width at half maximum