9040 *Biochemistry* **2009,** *48,* 9040–9046
DOI: 10.1021/bi900938z

BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

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[†]

Humeyra Taskent-Sezgin,[‡] Juah Chung,[‡] Vadim Patsalo,[§] Shigeki J. Miyake-Stoner,[∥] Andrew M. Miller,[∥] Scott H. Brewer,[∥] Ryan A. Mehl,[∥] David F. Green,[‡,§,⊥] Daniel P. Raleigh,*[,‡,⊥] and Isaac Carrico*[,‡,⊥]

[‡]*Department of Chemistry, State University of New York at Stony Brook, Stony Brook, New York 11794,* [§]*Department of Applied Mathematics and Statistics, State University of New York at Stony Brook, Stony Brook, New York 11794,* [∥]*Department of Chemistry, Franklin and Marshall College, Lancaster, Pennsylvania 17603, and* [⊥]*Institute of Chemical Biology and Drug Discovery at State University of New York at Stony Brook, Stony Brook, New York 11794*

Received June 3, 2009; Revised Manuscript Received August 4, 2009

ABSTRACT: The use of noncoded amino acids as spectroscopic probes of protein folding and function is growing rapidly, in large part because of advances in the methodology for their incorporation. Recently *p*-cyanophenylalanine has been employed as a fluorescence and IR probe, as well as a FRET probe to study protein folding, protein−membrane interactions, protein−protein interactions and amyloid formation. The probe has been shown to be exquisitely sensitive to hydrogen bonding interactions involving the cyano group, and its fluorescence quantum yield increases dramatically when it is hydrogen bonded. However, a detailed understanding of the factors which influence its fluorescence is required to be able to use this popular probe accurately. Here we demonstrate the recombinant incorporation of *p*-cyanophenylalanine in the N-terminal domain of the ribosomal protein L9. Native state fluorescence is very low, which suggests that the group is sequestered from solvent; however, IR measurements and molecular dynamics simulations show that the cyano group is exposed to solvent and forms hydrogen bonds to water. Analysis of mutant proteins and model peptides demonstrates that the reduced native state fluorescence is caused by the effective quenching of *p*-cyanophenylalanine fluorescence via FRET to tyrosine side-chains. The implications for the interpretation of *p*-cyanophenylalanine fluorescence measurements and FRET studies are discussed.

The experimental study of protein folding remains an area of intense interest, owing in part to the realization that protein misfolding can play a major role in certain diseases (*1*). The increasing interplay between experimental and computational investigations of folding has also helped to refocus interest in the area (*2*, *3*). Most experimental studies of folding rely on changes in the fluorescence of intrinsic fluorophores such as Trp or Tyr (*4*, *5*). Recently, *p*-cyanophenylalanine (F_{CN}[1]) has been shown to have unique fluorescence properties. F_{CN} can be selectively excited at 240 nm, and its emission maximum is 290 nm; its fluorescence quantum yield depends on hydrogen bonding interactions involving the cyano (CN) group (*6*). F_{CN} fluorescence can be selectively detected in the presence of Trp and Tyr, and the group can be tolerated both in a hydrophobic environment and in a hydrophilic environment, making it a conservative substitution for Tyr, Phe, and Trp.

Gai and co-workers have pioneered the use of F_{CN} as a fluorescence probe and have also demonstrated that F_{CN}−Trp can be used as a FRET pair with a Förster distance (R_0) of 16 Å (*6*). This is a convenient R_0 for studies of protein folding. Gai and co-workers have utilized F_{CN} fluorescence and F_{CN}−Trp FRET to study protein−membrane (*7*, *8*) and protein−protein interactions (*6*), while we have used F_{CN} fluorescence to probe amyloid formation (*9*) and protein folding (*10*, *11*). These studies have led to increasing interest in the use of F_{CN}. The CN group also provides a useful IR probe (*12*). Due to the sensitivity of its stretching frequency to solvation and electric field effects, the CN group can serve as a reporter of the local environment, and Boxer and co-workers have used CN groups in Stark tuning experiments to map protein electric fields (*13*).

F_{CN} fluorescence intensity is traditionally interpreted to reflect the presence or absence of H-bonding interactions with the CN group but little is known about other factors that can influence the quantum yield such as quenching by amino acid side chains. A deeper understanding of F_{CN} fluorescence is clearly required given the utility of the group as a probe of protein structure, protein−ligand interactions and protein−membrane interactions. Here we use a combination of fluorescence, infrared (IR)

[†]Grant sponsor: NSF CHE0806433 to D.P.R., F&M Hackman and Eyler funds; Research Corporation (CC7352) to S.H.B., NSF-MCB-0448297 to R.A.M., and Startup support from the Vice President for Research at SUNY at Stony Brook to I.C.

*To whom correspondence should be addressed. D.P.R: phone, 631-632-9547; fax, 631-632-7960; e-mail, draleigh@notes.cc.sunysb.edu. I. C.: phone, 631-632-7935; fax, 631-632-7960; e-mail, isaac.carrico@sunysb.edu.

[1]Abbreviations: ATR-FTIR, attenuated total reflection Fourier transform infrared spectroscopy; CD, circular dichroism; CN, cyano; F_{CN}, *p*-cyanophenylalanine; FRET, fluorescence resonance energy transfer; MALDI-TOF MS, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; MD, molecular dynamics; NTL9, N-terminal domain of the ribosomal protein L9 composed of residues 1 to 56 of intact L9; NTL9-F5F_{CN}, NTL9 with a *p*-cyanophenylalanine at position 5.

 Published on Web 08/06/2009

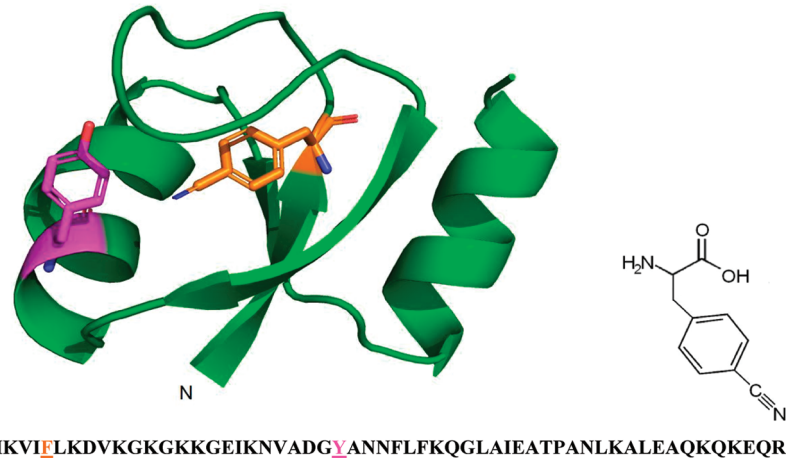

FIGURE 1: Ribbon diagram of NTL9 showing the location of the F5F_{CN} (orange) mutation and Y25 (magenta). The amino acid sequence of NTL9 is shown, F5 and Y25 are underlined. Inset: The structure of *p*-cyanophenylalanine (F_{CN}).

spectroscopy and molecular dynamics (MD) simulations to explore the factors which control F_{CN} fluorescence. As a model system we choose the N-terminal domain of the ribosomal protein L9 (NTL9). NTL9 is a 56-residue, mixed $\alpha-\beta$ protein and is the simplest example of the split $\beta-\alpha-\beta$ motif. Folding of the domain is reversible and two-state (*14*, *15*). Almost all of the published folding studies of NTL9 have made use of Tyr fluorescence since Tyr25 is the only fluorescent group in the domain. Tyr25 is on the solvent exposed face of the first helix in NTL9; however, the ring packs against the body of the protein and there is a measurable change in fluorescence upon unfolding (Figure 1). NTL9 can tolerate a wide range of substitutions in its hydrophobic core (*14*, *15*), and we recently prepared a derivative containing a Phe5 to F_{CN} substitution (NTL9-F5F_{CN}) by solid-phase peptide synthesis and have shown that the substitution does not perturb the structure (*10*). Thus NTL9 is an excellent model system to probe the origins of changes in F_{CN} fluorescence.

Here we show that the analysis of F_{CN} fluorescence is more complicated than initially thought and demonstrate that the fluorescence intensity cannot be interpreted solely in terms of hydrogen bonding to the CN group owing to complications caused by quenching from amino acid side-chains.

MATERIALS AND METHODS

Protein Expression, Purification and Characterization. The NTL9-WT gene was inserted into a pBAD vector. The codon for F5 in the NTL9 gene was replaced with the amber stop codon to allow coding for F_{CN} (*16*). The vector that is generated is denoted pBAD-NTL9-F5F_{CN}. A pDule vector carrying the gene for the aminoacyl-tRNA synthetase for the incorporation of F_{CN} into the growing protein chain was used (*11*, *17*). The two vectors, pBAD and pDule, were cotransformed into electrocompetent *Escherichia coli* DH10B cells. A 4 mL LB culture grown overnight was used as a seed for a sample of 750 mL of main autoinduction media containing 1 mM F_{CN}. The culture was grown for 24 h at 37 °C at 250 rpm (*18*). The proteins were purified with ion exchange chromatography followed by reverse-phase HPLC on a Vydac C8 preparative column as described previously (*14*). The protein, NTL9-F5F_{CN}, was characterized by LTQ Orbitrap XL mass spectrometer. The observed monoisotopic MW was 6241.448 Da, and calculated monoisotopic MW was 6241.443 Da. NTL9-F5F_{CN}Y25F was characterized by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). The observed average MW was 6228.70 Da, and calculated average MW was 6228.30 Da.

Peptide Synthesis and Characterization. The control peptides, GGF$_{CN}$AA and GGF$_{CN}$YA, were synthesized by standard Fmoc solid-phase peptide synthesis on an Applied Biosystems 433A peptide synthesizer. Both peptides have a free N-terminus, and the PAL-PEG-PS resin leads to an amidated C-terminus. The peptides were cleaved from the resin using 91% (v/v) trifluoroacetic acid (TFA), 3% (v/v) anisole, 3% (v/v) thioanisole, 3% (v/v) 1,2-ethanedithiol. The peptides were precipitated using cold ethyl ether, and scavengers were removed under vacuum. The crude peptide was then purified using reverse-phase HPLC on a Vydac C18 semipreparative column. An A−B gradient system was used with buffer A composed of 0.1% (v/v) solution of TFA and buffer B composed of 90% (v/v) isopropanol, 9.9% (v/v) water and 0.1% (v/v) TFA. The gradient was 10 min of buffer A wash followed by 0−90% B in 120 min. The peptides were characterized by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). For GGF$_{CN}$AA, observed MW was 446.10 Da and calculated MW was 446.45 Da; and for GGF$_{CN}$YA, observed MW was 538.57 Da and calculated MW was 538.55 Da.

Equilibrium Denaturation Experiments. Circular dichroism (CD) wavelength scans and CD-monitored denaturation experiments were conducted as previously described for NTL9 (*14*). Fluorescence-monitored denaturation experiments were conducted on an Applied Photon Technologies fluorimeter. The protein and peptide concentrations were determined by measuring the absorbance of the sample at 280 nm. The extinction coefficient for F_{CN} is 850 M^{-1} cm^{-1} (*6*), and for Tyr, it is 1490 M^{-1} cm^{-1}. Fluorescence- and CD-monitored experiments were performed at pH 5.4 and 25 °C. These experimental conditions were chosen to match previous studies. Samples were dissolved in 20 mM sodium acetate and 100 mM NaCl. For the fluorescence emission spectra, the F_{CN} group was excited at 240 nm and the emission signal was recorded in the range 250−450 nm. For the fluorescence-monitored urea denaturation experiments, the excitation wavelength for the F_{CN} group was 240 nm, and the emission signal was followed at 291 nm. The concentration of urea solutions was determined by

measuring the refractive index. Denaturation curves were fit to the following equation:

$$f = \frac{a_n + b_n[\text{denaturant}] + (a_d + b_d[\text{denaturant}])e^{-[\Delta G^\circ_U([\text{denaturant}]/RT)]}}{1 + e^{-[\Delta G^\circ_U(\text{denaturant})/RT]}} \quad (1)$$

where:

$$\Delta G^\circ_U([\text{denaturant}]) = \Delta G^\circ_U(H_2O) - m[\text{denaturant}] \quad (2)$$

f is the measured signal (ellipticity in CD, and fluorescence signal in fluorescence-monitored denaturation experiments), a_n is the intercept of the extrapolated curve in the pretransition region, b_n is the slope of this same curve. a_d is the intercept of the curve in the post-transition region, b_d is the slope in this region. ΔG°_U is the change in free energy of unfolding, T is the absolute temperature and R is the gas constant.

Equilibrium FTIR Measurements. Equilibrium FTIR absorbance spectra were recorded on a Bruker Vertex 70 FTIR spectrometer equipped with a globar source, KBr beamsplitter, a liquid nitrogen cooled mercury cadmium telluride (MCT) detector, and a Harrick BioATRcell II accessory. The spectra were the result of 512 scans and were recorded at a resolution of 1.0 cm^{-1}. The NTL9-F5F_{CN} protein was dissolved in a buffer containing 20 mM sodium acetate and 100 mM NaCl at a pH of 5.4, with or without 9.8 M urea. The protein concentration for the IR experiments was ~1.5 mM. The spectra were measured at 25 °C. The nitrile IR absorbance band was fit with a Gaussian line shape using Igor Pro (Wavemetrics, Inc.).

Molecular Dynamics Simulation System Setup. NTL9-WT coordinates (PDB: 2HBB) were retrieved from the Protein Data Bank (The Research Collaboratory for Structural Bioinformatics (RCSB), http://www.rcsb.org/pdb/). The last five residues of NTL9 are partially disordered and are not listed in the PDB file. Thus, the simulations were conducted on residues 1 to 51. Hydrogens were added with the HBUILD (*19*) module of CHARMM (*20*), and the system was minimized briefly using the PARAM22 force field (*21*). An orthorhombic box of pre-equilibrated TIP3P solvent was constructed 10 Å around the protein, and the system was charge-neutralized by random placement of sodium and chloride ions to a final concentration of approximately 145 mM. The final system contained 3607 TIP3P water molecules.

For NTL9-F5F_{CN}, Phe5 was mutated to F_{CN} by replacing the *para*-hydrogen with a nitrile, keeping all other atoms fixed. The F_{CN} side-chain was minimized while keeping the protein fixed, followed by unconstrained minimization. A box of solvent was then constructed and the system was neutralized by the addition of ions, as above.

Molecular dynamics calculations were run with NAMD (*22*). Briefly, the fully solvated system was minimized for 5000 steps, and equilibrated by heating to 300 K over 200 ps using Particle Mesh Ewald electrostatics and a time step of 2 fs. Production dynamics were run at 300 K for 90 ns at constant temperature and pressure (1 atm), with full electrostatics evaluated every time step.

RESULTS AND DISCUSSION

A ribbon diagram of NTL9 is displayed in Figure 1 and shows the location of F5, the site chosen for F_{CN} substitution, and Y25 which is the dominant natural fluorophore. F5 is a core residue and is located on the first β-strand. F_{CN} was introduced by expression in *E. coli* using a modified version of the methodology pioneered by the Schultz group (*11*, *16*, *18*, *23*). The plasmid (pDule), containing the gene for the aminoacyl t-RNA synthetase used to incorporate the unnatural phenylalanine variants, and the protein expression conditions were optimized in the Mehl laboratory (*11*, *18*). A yield of 7.5 mg of purified NTL9-F5F_{CN} was obtained from 0.75 L of culture supplemented with 1 mM F_{CN}. The incorporation of F_{CN} into the mutant, NTL9-F5F_{CN} was confirmed with the LTQ Orbitrap XL mass spectrometer. The monoisotopic mass distribution is consistent with essentially complete incorporation (Supporting Information).

Position 5 was chosen based on our results with chemically synthesized NTL9 variant which demonstrated that replacement of F5 with the F_{CN} group is a conservative mutation (*10*).

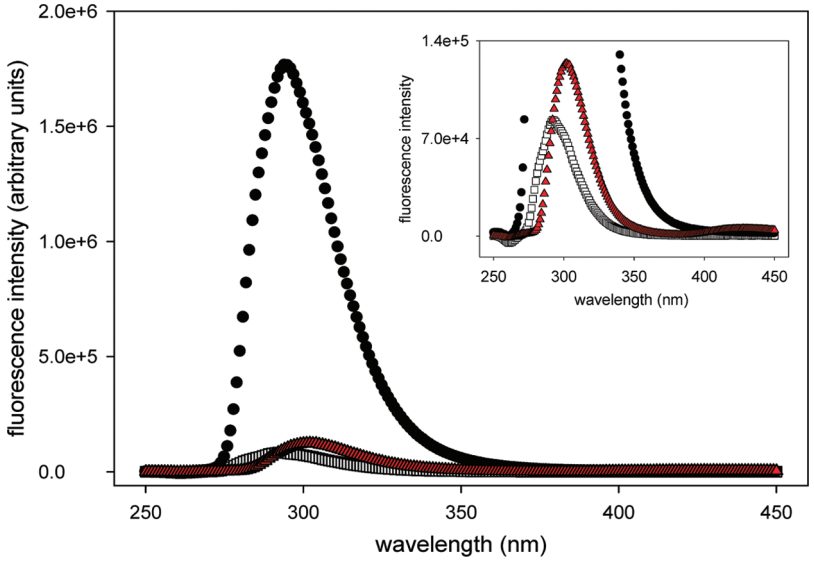

FIGURE 2: Fluorescence emission spectra of NTL9-F5F_{CN} in the folded (open squares), in the urea-induced unfolded state (filled circles) and wild-type NTL9 in the folded state (red triangles). The increase in the F_{CN} fluorescence signal upon going from the folded state to the unfolded state in NTL9-F5F_{CN} is traditionally thought to reflect increased solvation of the F_{CN} group. Inset: Expansion of the emission spectra of NTL9-F5F_{CN} and NTL9-WT in the folded state. Protein concentration was 15 μM. Experiments were conducted at pH 5.4 and 25 °C in 20 mM sodium acetate, 100 mM sodium chloride. The excitation wavelength was 240 nm for all experiments.

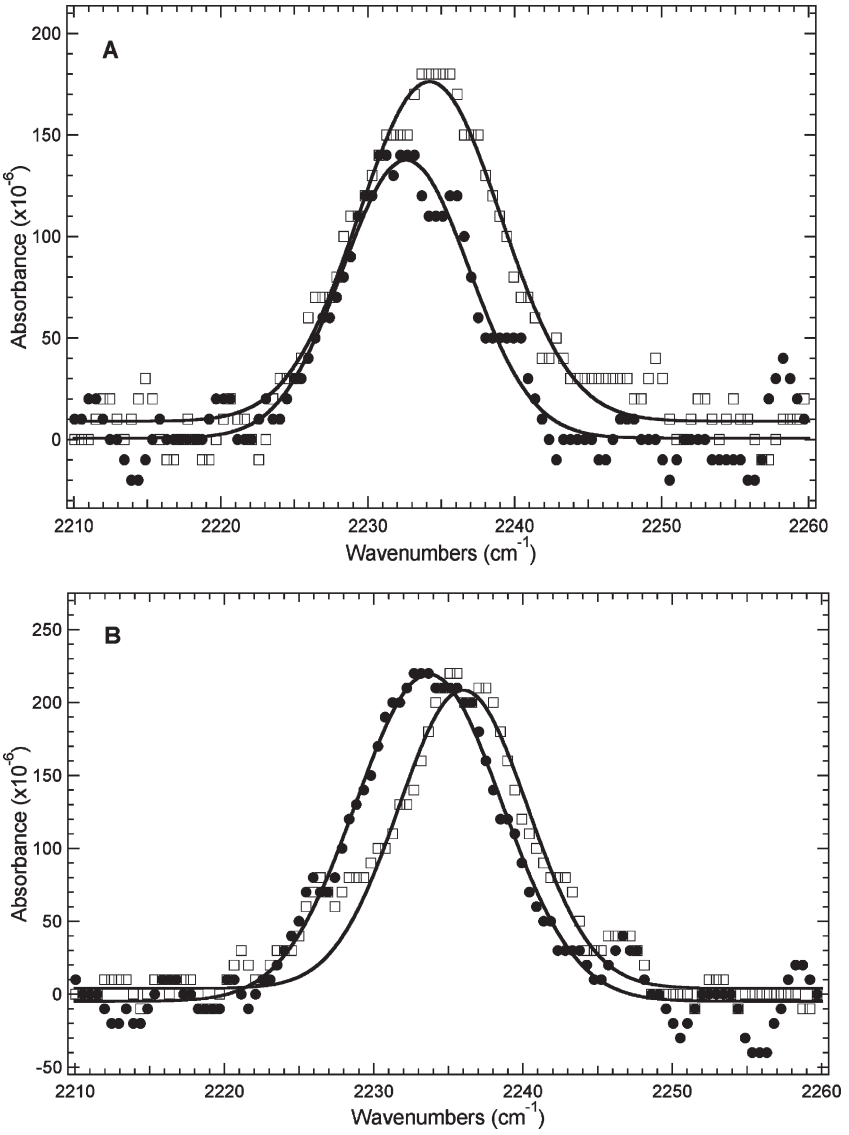

FIGURE 3: (A) FTIR spectra of NTL9-F5F_{CN} in the folded (open squares) and in the urea-induced unfolded state (filled circles). The CN stretching frequency was observed at 2234.2 cm^{-1} and 2232.6 cm^{-1} in the folded and in the unfolded state, respectively. This indicates that the CN group is in a similar environment in both states. The protein concentration was around 1.5 mM. (B) FTIR spectra of the control peptide, GGF$_{CN}$AA in the absence (open squares) and presence (filled circles) of urea. Samples were at pH 5.4 and 25 °C. Each spectrum was fit to a Gaussian line shape (solid curves).

The far-UV CD wavelength spectrum of the mutant, denoted here NTL9-F5F_{CN}, indicates that the secondary structure of this variant is similar to the wild-type protein (Supporting Information). We determined the stability of NTL9-F5F_{CN} by both CD and fluorescence monitored urea denaturation experiments (Supporting Information). The stability and the *m*-values determined by the two methods are identical, confirming that the modified domain folds cooperatively. Importantly, the stability and *m*-value are very close to those for wild-type NTL9, providing additional evidence that the F_{CN} substitution does not perturb the fold.

The fluorescence emission spectra of NTL9-F5F_{CN} in the folded and in the urea-induced unfolded state are shown in Figure 2. The spectrum of wild-type NTL9 in the folded state is included for comparison. An excitation wavelength of 240 nm was used for all three samples. The fluorescence intensity is drastically reduced in the folded state of NTL9-F5F_{CN} relative to the unfolded state, and the ratio is similar to what was previously reported for the chemically synthesized variant (*10*). The low fluorescence intensity of the F_{CN} group in the native state is, by the traditional interpretation, consistent with the burial of the CN group (*6*). However, there are some peculiarities which suggest that the situation may be more complex. For example, the fluorescence emission spectrum of the wild-type protein recorded using an excitation wavelength of 240 nm is actually more intense than the spectrum of the F5F_{CN} mutant (Figure 2). This is surprising since the mutant includes both F_{CN} and Tyr fluorophores.

To further investigate the origins of this effect, we used IR spectroscopy to probe the molecular environment of the CN group in order to obtain an independent assessment of the environment of the CN group. The CN stretching mode is a convenient spectroscopic marker since it appears in a transparent region of protein IR spectra, and its frequency is sensitive to hydrogen bonding and electric field effects. A blue-shifted CN stretching vibration is observed when the group is transferred from a hydrophobic environment to water (*12*). The vibrational frequency reported for the CN group of F_{CN} was 2237.2 cm^{-1} in

water, and the CN group of Fmoc-F_{CN} was 2228.8 cm^{-1} in THF (*12*). Figure 3a shows the ATR-FTIR (attenuated total reflection Fourier transform infrared) spectra of NTL9-F5F_{CN} in the absence and presence of urea. The nitrile stretching frequency of NTL9-F5F_{CN} in the absence of urea is 2234.2 cm^{-1}, suggesting a partially solvated nitrile group in the folded state of NTL9. The nitrile stretching frequency of NTL9-F5F_{CN} in the urea-induced unfolded state (9.8 M urea) is 2232.6 cm^{-1}, which is only a 1.6 cm^{-1} red shift relative to the folded state. A blue shift was expected upon unfolding since the nitrile moiety should become more solvated. However, the high concentration of urea present in the aqueous buffer lowers the dielectric of the buffer and may alter water structure. The net effect is a red shift in the frequency of an exposed nitrile group in 9.8 M urea relative to that observed in the native buffer. This effect is illustrated in Figure 3b, which shows the ATR-FTIR spectra of a control peptide, GGF$_{CN}$AA, in the absence and presence of urea. The nitrile stretching frequency of the control peptide shifted from 2236.0 cm^{-1} to 2233.7 cm^{-1} upon addition of 9.8 M urea, a red shift of 2.3 cm^{-1}. This peptide is disordered, and thus the CN group should be fully solvated in both the native and urea containing buffers. The 2.3 cm^{-1} shift represents a reasonable correction for the effects of 9.8 M urea. After correcting for this intrinsic urea dependence of the CN stretch, an estimated blue shift of ~0.7 cm^{-1} for the nitrile stretching frequency is obtained upon unfolding of NTL9-F5F_{CN}. Irrespective of the details of the urea induced correction, the key result is that the CN stretching frequency is very similar in the folded and unfolded states and is close to that observed for model compounds in water.

The fluorescence and IR studies lead to apparently incompatible conclusions. The very low fluorescence intensity in the folded state argues that the F_{CN} group is buried, while the IR data indicates that the CN group is exposed. Comparison of the fluorescence emission spectrum of wild-type NTL9 with NTL9-F5F_{CN} (recorded with an excitation wavelength of 240 nm) suggests, as noted above, that the tyrosine side chain may quench the fluorescence of the F_{CN} group in the native state. In the crystal structure of NTL9, the distance between the center of the Tyr25 ring and the center of the Phe5 ring is only 7.4 Å. The emission maximum of F_{CN} in NTL9-F5F_{CN}Y25F is 290 nm, and the full width of the emission spectrum at 50% intensity is about 30 nm, hence there is considerable overlap between the tyrosine absorption band and the F5F_{CN} emission spectrum in NTL9-F5F_{CN}. Given the significant spectral overlap, the relatively high fluorescence quantum yield of F_{CN} (comparable to Trp) and the close approach of F5F_{CN} and Y25, the most likely cause for the unexpectedly weak native state FCN fluorescence is quenching due to FRET to Y25.

To more rigorously test the possibility that Tyr25 quenches the F_{CN} fluorescence, we prepared an F5F_{CN}Y25F double mutant. Far-UV CD indicated that the double mutant has the same fold as the wild-type (Supporting Information). The stabilities and *m*-values calculated from the urea-induced unfolding experiments monitored by fluorescence and CD were very similar to the wild-type values (Supporting Information), indicating that the substitution is well tolerated. The fluorescence emission spectra of NTL9-F5F_{CN}Y25F in the native and urea-induced unfolded state are shown in Figure 4. An approximately 40-fold increase in the fluorescence emission signal of NTL9-F5F_{CN}-Y25F in the native state is detected relative to that observed for the native state of NTL9-F5F_{CN}. This directly validates the hypothesis that Tyr25 quenches the fluorescence signal of the

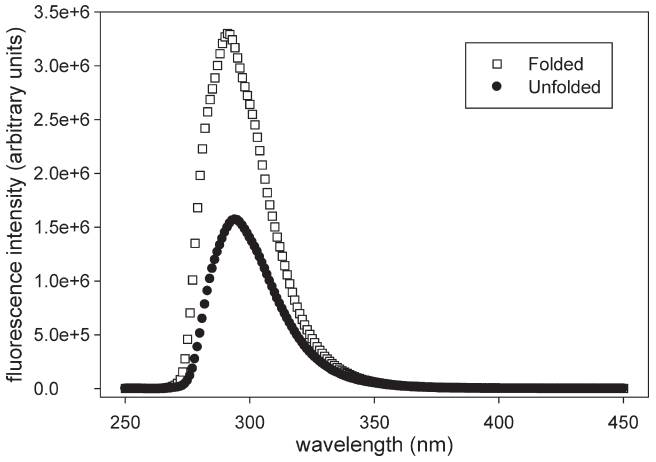

FIGURE 4: Fluorescence emission spectra of NTL9-F5F_{CN}Y25F in the folded (open squares) and in the urea-induced unfolded state (filled circles). The high fluorescence intensity in the folded state indicates that the F_{CN} group is not completely buried in NTL9-F5F_{CN}Y25F and forms H-bonds with the solvent. Protein concentration was 15 μM. Protein samples were in 20 mM sodium acetate, 100 mM sodium chloride, pH 5.4 at 25 °C.

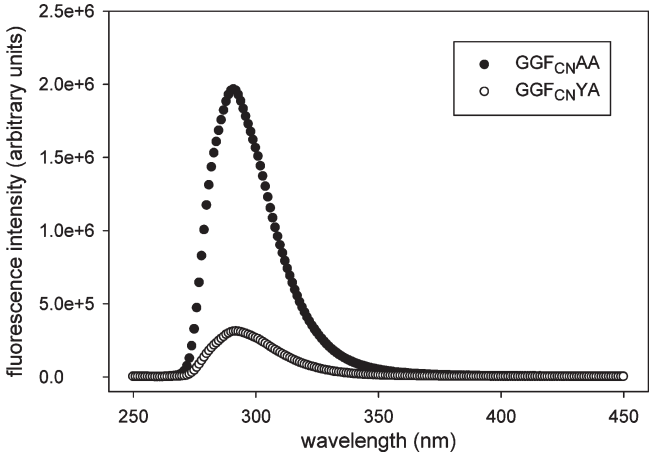

FIGURE 5: Fluorescence emission spectra of model peptides confirm that Tyr is an efficient quencher of F_{CN} fluorescence. GGF$_{CN}$AA (filled circles) and GGF$_{CN}$YA (open circles) in folding buffer. The decrease in the fluorescence signal intensity of F_{CN} (in GGF$_{CN}$AA) with the introduction of a Tyr (in GGF$_{CN}$YA) demonstrates that Tyr is quenching the fluorescence signal of F_{CN}. The peptide concentrations were 20 μM.

F_{CN} group in NTL9-F5F_{CN} via FRET. The intense fluorescence of F5F_{CN} in NTL9-F5F_{CN}Y25F in the folded state is consistent with solvent exposure of the CN group and is fully consistent with the FT-IR results. To further test the ability of Tyr to quench F_{CN} fluorescence, we synthesized a pair of small F_{CN} containing peptides. Figure 5 compares the fluorescence emission spectra of F_{CN} in the GGF$_{CN}$AA and GGF$_{CN}$YA peptides. The F_{CN} fluorescence is considerably reduced in the Tyr containing peptide confirming that Tyr is an effective quencher of F_{CN} fluorescence.

To assess whether the model suggested by these results was consistent with structural data, we conducted molecular dynamics (MD) simulations of wild-type NTL9 and NTL9-F5F_{CN} using the CHARMM (*20*) PARAM22 force field (*21*) and the TIP3P water model to probe for solvent hydrogen bond interactions with the F_{CN} group. It was first necessary to parametrize the F_{CN} group. Parameters for F_{CN} were modeled

by homology, using the Tyr residue of PARAM22 and the cyano group parameters from the 3-cyano-pyridine residue of the CHARMM pyridine force field (*24*). The parametrization placed a partial charge of −0.53*e* on the cyano nitrogen and +0.40*e* on the carbon of the cyano group.

The Cα backbone rmsd of the whole protein (residues 1 to 51) and of the N-terminal subdomain (residues 1 to 39) (Supporting Information) showed both proteins to be stably folded near the X-ray structure for the course of the 90 ns simulations. The simulations also appeared to be reasonably converged. The rmsd for the 1−39 core unit relative to the starting structure for the simulation, averaged over the last 45 ns of the MD trajectories, was 0.80 Å for wild-type and 0.79 Å for NTL9-F5F_{CN}. Chi angle analysis for the side chain at position 5 (Phe in wild-type and F_{CN} for NTL9-F5F_{CN}) showed that the two side chains adopt the same χ_1 and χ_2 angles (Supporting Information), further supporting the claim that the F5F_{CN} mutation is conservative, and that the side chain adopts the same orientation in both proteins. Using a hydrogen bond cutoff distance of 2.4 Å, the CN group was found to accept a hydrogen bond from surrounding water molecules in 55% of the trajectory frames during the course of the 90 ns MD trajectory. In addition, we calculated the distance and the relative orientation between the side chains at positions 5 and 25 (Supporting Information). The observed mean distance between F5F_{CN} and Y25 was 7.43 Å. The side chains were also found to adopt a fairly rigid relative orientation, despite the solvent accessibility of Y25. The close proximity of F5F_{CN} to Y25 side-chain and the observed H-bonding interactions between the CN group of the F5F_{CN} mutant with water in the native state are consistent with the IR data and support the hypothesis that the low fluorescence signal intensity in NTL9-F5F_{CN} mutant is not due to the burial of the CN group but rather caused by quenching due to the Tyr.

CONCLUSIONS

Unnatural amino acids are becoming increasingly popular spectroscopic probes as the methodology to incorporate them into proteins improves (*16*, *25*). Given its ease of incorporation and convenient spectroscopic properties we anticipate that the use of F_{CN} will continue to grow. Here, we have demonstrated the efficient recombinant incorporation of F_{CN} into a model protein. The combined fluorescence and IR investigations reported here clearly demonstrate that caution must be used when interpreting F_{CN} fluorescence in terms of solvent exposure and specific hydrogen bonding interactions involving the CN group. This study also demonstrates the power of combining IR and fluorescence to probe the local environment. This is one of the attractive features of an F_{CN} substitution. A second advantage of this probe is that IR spectra can be recorded in high concentrations of urea which facilitates studies of protein folding. Our demonstration that Tyr effectively quenches F_{CN} fluorescence has important implications for F_{CN}−Trp based FRET studies since conformational transitions which modulate the distance of a F_{CN}−Trp FRET pair could also modulate any F_{CN}−Tyr interactions. Thus the apparent F_{CN}−Trp FRET response could include a contribution from differential F_{CN}−Tyr FRET in the two states. Of course, F_{CN}−Tyr FRET can be used directly to probe conformational changes in proteins. We hope our analysis of F_{CN} quenching will motivate similar studies in other systems and will aid in the interpretation of the spectral response of this very useful probe. One can envisage more sophisticated applications involving proteins which contain Tyr, Trp and F_{CN} in which the probes are used to monitor relative chain motion between three sites. Finally, the parametrization of the F_{CN} group reported here should facilitate MD studies of F_{CN} containing proteins.

ACKNOWLEDGMENT

We thank the Stony Brook Proteomics center, particularly Dr. A. Koller, for help with mass measurements and Professor Erwin London for helpful discussion. We also acknowledge computing time awarded by the New York Center for Computational Science to D.F.G.

SUPPORTING INFORMATION AVAILABLE

Seven figures and one table. Orbitrap mass spectrum, far-UV wavelength scans of proteins, calculated Cα backbone rmsd of the wild-type and F5F_{CN} mutant, calculated probability density functions for the side chain at position 5 in wild-type and F5F_{CN} mutant, calculated probability density functions for distance and inter-residue dihedral angle for the side chains at positions 5 and 25 from molecular dynamics simulations, CD- and fluorescence-monitored urea denaturation of F5F_{CN} and F5F_{CN}Y25F and the thermodynamic parameters calculated from the denaturation experiments. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Selkoe, D. J. (2003) Folding proteins in fatal ways. *Nature 426*, 900–904.
2. Dyer, R. B. (2007) Ultrafast and downhill protein folding. *Curr. Opin. Struct. Biol. 17*, 38–47.
3. Kubelka, J., Hofrichter, J., and Eaton, W. A. (2004) The protein folding "speed limit". *Curr. Opin. Struct. Biol. 14*, 76–88.
4. Fersht, A. R. (2000) Structure and Mechanism in Protein Science - A Guide to Enzyme Catalysis and Protein Folding , W. H. Freeman & Company, New York.
5. Royer, C. A. (2006) Probing protein folding and conformational transitions with fluorescence. *Chem. Rev. 106*, 1769–1784.
6. Tucker, M. J., Oyola, R., and Gai, F. (2006) A novel fluorescent probe for protein binding and folding studies: p-cyano-phenylalanine. *Biopolymers 83*, 571–576.
7. Tang, J., Signarvic, R. S., DeGrado, W. F., and Gai, F. (2007) Role of helix nucleation in the kinetics of binding of Mastoparan X to phospholipid bilayer. *Biochemistry 46*, 13856–12863.
8. Tucker, M. J., Tang, J., and Gai, F. (2006) Probing the kinetics of membrane-mediated helix folding. *J. Phys. Chem. B 110*, 8105–8109.
9. Marek, P., Gupta, R., and Raleigh, D. P. (2008) The fluorescent amino acid p-cyanophenylalanine provides an intrinsic probe of amyloid formation. *ChemBioChem 9*, 1372–1374.
10. Aprilakis, K. N., Taskent, H., and Raleigh, D. P. (2007) Use of the novel fluorescent amino acid p-cyanophenylalanine offers a direct probe of hydrophobic core formation during the folding of the N-terminal domain of the ribosomal protein L9 and provides evidence for two-state folding. *Biochemistry 46*, 12308–12313.
11. Miyake-Stoner, S. J., Miller, A. M., Hammill, J. T., Peeler, J. C., Hess, K. R., Mehl, R. A., and Brewer, S. H. (2009) Probing protein folding using site-specifically encoded unnatural amino acids as FRET donors with tryptophan. *Biochemistry 48*, 5953–5962.
12. Getahun, Z., Huang, C. Y., Wang, T., De Leon, B., DeGrado, W. F., and Gai, F. (2003) Using nitrile-derivatized amino acids as infrared probes of local environment. *J. Am. Chem. Soc. 125*, 405–411.
13. Suydam, I. T., Snow, C. D., Pande, V. S., and Boxer, S. G. (2006) Electric fields at the active site of an enzyme: direct comparison of experiment with theory. *Science 313*, 200–204.
14. Anil, B., Sato, S., Cho, J. H., and Raleigh, D. P. (2005) Fine structure analysis of a protein folding transition state; distinguishing between hydrophobic stabilization and specific packing. *J. Mol. Biol. 354*, 693–705.
15. Horng, J. C., and Raleigh, D. P. (2003) Phi-Values beyond the ribosomally encoded amino acids: kinetic and thermodynamic

consequences of incorporating trifluoromethyl amino acids in a globular protein. *J. Am. Chem. Soc. 125*, 9286–9287.
16. Wang, L., Xie, J., and Schultz, P. G. (2006) Expanding the genetic code. *Annu. Rev. Biophys. Biomol. Struct. 35*, 225–249.
17. Jackson, J. C., Hammill, J. T., and Mehl, R. A. (2007) Site-specific incorporation of a (19)F-amino acid into proteins as an NMR probe for characterizing protein structure and reactivity. *J. Am. Chem. Soc. 129*, 1160–1166.
18. Hammill, J. T., Miyake-Stoner, S., Hazen, J. L., Jackson, J. C., and Mehl, R. A. (2007) Preparation of site-specifically labeled fluorinated proteins for 19F-NMR structural characterization. *Nat. Protoc. 2*, 2601–2607.
19. Brunger, A. T., and Karplus, M. (1988) Polar hydrogen positions in proteins: empirical energy placement and neutron diffraction comparison. *Proteins 4*, 148–156.
20. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., and Karplus, M. (1983) Charmm - a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem. 4*, 187–217.
21. MacKerell, A. D., Bashford, D., Bellott, M., Dunbrack, R. L., Evanseck, J. D., Field, M. J., Fischer, S., Gao, J., Guo, H., Ha, S., Joseph-McCarthy, D., Kuchnir, L., Kuczera, K., Lau, F. T. K., Mattos, C., Michnick, S., Ngo, T., Nguyen, D. T., Prodhom, B., Reiher, W. E., Roux, B., Schlenkrich, M., Smith, J. C., Stote, R., Straub, J., Watanabe, M., Wiorkiewicz-Kuczera, J., Yin, D., and Karplus, M. (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B 102*, 3586–3616.
22. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem. 26*, 1781–1802.
23. Wang, L., Brock, A., Herberich, B., and Schultz, P. G. (2001) Expanding the genetic code of Escherichia coli. *Science 292*, 498–500.
24. Yin, D. (1997) Ph.D. Thesis, Parameterization for empirical force field calculations and a theoretical study of membrane permeability of pyridine derivative, Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland.
25. Link, A. J., Mock, M. L., and Tirrell, D. A. (2003) Non-canonical amino acids in protein engineering. *Curr. Opin. Biotechnol. 14*, 603–609.