THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

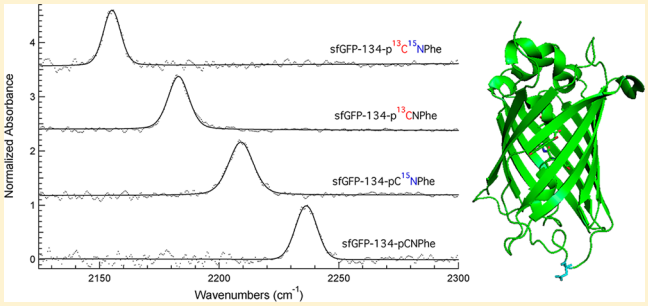
Expanding the Utility of 4-Cyano-L-Phenylalanine As a Vibrational Reporter of Protein Environments

Christopher G. Bazewicz, Jacob S. Lipkin, Emily E. Smith, Melanie T. Liskov, and Scott H. Brewer*

Department of Chemistry, Franklin & Marshall College, Lancaster, Pennsylvania 17604-3003, United States

Supporting Information

ABSTRACT: The ability to genetically incorporate amino acids modified with spectroscopic reporters site-specifically into proteins with high efficiency and fidelity has greatly enhanced the ability to probe local protein structure and dynamics. Here, we have synthesized the unnatural amino acid (UAA), 4-cyano-L-phenylalanine (pCNPhe), containing the nitrile vibrational reporter and three isotopomers (^{15}N, ^{13}C, ^{13}C^{15}N) of this UAA to enhance the ability of pCNPhe to study local protein environments. Each pCNPhe isotopic variant was genetically incorporated in an efficient, site-specific manner into superfolder green fluorescent protein (sfGFP) in response to an amber codon with high fidelity utilizing an engineered, orthogonal aminoacyl-tRNA synthetase. The isotopomers of 4-cyano-L-phenylalanine permitted the nitrile symmetric stretch vibration of these UAAs to be unambiguously assigned utilizing the magnitude and direction of the isotopic shift of this vibration. The sensitivity of the nitrile symmetric stretching frequency of each isotopic variant to the local environment was measured by individually incorporating the probes into two distinct local environments of sfGFP. The UAAs were also utilized in concert to probe multiple local environments in sfGFP simultaneously to increase the utility of 4-cyano-L-phenylalanine.



INTRODUCTION

Amino acids modified with spectroscopic reporters have great promise to significantly enhance the ability to study local protein structure and dynamics in a relatively noninvasive manner.[1−24] A number of these unnatural amino acids (UAAs) can be genetically incorporated into proteins site-specifically utilizing an orthogonal, engineered aminoacyl-tRNA synthetase with no inherent protein size limit.[1,2,4,5,16,19,20,22,23] The two most commonly utilized UAAs modified with vibrational reporters that have been genetically incorporated into proteins in this manner to study local protein structure and dynamics are 4-cyano-L-phenylalanine (pCNPhe, **1**)[4,5] and 4-azido-L-phenylalanine (pN_3Phe).[19,20] These UAAs utilize the nitrile symmetric stretch vibration of the two-atom nitrile group and the azide asymmetric stretch vibration of the three-atom azide group, respectively, as vibrational reporters of local protein environments. Both of these vibrational modes result in a relatively intense IR absorbance band (compared with the Amide I transition)[11] that is sensitive to local protein environments and appears in a relatively clear region of the IR spectrum.[2−6,19,20,25−35]

pCNPhe has been used to study the MLCK peptide−calmodulin complex,[3] the mastoparan-X peptide,[25,26] myoglobin,[4] the N-terminal domain of the ribosomal protein L9,[5] ribonuclease S,[27,28] the villin headpiece subdomain,[6,29] the human islet amyloid polypeptide,[30] the amyloid-β peptide (Aβ16−22),[31] and cytochrome *c*,[32] in which pCNPhe was incorporated either genetically or by synthetic methods. Similarly, pN_3Phe has been utilized to study the G-protein coupled receptor rhodopsin, where pN_3Phe was incorporated genetically using amber codon suppression technology.[19,20]

A recent direct comparison of a nitrile and azide modified nucleoside, 2′-azido-5-cyano-2′-deoxyuridine (N_3CNdU), has shown that the nitrile IR absorbance band is sharper (less broad) than the azide IR absorbance band, but it is slightly less sensitive to the local environment.[36] The peak extinction coefficient of the azide IR absorbance band in N_3CNdU was measured to be approximately twice as large as the nitrile IR absorbance band in an aqueous solution,[36] which is similar to the estimated peak extinction coefficient difference between pCNPhe and pN_3Phe.[2] The higher peak extinction coefficient and greater sensitivity to local environment of the azide oscillator compared with the nitrile oscillator suggest that pN_3Phe is the preferred vibrational UAA reporter. However, the azide group represents a greater possible structural perturbation than the nitrile group, and pN_3Phe has been shown to be photoreactive.[37,38] Both of these characteristics—namely, the photoreactivity of pN_3Phe—diminish the ability of this UAA to serve as an effective vibrational reporter of local protein environments. Therefore, we focused on enhancing the utility of the less invasive and more stable vibrational reporter, pCNPhe, which can be incorporated genetically into proteins

Received: July 11, 2012
Revised: August 7, 2012
Published: August 22, 2012

with site specificity to serve as a sensitive probe of the local protein environment.

We have synthesized 4-cyano-L-phenylalanine (**1**) and three isotopomers (^{15}N (**1a**), ^{13}C (**1b**), ^{13}C^{15}N (**1c**); see Figure 1) of

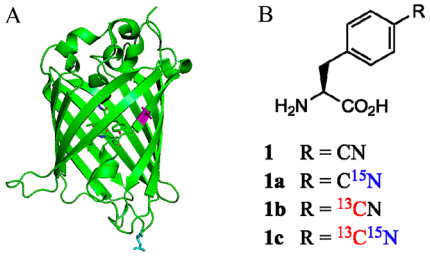


Figure 1. A. Structure of wt-sfGFP (PDB ID 2B3P) with site 134 (cyan) and site 150 (purple) highlighted. B. Structure of 4-cyano-L-phenylalanine (pCNPhe, **1**) and the ^{15}N (pC^{15}NPhe, **1a**), ^{13}C (p^{13}CNPhe, **1b**), and ^{13}C^{15}N (p^{13}C^{15}NPhe, **1c**) isotopically labeled versions of this unnatural amino acid.

this UAA to address the problem of the lower peak extinction coefficient of the nitrile IR absorbance band of pCNPhe (~220 M^{-1} cm^{-1} in THF)[2] compared with the azide IR absorbance band of pN_3Phe. Each of these UAAs (**1**, **1a**, **1b**, and **1c**) was successfully incorporated in an efficient, site-specific manner into the 247-residue monomeric protein, superfolder green fluorescent protein (sfGFP, Figure 1), in response to an amber codon with high fidelity utilizing an engineered, orthogonal aminoacyl-tRNA synthetase. The isotopic shift of the nitrile symmetric stretching frequency was used to unambiguously assign the nitrile symmetric stretch vibration of **1**, **1a**, **1b**, and **1c** in sfGFP. The sensitivity of the position of this band to local protein environments was also measured for **1**, **1a**, **1b**, and **1c** by incorporating each of these UAAs individually into a solvated and partially buried (dehydrated) site in the protein. The ability of pCNPhe to probe local environments was enhanced by utilizing each of these UAAs in concert to probe multiple distinct local protein environments in sfGFP simultaneously. In addition, future potential applications of **1**, **1a**, **1b**, and **1c** are discussed.

EXPERIMENTAL SECTION

General Information. Chemical reagents were purchased from Sigma-Aldrich and Acros Organics and used without further purification. ^{15}N-Labeled formamide (98% ^{15}N enrichment) and deuterium oxide (99.9% D enrichment) were purchased from Cambridge Isotope Laboratories. DH10B cells and *pBadA* were purchased from Invitrogen. All aqueous solutions were prepared with 18 MΩ-cm water.

NMR spectra were obtained at the following frequencies using a Varian INOVA 500 multinuclear Fourier transform NMR spectrometer: ^{1}H, 499.7 MHz; ^{13}C, 125 MHz; and ^{15}N, 50.5 MHz. Chemical shifts are reported in parts per million (ppm), and coupling constants (J) are reported in hertz (Hz). ^{1}H spectra in D_2O were referenced to the solvent peak at 4.67 ppm, ^{13}C NMR spectra in D_2O were referenced to 3-(trimethylsilyl)-1-propanesulfonic acid sodium salt, and ^{15}N spectra were referenced to a 100 mM solution of ^{15}N-labeled formamide in DMSO ($HCONH_2$ = 0.0 ppm).

Synthesis of 4-Cyano-L-Phenylalanine Hydrochloride (1·HCl). *N*-(*tert*-Butoxycarbonyl)-4-cyano-L-phenylalanine methyl ester (1.13 g, 3.73 mmol), synthesized as previously described,[39] was dissolved in a THF/H_2O (3:1, 43 mL) mixture in a round-bottom flask. Lithium hydroxide monohydrate (174 mg, 4.14 mmol) was added to the round-bottom flask, and the reaction mixture was stirred for 16 h at room temperature. The pH of the reaction mixture was adjusted to ~2.5 with 0.5 M sodium bisulfate, and the aqueous layer was extracted with ethyl acetate (2 × 54 mL). The organic layer was washed sequentially with water (64 mL) and brine (64 mL), dried over magnesium sulfate, filtered through a Celite cake, and concentrated in vacuo to yield *N*-(*tert*-butoxycarbonyl)-4-cyano-L-phenylalanine, a white solid (884 mg). *N*-(*tert*-Butoxycarbonyl)-4-cyano-L-phenylalanine (884 mg, 3.04 mmol) was subsequently dissolved in 2.5 M HCl in 1,4-dioxane (7.61 mL), and the resulting reaction mixture was stirred for 4 h at room temperature in a round-bottom flask. The reaction mixture was concentrated, and pentane was added to precipitate **1·HCl**. The white solid product was isolated by filtration to give 629 mg (75%). ^{1}H NMR (D_2O): δ 3.21 (dd, 1H, C_β–H_1, J = 7.8 Hz, J = 14.7 Hz), 3.32 (dd, 1H, C_β–H_2, J = 5.9 Hz, J = 14.7 Hz), 4.23 (dd, 1H, C_α–H, J = 6.1 Hz, J = 7.6 Hz), 7.40 (d, 2H, aromatic C–H, J = 8.3 Hz), 7.69 (d, 2H, aromatic C–H, J = 8.3 Hz).

Synthesis of ^{15}N-Labeled 4-Cyano-L-Phenylalanine Hydrochloride (1a·HCl). 1a·HCl was synthesized by the same procedure as **1·HCl**, except that ^{15}N-labeled *N*-(*tert*-butoxycarbonyl)-4-cyano-L-phenylalanine methyl ester, synthesized as previously described,[39] was used instead of the unlabeled version. ^{1}H NMR (D_2O): δ 3.22 (dd, 1H, C_β–H_1, J = 7.4 Hz, J = 14.7 Hz), 3.33 (dd, 1H, C_β–H_2, J = 5.9 Hz, J = 14.7 Hz), 4.25 (dd, 1H, C_α–H, J = 6.2 Hz, J = 7.6 Hz), 7.40 (d, 2H, aromatic C–H, J = 8.3 Hz), 7.69 (d, 2H, aromatic C–H, J = 8.4 Hz); ^{15}N NMR (D_2O): δ 134.3 (s, C^{15}N).

Synthesis of ^{13}C-Labeled 4-Cyano-L-Phenylalanine Hydrochloride (1b·HCl). 1b·HCl was synthesized by the same procedure as **1·HCl** except that ^{13}C-labeled *N*-(*tert*-butoxycarbonyl)-4-cyano-L-phenylalanine methyl ester, synthesized as previously described,[39] was used instead of the unlabeled version. ^{1}H NMR (D_2O): δ 3.21 (dd, 1H, C_β–H_1, J = 7.4 Hz, J = 14.7 Hz), 3.32 (dd, 1H, C_β–H_2, J = 5.9 Hz, J = 14.7 Hz), 4.23 (dd, 1H, C_α–H, J = 6.2 Hz, J = 7.6 Hz), 7.40 (d, 2H, aromatic C–H, J = 8.3 Hz), 7.69 (dd, 2H, aromatic C–H, J = 5.4 Hz, J = 8.3 Hz); ^{13}C NMR (D_2O): δ 119.5 (s, ^{13}CN).

Synthesis of ^{13}C^{15}N-Labeled 4-Cyano-L-Phenylalanine Hydrochloride (1c·HCl). 1c·HCl was synthesized by the same procedure as **1·HCl** except that ^{13}C^{15}N-labeled *N*-(*tert*-butoxycarbonyl)-4-cyano-L-phenylalanine methyl ester, synthesized as previously described,[39] was used instead of the unlabeled version. ^{1}H NMR (D_2O): δ 3.21 (dd, 1H, C_β–H_1, J = 7.4 Hz, J = 14.7 Hz), 3.32 (dd, 1H, C_β–H_2, J = 5.9 Hz, J = 14.7 Hz), 4.23 (dd, 1H, C_α–H, J = 6.1 Hz, J = 7.6 Hz), 7.39 (d, 2H, aromatic C–H, J = 8.3 Hz), 7.69 (dd, 2H, aromatic C–H, J = 5.4 Hz, J = 8.3 Hz); ^{13}C NMR (D_2O): δ 119.5 (d, ^{13}C^{15}N, J = 19.1 Hz); ^{15}N NMR (D_2O); δ 134.2 (d, ^{13}C^{15}N, J = 19.5 Hz).

Expression and Purification of sfGFP Constructs. The wild-type sfGFP (wt-sfGFP)[40] gene was inserted into *pBadA*, generating *pBad-sfGFP*. The gene was codon-optimized for *Escherichia coli* and contained a C-terminal 6-His affinity tag to aid in purification.[23,41] The codons for D134 and N150 were individually replaced by site-directed mutagenesis with the amber stop codon (TAG), generating *pBad-sfGFP-134TAG* and *pBad-sfGFP-150TAG*, respectively. Replacement of the codons for both D134 and N150 with the amber codon generated *pBad-sfGFP-134,150TAG*. The aminoacyl-tRNA synthetase for the incorporation of pCNPhe was inserted into

dx.doi.org/10.1021/jp306886s | J. Phys. Chem. B 2012, 116, 10824–10831

pDule, generating *pDule-pCNPhe*.[22] Each of these plasmids was obtained from Dr. Ryan A. Mehl (Oregon State University).

pBad-sf GFP was transformed into DH10B *E. coli* cells, while *pBad-sf GFP-134TAG*, *pBad-sf GFP-150TAG*, and *pBad-sf GFP-134,150TAG* were individually cotransformed with *pDule-pCNPhe* into DH10B *E. coli* cells. The transformed cells were used to inoculate 5 mL of noninducing media that was grown to saturation while shaking (250 rpm) at 37 °C. A portion (2.5 mL) of the cultured cells was used to inoculate 250 mL of autoinduction media. The autoinduction media contained **1**, **1a**, **1b**, or **1c** (or all) at a UAA concentration of 1 mM per amber codon, except for negative control experiments, in which the UAAs were excluded from the autoinduction media. The cells from the autoinduction media were collected by centrifugation after shaking at 37 °C for 24−30 h, and the expressed protein was purified using TALON cobalt ion-exchange chromatography (Clontech), similar to previous procedures.[23,41]

Site-specific incorporation of **1**, **1a**, **1b**, or **1c** into site 134 or 150 with high efficiency and fidelity was verified by SDS-PAGE and electrospray ionization quadrupole time-of-flight (ESI-Q-TOF) mass analysis. Typical yields for the sfGFP constructs containing one of the UAAs was 40 mg of purified protein per liter of autoinduction media calculated using the extinction coefficient of sfGFP at 488 nm.[40]

Equilibrium FTIR Measurements. Equilibrium FTIR absorbance spectra were recorded on a Bruker Vertex 70 FTIR spectrometer equipped with a globar source, KBr beamsplitter, and a liquid-nitrogen-cooled mercury cadmium telluride (MCT) detector. The spectra were recorded using a temperature-controlled transmission cell consisting of calcium fluoride windows with a path length of ∼100 μm. The temperature was measured using an embedded thermocouple in the transmission cell. The spectra were the result of 1024 scans recorded at a resolution of 1.0 cm^{-1}. The intensity normalized and baseline corrected absorbance spectra were fit to a line shape function that consisted of a linear combination of a Gaussian and Lorentzian function[42] in Igor Pro (Wavemetrics).

Density Functional Theory (DFT) Calculations. The DFT calculations of **1**, **1a**, **1b**, and **1c** with or without an explicit water molecule were carried out using the quantum chemical software package Gaussian 03.[43] The gas phase geometry optimizations and vibrational analyses were performed on a multiprocessor Mac Pro computer using the B3PW91 density functional[44,45] and a 6-31++G(d,p) basis set.[46,47] The structures were generated and the vibrational modes visualized using the graphical user interface GaussView 4.

RESULTS AND DISCUSSION

Structure of wt-sfGFP. The structure of the 247 residue protein, wt-sfGFP, is shown in Figure 1A with residues 134 (cyan) and 150 (purple) highlighted. This β-barrel monomeric protein consists of 47% β-sheet and 10% helical structure. Residue 134 is located in a loop region of the protein; residue 150 is located in a β-sheet. Previous solvent accessible surface area (SASA) calculations[23] have demonstrated that site 134 represents a fully solvated position in the protein, and site 150 represents a partially buried (desolvated) position in sfGFP due to neighboring amino acid side chains.

Incorporation of 1, 1a, 1b, and 1c into Site 134 or Site 150 of sfGFP. The unnatural amino acid 4-cyano-L-phenylalanine (**1**) and three newly synthesized isotopomers (^{15}N (**1a**), ^{13}C (**1b**), $^{13}C^{15}N$ (**1c**); Figure 1B) of this UAA were genetically incorporated in an efficient, site-specific manner into sfGFP in response to an amber codon with high fidelity utilizing an engineered, orthogonal aminoacyl-tRNA synthetase. The individual incorporation of **1**, **1a**, **1b**, and **1c** into site 134 of sfGFP resulted in the production of sfGFP-134-pCNPhe, sfGFP-134-p$C^{15}N$Phe, sfGFP-134-p^{13}CNPhe, and sfGFP-134-p$^{13}C^{15}N$Phe, respectively, whereas the protein constructs sfGFP-150-pCNPhe, sfGFP-150-p$C^{15}N$Phe, sfGFP-150-p^{13}CNPhe, and sfGFP-150-p$^{13}C^{15}N$Phe were produced when the corresponding UAAs were individually incorporated into site 150 of the protein. The incorporation of these UAAs into site 134 or 150 was verified by SDS-PAGE (see Figure 2) and ESI-Q-TOF mass analysis (see the Supporting Information). SDS-PAGE also confirmed the high fidelity of the UAA incorporation (Figure 2, lanes 7 and 12).

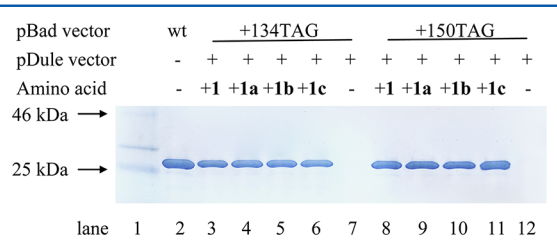


Figure 2. Coomassie blue stained tris−glycine SDS-PAGE illustrating efficient, site-specific incorporation of **1**, **1a**, **1b**, or **1c** with high fidelity into sfGFP. The protein constructs were expressed from *pBad-sf GFP* (wt-sfGFP, lane 2); *pBad-sf GFP-134TAG* and *pDule-pCNPhe* (lanes 3−7) in the presence (lanes 3−6) or absence (lane 7) of **1**, **1a**, **1b**, or **1c**, respectively; or *pBad-sf GFP-150TAG* and *pDule-pCNPhe* (lanes 8−12) in the presence (lanes 8−11) or absence (lane 12) of **1**, **1a**, **1b**, or **1c**, respectively.

Unambiguous Assignment of the Nitrile Symmetric Stretch Vibration of 1, 1a, 1b, and 1c in sfGFP. *Site 134.* The protein FTIR absorbance spectrum for sfGFP-134-pCNPhe in the region of the nitrile symmetric stretch vibration (2125−2300 cm^{-1}) is shown in Figure 3A. The spectrum shows a single absorbance band in this region at 2236.4 cm^{-1}. The position and profile of this band are indicative of the nitrile IR absorbance band of **1** incorporated in the 134 site of sfGFP arising from the nitrile symmetric stretch vibration of **1**. However, this assignment can be made unambiguously through the use of sfGFP-134-p$C^{15}N$Phe, sfGFP-134-p^{13}CNPhe, or sfGFP-134-p$^{13}C^{15}N$Phe, whose protein linear IR absorbance spectra are presented in Figure 3A. These spectra show a single, sharp absorbance band in this region at 2209.1, 2183.0, and 2155.2 cm^{-1}, respectively, that can be modeled by a single line shape function.

The experimental isotopic red shift of the nitrile symmetric stretch vibration between sfGFP-134-pCNPhe and sfGFP-134-p$C^{15}N$Phe is 27 cm^{-1}. This shift is in agreement with the predicted gas phase DFT calculated isotopic red shift (29 cm^{-1}) of the nitrile symmetric stretch vibration of **1** upon ^{15}N labeling of the nitrile group. Similarly, the isotopic red shift in the nitrile symmetric stretch vibration between sfGFP-134-pCNPhe and sfGFP-134-p^{13}CNPhe is 53 cm^{-1}, and the isotopic red shift is 81 cm^{-1} between sfGFP-134-pCNPhe and sfGFP-134-p$^{13}C^{15}N$Phe. Each of these isotopic shifts is in agreement with the direction and magnitude of the DFT calculated isotopic red shift of **1** in the gas phase upon ^{13}C and $^{13}C^{15}N$ labeling of the nitrile group (56 and 86 cm^{-1}, respectively).

dx.doi.org/10.1021/jp306886s | J. Phys. Chem. B 2012, 116, 10824−10831

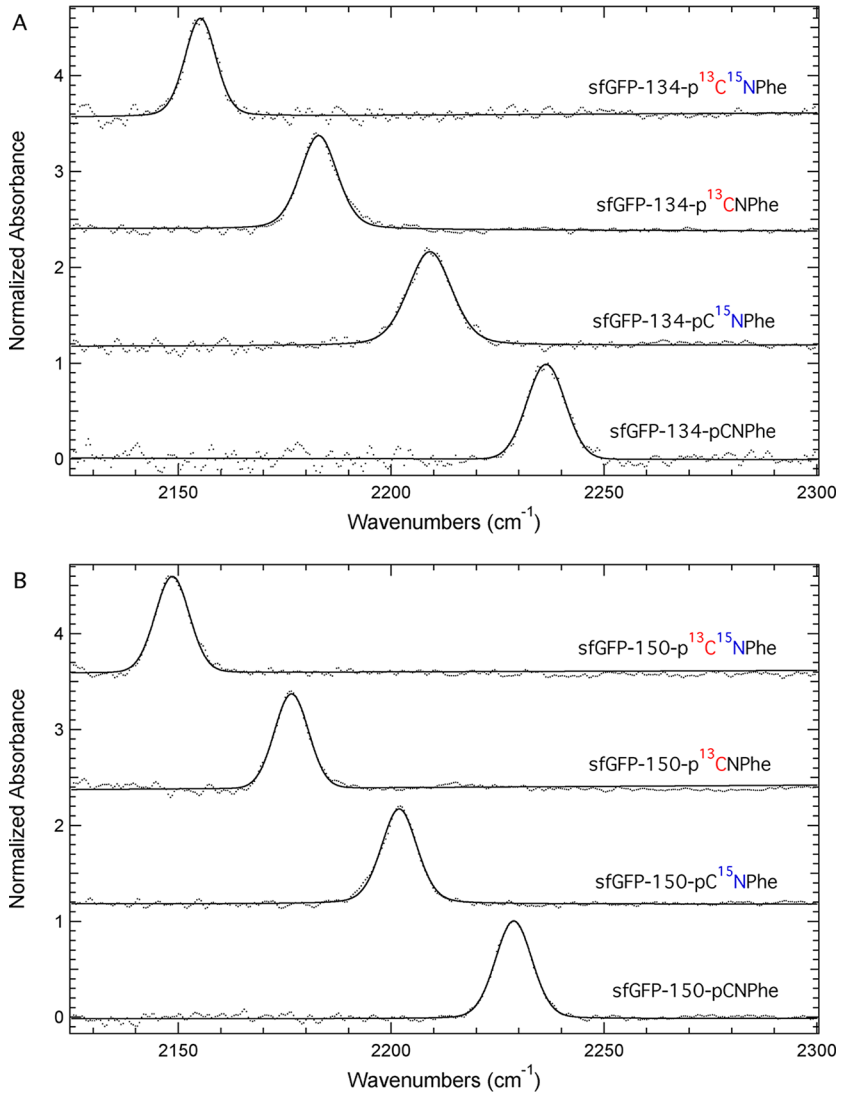


Figure 3. FTIR absorbance spectra of sfGFP constructs containing **1**, **1a**, **1b**, or **1c** at either site 134 (panel A) or site 150 (panel B) in the protein. The protein samples were dissolved to a concentration of ~1 mM in a pH 7.3 aqueous buffer containing 50 mM sodium phosphate and 150 mM sodium chloride. The spectra were recorded at 25 °C, baseline-corrected, intensity-normalized, and fit (solid curves) to a line shape function consisting of a linear combination of a Gaussian and Lorentzian function.

Thus, the comparison of the absorbance band positions in this region among sfGFP-134-pCNPhe, sfGFP-134-pC^{15}NPhe, sfGFP-134-p^{13}CNPhe, and sfGFP-134-p^{13}C^{15}NPhe permits the unambiguous assignment of these bands as resulting from the nitrile symmetric stretch vibration of **1**, **1a**, **1b**, and **1c**, respectively, incorporated at site 134 of sfGFP.

The positions of the nitrile symmetric stretch vibrations also suggest that site 134 in the protein is solvent-exposed, which is in agreement with previous SASA calculations.[23] For instance, the nitrile symmetric stretch vibration of sfGFP-134-pCNPhe (2236.4 cm^{-1}) occurs at a similar frequency compared with the nitrile symmetric stretch vibration of **1** dissolved in water (2237.2 cm^{-1}).[3] The free UAA in the aqueous solution provides a reference for the nitrile symmetric stretching frequency of **1** in a fully solvated (hydrated) environment. Thus, the comparison between the position of the nitrile symmetric stretch vibrations of sfGFP-134-pCNPhe and the free UAA dissolved in water provides evidence for a highly solvated nitrile group of **1** at position 134 in the protein.

It is also important to note that the nitrile IR absorbance bands for each of these protein constructs are resolved from each other, suggesting that **1**, **1a**, **1b**, and **1c** can be used in concert to probe multiple local protein environments simultaneously, as shown and discussed below.

Site 150. This methodology for utilizing isotopomers of **1** to unambiguously assign the nitrile symmetric stretch vibration in proteins was validated by individually incorporating **1**, **1a**, **1b**, and **1c** into site 150 of sfGFP. Figure 3B shows the FTIR absorbance spectrum of sfGFP-150-pCNPhe in the region of the nitrile symmetric stretch vibration (2125−2300 cm^{-1}). This spectrum shows a single, sharp band located at 2228.8 cm^{-1}. Figure 3B also includes the linear IR absorbance spectra of sfGFP-150-pC^{15}NPhe, sfGFP-150-p^{13}CNPhe, and sfGFP-150-p^{13}C^{15}NPhe. These spectra show a single, sharp absorbance band in this region at 2201.9, 2176.6, and 2148.6 cm^{-1}, respectively. The experimentally measured isotopic red shifts for the protein constructs are 27, 52, and 80 cm^{-1}, respectively, relative to sfGFP-150-pCNPhe. These isotopic shifts are in agreement with the DFT predicted isotopic shifts for the nitrile

dx.doi.org/10.1021/jp306886s | J. Phys. Chem. B 2012, 116, 10824−10831

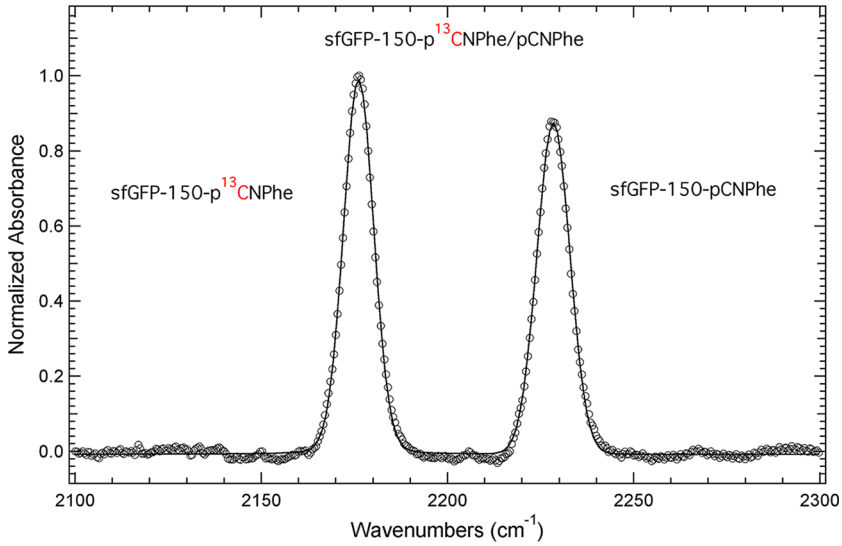


Figure 4. FTIR absorbance spectra of sfGFP containing either **1** or **1b** at site 150 in the protein. The protein mixture was obtained from DH10B cells cotransformed with *pBad-sf GFP-150TAG* and *pDule-pCNPhe* with approximately equal molar amounts of **1** and **1b** in the autoinduction media. The protein sample was dissolved to a concentration of ∼1 mM in a pH 7.3 aqueous buffer containing 50 mM sodium phosphate and 150 mM sodium chloride. The spectrum (open circles) was recorded at 25 °C, baseline-corrected, intensity-normalized, and fit (solid curve) to two line shape functions, each consisting of a linear combination of a Gaussian and Lorentzian function.

symmetric stretching frequency of the UAAs and thus indicate that these bands correspond to the nitrile symmetric stretch vibration of **1a**, **1b**, and **1c** in sfGFP at site 150, respectively. This result illustrates the ability of this methodology to unambiguously assign the nitrile symmetric stretch vibration in different local protein environments.

The position of the nitrile symmetric stretch vibration for sfGFP-150-pCNPhe (2228.8 cm^{-1}) suggests that the nitrile group of **1** is partially buried (desolvated) at site 150 in the protein. For instance, the position of the nitrile symmetric stretch vibration of the Fmoc-protected version of **1** in THF occurs at 2228.5 cm^{-1}.[3] This frequency provides a reference for the position of the nitrile symmetric stretch vibration corresponding to the nitrile group in a low dielectric environment that is not hydrated. The similarity between these nitrile symmetric stretching frequencies suggests that the nitrile group of **1** at site 150 is at least partially buried (dehydrated), which is in agreement with previous SASA calculations.[23]

The position of the nitrile symmetric stretch vibration in sfGFP-150-pCNPhe is 7.6 cm^{-1} red-shifted from the corresponding band in sfGFP-134-pCNPhe. The direction and magnitude of this shift are in qualitative agreement with simple DFT calculations exploring the effects of H-bonding on the nitrile symmetric stretching frequency. Specifically, the nitrile symmetric stretching frequency of **1** was calculated to red-shift 8.4 cm^{-1} upon removal of a single explicit water molecule in a H-bonding arrangement between the nitrile group and the water molecule characterized by a C≡N···H angle of 169° in the gas phase. This result is in qualitative agreement with more sophisticated theoretical and computational studies that have explored the effect of hydrogen bonding between water molecule(s) and the nitrile group on the position and profile of the nitrile IR absorbance band.[48−53] Similarly, the nitrile symmetric stretch vibration of **1a**, **1b**, and **1c** at site 150 was red-shifted 7.2, 6.4, and 6.6 cm^{-1}, respectively, relative to site 134. Thus, the nitrile symmetric stretch vibration of **1a**, **1b**, and **1c**, in addition to **1**, can be effectively utilized to probe distinct local protein environments. This result is critical for **1**, **1a**, **1b**, and **1c** to be used in concert to probe multiple local protein environments simultaneously.

The methodology previously discussed for the accurate assignment of the nitrile symmetric stretch vibration of **1**, **1a**, **1b**, and **1c** in proteins involves the measurement of at least two protein IR absorbance spectra. Since the nitrile IR absorbance bands of **1**, **1a**, **1b**, and **1c** are resolved from each other, the unambiguous assignment of this band can be performed in a single experiment. Figure 4 shows the protein FTIR spectrum of sfGFP produced with both **1** and **1b** present in autoinduction media containing DH10B *E. coli* cells cotransformed with *pBad-sf GFP-150TAG* and *pDule-pCNPhe*. This expression results in a mixture of sfGFP-150-pCNPhe and sfGFP-150-p^{13}CNPhe, in which the relative amount of each protein construct in the mixture is related to the molar ratio of **1** and **1b** in the autoinduction media. The IR spectrum of the purified protein mixture shows two IR absorbance bands in the region 2100−2300 cm^{-1}, separated by 52 cm^{-1}, as expected on the basis of previous measurements of the isotopic shift of the nitrile symmetric stretching frequency between **1** and **1b**. Thus, in a single measurement, the IR absorbance bands at 2228.5 and 2176.1 cm^{-1} can be accurately assigned to the nitrile symmetric stretch vibrations of sfGFP-150-pCNPhe and sfGFP-150-p^{13}CNPhe, respectively. Although **1** and **1b** were utilized here, any combination of **1**, **1a**, **1b**, and **1c** could potentially be used for the assignment of the nitrile symmetric stretch vibration of these UAAs without the need for two or more protein expressions, purifications, and IR spectra. Thus, this slightly modified procedure can be used to efficiently and effectively assign the nitrile symmetric stretch vibration of **1**, **1a**, **1b**, and **1c** in proteins.

Probing Multiple Local Protein Environments Simultaneously with 1, 1a, 1b, and 1c. 4-Cyano-L-phenylalanine and its three isotopomers can be utilized in concert to probe multiple distinct local protein environments simultaneously, since the nitrile IR absorbance bands for each of these UAAs are resolved from one another and each band is similarly

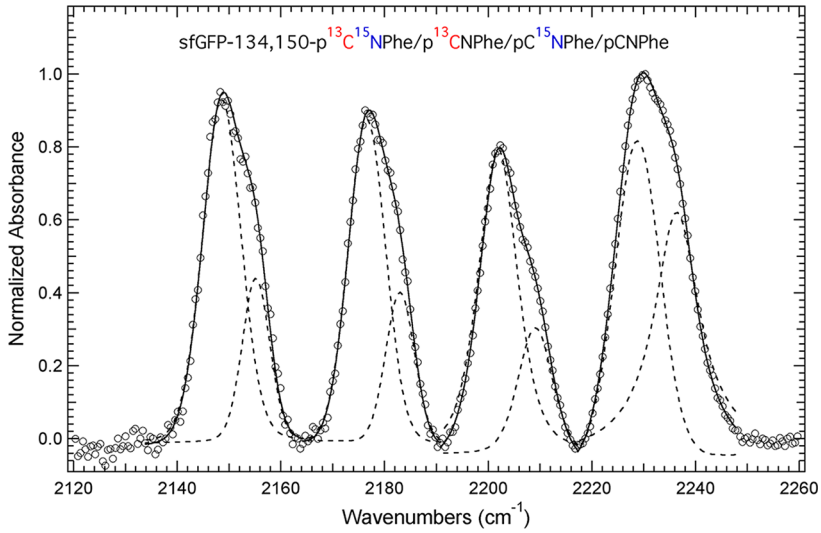


Figure 5. FTIR absorbance spectrum of sfGFP containing either **1**, **1a**, **1b**, or **1c** at sites 134 and 150 in the protein. The protein mixture was obtained from DH10B cells cotransformed with *pBad-sfGFP-134,150TAG* and *pDule-pCNPhe* with approximately equal molar amounts of **1**, **1a**, **1b**, and **1c** in the autoinduction media. The protein sample was dissolved to a concentration of ∼1 mM in a pH 7.3 aqueous buffer containing 50 mM sodium phosphate and 150 mM sodium chloride. The spectrum (open circles) was recorded at 25 °C, baseline-corrected, intensity-normalized, and fit (solid curve) to eight line shape functions (individual components, dashed curves; overall fit, solid curves), each consisting of a linear combination of a Gaussian and Lorentzian function.

sensitive to the local environment. This resolution is the result of the magnitude of the isotopic shift of the nitrile symmetric stretching frequency of **1a**, **1b**, and **1c** and the sharpness of the corresponding nitrile IR absorbance band. As an illustration of the ability to probe multiple local protein environments at the same time, each of the UAAs was simultaneously incorporated into two distinct sites in sfGFP. Specifically, each of the UAAs (**1**, **1a**, **1b**, and **1c**) was added in nearly equal molar amounts to autoinduction media containing DH10B *E. coli* cells cotransformed with *pBad-sfGFP-134,150TAG* and *pDule-pCNPhe*. This expression theoretically produced a mixture of 16 protein constructs, since each of the four UAAs could be incorporated at either site in the protein. The fidelity of the genetic incorporation of **1**, **1a**, **1b**, and **1c** into sites 134 and 150 of sfGFP was verified by SDS-PAGE (see Supporting Information Figure S2, lane 6).

Figure 5 shows the FTIR absorbance spectrum of the resulting protein mixture in the region 2120−2260 cm^{-1}, which contains four resolved nitrile IR absorbance bands corresponding to **1**, **1a**, **1b**, and **1c** in sfGFP going from high to low energy, respectively. Line shape analysis revealed that each of these bands consisted of two spectral components, as shown in Figure 5. The position of the high-frequency, lower-intensity spectral component for each nitrile IR absorbance band corresponds to the nitrile group in a solvent-exposed region of the protein (site 134), whereas the position of the low-frequency, higher-intensity component corresponds to the nitrile group in a partially buried environment in the protein (site 150). Thus, the local environments at site 134 and 150 were successfully probed simultaneously by the four UAAs. This capability significantly enhances the utility of 4-cyano-L-phenylalanine to serve as an efficient and effective vibrational reporter of local protein environments, in part by decreasing the number of experiments required to probe multiple protein environments.

The previous experiment focused on the genetic incorporation of **1**, **1a**, **1b**, and **1c** to probe multiple local protein environments simultaneously. In addition, the Fmoc-protected versions of these UAAs could be site-specifically incorporated at multiple positions in peptides through standard solid-phase peptide synthetic methods, in which each UAA would be incorporated at either a single site or multiple sites in the peptide. If each UAA is incorporated in only a single location, the profile of the nitrile IR absorbance bands could be better utilized to probe the local environment of the nitrile group in the protein.[6]

■ CONCLUSIONS

The utility of the vibrational reporter, 4-cyano-L-phenylalanine, has been substantially enhanced through the use of isotopic labels to effectively probe local protein environments. Specifically, pCNPhe and three isotopomers (^{15}N, ^{13}C, ^{13}C^{15}N) of this UAA were synthesized and genetically incorporated in an efficient, site-specific manner into sfGFP in response to an amber codon with high fidelity utilizing an engineered, orthogonal aminoacyl-tRNA synthetase. The isotopomers of **1** allowed for the unambiguous assignment of the nitrile symmetric stretch vibration in sfGFP by comparing the predicted and experimentally measured isotopic shifts of the UAAs. Each isotopomer was found to be an effective vibrational reporter of local environments, with the nitrile symmetric stretch vibration red-shifting ∼7 cm^{-1} from a solvated position (site 134) to a partially buried position (site 150) in sfGFP. The ability to effectively probe multiple local protein environments simultaneously in sfGFP using each of these UAAs was demonstrated by the genetic incorporation of **1**, **1a**, **1b**, and **1c** into sfGFP at sites 134 and 150. As noted, the Fmoc-protected versions of the UAAs could also be selectively incorporated into a peptide through standard solid-phase peptide synthetic methods to probe multiple local environments simultaneously, since the nitrile IR absorbance band of each of the isotopomers is resolved from that of the others.

An intriguing potential future use of **1**, **1a**, **1b**, and **1c** could be to utilize these UAAs as vibrational distance probes. Zanni and co-workers[54] have suggested that vibrational coupling

between unlabeled and ^{15}N-labeled nitrile groups of 5-cyano-2′-deoxyuridine can be used to infer distances between these oscillators in DNA oligomers. Therefore, the potential ability of synthesized isotopomers of pCNPhe to serve as vibrational distance probes will be explored in the future, which will require 2D IR experiments[6,28,29,55−60] to characterize any potential distance-dependent anharmonic coupling among the nitrile groups of **1**, **1a**, **1b**, and **1c** incorporated in either peptides or proteins.

The isotopomers of **1** could also be utilized as dual NMR/IR probes to aid in the correlation of the observed nitrile symmetric stretching frequency of these isotopomers to the local environment around the nitrile group in proteins.[35,61−63] For instance, **1b** has been recently coupled with ^{13}C NMR and IR spectroscopy.[35] Similarly, **1a** could be coupled with ^{15}N NMR, and **1c** could be coupled with both ^{13}C and ^{15}N NMR in addition to IR spectroscopy.

On the basis of the current and potential future applications of **1**, **1a**, **1b**, and **1c**, current work in our lab is underway to utilize isotopic labels to enhance the utility of current and new UAAs that contain a vibrational reporter, which can be incorporated into proteins genetically with site specificity.

■ ASSOCIATED CONTENT

Supporting Information

ESI-Q-TOF mass analysis of the sfGFP protein constructs, SDS-PAGE of sfGFP mixtures produced with multiple UAAs present in the autoinduction media, DFT-optimized structure of 4-cyano-L-phenylalanine with and without a water molecule in a H-bonding interaction with the nitrile group. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: (717) 358-4766. Fax: (717) 291-4343. E-mail: scott.brewer@fandm.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Lisa Mertzman for obtaining materials and supplies, Beth Buckwalter for acquiring NMR spectra, Kevin J. Hines for experimental contributions, and Carol Strausser for assistance in editing of the manuscript. This work was supported by F&M Hackman funds and NSF (CHE-1053946) to S.H.B.

■ REFERENCES

(1) Xie, J.; Schultz, P. G. *Nat. Rev. Mol. Cell Biol.* **2006**, *7*, 775−782.
(2) Waegele, M. M.; Culik, R. M.; Gai, F. *J. Phys. Chem. Lett.* **2011**, *2*, 2598−2609.
(3) Getahun, Z.; Huang, C. Y.; Wang, T.; De Leon, B.; DeGrado, W. F.; Gai, F. *J. Am. Chem. Soc.* **2003**, *125*, 405−411.
(4) Schultz, K. C.; Supekova, L.; Ryu, Y.; Xie, J.; Perera, R.; Schultz, P. G. *J. Am. Chem. Soc.* **2006**, *128*, 13984−13985.
(5) Taskent-Sezgin, H.; Chung, J.; Patsalo, V.; Miyake-Stoner, S. J.; Miller, A. M.; Brewer, S. H.; Mehl, R. A.; Green, D. F.; Raleigh, D. P.; Carrico, I. *Biochemistry* **2009**, *48*, 9040−9046.
(6) Urbanek, D. C.; Vorobyev, D. Y.; Serrano, A. L.; Gai, F.; Hochstrasser, R. M. *J. Phys. Chem. Lett.* **2010**, *1*, 3311−3315.
(7) Weeks, C. L.; Polishchuk, A.; Getahun, Z.; DeGrado, W. F.; Spiro, T. G. *J. Raman Spectrosc.* **2008**, *39*, 1606−1613.
(8) Fafarman, A. T.; Webb, L. J.; Chuang, J. I.; Boxer, S. G. *J. Am. Chem. Soc.* **2006**, *128*, 13356−13357.
(9) Sigala, P. A.; Fafarman, A. T.; Bogard, P. E.; Boxer, S. G.; Herschlag, D. *J. Am. Chem. Soc.* **2007**, *129*, 12104−12105.
(10) Fafarman, A. T.; Sigala, P. A.; Herschlag, D.; Boxer, S. G. *J. Am. Chem. Soc.* **2010**, *132*, 12811−12813.
(11) Oh, K. I.; Lee, J. H.; Joo, C.; Han, H.; Cho, M. *J. Phys. Chem. B* **2008**, *112*, 10352−10357.
(12) Stafford, A. J.; Ensign, D. L.; Webb, L. J. *J. Phys. Chem. B* **2010**, *114*, 15331−15344.
(13) McMahon, H. A.; Alfieri, K. N.; Clark, C. A. A.; Londergan, C. H. *J. Phys. Chem. Lett.* **2010**, *1*, 850−855.
(14) Taskent-Sezgin, H.; Chung, J. A.; Banerjee, P. S.; Nagarajan, S.; Dyer, R. B.; Carrico, I.; Raleigh, D. P. *Angew. Chem., Int. Ed.* **2010**, *49*, 7473−7475.
(15) Nagarajan, S.; Taskent-Sezgin, H.; Parul, D.; Carrico, I.; Raleigh, D. P.; Dyer, R. B. *J. Am. Chem. Soc.* **2011**, *133*, 20335−20340.
(16) Jackson, J. C.; Hammill, J. T.; Mehl, R. A. *J. Am. Chem. Soc.* **2007**, *129*, 1160−1166.
(17) Waegele, M. M.; Tucker, M. J.; Gai, F. *Chem. Phys. Lett.* **2009**, *478*, 249−253.
(18) Goldberg, J. M.; Batjargal, S.; Petersson, E. J. *J. Am. Chem. Soc.* **2010**, *132*, 14718−14720.
(19) Ye, S. X.; Huber, T.; Vogel, R.; Sakmar, T. P. *Nat. Chem. Biol.* **2009**, *5*, 397−399.
(20) Ye, S. X.; Zaitseva, E.; Caltabiano, G.; Schertler, G. F. X.; Sakmar, T. P.; Deupi, X.; Vogel, R. *Nature* **2010**, *464*, 1386−1389.
(21) van Dongen, S. F. M.; Teeuwen, R. L. M.; Nallani, M.; van Berkel, S. S.; Cornelissen, J. J. L. M.; Nolte, R. J. M.; van Hest, J. C. M. *Bioconjugate Chem.* **2009**, *20*, 20−23.
(22) Miyake-Stoner, S. J.; Miller, A. M.; Hammill, J. T.; Peeler, J. C.; Hess, K. R.; Mehl, R. A.; Brewer, S. H. *Biochemistry* **2009**, *48*, 5953−5962.
(23) Smith, E. E.; Linderman, B. Y.; Luskin, A. C.; Brewer, S. H. *J. Phys. Chem. B* **2011**, *115*, 2380−2385.
(24) Brewer, S. H.; Song, B. B.; Raleigh, D. P.; Dyer, R. B. *Biochemistry* **2007**, *46*, 3279−3285.
(25) Tucker, M. J.; Getahun, Z.; Nanda, V.; DeGrado, W. F.; Gai, F. *J. Am. Chem. Soc.* **2004**, *126*, 5078−5079.
(26) Tang, J.; Signarvic, R. S.; DeGrado, W. F.; Gai, F. *Biochemistry* **2007**, *46*, 13856−13863.
(27) Fafarman, A. T.; Boxer, S. G. *J. Phys. Chem. B* **2010**, *114*, 13536−13544.
(28) Bagchi, S.; Boxer, S. G.; Fayer, M. D. *J. Phys. Chem. B* **2012**, *116*, 4034−4042.
(29) Chung, J. K.; Thielges, M. C.; Fayer, M. D. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 3578−3583.
(30) Marek, P.; Mukherjee, S.; Zanni, M. T.; Raleigh, D. P. *J. Mol. Biol.* **2010**, *400*, 878−888.
(31) Inouye, H.; Gleason, K. A.; Zhang, D.; Decatur, S. M.; Kirschner, D. A. *Proteins* **2010**, *78*, 2306−2321.
(32) Zimmermann, J.; Thielges, M. C.; Seo, Y. J.; Dawson, P. E.; Romesberg, F. E. *Angew. Chem., Int. Ed.* **2011**, *50*, 8333−8337.
(33) Liu, J.; Strzalka, J.; Tronin, A.; Johansson, J. S.; Blasie, J. K. *Biophys. J.* **2009**, *96*, 4176−4187.
(34) Hu, W. H.; Webb, L. J. *J. Phys. Chem. Lett.* **2011**, *2*, 1925−1930.
(35) Bagchi, S.; Fried, S. D.; Boxer, S. G. *J. Am. Chem. Soc.* **2012**, *134*, 10373−10376.
(36) Gai, X. S.; Coutifaris, B. A.; Brewer, S. H.; Fenlon, E. E. *Phys. Chem. Chem. Phys.* **2011**, *13*, 5926−5930.
(37) Zhang, K. C.; Diehl, M. R.; Tirrell, D. A. *J. Am. Chem. Soc.* **2005**, *127*, 10136−10137.
(38) Carrico, I. S.; Maskarinec, S. A.; Heilshorn, S. C.; Mock, M. L.; Liu, J. C.; Nowatzki, P. J.; Franck, C.; Ravichandran, G.; Tirrell, D. A. *J. Am. Chem. Soc.* **2007**, *129*, 4874−4875.
(39) Bazewicz, C. G.; Lipkin, J. S.; Lozinak, K. A.; Watson, M. D.; Brewer, S. H. *Tetrahedron Lett.* **2011**, *52*, 6865−6868.
(40) Pedelacq, J. D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. *Nat. Biotechnol.* **2006**, *24*, 79−88.

(41) Miyake-Stoner, S. J.; Refakis, C. A.; Hammill, J. T.; Lusic, H.; Hazen, J. L.; Deiters, A.; Mehl, R. A. *Biochemistry* **2010**, *49*, 1667−1677.
(42) Maienschein-Cline, M. G.; Londergan, C. H. *J. Phys. Chem. A* **2007**, *111*, 10020−10025.
(43) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; Millam, J. M.; Iyengar, S. S.; Tomasi, J.; Barone, V.; Mennucci, B.; Cossi, M.; Scalmani, G.; Rega, N.; Petersson, G. A.; Nakatsuji, H.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Klene, M.; Li, X.; Knox, J. E.; Hratchian, H. P.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Ayala, P. Y.; Morokuma, K.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Zakrzewski, V. G.; Dapprich, S.; Daniels, A. D.; Strain, M. C.; Farkas, O.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Ortiz, J. V.; Cui, Q.; Baboul, A. G.; Clifford, S.; Cioslowski, J.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Gonzalez, C.; Pople, J. A. *Gaussian 03, Revision E.01*; Gaussian, Inc.: Wallingford, CT, 2004.
(44) Becke, A. D. *J. Chem. Phys.* **1993**, *98*, 5648−5652.
(45) Perdew, J. P.; Wang, Y. *Phys. Rev. B* **1992**, *45*, 13244−13249.
(46) Clark, T.; Chandrasekhar, J.; Spitznagel, G. W.; Schleyer, P. v. R. *J. Comput. Chem.* **1983**, *4*, 294−301.
(47) Hariharan, P. C.; Pople, J. A. *Theor. Chim. Acta* **1973**, *28*, 213−222.
(48) Choi, J. H.; Oh, K. I.; Lee, H.; Lee, C.; Cho, M. *J. Chem. Phys.* **2008**, *128*, 134506.
(49) Oh, K. I.; Choi, J. H.; Lee, J. H.; Han, J. B.; Lee, H.; Cho, M. *J. Chem. Phys.* **2008**, *128*, 154504.
(50) Choi, J. H.; Oh, K. I.; Cho, M. H. *J. Chem. Phys.* **2008**, *129*, 174512.
(51) Lindquist, B. A.; Corcelli, S. A. *J. Phys. Chem. B* **2008**, *112*, 6301−6303.
(52) Lindquist, B. A.; Haws, R. T.; Corcelli, S. A. *J. Phys. Chem. B* **2008**, *112*, 13991−14001.
(53) Lindquist, B. A.; Furse, K. E.; Corcelli, S. A. *Phys. Chem. Chem. Phys.* **2009**, *11*, 8119−8132.
(54) Krummel, A. T.; Zanni, M. T. *J. Phys. Chem. B* **2008**, *112*, 1336−1338.
(55) Cho, M. H. *Chem. Rev.* **2008**, *108*, 1331−1418.
(56) Ganim, Z.; Chung, H. S.; Smith, A. W.; Deflores, L. P.; Jones, K. C.; Tokmakoff, A. *Acc. Chem. Res.* **2008**, *41*, 432−441.
(57) Shim, S. H.; Gupta, R.; Ling, Y. L.; Strasfeld, D. B.; Raleigh, D. P.; Zanni, M. T. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 6614−6619.
(58) Kim, Y. S.; Hochstrasser, R. M. *J. Phys. Chem. B* **2009**, *113*, 8231−8251.
(59) Nydegger, M. W.; Dutta, S.; Cheatum, C. M. *J. Chem. Phys.* **2010**, *133*, 134506.
(60) Tucker, M. J.; Gai, X. S.; Fenlon, E. E.; Brewer, S. H.; Hochstrasser, R. M. *Phys. Chem. Chem. Phys.* **2011**, *13*, 2237−2241.
(61) Watson, M. D.; Gai, X. S.; Gillies, A. T.; Brewer, S. H.; Fenlon, E. E. *J. Phys. Chem. B* **2008**, *112*, 13188−13192.
(62) Gillies, A. T.; Gai, X. S.; Buckwalter, B. L.; Fenlon, E. E.; Brewer, S. H. *J. Phys. Chem. B* **2010**, *114*, 17136−17141.
(63) Gai, X. S.; Fenlon, E. E.; Brewer, S. H. *J. Phys. Chem. B* **2010**, *114*, 7958−7966.