ON THE DEVELOPMENT, CHARACTERIZATION, AND USE OF PROTEIN FLUORESCENCE AND INFRARED SPECTROSCOPIC PROBES

Mary Rose Hilaire

A DISSERTATION

in

Chemistry

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2017

Supervisor of Dissertation

_____

Dr. Feng Gai

Edmund J. and Louise W. Kahn Endowed Term Professor of Chemistry

Graduate Group Chairperson

_____

Dr. Gary A. Molander

Hirschmann-Makineni Professor of Chemistry

Dissertation Committee

Dr. Marsha I. Lester, Chair, Edmund J. Kahn Distinguished Professor of Chemistry

Dr. Zahra Fakhraai, Assistant Professor of Chemistry

Dr. Joseph E. Subotnik, Professor of Chemistry

To the memory of my godfather, Patrick A. Fisher

ACKNOWLEDGMENT

My time at the University of Pennsylvania has been one filled with great growth, learning, triumph, and achievement, which I have enjoyed immensely. I owe gratitude to everyone with whom I have had the pleasure of working.

First, I would like to thank my mentor, Dr. Feng Gai. His curiosity of science pushed me to work harder (and quicker!) to explore and understand the science we were studying. I am grateful for his constant support and encouragement. Because of him, I am a critical, thoughtful scientist. Additionally, I would like to thank the members of my dissertation committee, Drs. Marsha I. Lester, Zahra Fakhraai, and Joseph E. Subotnik, for their time, guidance, abundant advice, and motivating conversations over the past four years.

I would like to acknowledge the external collaborators that I have had the pleasure of working with: Drs. Aaron C. Robinson and Bertrand García-Moreno E. from Johns Hopkins University, Drs. Hyunil Jo and William F. DeGrado from the University of California San Francisco, Dr. Ivan Korendovych from Syracuse University, and Dr. Jinglin Fu from Rutgers University Camden. I appreciate the samples, time, and scientific discussions that I have had with each of them.

I feel lucky to have worked in the Gai Lab with a group of incredible scientists, including past members: Dr. Lev Chuntonov, Dr. Robert Culik, Dr. Ayanjeet Ghosh, Dr. Jianqiang Ma, Dr. Beatrice Markiewicz, Dr. Kwang-Im Oh, Dr. Ileana Pazos, Dr. Kathryn Smith-Dupont, and Dr. Wenkai Zhang and current group members: Rachel Abaskharon, Arusha Acharyya, Ismail Ahmed, Dr. Jian-Xin Chen, Dr. Bei Ding, Dr. Chun-Wei Lin, Debopreeti Mukherjee, Jeffrey Rodgers, Dr. Natalia Rubtsova, Dr. Thomas Troxler, and Dr. Patrick Walsh. I have had the pleasure of working closely with Ismail, Bei, and Tom

on projects, which has been an enjoyable experience. I am grateful to Ismail for his creative ideas, his synthetic skills, and stimulating conversation. Additionally, I am grateful to Tom for being an incredibly generous and patient mentor and for proofreading my thesis.

From the very beginning at Penn, I have had two great scientists and outstanding people as my friends, Rachel Abaskharon and Daniel Straus. I am thankful for our time spend working on homework in the library, discussing science over lunch, and many laughs shared during dinners with our spouses. Moreover, I have had the pleasure of sitting five feet away from Rachel for the past four years in lab. I couldn't ask for a more supportive friend. Thank you for pushing me to not only be a better scientist, but to be a better person!

Last but not least, I would like to thank my family for their continued love and support. First, I would like to thank my husband Kegan, for commiserating with me when I'm upset, celebrating my achievements, and trying so hard to understand my chemistry. Also, I need to thank my parents, James and Rose Ann Mintzer, as I would not have made it this far without their motivational speeches, constant support, and valued advice. To my siblings: my older brother Jimmy Mintzer and his wife Stephanie, my sister Joanna Mintzer, and my younger brothers Michael and Patrick Mintzer, I thank for all the laughter, teasing, and fun. Additionally, I am grateful to my aunt, Mary Kay Fisher, for many dinners over the years that served as a good distraction from my lab work. Finally, I would like to thank my in-laws, Stephen Hilaire and Maureen Lynch and Libby and Patrick Hilaire, for welcoming me to their family and treating me as one of their own. My family is the source of my motivation and I would be lost without them by my side. I am incredible blessed to be able to call all these people my family.

ABSTRACT

ON THE DEVELOPMENT, CHARACTERIZATION, AND USE OF PROTEIN FLUORESCENCE AND INFRARED SPECTROSCOPIC PROBES

Mary Rose Hilaire

Dr. Feng Gai

Proteins possess unique physical and chemical properties that allow them to carry out a wide variety of biological activities and functions. While it is generally understood that a protein's function is dictated by its structure and dynamics, arriving at a molecule-level understanding of the underlying structure-dynamics-function relationship still poses a challenging task in many cases. This is due, at least in part, to the fact that we lack the ability to take snapshots along the reaction coordinate of proteins with sufficient temporal and structural resolution. Therefore, to improve one's ability to acquire site-specific structural and/or environmental information of proteins via either infrared (IR) or fluorescence spectroscopy, the main focus of this thesis is to develop and characterize amino acid-based spectroscopic probes as well as to use such probes to study important biological questions. Specifically, we show that (1) *p*-cyanophenylalanine and selenomethionine constitute an efficient fluorophore-quencher pair, useful for characterizing protein conformational changes that occur on a short distance; (2) 4-cyanotryptophan is a novel blue fluorescent amino acid, applicable for biological imaging due to its unique photophysical properties; (3) the dielectric constant inside the hydrophobic interior of staphylococcal nuclease is about 10–15, significantly larger than previously assumed; and (4) a single mutation in a short segment of the protein transthyretin (i.e., 110–115) induces formation of amyloid fibrils consisting of both

β- and α-sheets, where the latter is a proposed structure in proteins, but has never been observed previously.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1: Introduction

Proteins are involved in virtually every cellular process, displaying an almost endless diversity in function. This biological function of a protein is primarily defined by its structure, which is programmed in its amino acid sequence, which is determined by its genetic code. Astonishingly, there are only 20 natural amino acid building blocks, which are distinguished by their side chain moieties.[1] Amino acids are connected to one another through dehydration reactions between the carboxylic acids and amine groups on the respective neighboring amino acids. Through interactions between the amino acid side chains, stable conformations of the amino acids form, which can be defined using their φ (C-N-C_α-C) or ψ (N-C_α-C-N) dihedral angles (Figure 1.1). Thus, common secondary structures develop, including α-helices and β-sheets, which are defined by specific backbone φ and ψ angles.[1] The overall three-dimensional arrangement of all the atoms in the protein is known as its tertiary structure, which takes into account the interactions between the amino acid side chains of the protein.[1]

How a protein spontaneously folds from this amino acid sequence to a compact, organized, three-dimensional structure has been the focus of scientists for decades, since the first observation that a protein can reversibly fold outside the confinement of a cell.[2,3] However, according to Levinthal's paradox, if all degrees of freedom of a protein are considered with no energetic bias towards any structure, a protein would sample an innumerable amount of structures before reaching its final functional structure, which led to the idea of protein folding pathways where a protein is directed to its lowest energy conformation(s).[4] In the so-called protein funnel perspective, there are many high free energy states and only a few low energy states, with barriers, transitions, traps, and intermediates located in-between states.[5,6] It is important, however, to note that a

protein's folded structure is not static and, as such, undergoes random conformational fluctuations. Additionally, this folded structure can be induced to change through its interactions with other biomolecules. For example, the calcium-modulating protein calmodulin, which regulates a variety of cellular processes including inflammation, metabolism, and cellular response, undergoes a conformational change upon binding of the protein myosin light chain kinase which plays a role in the contraction of smooth muscle.[7] Therefore, the understanding of these conformational flexibilities or dynamics, which are often vital to the function of the protein, and their relationship to a protein's structure and function have been another focus of research.[8–11] Research has made great strides in understanding a protein's structure-function-dynamics relationship; however, many fundamental questions remain, as the free energy funnel of a protein, and hence the conformational dynamics, has numerous degrees of freedom which complicate the investigation of all the factors at play.

Adding to an already complicated problem, a protein can misfold both *in vivo* and *in vitro*. In the cell, molecular chaperones (which are proteins themselves) prevent misfolding and aggregation by reducing the probability of these two events, thereby increasing the efficiency of the folding process.[12–14] Moreover, chaperones can also force misfolded proteins to refold.[14] However, if the protein is misfolded or aggregates, the function of the protein is lost, which in many cases leads to disease.[14–16] Alzheimer's disease, Parkinson's disease, and type II diabetes are just some of the many diseases linked to the aggregation of proteins in human tissue.[14–16] These aggregated proteins, referred to as amyloid fibrils, have common characteristics, as they are made up of several (~2-6) protofilaments where the protofilaments either are twisted together (7-13 nm wide) or associate edgewise to form long ribbons (2-5 nm thick, ~30 nm wide).[16] X-

ray diffraction has shown that the fibrils have a characteristic cross-β pattern, indicative of a structure where the fibrils consist of β-sheets whose strands run perpendicular to the fibril axis.[14] High-resolution structures of several fibrils has been characterized by transmission electron microscopy (TEM), atomic force microscopy (AFM), X-ray diffraction, and solid-state nuclear magnetic resonance (NMR) spectroscopy.[16] However, structural heterogeneity in samples is common, as the formation of fibrils is affected by various experimental conditions, including temperature, salt concentration, and pH.[17–19] As such, there is not as much structural information on smaller aggregates, including oligomers and protofilaments. The investigation into the mechanism of fibril formation has been difficult, as the thermodynamics and kinetics between the different states vary dependent on the experimental conditions. However, amyloid fibril formation typically follows a nucleated growth mechanism, which means that the kinetics of fibril formation has a lag phase followed by a rapid exponential growth phase.[16]

However, the ultimate goal of many experimental protein studies is to create a molecular movie of protein dynamics, with all the necessary mechanistic details of all the structural units (i.e., side chains, secondary, tertiary structures, etc.). This goal is difficult for a variety of reasons. First, the protein has many degrees of freedom that makes it difficult to experimentally monitor each one. Second, these protein motions occur on a variety of time scales.[9] For example, fast motions like bond vibrations and hydrogen bond formation takes place on the femtoseconds to picoseconds timescale, while much slower motions like tertiary structure formation occur on the seconds timescale.[20,21] Finally, the various techniques used to investigate protein motions only have certain temporal and structural resolution; common techniques include X-ray diffraction, NMR

relaxation, UV-Vis, fluorescence, Raman, and infrared (IR) spectroscopies, and molecular dynamic simulations. For instance, while X-ray diffraction affords single atom resolution, only a single snapshot in time is obtained, though recent efforts have expanded the temporal resolution of solution phase x-ray scattering.[9,22,23] On the other hand, infrared spectroscopy can afford very high temporal resolution (fs to ps), but at the expense of structural resolution.

One way to gain structural information in various spectroscopic techniques is to use site-specific probes, or markers, that report on a single location in the protein. In the context of proteins, a good probe is one that (1) has a high extinction coefficient (IR) or brightness (fluorescence) so that a low amount of sample can be used in studies; (2) absorbs in a region devoid of intrinsic protein absorptions; (3) is small in size, such that it minimally perturbs the structure of the protein; (4) is incorporable into the protein system of interest (either chemically or biologically); and (5) is sensitive to a particular property of proteins, i.e., local electrostatic field or hydration.[24,25]

In biological microscopy, by far the most widely used imaging probe is the green fluorescent protein (GFP) and its derivatives. The discovery and realization that GFP can be used as a fluorophore in the cellular environment has revolutionized many fields in biology.[26–31] For example, it has found tremendous utility in the study of gene expression, protein-protein interactions, and macromolecular localization and trafficking.[26,28,31,32] The exceeding utility of GFP as an *in vivo* fluorescent reporter is due to several reasons. First, it has a large extinction coefficient (ε) (~30,000 M^{-1} cm^{-1} at 395 nm) and a relatively high quantum yield (0.79) with an emission maximum at 504 nm, making it a bright (defined as the product of the extinction coefficient and the quantum yield) fluorophore.[26] Second, the absorption and fluorescence properties of its intrinsic

chromophore, which is formed from residues Ser65-Tyr66-Gly67 in wild-type GFP, can be tuned through mutations, allowing the generation of various GFP derivatives, such as enhanced GFP (EGFP),[33] enhanced blue fluorescence protein (EBFP),[34] and yellow fluorescence protein (YFP),[35,36] which can have better photostability and different absorption and emission maxima. Combined, these GFP variants significantly expand the spectral range of fluorescent proteins, allowing for multiplex fluorescence detections as well as the development of various fluorescence resonance energy transfer (FRET) applications.[27] Third, and perhaps most importantly, GFP can be expressed or co-expressed in cells, including those of bacterial, yeast, and mammalian, allowing for use as a tag in gene expression or as an indicator of its environment.[26] Despite its tremendous utility, one limitation of using GFP as a fluorescence reporter is that the chromophore needs a relatively long time (2-4 hours *in vivo*) to form, imposing a time limit on the appearance of GFP fluorescence and hence the time resolution of any corresponding spectroscopic and imaging applications.[37] Another drawback is that the chromophore needs molecular oxygen to mature, making measurement in anaerobic environments difficult, if not impossible. Additionally, the molecular weight of GFP (27 kDa) and its cylindrical shape (4.2 nm in length with a 2.4 nm diameter)[38] make it a rather large fluorescent probe, which could hinder potential measurements by altering the localization, function, or fold of the protein to which GFP is attached.[39] Moreover, the fluorescence lifetimes of GFP and its derivatives are typically in the range of 1-3 ns,[29] making them less ideal to be used in time-gated fluorescence imaging applications.

Thus, the overarching goal of this thesis is to develop and use spectroscopic and microscopic probes to investigate protein structure, function, and dynamics. To that end, the first part of this thesis focuses on the development and characterization of new

fluorescent probes that can be used in biological studies while the second half focuses on infrared probes to investigate protein electrostatics and structure. In Chapter 2, the theory and the methods that are essential to the understanding of this thesis are explained, including the topics of circular dichroism, fluorescence, and infrared spectroscopies and their application to the study of protein structure, function, and dynamics.

In Chapter 3, we focus on the development of a new fluorescence quencher, the unnatural amino acid selenomethionine, that can be used in conjunction with the fluorophore *p*-cyanophenylalanine. First, we characterize the steady-state and time-resolved fluorescence properties of a dipeptide consisting of the fluorophore and quencher. Then, we use Stern-Volmer quenching and calculations to determine the distance-dependent quenching rate equation that governs the process. Next, we discuss the mechanism of the quenching. Finally, we apply this pair to polyproline peptides to observe the conformational heterogeneity in this structural motif.

In Chapter 4, we discuss the development and characterization of a new blue fluorescent amino acid, 4-cyanotrpytophan. We first compare the absorption and emission properties of this new fluorophore to other common fluorescence standards to determine the fluorescence quantum yield. Next, we discuss the kinetics of its fluorescence lifetime decay and its photostability. We then incorporate this fluorophore into an antimicrobial peptide and use wide-field and confocal fluorescence microscopy to view mammalian cell death.

In Chapter 5, we characterize the photophysical properties of indole and its nitrile-derivatized analogs to describe the effect of the electron withdrawing nitrile moiety on the indole fluorophore. We first collect the UV absorption and the steady-state

fluorescence emission spectra in water of all the cyanoindoles and discuss the differences in terms of intensity, width, and maxima. We then measured the fluorescence decay kinetics of all compounds in a variety of solvents (alcohols, water, organics) in order to investigate various fluorescence quenching mechanisms.

In Chapter 6, we investigate the hydration and electrostatic environment of the hydrophobic core of the staphylococcal nuclease protein using two infrared frequencies: (1) the nitrile stretching vibration of a thiocyanate (-SCN) moiety and (2) the carbonyl stretching vibration of an ester carbonyl ($-CH_2(C=O)OCH_3$, referred to as $-D_M$). Using the -SCN moiety, we find that a location in the nucleotide binding pocket of the protein is partially hydrated while two other locations in the main hydrophobic cavity are devoid of water. Similar experiments using D_M-labeled mutants agree with these conclusions. Additionally, using a combination of Onsager Theory and experiments, we estimated the dielectric constant of the protein interior to be 15.1, which, to our knowledge, is the first experimentally-determined dielectric constant in a protein pocket, which has profound impacts in the understanding and predictability of protein electrostatics.

In Chapter 7, we describe the effect of a side chain mutation on the secondary structure of transthyretin fibrils. Through linear and nonlinear infrared spectroscopic methods, we discuss the differences between the amide I' region of wild-type and a mutant where the $-D_M$ probe is incorporated. We find that the mutation induces a change in structure from one consisting of only β-sheets to one of mixed α- and β-sheet character. Atomic force microscopy, attenuated total reflectance IR spectroscopy, and circular dichroism spectroscopy techniques support our analysis.

Finally, in Chapter 8, we present a Summary and Future Directions to the work presented.

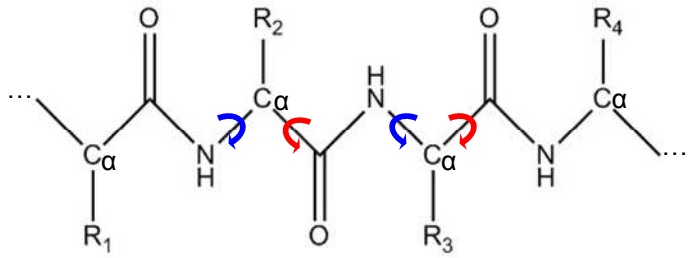

Figure 1.1: Schematic representation of the peptide backbone, where the blue and red arrows denote rotation about the φ and ψ dihedral angles, respectively and R_1, R_2, R_3, and R_4 represent various amino acid side chains.

2: Theory and Methods

2.1: CIRCULAR DICHROISM SPECTROSCOPY

Circular dichroism (CD) spectroscopy is an absorption spectroscopic technique that uses polarized light to investigate the structure of chiral molecules by measuring the difference between the absorption of left-circularly polarized light and that of right-circularly polarized light.[40] If the molecules are achiral or the light is not absorbed, there is no CD signal. UV CD spectroscopy is useful to investigate tertiary and secondary structure of peptides and proteins, as they are intrinsically chiral and have several chromophores that relate to its structure, including the peptide backbone (190-240 nm), aromatic amino acid side chains (~260-300 nm), and disulfide bonds (~260 nm).[40] Additional applications of CD spectroscopy include the investigation of protein-ligand interactions and folding intermediates and the thermodynamics and kinetics of protein folding.[41]

For each amide unit in the peptide backbone, there are two UV absorptions, the n→π* transition (transition 1) at 220 nm and the π→π* transition (transition 2) at 190 nm, that couple with one another, dependent on their individual distance and orientation relative to one another.[42,43] The extinction coefficient ($\Delta\varepsilon$) of this coupling is proportional to the interaction potential (V_{12}) and the rotational strength (R_{12}) between the respective electronic transition dipole moments (μ_1 and μ_2):[44,45]

$$\Delta\varepsilon \propto V_{12}R_{12} \tag{2.1.1}$$

where V_{12} is proportional to the dipole moments, the distance vector (r_{12}) and the angle (γ) between them, and the respective angles (α, β) between the dipole moments (μ_1, μ_2) and r_{12}:

$$V_{12} \propto \frac{\mu_1 \mu_2}{r_{12}^{\ 3}} (\cos\gamma - \cos\alpha \cos\beta) \tag{2.1.2}$$

and where R_{12} is proportional to:

$$R_{12} \propto \pm r_{12} \cdot (\mu_1 \times \mu_2) \tag{2.1.3}$$

Combining Eqs. (2.1.1) – (2.1.3) gives the following:

$$\Delta\varepsilon \propto \pm \frac{\mu_1^{\ 2} \mu_2^{\ 2}}{r_{12}^{\ 2}} \Lambda(\gamma, \alpha, \beta) \tag{2.1.4}$$

where $\Lambda(\gamma, \alpha, \beta)$ is a function of the spatial orientations of μ_1 and μ_2 and determines the sign of the coupling signal. After projection of μ_1 and μ_2 onto the same plane, a positive (negative) signal is generated from a clockwise (counterclockwise) rotation of the front dipole moment onto the one in the back.[44,45]

As such, each secondary structure has characteristic peaks in the far-UV. For example, α-helices have a positive peak located at 190 nm (52,000 cm^{-1}) and a negative doublet located at 208 (48,000 cm^{-1}) and 222 nm (52,000 cm^{-1}) while β-sheets have a

negative peak at 180 nm (55,000 cm^{-1}), a positive peak at 195 nm (51,000 cm^{-1}) and a negative peak at 215 nm (44,500 cm^{-1}). Using this information, the amount of each secondary structure in a given protein can be determined using a variety of available methods.[40] In addition, the coupling of aromatic residues can be used to investigate protein structure; for example, the aromatic residues tryptophan (Trp) and tyrosine (Tyr) have been used to investigate the stability of β-hairpin structures.[46–48] Additionally, nitrile-derivatized amino acids have also been used as extrinsic CD probes.[45]

2.2: FLUORESCENCE SPECTROSCOPY

Fluorescence spectroscopy is arguably the most common technique in biochemical and biophysical chemistry because it affords the following three advantages: (1) the fluorescence signal is extremely sensitive to the immediate environment (i.e. hydration, electrostatics, etc.) around the fluorophore; (2) low amounts of sample in the range of nano- to micromolar afford a high signal-to-noise ratio of the fluorescence signal; and (3) the fluorescence emission is on the nanosecond timescale, ideal to study conformational dynamics.[49] In this section, we will focus on the theory behind static and time-resolved fluorescence spectroscopy. Additionally, we will discuss the intrinsic and extrinsic fluorescence probes and their use in the investigation of protein structure, function, and dynamics.

2.2.1: Static Fluorescence Spectroscopy. The principle of fluorescence can be explained using a Jabłński diagram (Figure 2.1). When a molecule absorbs a photon of light, the molecule is promoted from its ground electronic state, S_0, to a vibrationally-excited state of its first electronic state, S_1. Then, in solution, the molecule typically

relaxes to its lowest vibrational level on S_1. From this level, several events can occur concurrently, including internal conversion, intersystem crossing, and photon emission. Through internal conversion, the molecule goes to an excited vibrational level in S_0 state that is at the same level as the ground vibrational level of S_1, where collisions with nearby molecules relax the molecule back to its ground vibrational level. Or, a molecule crosses from its singlet excited state to an excited vibrational level of its triplet state, T_1, where it would then relax to the lowest vibrational level. The signal of a molecule releasing a photon from its ground vibrational level of S_1 or of T_1 is known as fluorescence and phosphorescence, respectively. This thesis will focus on the fluorescence of molecules, though phosphorescence has been used to investigate the structure and dynamics of proteins.[50,51]

The fluorescence of a molecule typically has three characteristics: (1) as seen in Figure 2.1, the energy of emission is lower than that of absorption, known as the Stokes' shift,[52] which means that the fluorescence occurs at longer wavelengths than absorption; (2) the fluorescence spectrum is usually independent of the excitation wavelength, known as Kasha's rule;[53] and (3) the fluorescence spectrum is usually a mirror image of the absorption spectrum due to the Franck-Condon principle.[54] In the investigation of protein structure and dynamics, the most important property of fluorescence is the Stokes' shift, as part of this shift is sensitive to the environment surrounding the fluorophore. In other words, if the fluorophore changes in environment (i.e. from a hydrophobic to hydrophilic environment) as the protein folds, for example, the fluorescence spectrum of the fluorophore will change. Typically, the dipole moment of the fluorophore in the excited state (μ_e) is larger than that of the ground state (μ_g). After

the fluorophore is excited, a polar solvent will reorient its dipole moment to that μ_e, which causes a lowering of the energy of the ground vibrational level of S_1. Because the solvent relaxation rate is faster than the emission rate, the observed fluorescence is from a solvent-relaxed state. Therefore, as the strength of the solvent's dipole moment increases, the energy of this level decreases, which in turn lowers the fluorescence emission energy. To a first approximation, assuming that only solvent reorganization is responsible for the Stokes' shift, this difference between the absorption and emission energy can be approximately calculated using the Lippert Equation:[55]

$$\tilde{\nu}_A - \tilde{\nu}_F = \frac{2}{hc}\left(\frac{\varepsilon - 1}{2\varepsilon + 1} - \frac{n^2 - 1}{2n^2 + 1}\right)\frac{(\mu_e - \mu_g)^2}{a^3} + \text{constant} \qquad (2.2.1)$$

where $\tilde{\nu}_A$ and $\tilde{\nu}_F$ represent the absorption and emission wavenumbers, respectively, ε is the dielectric constant of the solvent, n is the index of refraction of the solvent, a^3 is the volume of the cavity in which the fluorophore resides, h is Planck's constant, and c is the speed of light. In the derivation of this equation, several assumptions are made. First, the fluorophore is treated as a spherical point dipole, which may not be the case depending on the shape of the fluorophore. Next, specific interactions with the solvent, including hydrogen bonding, acid-base chemistry, and charge-transfer interactions, are not considered. Additionally, the orientation of μ_e and μ_g is assumed to be the same. More complicated equations have been derived, which do not make these assumptions.[54] However, to a first approximation, Eq (2.2.1) is sufficient for many

systems. Deviations from this equation can indicate that additional interactions are present.[54]

In our setup, a Horiba Jobin Yvon Horiba Fluorolog 3.10 fluorometer is used to collect fluorescence emission spectra, using a 450 W xenon lamp as the excitation source. A specified emission wavelength is chosen using a monochromator. Emission over a range of wavelengths is collected using a monochromator and a photomultiplier tube (PMT) at a 90° angle to the excitation light, to prevent excitation light from hitting the detector. Additionally, time-based scans can be collected where the emission at a single wavelength is monitored as a function of time.

2.2.2: Time-Resolved Fluorescence. In addition to the static fluorescence spectrum of a fluorophore, the measurement of a fluorophore's fluorescence lifetime is useful in protein studies. To measure fluorescence lifetimes, τ, the fluorescence intensity decay, $I(t)$, is measured as a function of time and can be modeled by:

$$I(t) = \sum_{i=1}^{n} A_i e^{-\frac{t}{\tau_i}} \tag{2.2.2}$$

where A_i represents the amplitude of the fluorescence intensity of τ_i of the i^{th} component. In our lab, fluorescence lifetimes are determined in the time-domain, using a time correlated single photon counting (TCSPC) setup. Other methods for lifetime determination are beyond the scope of this thesis.[54] First, short pulses (i.e., much faster than the fluorescence lifetime on the order of 10-100 ps) of light are used to excite the

fluorophore repetitively. In our case, a home-built femtosecond Ti:Sapphire oscillator (800 nm, 850 MHz) is used to derive the excitation pulse. The arrival time of each laser pulse is monitored using a fast photodiode, which triggers a voltage ramp on the time-to-amplitude converter (TAC). When the first emitted photon from the sample fluorophore is detected by a fast PMT, this voltage ramp is triggered to stop. In this way, the difference in voltage is proportional to the amount of time the fluorophore stayed in the excited state from the excitation pulse. Repetitive excitation and emission detection builds up a histogram of counts over time. Once a significant amount of counts (>1,000,000) are collected, this photon arrival time distribution is representative of the fluorescence intensity decay of the fluorophore, $R(t)$. However, the time resolution of the instrument is limited to the response time in the electronics, PMT, and detection devices. An instrument response function (IRF) can be recorded to measure the shortest time window the setup can accomplish. For our instrument, the IRF is on the order of 50 ps, measured using a scattering sample with no excitation filters. Therefore, $R(t)$ is a convolution of the IRF and $I(t)$. Fitting of $R(t)$ revels the fluorescence lifetime(s) of the fluorophores. The number of lifetimes needed to fit the data is determined by minimization of the difference in the square of the residuals of the fit to the model (χ^2).

Non-single-exponential kinetics of a fluorophore can be caused by many things. First, if more than one fluorophore is present in the sample, the lifetimes could be indicative of each fluorophore, assuming they exist in differing environments. If only one fluorophore is present, multiexponential kinetics could indicate heterogeneity in the protein sample, meaning the protein could be sampling more than one conformation which results in a difference in lifetimes. These examples show the power of time-

resolved fluorescence measurements, as these situations would be difficult, if not impossible, to discern using conventional static fluorescence measurements.

2.2.3: Fluorescence Quantum Yield. To quantitatively compare the fluorescence intensity of one fluorophore to that of another, the fluorescence quantum yield (QY), defined as the number of emitted photons relative to the number of absorbed photons, is used. If we consider only radiative and non-radiative pathways of fluorescence emission, the following equations describe the absorption, non-radiative, and radiative processes occurring:

$$\begin{aligned} &\mathrm{F} + \mathrm{h}\nu \rightarrow \mathrm{F}^* \\ &\mathrm{F}^* \rightarrow \mathrm{F} + \mathrm{heat} \\ &\mathrm{F}^* \rightarrow \mathrm{F} + \mathrm{h}\nu \end{aligned} \qquad (2.2.3)$$

where F (F*) is the fluorophore in the ground (excited) state; each chemical reaction has their respective rate equation:

$$\begin{aligned} &\frac{d[F^*]}{dt} = k_a[F] \\ &-\frac{d[F^*]}{dt} = k_{nr}[F^*] \\ &-\frac{d[F^*]}{dt} = k_r[F^*] \end{aligned} \qquad (2.2.4)$$

where [X] denotes the concentration of X (X = F or F*) and k_Y denotes the rate constant of process Y, with Y = *a*, *nr*, or *r* denoting absorption, non-radiative, and radiative

processes, respectively. Under the steady-state approximation (i.e., under constant illumination), the rate of appearance of F* is equal to the rate of disappearance of F, giving:

$$k_A[F] = k_r[F^*] + k_{nr}[F^*] \tag{2.2.5}$$

Then, as the QY of the fluorophore is defined as:

$$QY = \frac{\text{emission rate}}{\text{absorption rate}} = \frac{k_r[F^*]}{k_a[F]} \tag{2.2.6}$$

Eq (2.2.5) can be substituted into Eq. (2.2.6), which yields:

$$QY = \frac{k_r}{k_r + k_{nr}} \tag{2.2.7}$$

where k_r is the radiative rate of decay of the excited S_1 and k_{nr} is the sum of all the non-radiative rates of relaxation to S_0. Considering that the measured fluorescence lifetime is defined as:

$$\tau = \frac{1}{k_r + k_{nr}} \tag{2.2.8}$$

we could define a natural fluorescence lifetime (denoted τ_0), which is the lifetime in the absence of any non-radiative relaxations as:

$$\tau_0 = \frac{1}{k_r} \tag{2.2.9}$$

Substituting Eq. (2.2.8) and (2.2.9) into Eq. (2.2.7) yields:

$$QY = \frac{\tau_0}{\tau} \tag{2.2.10}$$

Therefore, for a QY close to 1, there is little non-radiative relaxation as τ approaches τ_0.

The QY of an unknown fluorophore can be determined experimentally. One of the easiest ways is to quantitatively compare the steady-state fluorescence intensity of the unknown to that of a known reference under the same experimental conditions (i.e., temperature, excitation wavelength, emission wavelength, etc.), using the following equation:[54]

$$QY_S = QY_R \frac{I_S}{I_R} \frac{A_R}{A_S} \frac{{n_S}^2}{{n_R}^2} \tag{2.2.11}$$

where *I* is the integrated area of the fluorescence emission intensity, *A* is the optical density (OD) at the excitation wavelength, *n* is the index of refraction of the solvent, and S and R refer to the sample and reference, respectively. It is important that the chosen

reference absorb and emit in a similar region to that of the unknown sample. There are several cautions to note for the use of this equation. First, the OD of the sample and reference must be determined accurately, which means making sure the sample has a high OD compared to that of the background, typically in the range of 0.1-0.2. However, this OD may be too high to record an accurate fluorescence spectrum due to self-quenching of the molecules (See Section 2.2.4). Second, because the fluorescence intensity is dependent on the absorption of light, the lamp source must be stable over the time of the measurements. Ideally, the fluorescence spectra of the sample and reference would be recorded quickly, with a minimal time in between measurements. Therefore, to minimize these two concerns, the gradient method can be used, where a series of dilutions are used to measure the fluorescence intensity *I* as a function of absorbance *A*. This method is utilized in Chapter 4 to determine the QY of a new biological fluorophore.

2.2.4: Fluorescence Quenching. A process that decreases the fluorescence intensity of a sample is known as quenching. There are several quenching mechanisms in protein samples, which can be broadly classified as static or dynamic quenching. In static quenching processes, the quencher forms a nonfluoroescent complex with the fluorophore, before any excitation. Thus, after excitation, the total amount of fluorescence intensity is decreased, though the fluorescence lifetime of the fluorophore remains unchanged. On the other hand, dynamic quenching processes occur because of the diffusion of molecules, as the quencher encounters the excited fluorophore within the timescale of emission. In one type of dynamic quenching, called collisional quenching, the fluorescence intensity and the fluorescence lifetime of the sample are decreased as the excited state of the fluorophores is deactivated through collisions with

other solute molecules, including dissolved O_2, amines, dissolved salts in buffers, or even other fluorophores. To describe collisional quenching, the following equation can be used:

$$F^* + Q \rightarrow F + Q^* \tag{2.2.12}$$

where F and Q represent fluorophore and quencher, respectively, with the following rate equation:

$$-\frac{d[F^*]}{dt} = k_Q[F^*][Q] \tag{2.2.13}$$

where k_Q is the rate of quenching. Adding these equations to those in Eq. (2.2.3) and (2.2.4), under the steady-state approximation, gives:

$$k_A[F] = k_r[F^*] + k_{nr}[F^*] + k_Q[F^*][Q] \tag{2.2.14}$$

making the fluorescence QY in the presence of quencher (QY_Q):

$$QY_Q = \frac{k_r}{k_r + k_{nr} + k_Q[Q]} \tag{2.2.15}$$

Therefore, the ratio of QYs in the absence and presence of quencher is given as the so-called Stern-Volmer Equation:

$$\frac{QY_0}{QY_Q} = 1 + \frac{k_Q}{k_r + k_{nr}}[Q] = 1 + k_Q\tau_0[Q] \quad (2.2.16)$$

where τ_0 represents the fluorescence lifetime in the absence of quencher. Using this equation, the bimolecular rate constant of quenching k_Q can be determined by measuring the change of fluorescence signal as a function of concentration of quencher, assuming the fluorescence lifetime of the fluorophore is known. The Stern-Volmer equation (2.2.16) is used in Chapter 3 to explore the quenching mechanism of selenomethionine on *p*-cyanophenylalanine fluorescence.

2.2.5: Intrinsic Protein Fluorophores. Of the 20 natural amino acids, only three, tryptophan (Trp), tyrosine (Tyr), and phenylalanine (Phe), are fluorescent. All three absorb and fluoresce in the UV region, with an absorbance (emission) maximum centered at around 290 (350) nm for Trp, 275 (300) nm for Tyr, and 260 (280) nm for Phe (Table 2.1). However, Trp is almost exclusively used in biological studies of proteins, as the extinction coefficient and QY of Trp are larger than those of Phe and Tyr (Table 2.1). Additionally, the emission maximum, fluorescence intensity, and fluorescence lifetime of Trp are extremely sensitive to the environment,[49,56,57] with a peak around 350 nm in a hydrophilic environment compared to a peak around 335 nm in a hydrophobic environment. Coupled to the fact that the abundance of Trp in proteins is the lowest of any amino acid (1.4%),[1] Trp in proteins is frequently a site-specific probe of protein structure. The interpretation of Trp fluorescence kinetics can be difficult, however, as the decay does not follow single-exponential kinetics, due to different

rotamer conformations.[54] Moreover, Trp fluorescence can be quenched by the side chains of other amino acids[58–61] or statically quenched by oxygen and iodine or bromide anions.[62] As such, there are an innumerable amount of studies using Trp fluorescence as a probe of protein structure, function, and dynamics. To name only a recent few examples, Trp fluorescence has been used to investigate hard-to-study systems including membrane proteins[63] and amylogenic proteins,[64,65] to investigate ultrafast hydration dynamics between a protein and water,[66–68] and to investigate the end-to-end distances in peptides or proteins using electron transfer.[20] While there are far less studies using the fluorescence of Tyr and Phe, some examples include studying conformational changes of proteins[71] and measuring the ligand affinity to a protein.[72]

2.2.6: Extrinsic Protein Fluorophores. While there are a whole host of extrinsic fluorophores that can be attached or incorporated in protein samples, including organic dyes, fluorescent proteins, and quantum dots,[73] we will focus on the use of unnatural amino acids as protein fluorophores. A major advantage of unnatural amino acids compared to other extrinsic fluorophores is their structural similarity to amino acids, which is a smaller perturbation to the overall structure of proteins. Typical unnatural amino acid fluorophores are only different from a native amino acid by the addition of an atom or molecule. For example, cyanophenylalanines,[74] cyanotryptophans,[75,76] azatryptophans,[77,78] hydroxytryptophans,[79] and fluorotryptophans[80,81] have found use as protein fluorescent probes. As it applies to this thesis, we will focus on the nitrile-derivatized amino acids, particularly *p*-cyanophenylalanine (Phe_{CN}) and 5-, 6-, and 7-cyanotryptophan (hereafter referred to 5CN-Trp, 6CN-Trp, and 7CN-Trp, respectively).

The absorption and emission properties of Phe_{CN} are very different from those of Phe. First, the absorption maxima of Phe_{CN} are shifted to 233 nm (13,000 M^{-1} cm^{-1}) and 280 nm (850 M^{-1} cm^{-1}).[82] This allows Phe_{CN} to be selectively excited in the presence Phe, Trp, and Tyr, as none of them absorb at 233 nm. Additionally, the emission maximum is shifted to 295 nm in water, with a 5-fold increase in the QY compared to Phe,[82] meaning Phe_{CN} has a similar QY to that of Trp and Tyr. This emission maximum overlaps with the absorption maximum of Trp, making Phe_{CN}-Trp a good Förster resonance energy transfer (FRET) pair, where this overlap is required.[82] The fluorescence decay kinetics of Phe_{CN} follow single-exponential kinetics with a lifetime between ~4-8 ns depending on the solvent,[83] which is much longer than the average lifetime of Trp (~3 ns).[54] Thus, Phe_{CN} has been utilized in a variety of biological studies, including protein folding kinetics,[34] amyloid formation kinetics,[35] protein-peptide interactions,[82,86] and protein structure.[87,88] Recently, the photophysical properties of 2- and 3-cyanophenylalanine have also been explored.[89] In Chapter 3, we discuss an unnatural amino acid that quenches the fluorescence of Phe_{CN}.

Similarly, 5CN-, 6CN-, and 7CN-Trp have different photophysical properties over those of Trp, which can be advantageous in certain applications. While the absorption and emission spectra of 5CN-Trp are similar to those of Trp, the QY and fluorescence lifetime of 5CN-Trp are very sensitive to hydration.[76] For example, the integrated fluorescence intensity of 5CN-Trp in 1,4-dioxane is increased by a factor of 23 compared to that of 5CN-Trp in water. In the same solvents, the fluorescence intensity of Trp only increases by a factor of 1.3. Similarly, the average lifetime of 5CN-Trp decreases from 13.9 ns in dimethyl sulfoxide to 0.8 ns in water. This makes 5CN-Trp especially useful in applications where the fluorophore experiences a change in hydration in the folding of a

protein, for example. 6CN- and 7CN-Trp have red-shifted absorption and emission maxima compared to those of Trp.[75] An absorption maximum at 290 nm (10,300 M^{-1} cm^{-1}) and an emission maximum at 370 nm (QY = 0.53), all in methanol, were observed for 6CN-Trp while an absorption maximum at 310 nm (8,100 M^{-1} cm^{-1}) and an emission maximum at 390 nm (QY = 0.40) were observed for 7CN-Trp. With higher extinction coefficients and QY, these fluorophores are brighter than Trp. As such, they were used to investigate protein conformation fluctuations and DNA-protein interactions.[75] In Chapter 4, we discuss another cyanotryptophan that has useful absorption and emission properties for use in biological imaging applications. Additionally, in Chapter 5, we discuss the effect of nitrile substitution on the photophysical properties of indole.

2.3: INFRARED SPECTROSCOPY

Infrared (IR) spectroscopy is a popular biophysical and biochemical technique, as it is extremely sensitive to the composition and structure of biomolecules, including peptides and proteins, as the absorption of infrared light excites vibrational transitions in molecules.[90] While there are a large number of vibrational degrees of freedom in a biomolecule which can complicate its IR spectrum and the amount of sample needed for a measurement is more than that needed for fluorescence, IR spectroscopy has a number of advantages including high applicability to various molecules and short measuring time. In this thesis, we will focus on two techniques, Fourier transform infrared spectroscopy (FTIR) and attenuated total reflection infrared (ATR-FTIR) spectroscopy. We will then discuss common infrared probes used in the study of protein structure, function, and dynamics, including native probes such as protein backbone and

side chain vibrations, and nonnative probes such as side chain ester carbonyls and nitriles.

2.3.1: Fourier Transform Infrared Spectroscopy. In FTIR Spectroscopy, infrared light from a polychromatic source, typically a globar (1000-4000 cm^{-1}), passes through an interferometer, which produces an interferogram of infrared light that passes through the sample. The amount of light that is transmitted is detected, in our case by a liquid nitrogen-cooled MCT detector. Then, a Fourier transform converts this interferogram into a transmittance spectrum, $T(\lambda)$. The absorbance, $A(\lambda)$ of the sample is then:

$$A(\lambda) = \log\left(\frac{T_S(\lambda)}{T_R(\lambda)}\right) \tag{2.3.1}$$

where the subscripts S and R denote the sample and reference, respectively. In our case, the sample contains our peptide/protein in a buffer and the reference is the just the buffer without any peptide/protein. In this way, the entire infrared spectrum of a sample is collected at one time, instead of in a dispersive technique where each frequency is recorded separately. Compared to these measurements, FTIR spectroscopy has an improved signal-to-noise ratio, better frequency accuracy, and faster speed.[91]

2.3.2: Attenuated Total Reflection Infrared Spectroscopy. ATR-FTIR spectroscopy is a powerful technique that can determine structural information of an oriented sample, which is useful in many biophysical applications. A typical setup is shown in Figure 2.2.[92] First, a sample is deposited onto the surface of an ATR crystal, which is a trapezoidal

plate made of a material with a high index of refraction (i.e., germanium, zinc selenide, and gold). Infrared light is directed through the crystal. Dependent on the index of refraction of this crystal, there exists a critical angle at which the infrared light is totally internally reflected through the crystal until it reaches the end. These reflections produce a field outside the crystal, called an evanescent wave, in the direction normal to the crystal surface. This evanescent wave probes the sample and is detected after passing through the end of the crystal. The time-averaged electric field intensity, E, of the evanescent wave is described as a function of the distance from the surface:

$$E = E_0 e^{-\frac{z}{d_p}} \tag{2.3.1}$$

where E_0 is the time-averaged electric field intensity at the interface, z is the distance from the interface, and d_p is the penetration depth of the wave, which is given by:

$$d_p = \frac{\left(\frac{\lambda}{n_1}\right)}{2\pi\left[\sin^2\theta - \left(\frac{n_2}{n_1}\right)^2\right]} \tag{2.3.2}$$

where λ is the wavelength of light, n_1 and n_2 are the index of refraction of the ATR crystal and the sample, respectively, and θ is the incident angle of the light, i.e., the angle between the normal of the crystal surface and the infrared light. In our setup, where a germanium (Ge) ATR crystal (θ = 45°, n_1 = 4.0) is used, the d_p is on the order of

hundreds of nanometers. Therefore, ATF-FTIR spectroscopy is a surface technique, only measuring sample that is less than 1 μm away from the ATR crystal surface. Compared to FTIR where transmission, solution phase spectra of biomolecules are typically collected, ATR-FTIR spectra are collected on dry or hydrated films that are deposited onto the crystal.

To measure orientation of a sample, linearly-polarized light is used, as depicted in Figure 2.2. Absorbance spectra in both the parallel and perpendicular directions (with respect to the crystal surface) can be collected, as the absorbance of the sample is dependent on the magnitude of the *x* and *z* (parallel) and *y* (perpendicular) components of the transition dipole moment of the sample. To quantitatively describe the difference between the absorption of parallel and perpendicular light, the dichroic ratio (DR) is used, which is simply the ratio of the absorbance of parallel and perpendicular light (DR = $Abs_{\parallel}/Abs_{\perp}$). For an isotropic sample, the DR is always 2, as the absorption in the parallel direction (i.e., the Cartesian directions *x* and *z*) is twice that of the absorption in the perpendicular (i.e., the Cartesian direction y). However, if a protein sample is oriented on the surface, for example, a membrane protein supported in a bilayer membrane, the orientation of the protein dipole moment can be determined.[92] Because of its structural resolution, ATR-FTIR spectroscopy is a popular technique to study systems that can be difficult to study with other methods (i.e., CD, NMR, X-ray crystallography) including membrane proteins and amyloid fibrils and aggregates.[92–98]

2.3.3: Native Protein IR Probes. *Protein Backbone.* Protein FTIR spectra have a variety of bands, arising from both the side chains of amino acids and the protein backbone. That of the protein backbone, i.e., the peptide bond consisting of the –(C=O)-NH-

moiety, is a useful probe of protein secondary structure. The normal modes of this group include the amide A (~3300 cm^{-1}) and amide B (~3170 cm^{-1}), localized frequencies arising from the NH stretch; amide I (1600-1700 cm^{-1}), which arises mainly from the C=O stretch with minor contributions of an out-of-phase CN stretch, a CCN deformation, and a NH in-plane bend; amide II (~1550 cm^{-1}), which arises from the out-of-phase combination of the NH in-plane bend and the CN stretch; and amide III (1200-1400 cm^{-1}), which arises from the in-phase combination of the NH bend and the CN stretch,[99] among others.[100] Of these, the amide I is the most widely used, as it is sensitive to the secondary structure of proteins due to transition dipole coupling (TDC), which is an interaction between neighboring dipoles, whose coupling depends on the orientation of and the distance between the dipoles. TDC has two effects: (1) delocalization of the excited state due to energy transfer between oscillators, similar to fluorescence energy transfer and (2) splitting of the IR band due to in- or out-of-phase oscillations between the dipoles. As a result, the common secondary structures of proteins, including α-helices (~1650 cm^{-1}) and parallel and anti-parallel β-sheets (~1630 and 1680 cm^{-1}), have different characteristic IR frequencies[99] because of their differences in the orientations of their dipole moments.

Because of the strong background absorbance (extinction coefficient of 21.8 M^{-1} cm^{-1} at 1643.5 cm^{-1}) of H_2O in the amide I region of the IR spectrum,[101] D_2O is typically used as a solvent to look at the amide I region of proteins, as the deuterium shifts the bending frequency. When measured in D_2O, the amide I band is referred to as the amide I'. However, this necessitates that all exchangeable protons on the protein be exchange to deuterium before measurements, which is accomplished through repeat lyophilization

cycles where the protein is dissolved in a high concentration (~0.1 M) of a strong acid (e.g., DCl) in D_2O. Additionally, the maximum pathlength that one can use for measurements is dictated by the absorbance of solvent. For H_2O (D_2O), this maximum pathlength for the amide I (I') region is 25 (50) μm.

Because the amide I (I') peaks arise from coupling between backbone units and is delocalized across the units, there is no site-specificity to a single amide unit. However, isotope-labeling a single backbone unit to ^{13}C (or $^{13}C=^{18}O$) red-shifts the infrared frequency by 40 (~80) cm^{-1}, which separates this frequency from the main amide I (I') band.[102] In this way, site-specificity at any desired location in the protein is achieved, assuming the desired backbone unit can be sufficiently labeled.

Recent work using the protein backbone to elucidate structure has occurred in the amyloid fibril field. Isotope-labeling of the protein backbone in β-sheet amyloid fibrils, in combination with time-resolved two dimensional infrared spectroscopic techniques, has been used to distinguish whether a sheet is anti-parallel or parallel,[103,104] to investigate aggregation pathways,[105] and to observe dynamics of water molecules in fibrils.[21] Additionally, isotope-labeling has made great strides in determination of the structure of membrane proteins,[106–108] in the investigation of the mechanisms of ion channels,[109] in the measurement of protein folding kinetics,[110] and in probing protein-protein interactions,[111] to name just a few recent examples.

Protein Side Chains. Many side chains of amino acids are IR active and have a strong extinction coefficient, including carbonyl-containing amino acids like asparagine (Asn), aspartic acid (Asp), glutamine (Gln), and glutamic acid (Glu), ring-containing amino acids like Tyr and Trp, and thiol-containing amino acids like cysteine (Cys).[90] Because of IR spectra congestion, i.e. overlapping of different IR signals, the most

straightforward side chain vibrations to observe in experiments are those arising from the carbonyl of protonated carboxylic acids (i.e., those in Asp and Glu), which locate between 1710-1790 cm^{-1} in ^1H_2O and ^2H_2O and the thiol group of Cys, which locates between 2550-2600 cm^{-1} in the same solvents.[99] Because the carbonyl frequencies of Asp and Glu are sensitive to the protonation of the carboxylic acid, these amino acids have been used to investigate the proton pump bacteriorhodopsin.[112] Additionally, the carbonyl frequency is sensitive to hydrogen bonding and cation chelation. Thus, Asp and Glu have been used to study several calcium-binding protiens.[112] However, more widespread use of using native side chain probes is limited by spectral congestion and lack of site-specificity. In other words, because all the amino acids absorb the infrared light, it is not possible to site-specifically use one Asp as a local environmental probe when there are others in the protein, for example.

2.3.4: Nonnative Protein IR Probes. As site-specificity is hard to achieve using native side chain probes, one way in which to accomplish this goal is to incorporate a nonnative IR probe into the protein. In this section, we will focus on the three such nonnative side chain IR probes: the ester carbonyl (-(C=O)OCH_3), the nitrile (-C≡N), and the thiocyanate (-SC≡N) moieties.

Ester Carbonyl Side Chain. By methylation of the -OH moiety of the carboxylic acid groups on Asp and Glu, side chains containing ester carbonyls are produced, namely aspartic acid 4-methyl ester (hereafter referred to as -D_M) and glutamic acid 5-methyl ester (hereafter referred to as -E_M). These amino acids can be incorporated via solid phase peptide synthesis. Alternatively, the -CH_2COOCH_3 (i.e., the side chain of D_M) can be appended onto a Cys residue of a protein post-translationally.[113] Through both

methods of incorporation, site-specificity is achieved, (assuming there is only one reactive Cys in the protein). Additionally, the ester moiety shifts the frequency of the carbonyl to 1700-1800 cm^{-1}, a region that is safely outside that of the amide I (I') band and is devoid of other vibrations at neutral pH and has a large extinction coefficient in the range of 300-650 M^{-1} cm^{-1}.[25,114] The strength of the -D_M (or -E_M) probe is its sensitivity to local electrostatics and to hydration,[114] as the frequency of the carbonyl was found to vary linearly with the local electric field. In nonpolar solvents, as the dielectric constant of the solvent decreases, the frequency of -D_M and -E_M blue-shifts (e.g., the frequency in hexane (ε = 2.0) is 1756 cm^{-1} whereas the frequency in dimethyl sulfoxide (ε = 46.7) is 1738 cm^{-1} – a shift of 18 cm^{-1}). Additionally, the number of hydrogen bonds to the ester carbonyl moiety can be differentiated. For example, in methanol, three peaks are observed, corresponding to 0, 1, and 2 hydrogen bonds, with a red-shift in frequency as the number of hydrogen bonds increases. As such, the presence of water molecules in between the fibril stands was confirmed using the -D_M probe incorporated into aβ amyloid fibrils.[114]

Nitrile and Thiocyanate Side chains. Nitrile and thiocyanate side chains are useful IR probes in proteins, as the C≡N stretching frequency locates between 2050-2400 cm^{-1}, dependent on the R group to which it is attached, which is a region devoid of natural protein vibrations.[25] As nitriles connected to alkyl chains have a lower absorption cross-section (e.g., connected to an alanine side chain (-CH_2CN)) compared to that of nitriles connected to an aromatic side chain (e.g., Phe and Trp),[24] nitrile-derivatized aromatic amino acids are used frequently to study protein electrostatics over their alkyl counterparts, which can be incorporated into proteins using a variety of methods.[115,116] Additionally, as the nitrile moiety can be added post-translationally to a protein via Cys

cyanylation,[117] the thiocyanate IR probe is also a popular IR probe in protein investigations (we will referred to this thiocyanate probe hereafter as Cys_{CN}). This wide use is due to many factors. Because of the small size of the nitrile moiety, little to no perturbation of the structure is produced, especially when Phe_{CN} and 5CN-Trp are substituted for Phe, Trp, or Tyr in the native protein sequence and Cys_{CN} is used in place of Cys, leucine, isoleucine, and methionine. Additionally, the excitation coefficients of 5CN-Trp (160 M^{-1} cm^{-1}), Phe_{CN} (220 M^{-1} cm^{-1}), and Cys_{CN} (120 M^{-1} cm^{-1}) are relatively large.[25] Finally, the CN stretching frequency is sensitive to various specific (i.e., hydrogen bonding) and nonspecific interactions.[118] For example, upon going from a hydrophobic to an aqueous environment, the nitrile band becomes much broader in width and shifts its frequency a higher wavenumber. As such, Phe_{CN}, 5CN-Trp, and Cys_{CN} have been used widely in the investigation of protein structure, function, and dynamics.[25,118–127]

Table 2.1: Absorption and Emission Properties of Trp, Tyr, and Phe in Water

Amino Acid	Absorption Max (nm)	Extinction Coefficient (M^{-1} cm^{-1})	Emission Max (nm)	QY
Trp	290	5500	353	0.14
Tyr	275	1500	304	0.13
Phe	260	200	282	0.02

All Values obtained from Lakowicz.[54]

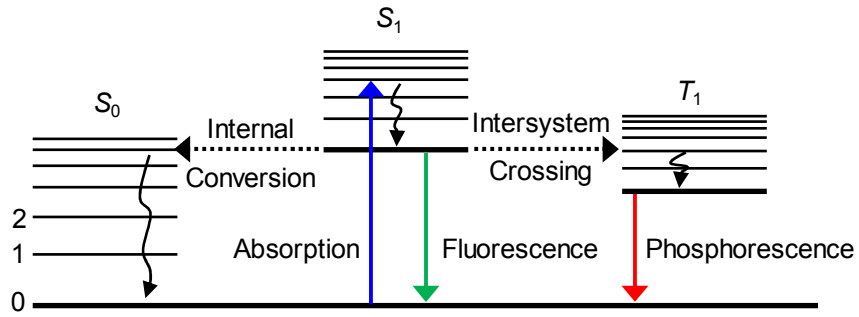

Figure 2.1: A form of a Jabłński Diagram, where S_x and T_x represent the x^{th} singlet and triplet electronic state and 0, 1, 2 … represent the vibrational levels in each electronic state. Internal conversion and intersystem crossing are represented by dotted lines while absorption (blue) of light and emission (fluorescence (green) and phosphorescence (red)) of light are represented by solid color lines, as indicated. Non-radiative relaxation is represented by curved, solid black lines.

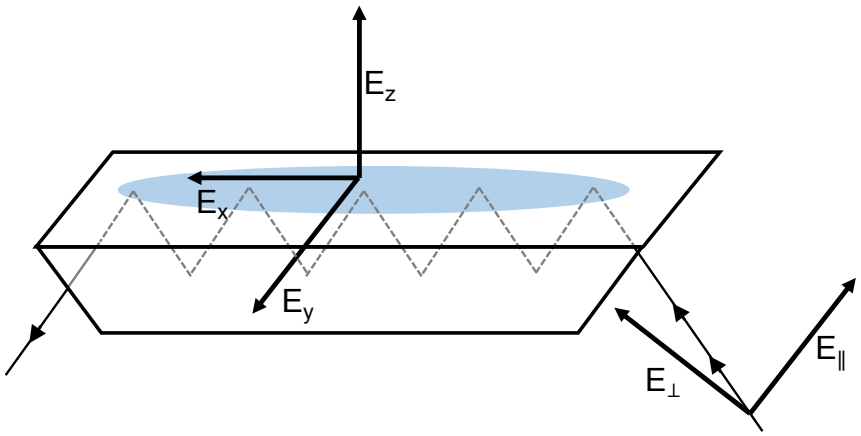

Figure 2.2: A schematic representation of the ATR-FTIR setup showing the infrared light pathway through the ATR crystal. The blue circle represents a sample on top of the crystal. $E_\parallel$ and $E_\perp$ represent parallel and perpendicularly polarized light, respectively, relative to the surface of the sample. E_x, E_y, and E_z represent the Cartesian coordinates of the sample.

3: *p*-Cyanophenylalanine and Selenomethionine Constitute a Useful Fluorophore-Quencher Pair for Short Distances Measurements: Application to Polyproline Peptides

3.1 ABSTRACT

The C≡N stretching frequency and fluorescence quantum yield of *p*-cyanophenylalanine are sensitive to environment. As such, this unnatural amino acid has found broad applications, ranging from studying how proteins fold to determining the local electric field of membranes. Herein, we demonstrate that the fluorescence of *p*-cyanophenylalanine can be quenched by selenomethionine through an electron transfer process occurring at short distances, thus further expanding its spectroscopic utility. Using this fluorophore-quencher pair, we are able to show that short polyproline peptides (1-4 prolines) are not rigid; instead, they sample a bimodal conformational distribution.

3.2 INTRODUCTION

The unnatural amino acid *p*-cyanophenylalanine (Phe_{CN}) is a versatile spectroscopic probe in proteins because (1) its C≡N stretching vibration is sensitive to the local environment and thus can serve as a site-specific infrared (IR) probe,[24,122,128,129] (2) its fluorescence quantum yield and lifetime in protic solvents, especially water, are significantly different than those in aprotic solvents,[83] (3) its fluorescence can be quenched by various metal ions,[130] (4) it can be used as an efficient fluorescence resonance energy transfer (FRET) donor to tryptophan (Trp) and other amino acid fluorophores,[82,131–135] (5) its fluorescence can be selectively excited by using an excitation wavelength of 240 nm, even in the presence of other fluorescent amino acids, i.e., phenylalanine (Phe), tyrosine (Tyr) and Trp,[74] (6) it can be easily incorporated into

biological molecules using various methods including amber codon suppression,[115] and (7) it has a similar size to Phe and Tyr, making substitution of these amino acids with Phe_{CN} minimally perturbing to the target protein's native structure. As such, Phe_{CN} has been utilized in many biological studies to investigate, for example, the thermodynamics and kinetics of protein folding,[84] protein conformation and conformational distribution,[136,137] protein-peptide interactions,[74,86] the kinetics and mechanism of amyloid formation,[85] and protein and membrane electrostatics.[138,139] To further extend the spectroscopic utility of Phe_{CN}, herein, we show that selenomethionine (SeMet), a naturally-occurring and non-fluorescent amino acid, is an efficient quencher of Phe_{CN} fluorescence and, thus, this fluorophore-quencher pair is ideally suited to probe protein conformational changes occurring over very short distances (<20 Å).

Quenching has been widely used to study the structure and dynamics of proteins. For example, the quenching of Trp fluorescence in proteins by amino acids, such as cysteine (Cys), methionine (Met), or by small molecules, such as acrylamide, has been used to determine the rate of contact formations in disordered peptides,[140] peptide binding to proteins,[59] and the degree of Trp residue solvent exposure.[141] Additionally, Trp has been used as a quencher of oxazine and rhodamine fluorophores through photoinduced electron transfer (PET).[142] FRET, depending on the donor-acceptor pair, typically occurs over a distance range of 20-100 Å.[54] Fluorescence quenching, on the other hand, requires overlap of the electron orbitals of the fluorophore and quencher, and thus occurs at a shorter length scale.[54] Therefore, for certain applications, fluorescence quenching is more desirable than FRET. Additionally, fluorescence quenching experiments are often simpler in practice, in comparison to

FRET measurements, since the quencher does not directly contribute to the fluorescence signal.

We hypothesize that Phe_{CN}-SeMet constitutes an efficient fluorophore-quencher pair based on a previous study which suggested that SeMet is capable of quenching the fluorescence of Trp in a protein environment.[59] To verify this hypothesis, we conducted both steady-state and time-resolved fluorescence measurements on a SeMet-Phe_{CN} dipeptide. Indeed, our results show that SeMet can efficiently quench Phe_{CN} fluorescence. Using the Stern-Volmer analysis and numerical calculations, we are able to further determine the underlying electron-transfer parameters that dictate the distance-dependent fluorescence quenching rate. Furthermore, we employed this fluorophore-quencher pair to investigate the conformational distribution of a series of short polyproline (Pro) peptides, i.e., SeMet-$(Pro)_n$-Phe_{CN} (n = 1-4), and found that two conformations are populated in each case. Taken together, these results lead us to believe that Phe_{CN}-SeMet is a useful fluorophore-quencher pair for protein structural and conformational studies. Additionally, using SeMet as a quencher in this regard affords the following advantages: (1) it can be synthetically incorporated into a peptide through solid phase peptide synthesis, or into a protein, through expression vectors[143] or cell-free expression,[144] and (2) it is a natural derivative of Met and, thus, a Met to SeMet substitution is expected to cause only minimum structural perturbation to the protein of interest.

3.3 MATERIALS AND METHODS

3.3.1 Sample Preparation. All peptides were prepared by standard 9-fluorenylmethoxy-carbonyl (Fmoc) solid phase peptide synthesis methods on a PS3 automated peptide

synthesizer (Protein Technologies, MA) and purified by reverse-phase high-performance liquid chromatography (HPLC). For each peptide, the mass and its identity were characterized using either liquid-chromatography mass spectrometry (LC-MS) or matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS). All peptide samples used in the fluorescence measurements were prepared by dissolving lyophilized peptide in Millipore water, and the pH of the solutions was around 5.3. The peptide concentration, determined optically using the absorbance of Phe_{CN} at 280 nm (ε = 850 M^{-1} cm^{-1}),[82] was between 10-25 μM for static measurements and 200 μM for the time-resolved experiments. For the Stern-Volmer measurements, the concentration of Phe_{CN} was kept constant at 165 μM.

3.3.2 UV-Vis Spectroscopy. UV-Vis spectra were collected on a Perkin Elmer Lambda 25 UV/Vis spectrometer using a 1 cm quartz cuvette. To prepare the free SeMet amino acid sample, an appropriate amount of SeMet was dissolved in Millipore water for a concentration of 100 μM.

3.3.3 Cyclic Voltammetry. Cyclic voltammetry measurements were collected on a CHI630E workstation (CH Instruments, Inc., Austin, TX). The glassy carbon working electrode, the Pt wire counter electrode, and the calomel reference electrode were contained in a Faraday cage. The free Phe_{CN} amino acid (500 μM) was dissolved in degassed phosphate buffer (pH 7.0) containing 0.2 M KCl. Measurements were conducted under argon to minimize O_2 reactions.

3.3.4 Static and Time-Resolved Fluorescence Measurements. All static fluorescence measurements, except the Stern-Volmer titrations, were collected on a Jobin Yvon Horiba Fluorolog 3.10 fluorometer at 25 °C in a 1 cm quartz cuvette with spectral resolution of 1.0 nm, an integration time of 2.0 s/nm, and an excitation wavelength of 270 nm. The Stern-Volmer experiments were conducted on an Agilent Technologies Cary Eclipse fluorometer under the same conditions. Time-resolved fluorescence measurements were collected on a time-correlated single photon counting (TCSPC) system with a 1 cm cuvette at 25 °C. The details of the TCSPC system have been described elsewhere.[145] Briefly, a home-built femtosecond Ti:Sapphire oscillator operating at 85 MHz was used to generate the 270 nm excitation pulse. Selective emission at 340 nm was monitored using a subtractive double monochromator with a MCP-PMT detector (Hamamatsu R2809U) and a TCSPC board (Becker and Hickl SPC-730). Fluorescence decays were deconvoluted with the experimental instrument response function (IRF) and were fit either to a single-exponential decay (Gly-Phe_{CN}-Gly) or a bi-exponential decay (SeMet-containing peptides) using FLUOFIT (Picoquant GmbH). For comparison, fluorescence decays were also fit using the maximum entropy method (MEM).[146] The optical density of the samples at the excitation wavelength of 270 nm was approximately 0.01 for the static measurements and 0.1 for the Stern-Volmer and time-resolved measurements.

3.3.5 Nanosecond Transient Absorption Spectroscopy. Excitation pulses of 266 nm (~1 mJ/pulse) were generated from the 4^{th} harmonic of a 10 Hz NdYAG laser (~10 ns pulse width) and slightly focused on the sample with a cylindrical UV lens (f = 100 mm). Details of the experimental setup are reported elsewhere.[147] The samples were prepared

by dissolving the desired amino acid in Millipore water. The samples were then degassed with argon for at least 30 minutes. The optical densities at 266 nm of the Phe_{CN} sample (165 μM), the Tyr sample (180 μM), and the Phe_{CN} (165 μM) and SeMet (8 mM) mixture were ~0.2, ~0.2, and ~0.3, respectively.

3.3.6 Molecular Dynamics (MD) Simulation. The starting geometry of Met-Phe_{CN} was produced using a *trans* peptide bond. MD simulations were carried out using the NAnoscale Molecular Dynamics (NAMD) program (version 2.7)[148] and the CHARM22 force field.[149] The force field parameters for Phe_{CN} were derived by Zoete and coworkers.[150] The peptide of interest was immersed in 5857 TIP3P water molecules[151] in a 40 Å x 40 Å x 40 Å cubic box with periodic boundary conditions and a cutoff of 12 Å for nonbonded interactions. After energy minimization of the entire system, a production run of 100 ns was performed with a pressure of 1 atm, maintained using the Nosé-Hoover Langevin piston method. The temperature was increased from 0 to 298 K with an increment of 20 K every 500 timesteps. The final temperature of 298 K was held constant and was controlled by the Langevin thermostat. Full electrostatics were evaluated at every time step via the Particle-Mesh Ewald (PME) method.[152] A 2 fs time step was used to integrate the equations of motion and the structural coordinates were saved every 1 ps for further analysis. The SHAKE algorithm was used to restrict the motion of all bonds involving hydrogen. Convergence of the simulation was checked by monitoring the distance between the sulfur atom in Met and the center-of-mass (COM) of the Phe_{CN} sidechain in 10 ns intervals. This distance distribution was then fit to the sum of two Gaussians. The simulation was run until the centers, amplitudes, and widths of

the two Gaussians were constant over a period of 50 ns. For this reason, only the frames from the last 50 ns of the full 100 ns trajectory were used for further analysis.

3.4 RESULTS AND DISCUSSION

3.4.1 Quenching PheCN Fluorescence by SeMet. In order to verify whether SeMet quenches Phe_{CN} fluorescence, we first compared the apparent fluorescence quantum yields of two peptides, SeMet-Phe_{CN} and Gly-Phe_{CN}-Gly. As shown (Figure 3.1), under the same experimental conditions, the Phe_{CN} fluorescence intensity of SeMet-Phe_{CN} is ten times less than that of Gly-Phe_{CN}-Gly, indicating that SeMet, when in close proximity, can efficiently quench the fluorescence of Phe_{CN}. Additionally, while it is known that the natural amino acid Met quenches Phe_{CN} fluorescence,[135] SeMet is a more effective quencher of Phe_{CN} fluorescence, which decreases the fluorescence by a factor of three when compared to a Met-Phe_{CN} peptide (Figure 3.1). To further quantify the efficiency of the SeMet quenching process, we carried out time-resolved fluorescence measurements. As shown (Figure 3.2), the fluorescence decay of Gly-Phe_{CN}-Gly fits well (i.e., $\chi^2 < 1.2$) to a single-exponential, with a lifetime of 7.5 ns (Table 3.1), in good agreement with previously reported fluorescence lifetime of Phe_{CN}.[83] Further assessment of this decay using the MEM yielded consistent results, which show a single lifetime distribution centered at 7.1 ns (inset of Figure 3.2). In comparison, the fluorescence decay of SeMet-Phe_{CN} is significantly faster than that of Gly-Phe_{CN}-Gly (Figure 3.2), further confirming the role of SeMet as a Phe_{CN} fluorescence quencher. Surprisingly, however, the fluorescence decay of SeMet-Phe_{CN} is best described by a bi-exponential function with the following lifetime (relative amplitude): 2.0 ns (47%) and 0.2 ns (53%). This deviation from single-exponential behavior, which is further verified by the MEM

analysis (inset of Figure 3.2), is indicative of two peptide conformations that have distinctively different separation distances between the sidechains of SeMet and Phe_{CN}.

3.4.2 Stern-Volmer Experiments. To determine the underlying quenching mechanism, we carried out further static fluorescence quenching experiments using free SeMet and Phe_{CN}. As shown, (Figure 3.3), the resultant Stern-Volmer plot, determined after correction of the inner filter effect of SeMet, exhibits an upward curvature, indicative of a distance-dependent quenching rate:[54]

$$k_Q(r) = k_0 \exp(-\beta(r - a_0)) \tag{3.1}$$

where a_0 is the donor-to-quencher separation distance when they are in van der Waals contact, k_0 is the corresponding quenching rate (i.e., when $r = a_0$), and β is a constant characteristic of the fluorophore-quencher pair.[153] The minimum value of the parameter a_0 was determined to be 6.23 Å from the sum of the van der Waals radii of the quencher, Se (1.90 Å), and the fluorophore, Phe_{CN} (4.33 Å). However, due to the solvent-cage and steric effects, the value of a_0 is likely larger. Thus, following previous practice,[154] we set the value of a_0 to be 7.0 Å. To solve for k_0 and β, we numerically fit the Stern-Volmer curve based on the methods described in the literature.[54,154–158] Briefly, the fluorescence decay signal, $I(t)$, is described as:

$$I(t) = I_0 \exp\left(-\frac{t}{\tau_0} - C_Q^0 \int_0^t k(t')dt' \right) \tag{3.2}$$

where τ_0 is the intrinsic fluorescence lifetime of the fluorophore, C_Q^0 is the bulk concentration of the quencher, and time-dependent quenching rate, $k(t)$, is calculated based on the following equation:

$$k(t) = \frac{4\pi}{C_Q^0}\int_{a_0}^{\infty} r^2 k_Q(r) C_Q(r,t) dr \tag{3.3}$$

where $C_Q(r, t)$ is the quencher concentration at a distance r and time t, which is governed by the rate of diffusion and the rate of quenching, given by:

$$\frac{\partial y(r,t)}{\partial t} = -D\nabla^2 y(r,t) - k_Q(r) y(r,t) \tag{3.4}$$

A normalized concentration of quencher, $y(r,t)$ was defined to simplify the calculation:

$$y(r,t) = C_Q(r,t) / C_Q^0 \tag{3.5}$$

Using appropriate initial conditions and boundary conditions,

$$y(r, t = 0) = 1 \tag{3.6}$$

$$\left(\frac{\partial y(r,t)}{\partial t}\right)_{r=a_0} = 0 \tag{3.7}$$

$$\lim_{r\to\infty} y(r,t) = 1 \tag{3.8}$$

we numerically solved for $C_Q(r, t)$ and then $k(t)$ and $I(t)$ for each C_Q^0, allowing us to fit the Stern-Volmer curve with k_0 = 42.6 ns^{-1} and β = 1.6 $Å^{-1}$, as shown (Figure 3.3). These parameters are comparable with values obtained for other distance-dependent quenching processes.[153,158,159] For example, for a fixed value of a_0 = 8.0 Å, k_0 and β were determined to be 60 ns^{-1} and 1.3 $Å^{-1}$, respectively, for the fluorescence quenching of p-Bis[2-(5-phenyloxazolyl)]benzene (POPOP) by 1,2,4-Trimethoxybenzene.[158]

3.4.3 Quenching Mechanism. There are several mechanisms by which SeMet could quench Phe_{CN} fluorescence, including FRET, electron transfer, and the heavy atom effect. As there is no overlap between the fluorescence emission spectrum of Phe_{CN} and the absorption spectrum of SeMet (Figure 3.4), a prerequisite for FRET to occur,[54] we do not believe that the quenching of the Phe_{CN} fluorescence by SeMet is through the mechanism of FRET. Furthermore, we tentatively rule out the possibility of the heavy atom effect, which is known to increase the intersystem crossing rate of the fluorophore,[158] as no detectable triplet state formation was observed for Phe_{CN} when a high concentration of SeMet (8 mM) was present (Figure 3.5). Thus, most likely SeMet quenches the Phe_{CN} fluorescence through an electron transfer mechanism. To further verify this possibility, we carried out cyclic voltammetry experiment on Phe_{CN}. As shown (Figure 3.6), the oxidation potential of Phe_{CN} was measured to be 0.921 V vs. NHE at pH 7.0, a value similar to that of other aromatic amino acids.[160] Because the reduction potential of SeMet was previously determined to be 1.21 V vs NHE at pH 7.0,[161] we can use the following Rehm and Weller equation[158] to estimate the free energy change (ΔG_{ET}) of the corresponding electron transfer reaction:

$$\Delta G_{ET} = E_{1/2}{}^{ox} - E_{1/2}{}^{red} - E_s + C \tag{3.9}$$

where $E_{1/2}^{ox}$ is the half-wave oxidation potential of the electron donor Phe_{CN}, $E_{1/2}^{red}$ is the half-wave reduction potential of the electron acceptor SeMet, E_s is the singlet excitation energy of the fluorophore Phe_{CN} (280 nm),[5] and *C* is the Coulomb energy change described by:

$$C = -\frac{e^2}{\varepsilon a 4\pi\varepsilon_0} \tag{3.10}$$

where ε is the dielectric constant of water and *a* is the closest distance separation (7 Å). As expected, ΔG_{ET} was estimated to be -4.7 eV, a value that is consistent with the notion that the Phe_{CN} fluorescence is quenched by SeMet via an electron transfer mechanism.[158]

3.4.4 MD Simulations. To investigate the origin of the non-single-exponential fluorescence decay of SeMet-Phe_{CN}, we carried out a 100 ns MD simulation on Met-Phe_{CN} in water at 298 K, as the force field parameters for SeMet were not available to us. While this will limit the extent to which direct and quantitative comparison can be made between simulation and experimental results, it will provide a qualitative description of the conformation distribution of SeMet-Phe_{CN}. To assess the distribution of the separation distance between the quencher and fluorophore, which is defined as the

distance between the sulfur atom in Met and the COM of the Phe_{CN} sidechain, a total of 200,000 MD frames, with adjacent frames separated by 0.5 ps, were saved. As shown (Figure 3.7), the resultant distance distribution from the last 50 ns of the trajectory consists of two peaks and can be fit by the sum of two Gaussians, centered at 9.0 Å (σ^2 = 1.9, relative amplitude = 45%) and 10.6 Å (σ^2 = 0.7, relative amplitude = 55%). Using these distances and the quenching parameters determined above from the Stern-Volmer analysis, we calculated the fluorescence decay lifetimes of these two components to be 3.4 and 0.5 ns. Comparing these values to those (i.e., 2.0 and 0.2 ns) determined experimentally for SeMet-Phe_{CN}, a qualitative agreement is observed with deviations likely due to using the Met force field instead of that for SeMet. Thus, these MD simulation results support the notion that the bi-exponential fluorescence decay of SeMet-Phe_{CN} arises from two distinguishable conformational ensembles.

3.4.5 Conformational Distribution of Short Polyproline Peptides. It is often assumed that polyproline adopts a rigid PPII, *all-trans* structure and, thus, can be used to determine the Förster distance of FRET pairs.[162] Though the *all-trans* conformation of polyproline is dominant in aqueous solution, there is evidence that other conformations can also populate.[69,70,163–167] For example, using single-molecule measurements, Best *et al*.[69] showed that the end-to-end distance of long polyproline chains (>8 Pro) can deviate significantly from the value predicted based on the corresponding PPII structure and demonstrated that as much as 30% of the internal peptide bonds could adopt the *cis* configuration in water. For short polyproline sequences (<6 Pro), computer simulations[168] have shown that the energy difference between the *all-cis* and *all-trans*

conformations in water is small (~1 kcal/mol), allowing both states to be populated. Moreover, nuclear magnetic resonance (NMR) measurements[163–165] have indicated that there could be as much as 65% *cis*-favored conformations for polyproline chains with 1-3 Pro residues. To further verify the notion that short polyproline peptides can adopt non-PPII structures in aqueous solution and also test the utility of the Phe_{CN}-SeMet fluorophore-quencher pair, we measured the Phe_{CN} fluorescence decay kinetics of the following peptides, SeMet-$(Pro)_n$-Phe_{CN} (n = 1-4).

For all four peptides, as shown (Figure 3.8), the fluorescence decay of Phe_{CN} is reasonably well fit (i.e., $\chi^2 < 1.2$) by a bi-exponential function, suggesting that there are at least two conformations present in the lifetime of the fluorophore. Further MEM analysis yielded consistent results. Interestingly, as indicated (Table 3.1), the longer lifetime (i.e., τ_1) in each case shows a clear increase with increasing the number of prolines, while the value of τ_2 does not show such a trend. These longer lifetimes, and therefore longer fluorophore-quencher separation distances, are likely due to the *all-trans* conformations of the peptides, as a single *cis*-bond can significantly reduce the end-to-end distance of polyproline chains.[69,70] More convincing evidence in support of this assignment is that the fluorophore-quencher separation distance calculated from the quenching rate (Table 3.1) shows a linear correlation with the C_α-C_α distance between the first and last amino acid when all the peptide bonds are *trans* (Figure 3.9). It should be noted that the increase of the calculated sidechain-to-sidechain distance with n is smaller than that of the C_α–C_α distance. This is most likely due to the fact that in these peptides the Phe_{CN} and SeMet sidechains sample a distribution of separation distances and the associated conformational transitions occur on a timescale that is faster than the

Phe_{CN} fluorescence lifetime. Since a shorter fluorophore–quencher separation distance would lead to a faster quenching rate, it would contribute more toward the decrease of the fluorescence lifetime. In other words, the calculated fluorophore–quencher separation distance in this case is skewed toward the short-distance side of the true separation distance distribution, leading to a smaller increase with *n*. On the other hand, the shorter lifetime components should arise from non-PPII conformations. An *all-cis* polyproline structure, with 1.9 Å per proline residue, is shorter than an *all-trans* structure, with 3.1 Å per proline residue. Varying conformations of *cis* and *trans* will give rise to these shorter lifetime components. Taken together, these results validate not only the utility of the Phe_{CN}-SeMet fluorophore-quencher pair in peptide or protein conformational studies, but also the notion that polyproline is not a rigid chain and caution should be taken when use it in distance determinations.

3.5 CONCLUSIONS

Fluorescence quenching through the mechanism of electron transfer requires the fluorophore and quencher to be in van der Waals contact, thus allowing measurement of very short distances. Here, we hypothesized and experimentally verified that Phe_{CN} and SeMet constitute such a fluorophore-quencher pair. Through the use of steady-state and time-resolved fluorescence measurements, we were able to determine the quenching rate constant to be $k_Q = k_0\exp(-\beta(r - a_0))$, where k_0 = 42.6 ns^{-1}, β = 1.6 $Å^{-1}$, and a_0 = 7.0 Å. To further demonstrate the utility of this fluorophore-quencher pair, we investigated the end-to-end distances of a series of short polyproline chains, SeMet-$(Pro)_n$-Phe_{CN} with n = 1-4. We found in all cases that the fluorescence decay of Phe_{CN} follows a bi-exponential function, indicating that short polyproline peptides can sample different

conformations. Because of their small sizes, we believe that this amino-acid based fluorophore-quencher pair will be valuable in structural, conformational, and dynamical studies of proteins.

3.6 ACKNOWLEDGEMENTS

The authors would like to thank Dr. Christopher C. Moser in the Department of Biochemistry and Molecular Biophysics at the University of Pennsylvania for the use the cyclic voltammetry instrument and the helpful insight into the electrochemistry. The authors gratefully acknowledge financial support from the National Institutes of Health (GM-065978 and P41-GM104605). M.R.H. is supported by a National Science Foundation Graduate Research Fellowship (DGE-1321851).

3.7 ORIGINAL PUBLICATION

This Chapter has been published and has been reproduced by permission of the PCCP Owner Societies from *Physical Chemistry Chemical Physics*: Mary Rose Mintzer, Thomas Troxler, Feng Gai, (2015) 17, 7881-7887 DOI: 10.1039/C5CP00050E.

Table 3.1: Fluorescence Decay Kinetics of SeMet-Containing Peptides. Fluorescence lifetime (τ) and relative amplitude (A) determined from fitting the fluorescence decay of each peptide to a bi-exponential function. For the polyproline peptides, the corresponding fluorophore-quencher separation distance (r) was calculated for each component using k_0 = 42.6 ns^{-1}, β = 1.6 Å^{-1}, and a_0 = 7.0 Å.

Peptide	τ_1 (ns)[a]	A_1(%)	r_1 (Å)	τ_2 (ns)[a]	A_2(%)	r_2 (Å)
Gly-Phe$_{CN}$-Gly	7.5	100	------	------	------	------
SeMet-(Pro)$_4$-Phe$_{CN}$	6.0	71	11.6	0.6	29	9.1
SeMet-(Pro)$_3$-Phe$_{CN}$	5.3	53	11.3	0.7	47	9.2
SeMet-(Pro)$_2$-Phe$_{CN}$	3.8	58	10.7	0.9	42	9.4
SeMet-Pro-Phe$_{CN}$	2.6	44	10.3	0.8	56	9.4
SeMet-Phe$_{CN}$	2.0	47	10.1	0.2	53	8.5

[a]The error in these values was estimated to be ±5%

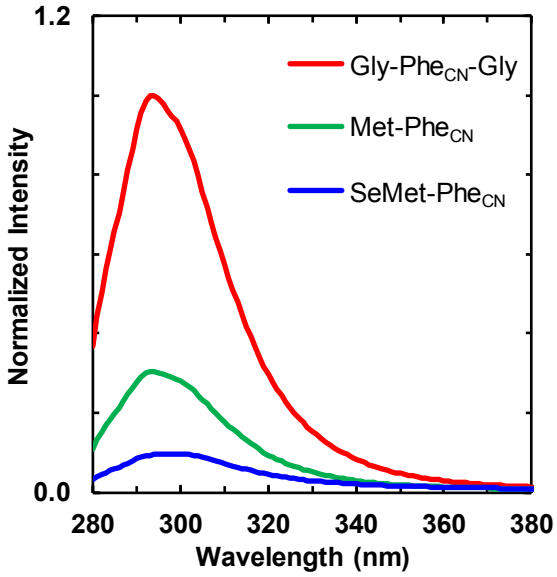

Figure 3.1: Normalized fluorescence spectra of Gly-Phe$_{CN}$-Gly, Met-Phe$_{CN}$, and SeMet-Phe$_{CN}$, as indicated.

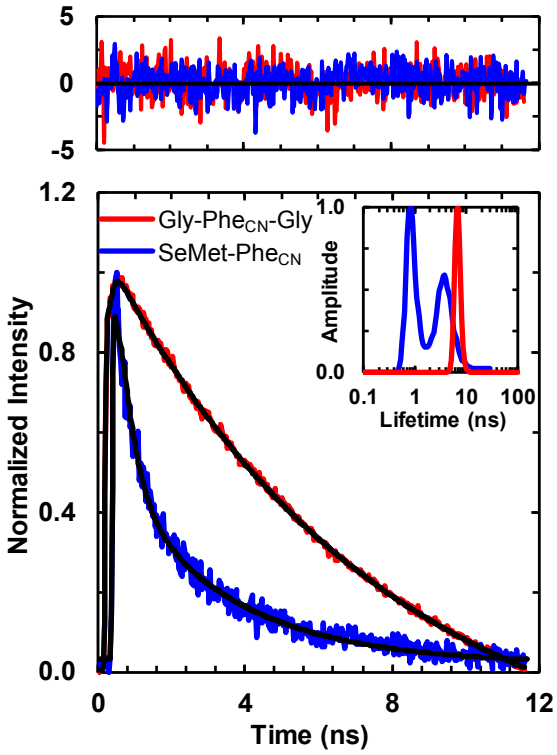

Figure 3.2: Fluorescence decay kinetics of Gly-Phe$_{CN}$-Gly and SeMet-Phe$_{CN}$. The smooth lines are fits of these data to a single-exponential function (Gly-Phe$_{CN}$-Gly) and a bi-exponential function (SeMet-Phe$_{CN}$). The residuals of the fits are shown in the top panel and the resulting fitting parameters are listed in Table 3.1. Shown in the inset are the lifetime distributions obtained from MEM analysis for both peptides.

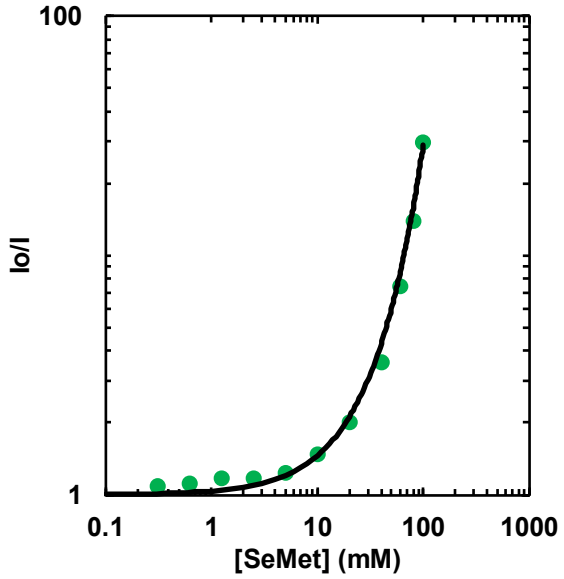

Figure 3.3: Relative fluorescence intensity of free Phe_{CN} versus concentration of free SeMet. The smooth line is a fit to the data using the approach discussed in the text and the following fitting parameters: k_0 = 42.6 ns^{-1}, β = 1.6 $Å^{-1}$ and a_0 = 7.0 Å.

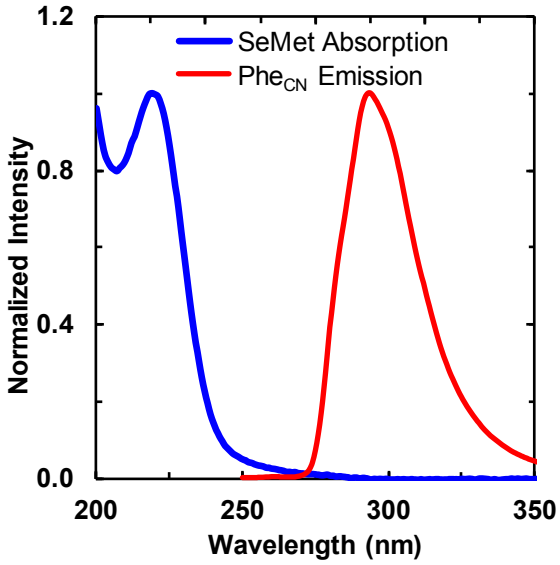

Figure 3.4: Absorption spectrum of free SeMet and emission spectrum of free Phe_{CN}, both in Millipore water, normalized to their respective maxima, 219 nm and 300 nm. There is no absorbance overlap of SeMet with the fluorescence emission of Phe_{CN}, making FRET improbable.

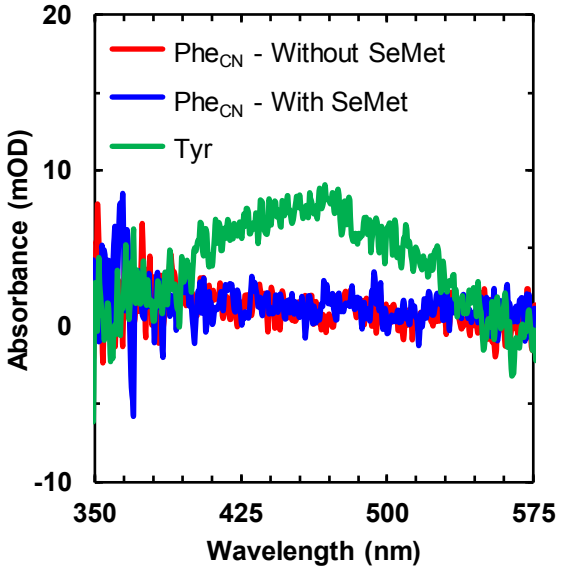

Figure 3.5: Nanosecond transient absorption measurements of the free amino acids Tyr, Phe$_{CN}$, and a Phe$_{CN}$/SeMet mixture at a delay of 1 μs. Phe$_{CN}$-With and -Without Quencher show no noticeable peaks, while Tyr contains a broad peak centered ~475 nm.

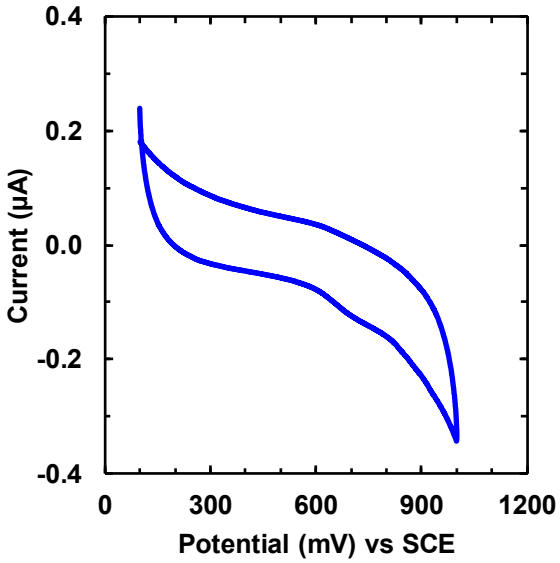

Figure 3.6: Cyclic Voltammogram of free Phe_{CN}, structure as shown, in 0.2 M KCl-containing phosphate buffer at pH 7.0 and room temperature, with a scan rate of 10 mV/s. The oxidation potential was approximately 680 mV vs. SCE (921 mV vs. NHE).

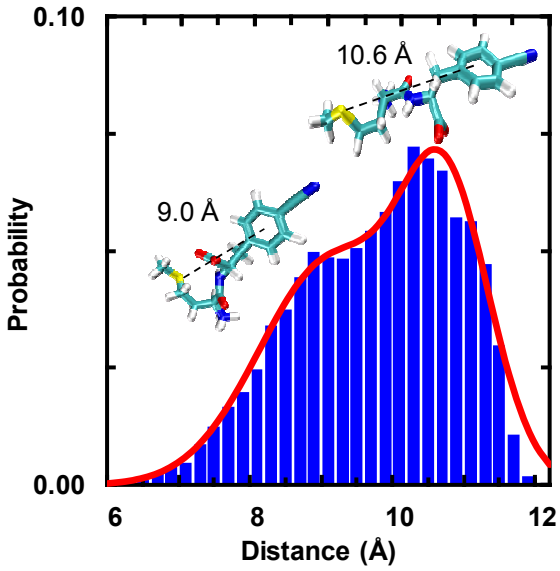

Figure 3.7: Probability distribution of the distance between S and the COM of Phe_{CN} sidechain obtained from MD simulation. The red line represents the best fit of this distribution to the sum of two Gaussians with centers at 9.0 Å (σ^2 = 1.9, relative amplitude = 45%) and 10.6 Å (σ^2 = 0.7, relative amplitude = 55%), respectively. A representative snapshot for each center distance is also shown.

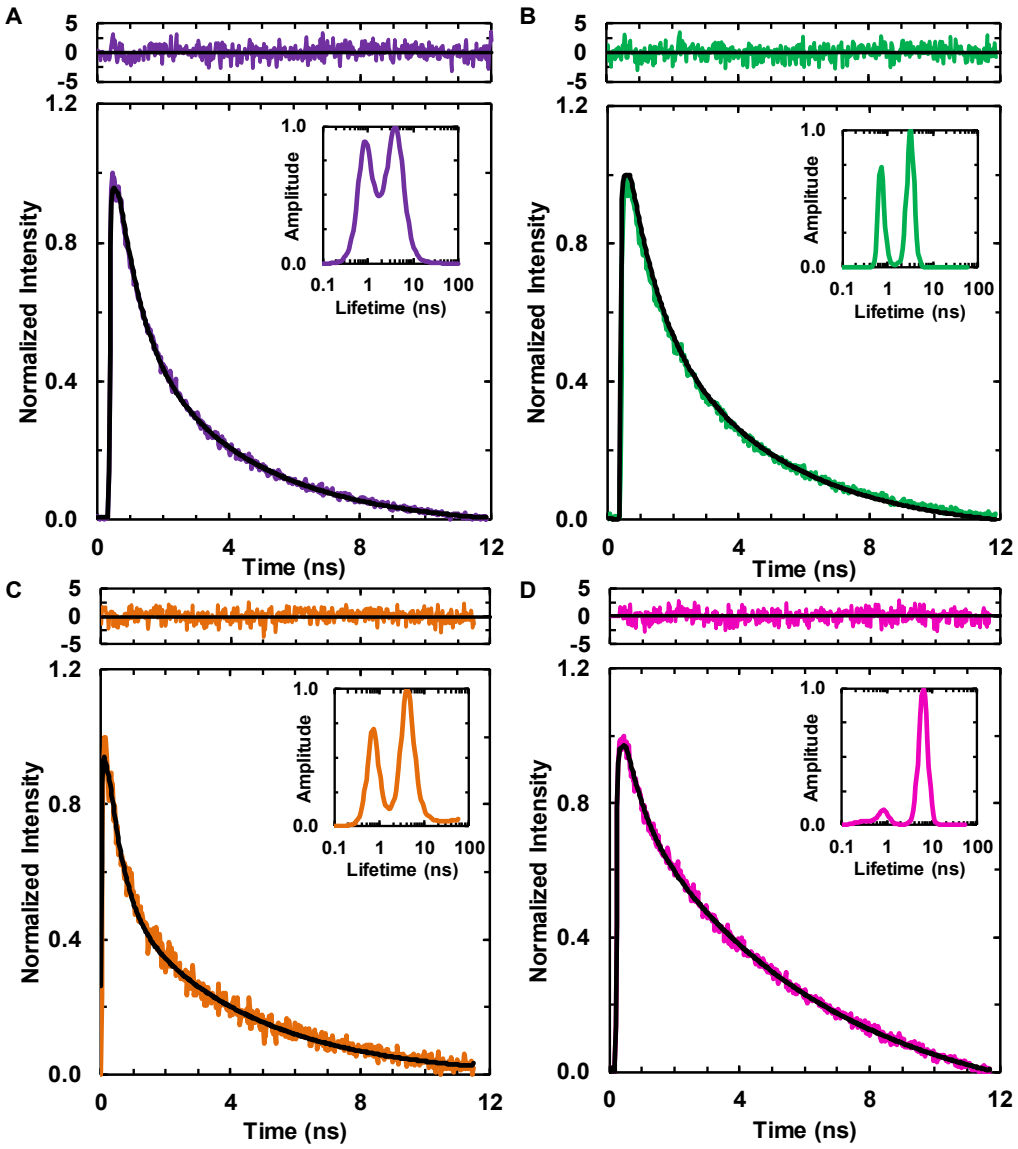

Figure 3.8: Fluorescence decay kinetics of SeMet-Pro-Phe$_{CN}$ (A), SeMet-(Pro)$_2$-Phe$_{CN}$ (B), SeMet-(Pro)$_3$-Phe$_{CN}$ (C), and SeMet-(Pro)$_4$-Phe$_{CN}$ (D). The smooth black line in each case represents the best fit of the corresponding decay to a bi-exponential function, with the corresponding residuals shown in the top panel and the fitting parameters given in Table 3.1. Additionally, shown in the inset is the bimodal lifetime distributions obtained from MEM analysis for each peptide.

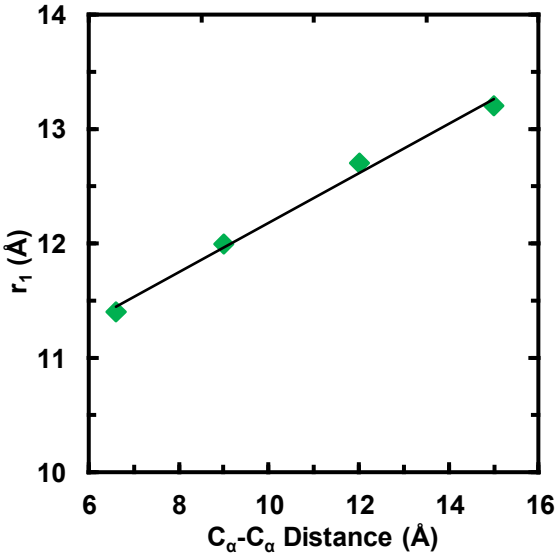

Figure 3.9: C_α-C_α distance of polyproline peptides as a function of fluorophore-quencher separation distance.

4: Blue Fluorescent Amino Acid for Biological Spectroscopy and Microscopy

4.1: ABSTRACT

Many fluorescent proteins are currently available for biological spectroscopy and imaging measurements, allowing a wide range of biochemical and biophysical processes and interactions to be studied at various length scales. However, in applications where a small fluorescence reporter is required or desirable, the choice of fluorophores is rather limited. As such, continued effort has been devoted to the development of amino acid-based fluorophores that do not require a specific environment and additional time to mature, have a large fluorescence quantum yield, long fluorescence lifetime, good photostability, and an emission spectrum in the visible region. Herein, we show that a tryptophan analog, 4-cyanotryptophan, which differs from tryptophan by only two atoms, is the smallest fluorescent amino acid that meets these requirements and has great potential to enable new *in vitro* and *in vivo* spectroscopic and microscopic measurements of proteins.

4.2: INTRODUCTION

Of the amino acids that are inherently responsible for the fluorescence of proteins, tyrosine (Tyr), tryptophan (Trp), and phenylalanine (Phe), Trp is the most widely-used fluorescence reporter of protein structure, function, and dynamics, as its fluorescence quantum yield (QY) is comparatively large and also sensitive to environment.[49,169,170] However, Trp absorbs/emits in the ultraviolet wavelength region and has low photostability. Combined, these factors render this naturally occurring fluorescent amino acid hardly useful as a fluorophore for single-molecule

measurements[171] and imaging applications, especially under *in vivo* conditions. For this reason, significant efforts have been devoted to identifying small unnatural fluorophores[77,78,172–176] that could overcome this limitation. So far, only naphthalene-based fluorophores, such as Prodan[173,174] and Aladan[176] have been shown to be useful in this regard. However, these fluorophores are not structurally based on any naturally-occurring amino acids and are larger in size than Trp, making them less attractive in applications where there are stringent requirements for the fluorophore size and structure. Therefore, the development of a smaller, ideally amino acid-based, fluorophore that does not require additional time to mature, have a large fluorescence quantum yield (QY), long fluorescence lifetime, good photostability, and an emission spectrum in the visible range, would enable new biological research. Herein, we show that a tryptophan (Trp) analog, 4-cyanotryptophan (4CN-Trp), whose structure is shown in Figure 4.1, meets these requirements and has great potential to expand biological fluorescence spectroscopy and microscopy into new territory.

Many past studies have focused on Trp-based unnatural amino acids, including azatryptophans[77,78,177] and various indole-ring substituted analogs,[75,76,178,179] aiming to identify new biological fluorophores. While some Trp analogs indeed exhibit improved fluorescent properties than Trp, none of them has found broad applications due to certain photophysical limitations. Recently, Benkovic, Hecht, and coworkers[75] showed that 6-cyanotryptophan (6CN-Trp) and 7-cyanotryptophan (7CN-Trp) have a significantly increased QY (about 0.5 in methanol) compared to that of Trp (about 0.15 in water), suggesting that cyanotryptophans are potential candidates that meet the aforementioned requirements. However, the emission maxima of 6CN-Trp and 7CN-Trp in methanol are centered at 370 and 390 nm, respectively, making them less attractive as visible

fluorophores. As shown (Figures 4.2 and 4.3), the absorption and emission spectra of 4-cyanoindole (4CNI), the sidechain of 4CN-Trp, are significantly shifted from those of indole (or Trp) and, in particular, an emission maximum of 405 nm suggests that 4CN-Trp could be a viable candidate for the aforementioned amino acid-based fluorophores. Indeed, our results confirm that 4CN-Trp emits in the blue region of the visible spectrum with a large QY (>0.8) and has a long fluorescence lifetime (12.6 ns), a reasonably large molar extinction coefficient (~6000 M^{-1} cm^{-1}), and good photostability. Taken together, these qualities make 4CN-Trp an attractive blue fluorescent amino acid (BFAA) for biological spectroscopy and microscopy.

4.3: MATERIALS AND METHODS

4.3.1: Synthesis of N-alpha-Boc-4-bromo-L-tryptophan methyl ester. 4-bromo-L-tryptophan (283 mg, 1 mmol) was dissolved in 20% MeOH in DCM (10 mL) and trimethylsilyldiazomethane (2 M in hexane) was added dropwise at room temperature. The solution was stirred for 3 hrs. and the volatiles were removed under reduced pressure. The residue was dissolved in anhydrous DCM (2 mL) then triethylamine (2 mmol) and Boc anhydride (0.23 mL, 1.1 mmol) were added. The mixture was stirred overnight and concentrated under reduced pressure. The residue was purified by silica gel column chromatography (0% to 10% MeOH in DCM) to yield N-alpha-Boc-4-bromo-L-tryptophan methyl ester (192 mg, 48%)

ESI-MS (positive): 397.7 (MH^+)

4.3.2: Synthesis of N-alpha-Boc-4-cyano-L-tryptophan methyl ester. A mixture of N-alpha-Boc-4-bromo-L-tryptophan methyl ester (192 mg, 0.48 mmol), $Zn(CN)_2$ (38 mg,

0.33 mmol), tBuXPhos Pd G3 (20 mg, 0.025 mmol) in THF (0.5 mL) and H_2O (2.5 mL, degassed by sonication) was vigorously stirred at 40 °C overnight under N_2. Then, a saturated $NaHCO_3$ solution (2 mL) and ethyl acetate (3 mL) were added, stirred for 5 min, and separated. The aqueous layer was extracted with ethyl acetate and the combined organic layer was dried over Na_2SO_4. The volatiles were removed under reduced pressure and the residue was purified by silica gel column chromatography (10% to 70% ethyl acetate in hexanes) to give the desired product (16 mg, 10%).

^{1}H NMR (300 MHz, DMSO-D6) δ ppm 11.53 (s, 1H), 7.70 (d, J=7.8 Hz, 1H), 7.49 (d, J= 7.1 Hz, 1H), 7.40 (s, 1H), 7.31 (d, J=7.4 Hz, 1H), 7.21 (t, J =7.8 Hz, 1H), 4.30-4.25 (m, 1H), 3.59 (s, 3H), 3.40-3.3 (m, 1H), 3.20-3.14 (m, 1H), 1.31 (s, 9H)

ESI-MS (positive): 366.4 (MNa^+)

4.3.3: Synthesis of N-alpha-acetyl-4-cyano-L-tryptophan methyl ester (4CN-Trp). N-alpha-Boc-4-cyano-L-tryptophan methyl ester (16 mg, 0.047 mmol) was treated with 50% trifluoroacetic acid in DCM (2 mL) for 3 hrs. The volatiles were removed under reduced pressure and the sample was lyophilized. The product was purified by reverse phase HPLC. N-terminal acetylation was accomplished by stirring the amino acid in acetic anhydride (5:1 molar ratio) in DMF for 1 hr. Water was then added to the solution and lyophilized. Purified amino acid was obtained by reverse phase HPLC.

ESI-MS (positive): 286 (MH^+)

4.3.4: Synthesis of N-alpha-acetyl-4-cyano-L-glycyl-glycyl-tryptophan methyl ester (Gly-Gly-4CN-Trp). The rest of the crude N-alpha-Boc-4-cyano-L-tryptophan methyl ester solid was dissolved in DMF (1 mL) and Fmoc-Gly-Gly-OH (20 mg, 0.056 mmol),

DIPEA (26 µL, 0.15 mmol), HATU (16 mg, 0.042 mmol) were added. The mixture was stirred at room temperature overnight and water (3 mL) was then added. The mixture was extracted with ethyl acetate (5 mL) and the organic layer was concentrated. The residue was dissolved in 30% diethylamine in acetonitrile (5 mL) and stirred for 1 hr. The volatiles were removed under reduced pressure and the residue was purified by reverse phase HPLC. N-terminal acetylation was accomplished by stirring the peptide in acetic anhydride (5:1 molar ratio) in DMF for 1 hr. Water was then added to the solution and lyophilized. Purified peptide was obtained by reverse phase HPLC.

ESI-MS (positive): 358 (MH^+)

4.3.5: Materials and Sample Preparation. L(-)-Tryptophan (99%, Acros Organics), indole (99+%, Sigma), 4-cyanoindole (4CNI) (97%, Acros Organics), 4-cyanoindole-3-acetic acid (4CNI-3AA) (98%, Ark Pharm), 9,10-diphenylanthracene (DPA) (98%, Acros Organics), and enhanced blue fluorescent protein (EBFP) (97%, BioVision, Inc.) were used as received. 4CN-Trp*-Gly and 4CN-Trp*-Met were synthesized manually and 4CN-Trp*-MpX (Sequence: 4CN-Trp*-INWKGIAAMAKKLL-NH_2) was synthesized on a Liberty Blue Automated Microwave Peptide Synthesizer (CEM Corporation); all utilized standard Fmoc-based solid phase peptide synthesis to couple the 4CN-Trp* fluorophore to the N-terminus of the peptide of interest via the carboxylic acid on 4CNI-3AA. Purification of peptides was achieved using reverse phase HPLC and identification and verification of the peptides were done using matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS). Stock solutions of Trp, 4CNI, and 4CNI-3AA were made by dissolving the respective compound in Millipore water and the solute concentration was determined by weight. The DPA samples in cyclohexane were

prepared similarly. The concentrations of 4CN-Trp*-containing peptides and 4CN-Trp samples in Millipore water were determined optically at 305 nm using the extinction coefficient determined for 4CNI-3AA (5780 M^{-1} cm^{-1} in water) (Figure 4.4).

4.3.6: Cell Culture. HEK293T/17 cells were seeded in an 8-well Lab-Tek chamber slide at 3,000 cells per well. Cells were grown overnight at 37 °C in a 5% CO_2 incubator. Before use, the growth media was removed by pipette from the cells and the cells were subsequently washed once with pH 7.4 phosphate buffer saline (PBS) (Thermo Fisher). An appropriate amount of a 500 μM stock solution of 4CN-Trp*-MpX in Millipore water was added to each well containing the cells in PBS to achieve the desired final peptide concentration (1 or 10 μM). The mixture was allowed to incubate for 1 minute before wide field imaging measurements. For cell samples used in the confocal microscopic measurement, the mixture was incubated ~10 hours at room temperature before use.

4.3.7: Absorption Measurements. UV-Vis absorption spectra were collected on a Jasco V-650 UV-Vis spectrophotometer using a 1.0 cm quartz cuvette at room temperature.

4.3.8: Static and Time-Resolved Fluorescence Measurements. All static fluorescence measurements were collected on a Jobin Yvon Horiba Fluorolog 3.10 fluorometer at 25 °C (or 37 °C, as noted) using a 1.0 cm quartz cuvette with spectral resolution of 1.0 nm and an integration time of 1.0 s/nm. In order to minimize the inner filter effect, self-quenching, and potential detector saturation, each and every solution used in the QY measurements was prepared by diluting a stock solution of high concentration (the

absorbance at the corresponding excitation wavelength was in the range of 0.1-0.5) 100-fold. QYs were determined as described in the text and for each case, the reported value corresponds to the average of three measurements using independent solutions. Additionally, for the QY determination of 4CNI-3AA, 4CN-Trp, and Gly-Gly-4CN-Trp at 325 nm, the gradient method was utilized where a series of dilutions were used to determine the ratio between the integrated area of the fluorescence, *I*, and the absorbance at the λ_{ex}, *A* (Figure 4.5). This ratio is then used in Equation 4.1 in the text to calculate the QY.

For the photobleaching experiments, EBFP and 4CN-Trp*-Gly solutions were prepared in water such that the integrated optical densities for both samples are the same (~0.5) in the spectral region of the excitation light (355 ± 2.5 nm), defined by the excitation slit width of the Fluorolog 3.10 fluorometer that uses a 450 W Xe arc lamp as the light source. Emission intensity was recorded at 450 ± 1 nm for EBFP and at 415 ± 1 nm for 4CN-Trp*-Gly every minute for 10 hrs.

Time-resolved fluorescence measurements were obtained on a time-correlated single photon counting (TCSPC) system with a 0.4 cm quartz cuvette at room temperature. The details of the TCSPC system have been described elsewhere.[76] Briefly, the 270 nm excitation pulse was generated by frequency-tripling of the fundamental output (800 nm) of a home-built femtosecond Ti:Sapphire oscillator (85 MHz). The repetition rate was decreased to 21 MHz using an electro-optical pulse picking system (Conoptics Inc.). Fluorescence decays were collected under the magic-angle polarization condition using a MCP-PMT detector (Hamamatsu R2809U) and a TCSPC board (Becker and Hickl SPC-730). Rejection of excitation light was accomplished by a 405/50 nm bandpass filter (Semrock) and a 300 nm longpass filter

(Semrock). Fluorescence decays were deconvoluted with the experimental instrument response function (IRF) and were fit to a single-exponential function using FLUOFIT (Picoquant GmbH). The OD of each sample was approximately 0.1 at the excitation wavelength (270 nm).

4.3.9: Fluorescence Imaging. Wide field fluorescence images were acquired at room temperature using an Olympus IX71 inverted microscope equipped with a 100 W Hg lamp, a 60× (0.9 NA) water objective, and HCImage Live software. Excitation light was selected using a 355/25 nm bandpass filter and the emission was isolated via a 420 nm longpass filter. The integration time for each image was 50 ms. Image processing and analysis were carried out using the ImageJ 1.50 software.[180] Confocal images were collected on a home-built stage-scanning confocal microscope at room temperature. Specifically, an inverted microscope (Eclipse TE300, Nikon) equipped with a 100× oil immersion objective (CFI S Fluor, Nikon) was used to condense the excitation light, which was derived from a UV LED light source (M340L4, Thorlabs). This LED produces CW light centered at 344 nm with a FWHM bandwidth of 10 nm. The light emitted from the LED was first collimated by a collimation adapter (SM1P25-A, Thorlabs), followed by further collimation of the resultant beam using a two-lens collimator with a 60 μm pinhole at the co-focal point of the two focusing lenses. The power of the incident light was measured to be approximately 6 μW before entering the microscope. A dichroic mirror with an edge wavelength of 349 nm (ZT349rdc, Chroma Technology) and a longpass filter with a cutoff wavelength of 365 nm (ET365lp, Chroma Technology) were used to allow the pass/rejection of the emission/excitation light. Fluorescence emission was then focused and passed through a 60 μm confocal pinhole and then measured by a single

photon counting detector (SPCM-AQRH-15, Excelitas Technologies) with an integration time of either 0.5 or 1.0 ms at each scanning position. The confocal image was acquired by a point scanning method, which was accomplished by a high-resolution XY microscope stage (P-734, Physik Instrumente) controlled by a piezo controller (E-501.00, Physik Instrumente).

4.4: RESULTS AND DISSCUSSION

4.4.1: Fluorescence Quantum Yield Measurements. To explore the feasibility of using 4CN-Trp as a biological fluorescence emitter, we first examined the QY of its fluorophore, 4CNI. As shown (Figure 4.2), the absorption spectrum of 4CNI in water is broader than and red-shifted from that of Trp, making it possible to prepare its electronically excited and fluorescent state using an excitation wavelength (λ_{ex}) of greater than 310 nm, without exciting any of the naturally-occurring fluorescent amino acids. This constitutes one of the key advantages of using 4CNI-based amino acid fluorophores. Similarly, the fluorescence spectrum of 4CNI in water is red-shifted from that of indole (Figure 4.3a), with a maximum emission wavelength of 405 nm. In addition, and perhaps more importantly, the integrated fluorescence intensity of 4CNI is much larger than that of indole, indicating that 4CNI has a much larger fluorescence QY than indole. To determine the QY of 4CNI, we quantitatively compared its fluorescence spectrum with that of Trp or 9,10-diphenylanthracene (DPA) (Figure 4.3b) using the following relationship:[54]

$$QY_S = QY_R \frac{I_S}{I_R} \frac{A_R}{A_S} \frac{n_S^2}{n_R^2} \tag{4.1}$$

where *I* is the integrated fluorescence intensity, *A* is the optical density of the fluorophore at λ_{ex}, *n* is the index of refraction of the solvent, and the subscripts S and R represent the sample and reference, respectively. Both Trp and DPA are commonly used fluorescence QY standards,[181] with a QY of 0.15 for Trp in water at λ_{ex} = 270 nm and 0.9 (0.97) for DPA in cyclohexane at λ_{ex} = 325 nm (355 nm). Therefore, for an λ_{ex} of 270-290 nm, we used Trp as the reference, whereas for an λ_{ex} of equal to or larger than 325 nm, we used DPA as the reference.

As indicated (Table 4.1), at λ_{ex} of 270 nm, the QY of 4CNI was determined to be 0.85 ± 0.13. To the best of our knowledge, this value is the largest among the reported QY values of all indole (the sidechain of Trp) derivatives, signifying the potential utility of 4CNI as a biological fluorophore. Moreover, the high QY of 4CNI is preserved at even longer excitation wavelengths (Table 4.1). Taken together, these results suggest that 4CN-Trp could be a bright enough fluorophore for biological single-molecule and imaging experiments, excited in a spectral region separated from that of other intrinsic fluorescent amino acids.

Because the photophysical properties of 4CNI could change when incorporated in an amino acid or peptide environment, we further quantified the absorption and emission spectra of 4CN-Trp, Gly-Gly-4CN-Trp, 4-cyanoindole-3-acetic acid (4CNI-3AA), and 4CN-Trp*-Gly. To avoid any potential effect of charged termini on the photophysics of the 4CNI fluorophore, both 4CN-Trp and Gly-Gly-4CN-Trp have an acetylated N-terminus and a methylated ($COOCH_3$) C-terminus. The reason that we included 4CNI-3AA is that (1) it is an analog of 4CN-Trp, (2) it is commercially available and relatively

inexpensive, and (3) it can be used to conveniently append a 4CNI fluorophore to the N-terminal end of a peptide or protein by forming an amide bond between the -COOH group of 4CNI-3AA and the -NH_2 group of the polypeptide. For example, a reaction between 4CNI-3AA and glycine (Gly) produces a dipeptide-like mimic (referred to as 4CN-Trp*-Gly), which is identical to the 4CN-Trp-Gly dipeptide, except that 4CN-Trp*-Gly contains one less carbon in its backbone and is missing the N-terminus (Figure 4.1). As shown (Figures 4.2 and 4.3), the absorption and emission spectra of 4CN-Trp, 4CNI-3AA, Gly-Gly-4CN-Trp, and 4CN-Trp*-Gly in water are very similar to each other, suggesting that 4CNI-3AA can be used as an alternative of 4CN-Trp. Furthermore, their absorption spectra extend further to the longer wavelengths in comparison to that of 4CNI, with a second maximum at 325 nm. These absorption characteristics are important as they render the use of various commercially-available lasers, such as the 355 nm YAG laser or the 325 nm HeCd laser commonly used in flow cytometry, to excite the 4CN-Trp fluorophore.

As indicated (Figure 4.3b), the fluorescence spectra of 4CN-Trp, 4CNI-3AA, Gly-Gly-4CN-Trp, and 4CN-Trp*-Gly have a peak wavelength of greater than 415 nm, making 4CN-Trp a true blue fluorescence emitter. Perhaps more importantly, their fluorescence QYs in water at 25 °C are larger than 0.80 when an λ_{ex} of 325 nm was used (Table 4.1), thus confirming the notion that 4CN-Trp is a highly emissive BFAA. Further measurements on 4CN-Trp*-Gly in a 20 mM phosphate buffer (pH 7.2) at 37 °C and in tetrahydrofuran (THF), an aprotic solvent commonly used to mimic the hydrophobic interior of proteins because of its low dielectric constant, show that the QY is only decreased to about 0.6 (Table 4.1) under these conditions. This is a key finding as it indicates that the strong fluorescence emission property 4CN-Trp is maintained

under physiological conditions and also in a dehydrated environment. It is well known that several amino acids,[58] including methionine (Met),[59] can quench Trp fluorescence. However, as shown (Table 4.1 and Figure 4.6a), the QY of 4CN-Trp*-Met is similar to that of 4CN-Trp*-Gly. This result suggests that the fluorescence QY of 4CN-Trp may not be as sensitive to nearby residues as Trp is. Thus, taken together, the above results suggest that the brightness of a 4CNI fluorophore in a protein will not be significantly affected by its location and can be attached to either a solvent-exposed or buried position.

4.4.2: Fluorescence Lifetime Measurements. In applications where the fluorophore of interest is imbedded in a heterogeneous and light-scattering medium, such as cells, time-based fluorescence measurements can be advantageous. Therefore, we further determined the fluorescence decay kinetics of 4CNI, 4CNI-3AA, 4CN-Trp, Gly-Gly-4CN-Trp, and 4CN-Trp*-Gly. As shown (Figure 4.7), 4CNI, 4CNI-3AA, and 4CN-Trp*-Gly in water afford single-exponential fluorescence decay kinetics with, more interestingly, a long lifetime (e.g., 12.6 ns for 4CN-Trp*-Gly) (Table 4.2). On the other hand, the fluorescence decay kinetics of 4CN-Trp and Gly-Gly-4CN-Trp consist of not only this long lifetime component (~13.6 ns), which dominates the amplitude (~95%), but also a shorter one (~1.8 ns) (Table 4.2). Regardless of the origin of this minor component, the prolonged fluorescence lifetimes of these model systems are consistent with their high QY values. In addition, they demonstrate that 4CN-Trp could be a useful fluorophore in time-gated fluorescence imaging applications as its long florescence lifetime allows more efficient elimination of the autofluorescence background (~6 ns) in experiments involving cells.[29] As shown (Table 4.2 and Figure 4.7d), the long fluorescence lifetime of 4CN-

Trp*-Gly (7.6 ns) is also maintained in THF, indicating that its fluorescence utility is not significantly compromised even when sequestered in a hydrophobic environment in proteins, making it a versatile fluorescence probe. In addition, the fluorescence lifetime of 4CN-Trp*-Met (12.2 ns) is nearly identical to that of 4CN-Trp*-Gly (Table 4.2 and Figure 4.6b), corroborating the conclusion reached above that Met does not quench the fluorescence of the 4CNI fluorophore.

4.4.3: Photostability Measurement. A key factor determining the utility of a fluorophore in single-molecule spectroscopy or confocal imaging application is its photostability. Therefore, we attempted to determine the photostability of the fluorophore in 4CN-Trp*-Gly. However, there are no standard practices to measure the photobleaching rate, as this rate depends upon many factors, including the source intensity (arc lamp or laser), the excitation filters, and the type of experiment being performed (e.g., single-molecule versus bulk),[182] which complicates comparisons between fluorophores measured at different times and places. Therefore, we compared the photobleaching kinetics of 4CN-Trp*-Gly and the enhanced blue fluorescent protein (EBFP)[34] under identical conditions (i.e., both samples have the same integrated absorbance at the excitation wavelength). As indicated (Figure 4.8a), after exposing the sample to 355 nm light, derived from a Jobin Yvon Horiba Fluorolog 3.10 fluorometer equipped with a 450 W xenon arc lamp with the excitation slit width set at 5 nm, the fluorescence intensity of 4CN-Trp*-Gly in water at 25 °C decreased only ~25% from its initial value in 10 hours, while that of EBFP decreased ~75% over the same time period. Thus, these results demonstrate that the 4CN-Trp fluorophore is more resistant to photobleaching than EBFP, a property very useful in measurements, such as time-lapse imaging, wherein the same set of

fluorophores are required to be exposed to excitation light for a relatively long period of time.

Furthermore, to provide a direct visual comparison between the fluorescence intensities of these two fluorophores, we took pictures of a 50 nM 4CN-Trp*-Gly sample and a 50 nM EBFP sample. As shown (Figure 4.8b), both samples emit blue light of comparable brightness when excited with the same light (λ_{ex} = 340 nm, derived from the aforementioned fluorometer), substantiating the notion that 4CN-Trp could be used as an alternative to fluorescent proteins.

4.4.4: Fluorescence Imaging Application. In a proof-of-principle imaging experiment, we used the 4CNI fluorophore to monitor the distribution of an antimicrobial peptide (AMP) inside HEK293T/17 cells. Specifically, we appended a 4CNI fluorophore to the N-terminus of an AMP, Mastoparan X (MpX),[183] via reacting 4CNI-3AA with MpX using the same synthesis method employed to make 4CN-Trp*-Gly (the resultant peptide is hereafter referred to as 4CN-Trp*-MpX). First, we used a commercial wide field microscope (Olympus IX71) equipped with a 60× water objective and a 100 W Hg lamp to test whether the fluorescence is strong enough under conditions commonly used in similar imaging applications (i.e., 1 - 10 μM peptide and ~3000 cells). In comparison to the control (Figures 4.9c and 4.9d), it is apparent that the fluorescence image (Figure 4.9f) acquired with 10 μM peptide allows a clear visualization of the distribution of 4CN-Trp*-MpX inside the HEK cells, which shows that this AMP distributes throughout the cell except in the nucleus. This result indicates, as also observed for other peptides,[184,185] that MpX is not only able to bind to cell membranes but can also permeate into the cytosol. While, as expected, the fluorescence image obtained with 1 μM peptide (Figure

4.9e) is much fainter, it still provides enough contrast to allow a rough visual inspection of the AMP distribution, in comparison to the control (Figures 4.9b and 4.9d). This is a remarkable result considering the fact that the commercial optical filter sets available on the microscope limited the shortest excitation wavelength to 355 nm, which locates at the tail region of the absorption spectrum of the amino acid fluorophore (Figure 4.2). In other words, an excitation wavelength of 330 nm, where 4CN-Trp exhibits a maximum absorbance, would substantially increase the image contrast or render the use of lower peptide concentrations. Furthermore, time-lapse microscopic measurements showed that useful images can still be obtained even after exposing the AMP-bound cells to the excitation light for over 15 minutes (Figure 4.10), consistent with the notion that 4CN-Trp is a highly photostable fluorophore.

It is known that AMPs can cause cell death via two mechanisms, necrosis and apoptosis.[186] Necrosis results in cell membrane rupture and loss of cellular content, whereas apoptosis leads to cell fragmentation and formation of smaller apoptotic bodies that still contain functional organelles. It is apparent that interaction with MpX leads to HEK cell death, as the morphology of the cell becomes more circular and shrinks in size upon increasing 4CN-Trp*-MpX concentration (Figure 4.9). To determine the mechanism of action of MpX toward HEK cells and to also verify the potential utility of 4CN-Trp in confocal microscopy, next we employed a home-built confocal microscope to image the 'products' arising from AMP-induced HEK cell death. For comparison, we used the same cell sample used to generate the wide field image in Figure 4.9e (i.e., in the presence of 1 μM 4CN-Trp*-MpX) but waited ~10 more hours before confocal measurements, to ensure that the entire cell population had died. As shown (Figure 4.11a), large-area confocal images acquired with low spatial resolution suggest that the HEK cells in this

case die via the apoptosis pathway, as manifested by formation of membrane blebs and small vesicle-like particles.[186] Further evidence supporting this notion comes from a confocal image acquired with a higher spatial resolution (Figure 4.11b), which clearly shows the existence of a micron-sized component that can interact with 4CN-Trp*-MpX within a circular bleb. It is worth noting that our confocal microscope utilized a 340 nm LED as the excitation source, which is less focusable and has a low power (>6 μW) compared to a coherent laser beam. Additionally, the photon counting detector in the current confocal setup is optimized to detect photons in the red region of the visible spectrum, with less than 20% efficiency in the fluorescence spectral region of 4CN-Trp*-MpX. Therefore, we believe that better 4CN-Trp fluorescence confocal images (i.e., better resolution and contrast) can be obtained when a coherent light excitation source and a detector optimized to detect blue fluorescence are used.

4.5: CONCLUSIONS

Because of its relatively large molar absorptivity and fluorescence QY, Trp has become the natural target or template of many previous efforts to develop amino acid-based fluorophores for biological applications. While several Trp-based fluorescent unnatural amino acids have been shown to offer improved spectroscopic utility than Trp[77,78,172], none of them has been proven to be useful for *in vivo* imaging applications. Herein, we have shown that by replacing the hydrogen atom at the 4th position of the indole ring of Trp with a nitrile group, the resultant unnatural amino acid, 4CN-Trp, exhibits unique photophysical properties: it has an absorption spectrum peaked at ~325 nm, an emission spectrum peaked at ~420 nm, a large fluorescence QY (0.8 – 0.9), a long fluorescence lifetime (13.7 ns), and good photostability. Combined, these properties

make 4CN-Trp an exceedingly useful fluorescent reporter for *in vitro* and *in vivo* spectroscopic and microscopic applications. While 4CN-Trp has not yet been biosynthetically incorporated into protein systems, we do not believe this is a difficult technological undertaking as various methods[75,178,187,188] have been developed to site-specifically incorporate other Trp analogues, such as 6CN-Trp and 7CN-Trp,[75] hydroxytryptophans,[178] azatryptophans,[178] and methyltryptophans,[188] into the proteins of interest. Furthermore, we believe that the current study is only the first step in exploring the potential utility of 4CN-Trp in biological research; we expect that future studies will investigate its applicability in single-molecule spectroscopy and multiphoton microscopy. In comparison to Trp, which poses several challenges in these types of applications,[189–191] 4CN-Trp has the advantages of being more photostable and having an excitation spectrum that is readily accessible by a typical femtosecond Ti:Sapphire laser (700-1000 nm). Finally, besides its attractive photophysical properties, the small size of 4CN-Trp, which only differs from its naturally occurring counterpart by two atoms, will make this BFAA especially useful in applications where a minimally perturbing fluorescence probe is required.

4.6: ACKNOWLEDGEMENTS

The authors gratefully acknowledge financial support from the National Institutes of Health (P41-GM104605). M.R.H. is supported by a National Science Foundation Graduate Research Fellowship (DGE-1321851). I.A.A. is supported by a NIH T32 Interdisciplinary Cardiovascular Training Grant (T32-HL007954). Additionally, the authors would like to acknowledge Dr. Samantha Wilner and Dr. Tobias Baumgart of the

University of Pennsylvania for generously providing the HEK293T/17 cells and assisting with the wide field imaging measurements.

4.7: ORIGINAL PUBLICATION

This Chapter has been submitted for publication: Mary Rose Hilaire,* Ismail A. Ahmed,* Chun-Wei Lin, Hyunil Jo, William F. DeGrado, and Feng Gai, (2017). (*Equal Contribution)

Table 4.1: Quantum Yields measured for 4CNI, 4CN-Trp, and their derivatives under different excitation wavelengths, temperatures (*T*) and solution conditions, as indicated.

Sample	Solvent	T (°C)	QY 270 nm	QY 280 nm	QY 280 nm	QY 325 nm	QY 355 nm
4CNI	H_2O	25	0.85 ± 0.13	0.86 ± 0.07	0.98 ± 0.08	0.78 ± 0.03	0.64 ± 0.07
4CNI-3AA	H_2O	25	0.71 ± 0.04	0.74 ± 0.04	0.83 ± 0.03	0.83 ± 0.05[a]	0.91 ± 0.14
4CN-Trp*-Gly	H_2O	25	0.73 ± 0.10	0.83 ± 0.04	0.96 ± 0.03	0.84 ± 0.09	0.82 ± 0.05
4CN-Trp*-Gly	THF	25	------	------	------	0.62 ± 0.01	0.56 ± 0.06
4CN-Trp*-Gly	Buffer[b]	37	------	------	------	0.64 ± 0.05	0.74 ± 0.03
4CN-Trp*-Met	H_2O	25	------	------	------	0.75[c]	------
4CN-Trp	H_2O	25	------	------	------	0.88 ± 0.06[a]	------
Gly-Gly-4CN-Trp	H_2O	25	------	------	------	0.89 ± 0.05[a]	------

[a]Measurements made using the gradient method as described in the Methods Section
[b]20 mM phosphate buffer, pH 7.2
[c]QY determined relative to 4CN-Trp*-Gly

Table 4.2: Fluorescence lifetimes (τ) measured for 4CNI, 4CN-Trp, and their derivatives using an excitation wavelength of 270 nm in the indicated solvent at 25 °C.

Sample	Solvent	τ (ns) (%Amp)
4CNI	H_2O	9.1 (100%)
4CNI-3AA	H_2O	13.5 (100%)
4CN-Trp*-Gly	H_2O	12.6 (100%)
4CN-Trp*-Gly	THF	7.6 (100%)
4CN-Trp*-Met	H_2O	12.2 (100%)
4CN-Trp	H_2O	13.7 (95%)
		1.9 (5%)
Gly-Gly-4CN-Trp	H_2O	13.6 (93%)
		1.7 (7%)

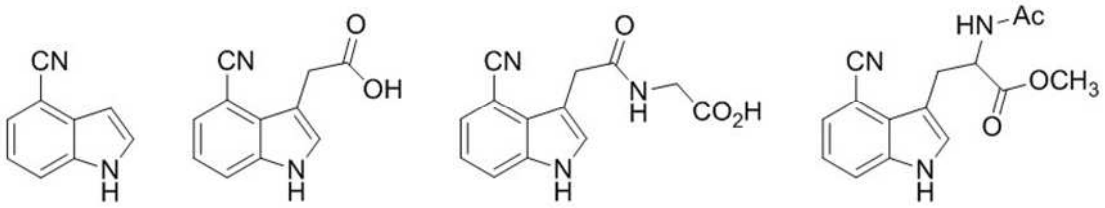

Figure 4.1: Structures of 4CNI, 4CNI-3AA, 4CN-Trp*-Gly, and 4CN-Trp, as indicated.

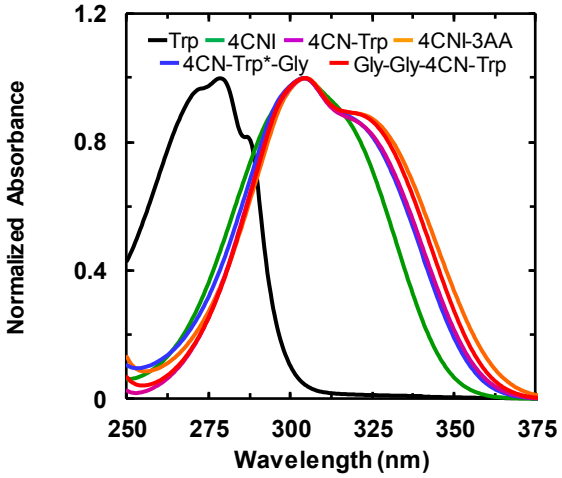

Figure 4.2: Absorption spectra of Trp, 4CNI, 4CNI-3AA, 4CN-Trp, 4CN-Trp*-Gly, and Gly-Gly-4CN-Trp, as indicated.

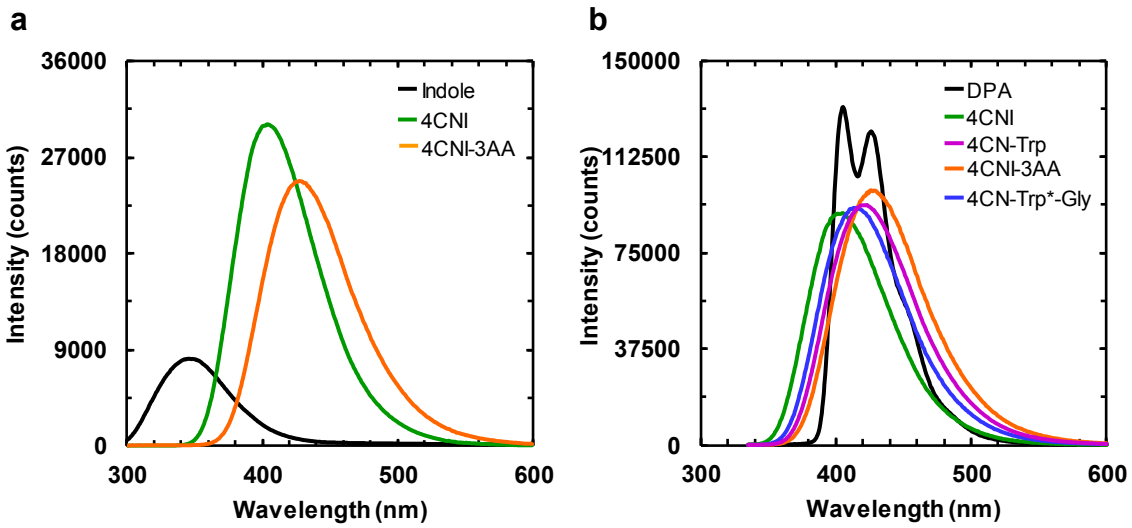

Figure 4.3: (a) Comparison of the fluorescence spectrum of indole with those of 4CNI and 4CNI-3AA. These spectra were obtained with an excitation wavelength of 270 nm at 25 °C and the optical densities of all the samples at 270 nm were equal (0.01). **(b)** Comparison of the fluorescence spectrum of DPA with those of 4CNI, 4CNI-3AA, 4CN-Trp, and 4CN-Trp*-Gly. These spectra were obtained with an excitation wavelength of 325 nm at 25 °C and the optical densities of all the samples at 325 nm were equal (0.01).

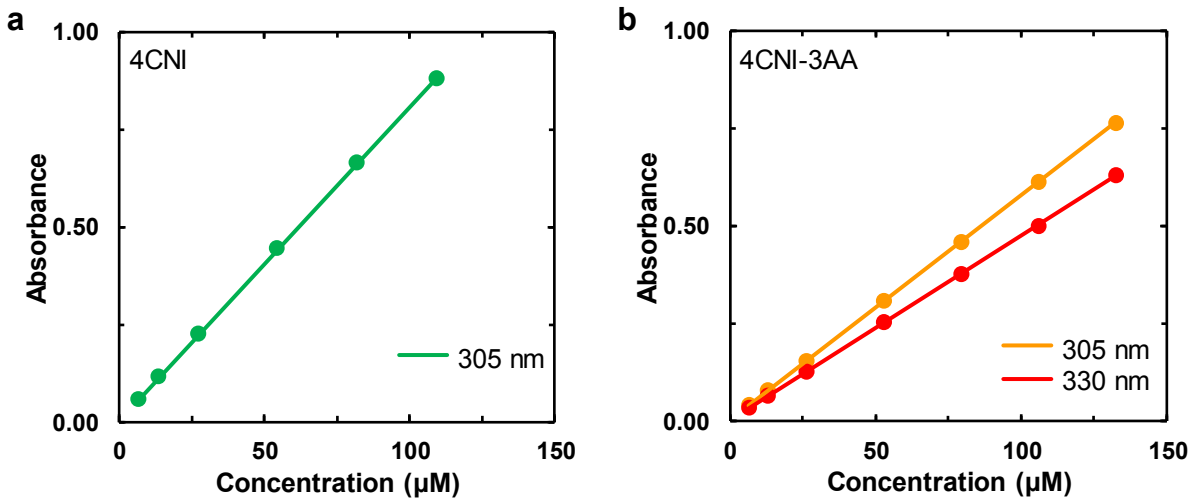

Figure 4.4: Beer's law plots of (**a**) 4CNI and (**b**) 4CNI-3AA at the indicated wavelengths. The straight line in each case corresponds to a linear regression of the respective data with a zero y-intercept, which yielded a molar extinction coefficient of 8056 ± 24 M^{-1} cm^{-1} for 4CNI at 305 nm, and 5778 ± 8 M^{-1} cm^{-1} and 4750 ± 8 M^{-1} cm^{-1} for 4CNI-3AA at 305 nm and 330 nm, respectively.

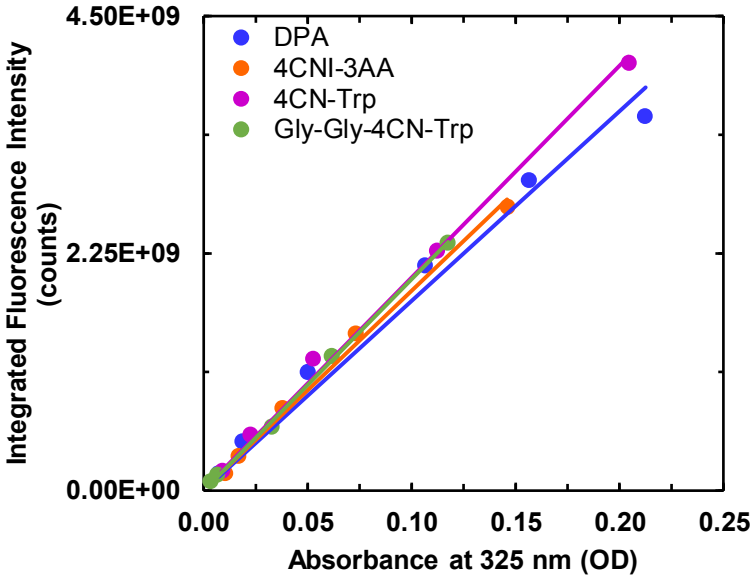

Figure 4.5: Integrated fluorescence intensity of DPA, 4CNI-3AA, 4CN-Trp, and Gly-Gly-4CN-Trp as a function of absorbance at 325 nm. The solid color line in each case corresponds to a linear regression of the respective data with a zero y-intercept, which yielded a slope of (1.80 ± 0.07) $\times 10^{10}$ for DPA, (1.89 ± 0.04) $\times 10^{10}$ for 4CNI-3AA, (2.02 ± 0.04) $\times 10^{10}$ for 4CN-Trp, and (2.00 ± 0.03) $\times 10^{10}$ for Gly-Gly-4CN-Trp. The calculated QYs at 325 nm are shown in Table 4.1.

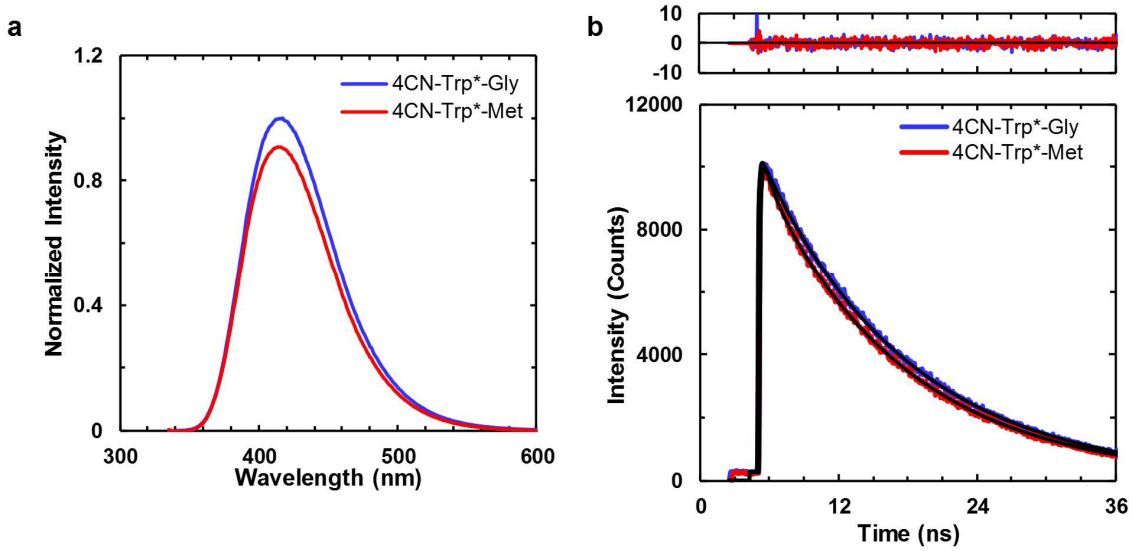

Figure 4.6: (a) Fluorescence spectra (λ_{ex} = 325 nm) and **(b)** fluorescence decay kinetics (λ_{ex} = 270 nm) of 4CN-Trp*-Gly and 4CN-Trp*-Met, as indicated. The optical densities of both samples at λ_{ex} were equal. Fitting each fluorescence decay in (**b**) to a single-exponential function (black) yielded a lifetime of 12.6 ns for 4CN-Trp*-Gly, and 12.2 ns for 4CN-Trp*-Met. The residuals to the fits are shown in the top panel.

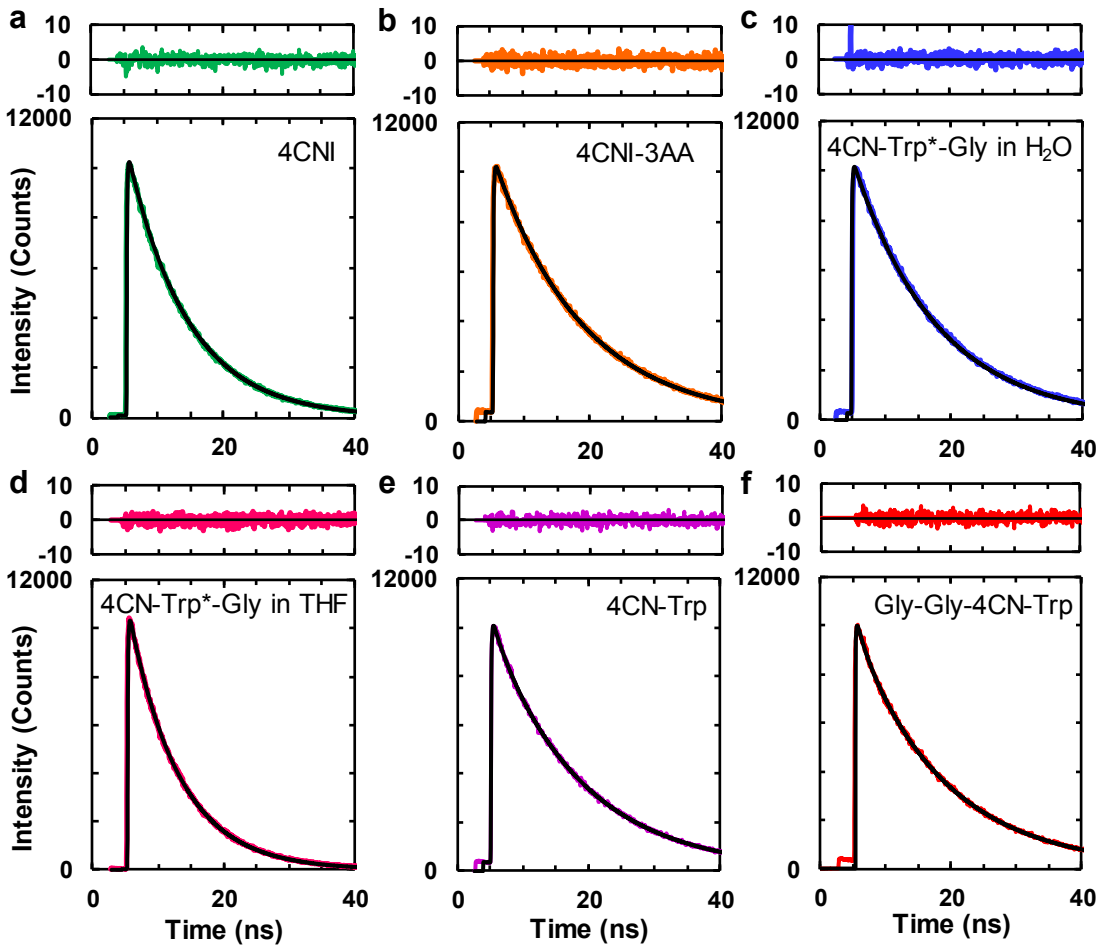

Figure 4.7: Fluorescence decay kinetics of **(a)** 4CNI, **(b)** 4CNI-3AA, **(c)** 4CN-Trp*-Gly **(d)** 4CN-Trp*-Gly in THF, **(e)** 4CN-Trp, and **(f)** Gly-Gly-4CN-Trp, all in H_2O except **(d)**. In each case, the smooth black line corresponds to the best fit of the respective data to a single-exponential (**a**-**d**) or a bi-exponential (**e**-**f**) function and the resultant lifetime(s) is given in Table 4.2. In addition, the top panel in each case is the corresponding residual plot.

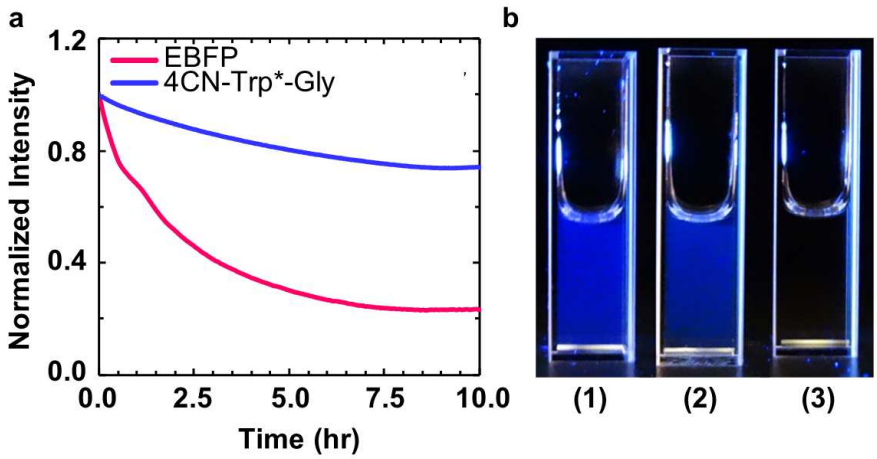

Figure 4.8: (a) Photobleaching kinetics of an EBFP sample (pink) and a 4CN-Trp*-Gly sample (blue). For both cases, the photobleaching light at 355 nm was derived from a Fluorolog 3.10 fluorometer with the excitation slit width set at 5 nm. **(b)** Pictures of a 50 nM 4CN-Trp*-Gly sample **(1)**, a 50 nM EBFP sample **(2)**, and water **(3)** under illumination of 340 nm light, taken using a Sony RX10 camera.

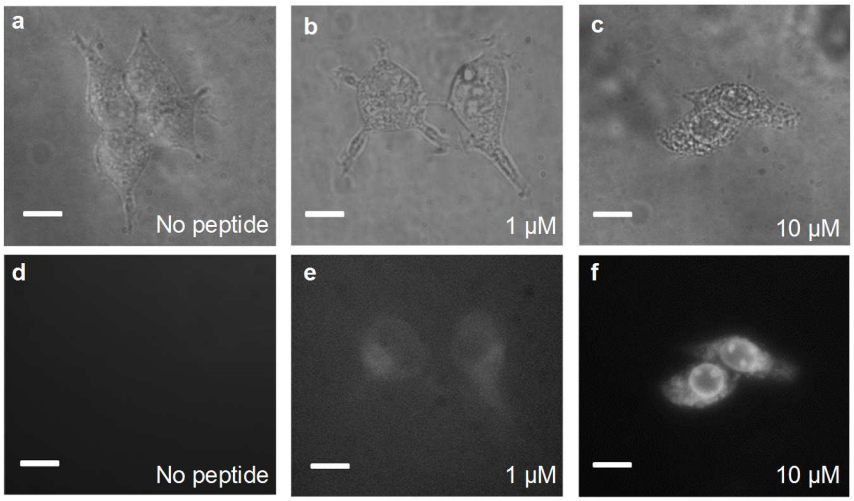

Figure 4.9: Bright-field (**a**, **b**, **c**) and fluorescence (**d**, **e**, **f**) images of HEK293T/17 cells obtained on an Olympus IX71 microscope, in the presence and absence of 4CN-Trp*-MpX at the indicated concentrations. The scale bar in each panel corresponds to 10 μm.

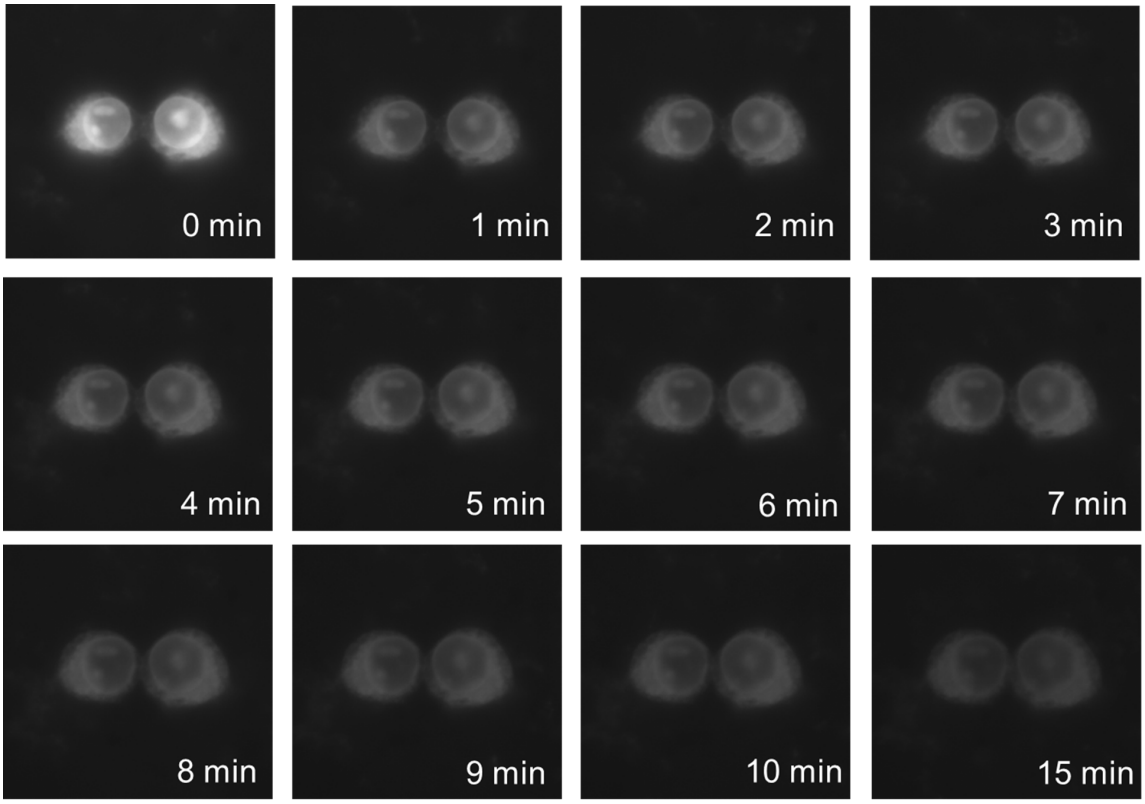

Figure 4.10: Fluorescence microscopic images of HEK293T/17 cells in the presence of 10 µM 4CN-Trp*-MpX obtained at different time points, as indicated. Between each measurement, the sample was continuously illuminated by the excitation light of the microscope. It is noticeable that the fluorescence intensity undergoes a large change between the first and second image. However, it is unclear whether this is caused by photobleaching or peptide redistribution between the solution and cellular phases.

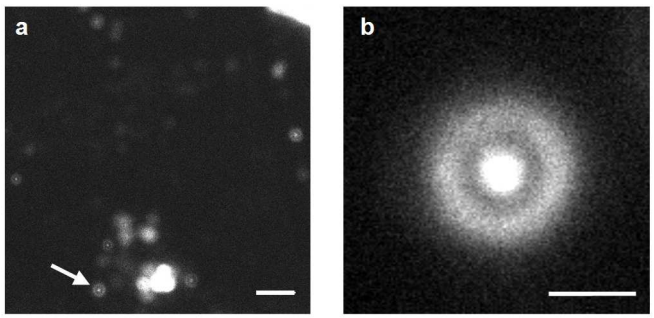

Figure 4.11: Representative confocal microscopic images of HEK293T/17 cells that have been incubated with 1 μM 4CN-Trp*-MpX for ~10 hours, obtained with a scanning step size of 100 nm (**a**) and 50 nm (**b**). The scale bar corresponds to 10 μm in (**a**) and 2 μm in (**b**). Many blebs and other cellular debris can be seen in (**a**), indicating that the HEK cells have died. The higher-resolution image in (**b**) corresponds to the bleb indicated by the white arrow in (**a**).

5: Solvent Dependence of the Fluorescence Lifetimes of Cyanoindoles

5.1: ABSTRACT

The understanding of the photophysical properties of the amino acid tryptophan and its side chain chromophore indole have been extensively studied. Recently, nitrile-derivatized tryptophans (indoles) have been proposed as fluorescence probe alternatives to tryptophan (indole) in the study of protein structure, function, and dynamics. Herein, we systematical study the effect of the nitrile moiety on the absorption and fluorescence properties of indole. Preliminary results suggest that 4-cyanoindole has a higher quantum yield and longer fluorescence lifetime in water due to the removal of a hydrogen bond interaction between the hydroxyl group of the solvent and the π-electron cloud of the aromatic ring that acts as a nonradiative relaxation pathway in other cyanoindoles derivatives.

5.2: INTRODUCTION

As a naturally occurring amino acid, tryptophan (Trp) is arguably the most widely used fluorescence probe in the investigation of protein structure, function, and dynamics. This is mainly due to the fact that, in comparison to other native fluorescent amino acids, Trp has a relatively large extinction coefficient and fluorescence quantum yield (QY) and that its fluorescent properties depend on its environment.[49,58] Despite its wide utility, however, in practice there are several limitations or drawbacks using Trp fluorescence. For example, (1) it often exhibits complex decay kinetics,[54] which complicate data interpretation; (2) its absorption spectrum overlaps with those of other aromatic amino acids, making it difficult to be selectively excited;[54] and (3) its fluorescence spectrum and QY typically do not change in a large extent in response to an environmental change, for instance, from a hydrated to a dehydrated one, resulting in a relatively low signal

contrast.[76] Therefore, many previous efforts have been dedicated to the development of Trp-based amino acid fluorophores that could overcome some or all of those shortcomings.[75–77,178,179,192] In particular, several recent studies have indicated that certain cyanotryptophans (CN-Trp), such as 4CN-Trp,[192] 5CN-Trp,[76] 6CN-Trp,[75] and 7CN-Trp,[75] could afford improved photophysical properties, compared to those of Trp, in the context of making it a better fluorophore in biological studies. However, except in the case of 5CN-Trp, a systematic examination of how the florescence decay kinetics of CN-Trp derivatives varies with environment (i.e., solvent) has not been done. To this end and as a first step towards this direction, we studied the fluorescence decay kinetics of six nitrile-derivatized indole variants (nCNI, where n = 2 – 7 and denotes the carbon position on the indole ring where the CN substitution is located, as shown in Figure 5.1) in a variety of solvents.

We chose these CNI derivatives for the following reasons: (1) unlike their amino acid counterparts, they are commercially available; (2) indole is an intercellular and interspecies bacterial signaling molecule that is involved in a variety of bacterial processes, including antibiotic resistance, biofilm formation, and acid tolerance, among others.[193–195] Therefore, by investigating the photophysical properties of these nitrile-derivatized indoles, we hope to identify a potential indole derivative that does not have the limitations of indole photophysics (i.e., the aforementioned complex decay kinetics, selective excitation, and hydration sensitivity) that can be used to investigate these bacterial processes; and (3) as there are only a few computational studies on one of these CNI derivatives (5CNI),[196–198] a photophysical investigation on a series of nitrile-derivatives of indole can be used to benchmark computational calculations of substituted-indole compounds.

We find that the absorption and emission spectra of the CNI compounds in H_2O are vastly different from indole, in terms of absorption and emission maxima and spectral shape and intensity. Additionally, compared to the other indole derivatives, 4CNI is unique, as its absorption and emission maxima are the most red-shifted from those of indole and its absorption and emission intensities are the largest of the measured compounds. The shifts in the emission maxima of all the CNI derivatives can be explained by the internal stark effect of indole, as the CN moiety is an electron-withdrawing group and pulls electron density away from the benzene or pyrrole ring, depending on the location of its substitution. Finally, we find that the fluorescence lifetimes of 4CNI, 5CNI, 6CNI, and 7CNI measured in solvents that have differing polarity and/or hydrogen bonding (H-bonding) ability is linearly correlated to $\sigma = \pi^* + \beta - \alpha$, where α, β, and π^* are the Kamlet-Taft parameters of the solvent that empirically characterize its H-bond donating ability, its H-bond accepting ability, and its polarizability, respectively. The fluorescence lifetimes of these compounds measured in H_2O, trifluoroethanol, and/or dimethyl sulfoxide are outliers to these linear correlations, likely due to preferential interactions with the indole ring.

5.3: MATERIALS AND METHODS

5.3.1: Materials and Sample Preparation. The compounds indole (Sigma) 2-cyanoindole (2CNI, 97%, Alfa Aesar), 3-cyanoindole (3CNI, 98.6%, Chem-Impex International, Inc.), 4-cyanoindole (4CNI, 97%, Acros Organics), 5-cyanoindole (5CNI, 99%, Acros Organics), 6-cyanoindole (6CNI, 98+%, Alfa Aesar), and 7-cyanoindole (7CNI, 95+%, Ark Pharm Inc.) were used as received. The solvents 1,4-dioxane (Alfa

Aesar), 2-propanol (Alfa Aesar), acetonitrile (ACN, Acros Organics), D_2O (99.9%, Cambridge Isotopes), dimethyl sulfoxide (DMSO, Acros Organics), ethanol (EtOH, Decon Laboratories, Inc.), methanol (MeOH, Acros Organics), tetrahydrofuran (THF, Acros Organics), and trifluoroethanol (TFE, Chem-Impex International Inc.) were of spectroscopic grade and used as received. Solutions of all indole compounds were made by dissolving the respective compound in the desired solvent, with an optical density (OD) at 270 nm between 0.1 – 0.2. Solute concentrations were determined optically using the following extinction coefficients (λ) 15693 ± 30 M^{-1} cm^{-1} (282 nm) for 2CNI, 6614 ± 17 M^{-1} cm^{-1} (276 nm) for 3CNI, 10400 ± 20 M^{-1} cm^{-1} (285 nm) for 6CNI, and 7845 ± 15 M^{-1} cm^{-1} (308 nm) for 7CNI (Figure 5.2). The extinction coefficient for 4CNI (8056 M^{-1} cm^{-1} at 305 nm)[192] and for 5CNI (5500 M^{-1} cm^{-1} at 280 nm)[76] have been reported previously.

5.3.2: Absorption Measurements. UV-Vis absorption spectra were collected on a Jasco V-650 UV-Vis spectrophotometer using a 1.0 cm quartz cuvette at room temperature.

5.3.3: Static and Time-Resolved Fluorescence Measurements. Static fluorescence spectra were collected on a Jobin Yvon Horiba Fluorolog 3.10 fluorometer at 25 °C in a 1.0 cm quartz cuvette at a resolution of 1.0 nm, an integration time of 1 s/nm, with an OD of 0.01 at the excitation wavelength (λ_{ex}) of 270 nm.

Time-resolved measurements were collected on a time-correlated single photon counting (TCSPC) system in a 0.4 cm quartz cuvette at room temperature with a

fluorophore OD at 270 nm in the range of 0.1 – 0.2. The details of the system have been described elsewhere.[76] A Ti:Sapphire oscillator (800 nm, 85 MHz) was used to derive the 270 nm excitation pulse through collinear third harmonic generation while an electro-optical pulse picking system (Conoptics, Inc.) was used to reduce the repetition rate of the pulse to 21 MHz. Selective emission was collected under magic angle polarization using a short band pass 355/45 nm filter (Semrock) and a long band pass 300 nm filter (Semrock) and was detected with a MCP-PMT detector (Hamamatsu R2809U) and a TCSPC board (Becker and Hickl SPC-730). The fluorescence decays were deconvoluted with the experimental instrument response function (IRF) and were fit to single-exponential functions using the FLUOFIT (Picoquant GmbH) program.

5.4: RESULTS AND DISCUSSION

5.4.1: Absorption Spectra. As shown (Figure 5.3), the position on the nitrile group on the indole ring has a distinct effect on the absorption spectra of the nitrile-derivatized indoles in water, as manifested by the differences in spectral shape, position of the absorption maxima (λ_{abs}) and extinction coefficient (ε) (Table 1). In comparison to that of indole, the λ_{abs} of all the nitrile-derivatives are red-shifted, with the smallest shifts (5 – 6 nm) observed for 3CNI and 5CNI, moderate shifts (12 – 15 nm) observed for 2CNI and 6CNI, and largest shifts (>35 nm) observed for 4CNI and 7CNI. Consistent with our results, Juszczak and coworkers observed a red-shift in λ_{abs} for substituted indoles in water for those with electron-withdrawing substituted moieties (i.e., CN, F) while a blue-shift was observed for those with electron-donating moieties (i.e., CH_3, OCH_3, OH).[199] Additionally, the spectral shape and ε of the CNI compounds are quite different, where

those of 3CNI and 5CNI are most similar to those of indole and those of 2CNI, 4CNI, 6CNI, and 7CNI are very different. Interestingly, 4CNI has a similar ε to indole, but is much broader in shape, indicating an increase in the oscillator strength of 4CNI compared to that of indole, assuming the underlying transitions of the molecules are similar. As the radiative rate constant is proportional to the square of the oscillator strength of the transition, 4CNI likely has a higher emission intensity and hence, a higher fluorescence quantum yield, than those of indole based upon these differences in their absorption spectra.

5.4.2: Fluorescence Spectra. As shown (Figure 5.4), the fluorescence spectra of the nitrile-derivatized indoles obtained in H_2O also show that the position of the nitrile group on the indole ring has a significant effect on the fluorescent properties of these fluorophores. First, these spectra reveal that 4CNI has a strikingly high apparent QY in H_2O, comparing to those of other nitrile-derivatized indoles. Second, 2CNI and 3CNI are essentially non-fluorescent when compared to 4CNI, which could be a result of the addition of an electron-withdrawing moiety to the pyrrole ring. Third, the fluorescence intensities of 5CNI, 6CNI, and 7CNI are all similar to one another and are much lower than that of indole. Fourth, the maximum emission wavelengths (λ_{em}) for all fluorophores except 3CNI are red-shifted compared to that of indole (Table 5.1). For example, λ_{em} of 2CNI is red-shifted 5 nm from that of indole, while the others (4CNI, 5CNI, 6CNI, and 7CNI) exhibit shifts between 30 – 60 nm. In the ^1L_a transition of indole, which is the lowest-energy singlet state of indole in polar solvents,[200] there are large decreases in the π-electron density at positions 1 and 3, accompanied by large increases at positions 4

and 7, with an approximate magnitude of 0.33*e* transferred from the pyrrole to the benzene ring.[201,202] Therefore, when the electron-withdrawing nitrile group is placed at positions 4 or 7 of indole, this charge transfer effect is enhanced, consistent with the largest red-shifts of the emission maxima. The emission maxima of 5CNI and 6CNI have a smaller red-shift than those of 4CNI and 7CNI, indicative of a smaller enhancement of the charge transfer. Conversely, the emission maximum of 3CNI is blue-shifted ~25 nm compared to that of indole, meaning that as the nitrile pulls electron density away from the pyrrole ring, the charge transfer effect is diminished. Consistent with our result, a red-shifted emission maximum was observed when an electron donating group was placed at position 3 in the indole ring.[203] Additionally, this effect is similar to what has been coined the internal Stark effect of Trp, where the shift in the emission maxima is determined by the local, average electric field projected onto the long axis of indole.[57,204] In other words, the emission maximum of indole red-shifts with a positive charge at the benzene end or a negative charge at the pyrrole end.

5.4.3: Fluorescence Lifetimes. The fluorescence lifetimes of these nitrile-derivatized indoles, except 5CNI,[76] have not been previously investigated. Therefore, in an attempt to provide a more comprehensive understanding of the effect of nitrile substitution on the photophysical properties of indole, we next measured the fluorescence decay kinetics of 2CNI, 3CNI, 4CNI, 6CNI, and 7CNI in a series of solvents that differ in polarity and/or hydrogen-bonding (H-bonding) ability. As indicated (Figure 5.5), the fluorescence decay kinetics of 2CNI, 3CNI, 4CNI, 6CNI, and 7CNI in H_2O can well described by a single-exponential function, with lifetimes (Table 5.2) consistent with the equilibrium measurements. For example, 4CNI has the longest lifetime of 9.1 ns while the lifetimes

of 2CNI and 3CNI are shorter than the time resolution of our experimental setup, with an IRF on the order of 50 ps. Interestingly, the fluorescence decay kinetics of 6CNI and 7CNI are quite fast as compared to those of indole (4.0 ns),[205] with lifetimes of 0.6 and 2.0 ns, respectively, which suggests that their amino acid forms would not be highly fluorescent in H_2O, similar to the fluorescence decay kinetics of 5CNI in H_2O[76] and in contrast to the study of Talukder et al.[75] who calculated the QY of 6CN-Trp (0.53) and 7CN-Trp (0.40) in MeOH.

The fluorescence lifetimes of the nitrile-derivatized indoles are dependent on solvent (Table 5.2). A cursory inspection of the results indicates that among all the solvents used, H_2O, TFE, and DMSO appear to have a more pronounced effect with 5CNI, 6CNI, and 7CNI exhibiting the longest (shortest) fluorescence lifetime in DMSO (H_2O and TFE). This is an interesting result as the fluorescence lifetime of indole in H_2O (4.0 ns)[205] is similar to that in DMSO (5.4 ns).[206] Traditionally, studies concerning how solvent affects the fluorescence properties of a fluorophore often seek to use an empirical solvent parameter, such as $E_T(30)$ or the Kamlet-Taft paramters (α, β, π^*), to correlate to the observed experimental properties of the fluorophore. Following this practice, we tried to find out whether a simple correlation exists between the fluorescence lifetime and an empirical solvent parameter for those nitrile-derivatized indoles. Recently, Zhang et al.[207] have shown that the nitrile stretching vibrational frequency of 5CNI exhibits a linear dependence on the solvent parameter $\sigma = \pi^* + \beta - \alpha$, where π^*, β, and α are the Kamlet-Taft parameters of the solvent that characterize its polarizability, its H-bond accepting ability, and its H-bond donating ability, respectively. In other words, the nitrile stretching frequency of 5CNI is not only sensitive to direct

interactions to the CN moiety with H-bond donating solvents (i.e., α > 0), but is also sensitive to indirect interactions with H-bond accepting solvents (i.e., β > 0) through its NH moiety and to the polarizability of the solvent (π^*).

Therefore, we looked at the correlation between the parameter σ and the fluorescence lifetime values of 5CNI (Figure 5.6). A linear correlation was observed for all solvents, excluding H_2O. This result substantiates the notion that the fluorescence lifetime of 5CNI is sensitive to direct interactions with its nitrile moiety and indirect interactions with the indole ring. Similarly, a linear correlation between the σ parameter and the fluorescence lifetimes of 6CNI (Figure 5.7), 7CNI (Figure 5.8), and 4CNI (Figure 5.9) was observed when the fluorescence lifetimes measured in TFE (6CNI and 7CNI), DMSO (4CNI and 7CNI), and/or H_2O (4CNI, 6CNI, and 7CNI) were excluded, indicating that these three solvents interact differently with the indole compounds than the other solvents.

As the fluorescence lifetimes of 4CNI and 7CNI measured in DMSO was excluded from the correlation, there is likely a preferential interaction between 4CNI/7CNI and DMSO that causes an increase in the fluorescence lifetimes of these molecules. Compared to all the other solvents, DMSO has an amphiphilic nature, where it can engage in H-bonds with H_2O through its oxygen atom and in hydrophobic interactions through its methyl groups.[208] As a result, in aqueous binary mixtures of DMSO, several studies have suggested that DMSO preferentially solvates hydrophobic residues of proteins, like Trp.[209–211] Likely, the methyl groups of DMSO are engaged in hydrophobic interactions with 4CNI/7CNI, increasing the fluorescence lifetime as differing quenching mechanisms are blocked. Similarly, the fluorescence lifetimes

measured in H_2O were excluded for all four compounds (i.e., 4CNI, 5CNI, 6CNI, and 7CNI), though 4CNI has a long lifetime in H_2O while 5CNI, 6CNI, and 7CNI have very short lifetimes (Table 5.2). Though excited-state proton transfer has been shown to quench the fluorescence lifetime of indole,[212] there appears to be little to no difference in the fluorescence lifetime of the indole derivatives in D_2O compared to those measured in H_2O, suggesting no excited-state proton transfer quenching process (Table 5.2). Markiewicz et al. attributed the fast decay kinetics of 5CNI in H_2O and TFE to a H-bonding interaction between the -OH group of the solvent and the π-electron cloud of the aromatic ring that increases the nonradiative rate of the fluorophore,[76] which is an interpretation that is consistent with the fast kinetics of 6CNI and 7CNI in H_2O and TFE. Conversely, as the distribution of the π-electron cloud is likely different for 4CNI compared to 5CNI, 6CNI, and 7CNI, this type of H-bonding interaction cannot occur, thus resulting in a longer fluorescence lifetime for 4CNI in H_2O and TFE.

5.5: CONCLUSIONS

The photophysical properties of indole and its nitrile-derivatized analogs were investigated using UV absorption and steady-state and time-resolved fluorescence measurements. 4CNI was shown to have the broadest absorption peak with a maximum that was the most red-shifted compared to those of indole. Additionally, in water, 4CNI has the highest QY and the most red-shifted emission maximum, results that are consistent with the charge transfer electronic transition of indole. Preliminary data suggests that the fluorescence lifetime of 4CNI in H_2O is different from those of other CNI compounds because a H-bonding interaction between the hydroxyl moiety of H_2O

and its π-electron cloud of the aromatic ring is not taking place, which decreases its nonradiative rate compared to the other CNI molecules. In order to investigate this series of cyanoindoles compounds further, extensive quantum mechanical calculations are suggested.

5.6: ACKNOWLEDGEMENTS

The authors gratefully acknowledge financial support from the National Institutes of Health (P41-GM104605). M.R.H. is supported by a National Science Foundation Graduate Research Fellowship (DGE-1321851).

This chapter has been adapted from a manuscript in preparation: Mary Rose Hilaire, Feng Gai (2017).

Table 5.1: Absorption maxima (λ_{abs}), extinction coefficient (ε) at λ_{abs}, and emission maxima (λ_{em}) in H_2O for indole and its nitrile derivatives.

Compound	λ_{abs} (nm)	ε (M^{-1} cm^{-1})	λ_{em} (nm)
Indole	270	6000[a]	345
2CNI	282	15693	350
3CNI	276	6614	320
4CNI	305	8056[b]	405
5CNI	275	5500[c]	390
6CNI	285	10400	375
7CNI	308	7845	400

[a]Value obtained from Creed[203]
[b]Value obtained from Hilaire et al.[192]
[c]Value obtained from Markiewicz et al.[76]

Table 5.2: Fluorescence lifetimes (τ) of indole and its nitrile-derivatives in various solvents, as indicated

Solvent	2CNI τ (ns)	3CNI τ (ns)	4CNI τ (ns)	5CNI τ (ns)[a]	6CNI τ (ns)	7CNI τ (ns)
THF	----	----	5.3	4.4	3.7	6.9
1,4-Dioxane	----	----	5.3	4.2	3.9	7.5
ACN	0.9	1.8	5.4	4.4	3.8	8.0
2-Propanol	----	----	5.7	4.6	3.7	9.0
EtOH	----	----	5.7	4.3	3.7	8.7
MeOH	----	----	5.8	2.9	3.5	8.2
TFE	----	----	6.3	0.1	0.8	1.2
DMSO	0.4	0.2	7.4	7.1	4.3	14.5
H_2O	< 0.05[b]	< 0.05[b]	9.1	0.3	0.6	2.0
D_2O	----	----	9.3	0.4	0.8	2.3

[a]Fluorescence lifetimes of 5CNI in all solvents, except 2-propanol, were obtained from Markiewicz et al.[76] The reported values for 5CNI in H_2O, D_2O, and TFE are amplitude-averaged lifetimes based upon bi-exponential kinetics.
[b]Fluorescence decay kinetics were on the same timescale as the IRF (~50 ps).

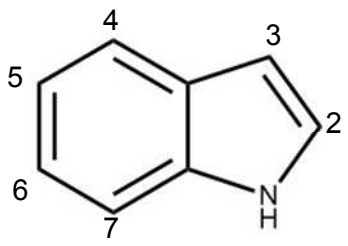

Figure 5.1: Structure of indole, with the carbon numbering scheme labeled.

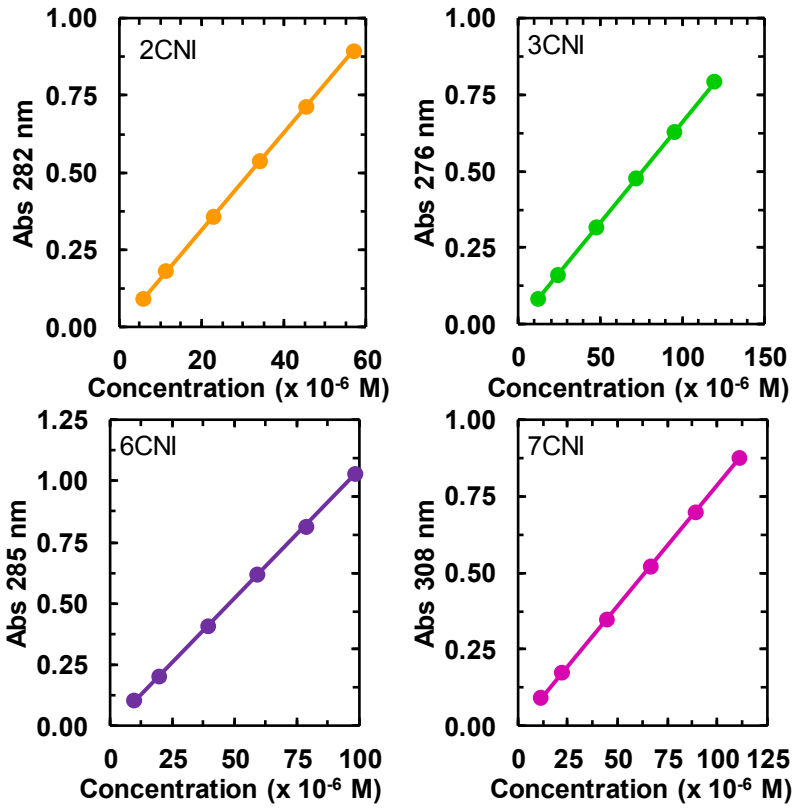

Figure 5.2: Beer's Law Plots of 2CNI, 3CNI, 6CNI, and 7CNI, as indicated. The smooth line in each case is a linear regression with a zero y-intercept with a slope equal to 15693 ± 30 M^{-1} cm^{-1} for 2CNI, 6614 ± 17 M^{-1} cm^{-1} for 3CNI, 10400 ± 20 M^{-1} cm^{-1} for 6CNI, and 7845 ± 15 M^{-1} cm^{-1} for 7CNI.

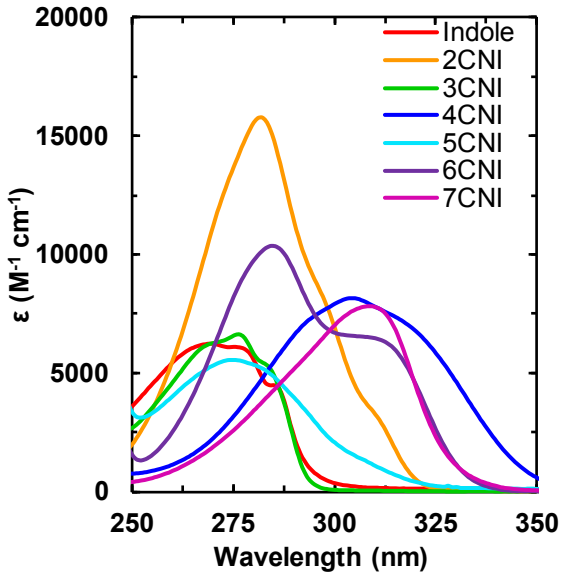

Figure 5.3: Absorption spectra of indole and its nitrile derivatives in H_2O, as indicated.

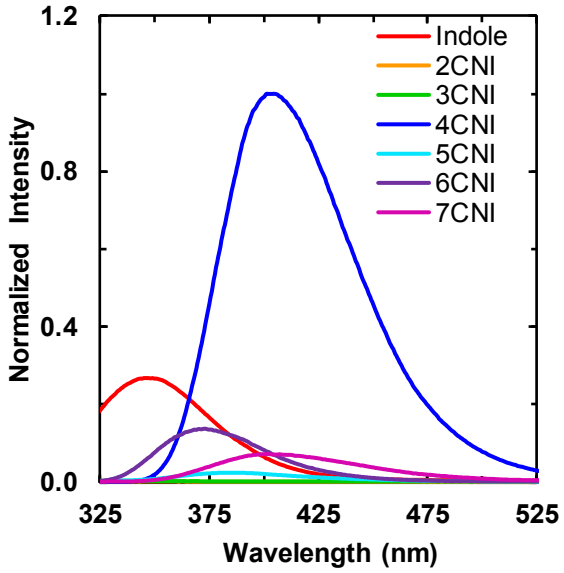

Figure 5.4: Emission spectra (λ_{ex} = 270 nm) of indole and its nitrile derivatives in H_2O, as indicated. The emission intensity was normalized with respect to that of 4CNI.

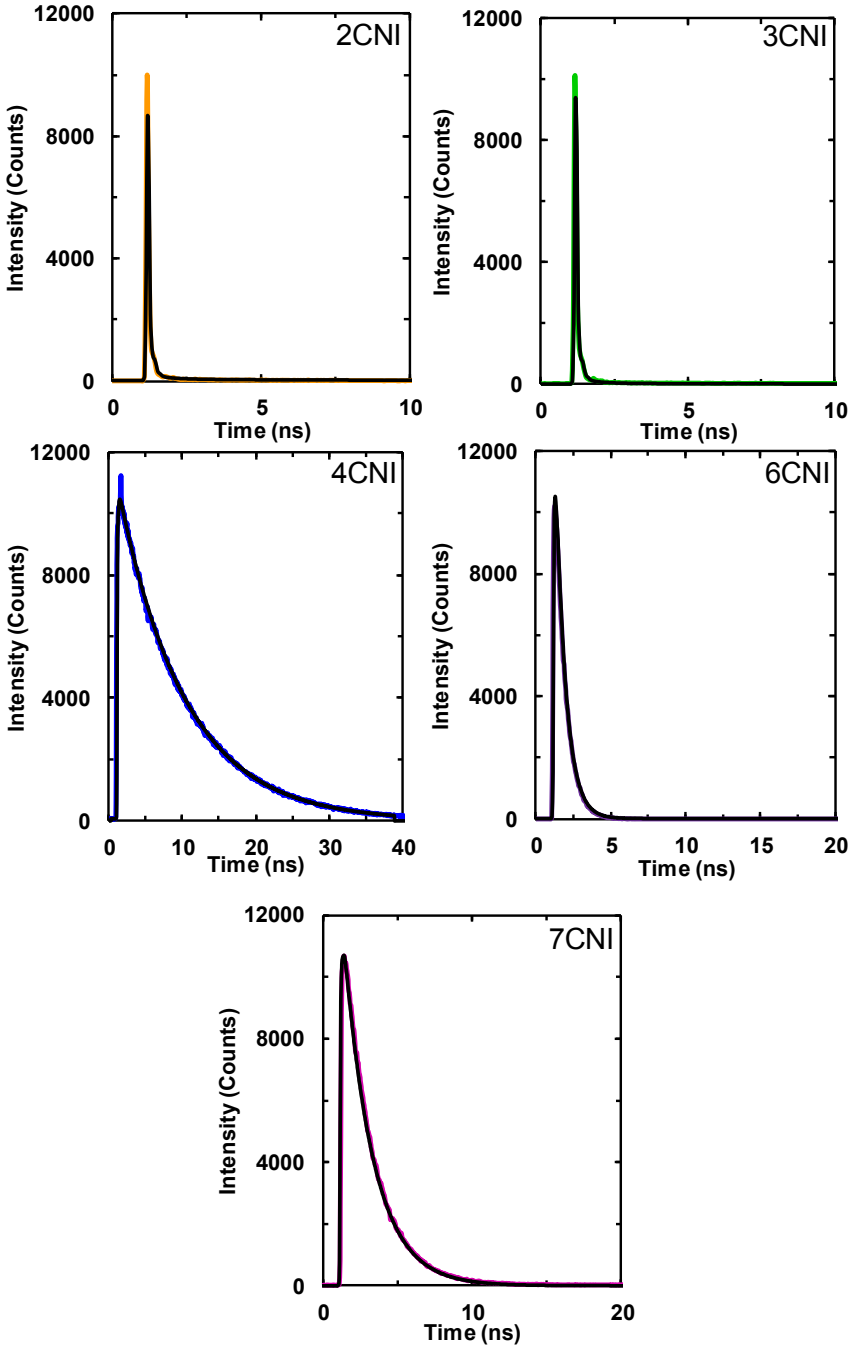

Figure 5.5: Fluorescence decay kinetics of 2CNI, 3CNI, 4CNI, 6CNI, and 7CNI in water, as indicated. In each panel, the black line represents a single-exponential fit to the data (colored line), with the corresponding lifetimes shown in Table 5.2.

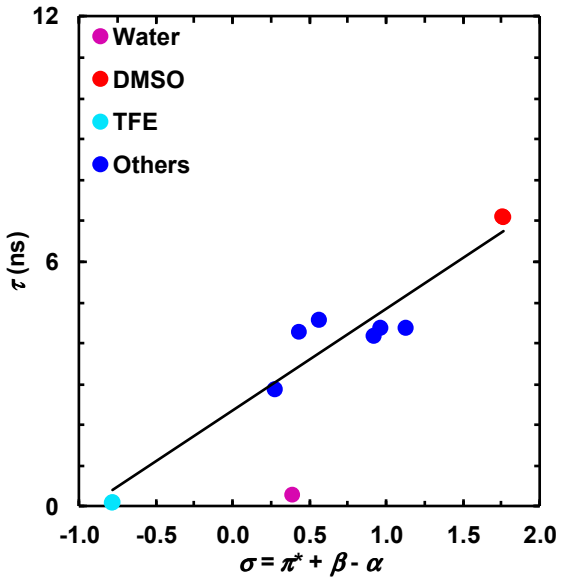

Figure 5.6: Dependence of 5CNI fluorescence lifetime on the empirical solvent parameter σ. The black line corresponds to a linear fit of the data, excluding that obtained in H_2O.

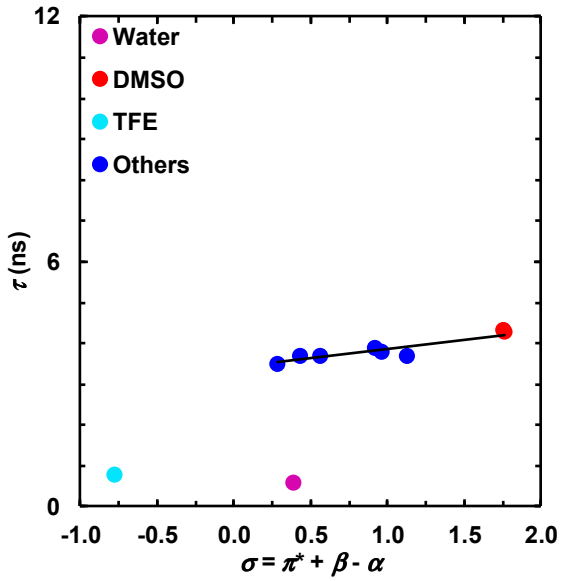

Figure 5.7: Dependence of 6CNI fluorescence lifetime on the empirical solvent parameter σ. The black line corresponds to a linear fit of the data, excluding those obtained in H_2O and TFE.

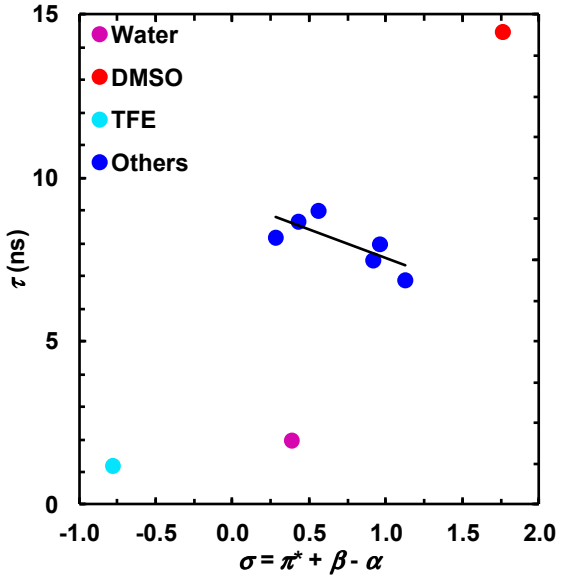

Figure 5.8 Dependence of 7CNI fluorescence lifetime on the empirical solvent parameter σ. The black line corresponds to a linear fit of the data, excluding those obtained in H_2O, TFE, and DMSO.

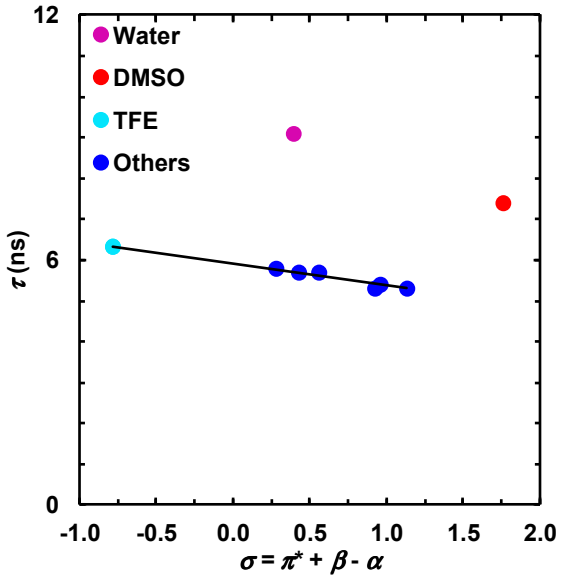

Figure 5.9: Dependence of 4CNI fluorescence lifetime on the empirical solvent parameter σ. The black line corresponds to a linear fit of the data, excluding those obtained in H_2O and DMSO.

6: Infrared Assessment of the Hydration and Electrostatic Environment of the Hydrophobic Core in Staphylococcal Nuclease

6.1: ABSTRACT

The electrostatic environment of the protein interior is involved in many biological processes, including enzymatic catalysis and redox reactions. Thus, much work has focused on quantitatively describing the hydration and electrostatics of hydrophobic cavities in proteins. Herein, we utilized linear infrared spectroscopic methods to investigate the hydrophobic core of staphylococcal nuclease. Through post-translational modifications, we site-specifically incorporate two infrared probes, a thiocyanate and an ester carbonyl moiety into the core of the protein, whose stretching frequencies then reported on the hydration and electrostatics in these environments. Water was found to penetrate at position L38 in the protein, while the V23 and V66 positions were found to be dehydrated and well-ordered. Additionally, the dielectric constant at these positions was estimated to be 15, the first experimentally-determined value for a dielectric constant of a protein interior. This value will help validate and develop predictive methods for electrostatic measurements in proteins.

6.2: INTRODUCTION

The electrostatics in the hydrophobic interior of proteins play a crucial role in many biophysical and biochemical processes, including electron and proton transfer, enzymatic catalysis, and redox reactions.[125,213–216] Thus, numerous experimental[217–220] and computational[221–224] studies have been carried out in the past, aiming to provide a quantitative and predictive understanding of this important property of proteins. For

example, the previous studies by García-Moreno and coworkers indicated that ionizable residues (e.g., Glu and Lys) in the hydrophobic cavities of staphylococcal nuclease (SNase) titrate with anomalous pK_a values when compared to their bulk solvent values, suggesting that those hydrophobic cavities have a higher than expected polarizability.[217,218] In other words, in order to reproduce these pK_a values using structure-based calculations, the protein interior needs to be treated with an effective dielectric constant in the range of 6-10, which is much higher than that (i.e., 2-4) measured with dry protein powders. This unusual ability of proteins to tolerate a polar residue in their hydrophobic interior could arise from the presence of water molecules near the ionizable groups[225–228] or local structural rearrangements upon ioniziation.[229–231] Thus, systematic crystallographic studies of on a wide range of SNase mutants (i.e., V66Q, V66N, V66Y, V66E, V66D, V66K, I92E, I92K) have been carried out, aiming to determine the molecular origin of this phenomenon by investigating two sites in the hydrophobic cavity of SNase that differ in the packing microenvironment surrounding each position.[232] It was found that at cryogenic temperatures (100 K) internal water molecules were present in all of those mutants except V66K and I92K, although at room temperatures (298 K) internal water was only detected in the crystals of V66E and V66Y.[233] However, internal waters could still exist in other mutants; but they are disordered, hence evading detection by crystallography. Furthermore, in solution water molecules can transiently penetrate into a protein's hydrophobic interior[234–236], a scenario that is even more difficult to assess.

Herein, we use infrared (IR) spectroscopy and two IR reporters, the C≡N (nitrile) stretching vibration of a thiocyanate[20,237] (-SCN) group and the C=O stretching vibration of an ester carbonyl[114] (-CH_2COOCH_3, hereafter referred to as -D_M), both of which can

be site-specifically incorporated into proteins, to directly probe whether or not water is present in the hydrophobic cavities of an acid-stable variant, Δ+PHS, of the SNase protein. In addition, another important goal is to experimentally measure the dielectric constant in the hydrophobic core of this protein, as only the apparent dielectric constants in varying locations of SNase have been reported.[217,218] Both -SCN and -D_M are established IR probes of protein local hydration and electrostatic status as their respective nitrile and carbonyl stretching vibrational bands are sensitive to various specific interactions, such as hydrogen-bonding, and nonspecific interactions.[20,114,237,238] For example, upon going from a hydrophobic to an aqueous environment, both bands become much broader and shift their frequencies (to a higher/lower wavenumber for the nitrile/carbonyl band). Thus, these moieties are ideal probes of protein internal waters. Additionally, both the -SCN and -D_M moieties can be post-translationally incorporated into proteins via chemical modification (i.e., cyanylation[117,239] or ligation[113]) of a cysteine (Cys) residue. Moreover, as SNase does not contain any natural Cys residues, the probes can be site-specifically incorporated by simply mutating the amino acid positions of interest in the protein sequence to a Cys. We chose the mutations, V23C, L38C, and V66C because they are well-studied positions, with both V23 and V66 part of the hydrophobic core of SNase,[240] and L38 located at the bottom of the nucleotide binding pocket,[67] a site that is less buried.[241] Thus, the IR-probe-labeled variants of these mutants will allow us to site-specifically monitor the hydration and electrostatic environments in both the hydrophobic core and binding pocket of Δ+PHS.

6.3: MATERIALS AND METHODS

6.3.1: Protein Samples. Three mutants of Δ+PHS with single-point mutations V23C, L38C, and V66C were used as received from the Garcia-Moreno Lab at Johns Hopkins University.

6.3.2: Preparation of SCN-labeled Protein. The cyanylation reaction was modified from previous work.[117] Dry protein was first dissolved in Tris buffer (20 mM, pH 8.0) with 6.0 M guanidine hydrochloride (GuHCl) and 3x molar excess of dithiothreitol (DTT). The protein was then denatured at 37°C for 40 minutes before any reagents were added. After denaturing, the –SH sidechain of Cys was activated using Ellman's reagent (5,5'-dithiobis-(2-nitrobenzoic acid)) (40 mM in same Tris buffer) in 4x molar excess of the total thiol amount. After 30 minutes, $K^{13}C^{15}N$ (100 mM in 1 M Tris buffer, pH 7.1) was added to produce the $-S^{13}C^{15}N$ sidechain. The reaction was left to proceed for twenty minutes before being quenched with acetic acid. The protein solution was then purified using reverse phase HPLC with a Vydac C4 column. After lyophilization, the protein samples were exchanged using 0.01 M DCl in D_2O in three rounds of lyophilization to remove residual trifluoroacetic acid (TFA) that was present in the samples. The pD of the samples was then adjusted to neutral pD using 0.1 M NaOD.

6.3.3: Preparation of SDM-labeled Protein. Lyophilized protein was first dissolved in cacodylate buffer (50 mM, pH 5.0) with 6.0 M urea. Then, 100 mM methyl bromoacetate (MBA) in acetonitrile (ACN) was added to the protein solution for a final MBA concentration of 10 mM and a final protein concentration of 1.0 mM. The reaction was allowed to proceed for three hours at 37°C. The reaction was then immediately purified

using reverse phase HPLC with a Vydac C4 column. After lyophilization, the protein samples were exchanged using 0.01 M DCl in D_2O in three rounds of lyophilization to remove residual TFA that was present in the samples. The pD of the samples was then adjusted to neutral pD using 0.1 M NaOD.

6.3.4: Circular Dichroism (CD) Measurements. CD spectra were obtained with an Aviv 62 DS spectrometer (Lakewood, NJ) using a 1.0 mm quartz cuvette at a protein concentration of ~15 μM in 20 mM cacodylate buffer in H_2O, pH 7.2 for all measurements.

6.3.5: Fourier Transform Infrared (FTIR) Spectroscopic Measurements. FTIR spectra were recorded using a Thermo Nicolet Magna 860 FTIR with an MCT detector at a spectral resolution of 1 cm^{-1}. A homebuilt IR cell consisting of two CaF_2 windows and a 50 μm (or a 75 μm for the temperature-dependent measurements) Teflon spacer divided into two compartments where one side was used for the reference buffer and the other for the protein sample, was used to collect single beam scans of both the reference and the sample with a translation stage that moved each side in and out of the IR light. A total of 512 single beams scans were collected and averaged. The final sample absorbance was calculated by taking the log of the ratio of the reference intensity to the sample intensity. For the temperature-dependent measurements, the temperature was controlled using a Haake K20 temperature controller. The temperature was ramped from ~5 – 50 °C in ~7 °C increments. The sample was allowed to equilibrate at the desired temperature for ten minutes before the FTIR measurement was collected. All other FTIR

measurements occurred at room temperature. For all measurements, the concentration of protein was between 3-5 mM in 20 mM cacodylate buffer in D_2O, pD 7.2.

6.3.6: FTIR Fitting Protocol. Peak frequencies and widths of SCN-labeled samples were determined by a least-squares minimization using the Excel Solver program to a Pseudo-Voigt function, i.e., a linear combination of Lorentzian and Gaussian functions that have the same width, defined as:

$$Abs(\tilde{\nu}) = Abs_0 + A\left(\left(m\frac{1}{2\pi}\frac{FWHM}{(\tilde{\nu}-\tilde{\nu}_0)^2+\left(\frac{FWHM}{2}\right)^2}\right)+(1-m)\left(\frac{2\sqrt{\ln 2}}{\sqrt{\pi}FWHM}e^{-\frac{4\ln 2(\tilde{\nu}-\tilde{\nu}_0)^2}{FWHM^2}}\right)\right) \quad (6.1)$$

where $Abs(\tilde{\nu})$ is the IR absorbance as a function of wavenumber $\tilde{\nu}$, Abs_0 is the offset from zero, A is the amplitude of the peak, m is a variable between 0 and 1 representing the percentage of the Lorentzian function, and $\tilde{\nu}_0$ and FWHM are the central frequency and the full width at half maximum of the IR peak, a method that has been previously used for SCN IR peaks.[237] Similarly, peak frequencies and widths of SD_M-labeled samples were determined by a least-squares minimization using the Solver program in Excel to a Gaussian function of the following form:

$$Abs(\tilde{\nu}) = Abs_0 + A\left(\frac{2\sqrt{\ln 2}}{\sqrt{\pi}FWHM}e^{-\frac{4\ln 2(\tilde{\nu}-\tilde{\nu}_0)^2}{FWHM^2}}\right) \quad (6.2)$$

where the variables are the same as defined above.

6.3.7: Calculation of Protein Cavity Dielectric Constant. The Onsager field ($F_{Onsager}$) equation,[242] given as:

$$F_{Onsager} = \frac{\mu_0}{4\pi\varepsilon_0 a^3}\left[\frac{2(\varepsilon-1)(n^2+2)}{3(n^2+2\varepsilon)}\right] \tag{6.3}$$

where μ_0 is the gas-phase permanent dipole moment of the solute methyl acetate, n is the index of refraction of methyl acetate, and a is the Onsager spherical cavity radius, determined from the molecular weight and density of methyl acetate with values of μ_0 = 1.9536 D, n = 1.361, and a^3 = 132 Å^3, and the C=O stretching frequency of -D_M dependence on the Onsager field:[114]

$$\upsilon_{C=O}\left(cm^{-1}\right) = 4.66F_{Onsager} + 1766.3 \tag{6.4}$$

which holds true in the absence of hydrogen bonds, were used to calculate the dielectric constant, ε, inside the protein cavity.

6.4: RESULTS AND DISSCUSSION

6.4.1: Circular Dichroism (CD) Measurements. In order to confirm that the introduction of either the -SCN or - D_M probe did not significantly perturb the overall structure of the protein, far-UV CD spectra of the SNase variants (hereafter referred to as V23C-

$S^{13}C^{15}N$, L38C-$S^{13}C^{15}N$, V66C-$S^{13}C^{15}N$, V23C-SD_M, and V66C-SD_M, respectively) were collected. As shown (Figure 6.1), the distinctive fold of the unlabeled protein samples is preserved after the cyanylation or ligation. We chose to prepare $S^{13}C^{15}N$-labeled samples, as opposed to $S^{12}C^{14}N$-labeled samples, to red-shift the –SCN frequency to an infrared region with a lower solvent background.[20]

6.4.2: SCN Frequency as Probe of Hydration. As shown (Figure 6.2 and Table 6.1), the nitrile stretching bands of the three cyanylated mutants locate at about 2080 cm^{-1}, consistent with a previous study.[20] It is immediately clear that the nitrile band of L38C-$S^{13}C^{15}N$ is blue-shifted from and much broader than those of V66C-$S^{13}C^{15}N$ and V23C-$S^{13}C^{15}N$. On the other hand, it is more similar to the nitrile band of a short peptide (YGGC*GG, where C* represents C-$S^{13}C^{15}N$), wherein the nitrile IR probe is expected to be fully hydrated. Thus, taken together, these results suggest that the -SCN group in L38C-$S^{13}C^{15}N$ is situated in an environment that is readily accessible by water molecules. To further verify this notion, we measured the temperature dependence of the nitrile stretching frequency of L38C-$S^{13}C^{15}N$. Romesberg and coworkers[243] have shown that when a -SCN group is hydrated and engaged in hydrogen-bonding interactions with water, the center frequency of the nitrile band will exhibit a linear dependence on temperature with a slope of -38.4×10^{-3} $cm^{-1}/°C$ (measured using the model compound CH_3SCN in water). As shown (Figure 6.3 and Table 6.1), the nitrile frequency of L38C-$S^{13}C^{15}N$ has a similar temperature dependence with a slope of -32.4 $\pm$ 0.8 $\times 10^{-3}$ $cm^{-1}/°C$, thus providing additional evidence to support our conclusion. Furthermore, our finding is entirely consistent with the studies of Zhong and coworkers,[67,244] which used tryptophan (Trp) fluorescence of a L38W mutant to show

that internal water is present at this location. Additionally, our result is also consistent with the study of Harms *et al.*,[241] which found that a Lys residue at this location (L38K) titrates with a normal pK_a value due to the presence of water molecules.

On the contrary, the nitrile stretching band of V66C-S^{13}C^{15}N is further red-shifted from that of the control peptide and becomes very narrow (Figure 6.2 and Table 6.1). The exceedingly narrow bandwidth (~3.2 cm^{-1}) of this vibrational transition provides strong evidence that the nitrile probe in this case is situated in a completely dehydrated environment as the presence of any water would lead to fluctuations in the local electrostatic field and hence a broadening of the vibrational band.[245] This assessment is consistent with the study of Boxer and co-workers,[238,246] which showed that when a -S^{13}C^{15}N probe was placed in the interior of the ketosteroid isomerase protein (i.e., M105C-S^{13}C^{15}N), where the probe was away from any hydrogen bonding sources, i.e., solvent or side chains, a similar narrow nitrile band (~3.5 cm^{-1}) was developed. Further measurement on the temperature dependence of the nitrile stretching frequency of V66C-S^{13}C^{15}N also corroborates the above conclusion as the slope of the corresponding linear temperature dependence is -8.4 × 10^{-3}cm^{-1}/°C (Figure 6.3 and Table 6.1), a value similar to –SCN incorporated into the murine adaptor protein Crk-II (M181C-S^{12}C^{14}N), where the –SCN moiety is adjacent to a hydrophobic pocket consisting of Phe and Trp residues. A Trp residue incorporated at this location exhibits a fluorescence emission maximum that is consistent with deep burial in the hydrophobic core,[244] which supports our conclusions.

Interestingly, while the nitrile band of V23C-S^{13}C^{15}N also exhibits a narrow bandwidth (~4.6 cm^{-1}), its center frequency is blue-shifted from that of V66C-S^{13}C^{15}N (Figure 6.2 and Table 6.1). This blue-shift could arise from two different scenarios: 1) the

-SCN group in V66C-$S^{13}C^{15}N$ is situated in a more polar environment compared to that of V23C-$S^{13}C^{15}N$, as the nitrile frequency of -SCN is known to shift to the red with increasing solvent polarity (for aprotic solvents);[237] and 2) the nitrile group in V23C-$S^{13}C^{15}N$ is hydrogen-bonded to either an internal water or a nearby side chain. To differentiate these two possibilities, we carried out temperature dependent studies. As indicated (Figure 6.3 and Table 6.1), the nitrile stretching frequency of V23C-$S^{13}C^{15}N$ shows a slightly larger change with temperature than that of V66C-$S^{13}C^{15}N$, with a slope of -12.8 ± 0.6 x 10^{-3} cm^{-1}/°C. The latter is comparable to that of the value obtained for CH_3SCN in isopropyl alcohol (-15.1 x 10^{-3} cm^{-1}/°C), suggesting that the aforementioned blue-shift is caused by a relatively weak hydrogen bond (i.e., weaker than that with water) formed between the nitrile group and a nearby side chain. Based on the crystal structure of ∆+PHS (pdb 3bdc), the most likely candidate of hydrogen bonding donor is the –OH moiety of a nearby threonine residue at position 62. The results obtained with the –$S^{13}C^{15}N$ probe provide clear evidence that there is no internal water near V23 and V66.

6.4.3: SD_M Probe of Electrostatics. To further verify this notion and, perhaps more importantly, to determine the dielectric constant in the hydrophobic core of SNase, we carried out similar studies using a different IR probe, the C=O stretching vibration an ester (i.e., D_M). In comparison to the nitrile stretching vibration of –SCN, the frequency of this carbonyl probe is easier to interpreted as it exhibits a linear dependence on the local electrostatic field.[114] In addition, Pazos *et al*.[114] have shown that when situated in a non-hydrogen-bonding environment, the C=O stretching frequency of -D_M exhibits a linear dependence on the Onsager field ($F_{Onsager}$), given in Equation (5.4). Since $F_{Onsager}$

depends on the dielectric constant, ε, of the local environment (or solvent), as shown in Equation (5.3), this frequency-field relationship of -D_M makes it possible to use its C=O stretching frequency to experimentally characterize ε in the interior of a protein.[114]

As shown (Figure 6.4), the peak frequencies of V23C-SD_M and V66C-SD_M are nearly identical, at ~1742.9 cm^{-1}. According to the study of Pazos *et al*.,[114] which showed that hydrogen bonds shift the ester carbonyl frequency to 1700-1730 cm^{-1}, this frequency indicates that in both mutants the ester carbonyl of –SD_M is not solvated by internal water, thus further corroborating the notion that both V23 and V66 are situated in a dehydrated environment. Additionally, it does not seem that V23C-SD_M is interacting with the Thr residue, as this hydrogen bond should shift the frequency. Likely, the orientation of this side chain compared to that of –SCN does not allow for the formation of a hydrogen bond. Furthermore, the bandwidth of the V23C-SD_M mutant (~7.7 cm^{-1}) is narrower than the bandwidth of V66C-SD_M (~10.6 cm^{-1}); while both are quite narrow, as compared to the FWHMs measured of a model compound methyl acetate in various solvents,[114] this is the opposite trend than was observed with the –$S^{13}C^{15}N$ mutants, indicating that the packing environment around the –$S^{13}C^{15}N$ and –SD_M side chains is likely different. Nonetheless, the bandwidths of the peaks are consistent with a dehydrated, structured environment.

Finally, using the frequency value of 1742.9 cm^{-1} and Equations (5.3) and (5.4), we calculated the dielectric constant in the hydrophobic interior of the SNase at positions 23 and 66 to be 15.1, using parameters for the model compound methyl acetate. This value is higher than most common nonpolar solvents and is similar to the dielectric constants of bulky alcohols (i.e., tert-butanol (12.4), benzyl alcohol (13.1), and isopropyl

alcohol (17.9)), which is surprising given the makeup of the nonpolar residues lining the cavity. However, ε = 15.1 is only an estimation, as Equation (5.4) does not account for the sulfur atom that is also present on the side chain. Likely, the actual dielectric constant is lower, as the sulfur shields the field surrounding the probe. Nonetheless, this dielectric constant value is consistent with the literature, as Goh et al.[247] calculated the dielectric constant of Δ+PHS and several of its variants to be in the range of 20-30, with a value of ~10 for the microenvironment surrounding position 66. Additionally, Mellor *et al*.[219] estimated the protein core of the protein β-lactoglobulin to be in the range of 6-7 through dielectric spectroscopy and molecular modeling and calculations. It should be noted, however, that this value assumes that the permittivity of the protein core is the same throughout the protein and therefore, should be viewed as an average dielectric constant of the protein core. In other words, there would be locations within the core that had lower and higher dielectric constants, with an average value in the range of 6-7. Finally, our value of 15.1 is higher than the reference dielectric constant that is used in simulation for the protein interior.[248,249] We believe this result can act as a benchmark for simulations in order to predict the electrostatic environment in the interior of a protein.

6.5: CONCLUSIONS

A protein's electrostatic environment, particularly that of its hydrophobic interior, plays an important role in many biochemical and biophysical processes. In order to describe this environment, we utilized two infrared probes, the nitrile stretch of the -SCN moiety and the carbonyl stretch of an ester carbonyl -SD_M, to describe the electrostatics around three sites in mutants of the SNase protein: V23, L38, and V66. Using the -SCN

probe, the microenvironment around the V23 and V66 sites was found to be consistent with an ordered, dehydrated environment. On the other hand, the microenvironment of L38 was partially-hydrated. The IR frequencies of V23C-SD_M and V66C-SD_M collaborated the notion that V23 and V66 were dehydrated. Using a frequency-field relationship, Onsager theory, and the IR frequency of V66C-SD_M, the dielectric constant at V23 and V66 was estimated as 15.1, a value that is likely higher than the actual dielectric constant, as the sulfur atom in the side chain would shield the electrostatic field. Nonetheless, this value, to our knowledge, is the first experimental determination of the dielectric constant in the protein's hydrophobic core, which is crucial to the understanding of protein electrostatic environments and their effect on protein biological function.

6.6: ACKNOWLEDGEMENTS

The authors gratefully acknowledge financial support from the National Institutes of Health (P41-GM104605). M.R.H. is supported by a National Science Foundation Graduate Research Fellowship (DGE-1321851).

This chapter has been adapted from a manuscript in preparation: Mary Rose Hilaire, Aaron C. Robinson, Bertrand García-Moreno E., Feng Gai (2017).

Table 6.1: Characteristics of the Nitrile Band: Peak Frequency (ν_o), Bandwidth (FWHM), and Temperature Dependence of ν_o (Slope)

Protein	ν_o (cm^{-1})	FWHM (cm^{-1})	Slope ($\times 10^{-3}$ cm^{-1}/°C)
V23C-$S^{13}C^{15}N$	2081.1	4.6	-12.8 ± 0.6
L38C-$S^{13}C^{15}N$	2082.3	9.3	-32.4 ± 0.8
V66C-$S^{13}C^{15}N$	2078.1	3.2	-8.4 ± 0.8
YGGC*GG	2084.5	10.8	-

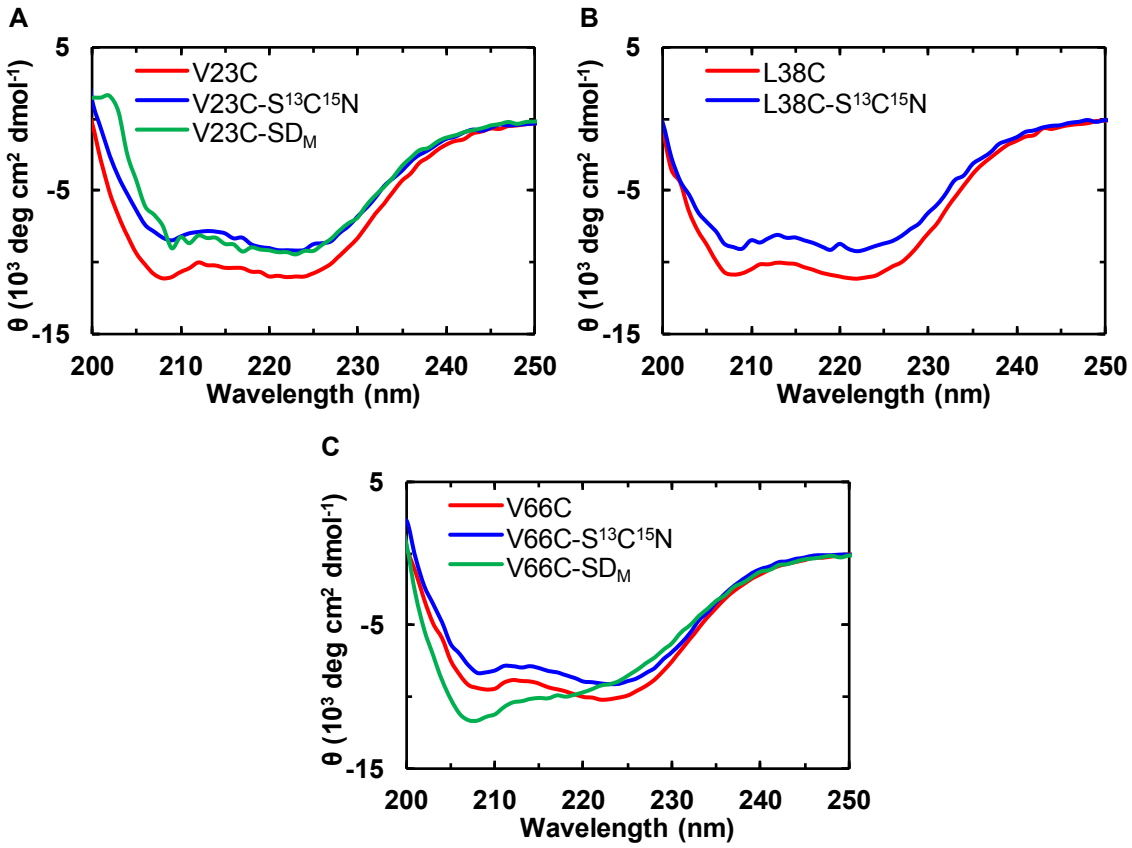

Figure 6.1: Far UV CD Spectra of the V23C variants (A), L38C variants (B), and V66C variants (C), as indicated.

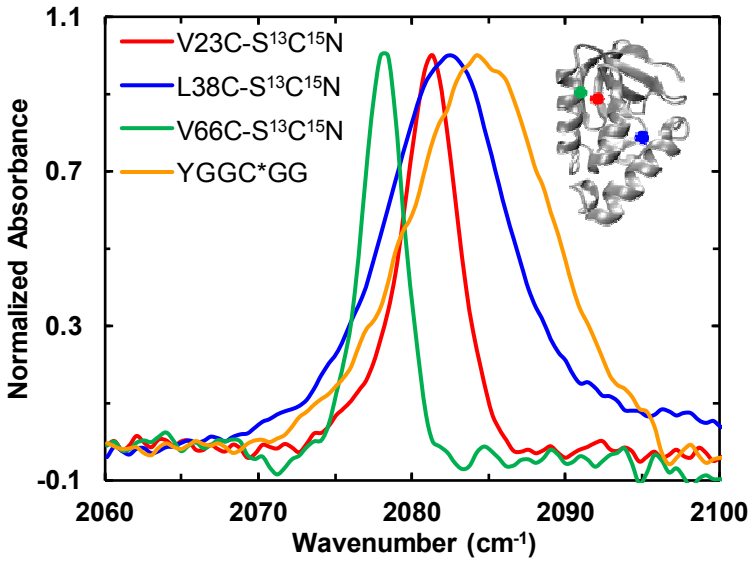

Figure 6.2: Nitrile stretching bands of V23C-$S^{13}C^{15}N$, L38C-$S^{13}C^{15}N$, V66C-$S^{13}C^{15}N$, and YGGC*GG in 20 mM cacodylate buffer in D_2O (pD = 7.5), as indicated. Fitting these spectra to a pseudo-Voigt profile yielded those spectral parameters presented in Table 6.1. Inset shows the location of the mutations in the Δ+PHS protein (pdb 3bdc).

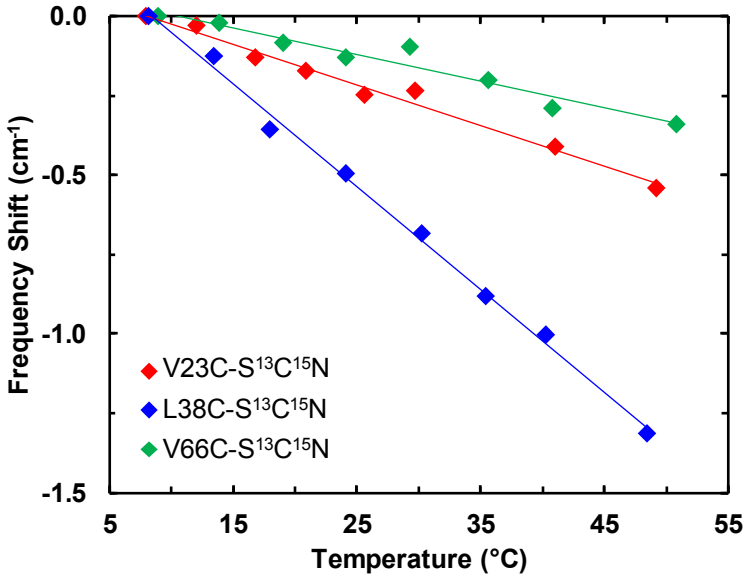

Figure 6.3: Temperature dependence of the nitrile stretching frequency of V23C-$S^{13}C^{15}N$, L38C-$S^{13}C^{15}N$, and V66C-$S^{13}C^{15}N$, in the format of a frequency shift from the value measured at the lowest temperature. The solid lines are linear fits to the data and the resultant slopes are listed in Table 6.1.

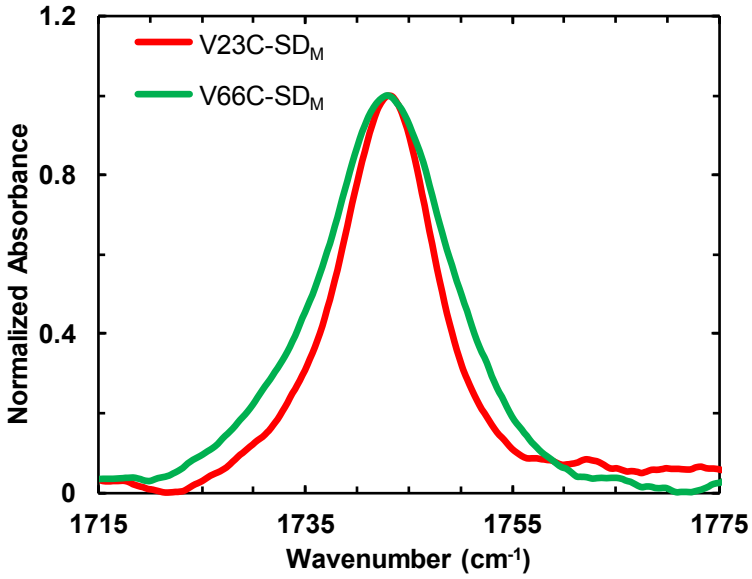

Figure 6.4: The C=O stretching vibrational bands of V23C-SD_M and V66C-SD_M in 20 mM cacodylate buffer in D_2O (pD = 7.2), as indicated. These bands are well fit to a Gaussian function with the following peak frequency (FWHM) parameters: 1742.9 cm^{-1} (7.7 cm^{-1}) for V23C-SD_M and 1742.7 cm^{-1} (10.6 cm^{-1}) for V66C-SD_M.

7: Observation of α-Sheets in the Amyloid Fibrils formed by a $TTR_{105-115}$ Mutant

7.1: ABSTRACT

Unlike the commonly encountered β-sheet secondary structure, α-sheets are not observed in proteins. While computer simulations suggest that the α-sheet could exist in amyloid oligomers and/or fibrils, to the best of our knowledge, so far there has been no experimental validation of this prediction. Herein, we combine several spectroscopic techniques to characterize the fibrils formed by a mutant of $TTR_{105-115}$ and find strong evidence that replacing Leu111 in the amyloidogenic peptide $TTR_{105-115}$ to an unnatural amino acid, aspartic acid 4-methyl ester, leads to formation of amyloid fibrils consisting of both β- and α-sheets.

7.2: INTRODUCTION

Amyloids formed by the transthyretin (TTR) protein have been linked to various diseases, including Senile Systemic Amyloidosis, Familial Amyloid Polyneuropathy and Cardiomyopathy, and Central Nervous System Selective Amyloidosis.[15,16,250] As such, numerous studies have focused on TTR amyloid fibrils, ranging from structure determination to elucidation of the underlying fibrilization mechanism.[251–254] In particular, an 11-residue segment of TTR, $TTR_{105-115}$ (sequence: YTIAALLSPYS where the numbering is kept consistent with the full-length protein) has been widely used as a model system in this regard, as this segment is known to form similar 'cross-β' amyloid fibrils as the full-length protein.[255,256] Recently, Griffin, Dobson, and co-workers have characterized the atomic structure of $TTR_{105-115}$ fibrils,[256] using a combination of structural

techniques, including magic angle spinning nuclear magnetic resonance (NMR), cyroelectron microscopy (cryo-EM), X-ray diffraction, scanning transmission electron microscopy (STEM), and atomic force microscopy (AFM). They found that these fibrils are formed from four, six, or eight two-sheet protofilaments which are aligned in a head-to-tail manner, with each protofilament consisting of parallel, in-register β-sheets that stack in an antiparallel fashion. In addition, their structure models suggested that structured water exists in the fibril core, which initially motivated us to introduce a site-specific infrared (IR) probe into the sequence of $TTR_{105-115}$ and use both linear and nonlinear IR spectroscopic techniques to assess the dynamics of such confined or structured water, a topic that has attracted considerable attention in the past few years. Specifically, we mutated Leu111 to aspartic acid 4-methyl ester (D_M), which is a well-established protein IR hydration probe.[114] The choice of this mutation is based on the fibril structure, which shows that the Leu111 side chains are either buried in a water-inaccessible region or exposed to confined water.[256] Much to our surprise, our spectroscopic results suggest that this $TTR_{105-115}$ mutant (hereafter referred to as TTR-111D_M) forms fibrils that are distinctly different from those formed by the wild-type peptide, consisting of both α- and β-sheets.

The α-sheet structure was first proposed by Pauling and Corey as a secondary structural element of proteins.[257] As indicated (Figure 7.1), the structure of an ideal α-sheet is similar to that of an ideal β-sheet in that both have the same meridional repeat distance (4.75 Å) and the same average hydrogen bond distance (2.3 Å); however, they differ in the orientation of their respective backbone NH and CO groups. In the β-sheet structure, the NH and CO groups alternate on either side of the sheet, while all the NH

(CO) groups in an α-sheet are aligned on one side of the peptide backbone plane.[257–259] Therefore, each NH/CO group in an α-sheet would form two hydrogen-bonds (H-bonds) with two CO/NH groups on a neighboring stand.[258] Comparably, β-sheets are more stable than α-sheets, which lead Pauling and Corey to reject the α-sheet structure as an "important" configuration in proteins.[258,260] Indeed, no α-sheet structures have been found in native proteins;[258] and an α-sheet conformation has only been observed in short peptides consisting of alternating D- and L-amino acids.[261–263] Interestingly, however, α-sheet conformations are frequently observed in molecular dynamics (MD) simulations of amyloid-forming proteins, including TTR, β_2-microglobulin, lysozyme, the prion protein, and polyglutamine, particularly in the regions of the protein which are considered the most amyloidogenic experimentally.[258,259,264–266] In this regard, the α-sheet has been postulated to be a common intermediate state in the formation of amyloid fibrils and, also, the underlying structure of the toxic amyloid oligomers.[258] Nevertheless, to the best of our knowledge, this structure has never been observed experimentally to exist in either transiently populated amyloid oligomers or mature fibrils, making it an elusive target for further studies. Thus, the experimental confirmation of this structure in amyloid fibrils will open new avenues for providing a better understanding of the biological role of α-sheets.

7.3: METHODS AND MATERIALS

7.3.1: Sample Preparation. The $TTR_{105\text{-}115}$ (Sequence: NH_2-YTIAALLSPYS-$CONH_2$) and TTR-111D_M (Sequence: NH_2-YTIAAL-D_M-SPYS-$CONH_2$) peptides were synthesized on a Liberty Blue automated microwave peptide synthesizer (CEM, Matthews, NC) using

Fmoc-based solid phase peptide synthesis. Peptides were purified by reverse-phase high performance liquid chromatography (HPLC) and were identified by matrix assisted laser desorption ionization mass spectroscopy (MALDI-MS). Trifluoroacetic acid (TFA) removal and H-D exchange were achieved by multiple rounds of lyophilization. The peptides were then monomerized by dissolving in 1,1,1,3,3,3-Hexafluoro-2-propanol (HFIP) at a concentration of 1 mM. HFIP was removed using N_2 gas. The peptides were then lyophilized to remove any residual HFIP. The peptides were aggregated following previous practice.[256] First, the samples were dissolved in 10% ACN in D_2O at a concentration of 4 mM and a pD of 2.0. They were heated at 37 °C for 6 hrs. (mutant) or 48 hrs. (wild-type) then stored at 25 °C in a dry box until aggregated (~1 week for wild-type and <1 day for mutant).

7.3.2: Atomic Force Microscopic Measurements. Atomic force microscopic (AFM) samples were prepared by placing ~5 µL of aggregated sample (4 mM) on clean mica. Then, 200 µL of Millipore water was used to rinse the mica. The mica was then dried using N_2 gas. AFM images were obtained on a Bruker Dimension Icon Atomic Force Microscope and were processed using the free Gwyddion software.

7.3.3: Fourier Transform Infrared Spectroscopic Measurements. Fourier Transform Infrared (FTIR) spectra were collected on a Nicolet 6700 FTIR spectrometer with a spectral resolution of 1 cm^{-1} and an average of 256 scans. The absorbance of the peptide samples (4 mM monomer concentration in 10% ACN in D_2O, pD 2) were measured between two CaF_2 windows with a 50 µm Teflon spacer, where a background spectrum of the aggregation solvent (10% ACN in D_2O, pD 2) was subtracted.

Attenuated total reflectance Fourier transform IR (ATR-FTIR) spectra were collect using a Horizon ATR accessory (Harrick Scientific Products, Inc.) for the Nicolet 6700 spectrometer on a germanium (Ge) crystal. The aggregated sample (5 μL of 4 mM) was added to the crystal and dried under vacuum for at least 24 hours before measurement. The absorbance shown was the average of 256 scans where the background spectrum of a clean crystal was subtracted.

7.3.4: Two-dimensional IR Measurements. 2D IR spectra were collected on a homebuilt 2D IR setup that has been described in detail previously, in a sample cell consisting of two CaF_2 windows with a 25 μm Teflon spacer.[267] Briefly, three femtosecond IR beams (k_1, k_2, and k_3, respectively) are focused onto the sample in a box-car geometry to generate the third-order response. This response was heterodyned with the local oscillator (LO) pulse and then the combined signal was dispersed by a monochromator and detected with a 64-pixel liquid nitrogen-cooled array detector (Infrared Associates, FL). By scanning the time delay (τ) at 2 fs steps between k_1 and k_2 pulses from -3.5 to 4.0 ps, the rephrasing and nonrephasing data were measured. The delay time (T) between k_1/k_2 and k_3 pulses was 0 ps.

7.3.5: Circular Dichroism Measurements. Circular dichroism (CD) spectra were collected on an Aviv CD 62A DS spectrometer with an averaging time of 20 s at a monomer concentration of 4 mM in 10% ACN in D_2O, pD 2 between two CaF_2 windows with a 6 μm spacer (fibrils) or in a quartz cuvette with a pathlength of 1 mM (monomers). A background spectrum of the aggregation solvent (10% ACN in D_2O, pD 2), measured

under the same conditions as the sample, was subtracted from the measured sample spectrum to produce the final spectrum.

7.4: RESULTS AND DISCUSSION

7.4.1: Linear and Nonlinear Infrared Spectroscopic Measurements. As shown (Figure 7.2), like the wild type $TTR_{105-115}$ peptide,[256] TTR-111D_M readily aggregates and forms fibrils at concentrations of a few mM. However, the amide I′ (amide I in D_2O) bands of these peptides, when aggregated, show significant differences (Figure 7.3). The amide I band, which arises mainly from the backbone C=O stretching vibrations, is a well-established IR reporter of protein conformations. This is because the amide I band of a polypeptide manifests the underlying amide I vibrational coupling, a process depending on its structure.[90] For example, the amide I′ band of amyloid fibrils consisting of β-sheets is characterized by a pair of narrow peaks that have very different intensities.[98] The strong peak is typically located at a frequency less than 1630 cm^{-1}, whereas the position of the weak one depends on the arrangement of the β-sheets. For amyloid aggregates composed of antiparallel but not parallel β-sheets, the weak peak is typically observed to locate at a frequency of 1680 cm^{-1} or higher.[90] Therefore, the FTIR results in Figure 7.3 provide the first evidence indicating that the Leu111 to D_M mutation in $TTR_{105-115}$ leads to a change in the underlying fibril structures. Consistent with the fibril structures of Fitzpatrick *et al*.,[256] the amide I′ band of an aggregated $TTR_{105-115}$ sample shows characteristics of parallel β-sheets,[98] with a pair of narrow peaks at ~1625 (strong) and ~1665 cm^{-1} (weak), respectively. Further two-dimensional infrared (2D IR) measurements confirm that these two bands result from vibrational coupling, but not

from heterogeneity of the fibril sample, as a cross peak exists at early waiting times (e.g., T = 0) (Figure 7.4).

In comparison (Figure 7.3), the FTIR spectrum of an aggregated TTR-111D_M sample in the amide I′ region contains more resolvable spectral features. As expected, the ester carbonyl of the D_M side chain gives rise to an additional IR band outside the amide I′ profile, which can be well fit by a Gaussian function with a peak frequency of 1744.8 cm^{-1} and a width (i.e., FWHM) of 7.2 cm^{-1} (Figure 7.5). According to the study of Pazos *et al.*,[114] this frequency indicates that the ester carbonyls in the fibrils are situated in a dehydrated environment, an unexpected result. Based on the fibril structures of $TTR_{105\text{-}115}$, 50% of the side chains at position 111 would face towards the hydrated core of the fibril and, hence, should produce an IR band at a frequency of 1730 cm^{-1} or lower due to H-bonding interactions with water. To verify that the fibrils produced by both the wild-type and mutant $TTR_{105\text{-}115}$ peptides indeed contain confined water, we measured the FTIR spectra of the respective fibril samples in the form of a dry film. As shown (Figure 7.6), the data clearly show that even for fibril samples that have been dried on the surface of a Ge crystal under vacuum for at least 24 hours, the D_2O IR band[21] at ~2575 cm^{-1} is still present. This result indicates that fibrils formed by both the wild-type and mutant TTR peptides can host trapped water, as initially suggested.[256] Thus, the fact that the D_M side chain in the TTR-111D_M fibrils is not located in a hydrated environment suggests that the corresponding Leu111 to D_M mutation changes the fibril structure(s). A more convincing evidence in support of this notion comes from the amide I′ band of the mutant sample, which consists of four resolvable, narrow peaks, at ~1616, ~1652, ~1662, and ~1672 cm^{-1}, respectively. The peaks at 1616 and 1662 cm^{-1} are similar to those observed in $TTR_{105\text{-}115}$ fibrils, with a modest shift toward lower wavenumbers. This

similarity suggests that the fibrils formed by TTR-111D_M contain parallel β-sheets. The question is then what produces the other two spectral features.

One possibility is that the 1652 and 1672 cm^{-1} peaks arise from non-aggregated peptides and residual trifluoroacetic acid (TFA) in the fibril sample remaining after the DCl exchange, respectively. This is because the amide I′ band of a disorder peptide is centered around 1650 cm^{-1} and TFA, which was used in the peptide synthesis and purification, exhibits a strong IR signal at 1672 cm^{-1}. To verify this possibility, we followed the aggregation/fibrillization process of the TTR-111D_M peptide by measuring its FTIR spectra as a function of time. As shown (Figure 7.7), the amide I′ band obtained at 'zero time' is dominated by a broad feature centered at ~1650 cm^{-1}, which directly rules out the possibility that the 1672 cm^{-1} peak originates from TFA. Furthermore, all of the fine spectral features are developed concomitantly, at the expense of the broad 1650 cm^{-1} feature, which, therefore, argues against the idea that the 1652 cm^{-1} peak corresponds to non-aggregated peptides. This argument is further corroborated by the ester carbonyl signal of the D_M side chain. If a non-negligible amount of peptide monomers were still present, one would expect the presence of a broad peak at 1725 cm^{-1}, as that observed at 'time zero', in the spectrum collected at 315 minutes, which is not observed (Figure 7.7 inset). In fact, the fairly narrow bandwidth of the ester carbonyl band of the fully aggregated sample provides strong evidence indicating that the D_M side chains in the fibrils sample a rather homogenous environment and that those narrow amide I′ features originate from the underlying fibril structure, rather than from any sample heterogeneity.

Theoretical calculations have predicted that the amide I′ frequency of the α-sheet is different from its β-sheet counterpart. For example, the density functional theoretical

study of Torii[268] indicated that the α-sheet conformation of a glycine dipeptide gives rise to a strong amide I′ peak between 1670 - 1690 cm^{-1} and a less intense one between 1630 - 1650 cm^{-1}. Similarly, the quantum mechanics calculation of Huo and coworkers[269] on the amide I band of an alanine dipeptide predicted that the α-sheet conformation would exhibit a strong peak at 1653 cm^{-1}, with two smaller ones at 1623 and 1681 cm^{-1}. Therefore, based on these theoretical predictions, a more plausible interpretation of the atypical amide I′ band shape of the TTR-111D_M fibrils is that it contains contributions from both β- and α-sheets. Because of the narrowness of the ester carbonyl vibrational band, it is unlikely that the aggregated TTR-111D_M sample is composed of two types of fibrils, with one consisting of β-sheets and the other of α-sheets. Therefore, we hypothesize that the fibrils are assembled from protofibrils consisting of both β- and α-sheets. To further validate this notion, we carried out 2D IR measurements on fibrils formed by TTR-111D_M. As shown (Figure 7.8), the 2D IR spectrum of a fully aggregated TTR-111D_M sample at T = 0 show cross peaks between the 1616 and 1662 cm^{-1} peaks as well as between the 1652 and 1672 cm^{-1} bands. These results therefore indicate that the frequencies arise from a single structure and not from two different fibril structures.

The protein amide A band mainly arises from the backbone N-D stretching vibrations (in D_2O). Therefore, the difference between the backbone-backbone H-bonding patterns of the β-and α-sheets is expected to affect not only their amide I′ bands, as observed (Figure 7.3), but also their amide A vibrations. Indeed, the aggregated $TTR_{105-115}$ and TTR-111D_M samples show distinct differences in their amide A bands (Figure 7.6), providing additional evidence supporting the notion that the underlying fibrillar structures of these two peptides are not identical.

7.4.2: Circular Dichroism Spectra. Further supporting evidence comes from the far-UV circular dichroism (CD) spectra of $TTR_{105-115}$ and TTR-$111D_M$. As indicated (Figure 7.9), the CD spectrum of an aggregated $TTR_{105-115}$ sample is similar to that observed previously.[270] However, in comparison, the CD spectrum of an aggregated TTR-$111D_M$ sample is not only different in shape from that of $TTR_{105-115}$ but also exhibits a giant CD signal at 205 nm. This is an amazing result, as, to the best of our knowledge, such a giant CD signal has never been observed in a peptide or protein system. It is well known that certain amino acids, such as tryptophan, can produce an additional CD signals (i.e., a CD couplet) in the far-UV spectral region due to exciton couplings.[271] Therefore, the giant CD signal observed for the TTR-$111D_M$ aggregates is most likely due to additional contributions from the DM side chains, as methyl acetate, a model compound for the side chain of D_M, exhibits a strong far-UV absorption band, peaked at ~205 nm in water (Figure 7.10). Since the magnitude of an exciton CD couplet is sensitively dependent on the relative positions (i.e., distance and orientation) of the two coupled chromophores,[44] the much enhanced CD signal at 205 nm suggests that in the fibrils, the D_M side chains from two in-register sheets must be in close proximity. In support of this picture, the far UV CD spectrum of TTR-$111D_M$ monomers only show typical spectral features associated with unstructured peptides (Figure 7.11). It is also possible that the giant CD signal is due to exciton couplings between the electronic transitions arising from the peptide backbone and the D_M side chains. Yet, another possible scenario is that the underlying fibril structure supports a supramolecular chirality, leading to a large increase in the CD signal, as has been observed for chiral helical stacks.[272–275] Irrespective of the

interpretation, these CD spectra once again indicate that TTT-111D_M form fibrils with a distinct structure.

7.4.3: Atomic Force Microscopic Measurements. Finally, the most direct evidence indicating that the fibril structures of the wild-type and mutant aggregates are dissimilar comes from the AFM images of individual fibrils formed by $TTR_{105-115}$ and TTR-111D_M. It is apparent that the dimension perpendicular to the long fibril axis (hereafter referred to as ***d***) is larger for the TTR-111D_M fibrils than for the $TTR_{105-115}$ fibrils (Figure 7.12). A more quantitative assessment further substantiates this notion as it indicates that the height (***h***) of the $TTR_{105-115}$ fibrils is about 10 - 12 nm, which is on the same order as measured previously,[256] whereas the TTR-111D_M fibrils have a ***h*** of about 4-6 nm and a ***d*** of about 50 nm. This width-to-height ratio seems to suggest that the TTR-111D_M fibrils adopt a ribbon-like structure; however, the fact that these fibrils contain confined water (Figure 7.6) argues against this picture. Furthermore, a two-dimensional fast Fourier transform (2D-FFT) analysis of individual fibrils (Figure 7.13) reveals that the wild-type fibrils have an intrinsic twist along the long fibril axis, with a periodicity of ~83 nm, a value similar to that reported in other studies for $TTR_{105-115}$ (95 ± 10 nm).[256] On the contrary, the corresponding 2D-FFT analysis of the TTR-111D_M fibrils indicates that they have no detectable twists or pitches (Figure 7.13). This finding not only provides definitive evidence that the $TTR_{105-115}$ and TTR-111D_M fibrils have different structures, but is also interesting as, except fibrils with a ribbon structure, non-twisting fibrils are rarely observed. One known example is the study of Usov et al,[276] which showed that the protein bovine serum albumin can form fibrils with a height of ~4 nm, among other fibril morphologies.

7.5: CONCLUSIONS

In summary, we have employed a combination of biophysical techniques to characterize the structures of the fibrils formed by $TTR_{105-115}$ and a Leu111 to D_M mutant (TTR-111D_M). We found that the TTR-111D_M fibrils exhibit quite unique spectroscopic and morphological properties: (1) their D_M side chains give rise to a single and narrow IR band at ~1745 cm^{-1}, signifying a rather homogenous fibril population; (2) their amide I′ band consists of two pairs of vibrationally coupled peaks; (3) they produce a giant far-UV CD signal; and (4) they have no detectable pitch or twist along the long fibril axis. Taken together, these results not only indicate that the fibrils formed by TTR-111D_M are structurally different from those formed by the wild-type peptide, but are also best explained by a description that the TTR-111D_M fibrils consist of both β- and α-sheets, according to previous simulations studies. Because of the elusiveness of α-sheet and its potential role in amyloid diseases, we hope that our work will inspire further structural studies on the TTR-111D_M fibrils, using computational and high-resolution structural methods.

7.6: ACKNOWLEDGMENTS

We thank the National Institutes of Health (P41-GM104605) for funding. M.R.H. is supported by a National Science Foundation Graduate Research Fellowship (DGE-1321851).

This chapter has been adapted from a manuscript in preparation: Mary Rose Hilaire,* Bei Ding,* Jianxin Chen, Feng Gai (2017). (*Equal Contribution)

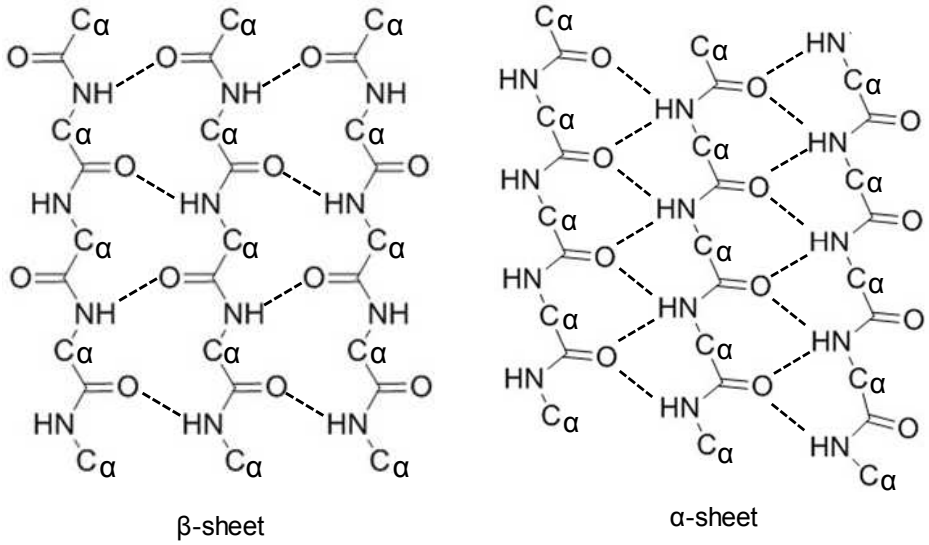

Figure 7.1: Schematic representation of the peptide backbone in β-sheets (left) and α-sheets (right), showing the orientation of the backbone atoms and the hydrogen bonding pattern between them.

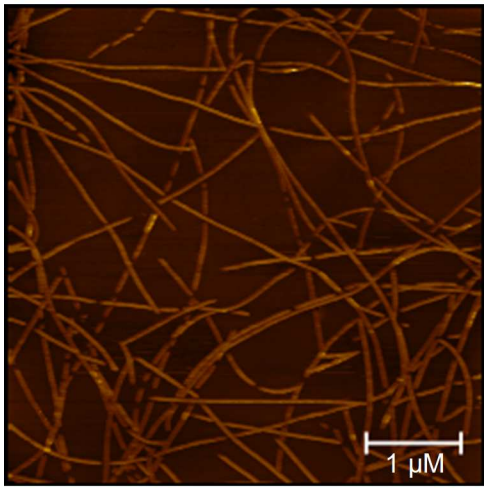

Figure 7.2: AFM Image of TTR-111D$_M$ fibrils.

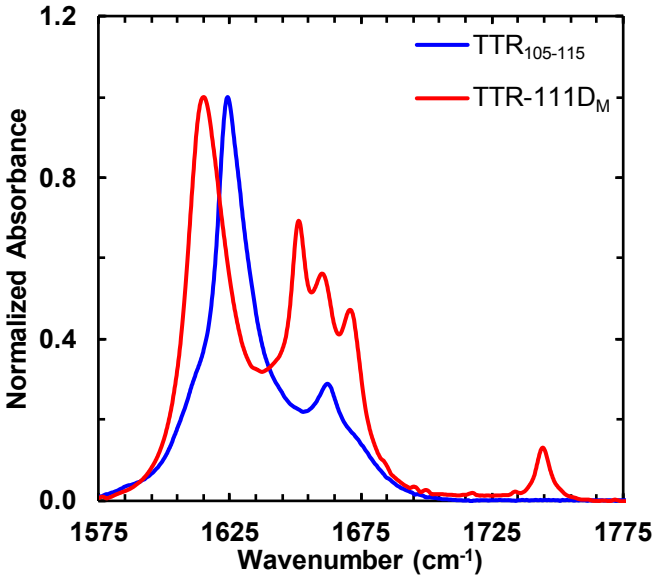

Figure 7.3: Comparison of the normalized FTIR spectra of $TTR_{105-115}$ and $TTR\text{-}111D_M$ fibrils, as indicated.

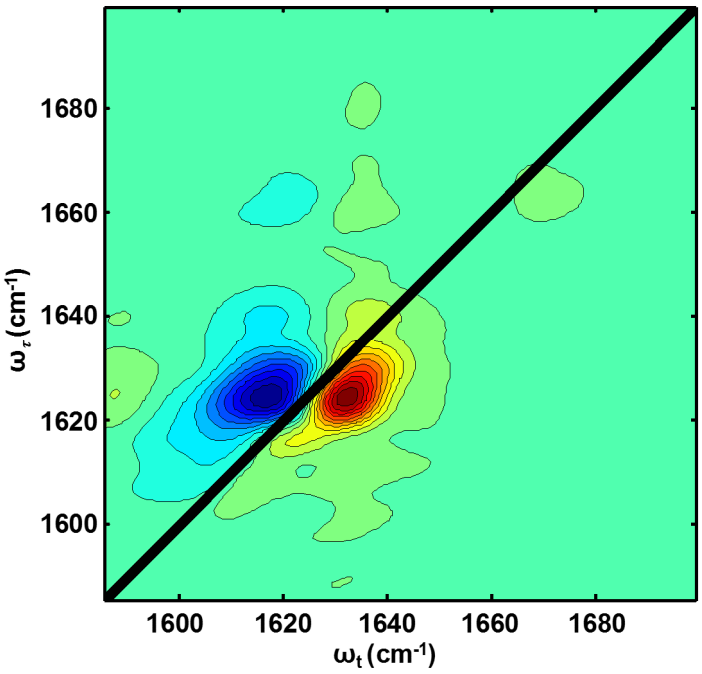

Figure 7.4: 2D IR Plot of the Amide I' Region of $TTR_{105-115}$ fibrils at $T = 0$.

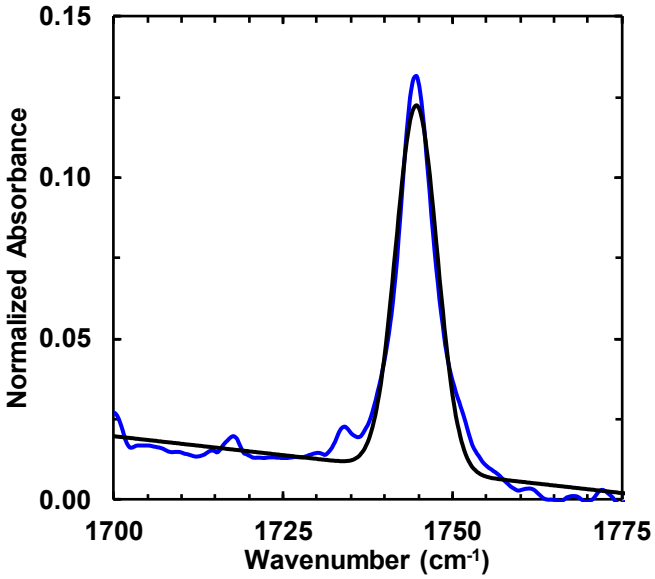

Figure 7.5: Ester carbonyl stretching frequency of TTR-111D$_M$ fibrils (blue) with a fit (black) to a Gaussian function with a center frequency of 1744.8 cm^{-1} and FWHM of 7.2 cm^{-1}.

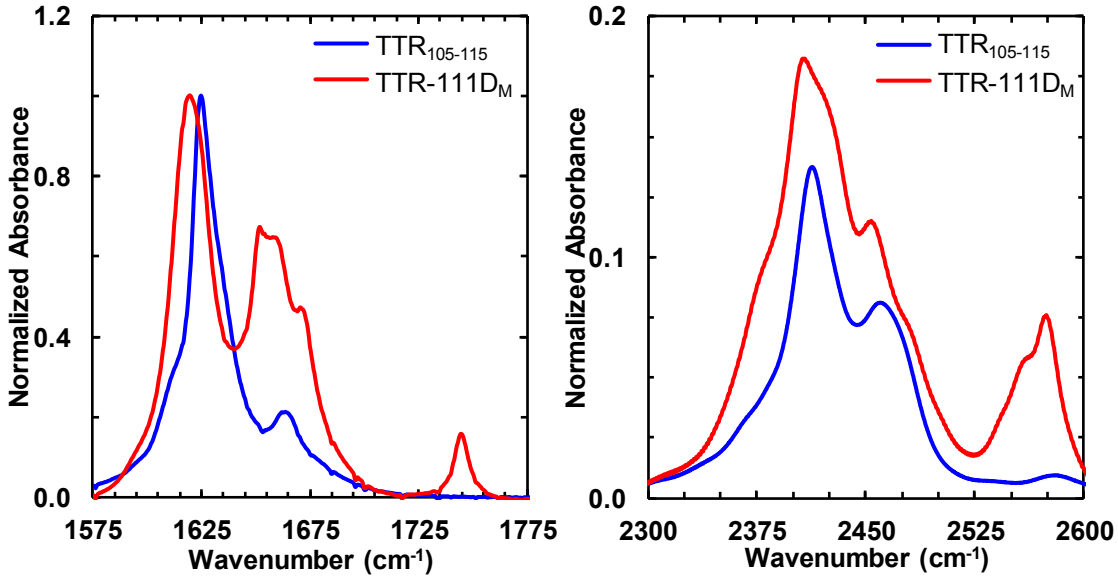

Figure 7.6: ATR-FTIR spectra of wild-type $TTR_{105-115}$ (blue) and TTR-$111D_M$ (red) fibrils as dry films in the amide I' (left) and amide A (right) regions, normalized to their respective maxima in the amide I' region. The dry film amide I' regions of both peptides are comparable to those measured in solution (Figure 7.3).

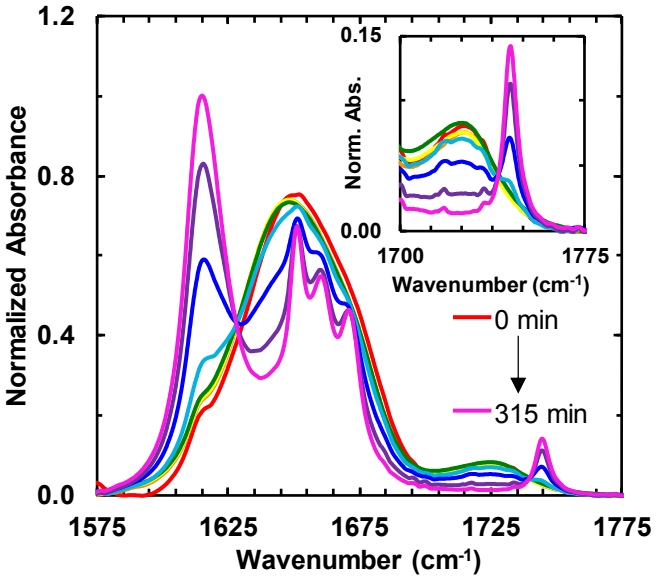

Figure 7.7: Time-dependent FTIR spectra of TTR-111D$_M$, from 0 to 315 minutes with an interval of 45 minutes. Shown in the inset are the corresponding spectra in the ester vibrational band region.

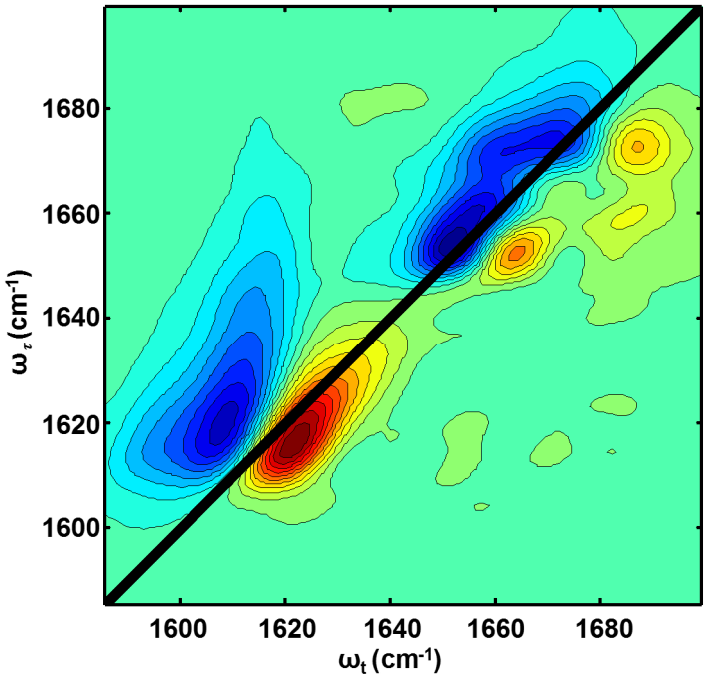

Figure 7.8: 2D IR Plot of the Amide I' Region of TTR-111D_M fibrils at T = 0.

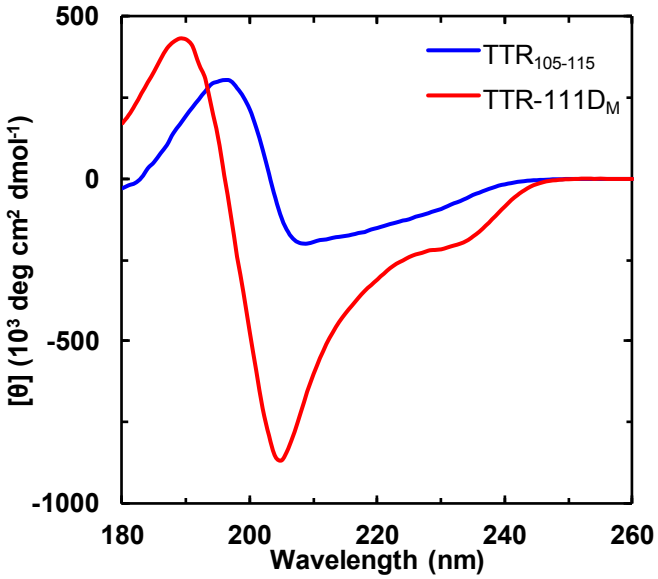

Figure 7.9: CD spectra of aggregated $TTR_{105-115}$ and $TTR\text{-}111D_M$ samples, measured using a sample holder with a pathlength of 6 μm. The mean residue ellipticity was calculated using the peptide monomer concentration (4.0 mM).

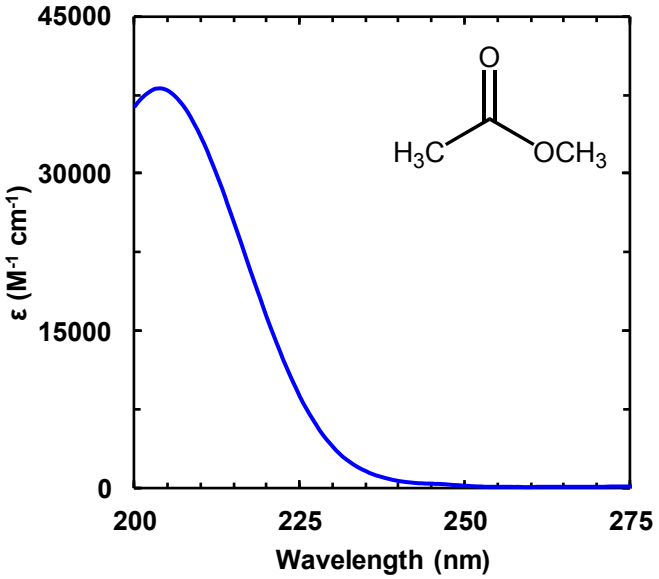

Figure 7.10: UV spectrum of methyl acetate, a model compound for the D_M side chain. The absorption maximum locates at 205 nm (~38,000 M^{-1} cm^{-1}).

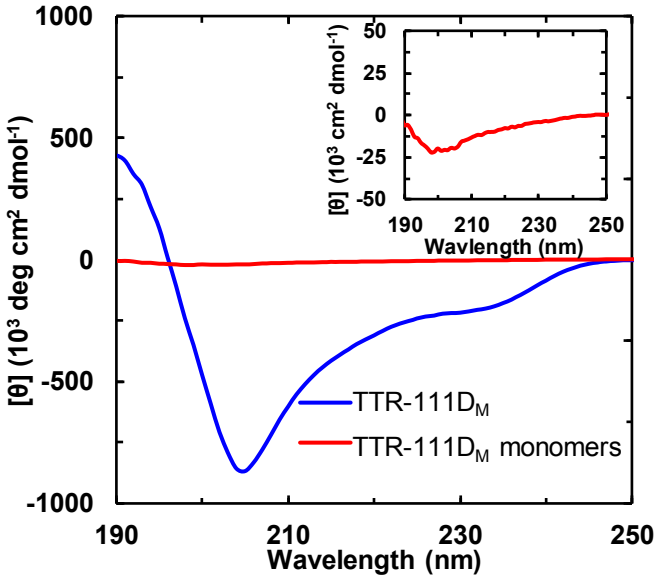

Figure 7.11: CD spectra of aggregated TTR-111D$_M$ (blue) and TTR-111D$_M$ (red) monomer samples, measured using a sample holder with a pathlength of 6 μm and 1 mm, respectively. The mean residue ellipticity was calculated using the peptide monomer concentration (4.0 mM). The inset shows the TTR-111D$_M$ monomers on a smaller scale, to show the negative peak at 200 nm, consistent with a random coil structure.

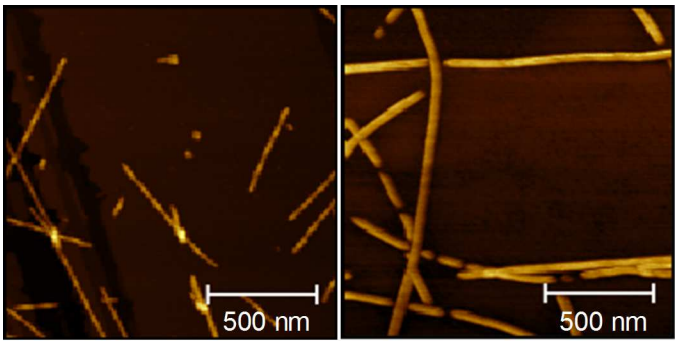

Figure 7.12: AFM images of amyloid fibrils formed by $TTR_{105-115}$ (left) and TTR-$111D_M$ (right).

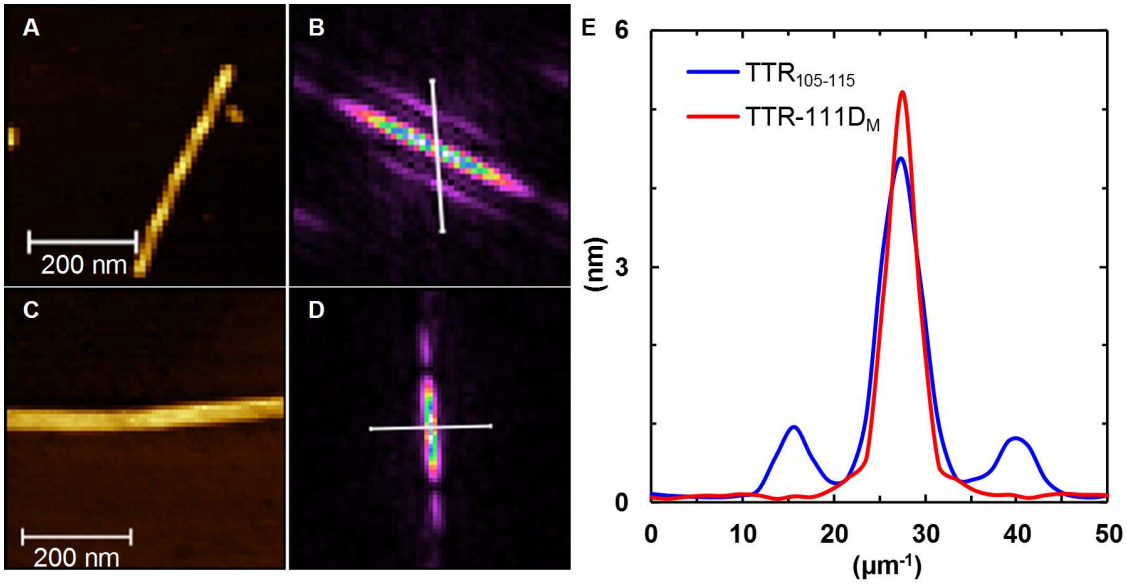

Figure 7.13: AFM images of $TTR_{105-115}$ (**A**) and $TTR\text{-}111D_M$ (**C**) fibrils with their respective 2D-FFT plot (**B**, **D**). Cuts along the white lines in **B** and **D** are shown in (**E**) for both fibrils. The distance between the peaks in **E** correspond to the pitch of the fibrils. For the $TTR\text{-}111D_M$ fibrils, there is only one peak in **E**, which means there is no measurable twist in the fibrils. On the other hand, the $TTR_{105-115}$ fibrils in **E** have three peaks at 15.8, 27.3, and 40.0 μm^{-1}, which corresponds to an average pitch of 83 nm.

8: Summary and Future Directions

Proteins are responsible for a wide variety of cellular activities and functions due to their unique physical and chemical properties that are encoded in their structure. A molecular-level understanding of the relationship between a protein's structure, dynamics, and function is a challenging problem, as the many degrees of freedom in a protein limit our ability to monitor the necessary motions of the protein in question with sufficient temporal and structural resolution. To improve this ability, one practice in fluorescence and infrared spectroscopic measurements is to employ an unnatural amino acid that is sensitive to the structure and/or the environment of a protein. Therefore, continuing this line of research, the goal of this thesis is to characterize novel amino acid-based spectroscopic probes as well as to use these probes to investigate important biological questions about proteins.

The unnatural amino acid *p*-cyanophenylalanine (Phe_{CN}) has become a versatile spectroscopic probe in the investigation of protein structure, dynamics, and function since its introduction fifteen years ago as an infrared[24] and fluorescence[82] probe. As described in detail in Chapter 3, in order to expand the spectroscopic utility of this fluorophore, we paired Phe_{CN} with an unnatural amino acid quencher, selenomethionine (SeMet). We found that SeMet quenches Phe_{CN} fluorescence through an electron transfer process in a distance-dependent manner and the quenching rate constant, k_Q, can be described by: $k_Q = k_0 \exp(-\beta(r - a_0))$, where a_0 is equal to 7.0 Å and is the contact distance between Phe_{CN} and SeMet, k_0 is equal to 42.6 ns^{-1} and is the quenching rate at the distance a_0, and β is equal to 1.6 $Å^{-1}$ and is the characteristic quenching constant. The utility of this novel fluorophore-quencher pair is further demonstrated by applying it to understand conformational distributions of a series of polyproline peptides.

Recently, Raleigh and coworkers have extended our work to other applications, showing that the Phe_{CN}-SeMet pair can be used to investigate local helical[277] and coiled-coil[278] structures. A potential future application of this system is to utilize it to interrogate the structure of amyloid fibrils. For example, by introducing the Phe_{CN} fluorophore and the SeMet quencher onto the N- and C- termini, respectively, of an amyloid-forming peptide, the arrangement of the β-strands in the fibrils could be determined. In antiparallel β-sheets where the strands stack in opposing directions, the fluorescence of Phe_{CN} in one strand would be quenched by SeMet on an adjacent strand. On the other hand, in parallel β-sheets where the strands stack in the same direction, no quenching would occur. In this way, structural information about the fibrils could be determined using a low amount of sample.

Additionally, Raleigh and coworkers also found that SeMet is an efficient quencher of tryptophan (Trp) fluorescence which can be used to probe the formation of α-helices and β-sheets,[279] similar previous studies.[59] We have collected preliminary evidence which indicates that SeMet also quenches the fluorescence of 4-cyanotryptophan (4CN-Trp) (Figure 8.1). Taken together with the previous findings, we hypothesize that SeMet could act as a universal quencher of commonly-used fluorescence probes, including organic dyes and fluorescence proteins. To verify this theory, a systematic study of fluorophores that span the UV and visible region would need to be performed. Assuming the hypothesis is true, SeMet could then be used as a quencher in single molecule-based experiments, including fluorescence correlation spectroscopic (FCS) measurements, where FRET pairs are typically utilized. In these types of experiments, a quencher could simplify the interpretation, as there would only

be one fluorophore rather than two whose individual contributions can be hard to determine.

The development of new fluorophores that are tailored to investigate a specific biological process, either *in vitro* or *in vivo*, has been at the forefront of protein science research.[77,78,172–176] In that vein, as discussed in Chapter 4, we characterize the photophysical properties of 4CN-Trp, a novel fluorescence probe and we describe in Chapter 5 the fluorescence decay kinetics of all the nitrile derivatives of indole, the side chain of Trp. We found that 4CN-Trp has an absorption maximum at 325 nm, an emission maximum at 420 nm, a quantum yield between 0.8-0.9, a fluorescence lifetime of 13.7 ns, and good photostability, properties that make 4CN-Trp an ideal fluorophore in the blue region of the spectrum for use in biological imaging and spectroscopy. We showed that 4CN-Trp can be utilized in biological imaging techniques like wide field and confocal microscopies to view mammalian cell death.

Furthermore, 4CN-Trp has tremendous potential in a variety of applications, including single molecule-based measurements like FCS, two photon absorption imaging, and super resolution fluorescence techniques like total internal reflection fluorescence (TIRF) microscopy. FCS currently relies on the use of an organic dye or fluorescent protein tethered to the protein of interest. 4CN-Trp is a promising alternative to these fluorophores, particularly when the size of the fluorophore is essential, as is the case in the investigation of membrane proteins. Additionally, 4CN-Trp absorbs in the near-UV region of the spectrum, which could be damaging to cells. However, dependent on its two photon absorption cross section, 4CN-Trp could be used in two photon absorption imaging, where the fluorophore could be excited using femtosecond pulses of

700 nm light derived from a standard Ti:Sapphire laser, instead of the damaging UV light.

In an effort to understand how the photophysical properties of 4CN-Trp are derived, in Chapter 5, we explore the absorption and emission properties of all the nitrile derivative of indole in a variety of solvents. Preliminary results indicate that carbon 4 in the indole side chain is a special position compared to the others, though a complete understanding of this phenomenon is lacking. In order to gain this understanding, future work includes performing high-level quantum mechanical calculations in order to investigate the effect the two lowest excited states and/or vibrational coupling have on the photophysical properties. With this understanding, we hope to develop the 'best' amino acid-based fluorophore, which would have all the properties of 4CN-Trp but absorb in the visible region of the spectrum and have an improved extinction coefficient. We believe that the addition of a stronger electron-withdrawing group than the nitrile moiety to the 4 position would red-shift the absorption and emission maxima of indole further than 4CN-Trp, though we cannot predict how this change would affect the fluorescence intensity and lifetime without the aforementioned calculations.

In the later chapters, we shift our focus to the application of site-specific infrared probes. In Chapter 6, we utilized two side chain infrared probes: the -SCN moiety and the $-CH_2COOCH_3$ ($-D_M$) moiety, both of which can be conveniently appended onto a cysteine residue in a protein,[113,117] to investigate the hydration status and the electrostatic environment of the hydrophobic interior of the protein staphylococcal nuclease. The electrostatics in the hydrophobic interior of proteins have been widely studied, as they play a large role in a variety of biophysical and biochemical processes.[217–224] First, to understand the hydration status of three positions in the

protein: V23, L38, and V66, we incorporated the -SCN probe into these positions. Through FTIR measurements, we found that only the L38 position was exposed, at least partially, to water molecules while both V23 and V66 positions were devoid of water. Utilizing the Onsager Field equation[242] and the infrared frequency of the -D_M probe incorporated at the V66 position, the dielectric constant in the cavity at this position was determined experimentally to be 15.1. A future goal of this work to explore more positions in the protein (>20) in order to completely describe every local electrostatic environment of the protein.

Finally, in Chapter 7, we show that a leucine to aspartic acid methyl ester (i.e. -D_M) mutation at position 111 in a short segment of the amyloidogenic transthyretin protein suggests a new fibrillar structure consisting of both β- and α-sheets, where the latter structure, though proposed as a secondary structure of proteins and an on-pathway intermediate to the formation of amyloid fibrils,[257,258] has not been observed experimentally. FTIR spectra of both the mutant (TTR-111D_M) and wild-type ($TTR_{105-115}$) fibrils reveal differences in the amide I' bands of the fibrils, particularly the formation of narrow bands centered at 1652 and 1672 cm^{-1} in the mutant fibrils. Further nonlinear infrared methods on TTR-111D_M fibrils confirm two pairs of vibrational couplings between amide I' peaks, indicative of a single structure and sample homogeneity. Additionally, CD measurements on both fibrils not only show a difference in structure but also that additional electronic coupling occurs in the TTR-111D_M fibrils between the -D_M side chains in adjacent strands. Through AFM images, the TTR-111D_M fibrils are found to have no detectable twist along their fibril axis in stark contrast to the ~85 nm periodicity that was observed for $TTR_{105-115}$ fibrils.

Potential future directions of this project include a further investigation into the electronic coupling of the -D_M side chains, as this could be used as a side chain CD probe of fibrillar structure, similar to the development of 5-cyanotryptophan and Phe_{CN} as CD probes of protein structure.[45] Additionally, as α-sheets have been proposed as the toxic conformer of amyloid fibrils,[258] an investigation into the toxicity of TTR-111D_M fibrils compared to that of $TTR_{105\text{-}115}$ fibrils would be an interesting study. Particularly, because a leucine to methionine mutation at the 111 position in the TTR protein is associated with the disease familial amyloid cardiomyopathy in patients of Dutch ancestry,[280] proposed future experiments could investigate whether the difference in structure is caused by the position of the mutation or by the D_M side chain itself. Preliminary AFM images (Figure 8.2) of TTR-110D_M, fibrils, which contain a leucine to D_M mutation at the 110 position, are similar to those of wild-type $TTR_{105\text{-}115}$ fibrils, suggesting that the change in structure is due to the position of the mutation rather than the mutation itself. Further experiments are needed to confirm these findings, including linear and nonlinear infrared measurements and high resolution AFM images.

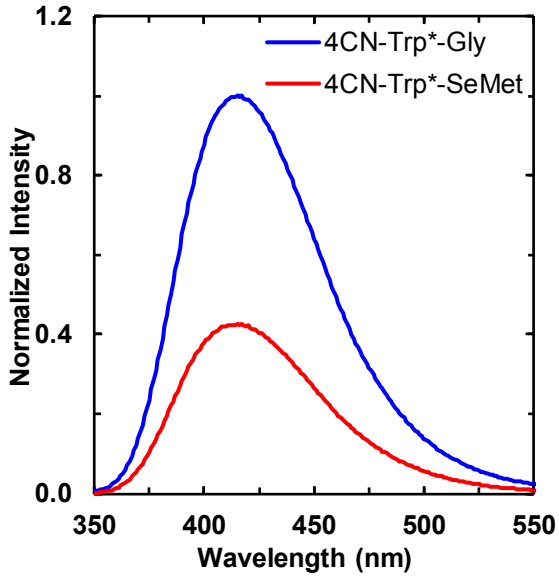

Figure 8.1: Fluorescence spectra of 4CN-Trp*-Gly (blue) and 4CN-Trp*-SeMet (red), indicating that SeMet quenches the fluorescence of 4CN-Trp. These spectra were obtained with an excitation wavelength of 325 nm at 25 °C and the optical densities of all the samples at 325 nm were equal (0.003).

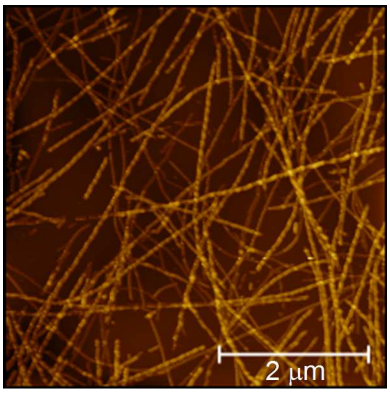

Figure 8.2: AFM Image of TTR-110D$_M$ fibrils. Twisting of the fibrils is clearly observed, which is similar to TTR$_{105-115}$ fibrils but different from TTR-111D$_M$ fibrils.

REFERENCES

(1) Nelson, D. L.; Cox, M. M. *Lehninger Principles of Biochemistry*, Fifth Edit.; W.H. Freeman and Company: New York, NY, U.S.A., 2008.

(2) Schechter, A. N.; Chen, R. F.; Anfinsen, C. B. *Science* **1970**, *167* (3919), 886–887.

(3) Anfinsen, C. B. *Science* **1973**, *181* (4096), 223–230.

(4) Levinthal, C. *J. Chim. Phys. Physico-Chimie Biol.* **1968**, *65* (1), 44–45.

(5) Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. *Annu. Rev. Phys. Chem.* **1997**, *48* (1), 545–600.

(6) Dill, K. A.; Ozkan, S. B.; Shell, M. S.; Weikl, T. R. *Annu. Rev. Biophys.* **2008**, *37*, 289–316.

(7) Ikura, M.; Clore, G. M.; Gronenborn, A. M.; Zhu, G.; Klee, C. B.; Bax, A. *Science* **1992**, *256*, 632–638.

(8) Warshel, A. *Acc. Chem. Res.* **1981**, *14* (9), 284–290.

(9) Henzler-Wildman, K.; Kern, D. *Nature* **2007**, *450* (7172), 964–972.

(10) Hanoian, P.; Liu, C. T.; Hammes-Schiffer, S.; Benkovic, S. *Acc. Chem. Res.* **2015**, *48* (2), 482–489.

(11) Whitford, P. C.; Onuchic, J. N. *Curr. Opin. Struct. Biol.* **2015**, *30*, 57–62.

(12) Ellis, R. J. *Curr. Opin. Struct. Biol.* **2001**, *11* (1), 114–119.

(13) Hartl, F. U.; Hayer-Hartl, M. *Science* **2002**, *295* (5561), 1852–1858.

(14) Dobson, C. M. *Nature* **2003**, *426*, 884–890.

(15) Selkoe, D. J. *Nature* **2003**, *426* (6968), 900–904.

(16) Chiti, F.; Dobson, C. M. *Annu. Rev. Biochem.* **2006**, *75*, 333–366.

(17) Jansen, R.; Grudzielanek, S.; Dzwolak, W.; Winter, R. *J. Mol. Biol.* **2004**, *338* (2),

203–206.

(18) Petkova, A. T.; Leapman, R. D.; Guo, Z.; Yau, W.-M.; Mattson, M. P.; Tycko, R. *Science* **2005**, *307* (5707), 262–265.

(19) Kurouski, D.; Lu, X.; Popova, L.; Wan, W.; Shanmugasundaram, M.; Stubbs, G.; Dukor, R. K.; Lednev, I. K.; Nafie, L. A. *J. Am. Chem. Soc.* **2014**, *136* (6), 2302–2312.

(20) van Wilderen, L. J. G. W.; Kern-Michler, D.; Müller-Werkmeister, H. M.; Bredenbeck, J. *Phys. Chem. Chem. Phys.* **2014**, *16*, 19643–19653.

(21) Kim, Y. S.; Liu, L.; Axelsen, P. H.; Hochstrasser, R. M. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (42), 17751–17756.

(22) Cho, H. S.; Dashdorj, N.; Schotte, F.; Graber, T.; Henning, R.; Anfinrud, P. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107* (16), 7281–7286.

(23) Kim, J. G.; Kim, T. W.; Kim, J.; Ihee, H. *Acc. Chem. Res.* **2015**, *48* (8), 2200–2208.

(24) Getahun, Z.; Huang, C.-Y. Y.; Wang, T.; De León, B.; DeGrado, W. F.; Gai, F. *J. Am. Chem. Soc.* **2003**, *125* (2), 405–411.

(25) Ma, J.; Pazos, I. M.; Zhang, W.; Culik, R. M.; Gai, F. *Annu. Rev. Phys. Chem.* **2015**, *66*, 357–377.

(26) Tsien, R. Y. *Annu. Rev. Biochem.* **1998**, *67*, 509–544.

(27) Zimmer, M. *Chem. Rev.* **2002**, *102* (3), 759–782.

(28) Day, R. N.; Davidson, M. W. *Chem. Soc. Rev.* **2009**, *38*, 2887–2921.

(29) Berezin, M. Y.; Achilefu, S. *Chem. Rev.* **2010**, *110* (5), 2641–2684.

(30) Chudakov, D. M.; Matz, M. V; Lukyanov, S.; Lukyanov, K. A. *Physiol. Rev.* **2010**, *90*, 1103–1163.

(31) Mishin, A. S.; Belousov, V. V.; Solntsev, K. M.; Lukyanov, K. A. *Curr. Opin. Chem. Biol.* **2015**, *27*, 1–9.

(32) Enterina, J. R.; Wu, L.; Campbell, R. E. *Curr. Opin. Chem. Biol.* **2015**, *27*, 10–17.

(33) Yang, T. T.; Cheng, L.; Kain, S. R. *Nucleic Acids Res.* **1996**, *24* (22), 4592–4593.

(34) Yang, T.-T.; Sinai, P.; Green, G.; Kitts, P. A.; Chen, Y.-T.; Lybarger, L.; Chervenak, R.; Patterson, G. H.; Piston, D. W.; Kain, S. R. *J. Biol. Chem.* **1998**, *273* (14), 8212–8216.

(35) Ormö, M.; Cubitt, A. B.; Kallio, K.; Gross, L. A.; Tsien, R. Y.; Remington, S. J. *Science* **1996**, *273* (5280), 1392–1395.

(36) Elsliger, M.-A.; Wachter, R. M.; Hanson, G. T.; Kallio, K.; Remington, S. J. *Biochemistry* **1999**, *38* (17), 5296–5301.

(37) Reid, B. G.; Flynn, G. C. *Biochemistry* **1997**, *36* (22), 6786–6791.

(38) Hink, M. A.; Griep, R. A.; Borst, J. W.; Van Hoek, A.; Eppink, M. H. M.; Schots, A.; Visser, A. J. W. G. *J. Biol. Chem.* **2000**, *275* (23), 17556–17560.

(39) Snapp, E. *Curr. Protoc. Cell Biol.* **2005**, *27*, 1–13.

(40) Kelly, S. M.; Jess, T. J.; Price, N. C. *Biochim. Biophys. Acta* **2005**, *1751* (2), 119–139.

(41) Greenfield, N. J. *Trends Anal. Chem.* **1999**, *18* (4), 236–244.

(42) Bayley, P. M. *Prog. Biophys. Mol. Biol.* **1973**, *27*, 1–76.

(43) Jiang, J.; Abramavicius, D.; Bulheller, B. M.; Hirst, J. D.; Mukamel, S. *J. Phys. Chem. B* **2010**, *114* (24), 8270–8277.

(44) Berova, N.; Di Bari, L.; Pescitelli, G. *Chem. Soc. Rev.* **2007**, *36* (6), 914–931.

(45) Mukherjee, D.; Gai, F. *Anal. Biochem.* **2016**, *507*, 74–78.

(46) Cochran, A. G.; Skelton, N. J.; Starovasnik, M. A. *Proc. Natl. Acad. Sci. U. S. A.*

2001, *98* (10), 5578–5583.

(47) Wu, L.; McElheny, D.; Huang, R.; Keiderling, T. A. *Biochemistry* **2009**, *48* (43), 10362–10371.

(48) Wu, L.; McElheny, D.; Takekiyo, T.; Keiderling, T. A. *Biochemistry* **2010**, *49* (22), 4705–4714.

(49) Royer, C. A. *Chem. Rev.* **2006**, *106* (5), 1769–1784.

(50) Papp, S.; Vanderkooi, J. M. *Photochem. Photobiol.* **1989**, *49* (6), 775–784.

(51) Gonnelli, M.; Strambini, G. B. *Biochemistry* **1995**, *34* (42), 13847–13857.

(52) Stokes, G. *Phil. Trans. R. Soc. Lond.* **1852**, *142*, 463–562.

(53) Kasha, M. *Discuss. Faraday Soc.* **1950**, *9*, 14–19.

(54) Lakowicz, J. R. *Principles of Fluorescence Spectroscopy*, Second Edi.; Springer: New York, NY, U.S.A., 1999.

(55) Mataga, N.; Kaifu, Y.; Koizumi, M. *Bull. Chem. Soc. Jpn.* **1956**, *29*, 465–470.

(56) Beechem, J. M.; Brand, L. *Annu. Rev. Biochem.* **1985**, *54*, 43–71.

(57) Vivian, J. T.; Callis, P. R. *Biophys. J.* **2001**, *80* (5), 2093–2109.

(58) Chen, Y.; Barkley, M. D. *Biochemistry* **1998**, *37* (28), 9976–9982.

(59) Yuan, T.; Weljie, A. M.; Vogel, H. J. *Biochemistry* **1998**, *37* (9), 3187–3195.

(60) Qiu, W.; Li, T.; Zhang, L.; Yang, Y.; Kao, Y. T.; Wang, L.; Zhong, D. *Chem. Phys.* **2008**, *350*, 154–164.

(61) Alston, R. W.; Lasagna, M.; Grimsley, G. R.; Scholtz, J. M.; Reinhart, G. D.; Pace, C. N. *Biophys. J.* **2008**, *94* (6), 2280–2287.

(62) Eftink, M. R.; Ghiron, C. A. *Anal. Biochem.* **1981**, *114* (2), 199–227.

(63) Clayton, A. H.; Sawyer, W. H. *Eur. Biophys. J.* **2002**, *31* (1), 9–13.

(64) Dusa, A.; Kaylor, J.; Edridge, S.; Bodner, N.; Hong, D. P.; Fink, A. L. *Biochemistry*

2006, *45* (8), 2752–2760.

(65) van Rooijen, B. D.; van Leijenhorst-Groener, K. A.; Claessens, M. M. A. E.; Subramaniam, V. *J. Mol. Biol.* **2009**, *394* (5), 826–833.

(66) Peon, J.; Pal, S. K.; Zewail, A. H. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (17), 10964–10969.

(67) Gao, G.-Y.; Li, Y.; Wang, W.; Wang, S.-F.; Zhong, D.; Gong, Q.-H. *Chin. Phys. B* **2015**, *24*, 18201.

(68) Qin, Y.; Wang, L.; Zhong, D. *Proc. Natl. Acad. Sci. U. S. A.* **2016**, *113* (30), 8424–8429.

(69) Best, R. B.; Merchant, K. A.; Gopich, I. V; Schuler, B.; Bax, A.; Eaton, W. A. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 18964–18969.

(70) Doose, S.; Neuweiler, H.; Barsch, H.; Sauer, M. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 17400–17405.

(71) Zhdanova, N. G.; Shirshin, E. A.; Maksimov, E. G.; Panchishin, I. M.; Saletsky, A. M.; Fadeev, V. V. *Photochem. Photobiol. Sci.* **2015**, *14* (5), 897–908.

(72) VanScyoc, W.; Shea, M. *Protein Sci.* **2001**, *10* (9), 1758–1768.

(73) Giepmans, B. N.; Adams, S.; Ellisman, M.; Tsien, R. Y. *Science* **2006**, *312*, 217–224.

(74) Tucker, M.; Oyola, R.; Gai, F. *Biopolymers* **2006**, *83*, 571–576.

(75) Talukder, P.; Chen, S.; Roy, B.; Yakovchuk, P.; Spiering, M. M.; Alam, M. P.; Madathil, M. M.; Bhattacharya, C.; Benkovic, S. J.; Hecht, S. M. *Biochemistry* **2015**, *54*, 7457–7469.

(76) Markiewicz, B. N.; Mukherjee, D.; Troxler, T.; Gai, F. *J. Phys. Chem. B* **2016**, *120*, 936–944.

(77) Smirnov, A. V.; English, D. S.; Rich, R. L.; Lane, J.; Teyton, L.; Schwabacher, A. W.; Luo, S.; Thornburg, R. W.; Petrich, J. W. *J. Phys. Chem. B* **1997**, *101* (15), 2758–2769.

(78) Lepthien, S.; Hoesl, M. G.; Merkel, L.; Budisa, N. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (42), 16095–16100.

(79) Wong, C. Y.; Eftink, M. R. *Protein Sci.* **1997**, *6* (3), 689–697.

(80) Broos, J.; Maddalena, F.; Hesp, B. H. *J. Am. Chem. Soc.* **2004**, *126* (1), 22–23.

(81) Xu, J.; Chen, B.; Callis, P.; Muino, P. L.; Rozeboom, H.; Broos, J.; Toptygin, D.; Brand, L.; Knutson, J. R. *J. Phys. Chem. B* **2015**, *119* (11), 4230–4239.

(82) Tucker, M. J.; Oyola, R.; Gai, F. *J. Phys. Chem. B* **2005**, *109* (10), 4788–4795.

(83) Serrano, A. L.; Troxler, T.; Tucker, M. J.; Gai, F. *Chem. Phys. Lett.* **2010**, *487*, 303–306.

(84) Aprilakis, K. N.; Taskent, H.; Raleigh, D. P. *Biochemistry* **2007**, *46* (43), 12308–12313.

(85) Marek, P.; Gupta, R.; Raleigh, D. P. *Chembiochem* **2008**, *9* (9), 1372–1374.

(86) Liu, J.; Strzalka, J.; Tronin, A.; Johansson, J. S.; Blasie, J. K. *Biophys. J.* **2009**, *96* (10), 4176–4187.

(87) Xhindoli, D.; Morgera, F.; Zinth, U.; Rizzo, R.; Pacor, S.; Tossi, A. *Biochem. J.* **2015**, *3* (465), 443–457.

(88) Ries, L. K.; Schmid, F. X.; Schmidpeter, P. A. M. *Biochemistry* **2016**, *55*, 6739–6742.

(89) Martin, J. P.; Fetto, N. R.; Tucker, M. J. *Phys. Chem. Chem. Phys.* **2016**, *18* (30), 20750–20757.

(90) Barth, A. *Biochim. Biophys. Acta - Bioenerg.* **2007**, *1767* (9), 1073–1101.

(91) Harris, D. C. *Quantitative Chemical Analysis*, Eighth Edi.; W.H. Freeman and Company: New York, NY, U.S.A., 2010.

(92) Goormaghtigh, E.; Raussens, V.; Ruysschaert, J.-M. *Biochim Biophys Acta* **1999**, *1422* (2), 105–185.

(93) Arrondo, J. L. R.; Goñi, F. M. *Prog. Biophys. Mol. Biol.* **1999**, *72* (4), 367–405.

(94) Picard, F.; Buffeteau, T.; Desbat, B.; Auger, M.; Pézolet, M. *Biophys. J.* **1999**, *76* (1), 539–551.

(95) Fink, A. L.; Seshadri, S.; Khurana, R.; Oberg, K. A. In *Infrared Analysis of Peptides and Proteins*; 1999; Vol. 750, pp 132–144.

(96) Marsh, D.; Müller, M.; Schmitt, F.-J. *Biophys. J.* **2000**, *78* (5), 2499–2510.

(97) Vigano, C.; Manciu, L.; Buyse, F.; Goormaghtigh, E.; Ruysschaert, J. M. *Protein Sci.* **2000**, *55* (5), 373–380.

(98) Sarroukh, R.; Goormaghtigh, E.; Ruysschaert, J. M.; Raussens, V. *Biochim. Biophys. Acta - Biomembr.* **2013**, *1828* (10), 2328–2338.

(99) Barth, A.; Zscherp, C. *Q. Rev. Biophys.* **2002**, *35* (4), 369–430.

(100) Krimm, S.; Bandekar, J. *Adv. Protein Chem.* **1986**, *38*, 181–364.

(101) Venyaminov, S. Y.; Prendergast, F. G. *Anal. Biochem.* **1997**, *248* (2), 234–245.

(102) Decatur, S. M. *Acc. Chem. Res.* **2006**, *39* (3), 169–175.

(103) Woys, A. M.; Almeida, A. M.; Wang, L.; Chiu, C. C.; McGovern, M.; De Pablo, J. J.; Skinner, J. L.; Gellman, S. H.; Zanni, M. T. *J. Am. Chem. Soc.* **2012**, *134* (46), 19118–19128.

(104) Moran, S. D.; Zanni, M. T. *J. Phys. Chem. Lett.* **2014**, *5* (11), 1984–1993.

(105) 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* (16), 6614–6619.

(106) Paul, C.; Wang, J.; Wimley, W. C.; Hochstrasser, R. M.; Axelsen, P. H. *J. Am. Chem. Soc.* **2004**, *126* (18), 5843–5850.

(107) Arkin, I. T. *Curr. Opin. Chem. Biol.* **2006**, *10*, 394–401.

(108) Manor, J.; Arbely, E.; Beerlink, A.; Akkawi, M.; Arkin, I. T. *J. Phys. Chem. Lett.* **2014**, *5* (15), 2573–2579.

(109) Kratochvil, H.; Carr, J.; Matulef, K.; Annen, A.; Li, H.; Maj, M.; Ostmeyer, J.; Serrano, A. L.; Raghuraman, H.; Moran, S. D.; Skinner, J. L.; Parozo, E.; Roux, B.; Valiyaveetil, F.; Zanni, M. T. *Science* **2016**, *353* (6303), 1040–1044.

(110) Jones, K. C.; Peng, C. S.; Tokmakoff, A. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (8), 2828–2833.

(111) Haris, P. I. *Biochim. Biophys. Acta* **2013**, *1828* (10), 2265–2271.

(112) Barth, A. *Prog. Biophys. Mol. Biol.* **2000**, *74*, 141–173.

(113) Ahmed, I. A.; Gai, F. *Protein Sci.* **2017**, *26* (2), 375–381.

(114) Pazos, I. M.; Ghosh, A.; Tucker, M. J.; Gai, F. *Angew. Chemie - Int. Ed.* **2014**, *53* (24), 6080–6084.

(115) Schultz, K. C.; Supekova, L.; Ryu, Y.; Xie, J.; Perera, R.; Schultz, P. G. *J. Am. Chem. Soc.* **2006**, *128* (43), 13984–13985.

(116) Jo, H.; Culik, R. M.; Korendovych, I. V; Degrado, W. F.; Gai, F. *Biochemistry* **2010**, *49* (49), 10354–10356.

(117) Fafarman, A. T.; Webb, L. J.; Chuang, J. I.; Boxer, S. G. *J. Am. Chem. Soc.* **2006**, *128*, 13356–13357.

(118) Kim, H.; Cho, M. *Acc. Chem. Res.* **2013**, *113* (8), 5817–5847.

(119) Waegele, M. M.; Tucker, M. J.; Gai, F. *Chem. Phys. Lett.* **2009**, *478*, 249–253.

(120) Edelstein, L.; Stetz, M. A.; McMahon, H. A.; Londergan, C. H. *J. Phys. Chem. B*

2010, *114*, 4931–4936.

(121) McMahon, H. A.; Alfieri, K. N.; Clark, K. A. A.; Londergan, C. H. *J. Phys. Chem. Lett.* **2010**, *1*, 850–855.

(122) Waegele, M. M.; Culik, R. M.; Gai, F. *J. Phys. Chem. Lett.* **2011**, *2*, 2598–2609.

(123) Fafarman, A. T.; Sigala, P. A.; Schwans, J. P.; Fenn, T. D.; Herschlag, D.; Boxer, S. G. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, E299–E308.

(124) Stafford, A. J.; Walker, D. M.; Webb, L. J. *Biochemistry* **2012**, *51*, 2757–2767.

(125) Sigala, P. A.; Fafarman, A. T.; Schwans, J. P.; Fried, S. D.; Fenn, T. D.; Caaveiro, J. M. M.; Pybus, B.; Ringe, D.; Petsko, G. A.; Boxer, S. G.; Herschlag, D. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, E2552–E2561.

(126) Fried, S. D.; Boxer, S. G. *Acc. Chem. Res.* **2015**, *48*, 998–1006.

(127) Markiewicz, B. N.; Lemmin, T.; Zhang, W.; Ahmed, I. A.; Jo, H.; Fiorin, G.; Troxler, T.; DeGrado, W. F.; Gai, F. *Phys. Chem. Chem. Phys.* **2016**, *18* (18), 28939–28950.

(128) Lindquist, B. A.; Furse, K. E.; Corcelli, S. A. *Phys. Chem. Chem. Phys.* **2009**, *11* (37), 8119–8132.

(129) Bazewicz, C. G.; Lipkin, J. S.; Smith, E. E.; Liskov, M. T.; Brewer, S. H. *J. Phys. Chem. B* **2012**, *116*, 10824–10831.

(130) Pazos, I. M.; Roesch, R. M.; Gai, F. *Chem. Phys. Lett.* **2013**, *563*, 93–96.

(131) Glasscock, J. M.; Zhu, Y.; Chowdhury, P.; Tang, J.; Gai, F. *Biochemistry* **2008**, *47* (42), 11070–11076.

(132) 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* (25), 5953–5962.

(133) 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* (38), 9040–9046.

(134) Goldberg, J. M.; Batjargal, S.; Petersson, E. J. *J. Am. Chem. Soc.* **2010**, *132* (42), 14718–14720.

(135) Taskent-Sezgin, H.; Marek, P.; Thomas, R.; Goldberg, D.; Chung, J.; Carrico, I.; Raleigh, D. P. *Biochemistry* **2010**, *49* (29), 6290–6295.

(136) Urbanek, D. C.; Vorobyev, D. Y.; Serrano, A. L.; Gai, F.; Hochstrasser, R. M. *J. Phys. Chem. Lett.* **2010**, *1*, 3311–3315.

(137) Chung, J. K.; Thielges, M. C.; Fayer, M. D. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (9), 3578–3583.

(138) Tang, J.; Yin, H.; Qiu, J.; Tucker, M. J.; DeGrado, W. F.; Gai, F. *J. Am. Chem. Soc.* **2009**, *131* (11), 3816–3817.

(139) Hu, W.; Webb, L. J. *J. Phys. Chem. Lett.* **2011**, *2* (15), 1925–1930.

(140) Lapidus, L. J.; Steinbach, P. J.; Eaton, W. A.; Szabo, A.; Hofrichter, J. *J. Phys. Chem. B* **2002**, *106* (44), 11628–11640.

(141) Eftink, M. R.; Ghiron, C. A. *Biochemistry* **1976**, *15* (3), 672–680.

(142) Doose, S.; Neuweiler, H.; Sauer, M. *Chemphyschem* **2009**, *10* (9–10), 1389–1398.

(143) Zhang, M.; Vogel, H. *J. Mol. Biol.* **1994**, *239*, 545–554.

(144) Kigawa, T.; Yamaguchi-Nunokawa, E.; Kodama, K.; Matsuda, T.; Yabuki, T.; Matsuda, N.; Ishitani, R.; Nureki, O.; Yokoyama, S. *J. Struct. Funct. Genomics* **2002**, *2* (1), 29–35.

(145) Hill, P. A.; Wei, Q.; Troxler, T.; Dmochowski, I. J. *J. Am. Chem. Soc.* **2009**, *131* (8), 3069–3077.

(146) Wu, Y.; Kondrashkina, E.; Kayatekin, C.; Matthews, C. R.; Bilsel, O. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (36), 13367–13372.

(147) Papp, S.; Vanderkooi, J. M.; Owen, C. S.; Holtom, G. R.; Phillips, C. M. *Biophys. J.* **1990**, *58*, 177–186.

(148) Kalé, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. *J. Comput. Phys.* **1999**, *151*, 283–312.

(149) 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.; 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.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102* (18), 3586–3616.

(150) Gfeller, D.; Michielin, O.; Zoete, V. *Nucleic Acids Res.* **2013**, *41* (D1), D327–D332.

(151) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79* (2), 926.

(152) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092.

(153) Marcus, R.; Sutin, N. *Biochim. Biophys. Acta-Reviews Bioenerg.* **1985**, *811*, 265–322.

(154) Lakowicz, J. R.; Kuśba, J.; Szmacinski, H.; Johnson, M. L.; Gryczynski, I. *Chem. Phys. Lett.* **1993**, *206*, 455–463.

(155) Kuśba, J.; Lakowicz, J. *Methods Enzymol.* **1994**, *240* (1986), 216–262.

(156) Lakowicz, J. R.; Zelent, B.; Gryczynski, I.; Kuśba, J.; Johnson, M. L. *Photochem. Photobiol.* **1994**, *60* (3), 205–214.

(157) Lakowicz, J. R.; Zelent, B.; Kuśba, J.; Gryczynski, I. *J. Fluoresc.* **1996**, *6* (4), 187–194.

(158) Zelent, B.; Gryczynski, I.; Johnson, M. L.; Lakowicz, J. R. *J. Phys. Chem.* **1996**, *100*, 18592–18602.

(159) Moser, C.; Keske, J.; Warncke, K.; Farid, R.; Dutton, P. *Nature* **1992**, *355*, 796–802.

(160) Harriman, A. *J. Phys. Chem.* **1987**, *91* (3), 6102–6104.

(161) Mishra, B.; Sharma, a.; Naumov, S.; Priyadarsini, K. I. *J. Phys. Chem. B* **2009**, *113*, 7709–7715.

(162) Stryer, L.; Haugland, R. P. *Proc. Natl. Acad. Sci. U. S. A.* **1967**, *58*, 719–726.

(163) Deber, C.; Bovey, F.; Carver, J.; Blout, E. *J. Am. Chem. Soc.* **1970**, *92*, 6191–6198.

(164) Tonelli, A. *J. Am. Chem. Soc.* **1970**, *92*, 6187–6190.

(165) Chiu, H. C.; Bersohn, R. *Biopolymers* **1977**, *16* (2), 277–288.

(166) Schuler, B.; Lipman, E. A.; Steinbach, P. J.; Kumke, M.; Eaton, W. A. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (8), 2754–2759.

(167) Radhakrishnan, A.; Vitalis, A.; Mao, A. H.; Steffen, A. T.; Pappu, R. V. *J. Phys. Chem. B* **2012**, *116*, 6862–6871.

(168) Zhong, H.; Carlson, H. A. *J. Chem. Theory Comput.* **2006**, *2*, 342–353.

(169) Zhong, D. In *Advances in Chemical Physics*; S. A. Rice, Ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2009; Vol. 143, pp 83–149.

(170) Callis, P. R. *J. Mol. Struct.* **2014**, *1077*, 22–29.

(171) Roy, R.; Hohng, S.; Ha, T. *Nat. Methods* **2008**, *5* (6), 507–516.

(172) Moroz, Y. S.; Binder, W.; Nygren, P.; Caputo, G. A.; Korendovych, I. V. *Chem.*

Commun. **2013**, *49* (5), 490–492.

(173) Wang, J.; Xie, J.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103* (26), 9785–9789.

(174) Chatterjee, A.; Guo, J.; Lee, H. S.; Schultz, P. G. *J. Am. Chem. Soc.* **2013**, *135* (34), 12540–12543.

(175) Hamada, H.; Kameshima, N.; Szymańska, A.; Wegner, K.; Łankiewicz, L.; Shinohara, H.; Taki, M.; Sisido, M. *Bioorganic Med. Chem.* **2005**, *13* (10), 3379–3384.

(176) Cohen, B. E.; McAnaney, T. B.; Park, E. S.; Jan, Y. N.; Boxer, S. G.; Jan, L. Y. *Science* **2002**, *296*, 1700–1703.

(177) Talukder, P.; Chen, S.; Liu, C. T.; Baldwin, E. A.; Benkovic, S. J.; Hecht, S. M. *Bioorganic Med. Chem.* **2014**, *22* (21), 5924–5934.

(178) Ross, J. B. A.; Szabo, A. G.; Hogue, C. W. V. In *Methods in Enzymology*; 1997; Vol. 278, p 151.

(179) Lotte, K.; Plessow, R.; Brockhinke, A. *Photochem. Photobiol. Sci.* **2004**, *3* (4), 348–359.

(180) Schneider, C. A.; Rasband, W. S.; Eliceiri, K. W. *Nat. Methods* **2012**, *9* (7), 671–675.

(181) Suzuki, K.; Kobayashi, A.; Kaneko, S.; Takehira, K.; Yoshihara, T.; Ishida, H.; Shiina, Y.; Oishi, S.; Tobita, S. *Phys. Chem. Chem. Phys.* **2009**, *11*, 9850–9860.

(182) Shaner, N. C.; Steinbach, P. A.; Tsien, R. Y. *Nat. Methods* **2005**, *2* (12), 905–909.

(183) Wakamatsu, K.; Okada, A.; Miyazawa, T.; Ohya, M.; Higashijima, T. *Biochemistry* **1992**, *31* (24), 5654–5660.

(184) Shai, Y. *Biochim. Biophys. Acta* **1999**, *1462*, 55–70.

(185) Henriques, S. T.; Melo, M. N.; Castanho, M. a R. B. *Biochem. J.* **2006**, *399* (1), 1–7.

(186) Paredes-Gamero, E. J.; Martins, M. N. C.; Cappabianco, F. A. M.; Ide, J. S.; Miranda, A. *Biochim. Biophys. Acta - Gen. Subj.* **2012**, *1820* (7), 1062–1072.

(187) Talukder, P.; Chen, S.; Arce, P. M.; Hecht, S. M. *Org. Lett.* **2014**, *16* (2), 556–559.

(188) Francis, D.; Winn, M.; Latham, J.; Greaney, M. F.; Micklefield, J. *ChemBioChem* **2017**, *18*, 382–386.

(189) Lippitz, M.; Erker, W.; Decker, H.; Holde, K. E. van; Basché, T. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99* (5), 2772–2777.

(190) Rehms, A. A.; Callis, P. R. *Chem. Phys. Lett.* **1993**, *208* (3,4), 276–282.

(191) Palero, J.; Boer, V.; Vijverberg, J.; Gerritsen, H.; Sterenborg, H. J. C. M. *Opt. Express* **2005**, *13* (14), 5363–5368.

(192) Hilaire, M. R.; Ahmed, I. A.; Lin, C.-W.; Jo, H.; Degrado, W. F.; Gai, F. *Submitted* **2017**.

(193) Lee, J. H.; Lee, J. *FEMS Microbiol. Rev.* **2010**, *34* (4), 426–444.

(194) Melander, R. J.; Minvielle, M. J.; Melander, C. *Tetrahedron* **2014**, *70* (37), 6363–6372.

(195) Lee, J.-H.; Wood, T. K.; Lee, J. *Trends Microbiol.* **2015**, *23* (11), 707–718.

(196) Oeltermann, O.; Brand, C.; Engels, B.; Tatchen, J.; Schmitt, M. *Phys. Chem. Chem. Phys.* **2012**, *14* (29), 10266.

(197) Min, A.; Moon, C. J.; Ahn, A.; Lee, J. H.; Kim, S. K.; Choi, M. Y. *Chem. Phys. Lett.* **2016**, *658*, 63–70.

(198) Wilke, J.; Wilke, M.; Brand, C.; Meerts, W. L.; Schmitt, M. *ChemPhysChem* **2016**, 2736–2743.

(199) Meng, X.; Harricharran, T.; Juszczak, L. J. *Photochem. Photobiol.* **2013**, *89* (1), 40–50.

(200) Lami, H.; Glasser, N. *J. Chem. Phys.* **1986**, *84* (2), 597–604.

(201) Callis, P. R. *J. Chem. Phys.* **1991**, *95* (6), 4230–4240.

(202) Callis, P. R. *Methods Enzymol.* **1997**, *278*, 113–150.

(203) Creed, D. *Photochem. Photobiol.* **1984**, *39* (4), 577–583.

(204) Callis, P. R.; Burgess, B. K. *J. Phys. Chem. B* **1997**, *101* (46), 9429–9432.

(205) Chen, Y.; Liu, B.; Barkley, M. D. *J. Am. Chem. Soc.* **1995**, *117*, 5608–5609.

(206) Tine, A.; Valat, P.; Aaron, J. J. *J. Lumin.* **1986**, *36* (2), 109–113.

(207) Zhang, W.; Markiewicz, B. N.; Doerksen, R. S.; Smith III, A. B.; Gai, F. *Phys. Chem. Chem. Phys.* **2016**, *18* (18), 7027–7034.

(208) Lu, Z.; Manias, E.; Macdonald, D. D.; Lanagan, M. *J. Phys. Chem. A* **2009**, *113* (44), 12207–12214.

(209) Arakawa, T.; Kita, Y.; Timasheff, S. N. *Biophys. Chem.* **2007**, *131*, 62–70.

(210) Gasymov, O. K.; Abduragimov, A. R.; Glasgow, B. J. *Anal. Biochem.* **2008**, *374* (2), 386–395.

(211) Roy, S.; Jana, B.; Bagchi, B. *J. Chem. Phys.* **2012**, *136*, 115103–115103.

(212) Yu, H.-T.; Colucci, W. J.; Mclaughlin, M. L.; Barkley, M. D. *J. Am. Chem. Soc.* **1992**, *114*, 8449–8454.

(213) Åberg, A.; Nordlund, P.; Eklund, H. *Nature* **1993**, *361*, 276–278.

(214) von Ballmoos, C.; Wiedenmann, A.; Dimroth, P. *Annu. Rev. Biochem.* **2009**, *78*, 649–672.

(215) Ho, M.; Ménétret, J.-F.; Tsuruta, H.; Allen, K. N. *Nature* **2009**, *459*, 393–399.

(216) Madeo, J.; Mihajlovic, M.; Lazaridis, T.; Gunner, M. R. *J. Am. Chem. Soc.* **2011**,

133, 17375–17385.

(217) Isom, D. G.; Castañeda, C. A.; Cannon, B. R.; Velu, P. D.; García-Moreno E, B. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 16096–16100.

(218) Isom, D. G.; Castañeda, C. A.; Cannon, B. R.; García-Moreno E, B. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 5260–5265.

(219) Mellor, B. L.; Cortés, E. C.; Busath, D. D.; Mazzeo, B. A. *J. Phys. Chem. B* **2011**, *115* (10), 2205–2213.

(220) Aghera, N.; Dasgupta, I.; Udgaonkar, J. B. *Biochemistry* **2012**, *51* (45), 9058–9066.

(221) Sharp, K. A.; Honig, B. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 301–332.

(222) Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. *Biochim. Biophys. Acta* **2006**, *1764* (11), 1647–1676.

(223) Alexov, E.; Mehler, E. L.; Baker, N.; M. Baptista, A.; Huang, Y.; Milletti, F.; Erik Nielsen, J.; Farrell, D.; Carstensen, T.; Olsson, M. H. M.; Shen, J. K.; Warwicker, J.; Williams, S.; Word, J. M. *Proteins Struct. Funct. Bioinforma.* **2011**, *79* (12), 3260–3275.

(224) Warshel, A.; Dryga, A. *Proteins Struct. Funct. Bioinforma.* **2011**, *79* (12), 3469–3484.

(225) Lee, K. K.; Fitch, C. A.; Lecomte, J. T. J.; García-Moreno E, B. *Biochemistry* **2002**, *41* (17), 5656–5667.

(226) Denisov, V. P.; Schlessman, J. L.; García-Moreno E, B.; Halle, B. *Biophys. J.* **2004**, *87*, 3982–3994.

(227) Karp, D. A.; Gittis, A. G.; Stahley, M. R.; Fitch, C. A.; Stites, W. E.; García-Moreno E, B. *Biophys. J.* **2007**, *92*, 2041–2053.

(228) Wu, X.; Lee, J.; Brooks, B. R. *J. Phys. Chem. B* **2016**.

(229) Karp, D. A.; Stahley, M. R.; García-Moreno E, B. *Biochemistry* **2010**, *49*, 4138–4146.

(230) Richman, D. E.; Majumdar, A.; García-Moreno E, B. *Biochemistry* **2015**, *54*, 5888–5897.

(231) Zheng, Y.; Cui, Q. *Proteins* **2017**, *85*, 268–281.

(232) Nguyen, D. M.; Reynald, R. L.; Gittis, A. G.; Lattman, E. E. *J. Mol. Biol.* **2004**, *341* (2), 565–574.

(233) Schlessman, J. L.; Abe, C.; Gittis, A.; Karp, D. A.; Dolan, M. A.; García-Moreno E, B. *Biophys. J.* **2008**, *94*, 3208–3216.

(234) Ernst, J. A.; Clubb, R. T.; Zhou, H. X.; Gronenborn, A. M.; Clore, G. M. *Science* **1995**, *267*, 1813–1817.

(235) Otting, G.; Liepinsh, E.; Halle, B.; Frey, U. *Nat. Struct. Mol. Biol.* **1997**, *4* (5), 396–404.

(236) Yu, B.; Blaber, M.; Gronenborn, A. M.; Clore, G. M.; Caspar, D. L. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (1), 103–108.

(237) Maienschein-Cline, M. G.; Londergan, C. H. *J. Phys. Chem. A* **2007**, *111*, 10020–10025.

(238) Fafarman, A. T.; Sigala, P. A.; Herschlag, D.; Boxer, S. G. *J. Am. Chem. Soc.* **2010**, *132*, 12811–12813.

(239) Patchornik, A. *J. Am. Chem. Soc.* **1970**, *70*, 6969–6971.

(240) García-Moreno E, B.; Dwyer, J. J.; Gittis, A. G.; Lattman, E. E.; Spencer, D. S.; Stites, W. E. *Biophys. Chem.* **1997**, *64*, 211–224.

(241) Harms, M. J.; Schlessman, J. L.; Chimenti, M. S.; Sue, G. R.; Damjanovic, A.;

García-Moreno E, B. *Protein Sci.* **2008**, *17*, 833–845.

(242) Onsager, L. *J. Am. Chem. Soc.* **1936**, *58* (8), 1486–1493.

(243) Adhikary, R.; Zimmermann, J.; Dawson, P. E.; Romesberg, F. E. *Anal. Chem.* **2015**, *87*, 11561–11567.

(244) Gao, G.-Y.; Li, Y.; Wang, W.; Zhong, D.; Wang, S.-F.; Gong, Q.-H. *Chinese Phys. Lett.* **2015**, *32*, 48701.

(245) Laberge, M. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* **1998**, *1386* (2), 305–330.

(246) Jha, S. K.; Ji, M.; Gaffney, K. J.; Boxer, S. G. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 16612–16617.

(247) Goh, G. B.; García-Moreno E, B.; Brooks III, C. L. *J. Am. Chem. Soc.* **2011**, *133*, 20072–20075.

(248) Li, L.; Li, C.; Zhang, Z.; Alexov, E. *J. Chem. Theory Comput.* **2013**, *9* (4), 2126–2136.

(249) Wang, L.; Li, L.; Alexov, E. *Proteins Struct. Funct. Bioinforma.* **2015**, *83* (12), 2186–2197.

(250) Faria, T. Q.; Almeida, Z. L.; Cruz, P. F.; Jesus, C. S. H.; Castanheira, P.; Brito, R. M. M. *Phys. Chem. Chem. Phys.* **2015**, *17* (17), 7255–7263.

(251) Damas, A. M.; Saraiva, M. J. *J. Struct. Biol.* **2000**, *130* (2–3), 290–299.

(252) Hamilton, J. A.; Benson, M. D. *Cell. Mol. Life Sci.* **2001**, *58* (10), 1491–1521.

(253) Hurshman, A. R.; White, J. T.; Powers, E. T.; Kelly, J. W. *Biochemistry* **2004**, *43* (23), 7365–7381.

(254) Johnson, S. M.; Connelly, S.; Fearns, C.; Powers, E. T.; Kelly, J. W. *J. Mol. Biol.* **2012**, *421* (2–3), 185–203.

(255) Gustavsson, Å.; Engström, U.; Westermark, P. *Biochem. Biophys. Res. Commun.* **1991**, *175* (3), 1159–1164.

(256) Fitzpatrick, A. W. P.; Debelouchina, G. T.; Bayro, M. J.; Clare, D. K.; Caporini, M. A.; Bajaj, V. S.; Jaroniec, C. P.; Wang, L.; Ladizhansky, V.; Müller, S. A.; MacPhee, C. E.; Waudby, C. A.; Mott, H. R.; De Simone, A.; Knowles, T. P. J.; Saibil, H. R.; Vendruscolo, M.; Orlova, E. V; Griffin, R. G.; Dobson, C. M. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (14), 5468–5473.

(257) Pauling, L.; Corey, R. B. *Proc. Natl. Acad. Sci. U. S. A.* **1951**, *37* (4), 521–526.

(258) Daggett, V. *Acc. Chem. Res.* **2006**, *39* (9), 594–602.

(259) Yang, M.; Lei, M.; Yordanov, B.; Huo, S. *J. Phys. Chem. B* **2006**, *110* (12), 5829–5833.

(260) Pauling, L.; Corey, R. B. *Proc. Natl. Acad. Sci. U. S. A.* **1951**, *37* (11), 729–740.

(261) Hesselink, F. T.; Scheraga, H. A. *Macromolecules* **1972**, *5* (4), 455–463.

(262) De Santis, P.; Morosetti, S.; Rizzo, R. *Macromolecules* **1974**, *7* (1), 52–58.

(263) Heitz, F.; Detriche, G.; Vovelle, F.; Spach, G. *Macromolecules* **1981**, *14* (1), 47–50.

(264) Armen, R. S.; Alonso, D. O. V; Daggett, V. *Structure* **2004**, *12* (10), 1847–1863.

(265) Armen, R. S.; DeMarco, M. L.; Alonso, D. O. V; Daggett, V. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101* (32), 11622–11627.

(266) Armen, R. S.; Bernard, B. M.; Day, R.; Alonso, D. O. V; Daggett, V. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102* (38), 13433–13438.

(267) Ding, B.; Hilaire, M. R.; Gai, F. *J. Phys. Chem. B* **2016**, *120* (23), 5103–5113.

(268) Torii, H. *J. Phys. Chem. B* **2008**, *112* (29), 8737–8743.

(269) Wu, H.; Canfield, A.; Adhikari, J.; Huo, S. *J. Comput. Chem.* **2010**, *31* (6), 1216–

1223.

(270) MacPhee, C. E.; Dobson, C. M. *J. Am. Chem. Soc.* **2000**, *122* (51), 12707–12713.

(271) Grishina, I. B.; Woody, R. W. *Faraday Discuss.* **1994**, *99*, 245–262.

(272) Nuckolls, C.; Katz, T. J.; Castellanos, L. *J. Am. Chem. Soc.* **1996**, *118* (15), 3767–3768.

(273) Berlepsch, H. Von; Kirstein, S.; Bo, C. *J. Phys. Chem. B* **2003**, *107*, 9646–9654.

(274) Kim, O. K.; Je, J.; Jernigan, G.; Buckley, L.; Whitten, D. *J. Am. Chem. Soc.* **2006**, *128* (2), 510–516.

(275) Borzsonyi, G.; Beingessner, R. L.; Yamazaki, T.; Cho, J. Y.; Myles, A. J.; Malac, M.; Egerton, R.; Kawasaki, M.; Ishizuka, K.; Kovalenko, A.; Fenniri, H. *J. Am. Chem. Soc.* **2010**, *132* (43), 15136–15139.

(276) Usov, I.; Adamcik, J.; Mezzenga, R. *ACS Nano* **2013**, *7* (12), 10465–10474.

(277) Peran, I.; Watson, M. D.; Bilsel, O.; Raleigh, D. P. *Chem. Commun.* **2016**, *52* (10), 2055–2058.

(278) Watson, M. D.; Peran, I.; Raleigh, D. P. *Biochemistry* **2016**, *55* (26), 3685–3691.

(279) Watson, M. D.; Peran, I.; Zou, J.; Bilsel, O.; Raleigh, D. P. *Biochemistry* **2017**, *56*, 1085–1094.

(280) Ranløv, I.; Alves, I. L.; Ranløv, P. J.; Husby, G.; Costa, P. P.; Saraiva, M. J. M. *Am. J. Med.* **1992**, *93* (1), 3–8.