2019

The Development Of Unnatural Amino Acid-Based Probes And Methods For Biological Studies

Ismail Abd Al Azim Ahmed
University of Pennsylvania, i.ahmed1988@gmail.com

The Development Of Unnatural Amino Acid-Based Probes And Methods For Biological Studies

Abstract
Proteins form a diverse ensemble of dynamic structures to carry out all life-sustaining functions. Therefore, many efforts have gone into studying the structure-dynamics-function relationship of proteins using a wide range of techniques, including fluorescence and infrared (IR) spectroscopies. While very useful, intrinsic fluorescence and IR signals arising from the natural amino acid side chains within the protein are often insufficient or unable to provide the information needed to understand the biological question of interest. To this end, various extrinsic spectroscopic probes, such as fluorescent dyes, have been used to increase the information content in specific measurements and applications. However, incorporation of a foreign moiety into any protein unavoidably affects its native structure and dynamics; hence effort must be made to reduce such perturbation. In this regard, the overarching aim of this thesis is to develop novel spectroscopic probes based on scaffolds of natural amino acids (NAAs). Because of their small size and similarity to NAAs, such unnatural amino acid-based (UAA-based) probes are expected to be minimally perturbing. Specifically, we show that (1) 4-cyanotryptophan (4CN-Trp) is a blue fluorescent amino acid useful for fluorescence microscopy applications; (2) 4CN-Trp and DiO (a common dye used to stain membranes) are a useful FRET pair to study peptide-membrane interactions; (3) 4CN-Trp, and tryptophan constitutes a dual FRET-PET pair which was used to study peptide end-to-end termini interactions and protein ligand-binding; and (4) the functional group of 4CN-Trp, 4-cyanoindole can be used in the form of a nucleoside as a dual fluorescence-IR reporter for DNA-protein studies. Furthermore, we extended applications of previously known UAAs and showed (5) p-cyanophenylalanine is useful as a fluorescence-based pH sensor which we used to determine peptide pKa's and peptide membrane penetration kinetics and (6) we use a simple synthetic method for post-translationally installing an ester moiety on to proteins via cysteine alkylation as an UAA-based vibrational probe in proteins to study fibril formation and protein-ligand interactions

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Biochemistry & Molecular Biophysics

First Advisor
Feng Gai

Second Advisor
Sergei Vinogradov

Keywords
fluorescence, GFP, nucleoside, probe, site-specific, unnatural amino acid

Subject Categories
Biophysics | Chemistry

THE DEVELOPMENT OF UNNATURAL AMINO ACID-BASED PROBES AND

METHODS FOR BIOLOGICAL STUDIES

Ismail A. Ahmed

A DISSERTATION

in

Biochemistry and Molecular Biophysics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2019

Supervisor of Dissertation

_____

Feng Gai Ph.D.

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

Graduate Group Chairperson

_____

Kim A. Sharp Ph.D.

Associate Professor of Biochemistry and Biophysics

Dissertation Committee

Sergei Vinogradov Ph.D., Professor of Biochemistry and Biophysics

P. Leslie Dutton Ph.D., Eldridge Reeves Johnson Professor of Biochemistry and Biophysics

S. Walter Englander Ph.D., Professor of Biochemistry and Biophysics

Lin Guo Ph.D., Assistant Professor of Biochemistry and Molecular Biology, Jefferson University

Dedicated to My Beloved Family

ACKNOWLEDGMENT

Firstly, I would like to thank my Ph.D. advisor Dr. Feng Gai for the opportunity to join his lab. He welcomed me at a time when the doors of scientific research seemed to be closing in on me. He challenged me, he inspired me, and he supported me. He taught me to be rigorous, always ask questions, and to work hard¬–because like Kevin Durant always says, "hard work beats talent when talent fails to work hard."

I thank my thesis committee Dr. Sergei Vinogradov, Dr. P. Leslie Dutton, Dr. S. Walter Englander, and Dr. Lin Guo. It was an absolute pleasure having them serve on my committee. Our interactions were always pleasant, discussions always insightful, and they always supported me in any way I needed. Specifically, I wanted to recognize that Dr. Dutton has been a huge advocate for me since joining Penn and has helped me in more ways than I can enumerate. I am forever indebted.

I thoroughly enjoyed my time in the Gai lab where I constantly shared science and laughter with my lab mates. I appreciated working with Mary Rose Hilaire on developing a blue fluorescent unnatural amino acid which served as the foundation for my Ph.D. thesis work. Our abilities perfectly complemented each other to bring something truly beautiful to life. In the last two years of my Ph.D., the lab significantly decreased in size, but that only brought us (Debopreeti Mukherjee and Arusha Acharyya) closer. We supported each other, and we were there for each other in good and bad times. It certainly has been a pleasure. One of the highlights of my graduate experience was serving as a mentor for Christina Eng, an undergraduate researcher in our lab. We shared our stories, laughs, and overall had a great time. Being a mentor taught me to be more responsible and to be selfless. I am so proud of her accomplishments, and I am looking forward to seeing what she does in the future.

Dr. Thomas Troxler has served as a second mentor to me in the lab. He read all my grant proposals and gave me insightful feedback. He patiently fielded and answered all my questions in spectroscopy and physics. He helped me through various experiments and overall gave me valuable perception into the world of academia.

characteristics helped me through graduate school. Momma, look at me now! It is all because of you.

I cannot express enough the gratitude that I have for my dear wife, Jamila Hoque. She came into my life at a time where I had no idea if I could continue in science. She helped me stay the course and made me believe in myself through her unyielding belief in me and my dream. She grounded me; put up with the demanding hours I spent lab, loved and supported me unconditionally. We grew our young family together here at Penn, I will never forget this glorious chapter, and I am excited about our next chapters we will write together.

ABSTRACT

THE DEVELOPMENT OF UNNATURAL AMINO ACID-BASED PROBES AND

METHODS FOR BIOLOGICAL STUDIES

Ismail A. Ahmed

Dr. Feng Gai

Proteins form a diverse ensemble of dynamic structures to carry out all life-sustaining functions. Therefore, many efforts have gone into studying the structure-dynamics-function relationship of proteins using a wide range of techniques, including fluorescence and infrared (IR) spectroscopies. While very useful, intrinsic fluorescence and IR signals arising from the natural amino acid side chains within the protein are often insufficient or unable to provide the information needed to understand the biological question of interest. To this end, various extrinsic spectroscopic probes, such as fluorescent dyes, have been used to increase the information content in specific measurements and applications. However, incorporation of a foreign moiety into any protein unavoidably affects its native structure and dynamics; hence effort must be made to reduce such perturbation. In this regard, the overarching aim of this thesis is to develop novel spectroscopic probes based on scaffolds of natural amino acids (NAAs). Because of their small size and similarity to NAAs, such unnatural amino acid-based (UAA-based) probes are expected to be minimally perturbing. Specifically, we show that (1) 4-cyanotryptophan (4CN-Trp) is a blue fluorescent amino acid useful for fluorescence microscopy applications; (2) 4CN-Trp and DiO (a common dye used to stain membranes) are a useful FRET pair to study peptide-membrane interactions; (3) 4CN-Trp, and tryptophan constitutes a dual FRET-PET pair which was used to study peptide end-to-end termini interactions and protein ligand-binding; and (4) the functional group of 4CN-Trp, 4-cyanoindole can be used in the form of a nucleoside as a dual fluorescence-IR reporter for DNA-protein studies. Furthermore, we extended applications of previously known UAAs and showed (5) *p*-cyanophenylalanine is useful as a fluorescence-based pH sensor which we used to determine peptide pKa's and peptide membrane penetration kinetics and (6) we use a

simple synthetic method for post-translationally installing an ester moiety on to proteins via cysteine alkylation as an UAA-based vibrational probe in proteins to study fibril formation and protein-ligand interactions

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

xiv

1: INTRODUCTION

1.1: PROTEIN STRUCTURE AND FUNCTION IN BIOLOGY

According to the central dogma of biology, the genetic code which is made of DNA will encode RNA, which is subsequently translated into proteins. These proteins serve as one of the fundamental building blocks and drivers of life which are involved in essentially every cellular process. For example, proteins serve as enzymes that catalyze the chemical reactions that break down food (Rothman et al., 2002), as transporters of molecules such as oxygen throughout the body(Mairbäurl and Weber, 2012), to antibodies that are essential for immune responses in the fight against invading viruses (Burton, 2002). Proteins are made up of chains of amino acids (primary structure) linked together by covalent (peptide) bonds in combinations which are dictated by the genes that encode them. There are 20 naturally occurring amino acids that differ from the chemical composition of their side chains (functional groups). The functional groups of the amino acids within a protein chain will spontaneously interact with each other to give rise to secondary structures (α-helices and β-sheets) and consequently fold into a three-dimensional protein structure (tertiary structure). Owing to the diversity of the canonical amino acids and vast combinations of possible amino acid sequences the structural possibilities for proteins are endless. For this reason, we witness a great diversity in protein structure & functions (Lee et al., 2007; Moult and Melamud, 2000) which exemplifies the common overarching theme of "structure determines function" in biology.

While proteins play so many roles pivotal to maintaining life, sometimes mutations in the genes that encode them or other factors such as environmental stress can cause proteins to misfold (Englander, 2007). Even the slightest misfolding of a protein can cause deleterious effects which manifest in the form of disease (Hartl, 2017; Scheckel and Aguzzi, 2018). A classic example of this is in sickle-cell disease, where a single point mutation in the gene that encodes hemoglobin, a protein that transports oxygen, causes a major change in the structure of the protein (Kato et al., 2018; Pace et al., 2012). This change in structure promotes aggregation of

hemoglobin causing red blood cells to take a sickle shape and thus decreases the cells ability to transport blood throughout the body leading to oxygen deficiency in a patient (Kato et al., 2018; Pace et al., 2012). Other examples of protein misfolding are ones that form aggregates called amyloid fibrils, which are typically associated with diseases such as Alzheimer's, Parkinson's, and type II diabetes. In these cases, when the proteins misfold into amyloids, the functions of the proteins are lost. To deal with this issue, there are molecular quality control mechanisms within a cell to deal with protein misfolding such as chaperone proteins which will attempt to refold the misfolded protein correctly or other mechanisms where the cell detects and destroy defective protein (Saibil, 2013). Despite these biological quality control mechanisms, sometimes protein misfolding cannot be overcome and will eventually lead to some form of disease, and potentially cell death.

To this end, there have been many, and still ongoing, efforts to determine the structure and function of physiological and pathological proteins alike using various – in vitro as well as in vivo - experimental methods. The majority of these methods rely on some variation of imaging (microscopy) and/or measurements of energy absorption or emission (spectroscopy). Each of these techniques utilizes some form of electromagnetic radiation; summarized in **Table 1.1** various common methods used to study proteins (and other biomolecules) are outlined on the electromagnetic spectrum and matched with each type of electromagnetic radiation it uses. Each of these techniques has its advantages and disadvantages. For example, crystallography can provide the structure of a protein at atomic resolution but is only a static snap-shot of the "dynamic" protein which lacks temporal resolution and cannot be done in the proteins native environment, within the cell (Acharya and Lloyd, 2005; Shi, 2014). Conversely, infrared spectroscopy can be utilized to study proteins with a high temporal resolution on the femtosecond and picosecond timescale, but lack structural resolution (Barth, 2007a; Cho*, 2008; Hunt, 2009; Thielges and Fayer, 2012). However, a great amount of information about a system (proteins) can be obtained by using these methods separately or in combination. The combination of

2

experimental methods can paint us a detailed picture of the structure, function, and dynamics of proteins and other biomolecules.

1.2: FLUORESCENCE MICROSCOPY

One of the best methods to track proteins within the cellular environment is using microscopy (Thorn, 2016). The simplest form of optical microscopy is bright-field microscopy, where the sample is illuminated via transmitted white light from below and observed from above. The image has the typical appearance of a dark sample on a bright background, hence the name. While this method is simple to use and requires minimal sample preparation, its obvious limitation is the low contrast of most biological samples. However, staining methods can be used to increase the contrast of the sample versus background for more defined images. Nevertheless, bright-field imaging has generated a plethora of new knowledge and has been used to visualize and image from tissue to individual cells and organelles with distinct morphologies such as the nucleus and endoplasmic reticulum. However, to distinguish particular and/or multiple components within the cell such as proteins is challenging using bright-field microscopy.

Staining samples or intrinsic color arising from the sample itself, like in the case of chlorophyll can enhance the contrast of the sample. However, most staining methods are either not specific or too specific, as in "everything" or just "one component" is labeled. A method developed to overcome some of the listed limitations is fluorescence microscopy (Combs, 2010; Lichtman and Conchello, 2005; Thorn, 2016). Fluorescence microscopy uses an optical microscope to detect fluorescence instead of, scattering, reflection, and absorption. To use fluorescence microscopy the objects of interest must fluoresce (Combs, 2010; Lichtman and Conchello, 2005; Thorn, 2016). Fluorescence is the emission of light that occurs within nanoseconds after a molecule absorbs light. Briefly, as seen in **Figure 1.1**, when a molecule absorbs a photon of light, the molecule is promoted form its ground electronic state to a high-energy electronically-excited state, usually the first electronic state (Lakowicz, 1999). Then in solution, the molecule will typically relax to its lowest vibrational state and then could

subsequently relax to its ground state (**Figure 1.1**). Vibrational relaxation and fluorescence emission are the chief ways a fluorophore returns to its low-energy ground state, but other potential concurrent events can occur such as internal conversion and/or intersystem crossing (**Figure 1.1**). Typically the fluorescence emission spectrum of a given molecule is of shorter wavelength than its absorbance spectrum. The difference between the maxima of the exciting and emitted wavelengths is known as the Stokes shift which arises from the energy loss in between the time a photon is absorbed and when a new one is emitted (Lakowicz, 1999, Stokes, 1852). The Stokes shift is the critical property that makes fluorescence so powerful because it allows for completely filtering out the exciting (scattering) light without blocking the emission fluorescence (Lakowicz, 1999). Therefore, it is possible to see only the objects/molecules that are fluorescent. This approach for increasing contrast of a sample is superior to bright-field and phase-contrast methods where objects/molecules are stained with agents that absorb light. Additionally, it is important to note that the use of fluorescence is compatible with many versions of microscopy (Combs, 2010; Lichtman and Conchello, 2005; Thorn, 2016), each having their distinct advantages, from epifluorescence microscopy, 2-photon microscopy, confocal microscopy, super-resolution, expansion microscopy, light-sheet microscopy and more.

1.3: FLUOROPHORES FOR MICROSCOPY

To use fluorescent microscopy to image an object/molecule such as proteins, it must contain a fluorophore (molecules used by virtue of their fluorescent properties). Generally, the outermost electron orbitals (HOMO/LUMO) of the fluorophore molecule governs both its fluorescent properties such the wavelengths of absorption and emission as well as fluorescence efficiency (quantum yield) which is essentially the ratio of the number of photons emitted by fluorescence relative to the number of photons it has absorbed (Lakowicz, 1999). In the case of proteins, they can contain one or more of the three naturally occurring amino acids which are fluorescent, phenylalanine, tyrosine, and, tryptophan (Lakowicz, 1999; Suzuki et al., 2009). All three of these amino acids absorb and fluoresce in the UV region of the spectra, with

absorbance/emission maxima of 260/280 nm for phenylalanine, 275/300 nm for tyrosine, and 280/350 nm for tryptophan (**Table 1.2**). Additionally, these naturally occurring fluorophores have relatively low quantum yield such that the brightest of them, tryptophan, has a quantum yield of 0.14 (**Table 1.2**). These features make them less suitable for fluorescence microscopy which is generally optimized for bright fluorophores which emit in the visible region of the spectra (ca. 380 – 740 nm). To this end, to image proteins using fluorescence microscopy, a fluorophore must be appended to the protein of interest.

Many organic molecules have intrinsic fluorescence (autofluorescence), and like the abovementioned amino acids, few are useful for specific labeling of proteins and other components in biological systems using fluorescence microscopy applications (Lavis and Raines, 2014; Martynov et al., 2016). The typical approach to fluorescence microscopy is to take advantage of synthesized compounds with ideal fluorescent properties. Characteristically, these compounds commonly have some degree of conjugated double bonds and are often aromatic (have ring structures) with pi bonds that distribute the outer orbital electron density over a wide area (Lakowicz, 1999; Lavis and Raines, 2014; Lichtman and Conchello, 2005). These types of structures are optimal for fluorescence microscopy because the energy differences between the excited state and ground state orbitals are small enough that relatively low-energy photons of visible light can be used to excite electrons into their excited states. In general, the more conjugated bonds the molecule has, the required excitation energy can be lower and therefore redder (longer) the excitation wavelength can be. Additionally, the fluorescence efficiency of a fluorophore, as measured by fluorescent quantum yield, generally increases with the number of pi bonds (Lakowicz, 1999; Lichtman and Conchello, 2005). Using these general rules, synthetic chemists have developed many thousands of fluorescent probes that provide a means of labeling many aspects of biological systems including individual proteins.

Although synthetic fluorophores have been vital for studying a wide range of cell biological questions, there are some drawbacks such as delivery and specific-labeling in the cell,

especially for in-vivo experiments. While there are approaches to address these issues, the discovery of the green fluorescent protein (GFP) was revolutionary for biological research (Berezin and Achilefu, 2010; Shaner et al., 2005; Tsien, 1998). This discovery ushered in the development of a large family of fluorescent proteins with various characteristics (e.g., different colors, photo-stability, quantum yield, etc.) that can simply be genetically encoded as fluorescent tags onto the proteins of interest. For this reason, fluorescent proteins are by far the most widely used fluorophores in biological microscopy (Chudakov et al., 2010; Day and Davidson, 2009). Examples have enabled the selective labeling and analysis of single proteins, cellular organelles, and even whole cells and have been used to study gene expression, protein localization (protein-protein co-localization), protein trafficking, engineered into fluorescent biosensors, and more. Key features of fluorescent proteins that fuel its use are: (1) high extinction coefficients, (2) relatively high quantum yields, (3) multiple colors in the visible region of the spectra (4) able to be genetically encoded, and (5) can be engineered for new properties.

Fluorescent proteins derive their fluorescence from the formation of intrinsic chromophores (fluorophores). For example, GFP forms its chromophore from a 3 amino acid sequence Ser65-Tyr66-Gly-67 and fluorescence properties of GFP can be tuned by mutations throughout the protein (Tsien, 1998). For example, such mutations have enabled the generation of various derivatives with better photophysical properties and/or different colors. In combining the use of various colored fluorescent proteins enable multiplexing (e.g., labeling several different components at the same time) and the application of fluorescence resonance energy transfer (FRET) (Berezin and Achilefu, 2010; Chudakov et al., 2010; Day and Davidson, 2009; Shaner et al., 2005). Despite all these highly desirable features ideal for fluorescence microscopy, fluorescent proteins require maturation of the chromophore and protein. This is because the chromophore has to be generated by post-translational reactions and must also be sequestered into the hydrophobic barrel structure of the protein for it to fluoresce (it is relatively non-fluorescent in a hydrophilic environment). Therefore, this imposes a time constraint on the order of hours to see the appearance of its fluorescence. For example, following a protein from the time

of translation would be limited by the time it takes for the fluorescence protein to maturate. Another caveat is that fluorescent proteins can at times be too large (>25kDa) for the proteins they are labeling and may cause perturbations in native structure resulting in altered or deleterious effects to the proteins native function.

1.4: UNNATURAL AMINO ACIDS FOR FLUORESCENCE MICROSCOPY

Both synthetic fluorophores and fluorescent proteins have their respective advantages and disadvantages. However, one can imagine extracting the advantages of both and combining them. This is afforded in the form of unnatural amino acids. Unnatural amino acids are essentially amino acids with chemically unique side chains, which in this case would be a fluorophore. Conceptually, in essence, this is the addition of a fluorophore to an amino acid backbone. The use of unnatural amino acids merges the advantageous photophysical properties of synthetic fluorophores with the ability to be directly incorporated into the sequence of a protein similar to tagging a protein of interest with fluorescent proteins. To this end, various efforts lead to the generation of visibly fluorescent unnatural amino acids such as dansylalanine (Summerer et al., 2006), L-(7-hydroxycoumarin-4-yl) ethylglycine (Wang et al., 2006), acridon-2-ylalanine (Speight et al., 2013), and others which are compatible with fluorescence microscopy. Furthermore, this concept has driven the expansion of the genetic code to incorporate fluorescent and other unnatural amino acids through amber codon suppression, a method pioneered by Peter Schultz (Chatterjee et al., 2013a; Lang and Chin, 2014; Liu et al., 1997; Summerer et al., 2006; Wang et al., 2006; Xie and Schultz, 2006).

1.4: UNNATURAL AMINO ACIDS AND BIOLOGICAL SPECTROSCOPY

To date, unnatural amino acids have found various utilities in biological research in addition to being amino acid based-fluorophores. Some examples of this include photo-controlled amino acids (Hoppmann et al., 2014, 2015; Peters et al., 2009) and as handles for molecular modifications, including dyes (Kim et al., 2013; Patil and Luzzio, 2017; Shieh and Bertozzi, 2014;

Tyagi and Lemke, 2013). In this regard, the possibilities for unnatural amino acids and their applications are vast as they are limited only by the functional group diversity available through synthetic chemistry. In addition to being useful for biological microscopy, unnatural amino acids provide functional groups that are spectroscopically unique. Therefore, they can be leveraged as useful probes for biological spectroscopy (Goldberg et al., 2012; Ma et al., 2015; Summerer et al., 2006; Waegele et al., 2011a) which include fluorescence spectroscopy, Fourier-transform infrared spectroscopy (FTIR), nuclear magnetic resonance spectroscopy (NMR), Raman spectroscopy, and more. For the purpose of this thesis, we will primarily focus on developing and using unnatural amino acids in the context of fluorescence and FTIR spectroscopies.

The structural integrity of a folded protein is maintained by many weak interactions and, as a result, it may be subject to local and global conformational fluctuations (or folding and unfolding) (Kim and Baldwin, 1990), and its thermodynamic stability may vary across regions or domains. Also, a protein may be intrinsically disordered, sampling a dynamic conformational ensemble without forming well-defined and long-lasting tertiary contacts (Garcia-Mira et al., 2002; Tompa, 2012). Furthermore, to carry out its function, a protein may need to transiently vary its conformation, either locally or globally, in response to a functional stimulus (e.g., a binding event or change in an environmental property). Therefore, to provide a holistic description of the conformational energy landscape of the protein system in question, either from a protein folding or functional point of view, we need experimental methods that can characterize protein structure, environment, dynamics and stability in a site-specific manner. While NMR spectroscopy is perhaps the method of choice in this regard, its required expertise, cost and sample requirement often limit its widespread use in the field. Alternatively, commonly used biophysical techniques, including infrared (IR) and fluorescence spectroscopies, are useful for assessing various protein structural and thermodynamic properties with site specificity.

In addition to circular dichroism (CD) (Greenfield, 2006), the most commonly used method to determine protein secondary structure and stability is IR. The amide I band, which

arises from protein's amide backbone units, is widely used as an IR reporter of protein structure and stability (Barth, 2007a; Yang et al., 2015). However, the structural sensitivity of these intrinsic protein spectroscopic signals originates from the dependence of the underlying vibrational couplings across multiple sites on backbone configurations and, therefore, like CD they cannot be used to reveal environmental or structural details at a specific amino acid site. Therefore, the information obtained is rather global. One intrinsic spectroscopic signal that does offer this ability is tryptophan fluorescence. Nevertheless, the less desirable photophysical properties of tryptophan often limit its practical utility in protein biophysical studies.

To increase the information content obtainable via either IR or fluorescence spectroscopy, one strategy is to introduce one or multiple external labels into the protein system of interest that possess unique spectroscopic properties that are not only distinguishable from the rest of the protein but are also dependent on the local environment (Ma et al., 2015; Waegele et al., 2011b). However, besides the isotope-editing strategy used in IR spectroscopy, incorporation of a foreign moiety into the protein sequence in question could induce significant and undesirable perturbations in the native structure and function. Therefore, recently there has been an increased interest in the development of unnatural amino acid-based spectroscopic probes and their applications to study various biochemical and biophysical questions in a site-specific manner. This is because unnatural amino acid can be easily incorporated into proteins using either chemical or genetic methods and, perhaps more importantly, when replacing a native amino acid of similar size and structure, the potential structural perturbation by an unnatural amino acid can be minimized. Site-specific probes are useful to report on a single location within a protein to reveal structural and dynamic information. Characteristics of a useful probe for protein spectroscopy is one that has (1) large signal (brightness for fluorescence and a large extinction coefficient for IR) so that large amounts of sample is not required for measurements; (2) an absorption and emission that does not overlap with regions devoid of intrinsic protein absorption and emission; (3) is small in size, such that it is minimally perturbing to protein

structure; (4) is incorporable into protein systems (chemically or biologically); and is sensitive to changes of a particular property of proteins, i.e., structure, environment, etc.

Below, we describe the approach of using unnatural amino acid-based spectroscopic probes, with a focus on their utility to reveal site-specific environmental, dynamic, structural or stability information via commonly used spectroscopic techniques, including fluorescence and IR spectroscopies.

1.5: PROTEIN FLUORESCENCE SPECTROSCOPY

Fluorescence spectroscopy is widely used in the study of protein structure and stability. As stated previously, this is due in part to the fact that the three aromatic amino acids (i.e., tryptophan, tyrosine, and phenylalanine) are fluorescent when excited with UV light and, depending on their specific location within a protein, their fluorescence properties can change (Lakowicz, 1999). Below we briefly give an overview of fluorescent properties that are commonly exploited when studying proteins using fluorescence spectroscopy.

(1) Stokes Shift − As previously described (section 1.2), when a fluorophore absorbs a photon of light, it is promoted form its ground electronic state to a high-energy electronically-excited state (usually the first electronic state). Then in solution, the molecule will typically relax to its lowest vibrational state and then could subsequently relax to its ground state (**Figure 1.1)**. There are several pathways for any molecule to return to its low-energy ground state, in the case of a fluorophore, fluorescence emission is most prevalent. The resultant emission spectrum of a fluorophore is generally of shorter wavelength than the absorbance spectrum of the same fluorophore and is a mirror of its corresponding absorption spectrum (Franck–Condon principle) (Lakowicz, 1999). The difference between the maxima of the exciting and emitted wavelengths is known as the Stokes shift which arises from the energy loss in between the time a photon is absorbed and when a new one is emitted. This difference between the absorption and emission energy for a given fluorophore can be approximated using the Lippert Equation (Lakowicz, 1999):

$$v_\text{A} - v_\text{F} = \frac{2}{hc}\left(\frac{\varepsilon-1}{2\varepsilon+1} - \frac{n^2-1}{2n^2+1}\right)\frac{\left(\mu_\text{e} - \mu_\text{g}\right)^2}{a^3} + \text{constant} \,,$$

(**1.5.1**)

Where $v_\text{A} - v_\text{F}$ represent the absorbtion 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 Plank's constant, and c is the speed of light.

In general, the Stokes shift also referred to as spectral shift is the result of two possible events after the fluorophore is excited from the ground state to the excited state. The first being vibrational relaxation or dissipation and the second being solvent reorganization. Following photo-excitation, there is a rapid shift in the dipole moment of the fluorophore, and if the fluorophore is in a polar solvent (such as water), solvent molecules must reorganize to stabilize the excited state. The stabilization of the excited state results in a red-shift of the fluorescence (spectral shift). This stabilization will be far less in a more nonpolar environment such as in a nonpolar solvent or the interior of a protein and thus result in less of a red-shift. For example, tryptophan blue-shifts (Chen and Barkley; Vivian and Callis, 2001) relative to water in apolar solvents (hydrophobic environment) which is a manifestation of the less stabilizing effect of a less polar solvent on the fluorophore's excited state, whose permanent dipole moment is smaller from the corresponding ground-state value. In the cases where the fluorophore is red-shifted the permanent dipole moment is larger from the corresponding ground-state value. In addition to solvent (local environment), pH, temperature, and interactions between the fluorophore and surrounding solvent molecules such as hydrogen bonding can also affect its spectral shift (Lakowicz, 1999). Therefore, the maxima of the fluorescence emission peak (spectral shift) are telling of the fluorophores surrounding environment. This property has been exploited to measure protein-molecule interactions, protein-protein interactions, protein conformational change, local hydration and more.

(2) Fluorescence Intensity – Similar to the Stokes shift, changes in fluorescence intensity can be used to assess protein-molecule interactions, protein-protein interactions, protein conformational change, local hydration and more. This is because the fluorescence intensity of a given fluorophore can be affected by its solvent, pH, ion concentration, temperature, viscosity, and more (Lakowicz, 1999). Currently, there is no central rule to predict how solvent properties will affect the fluorescence intensity of a given fluorophore. It is molecule dependent, and in order to establish these trends, solvent studies with individual fluorophores are necessary. For example, in addition to blue-shifting its maximum emission in a hydrophobic environment, tryptophan will also increase its fluorescence intensity in an apolar environment. Each dye is unique to its chemical structure, and therefore the emission of some fluorophores is not significantly affected by solvent (local environment). This essentially depends on the interactions of the dye with the molecules of the solvent such as hydrogen bonding or lack thereof. In this regard fluorophores that have emission with significant dependence on solvent are most useful.

(3) Fluorescence Lifetime Decay – The fluorescence lifetime is another property useful in fluorescence-based studies (Lakowicz, 1999). The fluorescence lifetime is the average time a fluorophore stays in its excited state before emitting a photon. Solvent (local environment) can affect the fluorescence lifetime of many fluorophores and can be used in the same manner as the Stokes shift and fluorescence intensity of a fluorophore. The fluorescence lifetime, τ, can be measured from the fluorescence intensity decay, $I(t)$, which is measured as a function of time and can be modeled by (Lakowicz, 1999):

$$I(t) = \sum_{i=1}^{n} A_i e^{-\frac{t}{\tau_i}}, \tag{1.5.2}$$

where A_i represents the amplitude of the fluoresence of τ_i of the ith component. In practice, there is a repetitive excitation of the fluorophore. The fluorescence decay of the fluorophore is plotted versus time and subsequently by fitting the fluorescence decay reveals the fluorescence

lifetime(s) of a given fluorophore. Changes in fluorescence lifetime can be monitored to assess protein-molecule interactions, protein-protein interactions, protein confomational change, local hydration and more.

(4) Fluorescence energy transfer (FRET) − FRET is a mechanism for energy transfer between fluorophores (Clegg, 1996; Förster, 1948, 1965). A donor fluorophore, initially in its electronic excited state, will transfer energy to an acceptor chromophore through non-radiative dipole-dipole coupling. The efficiency (*E*) of this energy transfer is inversely proportional to the sixth power of the distance between donor and acceptor. This distance-dependent property makes FRET sensitive to small changes in the distance between the donor and acceptor fluorophores. This is described in the following equation (Lakowicz, 1999):

$$E = \frac{1}{1+\left(\frac{r}{R_0}\right)^6} \, , \qquad\qquad (1.5.3)$$

where r is the donor-to-acceptor distance and R_0 is the Förster distance of the donor and acceptor pair, which is the distance at which the energy transfer efficiency is 50%. The R_0 is calculated using the following equation (Lakowicz, 1999):

$$R_0^6 = \left(\frac{9000(ln10)\kappa^2 Q_D}{128\pi^5 N \eta^4}\right) J(\lambda) \, , \qquad\qquad (1.5.4)$$

where κ_2 is the orientation factor, assumed to be 2/3 in an isotropic environment, Q_D is the fluorescence quantum yield of the donor in the absence of the acceptor, *N* is the Avogadro's number, η is the refractive index of the medium, and *J(λ)* is the spectral overlap integral of the donor emission and the emission of the acceptor. Therefore, for FRET to occur, the emission spectrum of the donor must overlap with the absorption spectrum of the acceptor. This is determined in the following equation (Lakowicz, 1999):

$$J(\lambda) = \int_0^\infty F_D(\lambda)\varepsilon_A(\lambda)\lambda^4 d\lambda,$$ (**1.5.5**)

where $F_D(\lambda)$ is the area-normalized emission spectrum of the donor and $\varepsilon_A(\lambda)$ is the wavelength-dependent molar absorption coefficient of the acceptor.

FRET experiments are generally designed in a way that the donor is one color and the acceptor is another. When they are close together, within the Förster distance, the acceptor will fluoresce via the emission of the donor through energy transfer. However, as the donor and acceptor separate the donor will emit. Using different colors allow for separation (using filters) of the signal from two different sources. Because of this, FRET has become very useful to study conformational changes of proteins and protein interactions (by proximity) within a range of 10 to 100 Å. FRET is used in all spheres of biological research (Lakowicz, 1999). It is commonly exploited to build protein-based biosensors for various analytes (Giuliano and Taylor, 1998; Wang et al., 2018). When the sensor does not interact with its analyte, it fluoresces in one color, and then upon binding its analyte, it fluoresces in another due to a conformational change in the sensor to separate the fluorophores. For example, a commonly exploited fluorescent protein FRET pair is CFP and YFP (Shaner et al., 2005).

(5)Fluorescence Quenching – Quench is a process where the fluorescence intensity of a fluorophore is deceased via multiple mechanisms. There are several mechanisms for fluorescence quenching which are broadly classified into static or dynamic quenching. Fluorescence quenching requires a quencher molecule. In a static quenching process, before excitation, the quencher will form a complex with the fluorophore that is non-fluorescent. This causes a decrease in the fluorescence intensity but not the fluorescence lifetime. Conversely, a dynamic quenching process occurs upon the interaction of an excited fluorophore and quencher molecule via diffusion. One form of dynamic quenching is collisional quenching where the fluorescence intensity and lifetime of the fluorophore is decreased through collisions of the fluorophore with the quencher. A common collision quenching process is electron transfer (ET),

where the process of quenching occurs when an excited electron is transferred from a donor (quencher) to an acceptor fluorophore (Lakowicz, 1999; Lakowicz et al., 1994). The distance-dependent quenching rate through ET can be adequately described by the following equation (Lakowicz, 1999):

$$k_{\mathrm{Q}}(r) = k_0 e^{-\beta(r-a_0)} \,, \tag{1.5.6}$$

where r is the fluorophore-quencher separation distance, k_0 is the quenching rate when the pair are in van der Waals contact (when $r = a_0$), and β is a constant characteristic of the particular fluorophore-quencher interaction. While the rate of FRET depends on $1/r^6$, where r is the separation distance between the fluorophore and modulator (i.e., acceptor in FRET and quencher in ET), the rate of ET shows an exponential dependence on r (Lakowicz, 1999). Therefore, the quencher must be close to modulate the fluorophores fluorescence intensity and lifetime. This property is useful to measure small distant contacts (r<20 Å) and interactions between dyes to assess conformational changes, protein-ligand interactions and more (Doose et al., 2009; Lakowicz, 1999; Lakowicz et al., 1994).

1.6: UNNATURAL AMINO ACIDS FOR FLUORESCENCE SPECTROSCOPY

While very useful, the intrinsic protein fluorescence arising from the natural amino acids suffer from various limitations. For example, the practical utility of tryptophan as a fluorescence probe is often limited by the following factors (Adams et al., 2002; and Barkley, 1998; Paul D. Adams et al., 2002; Petrich et al., 1983; Yu Chen et al., 1996): (1) its fluorescence cannot be selectively excited when more than one tryptophan or other aromatic residues are present, (2) it has a relatively low fluorescence quantum yield (QY), (3) its fluorescence spectrum and QY only exhibit a modest dependence on environment, and (4) its excited-state decay kinetics are complex and can be affected by many different mechanisms. Therefore, great efforts have been

devoted to developing unnatural amino acid-based fluorophores that possess considerable improved photophysical properties and hence offer enhanced utility over those of naturally occurring fluorescent amino acids. While one approach to building fluorescent unnatural amino acids is to append an amino acid backbone to a known organic dye, those at times are not ideal. Therefore, these extrinsic fluorophores in this study are simple derivatives of tryptophan, phenylalanine or tyrosine, hence having the smallest size in comparison to those commonly used bulky organic dyes or fluorescent proteins. This characteristic can be quite advantageous in applications where any potential structural perturbation should be minimized. Below we discuss factors concerning the selection of fluorescent unnatural amino acid probes.

(1) Size and structure – Incorporation of any unnatural amino acid will unavoidably perturb the structure and/or other property of the protein system in question. The question to ask is in a given experiment is, will the unnatural side chain be structurally perturbative? For this reason, all of the fluorescent unnatural amino acids we employ are simple derivatives of naturally occurring amino acids such as tryptophan, tyrosine, or phenylalanine. Therefore, ideally, they should replace their respective native residues. If that is not possible, the chosen unnatural amino acid should be placed at a location where it is unlikely to induce a large structural change or impede the native function.

(2) Brightness –For an unnatural amino acid to be useful as a fluorescent probe it must have a certain brightness (B) which is the product of the molar extinction coefficient (ε) at the excitation wavelength (λ_{ex}) and the corresponding QY. For applications where a low protein concentration is required, a probe that has a larger B value is most desirable (if it also meets other requirements).

(3) Sensitivity to the environment – The applicability of a fluorescent probe in studying protein structure and stability stems from the environmental dependence of its fluorescence properties, including QY, peak emission wavelength (λ_{em}), and lifetime (τ_F). In other words, for an ideal

fluorophore one or all of those properties should undergo a drastic change, hence providing a large signal contrast, upon going from a hydrated to a dehydrated environment. Besides applications to study protein conformational changes, another widely pursued utility of fluorescence spectroscopy is to determine protein distribution or localization in the cellular environment.

(4) Excitation wavelength – When multiple fluorescent amino acids are present in the protein system in question, the ability to selectively excite the unnatural amino acid fluorophore is crucial for attaining site-specific information.

(5) Incorporation method – There are several methods available to incorporate unnatural amino acids into peptides and proteins discussed in detail later.

In addition to using these unnatural amino acids as stand-alone fluorophores, they can also form novel fluorescence donor-acceptor pairs with other natural and unnatural amino acids. These donor-acceptor pairs are most useful where the fluorescence intensity of the donor is quenched by the acceptor through the mechanism of either fluorescence resonance energy transfer (FRET) or electron transfer (ET), are exceedingly useful in providing site-specific structural information as the quenching efficiency is distance dependent (Doose et al., 2009; Rogers et al., 2011). However, in comparison to commonly used FRET pairs composed of fluorescent dyes or proteins, these unnatural amino acids-based FRET pairs have a smaller Förster distances (R_0) and hence are more amenable for site-specific interrogation of local interactions. For example, denaturant concentration dependent FRET signals can be used to determine protein folding stability (Glasscock et al., 2008; Rogers et al., 2010).

1.7: PROTEIN INFRARED SPECTROSCOPY

IR spectroscopy (both linear and nonlinear) is an exceedingly useful tool to study the structure and dynamics of biological molecules. This is because the IR spectrum of a molecule is

a distinct manifestation of its chemical identity and structure, as well as its environment. Also, since molecular vibrations occur on the femtosecond timescale, IR spectroscopy has no temporal resolution limit in the study of biological processes. Following the change in lineshape and center frequency of an IR signature peak is a useful means of studying a molecule in its environment. Below we briefly give an overview of IR properties that are commonly exploited when studying proteins using IR spectroscopy.

(1) Change in Frequency – Like fluorescence, various IR sensitive functional groups can experience solvent-induced frequency shift. Typically, increased hydrogen bonding of a vibrational probe will cause a shift in the IR spectrum and vice versa. This can be used to monitor change in the local environment of the probe and the probes interactions with its local environment (Serrano et al., 2012a). Additional, in some cases the frequency change can also be correlated with electrostatic changes (external electric field from the solvent) experienced by the IR probe (Chattopadhyay and Boxer, 1995; Pazos et al., 2014a). This phenomenon is called the Vibrational Stark effect. In general, the change in transition frequency, Δv, due to an externally applied field $\mathbf{F}$ (i.e., solvent), is given by (Boxer, 2009; Chattopadhyay and Boxer, 1995):

$$h\Delta v = \Delta\boldsymbol{\mu} \cdot \mathbf{F} - \mathbf{F} \cdot \Delta\boldsymbol{\alpha} \cdot \mathbf{F}/2, \qquad\qquad (\textbf{1.7.1})$$

where $\Delta\boldsymbol{\mu}$ and $\Delta\boldsymbol{\alpha}$ are changes in dipole moment and polarizability, respectively, between the states involved.

(2) Change in Bandwith – Similar to frequency shift, the bandwidth of an IR spectrum can change due to solvent by either increasing or decreasing. Additionally, the homogeneous or heterogeneous nature of the IR peak of interest can change due to solvent and other effects (Serrano et al., 2012a). Understanding these trends for specific probes make them useful to study proteins interactions with its local environment and other molecules.

1.8: UNNATURAL AMINO ACIDS FOR INFRARED SPECTROSCOPY

However, the IR spectrum of a native protein is typically dominated by spectral features that arise from overlapping vibrational transitions, due to degeneracy and/or vibrational couplings, hence making many of the naturally occurring IR bands of proteins unsuitable or unusable to reveal any site-specific information. Therefore, to increase the information content obtainable via IR spectroscopy, a straightforward approach is to label the protein in question with an appropriate external IR probe, most commonly in the form of an unnatural amino acid, in a frequency range that is "free" from spectral overlap with rest of the protein (Ma et al., 2015; Waegele et al., 2011b). This approach, in conjunction with different vibrational spectroscopic techniques, has been used to study a wide variety of protein biochemical and biophysical properties in a site-specific manner, including electrostatic field, hydration status and dynamics, hydrogen-bonding status and dynamics, conformation and conformational stability, intermolecular interactions, ligand binding, and amyloid formation. Currently, there are a large number of unnatural amino acids that have been shown to be useful as site-specific IR probes of proteins (Ma et al., 2015; Waegele et al., 2011b). However, these probes differ in size, vibrational property, sensitivity to the environment, and the ease of being incorporated into proteins. Therefore, similar to using unnatural amino acid-based fluorescent probes the following factors need to be considered when choosing an unnatural amino acid IR probe for any given application.

(1) Size and structure – Incorporation of any unnatural amino acid will unavoidably perturb the structure and/or other property of the protein system in question. Therefore, to minimize any potential perturbations, the chosen unnatural amino acid should be similar, in both structure and size, to the targeted native residue for mutation or substitution.

(2) Sensitivity to the environment – For a vibrational mode to be useful as a site-specific spectroscopic reporter, one of its properties, including peak frequency (ν), bandwidth (w), molar absorptivity (ε), and lifetime (τ), should exhibit a dependence on its environment. In practice, this

dependence on its environment is often evaluated by the difference of frequencies ($|\Delta\nu|$) measured in H_2O/D_2O and an organic solvent, such as tetrahydrofuran (THF) or dimethyl sulfoxide (DMSO). If the width of the IR band in question is sufficiently narrow, a larger $|\Delta\nu|$ would mean a higher sensitivity and hence a larger dynamic range.

(3) Frequency range and solvent – Water is an essential solvent for biological molecules. However, water absorbs ubiquitously in the mid-IR region, hence producing a solvent background in all IR measurements. Due to the isotopic effect, the vibrational spectrum of D_2O is shifted from that of H_2O; to reduce this solvent background, therefore one should consider using D_2O as the solvent in these cases.

(4) Sample concentration – Due to constraints arising from protein availability, solubility, and aggregation, sample concentration is the main factor that determines which probe is better suited for a specific application. When structural perturbation is not a major concern, using probes that have a high extinction coefficient would help lower the amount of protein sample used in a single experiment.

(5) Application – Depending on the nature of the question being studied, one probe can be better suited than others. While most unnatural amino acids with IR signatures that do not overlap with the native protein signals, when placed at the right location in a protein, can be used to follow a well-defined structural transition, such as protein folding and unfolding, due to the dependence of their frequencies on the environment. Additionally, some of unnatural amino acid probes can also be used to provide quantitative information about local electric field at the moment because the corresponding frequency-field maps or Stark tuning rates have been determined.

(5) Incorporation method – There are several methods available to incorporate unnatural amino acids into peptides and proteins are discussed in detail below.

1.8: METHODS FOR UNNATURAL AMINO ACID INCORPORATION

To use unnatural amino acids as small probes for protein microscopy and spectroscopy applications, they must be incorporable into the protein of choice. The methods for incorporating unnatural amino acids primarily fall under two main routes, chemical or genetic. While these methods previously required expertise, great efforts have been made to make it universally accessible. Here, we summarize the most robust and useful techniques in the field to introduce unnatural amino acids into proteins.

The direct chemical modification of one or more natural amino acids is the most common way to introduce unnatural amino acids within a protein. In nature, enzymes commonly introduce post-translational modifications of residues within the proteome. Instead of using enzymes, we can modify amino acids within a protein using facile protein compatible reactions in vitro. Classically, researchers have covalently modified natural amino acids such as cysteine and lysine through various types of chemical reactions (Spicer and Davis, 2014a). Direct modification reactions target three major functional groups: thiols (cysteine, methionine), amines (lysine, arginine, glutamic acid, and aspartic acid) and hydroxyl groups (tyrosine and serine). One of the earliest and most popular methods of directly modifying amino acids in a protein is the targeting of free cysteine residues through substitution with haloalkyl substrates containing the functional groups of interest (Chalker et al., 2009). While this approach does successfully modify the amino acid, it does present the potential for cross-reactivity with other nucleophilic side chains such as lysine and histidine. To prevent potential cross-reactivity, pH levels can be optimized for selective reaction to cysteine over other nucleophilic amino acids (Crankshaw and Grant, 1996). Another major strategy to modify amino cysteine is using maleimide–thiol coupling which is commonly used for scalable chemical modification of proteins through cysteine residues (Moore and Ward, 1956; Smith et al., 2010; Witter and Tuppy, 1960). Specifically, functionalized maleimides serve as Michael acceptors, in a reaction between a maleimide and a thiolate to form a thiosuccinimide bond at the cysteine residue to incorporate the label of choice. Finally, because cysteine exists in

low abundance in proteins, it is more amenable to site-specific control of amino acid modification, making it an attractive target candidate (Clark and Lowe, 1977).

These reactions are useful for appending fluorophores, other spectroscopically active molecules (i.e. IR, CD, NMR), affinity tags, posttranslational modifications and more. While direct modification, in essence, achieves the formation of an unnatural amino acid, but when aiming for a subtle modification, the intermediary bonds between the α-carbon and the desired functional group can either be bulky like in the case of maleimide or even in the case of labeled cysteine the presence of a sulfur molecule can be undesirable. Therefore, using a posttranslational mutagenesis approach developed by Davis and coworkers mitigates this problem through a two-step reaction method which works at ambient conditions (Wright and Davis, 2017; Wright et al., 2016). The first step requires desulfurization of the cysteine to dehydroalanine which is then further reacted with a haloalkyl substrate containing the functional group of choice through free radical chemistry to form a natural amino acid, an unnatural amino acid or posttranslationally modified residue through a newly formed $C(sp^3) - C(sp^3)$ bond (Wright et al., 2016). While most direct chemical modification methods work *in vitro*, some methods will work *in vivo* as well. The major caveat for direct modification is the selectivity, such as labeling a specific lysine when there is more than one. Since free cysteines have low abundance in proteins, it is a good candidate for a single modification. However, modifying multiple cysteines with different molecules is challenging. To circumvent this, synthetic, a semi-synthetic approach (discussed below) and the use of other unnatural amino acids, like azidoalanine as a handle can be useful. Additionally, modifying the area around cysteine with specific residues to change the local environment around the cysteine will allow for selectivity for as used in π-Clamp-mediated cysteine conjugation.

An alternative method to chemically modifying the amino acid side chains in biologically expressed proteins is to synthesize them using solid-phase peptide synthesis chemically. In general solid-phase peptide synthesis facilitates the production of peptides which are difficult to express in bacteria (or other organisms), the incorporation of unnatural amino acids,

peptide/protein backbone modification, and the synthesis of proteins containing D-amino acids. Specifically, this is done by employing a series of reactions to couple amino acids one at a time through amide bonds. Solid-phase peptide synthesis is most appropriate for peptides and small proteins. Although this is a caveat of this method, advances in the field for currently permit lengths of up to 100-150 amino acids in a single round of automated solid-phase synthesis. For solid-phase synthesis, the N-terminus of the amino acid needs to be protected (Fmoc, Boc, etc.) To prevent unwanted side reactions with the various amino acid side chains, protecting group strategies are usually necessary. Solid-phase peptide synthesis generally starts at the C-terminus and proceeds toward the N-terminus. A typical solid-phase peptide synthesis reaction cycle includes deprotection of the amino acid protecting group, washing, coupling, and post-coupling washing steps. These cycles are repeated to add on a single amino acid at a time until the completion of the peptide. Unnatural amino acids can be used just as easily as a natural amino acid, so long as it is available with the form of N-terminus protected, such as with Fmoc or Boc protecting groups. There are many unnatural amino acids commercially available, and a great deal of effort goes into exploring the synthesis of new ones.

Native chemical ligation is a technique that is useful to overcome the limitations of protein length generally posed by peptide synthesis (Kulkarni et al., 2018). Native chemical ligation is a method that chemically stitches together a larger protein from its smaller peptide fragments which can be expressed, synthesized or a combination of the two. Therefore, allowing for incorporation of as many unnatural amino acids and post-translational modifications as desired in a site-specific manner. This specifically works by incorporating the unnatural amino acids and post-translational modifications into peptides by conventional solid phase peptide synthesis which can then be stitched together modularly with either other peptides or larger sequences which have were expressed in cells (Dawson et al., 1994; Kumar et al., 2011; Torbeev and Kent, 2007). This is achieved specifically using an N-terminal cysteine which can be ligated with a peptide thioester in a chemoselective manner via a native chemical ligation to generate a full-length protein with the desired unnatural amino acids and post-translational modifications (Dawson et al., 1994; Kumar

et al., 2011; Torbeev and Kent, 2007). After chemical synthesis / semi-synthesis, all free cysteines are desulfurized to form alanine, which is key to achieving native-like peptide sequence without the introduction of additional cysteines. While alanine was the junction of choice for ligation, Phe, Val, Leu, Thr, Lys, Pro, and Gln junctions are now possible (Crich and Banerjee, 2007; Haase et al., 2008; Harpaz et al., 2010; Shang et al., 2011; Siman et al., 2012). While this strategy is very powerful, it can be highly involved. Recently, Brik and coworkers have demonstrated a straight forward method for native chemical ligation in a single test tube with one purification step to achieve site-specific ubiquitination of the protein α-synuclein (Fauvet et al., 2013; Haj-Yahya et al., 2013). This methodology can be useful to generate proteins with other post-translational modifications (glycosylation, phosphorylation, methylation, etc.) and unnatural amino acids as site-specific spectroscopic probes.

An alternative to chemical methods for incorporating unnatural amino acids into proteins is to encode them into the protein genetically. To date, many unnatural amino acids can be incorporated genetically into proteins (Dumas et al., 2015) by the use of orthogonal tRNA synthase and amber codon suppression methods pioneered by Schultz and Chin (Kim et al., 2013; Lang and Chin, 2014; Liu et al., 1997). This method makes use of appropriating the amber stop codon to now recruit and corresponds to a new tRNA that is coupled with the unnatural amino acid of choice. This approach requires involved selection methods and directed evolution of tRNA synthetase to recognition and binding the desired unnatural amino acid to charge it to the tRNA. Additionally, the efficiency of unnatural amino acid incorporation is known to be relatively low, but getting better. However, this method uniquely allows for site-specific incorporation of unnatural amino acids across several types of cell such as *E. coli* (Wang et al., 2001), yeast (Wang and Wang, 2012), and mammalian cells (Liu et al., 2007) as well as some live animals. Genetically encoded unnatural amino acids have been used as selective reactive group handles, spectroscopic probes, natural post-translational modifications, and photo-reactive groups such as photocaged amino acids and photo-crosslinkers. Moreover, methods have been developed to incorporate multiple unique unnatural amino acids in a single protein (Chatterjee et al., 2013b).

Generating new tRNA synthetase and unnatural amino acid pairs can be highly involved and specialized. Therefore, a cell-free biosynthetic method (Kuruma and Ueda, 2015; Talukder et al., 2015a) that combines chemical and genetic methods to incorporate unnatural amino acids into large proteins in vitro can be simpler. This method involves chemically charging suppressor tRNAs with the unnatural amino acid of interest and has been recently used by Hecht and coworkers to incorporate two unnatural amino acids into proteins(Chatterjee et al., 2013b; Talukder et al., 2015a). In general, the unnatural amino acid of choice is coupled to pdCpA dinucleotide in a multistep process, which is then used to construct amber suppressor tRNAs which are aminoacylated with the dinucleotide containing the desired unnatural amino acid. After that, the misacylated suppressor tRNAs and the plasmids encoding the protein of choice with the amber mutation are added together for cell-free protein synthesis using kits like the PURE system for in vitro protein production[29]. Cell-free incorporation is very useful when it is hard to generate the desired tRNA synthetase and unnatural amino acid pair. For example, it would be challenging to generate selectivity for isotope labeled amino acids within a cell.

1.9: OVERVIEW OF THESIS CHAPTERS

Together the various methods from spectroscopy to the incorporation of unnatural amino acids enable the wide use of these unnatural probes in protein-based studies, both in vitro and in vivo. As described, the currently available unnatural amino acids are valuable as spectroscopic probes useful to reveal site-specific environmental, dynamic, structural or stability information via commonly used spectroscopic methods. Additionally, unnatural amino acids are advantageous as minimal fluorescent tags of proteins for microscopy within the cellular environment. Furthermore, in the same manner, this concept can also be further extended to study the other biologically relevant macromolecules such as DNA, RNA, and lipids.

The unifying theme for this thesis is the development, characterization, and application of subtly modified unnatural amino acids as probes for biological microscopy and spectroscopy. The first part of the thesis focuses on the discovery and application of novel tryptophan derivatives

while the second part focuses on innovative ways of extending the applications of previously known unnatural amino acids.

In Chapter 2, we surveyed the absorption and emission properties of all isomeric cyanoindoles which led us to the discovery of a novel blue fluorescent amino acid, 4-cyanotryptophan with the potential to be useful in microscopy applications. To verify this, we characterized the photophysical properties of this new amino acid-based fluorophore, which included measuring its absorption spectrum, fluorescence spectrum, molar absorptivity, quantum yield (in various solvents and conditions), fluorescence lifetime decay (in various solvents and conditions), and its photostability in comparison to EBFP (enhanced blue fluorescent protein). To demonstrate and validate its utility in biological microscopy applications, we incorporated this fluorophore into an antimicrobial peptide which we then treated mammalian cells to observe apoptosis using wide-field and confocal fluorescence microscopy.

In Chapter 3, we extend the application of 4-cyanotryptophan to fluorescence spectroscopy. Specifically, as useful FRET pair with the DiO, a common lipophilic dye used to stain membranes. To verify this, we collect absorption and fluorescence emission spectra of both the donor (4-cyanotryptophan) and the acceptor (DiO) to demonstrate the necessary spectral overlap and calculated the Förster distance for the pair. As a proof of concept application, we incorporated 4-cyanotryptophan into a pH-dependent membrane-binding peptide which we first used to image cell-membrane-bound peptides. Then we demonstrated its utility as a FRET pair by determining the binding constant of peptide bound to a model membrane. Additionally, as a control, we show pH essentially no effect on 4-cyanotryptophan fluorescence intensity by titrating a model tripeptide containing 4-cyanotryptophan from pH 4 to pH 10. Furthermore, we detail an expedient synthetic route using enzymatic resolution to make the L-form of 4-cyanotyprophan. The product was characterized extensively by NMR, high-resolution mass spectrometry, optical rotation, circular dichroism, and high-performance liquid chromatography.

In Chapter 4, we focus on further building upon the utility of 4-cyanotryptophan for applications in fluorescence spectroscopy. In this study, we demonstrate that tryptophan (donor) and 4-cyanotryptophan (acceptor) is a FRET pair as indicated by the significant spectral overlap and the calculated Förster distance between the donor and acceptor. More interestingly, the pair also serves as a fluorophore-quencher pair likely via photo-induced electron transfer (PET). This notion was first supported by steady-state fluorescence and time-resolved measurements of model peptides with varying distances between the fluorophore (4-cyanotryptophan) and the quencher (tryptophan). Then, using a Stern-Volmer quenching experiment and calculations, we determine the distance-dependent quenching rate equation for this fluorescence quenching process. Additionally, we show that the only natural amino acid that can significantly quench 4-cyanotryptophan fluorescence is tryptophan. Moreover, toggling between the FRET and PET property of this pair is conveniently dictated by the excitation wavelength, and therefore you get two separate measurements (on different length scales) with a single sample. Lastly, we show we can exploit the dual FRET-PET property of this pair to study peptide end-to-end interactions and protein-ligand interactions.

In Chapter 5, we repurposed the side chain of 4-cyanotryptophan as a nucleoside (4-cyanoindole-2′-deoxyribonucleoside) to use in DNA based applications. First, we used steady-state fluorescence to demonstrate the nucleoside had similar photophysical properties to the amino acid version of itself. Then we validated the notion that the natural nucleoside guanine can quench 4-cyanoindole-2′-deoxyribonucleoside through the PET mechanism. As a proof of principle, we demonstrate that the fluorescence-quencher pair formed by 4-cyanoindole-2′-deoxyribonucleoside and guanine can be used to quantitatively assess the binding affinity of a single-stranded DNA to a transport (bovine serum albumin) via fluorescence spectroscopy. The binding results were then corroborated using isothermal calorimetry. Then we further show that the C≡N stretching frequency of 4-cyanoindole-2′-deoxyribonucleoside is sensitive to the local environment, making it a useful site-specific infrared probe of oligonucleotides. Lastly, we use the

infrared properties of the probe to interrogate the electrostatics and environment of the single-stranded DNA to bovine serum albumin.

In Chapter 6, we expand the spectroscopic utility of a well-known infrared and fluorescence probe, *p*-cyanophenylalanine, to show that it can be used as a pH sensor. To demonstrate this, we carried out a series of steady-state fluorescence measurements of model peptides which we titrated from acid to basic pH. By varying the position of *p*-cyanophenylalanine within the peptide showed us that the fluorescence quantum yield of this unnatural amino acid, when placed at or near the N-terminal end of a polypeptide, depends on the protonation status of the N-terminal amino group of the peptide. Then using the pH sensing-property of *p*-cyanophenylalanine, we determined the N-terminal pKa values of nine tripeptides. Lastly, to demonstrate a biological application we utilized the *p*-cyanophenylalanine to probe the membrane penetration kinetics of a cell-penetrating peptide TAT as it transitioned from an acidic environment of the bulk solution to a basic environment within a vesicle and vice versa.

In Chapter 7, we demonstrate a simple method to incorporate an ester moiety onto cysteine. The ester carbonyl stretching vibration is a sensitive and convenient infrared (IR) probe of protein electrostatics and hydration. We devised that we could use a cysteine alkylation reaction to achieve this goal. We demonstrate the feasibility of this simple method by successfully incorporating a methyl ester group ($-CH_2-COOCH_3$) into a model peptide, two amyloid-forming peptides derived from the insulin B chain and Aβ, and bovine serum albumin (BSA). We then carried out IR studies with those peptide and protein systems to further confirm the utility of this vibrational probe in monitoring, for example, the structural integrity of amyloid fibrils and ligand binding-induced changes in protein local hydration status.

To conclude, in chapter 8, we summarize the work detailed in this thesis and discuss the future directions inspired by this body of work.

Table 1.1: Common Spectroscopic Techniques and their Associated Regions of the Electromagnetic Spectrum.

Method	Region of the Electromagnetic Spectrum
Crystallography	X-ray
X-ray Spectroscopy	X-ray
Electron Microscopy	X-ray
UV-Vis Spectroscopy	UV / Visible
Fluorescence Microscopy	Visible
IR Spectroscopy	Infrared
Nuclear Magnetic Resonance	Microwave

Table 1.2: 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 were obtained from (Lakowicz, 1999).
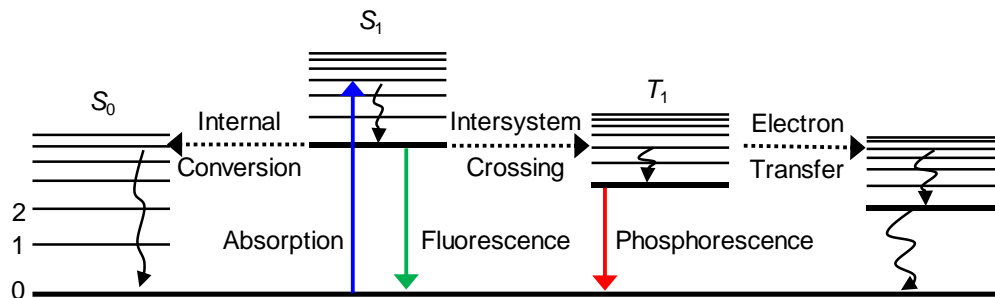
Figure 1.1: A form of a Jabłoński Diagram, where S_x and T_x represent the xth singlet and triplet electronic state and 0, 1, 2…., represent the vibrational levels in each electronic state. Internal conversion and intersystem crossing are represented by dotted lines while absorption (blue) of light and 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.

2: Blue Fluorescent Amino Acid for Biological Spectroscopy and Microscopy

2.1: 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.

2.2: 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 is also sensitive to environment (Callis et al., 2014; Royer et al., 2006; Zhong et al., 2009). 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 (Roy et al., 2008)

and imaging applications, especially under *in vivo* conditions. For this reason, significant efforts have been devoted to identifying small unnatural fluorophores (Chatterjee et al., 2013; Cohen et al., 2002; Hamada et al., 2005; Lepthien et al., 2008; Moroz et al., 2013; Summerer et al., 2006) that could overcome this limitation. So far, only naphthalene-based fluorophores, such as Prodan (Chatterjee et al., 2013; Summerer et al., 2006) and Aladan (Cohen et al., 2002) are useful in this regard. However, these fluorophores are not structurally based on any naturally-occurring amino acids and are larger 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, a 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 Trp analog, 4-cyanotryptophan (4CN-Trp), 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 (Lepthien et al., 2008; Talukder et al., 2015) and various indole-ring substituted analogs (Lotte et al., 2004; Markiewicz et al., 2016; Ross et al., 1997), 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 (Talukder, 2014) 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 (**Figure 2.1**), the absorption and emission spectra of 4-cyano-indole (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 (13.7 ns), a reasonably large molar extinction coefficient (ca. 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.

2.3: EXPERIMENTAL SECTION

Material and Methods. Small molecules were purchased from either Acros Organics (Trp, 4CNI, and DPA), Ark Pharm (4CNI-3AA), or Sigma (indole) and used as received. To the best of our knowledge, 4-cyano-L-tryptophan was currently not available commercially. Therefore, we synthesized 4-cyano-L-tryptophan by first synthesizing N-alpha-Boc-4-bromo-L-tryptophan methyl ester, which was prepared from the readily available 4-bromo-L-tryptophan by methylation using $TMSCH_2N_2$ followed by Boc protection using the reagents Boc_2O and triethylamine. Palladium-catalyzed cyanation of the N-alpha-Boc-4-bromo-L-tryptophan methyl ester was performed under mild condition (40 °C, overnight) with the ligand developed by Cohen and Buchwald (Cohen and Buchwald, 2015). Subsequent Boc-deprotection by TFA (50% in DCM) provided the amine, which was acetylated using acetic anhydride in DMF. High-performance liquid chromatography (HPLC) furnished the purified 4CN-tryptophan derivative as the acetyl, methyl ester (designated 4CN-Trp). For the solution phase peptide synthesis, the amine of 4-cyano-L-tryptophan methyl ester was coupled to Fmoc-Gly-Gly-OH using HCTU and DIPEA. The Fmoc group was removed by diethylamine, and the amide was acetylated using acetic anhydride in DMF to furnish the N-acetyl methyl ester of the tripeptide Gly-Gly-4CN-Trp after HPLC purification. Steady-state fluorescence spectra were collected on a Jobin Yvon Horiba Fluorolog 3.10 fluorometer. Fluorescence decay kinetics were collected on a home-built time-correlated single photon counting setup (TCSPC) using an excitation of 270 nm. Wide-field fluorescence images were acquired on an Olympus IX71 microscope equipped with a 100 W Hg lamp and a 60× water objective. Confocal images were acquired on a home-built stage-scanning confocal microscope. Specific details of 4CN-Trp

and Gly-Gly-4CN-Trp synthesis, sample preparation, fluorescence measurements, and imaging experiments are given in the following section.

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 was 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) was 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$^+$)

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 was 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-D$_6$) 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$^+$)

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 was 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. The sample was then dissolved in H_2O and lyophilized to remove the DMF and water. The purified amino acid was obtained by reverse phase HPLC.

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

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 was 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. The sample was then dissolved in H_2O and lyophilized to remove the DMF and water. The purified peptide was obtained by reverse phase HPLC.

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

Materials and Sample Preparation. L(-)-Tryptophan (99%, Acros Organics), 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 2.2**).

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 ca. 10 hours at room temperature before use.

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.

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 a 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, 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 employed 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 2.3**). This ratio is then used in Equation 2.4.1 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 (ca. 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 (Markiewicz et al., 2016). Briefly, the 270 nm excitation pulse was generated by collinear 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 an 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 either a single-exponential or bi-exponential function using FLUOFIT (Picoquant GmbH). The OD of each sample was approximately 0.1 at the excitation wavelength (270 nm).

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 (Schneider et al., 2012). 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).

2.4: RESULTS AND DISCUSSION

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, 4-cyano-indole (4CNI). As shown (**Figure 2.1b**), 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 2.1c**), with a maximum emission wavelength of 405 nm. Also, 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 2.1d**) using the following relationship (Lakowicz, 1999):

$$QY_S = QY_R \frac{I_S}{I_R} \frac{A_R}{A_S} \frac{n_S^2}{n_R^2} \, , \tag{2.4.1}$$

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 (Suzuki et al., 2009), with a QY of 0.15 for Trp in water at $\lambda_{ex} = 270$ nm and 0.9 (0.97) for DPA in cyclohexane at $\lambda_{ex} = 325$ nm (355 nm). Therefore, for a λ_{ex} of 270 nm, we used Trp as the reference, whereas for a λ_{ex} of equal to or larger than 325 nm, we used DPA as the reference.

As indicated (**Table 2.1**), 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, determined using DPA as a reference (**Table 2.1**). 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-cyano-indole-3-acetic acid (4CNI-3AA), and 4CN-Trp*-Gly (For

nomenclature see **Figure 2.1a**; 4CN-Trp and Gly-Gly-4CN-Trp were evaluated as the corresponding N-acetyl methyl esters, while the other peptides are assessed as the free carboxylates). 4CN-Trp was synthesized as described in the experimental section, while the corresponding 4CN-Trp[*] peptide derivatives were prepared from the commercially available and relatively inexpensive 4CNI-3AA, which can be conveniently appended to the N-terminal end of a peptide (**Figure 2.1a**). As shown (**Figure 2.1**), 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 2.1c**), 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 the 4CN-Trp chromophore a true blue fluorescence emitter. Perhaps more importantly, their fluorescence QYs in water at 25 $^{\circ}$C are larger than 0.80 when a λ_{ex} of 325 nm was used (**Table 2.1**), confirming 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 2.1**) under this condition. 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 (Chen et al., 1998), including methionine (Met) (Yuan et al., 1998), can quench Trp fluorescence. However, as shown (**Table 2.1** and **Figure 2.4a**), 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. Taken together, the above results suggest that the

41

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.

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 2.4**), 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 2.1**). The fluorescence decay kinetics of 4CN-Trp and Gly-Gly-4CN-Trp show a predominant long lifetime component (ca. 13.6 ns; ca. 95%), and a small 5% shorter lifetime component (ca. 1.8 ns) (**Table 2.1**). It is possible that this double-exponential decay behavior arises from different 4CN-Trp rotamers, as proposed for interpreting the non-single-exponential fluorescence decay kinetics of Trp (Lakowicz, 1999). However, the fact that both the C- and N-termini of 4CN-Trp and Gly-Gly-4CN-Trp used in the current study are capped and that the fluorescence of 4CN-Trp*-Gly, which has an amide bond, decays in a single-exponential manner argue against this possibility. Therefore, we tentatively attribute the shorter lifetime component to an impurity that cannot be removed by the purification methods employed. One likely candidate for this impurity is Trp, a byproduct of the synthesis of 4CN-Trp. 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 (ca. 6 ns) in experiments involving cells (Berezin, 2010). As shown (**Table 2.1** and **Figure 2.4**), 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 2.1** and **Figure 2.5b**), corroborating

the conclusion reached above that Met does not quench the fluorescence of the 4CNI fluorophore.

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) (Shaner et al., 2005), 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) (Yang et al., 1998) under identical conditions (i.e., both samples have the same integrated absorbance at the excitation wavelength). As indicated (**Figure 2.6a**), 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 ca. 25% from its initial value in 10 hours, while that of EBFP decreased ca. 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 2.6b**), 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.

Fluorescence Imaging Application. In a proof-of-principle imaging experiment, we used the 4CNI fluorophore to monitor the distribution of a cytotoxic and antimicrobial peptide (AMP) inside

HEK293T/17 cells. Specifically, we appended a 4CNI fluorophore to the N-terminus of an AMP, Mastoparan X (MpX) (Wakamatsu et al., 1992), 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 ca. 3000 cells). In comparison to the control (**Figures 2.7c and 2.7d),** it is apparent that the fluorescence image (**Figure 2.7f**) 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 (Henriques et al., 2006; Shai et al., 1999), 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 2.7e**) is much fainter, it still provides enough contrast to allow a rough visual inspection of the AMP distribution, in comparison to the control (**Figures 2.7b and 2.7d**). 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 2.1b**). 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 2.8**), 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 (Paredes-Gamero et al., 2012). 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 2.7**). 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 2.7e** (i.e., in the presence of 1 μM 4CN-Trp*-MpX) but waited ca. 10 more hours before confocal measurements, to ensure cell death throughout the entire population. As shown (**Figure 2.9a**), 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 the formation of membrane blebs and small vesicle-like particles (Paredes-Gamero et al., 2012). Further evidence supporting this notion comes from a confocal image acquired with a higher spatial resolution (**Figure 2.9b**), which clearly shows the existence of a micron-sized components that can interact with 4CN-Trp*-MpX within a circular bleb. It is noteworthy that our confocal microscope utilized a 340 nm LED as the excitation source, which is less focusable and has 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.

2.5: 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 (Lepthien et al., 2008; Moroz et al., 2013), 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 ca. 325 nm, an emission spectrum peaked at ca. 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 (Ross et al., 1997; Talukder et al., 2014; Talukder et al., 2015) have been developed to site-specifically incorporate other Trp analogues, such as 6CN-Trp and 7CN-Trp (Talukder et al., 2015), hydroxytryptophans (Ross et al., 1997), azatryptophans (Ross et al., 1997), and methyltryptophans (Francis et al., 2017), 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 (Lippitz et al., 2002; Palero et al., 2005; Rehms et al., 1993), 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). Moreover, 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. Finally, it is worth mentioning that further experimental and theoretical studies are needed to understand why 4CN-Trp possesses rather unique photophysics, in comparison to other tryptophan derivatives.

Table 2.1: Fluorescence lifetime (τ) and QYs measured for 4CNI, 4CN-Trp, and their derivatives under different excitation wavelengths, temperatures (*T*) and solution conditions, as indicated.

Sample	Solvent	*T* (°C)	τ (ns) (%Amp)	QY 270 nm	QY 325 nm	QY 355 nm
4CNI	H_2O	25	9.1 (100%)	0.85 ± 0.13	0.78 ± 0.03	0.64 ± 0.07
4CNI-3AA	H_2O	25	13.5 (100%)	0.71 ± 0.04	0.83 ± 0.05[a]	0.91 ± 0.14
4CN-Trp*-Gly	H_2O	25	12.6 (100%)	0.73 ± 0.10	0.84 ± 0.09	0.82 ± 0.05
4CN-Trp*-Gly	THF	25	7.6 (100%)	------	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	12.2 (100%)	------	0.75[c]	------
4CN-Trp	H_2O	25	13.7 (95%) 1.9 (5%)	------	0.88 ± 0.06[a]	------
Gly-Gly-4CN-Trp	H_2O	25	13.6 (93%) 1.7 (7%)	------	0.89 ± 0.05[a]	------

[a]Measurements made using the gradient method as described in the experimental section

[b]20 mM phosphate buffer, pH 7.2

[c]QY determined relative to 4CN-Trp*-Gly
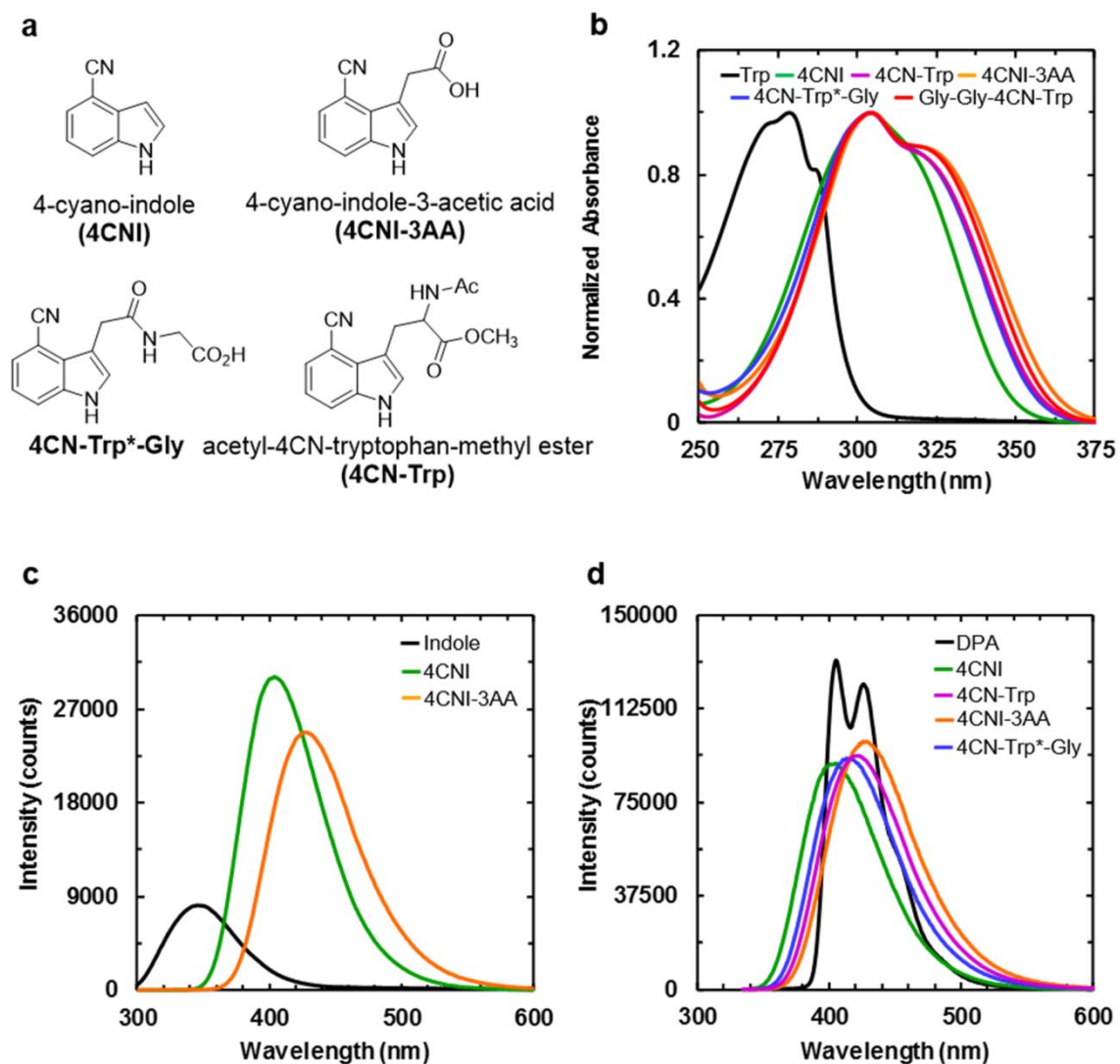
Figure 2.1: (a) Structures of 4CNI, 4CNI-3AA, 4CN-Trp, and 4CN-Trp*-Gly, as indicated. **(b)** Absorption spectra of Trp, 4CNI, 4CNI-3AA, 4CN-Trp, 4CN-Trp*-Gly, and Gly-Gly-4CN-Trp, as indicated. **(c)** Comparison of the fluorescence spectrum of indole with those of 4CNI and 4CNI-3AA. These spectra 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. **(d)** 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.

48
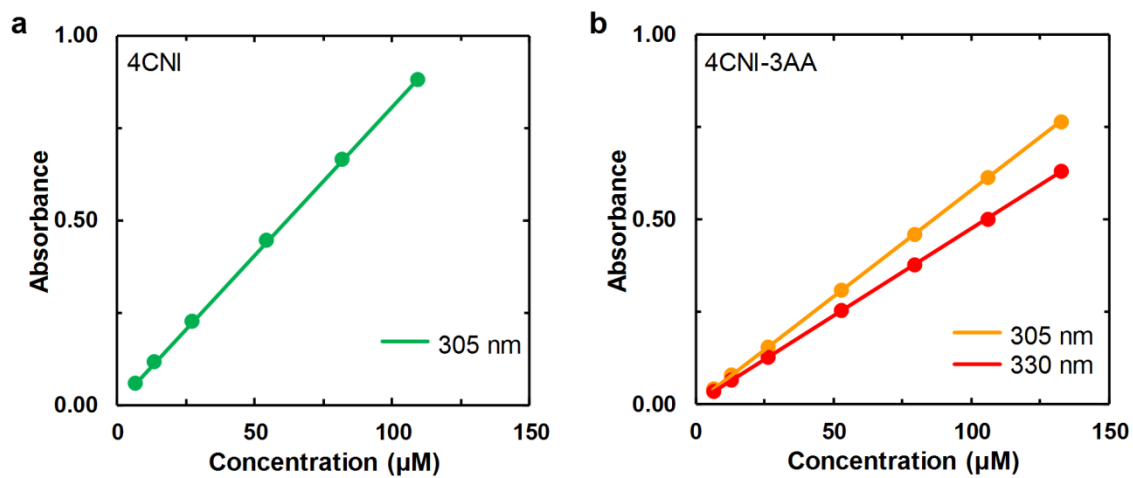
Figure 2.2: Beer's law plots of (**a**) 4CNI and (**b**) 4CNI-3AA at the indicated wavelengths. The straight line in each case corresponds to a linear regression of the respective data with a zero y-intercept, which yielded a molar extinction coefficient of 8056 ± 24 M^{-1} cm^{-1} for 4CNI at 305 nm, and 5778 ± 8 M^{-1} cm^{-1} and 4750 ± 8 M^{-1} cm^{-1} for 4CNI-3AA at 305 nm and 330 nm, respectively.
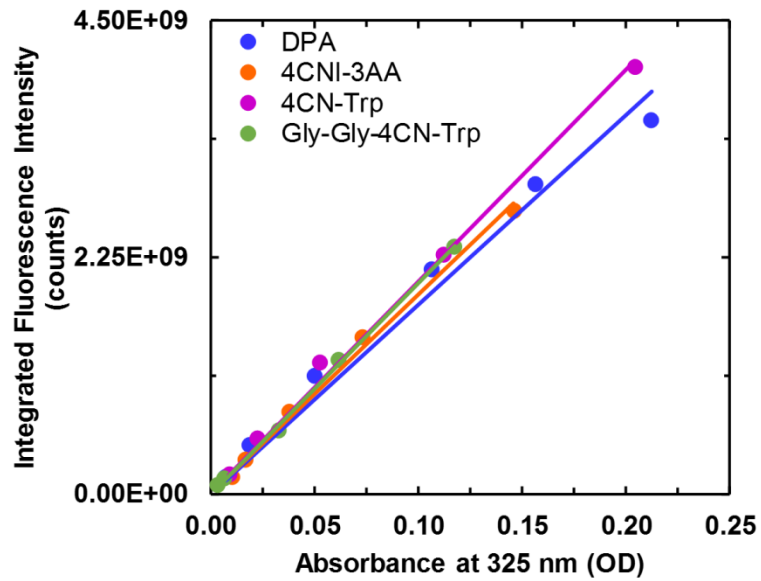
Figure 2.3: Integrated fluorescence intensity of DPA, 4CNI-3AA, 4CN-Trp, and Gly-Gly-4CN-Trp as a function of absorbance at 325 nm. The solid color line in each case corresponds to a linear regression of the respective data with a zero y-intercept, which yielded a slope of $(1.80 \pm 0.07) \times 10^{10}$ for DPA, $(1.89 \pm 0.04) \times 10^{10}$ for 4CNI-3AA, $(2.02 \pm 0.04) \times 10^{10}$ for 4CN-Trp, and $(2.00 \pm 0.03) \times 10^{10}$ for Gly-Gly-4CN-Trp. The calculated QYs at 325 nm are shown in Table 2.1.
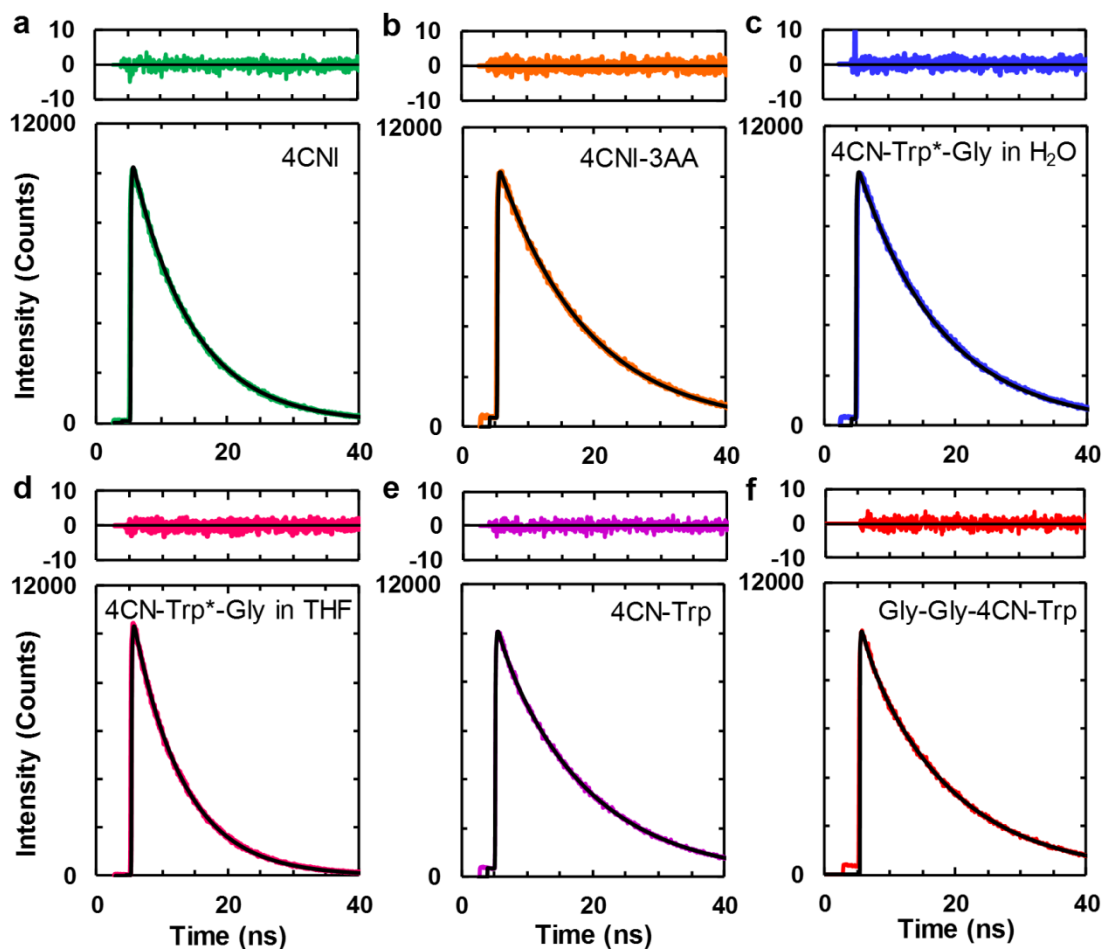
Figure 2.4: Fluorescence decay kinetics of **(a)** 4CNI, (b) 4CNI-3AA, (c) 4CN-Trp*-Gly **(d)** 4CN-Trp*-Gly in THF, **(e)** 4CN-Trp, and **(f)** Gly-Gly-4CN-Trp, all in H_2O except **(d)**. In each case, the smooth black line corresponds to the best fit of the respective data to a single-exponential (**a**-**d**) or a bi-exponential (**e**-**f**) function and the resultant lifetime(s) is given in Table 2.1. Also, the top panel in each case is the corresponding residual plot.

51
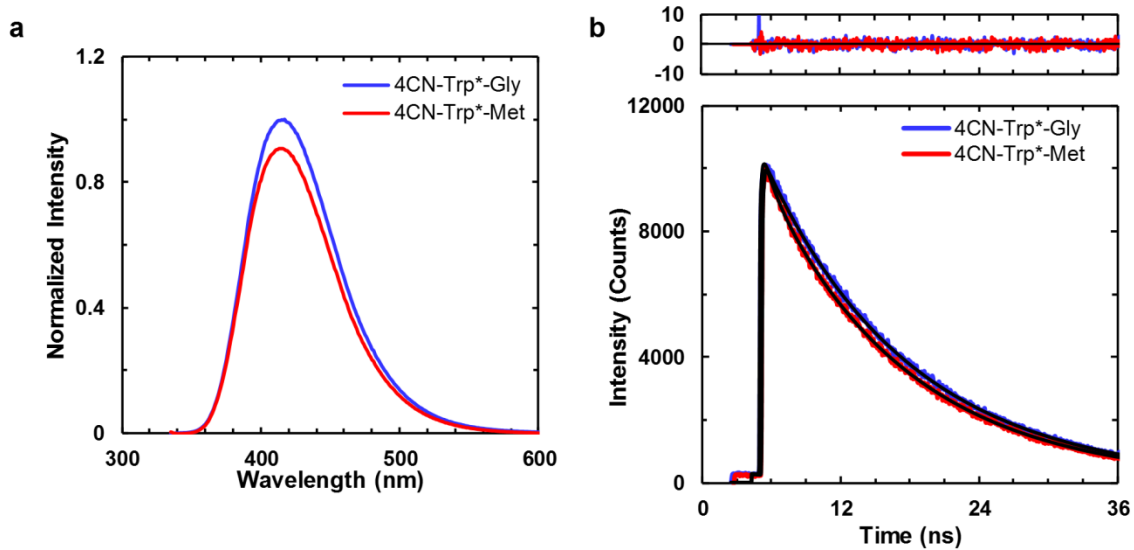
Figure 2.5: (a) Fluorescence spectra (λ_{ex} = 325 nm) and **(b)** fluorescence decay kinetics (λ_{ex} = 270 nm) of 4CN-Trp*-Gly and 4CN-Trp*-Met, as indicated. The optical densities of both samples at λ_{ex} were equal. Fitting each fluorescence decay in (**b**) to a single-exponential function (black) 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.
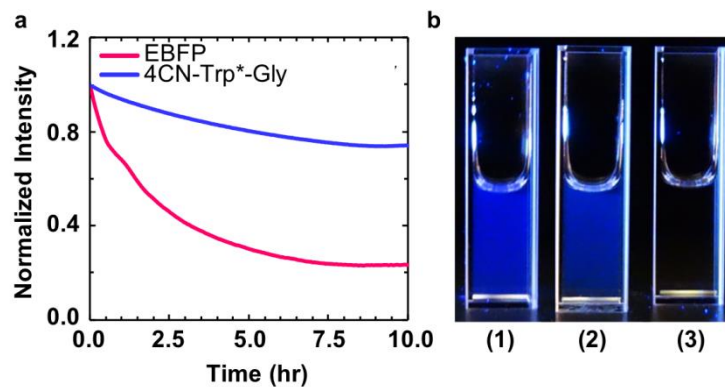
Figure 2.6: (a) Photobleaching kinetics of an EBFP sample (pink) and a 4CN-Trp*-Gly sample (blue). For both cases, the photobleaching light at 355 nm was derived from a Fluorolog 3.10 fluorometer with the excitation slit width set at 5 nm. **(b)** Pictures of a 50 nM 4CN-Trp*-Gly sample (1), a 50 nM EBFP sample (2), and water (3) under the illumination of 340 nm light, taken using a Sony RX10 camera.
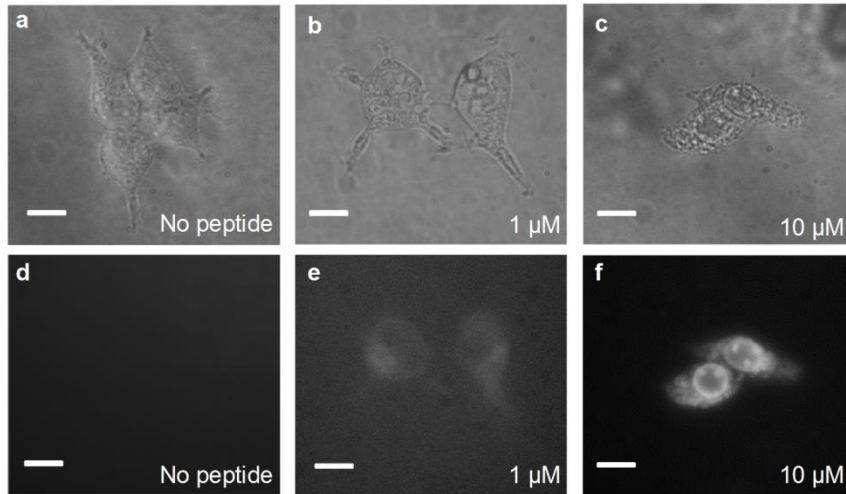
Figure 2.7: Bright-field (**a**, **b**, **c**) and fluorescence (**d**, **e**, **f**) images of HEK293T/17 cells obtained on an Olympus IX71 microscope, in the presence and absence of 4CN-Trp*-MpX at the indicated concentrations. The scale bar in each panel corresponds to 10 μm.
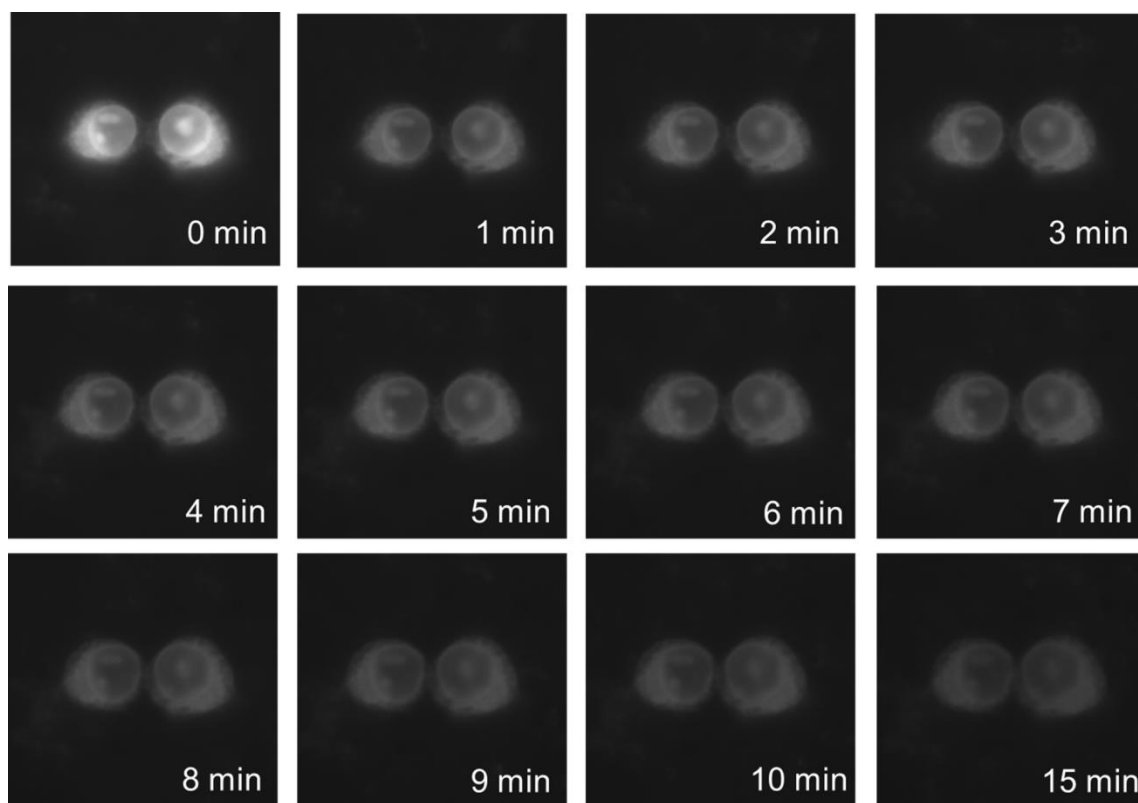
Figure 2.8: Fluorescence microscopic images of HEK293T/17 cells in the presence of 10 μM 4CN-Trp*-MpX obtained at different time points, as indicated. Between each measurement, the sample was continuously illuminated by the excitation light of the microscope. It is noticeable that the fluorescence intensity 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.
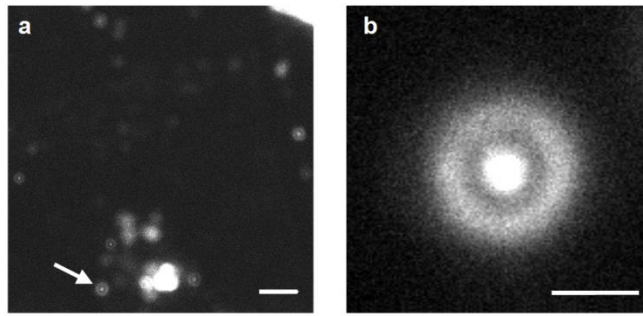
Figure 2.9: Representative confocal microscopic images of HEK293T/17 cells that have been incubated with 1 µM 4CN-Trp*-MpX for ca. 10 hours, obtained with a scanning step size of 100 nm (**a**) and 50 nm (**b**). The scale bar corresponds to 10 µm in (**a**) and 2 µm in (**b**). Many blebs and other 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**).

3: Synthesis and Application of the Blue Fluorescent Amino Acid L-4-Cyanotryptophan to Assess Peptide-Membrane Interactions

3.1: ABSTRACT

Recently, L-4-cyanotryptophan has been shown to be an efficient blue fluorescence emitter, with the potential to enable novel applications in biological spectroscopy and microscopy. However, lack of facile synthetic routes to this unnatural amino acid limits its wide use. Herein, we describe an expedient approach to synthesize Fmoc protected L-4-cyanotryptophan with high optical purity (>99%). Additionally, we test the utility of this blue fluorophore in imaging cell-membrane-bound peptides and in determining peptide-membrane binding constants.

3.1: INTRODUCTION

Fluorescence spectroscopy and microscopy are two of the most prevalent tools used in biochemistry, biophysics, and biology, ranging from studying the conformational dynamics of proteins in vitro to imaging live cells. Since most biological molecules do not afford useful fluorescence properties in this regard, the majority of fluorescence measurements rely on the use of one or multiple extrinsic fluorophores. Therefore, there has been continuous effort to develop new biological fluorescence reporters that can meet different requirements and goals. For example, over the past few decades many analogues of tryptophan (Trp), which is the most fluorescent native amino acid in proteins, have been examined and explored (Hilaire et al., 2017; Lepthien et al., 2008; Lotte et al., 2004; Markiewicz et al., 2016; Moroz et al., 2013; Ross et al., 1997; Smirnov et al., 1997; Talukder et al., 2015), for the purpose of identifying Trp-based, unnatural amino acid (UAA) fluorophores that have improved photophyiscal properties over Trp. While Trp is a useful and convenient intrinsic fluorescence probe of protein structure and dynamics, its

57

application is limited to in vitro spectroscopic studies. This is because (1) its fluorescence quantum yield (QY) is relatively low (<0.2), (2) its fluorescence decay kinetics are complex, (3) its absorption spectrum is in the ultraviolet (UV) region (i.e., <300 nm), and (4) it has low photostability.

Recently, Hilaire *et al.* (Hilaire et al., 2017) have shown that a simple derivative of Trp, L-4-cyanotryptophans (L-4CN-Trp), exhibits rather a unique absorption and emission properties, making it useful in biological spectroscopic and imaging applications. In comparison to Trp, the much-expanded utility of L-4CN-Trp is due to the following factors: (1) it has a larger fluorescence QY (>0.8 in aqueous solution) and, a longer fluorescence lifetime (ca. 13 ns); (2) its absorption and emission spectra are red-shifted, resulting in fluorescence excitation and detection conditions accessible by commercial fluorescence microscopes; and (3) it is more photostable. While the study of Hilaire et al. has validated the potential biological utility of this blue fluorescence amino acid, its wide use is limited by lack of a facile method to reliably synthesize L-4CN-Trp and the corresponding Fmoc- or Boc-protected L-4CN-Trp for solid-phase peptide synthesis.

Herein, we aim to describe a simple, high-yielding, and cost-effective synthetic route for L-4CN-Trp and Fmoc-L-4CN-Trp and demonstrate the utility of L-4CN-Trp fluorescence in the study of peptide-membrane interactions, under both in vitro and in vivo conditions. Indeed, several methods for syntheses of 4CN-Trp or its derivatives have been reported in the literature and (**Figure 3.1** and **Table 3.1**) (Bartoccini et al., 2016; Boville et al., 2018; Hilaire et al., 2017; Romney et al., 2017; van Wilderen et al., 2018; Winn et al., 2018). For example, Bartoccini *et al.* (Bartoccini, 2016) described the synthesis of N-acetyl-4CN-Trp methyl ester from 4-pinacol boronated ester **1**. However, the product was a racemic mixture and the yield was low. Recently, another racemic synthesis of N-acetyl-4CN-Trp from 4-cyanoindole **2** was reported.18 Since the experimental details were not described in that study, we attempted to reproduce its result by the implementation of

similar reaction conditions (Blaser et al., 2008; Yokoyama et al., 1999, 2004) known for tryptophan analogs but the desired product was obtained in very low yield in our hands. In the study of Hilaire et al. (Hilaire et al., 2017), a short synthetic route to produce 4CN-Trp methyl ester from 4-bromo-L-tryptophan **3** in optically pure form was described (Hilaire et al., 2017). However, this approach is not ideal in practice due to the low yield in the Pd-catalyzed cyanation step, the use of toxic reagent, and the high cost of the starting material. Alternatively, L-4CN-Trp can be enzymatically synthesized from 4-cyanoindole **2** and serine by tryptophan synthase (TrpS) or a variant of tryptophan synthase subunit B (TrpB) which was recently generated through directed evolution by Arnold and others (Boville et al., 2018; Romney et al., 2017; Winn et al., 2018). While very useful, this genetic approach is limited to labs with expertise in protein expression as well as the availability of the corresponding enzyme. Thus, devising an efficient and expedient chemical synthesis method of enantiomerically pure L-4CN-Trp is highly desirable. In addition to the synthesis test the utility of this blue fluorophore as FRET pair with a membrane staining dye to image cell-membrane-bound peptides and to determine peptide-membrane binding constants.

3.3: EXPERIMENTAL SECTION

General. All reagents were used as received from commercial sources. Reactions were monitored through thin layer chromatography (TLC) on 0.25-mm SiliCycle silica gel plates and visualized under UV light. Flash column chromatography (FCC) was performed using Combiflash®Rf + Lumen UV-VIS/ELSD. NMR spectra were recorded using Bruker Avance-300 calibrated to CD_3OD (3.31 and 49.0 ppm for ^1H and ^{13}C NMR spectra, respectively), $(CD_3)_2CO$ (2.05 ppm for ^1H and 29.8, 206.3 ppm ^{13}C NMR spectra, respectively), D_2O (4.79 ppm for ^1H) as the internal reference. ^1H NMR spectral data are reported in terms of chemical shift (δ, ppm), multiplicity, coupling constant (Hz), and integration. ^{13}C NMR spectral data are reported in terms of chemical shift (δ, ppm). The following abbreviations indicate the multiplicities: s, singlet; d,

doublet; t, triplet; q, quartet; m, multiplet. Optical rotation was measured on a Jasco P2000 polarimeter and CD experiment was performed on a Jasco J-810 spectropolarimeter.

Synthesis of N-acetyl Amino Ester 5. To a solution of amino ester **4** (10.3 g, 40.0 mmol) in dry DCM (300 mL) under inert atmosphere at 0 ^{o}C were added dropwise and successively Et$_3$N (11.1 mL, 3.0 equiv) and acetyl chloride (3.2 mL, 1.1 equiv). The reaction mixture was stirred at 0 ^{o}C for 1 h and then allowed to stir for overnight at rt. The solution was quenched with a solution of sat. aq. NaHCO$_3$ and the resulting mixture was extracted with ethyl acetate. The combined organic layers were washed with brine, dried over Na$_2$SO$_4$, and concentrated under reduced pressure. The residue was purified by the column chromatography using CombiFlash®Rf + Lumen UV-VIS/ELSD (Teledyne ISCO) and a RediSep column [silica 120 g (60-Å pore size, 20–40 μm)] to provide the N-acetyl amino ester **5** (11.0 g, 92%). Ethyl 2-acetamido-3-(4-cyano-1H-indol-3-yl)propanoate **5**: Pale yellow solid; R_f = 0.52 (hexane/EtOAc, 3:1); ^{1}H NMR (300 MHz, CD$_3$OD) δ 7.68–7.65 (m, 1H), 7.46–7.43 (m, 1H), 7.32 (s, 1H), 7.24–7.19 (m, 1H), 4.81 (dd, J = 6.7, 8.3 Hz, 1H), 4.10 (q, J = 7.1 Hz, 2H), 3.55 (dd, J = 6.4, 14.7 Hz, 1H), 3.36 (dd, J = 8.2, 14.7 Hz, 1H), 1.95 (s, 3H), 1.14 (t, J = 7.1 Hz, 3H); ^{13}C NMR (75 MHz, CD$_3$OD) δ 172.1, 171.8, 136.9, 127.0, 126.4, 125.5, 120.7, 119.2, 116.4, 109.4, 100.6, 60.8, 53.7, 26.7, 20.9, 12.9.

Synthesis of N-acetyl Amino Acid 6: A mixture of the ester **5** (4.8 g, 16 mmol) in ethanol (50 mL) and LiOH (3.0 equiv, 1 M solution) stirred at rt overnight. The ethanol was evaporated in vacuo. The residue was diluted with water (50 mL), acidified to pH 2 using aqueous HCl (1N), and extracted with ethyl acetate. The combined organic phases were washed with NaHCO$_3$, brine and concentrated under reduced pressure to yield the acid (3.86g, 89%). 2-acetamido-3-(4-cyano-1H-indol-3-yl)propanoic acid **6**: Pale yellow solid; R_f = 0.30 (DCM/Methanol, 10:1); ^{1}H NMR (300 MHz, CD$_3$OD) δ 7.68–7.65 (m, 1H), 7.46–7.38 (m, 1H), 7.38 (s, 1H), 7.23–7.18 (m, 1H), 4.81 (dd, J = 5.1, 8.3 Hz, 1H), 3.63 (dd, J = 5.1, 15.1 Hz, 1H), 3.36 (dd, J = 8.4, 15.1 Hz, 1H), 1.93 (s, 3H); ^{13}C NMR (75 MHz, CD$_3$OD) δ 173.7, 171.8, 136.9, 126.9, 126.4, 125.5, 120.7, 119.2, 116.4, 109.7, 100.5, 53.2, 26.8, 21.0.

Synthesis of L-4CN-Trp 7 and Fmoc-L-4CN-Trp 8: Compound **6** (0.51 g, 2.0 mmol) was dissolved in PBS buffer solution (0.1 M, 25 mL) containing $CoCl_2 \bullet H_2O$ (0.125 mM) at pH 8.0. Amano acylase (0.5 g, >30,000 U/g) was added with stirring. The mixture was placed in a shaker at 37 °C for 48 h and then quenched by adjusting the pH to 5 with 1M HCl solution. After centrifugation, the precipitate was isolated and lyophilized to obtain crude L-4CN-Trp **7**, which was subsequently used in the next step without purification. A small amount of sample was isolated by RP-HPLC for analytical analysis, whose spectroscopic data is in good agreement with literature values (Winn, 2018). (S)-2-amino-3-(4-cyano-1H-indol-3-yl)propanoic acid **7,** white solid; ^{1}H NMR (300 MHz, D_2O + 100 mM DCl) δ 7.42 (d, J = 8.01 Hz, 1H), 7.18 (d, J = 7.2 Hz, 1H), 7.11 (s, 1H), 6.94 (t, J = 7.89 Hz, 1H), 4.06 (dd, J = 5.6, 9.6 Hz, 1H), 3.31 (dd, J = 5.6, 15.3 Hz, 1H), 2.97 (dd, J = 9.6, 15.3 Hz, 1H); ^{13}C NMR (75 MHz, D_2O + 100 mM DCl) δ 170.9, 136.2, 128.5, 126.3, 125.2, 121.3, 119.8, 117.4, 105.7, 98.9, 53.4, 25.4; HRMS (ESI) m/z 230.0909 [$(M^+H)^+$; calcd for $C_{12}H_{13}N_3O_2^+$ $(M^+H)^+$: 230.0930]; $[\alpha]_D^{25}$=-209.64 (c= 2.5 mg/mL, in 1 M HCl).

The aforementioned crude **7** was dissolved in an aqueous Na_2CO_3 solution (0.1 M, 40 mL) followed by the addition of 9-fluorenylmethyl succinimidyl carbonate (0.67 g, 2.0 mmol) in THF (20 mL). The mixture was stirred for 2 h at rt. THF was removed under vacuo and the crude mixture was washed with Et_2O. The pH of the aqueous layer was adjusted to 2 using 3 M aqueous HCl and was extracted with DCM. The organic extracts were combined and washed with brine, dried over Na_2SO_4 and concentrated *in vacuo*. The residue was purified by the column chromatography to yield the product (280 mg, 62% of the theoretical yield over two steps). (S)-2-(((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-3-(4-cyano-1H-indol-3-yl)propanoic acid **8** Pale yellow solid; R_f = 0.20 (DCM/methanol, 10:1); ^{1}H NMR (300 MHz, $(CD_3)_2CO$) δ 10.74 (s, 1H), 7.86–7.84 (m, 2H), 7.78–7.75 (m, 1H), 7.67–7.64 (m, 2H), 7.55–7.50 (m, 2H), 7.43–7.38 (m, 2H), 7.32–7.24 (m, 3H), 6.88 (d, J = 8.0 Hz, 1H), 4.81–4.73 (m, 1H), 4.28–4.18 (m, 3H), 3.77–3.71 (m, 1H), 3.50–3.42 (m, 1H); ^{13}C NMR (75 MHz, $(CD_3)_2CO$) δ 172.8, 156.1, 144.1, 141.1, 136.9,

127.6, 127.4, 127.0, 126.7, 125.7, 125.3 (2C), 121.1, 119.9, 119.3, 116.7, 110.5, 101.4, 66.3, 54.6, 47.0, 27.0. HRMS (ESI) m/z 452.1605 [(M$^+$H)$^+$; calcd for $C_{27}H_{22}N_3O_4^+$ (M$^+$H)$^+$: 452.1610].

Determination of Optical Purity: Marfey's reagent (1-fluoro-2-4-dinitrophenyl-5-L-alanine amide) was used as a solution in acetone (33 mM). In a 2 mL vial, the amino acid (a racemic amino acid that was obtained by the hydrolysis (1M LiOH in THF-H$_2$O) of compound **4** or crude L-amino acid from the above resolution (0.5 μmol) was dissolved in 1 M aqueous NaHCO$_3$ (140 μL). Marfey's reagent (60 μL, 2 μmol) was added, then the vial was placed in an incubator (rpm 900) at 37 ^{o}C. After 1 h, the reaction mixture was allowed to cool to rt and then diluted with 1:1 water /acetonitrile (550 μL) and analyzed by HPLC [solvent A (0.1% TFA in water) and B (0.1% TFA, 1% water in acetonitrile), Vydac$^®$ 214TP, 5μm, C4, 300Å, 4.6 mm i.d. x 250 mm), wavelength 330 nm with a gradient of 5–95% B over 35 min].

Solid-phase Peptide Synthesis of L-4CN-Trp Labeled pHLIP: The sequence synthesized: NH$_2$-GGEQNPIYW$_{CN}$ARYADWLFTTPLLLLDLALLVDADEGT-CO$_2$H, where W$_{CN}$ stands for L-4CN-Trp. The peptide was synthesized on a 0.1 mmol scale on the preloaded Thr(OtBu)-HMPB-ChemMatrix resin (0.5 meq/g) using a Biotage Initiator + Alstra peptide synthesizer. The deprotection was carried out for 5 min at 70 ºC with 4.5 mL 20% 4-methylpiperidine in DMF. A standard coupling step (for all amino acids except for Fmoc-protected 4-CN-Trp) was done for 5 min at 75 ºC with 5 equivalents of Fmoc-protected amino acids, 4.98 equivalents of HCTU, and 10 equivalents of DIPEA (relative to the amino groups on resin) in DMF. For L-Fmoc-protected 4-CN-Trp, a coupling reaction was done for 5 min at 75 ºC with 1.5 equivalents of Fmoc-protected amino acids, 1.49 equivalents of HCTU, and 3 equivalents of DIPEA (relative to the amino groups on resin) in DMF. Peptide cleavage was carried out in the presence of TFA//TIS/H$_2$O (95:2.5:2.5, v/v) for 2 h at rt. The crude peptide was obtained after precipitation in cold diethyl ether and purified by RP-HPLC. (Vydac C4 214TP1022) using solvent A (0.1% TFA in water) and B (0.1% TFA, 1% water in acetonitrile). After 5 min equilibration with 5% B at a flow rate of 5 mL/min, a gradient of 5–70% B in 35 min was used. The chemical entity and purity of synthesized peptides

were verified by a Shimazu AXIMA MALDI-TOF mass spectrometer and an HP 1100 analytical HPLC system, respectively. MS (MALDI-TOF): m/z 4077.73 (calcd $[M^+H]^+$ = 4077.51).

Solid-phase Peptide Synthesis of Ac-Gly-L-4CN-Trp-Gly: The peptide was synthesized on a 0.05 mmol scale using a Biotage Initiator+ Alstra peptide synthesizer. A typical solid-phase peptide synthesis reaction cycle includes Fmoc deprotection, washing, coupling, and post-coupling washing steps. The deprotection step was carried out for 5 min at 70 ºC with 4.5 mL 20% 4-methylpiperidine in DMF. A standard coupling step (for all amino acids except for Fmoc-protected L-4CN-Trp) was done for 5 min at 75 ºC with 5 equivalents of Fmoc-protected amino acids, 4.98 equivalents of HCTU, and 10 equivalents of DIPEA (relative to the amino groups on resin) in DMF. For Fmoc-protected L-4CN-Trp, a coupling reaction was done for 5 min at 75 ºC with 1.5 equivalents of Fmoc-protected amino acids, 1.49 equivalents of HCTU, and 3 equivalents of DIPEA (relative to the amino groups on resin) in DMF. Acetylation was done by using 5 equivalents of Ac_2O and 10 equivalents of DIPEA. Peptide cleavage was carried out in the presence of TFA/TIS/H_2O (95:2.5:2.5, v/v) for 2 h at rt. The crude peptide was obtained after precipitation in cold diethyl ether. The chemical entity and purity of synthesized peptides were verified by a Shimazu mass spectrometer and an HP 1100 analytical HPLC system, respectively. MS (ESI): m/z 385.3 (calcd $[M^+H]^+$ = 385.4).

Cell Culture: HeLa cells were grown overnight at 37 ºC on slides to about 60% confluency in DMEM medium containing 10% FBS and 4 mM L-Glutamine. At that point, the supernatant was aspirated, and 0.5 mL of a stock 4CN-Trp-pHLIP solution (in PBS buffer at the desired pH) was added to each slide with a final peptide concentration of 10 μM. Then the peptide treated cells were incubated at rt for 1 h, followed by washing with the corresponding PBS buffer 3✗ and fixed for 30 minutes with 2% formaldehyde in the corresponding PBS. Subsequently, each slide was again washed 3✗ with the corresponding PBS buffer and air-dried. Finally, two drops of 50% glycerol were then added to each slide, which was covered with a cover slip and the edges were sealed with clear nail polish.

Cell Imaging: Cell images were acquired on a commercial widefield fluorescence microscope (Leica DM6000) equipped with a 20X dry objective using a standard DAPI filter (excitation bandwidth: 325 – 375 nm, emission bandwidth: 435 nm – 485 nm). The integration time for each image was 50 ms, and data/image processing was carried out using ImageJ 1.5 software (Schneider et al., 2012).

Preparation of Large Unilamellar Vesicles (LUVs): The 100-nm LUVs used in the peptide-membrane binding experiments were composed of 99% POPC (purchased from Avanti Polar Lipids Inc., AL) and 1% of 3,3'-dioctadecyloxacarbocyanine perchlorate (DiO, purchased from Thermo Fisher Scientific), which were prepared following previously published procedures (Aurora et al., 1985). Briefly, an appropriate amount of DiO was added to a stock lipid solution in chloroform, and the resultant mixture was put to shake at rt for 30 minutes. This lipid solution was then allowed to dry under a flow of nitrogen to form a lipid film, which was followed by a 30-minute lyophilization to remove any remaining solvent. The resultant lipid film was then rehydrated in the desired buffers. This sample was then subjected to 7 rounds of slow vortexing, freezing, and thawing. The resulting vesicle solution was then extruded 11 times through an extruder (Avanti Polar Lipids Inc., AL) equipped with a 100 nm membrane. After extrusion, the LUV solution was diluted to 1.0 mM (lipid concentration) with desired buffers.

Fluorescence Measurements: All fluorescence spectra were collected on a Jobin Yvon Horiba Fluorolog 3.10 spectrofluorometer using a 1 cm quartz cuvette with a 1.0 nm resolution, 1 nm excitation/emission slit, an integration time of 1.0 nm/s, and an excitation wavelength of 320 nm at 20°C. Fit for fluorescence binding curve is detailed in the below.

Förster Distance Calculation: The Förster distance (R_0) of the FRET pair, L-4CN-Trp (donor) and DiO (acceptor), is calculated using the following equation (Lakowicz, 1999):

$$R_0^6 = \left(\frac{9000(ln10)\kappa^2 Q_D}{128\pi^5 N\eta^4}\right)J(\lambda),$$ (**3.3.1**)

where κ^2 is the orientation factor, assumed to be 2/3, Q_D is the fluorescence quantum yield of the donor in the absence of the acceptor which is 0.85 for L-4CN-Trp, N is the Avogadro's number, η is the refractive index of the medium (1.33 for MeOH), and $J(\lambda)$ is the overlap integral, determined by the following equation (Lakowicz, 1999):

$$J(\lambda) = \int_0^\infty F_D(\lambda)\varepsilon_A(\lambda)\lambda^4 d\lambda,$$ (**3.3.2**)

where $F_D(\lambda)$ is the area-normalized emission spectrum of the donor and $\varepsilon_A(\lambda)$ is the wavelength-dependent molar absorption coefficient of the acceptor.

Data Fitting: The fluorescence binding curve (i.e., Figure 3.6), obtained from the intensity at 512 nm of the fluorescence spectrum of every peptide-membrane mixture (i.e., Figure 3.7), was fit to the following membrane binding model (Reshetnyak et al., 2008):

$$P + \mathrm{n}L \rightleftharpoons PL_\mathrm{n}.$$
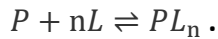
where P and L represent peptide and lipid, respectively. Following Engelman and coworkers (Reshetnyak et al., 2008), we assumed that n = 50 (i.e., every membrane-bound peptide is solvated by 50 lipids). For a given set of initial peptide ($[P_0]$ and lipid ($[L_0]$) concentrations, one can easily show that the equilibrium concentration of PL_n ($[PL_\mathrm{n}]_{eq}$ is:

$$[\mathrm{PL_n}]_{eq} = \frac{(K_d+[P_0]+[L_0]^*)-\sqrt{(K_d+[P_0]+[L_0]^*)^2-4[P_0][L_0]^*}}{2},$$ (3.3.3)

where K_d is the dissociation equilibrium constant and $[L_0]^* = [L_0]/50$. Because only the membrane-bound peptides can contribute to the observed FRET signal, the binding curve can be described by the following equation:

$$I = I_m \times [\mathrm{PL_n}]_{eq}.$$ (3.3.4)

In the fitting, K_D and I_m were treated as variables.

3.4: RESULTS AND DISCUSSION

Synthesis of L-4-Cyanotryptophan. Among the available methods for preparation of enantiomerically enriched α-amino acids (Nájera et al., 2007), enzymatic hydrolysis of racemic N-acyl amino acid via an acylase has the advantage of being able to mass-produce optically pure, non-proteogenic L-amino acid products (Chenault et al., 1989; Duthaler et al., 1994). In addition, production of D-amino acid may be achievable by using of a D-acylase (Tsai et al., 1992). Thus, we set to synthesize the required resolution substrate N-acyl amino acid **6** starting from the previously known amine **4** (**Figure 3.2**) by conventional chemistry. The racemic amine **4** was easily prepared from 4-cyanoindole **2** following a well-established 3-step procedure (i.e., Mannich reaction, alkylation by ethyl nitroacetate, and Zn/AcOH reduction). This approach has the advantage of eliminating the use of toxic reagents and enables the preparation of the desired product on a multi-gram scale with high efficiency. Next, acetylation provided the N-acetyl ester **5** in 92% yield, and the mild hydrolysis by LiOH furnished the racemic N-acetyl carboxylic acid **6** in

66

89% yield. The pivotal step, enzymatic resolution of the racemic N-acetyl amino acid **6** proceeded smoothly by Amano acylase to provide the desired L-4CN-Trp **7**. It is worth noting that the solubility of L-4CN-Trp **7** in aqueous media is negligible and the crude amino acid product could be easily separated from the unreacted material by centrifugation. Production of L-4CN-Trp **7** with a high enantiomeric excess (over 99%) was confirmed by HPLC analysis of Nα-(2,4-Dinitro-5-fluorophenyl)-L-alaninamide (FDAA) derivatives (**Figure 3.3**) Finally, Fmoc protection of L-4CN-Trp **7** under conventional reaction condition provided Fmoc-L-4CN-Trp **8** in good yield (62 % over the two steps).

L-4CN-Trp Labeled pHLIP Membrane Binding Study. To further explore the biological utility of L-4CN-Trp, we incorporated it into the pH-(Low) Insertion Peptide (pHLIP) and studied the interaction of the mutant peptide with model and cell membranes using both fluorescence spectroscopic and imaging techniques. The pHLIP peptide (sequence: GGEQNPIYWARYADWLFTTPLLLLDLALLVDADEGT) is a pH-dependent, membrane-interacting peptide designed and well-characterized by Engelman and coworkers (Andreev et al., 2007; Hunt et al., 1997; Reshetnyak et al., 2006, 2007; Zoonens et al., 2008). Specifically, the pHLIP peptide weakly binds to membranes as unstructured monomers at neutral pH, whereas at pH < 6.5 it forms a transmembrane α-helix. Specifically, we replaced the N-terminal Trp (i.e., Trp9) of pHLIP with L-4CN-Trp. Synthesis of this 4CN-Trp-containing pHLIP peptide (hereafter referred to as 4CN-Trp-pHLIP) via microwave assisted solid phase peptide synthesis using Fmoc-L-4CN-Trp 8 produced the desired pHLIP mutant with a yield that is comparable to that of synthesizing the wide-type pHLIP, hence providing additional support of the success of the synthesis protocol for **7** and **8**.

The interaction between 4CN-Trp-pHLIP and membranes was first investigated via fluorescence microscopy (**Figure 3.4**). HeLa cells, which were incubated with 4CN-Trp-pHLIP (10 μM) in either pH 7.4 or pH 6.0 PBS buffer for 30 minutes and then thoroughly washed 3 times with the corresponding buffer clearly show that the fluorescence intensity of L-4CN-Trp increases

with decreasing pH. These results, which are similar to those obtained with pHLIP peptides labeled with a fluorescent dye (Reshetnyak et al., 2006), not only show that the Trp to L-4CN-Trp mutation does not change the pH-responsive membrane-binding property of pHLIP, but also indicates that the fluorescence of L-4CN-Trp, which was collected through a Leica DM6000 widefield fluorescence microscope using a standard DAPI filter set (i.e., excitation and emission optical filters designed for the fluorescent dye DAPI, a commonly used stain for cell nucleus), is bright enough for biological imaging applications. Next, we sought to demonstrate that the fluorescence of L-4CN-Trp can be used to quantify the peptide-membrane interaction of interest, using 4CN-Trp-pHLIP as a model. The absorption spectrum of 3,3'-dioctadecyloxacarbocyanine perchlorate (DiO), a green lipophilic dye, overlaps significantly with the emission spectra of L-4CN-Trp (**Figure 3.5**). This, along with the calculated ca. 62 Å Förster distance (R_0) in MeOH (see details in experimental section), indicates that L-4CN-Trp and DiO can be used an efficient fluorescence resonance energy transfer (FRET) pair. Because DiO is a universal membrane stain (Honig et al., 1986), we propose that this FRET pair can be used to assess peptide-membrane interactions. Our working hypothesis is that binding of an L-4CN-Trp-containing peptide to a membrane that is stained with DiO will lead to an observable FRET signal that can be used to assess the underlying binding thermodynamics and kinetics. To test this hypothesis, we first compared the fluorescence spectra of 4CN-Trp-pHLIP (1.0 µM) collected in water (no DiO was added due to its negligible solubility) and in aqueous solutions containing DiO-stained (1%), 100-nm, large unilamellar vesicles (LUVs) made of POPC (1.0 mM) (**Figure 3.6**).

When excited at 320 nm, where DiO has a negligible absorbance (**Figure 3.5**), only the peptide-containing samples show observable FRET signals, indicating peptide binding. Also, the FRET signal is significantly increased at acidic pH. Since the fluorescence intensity of the donor (i.e., L-4CN-Trp) is essentially independent of pH in the range of the experiments (**Figure 3.7**), this result is therefore consistent with the designed pH-responsive behavior of pHLIP. To further determine the binding constant of 4CN-Trp-pHLIP to POPC membranes at pH 4.0, we collected the fluorescence spectra of a series of solutions containing 1.0 mM POPC (in the form of 100-nm

68

LUVs stained with 1% DiO) and 4CN-Trp-pHLIP of varying concentrations ([4CN-Trp-pHLIP]). As indicated (**Figure 3.8**) the FRET signal increases with increasing [4CN-Trp-pHLIP] and starts to level off at ca. 30 μM, suggesting that the dissociation constant (K_d) in the range of 20-50 μM. Indeed, fitting the corresponding fluorescence signals (i.e., intensities at 512 nm) to a binding mode previously used by Engelman and co-workers (Reshetnyak et al., 2008), which assumed that 1 pHLIP peptide interacts with 50 lipid molecules, yielded a K_d of ca. 14 μM (**Figure 3.9**). This value is in agreement with that previously reported for pHLIP (i.e., K_d = 8 μM at low pH) (Reshetnyak et al., 2008), hence validating the usefulness the 4CN-Trp-Dio FRET pair for quantitatively assessing peptide-membrane binding interactions.

While L-4CN-Trp can be incorporated into peptides via solid-phase peptide synthesis, its utility would be significantly expanded if it can also be genetically incorporated into proteins through the protein synthesis machinery of the cell. While the development of such a method is beyond the scope of the current study, we tested the toxicity of L-4CN-Trp to cells. This is because tryptophan analogs could intervene in the tryptophan metabolic pathway, leading to cell death. As indicated (**Figure 3.10**) at 250 μM L-4CN-Trp does not show any significant growth inhibition of *E. coli* cells, suggesting that this UAA is amenable for cellular applications.

3.5: CONCLUSIONS

In conclusion, we developed an improved method for the selective synthesis of L-4CN-Trp and its Fmoc-protected version that can be used for incorporation of this blue fluorescence amino acid into polypeptides via solid phase peptide synthesis. In addition, we demonstrated that this UAA fluorophore is a viable fluorescence reporter to monitor the binding of peptides to cell membranes. Moreover, we devised and validated a new fluorescence method to assess peptide-membrane interactions, which is based on the application of a FRET pair (R_0 = 62 Å) consisting of L-4CN-Trp (donor) and a universal membrane stain, DiO (acceptor).

Table 3.1: Summary of reported syntheses of 4CN-Trp and its derivatives.

Reference	Key reaction	Advantages
J. Org. Chem., 2018, **83,** 7447.	TrpB variant, aqueous buffer	Green chemistry, use of readily available starting material, high enantiomeric purity
J. Am. Chem. Soc., 2017, **139,** 10769.	TrpB variant · >20 Trp analogs up to 99% yield up to 99% ee	Green chemistry, use of readily available starting material, synthesis of Trp analogues with various substitution
Angew. Chem., Int. Ed., 2018, **57,** 6830.	StTrpS · PLP	Green chemistry, use of readily available starting material, high enantiomeric purity
Org. Biomol. Chem., 2016, **14,** 10095.	Suzuki-Miyaura	Gram-scale and rapid synthesis
Phys. Chem. Chem. Phys., 2018, **20,** 19906.	CN at 4- or 5-position · Ac_2O, AcOH, rt	Facile method to synthesis a variety of Trp derivatives
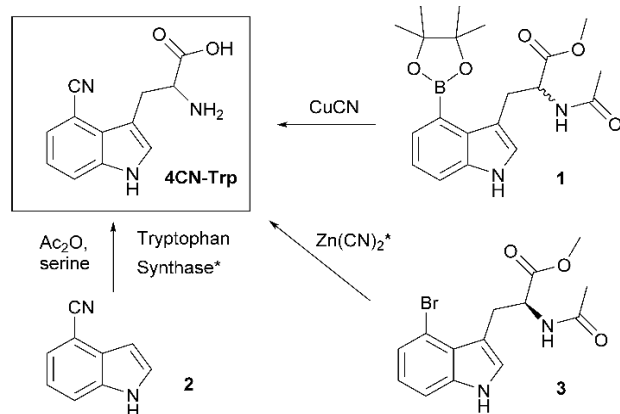
Figure 3.1: Previous synthetic approaches for 4CN-Trp and its derivatives. * denotes a route to enantiomerically enriched 4CN-Trp.
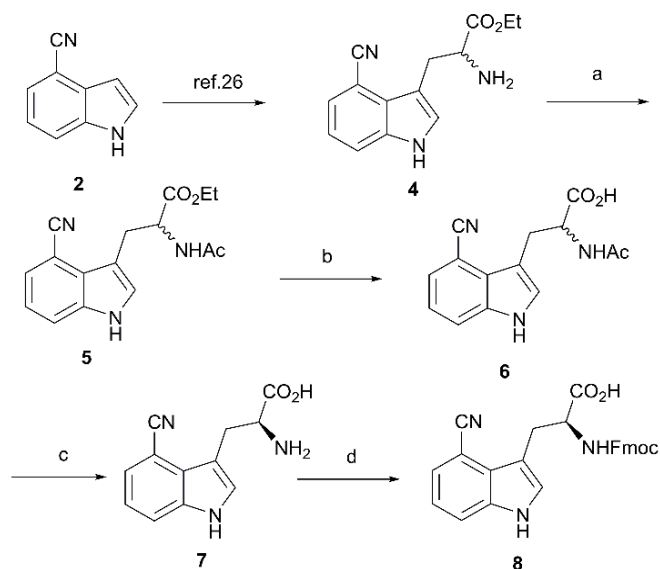
Figure 3.2: Synthetic route to L-4CN-Trp 7 and Fmoc-L-4CN-Trp 8. Reagents and conditions: a. AcCl, Et3N, 92%; b. LiOH, EtOH, 89%; c. Amano acylase, pH = 8.0; d. Fmoc-Osu, NaHCO3, 62% over c and d.
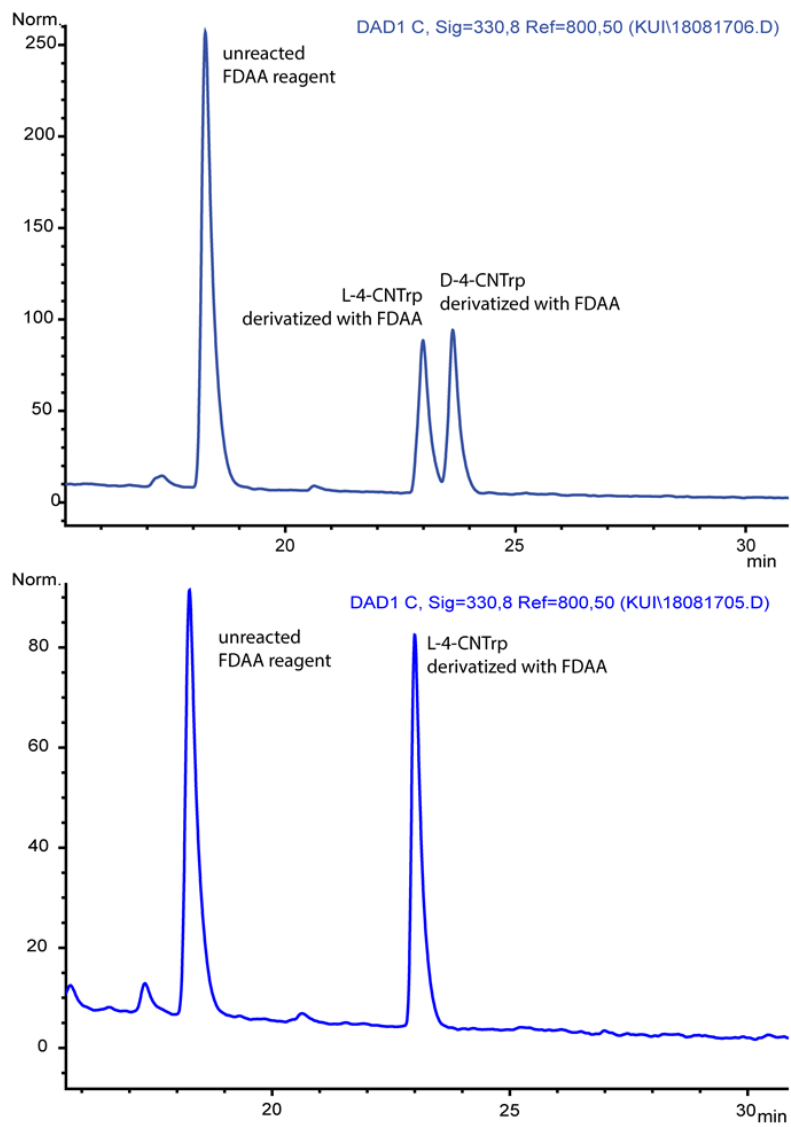
Figure 3.3: HPLC profile of FDAA derivatized racemic 4CN-Trp (Top) and L-4CN-Trp (Bottom).
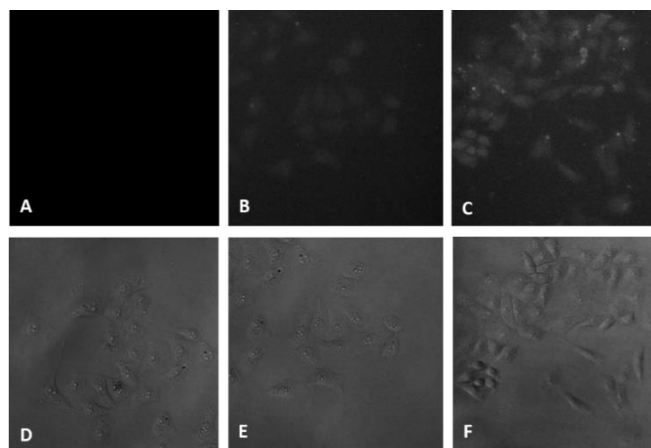
Figure 3.4: Widefield fluorescence images of HeLa cells treated with PBS buffer (A), L-4CN-Trp-pHLIP at pH 7.4 (B), and 4CN-Trp-pHLIP at pH 6.0 (C). D-F are their corresponding brightfield images.
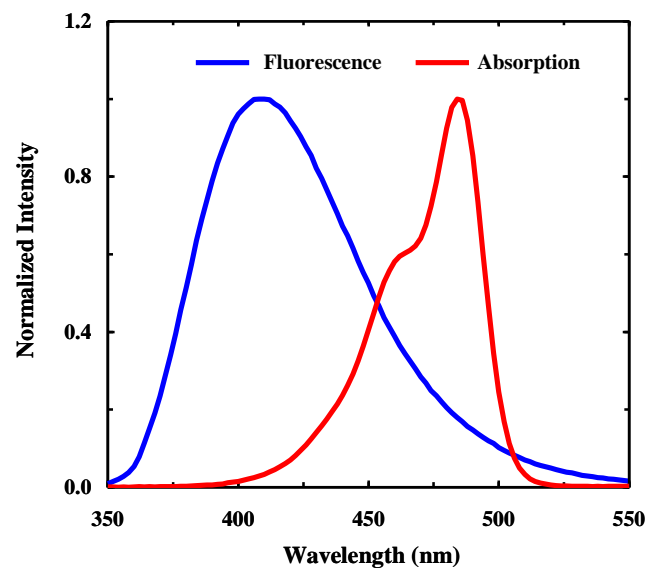
Figure 3.5: Normalized fluorescence spectrum of L-4CN-Trp (blue) and absorption spectrum of

DiO (red) in methanol.
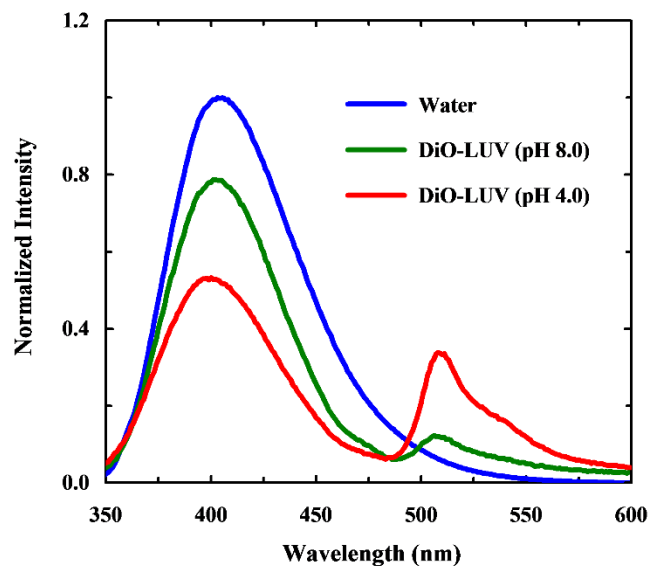
Figure 3.6: Normalized fluorescence spectra of 4CN-Trp-pHLIP collected in water and DiO stained POPC LUVs of different pH values, as indicated. The excitation wavelength was 320 nm.
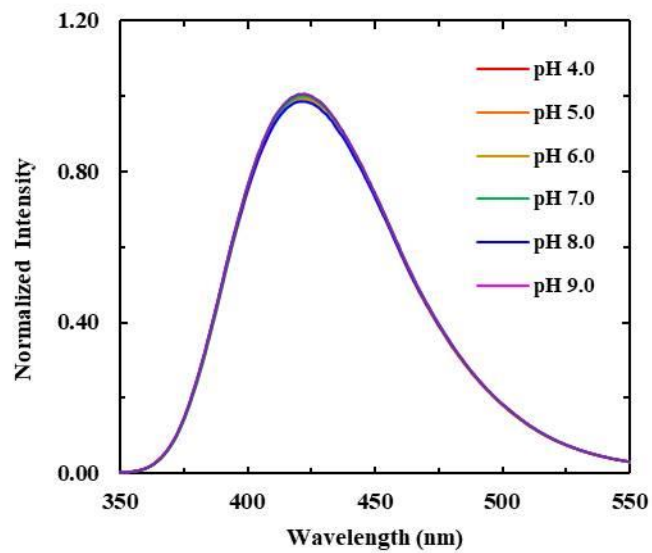
Figure 3.7: Normalized fluorescence spectra of Ac-G-L-4CN-Trp-G at different pH values (as indicated), using the spectrum obtained at pH 7.0 as the reference. The concentration was 10 µM for each sample, and the excitation wavelength was 320 nm.
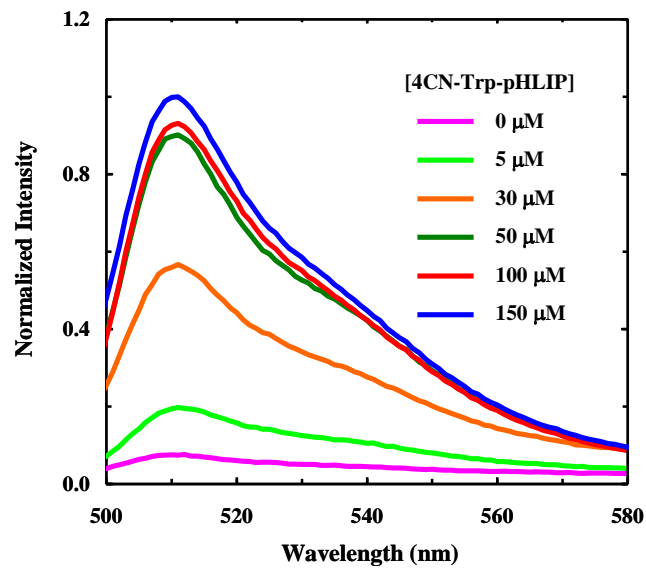
Figure 3.8: Normalized fluorescence spectra (in the DiO emission wavelength region) of mixtures of DiO-stained POPC (1.0 mM) LUVs and 4CN-Trp-pHLIP of different concentrations, as indicated. The excitation wavelength was 320 nm.
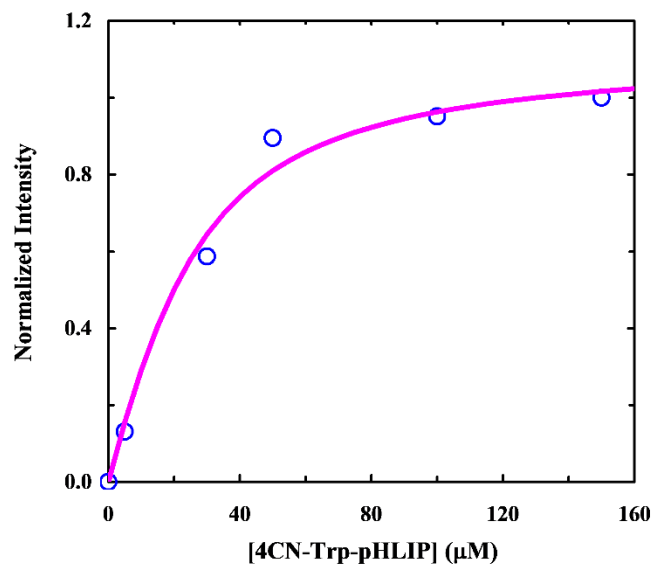
Figure 3.9: FRET intensity of DiO at 512 nm as a function of 4CN-Trp-pHLIP concentration. Smooth line is the fit of these data to a binding model described in the text.
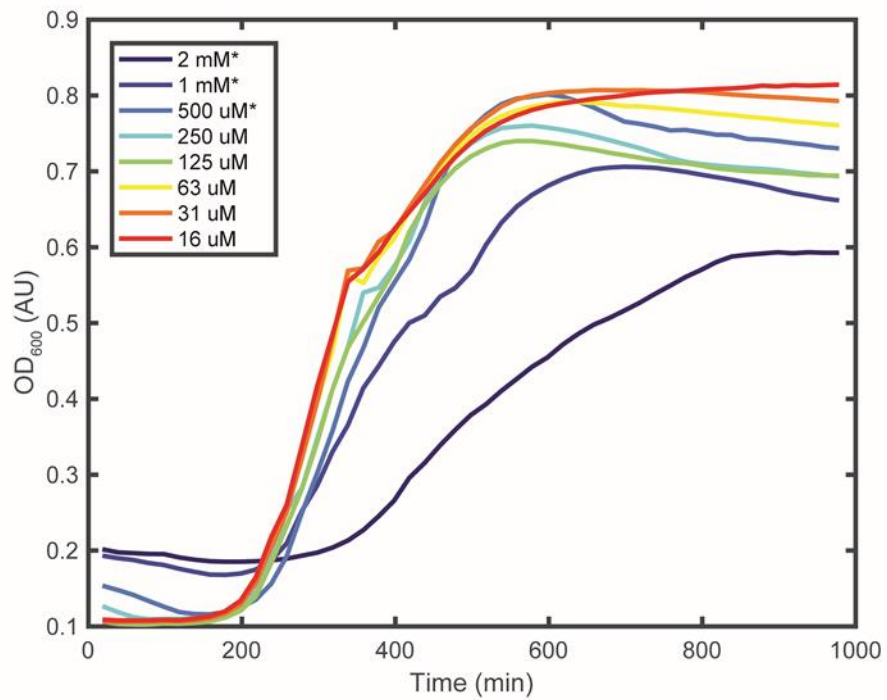
Figure 3.10: *E. coli* growth curves in the presence of different concentrations of L-4CN-Trp, which suggest that at concentrations of below 250 µM, L-4CN-Trp has no significant effect on the growth of *E. coli* cells. It is worth noting that for experiments carried out at higher concentrations, marked with an asterisk (*), the results are not conclusive because of the poor solubility of the compound at these concentrations.
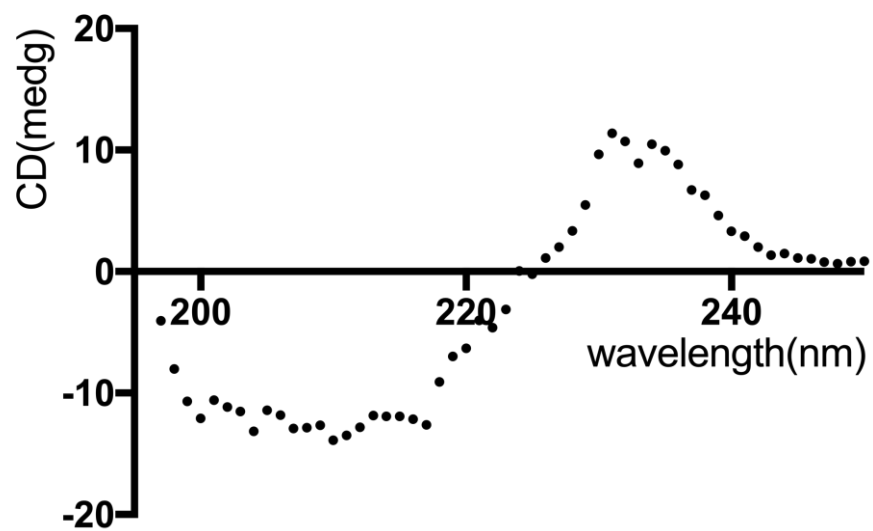
Figure 3.11: CD spectrum of L-4CN-Trp 7 (2 mM, in 0.1 M HCl)
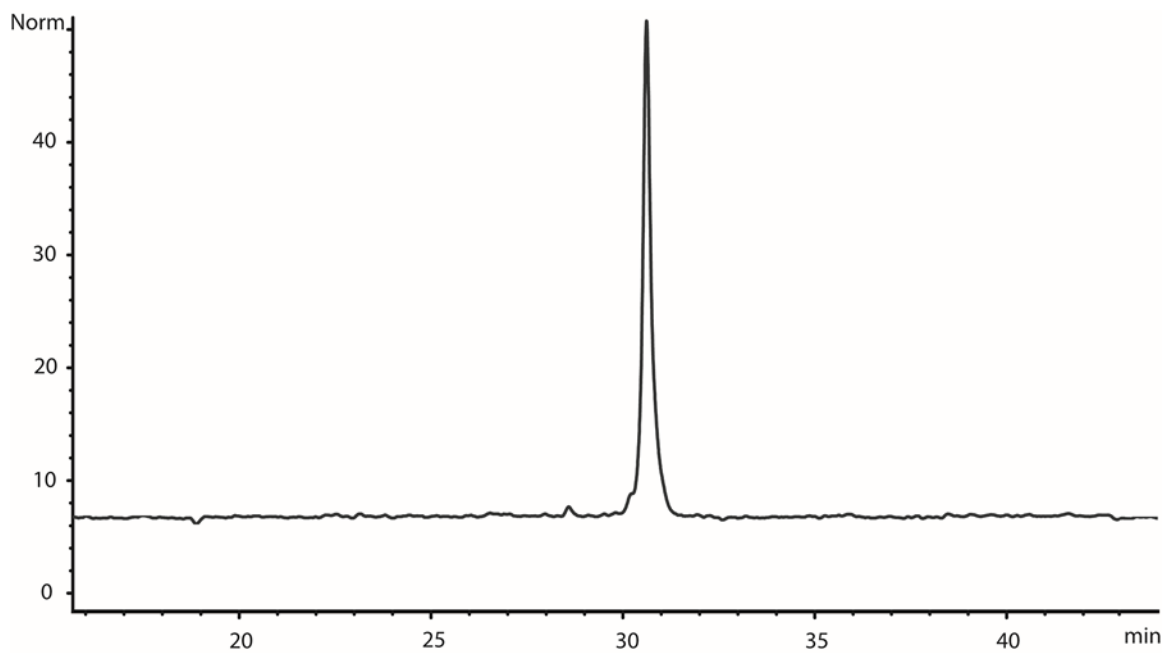
Figure 3.12: HPLC profile of purified 4CN-Trp-pHLIP.
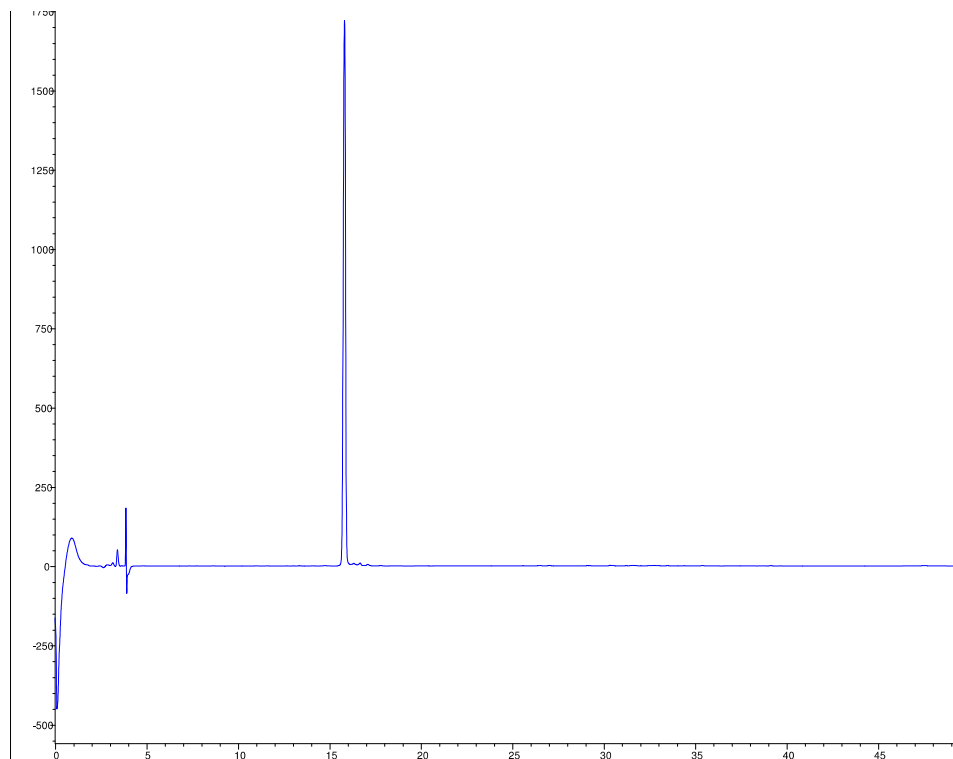
Figure 3.13: HPLC profile of purified Ac-G-L-4CN-Trp-G.
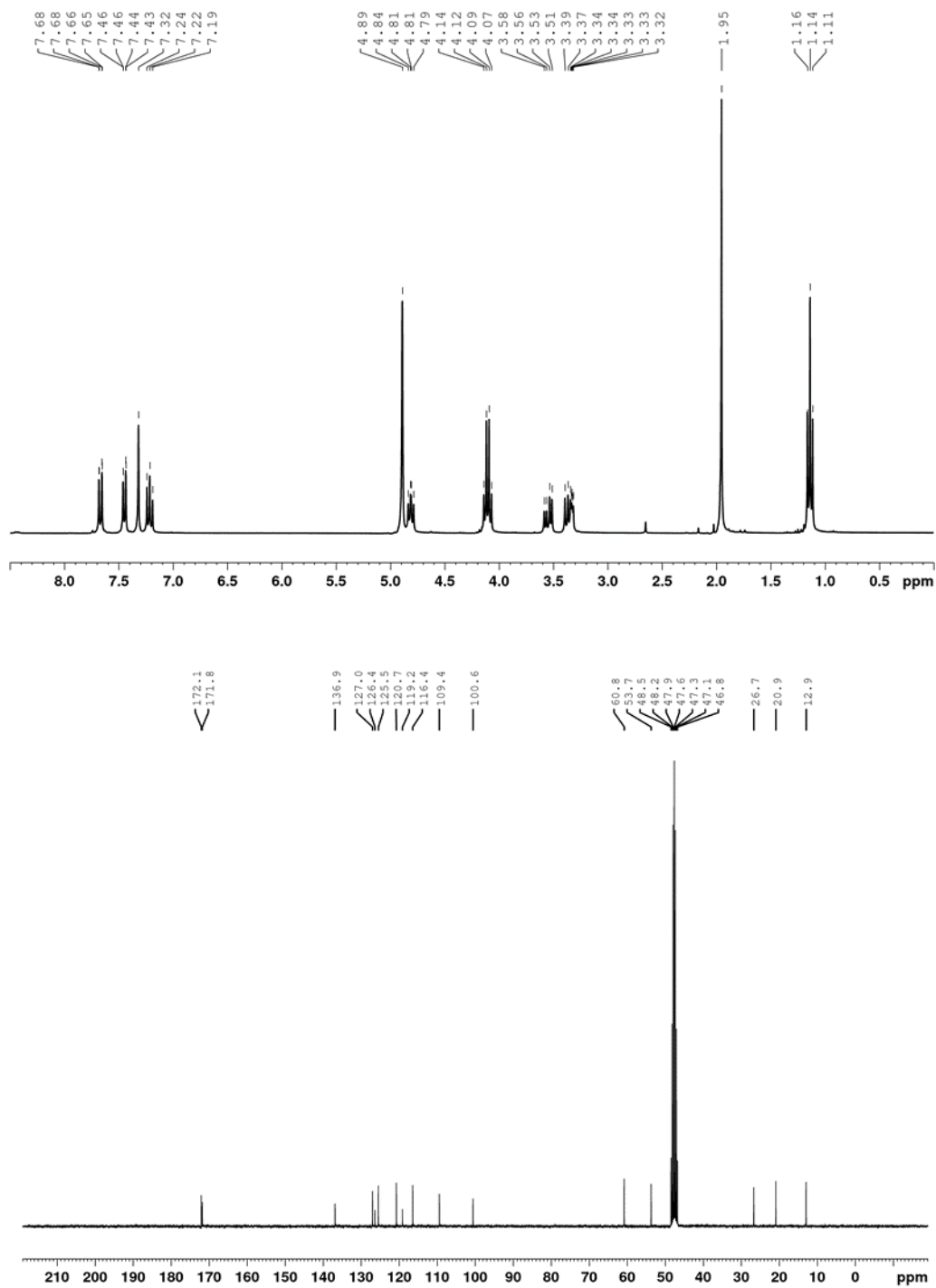
Figure 3.14: ^{1}H NMR and ^{13}C NMR spectrum of **5**.
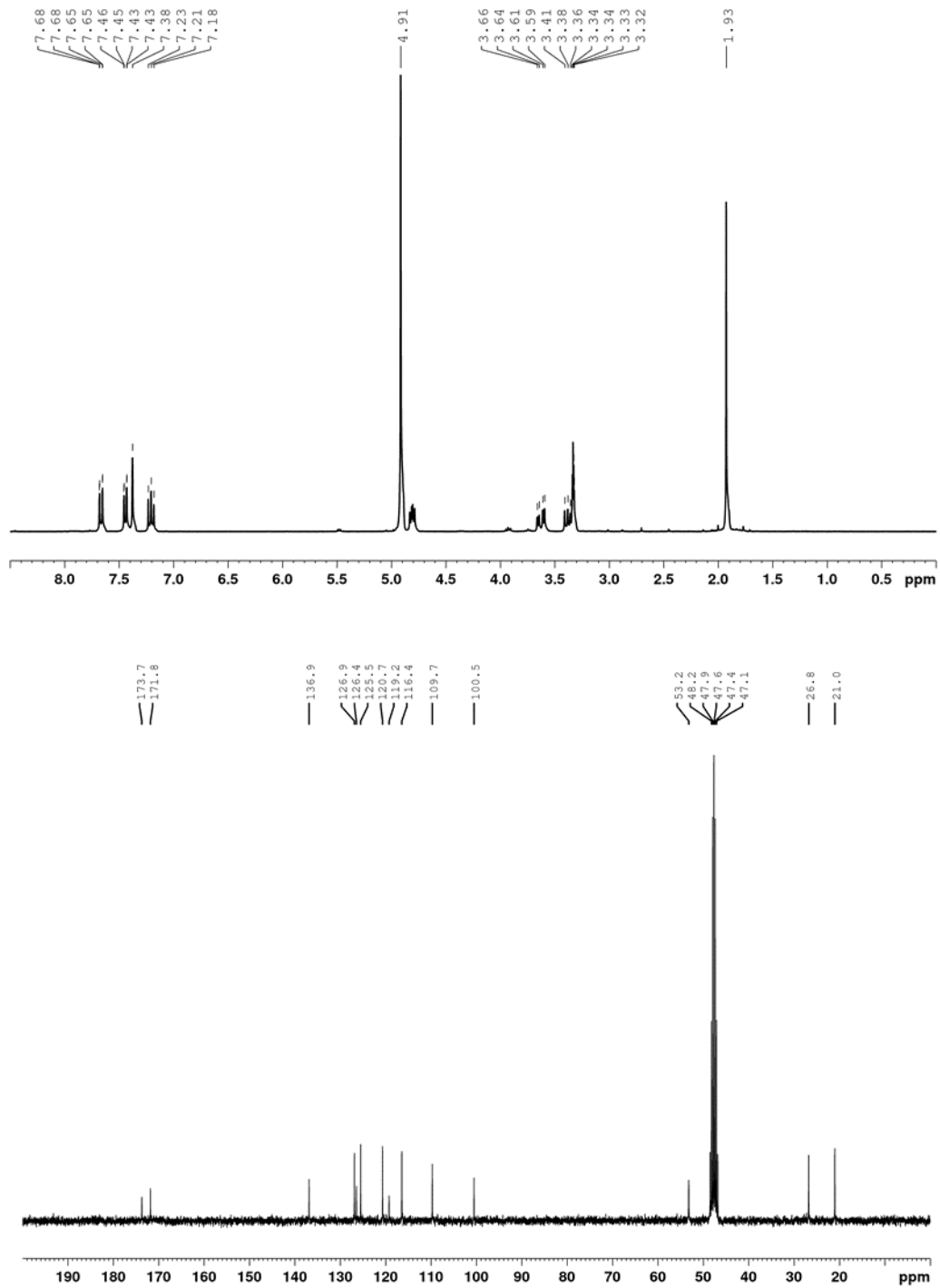
Figure 3.15: ^{1}H NMR and ^{13}C NMR spectrum of **6**.
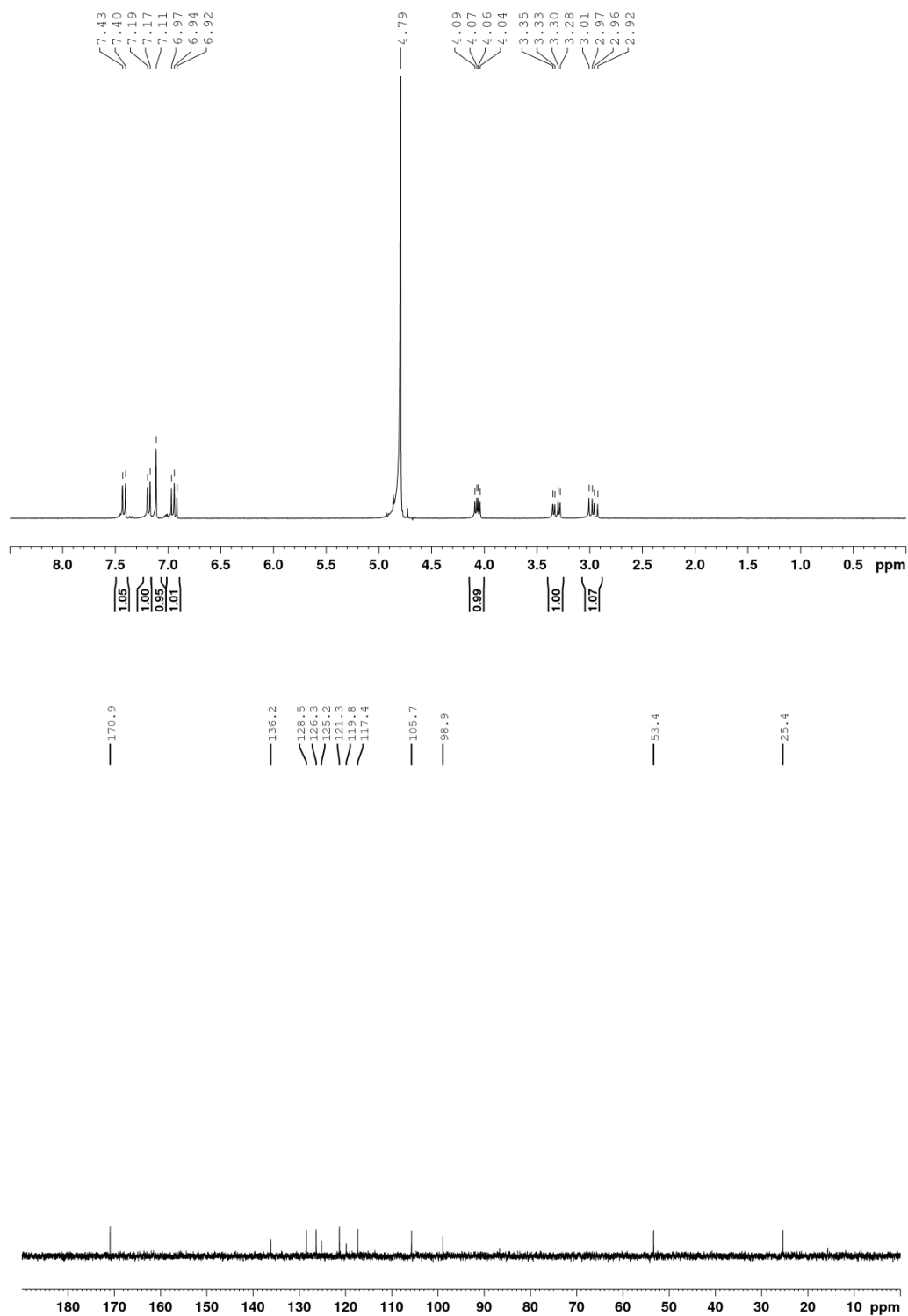
Figure 3.16: ^{1}H NMR and ^{13}C NMR spectrum of **7**.
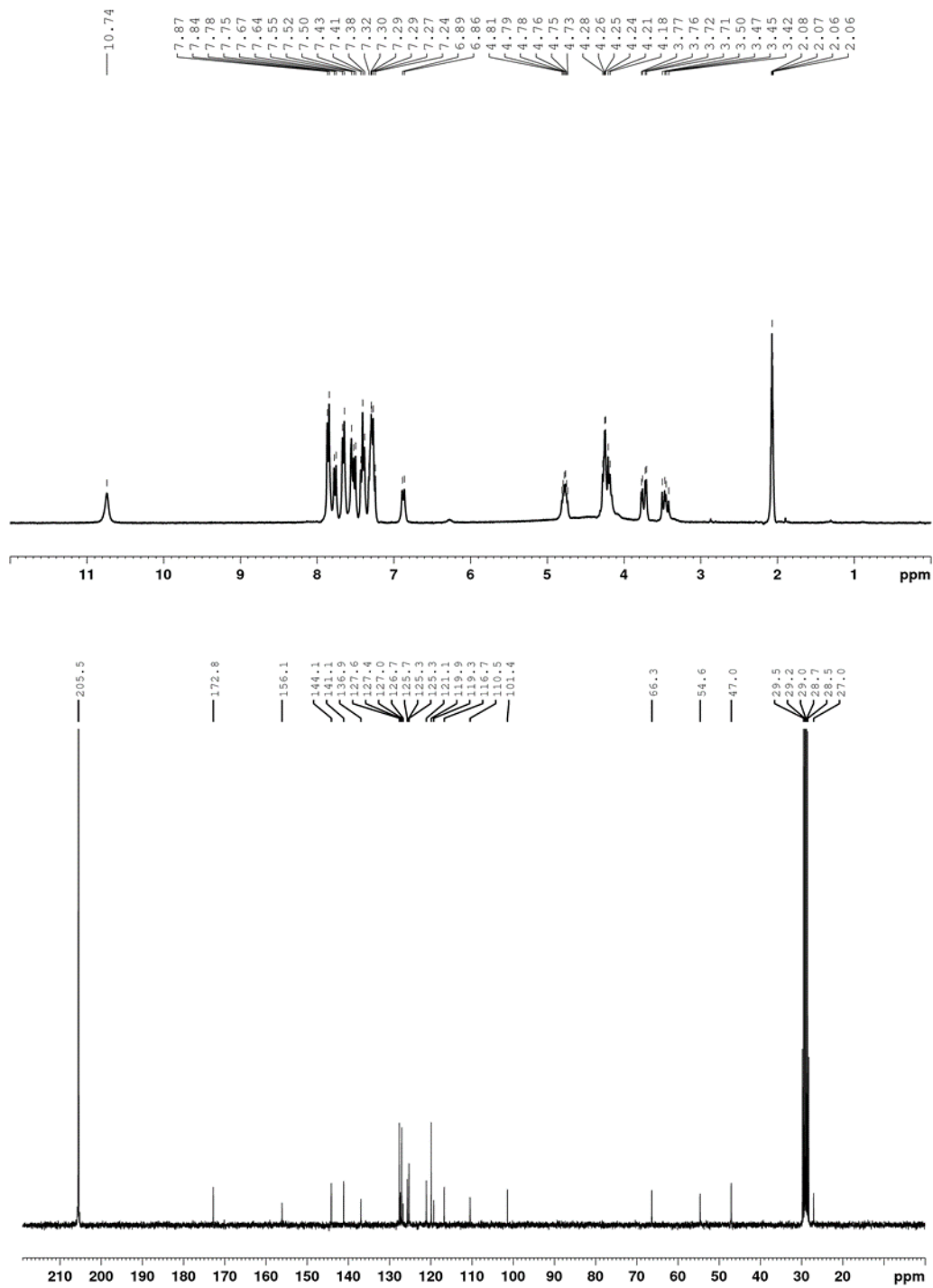
Figure 3.17: ^{1}H NMR ^{13}C NMR spectrum of **8**.

4: PET and FRET Utility of an Amino Acid Pair: Tryptophan and 4-Cyanotryptophan

This chapter was submitted for publication to Phys. Chem. Chem. Phys.

4.1: ABSTRACT

Methods based on fluorescence resonance energy transfer (FRET) and photo-induced electron transfer (PET) are widely used in the biological sciences, employing mostly dye-based FRET and PET pairs. While very useful and important, dye-based reporters are not always applicable without concern, for example, in cases where the fluorophore size needs to be minimized. Therefore, development and characterization of smaller, ideally amino acid-based PET and FRET pairs will expand the biological spectroscopy toolbox to enable new applications. Herein, we show that, depending on the excitation wavelength, tryptophan and 4-cyanotrptophan can interact with each other via the mechanism of either energy or electron transfer, hence constituting a dual FRET and PET pair. The biological utility of this amino acid pair is further demonstrated by applying it to study the end-to-end collision rate of a short peptide, the mode of interaction between a ligand and BSA, and the activity of a protease.

4.2: INTRODUCTION

Fluorescence emission is one of the most utilized physical properties in biological studies, enabling the assessment of a wide range of phenomena and problems. For example, it is especially useful to reveal information about the structure, dynamics, and interactions of biological molecules, both *in vitro* and *in vivo*, when the fluorescence signal of interest exhibits a dependence on a specific distance parameter. In this context, fluorescence intensity modulation based on fluorescence resonance energy transfer (FRET) (Clegg et al., 1996; Förster, 1948, 1965) or photo-induced electron transfer (PET) (Neuweiler et al., 2005; Seidel et al., 1996) is most widely exploited. The rate of FRET (Lakowicz, 1999) depends on $1/r^6$, where r is the separation distance between the fluorophore and modulator (i.e., acceptor in FRET and quencher

in PET), whereas the rate of PET (Lakowicz, 1999) shows an exponential dependence on r. Therefore, in practice methods based on these two mechanisms are often used separately. As a result, the currently available fluorophore-modulator pairs are designed to perform a specific function, either FRET or PET. Hence, it would be quite useful to devise fluorophore-modulator pairs that can do both, as such pairs would provide not only convenience but also allow extraction of different types of information from a single fluorophore-modulator system. For example, when tracking a conformational motion that brings two initially separated sites to proximity, FRET can be used to monitor the early position of the trajectory, while PET is useful to provide information at later times when r becomes too short for FRET to capture any changes. Herein, we show that the unnatural amino acid (UAA) 4-cyanotryptophan (4CN-Trp) and endogenous tryptophan (Trp) constitute a dual FRET and PET pair.

In protein conformational studies, using an amino acid-based FRET or PET pair offers additional advantages, including (1) less perturbation to the native system (Waegele et al., 2011), (2) ease of incorporation, and (3) site-specificity (Serrano et al., 2012). In contrast, dye-based FRET and PET pairs are often bulky and more hydrophobic in comparison to amino acid sidechains and, therefore, likely to introduce adverse structural perturbations to the protein. For example, several recent studies showed that many commonly used fluorescent dyes could interact with each other, causing inaccurate results in protein conformational studies (Riback et al., 2018a, 2018b). Also, labeling a protein with two dyes through chemical reactions with amino acid sidechains could yield incomplete or unwanted products, hence leading to undesirable spectroscopic results (Karolin et al., 1998; Waegele et al., 2011).

Recently, Hilaire *et al.* (Hilaire et al., 2017) found that the absorption and emission spectra of 4CN-Trp (and its fluorophore 4CN-indole) are significantly red-shifted from those of Trp (and indole). As a result, the fluorescence spectrum of Trp exhibits a significant overlap with the absorption spectrum of 4CN-Trp. Therefore, it is apparent that Trp can be used as a FRET donor to 4CN-Trp. However, it is not obvious that Trp is also capable of modulating the fluorescence intensity of 4CN-Trp via the mechanism of PET. We hypothesize that this could occur based on

the fact that Trp can effectively quench the fluorescence of oxazine- and rhodamine-based dyes through this mechanism (Neuweiler et al., 2003, 2005, 2007). To validate this hypothesis, we carried out fluorescence titration experiments using free fluorophores as well as fluorescence intensity and lifetime measurements on a series of 4CN-Trp-containing peptides. Indeed, our results support the notion that Trp can quench the fluorescence of 4CN-Trp via PET and, by quantitatively analyzing the corresponding Stern-Volmer plot, we further determined the electron transfer rate constant, $k_{\mathrm{ET}}(r) = k_0 e^{-\beta(r-a_0)}$. Considering that 4CN-Trp is only a two-atom modification of Trp and that it has many desirable photophysical properties as a fluorescent UAA, such as high fluorescence quantum yield (QY) (>0.8), long fluorescence lifetime (13.7 ns), and good photostability (Hilaire et al., 2017, 2017), we expect that this amino acid pair will find various biological applications. For example, in three proof-of-principle studies, we demonstrated that it could be used to assess the end-to-end collision rate of peptides, protein-ligand interactions, and protease activity.

4.3: EXPERIMENTAL SECTION

Sample Preparation. Amino acids, N-acetyl-L-tryptophanamide (NATA), and 4-cyanoindole-3-acetic acid (4CNI-3AA) were purchased from Chem-Impex (Wood Dale, Il), Sigma (St. Louis, MO), and Ark Pharm (Arlington Heights, Il), respectively. All peptides were prepared by standard 9-fluorenylmethoxy-carbonyl (Fmoc) solid phase peptide synthesis method. Because the Fmoc protected 4-cyanotryptophan was not available commercially at the time of this study, the 4-cyanoindole fluorophore in all the peptides was introduced manually by coupling 4CNI-3AA to the N-terminus of the respective peptides through the formation of an amide bond. It has been shown that the photophysical property of the 4-cyanoindole fluorophore in such peptides, which only lacks the amino group at the N-terminus, is similar or identical to that of the 4-cyanoindole fluorophore in peptides made with 4-cyanotryptophan (Hilaire et al., 2017). Therefore, for convenience, all peptides were referred to as 4CN-Trp peptides. Each peptide was then purified by reverse-phase high-performance liquid chromatography on an Agilent Technologies 1260

Infinity (Santa Clara, CA) equipped with a Vydac C18 column (Berkshire, UK). The pH of the purified peptide solution was first titrated to ca. 7 using NaOH and then the resultant peptide solution was lyophilized. The mass of each peptide was verified using either a liquid-chromatography mass spectrometer from Waters (Milford, MA) or a matrix-assisted laser desorption/ionization mass spectrometer from Bruker (Billerica, MA). All peptide samples used in the fluorescence measurements were prepared by dissolving the lyophilized peptide in Millipore water (pH 7). The peptide concentration (10 – 35 µM) was determined optically using the absorbance of the 4-cyanoindole fluorophore at 325 nm with ε = 4,750 M^{-1} cm^{-1}.[1]

Static Fluorescence Measurement. All static fluorescence measurements were carried out on a Jobin Yvon Horiba Fluorolog 3.10 spectrofluorometer (Kyoto, Japan) at room temperature (unless indicated otherwise) using a 1 cm quartz cuvette, a 1.0 nm spectral resolution, and an integration time of 1.0 nm/s. The excitation wavelength for each measurement was either 270 or 330 nm, as indicated throughout the text.

Förster Distance Calculation. The Förster distance (R_0) of the Trp (donor) and 4CN-Trp (acceptor) FRET pair was calculated using the following equation (Lakowicz, 1999):

$$R_0^6 = \left(\frac{9000(ln10)\kappa^2 Q_D}{128\pi^5 N\eta^4}\right)J(\lambda),$$
(**4.2.1**)

where κ^2 is the orientation factor, assumed to be 2/3, Q_D is the fluorescence quantum yield of the donor in the absence of the acceptor which is 0.14 for Trp in water (Lakowicz, 1999; Suzuki, 2009), N is the Avogadro's number, η is the refractive index of the medium (1.33 for water), and $J(\lambda)$ is the overlap integral, determined by the following equation (Lakowicz, 1999):

$$J(\lambda) = \int_0^\infty F_D(\lambda)\varepsilon_A(\lambda)\lambda^4 d\lambda,$$
(**4.2.2**)

where $F_D(\lambda)$ is the area-normalized emission spectrum of the donor and $\varepsilon_A(\lambda)$ is the wavelength-dependent molar absorption coefficient of the acceptor.

Time-Resolved Fluorescence Measurements. Time-resolved fluorescence data were collected on a home-built time-correlated single photon counting (TCSPC) apparatus, the details of which have been described elsewhere (Markiewicz et al., 2016). Briefly, a Ti:Sapphire oscillator (800 nm, 85.0 MHz) was used to derive the 270 nm excitation pulse through third harmonic generation while an electro-optical pulse picking system (Conoptics, Inc.) was used to reduce the repetition rate of the pulses to 21 MHz if necessary. To minimize any inner filter effect, the excitation beam was positioned near the edge of the 0.4 cm quartz sample cuvette that faces the fluorescence collection optics. Selection of the 4CN-Trp emission under magic angle polarization condition was achieved by passing the fluorescence beam through a long bandpass 300 nm filter (Semrock FF01-300) and a 403/95 nm bandpass filter (Semrock) or a 450 nm longpass filter (Semrock). Detection of the fluorescence signal was done with an 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 either a single- or bi-exponential function using the FLUOFIT (Picoquant GmbH) program. All measurements were carried out at room temperature, and the fluorophore absorbance at 270 nm was in the range of 0.05 – 0.2.

Stern-Volmer Titration and Numerical Fitting. Static fluorescence quenching experiments (λ_{ex} = 330 nm) was carried out using free 4CNI-3AA and NATA. To determine the underlying quenching rate, we employed a diffusion model as described in detail elsewhere (Lakowicz, 1999; Mintzer et al., 2015) to numerically fit the experimental Stern-Volmer curve (Figure 4 in the main text). Briefly, the decaying fluorescence signal $I(t)$ is described by:

$$I(t) = I_0 \exp\left(-\frac{t}{\tau_0} - C_Q^0 \int_0^t k(t')\, d\,t'\right), \tag{4.2.3}$$

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

$$k(t) = \frac{4\pi}{C_Q^0} \int_{a_0}^{\infty} r^2 \, k_Q(r) C_Q(r,t) dr, \tag{4.2.4}$$

In the above equation, $C_Q(r,t)$ is the concentration of quencher at distance r from the fluorophore at time t, and $k_Q(r)$ is the distance-dependent quenching rate, defined as (Lakowicz, 1999):

$$k_Q(r) = k_0 e^{-\beta(r-a_0)}, \tag{4.2.5}$$

where k_0 is the quenching rate when the pair are in van der Waals contact (when $r = a_0$), and β is a constant. For this study, the value of a_0 was set to 7.0 Å, which corresponds to the sum of the van der Waals radii of two tryptophan sidechains. By defining a normalized concentration of the quencher as:

$$y(r,t) = C_Q(r,t)/C_Q^0, \tag{4.2.6}$$

we can model the time- and distance-dependent quenching using the following equation:

$$\frac{\partial}{\partial t} y(r,t) = -D\nabla^2 y(r,t) - k_Q(r)y(r,t), \tag{4.2.7}$$

We numerically solved for $C_Q(r,t)$ using the following initial conditions and boundary conditions:

$$y(r, t = 0) = 1, \tag{4.2.8}$$

$$\left(\frac{\partial}{\partial t} y(r, t)\right)_{r=a_0} = 0, \tag{4.2.9}$$

$$\lim_{r \to \infty} y(r, t) = 1, \tag{4.2.10}$$

and subsequently obtained $k(t)$ and $I(t)$ for each bulk quencher concentration C_Q^0. For the Stern-Volmer data presented in Figure 4, the best fit yielded the following parameters: $k_0 = 6.8 \text{ ns}^{-1}$, $\beta = 1.3 \text{ Å}^{-1}$, and $a_0 = 7.0 \text{ Å}$.

Molecular Dynamics Simulations. Molecular dynamics (MD) simulations on the 4CN-Trp-Trp peptide were carried out to characterize its conformational distribution in aqueous solution. First, the force field parameters for the 4-cyanoindole moiety were determined using the Force Field Toolkit Plugin (Mayne et al., 2013) v1.1 for VMD (Humphrey et al., 1996) v1.9.4 with initial charge, bond, and angle parameters from CHARMM36 (Best et al., 2012) and CGenFF v4.1 for Small Molecule Drug Design (Vanommeslaeghe et al., 2010). The peptide model was built using Vega ZZ (Pedretti et al., 2002) v3.1.1 with an initial structure of a fully extended conformation ($\phi = \varphi = 180°$) and then solvated using the Automatic PSF Generation Plugin v.1.3 in VMD. Simulations were carried out using the NAMD (Phillips et al., 2005) v.2.12 software package. Following a 1 ns equilibration run at 298 K and 1 atm in the NPT ensemble, subsequent production runs totaling 150 ns were performed on the equilibrated system at 298 K in the NVT ensemble. A trajectory was built by saving a snapshot every 500 fs, resulting in 300,000 total frames. Analysis of the MD trajectory was accomplished using VMD; for each frame, the minimum distance between any atoms in the 4-cyanoindole ring and the indole ring in the peptide was determined and used for constructing the distance distribution plot.

Trypsin Digest. The trypsin cleavage experiment was carried out at 37 °C, using 1 µM trypsin (Promega) to 15 µM peptide. Both samples were prepared in 50 mM tris buffer with 1 mM $CaCl_2$ at pH 7.8. Fluorescence spectra of the mixture (λ_{ex} = 330 nm) were measured at discrete reaction times, as indicated in the results section.

4.4. RESULTS AND DISCUSSION

FRET Property of the 4CN-Trp – Trp Pair. As indicated (**Figure 4.1**), the fluorescence spectrum of Trp overlaps with the absorption spectrum of 4CN-Trp, indicating that the fluorescent state of Trp (4CN-Trp) can be de-excited (excited) via the mechanism of FRET. Based on standard practice (see detail in SI), we determined the Förster distance (R_0) of this FRET pair to be 24.6 Å. For an ideal FRET pair, the excitation light should only excite the donor fluorescence. While this is not the case for the Trp and Trp-4CN pair, using an excitation wavelength of 270 nm nevertheless allows almost selective excitation of Trp (the donor) as its absorbance at this wavelength is about 4 times higher than that of 4CN-Trp (**Figure 4.2**). Similarly, for a PET pair, the excitation light should only populate the excited state(s) of the fluorophore, without exciting the quencher. As shown (**Figure 4.1**), selective excitation of 4CN-Trp, when Trp is present, can be achieved by using an excitation wavelength in the range of 310 – 360 nm, wherein the absorbance of Trp is negligible.

PET Property of the 4CN-Trp – Trp Pair. To show that Trp is the only amino acid that can significantly quench the fluorescence of 4CN-Trp, we studied the fluorescence property of a series of peptides with the sequence of 4CN-Trp-X (when X is a polar/charged residue) or 4CN-Trp-X-AAKKK (when X is a hydrophobic residue). In the latter case, the extra AAKKK segment is added to increase the peptide solubility. Specifically, the X residue represents one of the following amino acids: Gly, Met, His, Arg, Glu, Ser, Phe, Tyr, or Trp, and for simplicity, all these peptides are hereafter referred to as 4CN-Trp-X peptides, whose sequences are shown in **Table 4.1**. As shown (**Figure 4.3**), the fluorescence spectra of these peptides obtained under identical experimental conditions (e.g., concentration, temperature, and excitation wavelength) indeed

95

indicate that only Trp, when in close proximity, can efficiently quench the fluorescence of 4CN-Trp. This conclusion is further corroborated by fluorescence lifetimes measurements (**Figure 4.4** and **Figure 4.5**), which showed that the fluorescence decay of 4CN-Trp-Gly can be described by a single time constant of ca. 12.8 ns, close to the fluorescence lifetime of the free fluorophore (Hilaire, 2017), whereas that of 4CN-Trp-Trp not only is significantly faster but also requires a bi-exponential function to fit (**Table 4.2**). This deviation from single-exponential behavior is indicative of the existence of (at least) two peptide conformational ensembles that have distinctively different separation distances between the sidechains of Trp and 4CN-Trp (see below for further discussion). Additionally, since Trp has a negligible absorbance in the emission wavelength range of 4CN-Trp, this quenching effect cannot be attributed to FRET.

Following common practice, we carried out further static fluorescence quenching experiments using a compound that contains the 4-cyanoindole fluorophore of 4CN-Trp, 4-cyanoindole-3-acetic acid (4CNI-3AA), and N-acetyl-L-tryptophanamide (NATA), in an attempt to determine the underlying quenching mechanism. As shown (**Figure 4.6**), the resultant Stern-Volmer plot exhibits an upward curvature, indicative of a distance-dependent quenching process. A quantitative fitting of the Stern-Volmer curve using the diffusion model described in the literature (Lakowicz et al., 1994, Lakowicz 1999) and the numerical method detailed in the experimental section indicates that the following electron transfer rate equation (Lakowicz, 1999) can adequately describe the fluorescence quenching rate constant of NATA toward 4CN-Trp:

$$k_{\mathrm{ET}}(r) = k_0 e^{-\beta(r-a_0)} \tag{4.4.1}$$

with k_0 = 6.8 ns^{-1}, β = 1.3 Å^{-1}, and a_0 = 7.0 Å. Therefore, the Stern-Volmer quenching results support the notion that Trp quenches 4CN-Trp fluorescence via the mechanism of PET.

Eq. (4.4.1) provides a direct means to determine the fluorophore-quencher separation distance (r) if the fluorescence decay lifetime of the fluorophore is known. For example, for the 4CN-Trp-Trp peptide, which exhibits bi-exponential fluorescence decay kinetics with time

constants of 4.1 ns and 0.3 ns, respectively, the corresponding separation distances were calculated to be 6.9 and 4.6 Å (**Table 4.2**). To further substantiate this conclusion, we carried out molecular dynamics (MD) simulations to assess the conformational distribution of 4CN-Trp-Trp (see details in experimental section). As shown (**Figures 4.7, 4.8, 4.9**), the calculated conformational distribution as a function of *r*, which corresponds to the edge-to-edge distance between the closest atoms on the sidechains of 4CN-Trp and Trp, consists of two peaks, at 6.3 (32%) and 2.9 Å (68%), respectively. These values are in agreement with those calculated via Eq. (4.4.1) and hence support its applicability.

Eq. (4.4.1) indicates that efficient fluorescence quenching via PET can occur only when the fluorophore and quencher are at or near van der Waals contact (Lakowicz, 1999). Therefore, to check whether the 4CN-Trp and Trp pair exhibits such characteristics, we studied another two peptides where the fluorophore and quencher are separated by either one or two proline residues, serving as a rigid spacer. As shown (**Figure 4.10**), in comparison to that of 4CN-Trp-Trp, the fluorescence intensities of these proline-containing peptides are increased, with that of 4CN-Trp-Pro-Pro-Trp reaching to ca. 70% of the intensity of 4CN-Trp-Gly. The fluorescence decay kinetics (**Table 4.2** and **Figure 4.11**) are consistent with these results, which, taken together, confirm that efficient quenching of 4CN-Trp fluorescence by Trp occurs only when these two amino acids are close to each other.

Biological Applications. We believe that amino acid-based PET pairs, such as Trp and 4CN-Trp, will find various useful applications, especially in cases where using fluorescent dyes is prohibitive or undesirable. Below we describe three examples, demonstrating the biological utility of the Trp and 4CN-Trp pair. First, we employed it to probe the loop formation rate of a short peptide. Recently, Jacob *et al.* (Jacob et al., 2018) applied multiple techniques (e.g., FRET and PET) to assess the loop formation rate (k_L) of a series of flexible and unstructured L_1-$(GS)_n$-L_2 peptides, where L_1 and L_2 represent two different probes and n varies from 0 to 10. Their findings indicate that k_L depends not only on n but also on the identity of L_1 and L_2, highlighting the

importance of using non-interacting or weakly interacting probes in this type of experiments. In addition, their study provides valuable experimental data that can be used as a reference to validate the applicability of new PET and FRET pairs. Therefore, to make a direct comparison, we studied the 4CN-Trp-(GS)$_4$-Trp peptide, which is an analog of the Trp-(GS)$_4$-MR121 peptide of Jacob *et al.* (Jacob et al., 2018). As shown (**Figure 4.12**), compared to the fluorescence decay kinetics of a reference peptide that lacks the Trp quencher (sequence: 4CN-Trp-(GS)$_4$-Phe), the fluorescence decay of 4CN-Trp-(GS)$_4$-Trp is much faster, indicating that this peptide indeed samples an ensemble of dynamic conformations that can bring the terminal 4CN-Trp and Trp residues in close proximity for PET to occur. Specifically, the fluorescence lifetime of 4CN-Trp-(GS)$_4$-Trp is 5.6 ns, comparing to the 12.8 ns of 4CN-Trp-(GS)$_4$-Phe. Assuming that this change in the fluorescence lifetime of 4CN-Trp is only due to the addition of the PET quenching channel, the effective quenching rate constant (k_Q) is calculated to be 9.9×10^7 s^{-1}. As described by Jacob *et al.* (Jacob et al., 2018), k_Q is proportional to k_L although the exact proportionality factor depends on the model used. Interestingly, the k_Q value determined by Jacob *et al.* (Jacob et al., 2018) for Trp-(GS)$_4$-MR121 is faster, ca. 1.7×10^8 s^{-1}. We believe that this difference reflects the difference between 4CN-Trp and MR121, as the latter is larger and more hydrophobic. This notion is supported by previous studies showing that MR121 can preferentially interact with Trp and hence increases the loop formation rate (Fierz et al., 2007; Neuweiler et al., 2007). While 4CN-Trp and Trp will also likely show specific interactions, such as aromatic stacking, our result suggests that they are less perturbative to the native structure and dynamics of peptides, in comparison to commonly used hydrophobic dyes.

In the second case, we demonstrate the utility of the Trp and 4CN-Trp pair in the context of biomolecular interactions using bovine serum albumin (BSA) as a testbed. It is well known that BSA can interact with many small molecules (Peters, 1985; Sudlow, 1975) through one of its two major binding sites (i.e., Sudlow's site I and II). BSA contains two Trp residues, located at the 134 and 212 positions. However, based on its crystal structure (Majorek et al., 2012), Trp134 is relatively far away from both binding sites. Therefore, Trp212, which is in the binding pocket of

Sudlow's site I, has been commonly used as a fluorescence reporter to assess small molecule binding.[23] Similarly, this difference between sites I and II can be capitalized to determine the model of interaction of ligands that contain a 4CN-indole moiety (i.e., the fluorophore of 4CN-Trp) via fluorescence measurement. This is because only binding to site I can result in a significant quenching of the 4CN-indole fluorescence, due to PET. To test this notion, we studied the binding of 4CNI-3AA to BSA. We chose this molecule because it is a derivative of a known BSA-binding ligand, indole-3-acetic acid (IAA) (Bertuzzi et al., 1997). As shown (**Figure 4.13a**), in the presence of BSA the fluorescence spectrum of 4CNI-3AA (λ_{ex} = 330 nm) does not show a significant decrease in intensity; instead, it is blue-shifted from that of the free 4CNI-3AA. Therefore, these data indicate that 4CNI-3AA is bound to BSA, but not within the binding pocket of Sudlow's site I. This result is consistent with a previous study showing that IAA is bound to Sudlow's site II of BSA (Bertuzzi et al., 1997).

The fluorescence spectra obtained with λ_{ex} = 270 nm provide further evidence supporting the abovementioned binding interaction between 4CNI-3AA and BSA. As indicated (**Figure 4.13b**), in the presence of 4CNI-3AA the intrinsic fluorescence intensity of BSA (arising mostly from Trp and Tyr residues) is significantly decreased, accompanied by a significant increase in the fluorescence intensity of 4CNI-3AA. Together, these results indicate that the fluorescence of 4CNI-3AA is enhanced through a FRET mechanism, with Trp and/or Tyr residues being the FRET donors. Indeed, Trp212 is located at a position that is ca. 20 Å away from the center of Sudlow's site II, which is well within the Förster distance (R_0 = 24.6 Å) for efficient FRET to occur. Additionally, two Tyr residues, which are 7.8 Å and 11.5 Å away from the center of this binding site, could also contribute to the fluorescence enhancement of the reporter via FRET.

Finally, we show that 4CN-Trp and Trp can be used to assess protease activity. Proteases are responsible for activating proenzymes and degrading proteins amongst other functions (López-Otín et al., 2008), hence playing a key role in many biological pathways. As such, various fluorogenic assays have been developed to detect their proteolytic activities (Ong et

al., 2017). For *in vitro* studies, the available assays generally involve the use of short peptides labeled with two dyes, making them quite laborious to produce. Therefore, synthetic peptides containing a pair of 4CN-Trp and Trp would enjoy an advantage in this regard. In a proof-of-principle study, we designed a peptide (sequence: 4CN-Trp-Lys-Trp-Ala-Gly-Lys) to measure the proteolytic activity of trypsin via fluorescence spectroscopy. Trypsin is a serine protease found in the digestive system of many organisms and used to break down proteins by hydrolyzing the amide bond at the C-terminal side of a Lys or Arg residue (not followed by a proline). The design of this peptide is based on the idea that the fluorescence intensity of 4CN-Trp is low in the intact peptide, due to Trp quenching, whereas after peptide cleavage by trypsin, which leads to the formation of 4CN-Trp-Lys and Trp-Ala-Gly-Lys fragments, the 4CN-Trp fluorescence intensity will significantly increase due to the removal of the PET quenching pathway. The experimental results obtained in the presence and absence of trypsin indeed show that upon peptide cleavage the fluorescence intensity of 4CN-Trp increases more than 6 times (**Figure 4.14**), hence validating the utility of this amino acid PET pair in the study of protease function.

4.5: CONCLUSIONS

In summary, we demonstrate that Trp can serve as either a FRET donor or a PET quencher to 4CN-Trp, a newly found blue fluorescent unnatural amino acid with a large fluorescence quantum yield (>0.8). The following considerations motivate this work: (1) spectroscopic techniques based on FRET and PET are widely used to assess protein conformations, conformational dynamics, and interactions; (2) evidence from several recent studies (Riback et al., 2018a, 2018b) indicate that using dye-based fluorophores in FRET and PET applications can yield skewed results, due to specific or preferred probe-probe interactions. Therefore, development of amino acid-based FRET and PET pairs that are intrinsically less-perturbative and hence can minimize such pitfalls is needed; (3) the absorption spectrum of 4CN-Trp is significantly red-shifted from that of Trp, allowing selective excitation of its fluorescence (e.g., using λ_{ex} = 330–360 nm) or that of Trp (e.g., using λ_{ex} = 270 nm) in PET or FRET

applications; (4) 4CN-Trp is only one atom larger than Trp, making it less perturbative to proteins than fluorescent dyes; (5) 4CN-Trp can now be conveniently synthesized via chemical (Zhang et al., 2019) and biological (Boville et al., 2018) means; and (6) it is possible to incorporate 4CN-Trp into proteins genetically via amber codon suppression or chemically using a post-translational modification method (Ahmed and Gai, 2017; Wright et al., 2016). Given the unique photophysical property (Ahmed et al., 2019; Hilaire, 2017a, 2017b) of 4CN-Trp and the less-perturbating nature of this unnatural amino acid and Trp, we believe that this pair will find valuable applications in biochemistry and biophysics, especially in cases where using dye-based PET or FRET reporters is deemed inappropriate. We are particularly excited about the possibility of using this pair of amino acids in single-molecule fluorescence studies, such as PET-FCS (Doose et al., 2009 et al.; Li et al., 2006; Lin et al., 2017), as well as using it to study the functional and/or conformational dynamics of Trp-rich proteins (Liu et al., 2014).

Table 4.1: Fluorescence lifetime (τ) and relative amplitude (A) determined from fitting the fluorescence decay of each peptide in Figure 4.5 to a single or bi-exponential function.

Peptide Name	Peptide Sequence	τ (ns)	A (%)
4CN-Trp-Gly	4CN-Trp-G	12.8	100
4CN-Trp-His	4CN-Trp-H	12.5	100
4CN-Trp-Arg	4CN-Trp-R	12.5	100
4CN-Trp-Ser	4CN-Trp-S	12.6	100
4CN-Trp-Glu	4CN-Trp-E	12.6	100
4CN-Trp-Met	4CN-Trp-M	12.2	100
4CN-Trp-Phe	4CN-Trp-FAAKKK	11.5	100
4CN-Trp-Tyr	4CN-Trp-YAAKKK	10.5	100
4CN-Trp-Trp	4CN-Trp-WAAKKK	4.1	64
		0.3	36

Table 4.2. Fluorescence lifetime (τ) and relative amplitude (A%) determined from fitting the fluorescence decay kinetics of each peptide to a single or bi-exponential function. For each lifetime, the corresponding fluorophore-quencher separation distance (r) was calculated using Eq. (4.4.1).

	τ_1 (ns)	A_1	r_1 (Å)	τ_2 (ns)	A_2	r_2 (Å)
4CN-Trp-Gly	12.8	100	---	---	---	---
4CN-Trp-Trp	4.1	64	6.9	0.3	36	4.6
4CN-Trp-Pro-Trp	6.4	62	7.5	1.2	38	5.7
4CN-Trp-Pro-Pro-Trp	8.2	100	8.0	---	---	---
4CN-Trp-(GS)$_4$-Phe	12.8	100	---	---	---	---
4CN-Trp-(GS)$_4$-Trp	5.7	100	7.3	---	---	---
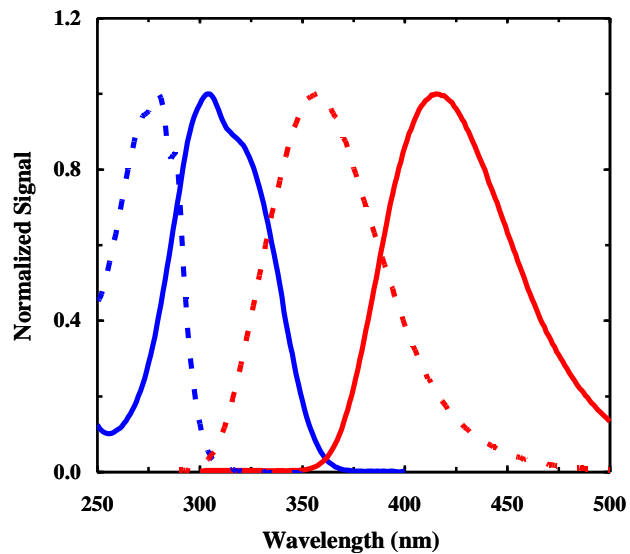
Figure 4.1: Normalized absorption (blue) and fluorescence spectra (red) of Gly-Trp-Gly (dashed line) and 4CN-Trp-Gly (solid line) peptides in water. The excitation wavelength for the fluorescence measurements was 270 nm.
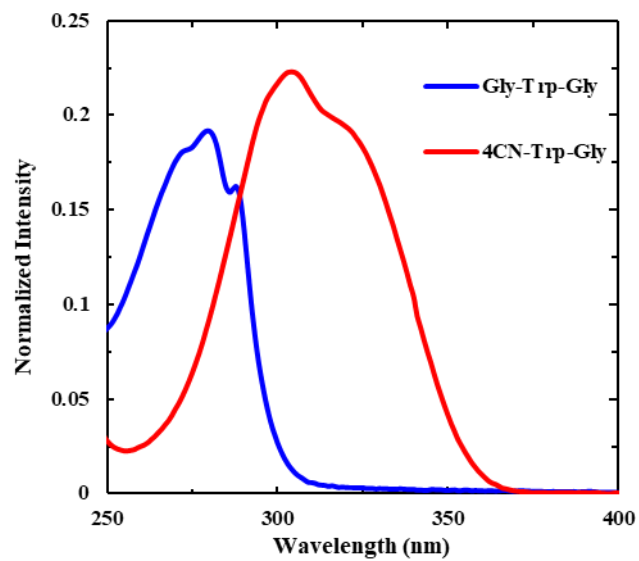
Figure 4.2: Absorption spectra of Gly-Trp-Gly (33 µM) and 4CN-Trp-Gly (33 µM) in water, as indicated.
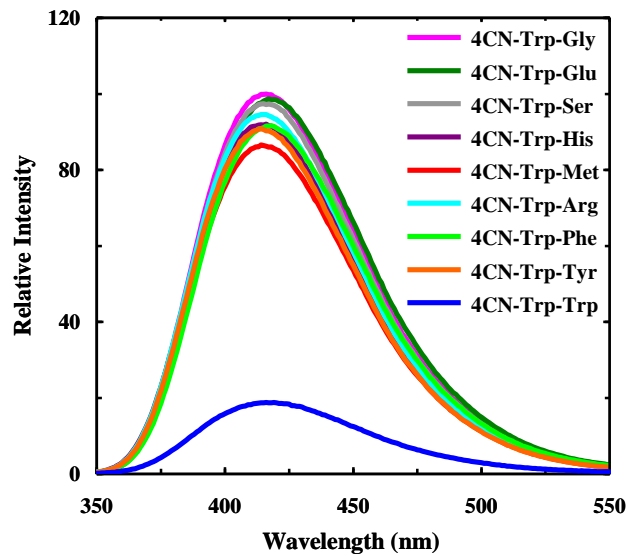
Figure 4.3: Fluorescence spectra of 4CN-Trp-X peptides in water (20 µM), as indicated. These spectra were collected under identical experimental conditions with a λ_{ex} = 330 nm and, for easy comparison, their intensities have been scaled relative to that of 4CN-Trp-Gly.
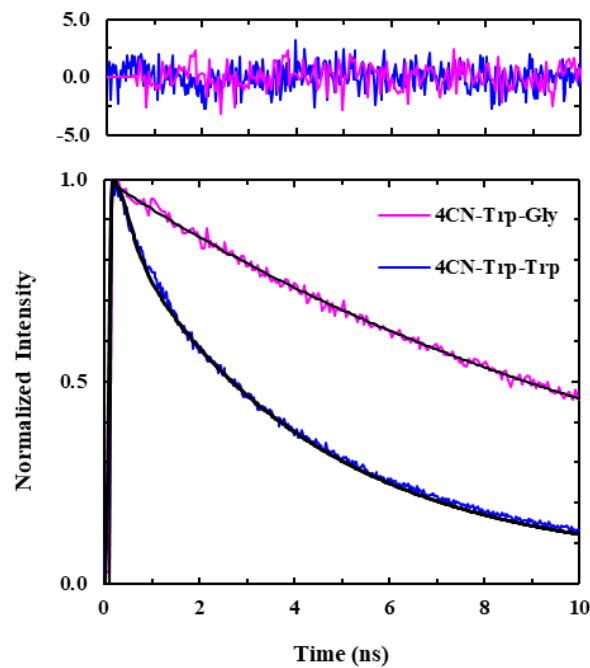
Figure 4.4: Comparison between the fluorescence decay kinetics of 4CN-Trp-Gly and 4CN-Trp-Trp, as indicated. For 4CN-Trp-Gly, the decay curve can be fit by a single-exponential function, whereas that of 4CN-Trp-Trp requires a bi-exponential function to fit. The corresponding fitting parameters are listed n Table 4.2. Shown in the top panel are the residuals of the respective fits (smooth lines).
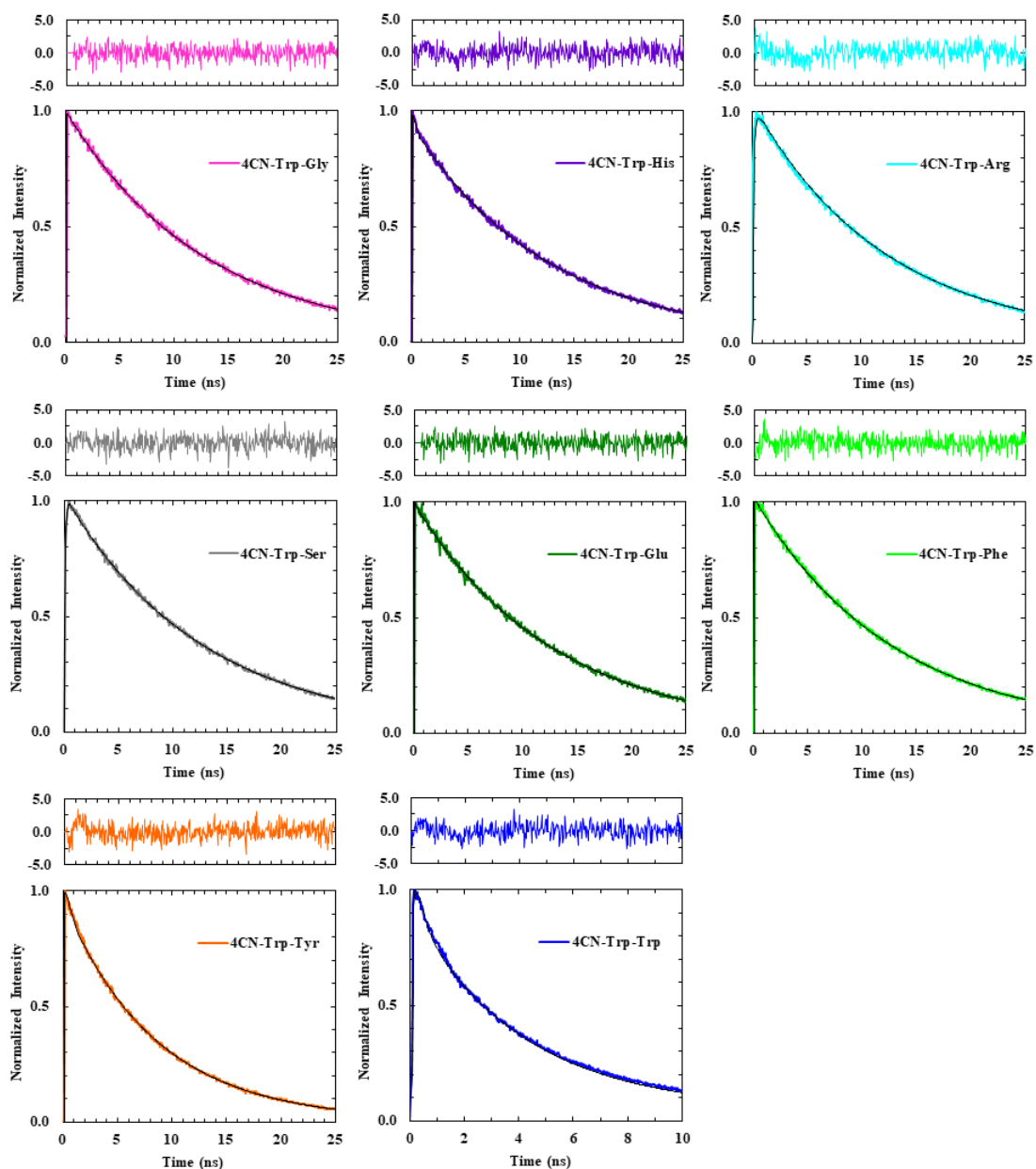
Figure 4.5: Fluorescence decay kinetics of 4CN-Trp-X peptides, as indicated. In each case, the smooth black line corresponds to the best fit of the respective data to a single-exponential or a bi-exponential function and the resultant lifetime(s) is given in Table 4.1. Residuals of the fits are on top of each respective panel.
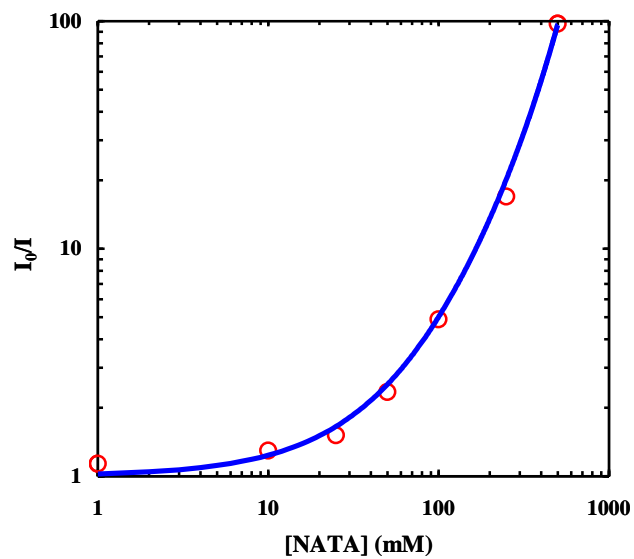
Figure 4.6: Relative fluorescence intensity of 4CNI-3AA (10 µM) as a function of the concentration of the quencher, NATA. These data were obtained with a λex = 330 nm and under the same fluorescence measurement conditions. Fitting these data to the model described in the text yielded the smooth line and the following parameters: k_0 = 6.8 ns^{-1}, β = 1.3 Å^{-1}, and a_0 = 7.0 Å.
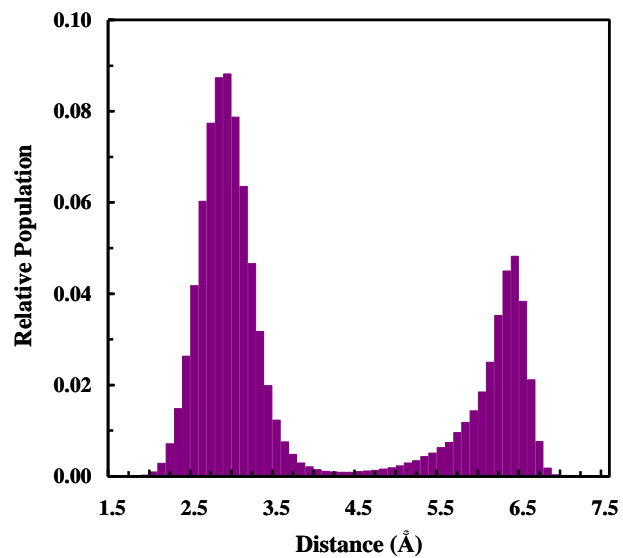
Figure 4.7: Distribution of the separation distance between 4CN-Trp and Trp residues in a 4CN-Trp-Trp peptide from MD simulations.
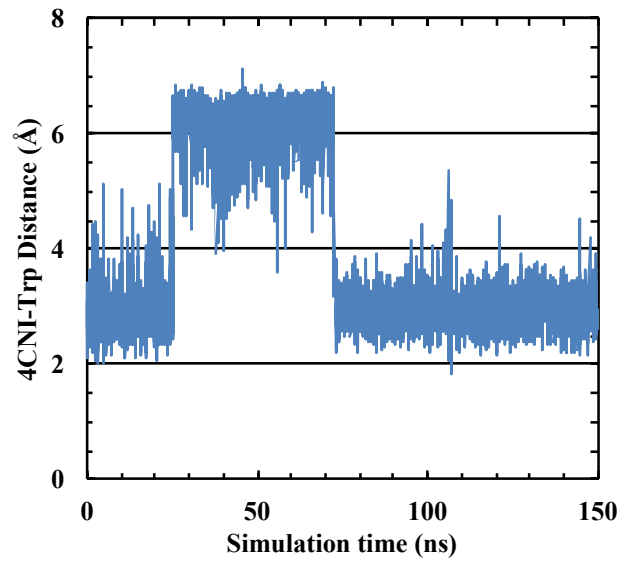
Figure 4.8: The trajectory of the separation distance between the two fluorophores in the 4CN-Trp-Trp peptide obtained from the MD simulations.
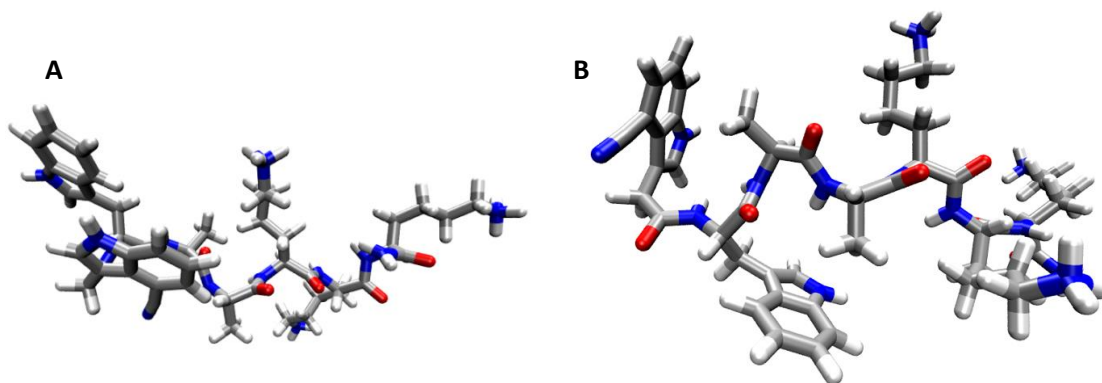
Figure 4.9: Representative structures of the two conformational populations of the 4CN-Trp-Trp peptide. The distances between the 4-cyanoindole and indole moieties are 2.96 Å (A) and 6.62 Å (B), respectively.
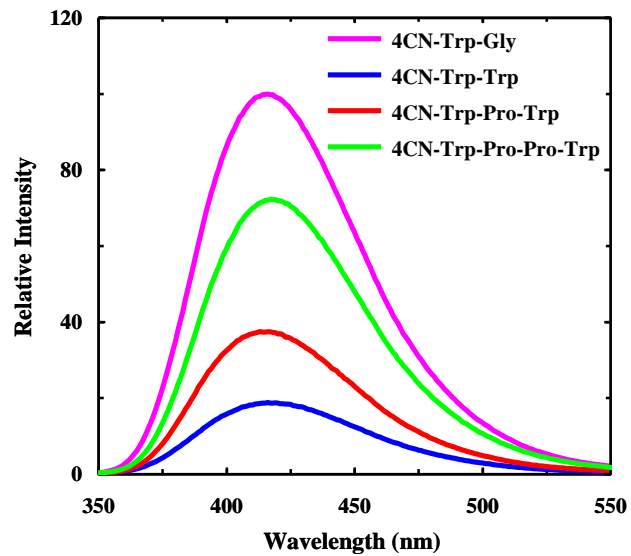
Figure 4.10: Fluorescence spectra of different 4CN-Trp-containing peptides in water (20 μM), as indicated. These spectra were collected under identical experimental conditions with a λ_{ex} = 330 nm and, for easy comparison, their intensities are scaled relative to that of 4CN-Trp-Gly.
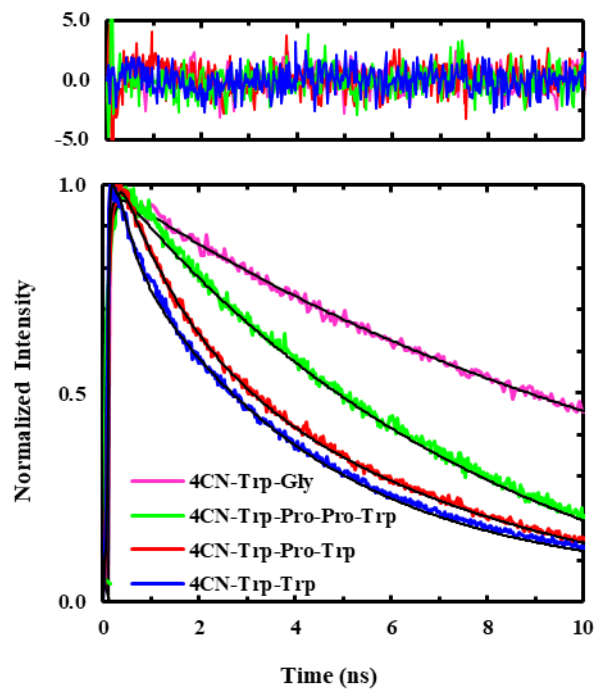
Figure 4.11: Fluorescence decay kinetics of 4CN-Trp-Gly, 4CN-Trp-Pro-Pro-Trp, 4CN-Trp-Pro-Trp, and 4CN-Trp-Trp, as indicated. The smooth line in each case correspond to the best fit of the data to either a single-exponential function (4CN-Trp-Gly and 4CN-Trp-Pro-Pro-Trp) or a bi-exponential function (4CN-Trp-Pro-Trp and 4CN-Trp-Trp) and the resultant lifetime(s) are listed in Table 4.2.
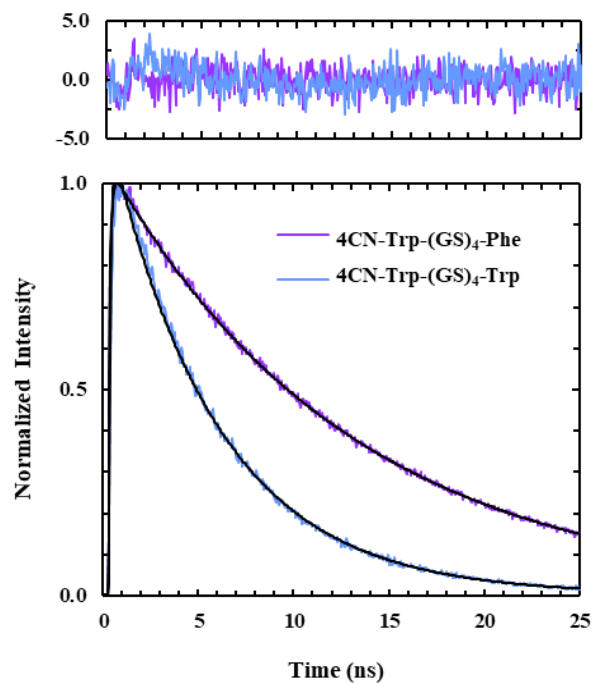
Figure 4.12: Fluorescence decay kinetics of 4CN-Trp-(GS)$_4$-Phe and 4CN-Trp-(GS)$_4$-Trp, as indicated. Fitting these kinetic traces to a single-exponential function yielded a decay time constant of 12.8 ns for 4CN-Trp-(GS)$_4$-Phe and 5.6 ns for 4CN-Trp-(GS)$_4$-Trp. Shown in the top panel are the corresponding residuals.
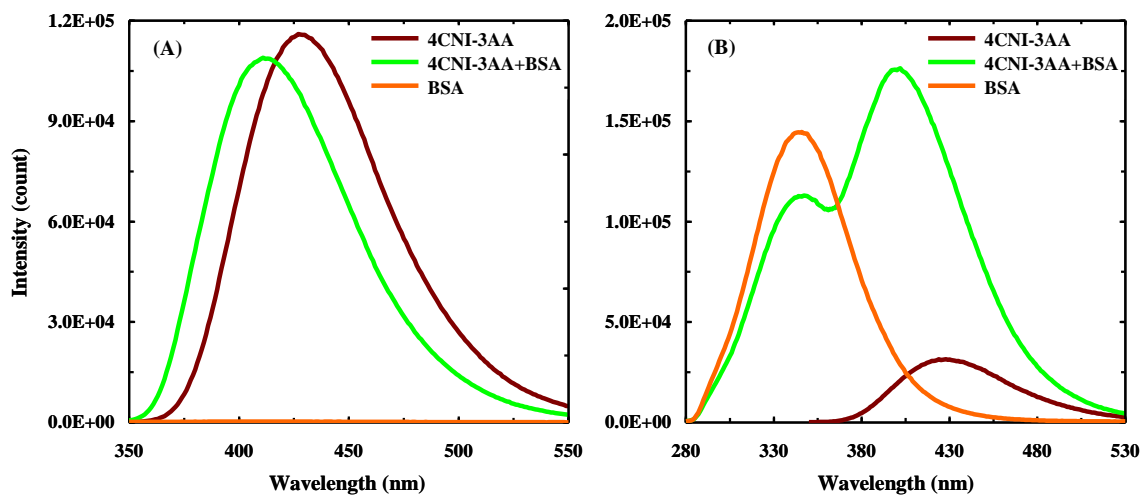
Figure 4.13: Fluorescence spectra of 4CNI-3AA (1 µM), BSA (100 µM), and the mixture of 4CNI-3AA (1 µM) and BSA (100 µM) obtained with an excitation wavelength of 330 nm (A) and 270 nm (B).
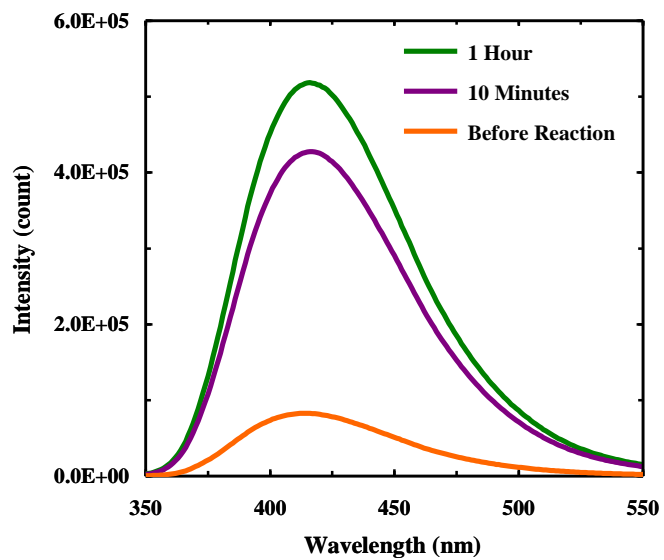
Figure 4.14: Fluorescence spectra of 4CN-Trp-Lys-Trp-Ala-Gly-Lys (15 μM) obtained before, 10 minutes, and 1 hour after the addition of 1 μM of trypsin (in 50 mM tris buffer, pH 7.8), as indicated. The excitation wavelength was 330 nm.

5: 4-Cyanoindole-2′-deoxyribonucleoside As a Dual Fluorescence and Infrared Probe of DNA Structure and Dynamics

5.1: ABSTRACT

Unnatural nucleosides possessing unique spectroscopic properties that mimic natural nucleobases in both size and chemical structure are ideally suited for spectroscopic measurements of DNA/RNA structure and dynamics in a site-specific manner. However, such unnatural nucleosides are scarce, which prompts us to explore the utility of a recently found unnatural nucleoside, 4-cyanoindole-2′-deoxyribonucleoside (4CNI-NS), as a site-specific spectroscopic probe of DNA. A recent study revealed that 4CNI-NS (Passow and Harki, 2018) is a universal nucleobase that maintains the high fluorescence quantum yield of 4-cyanoindole (Hilaire et al., 2017b) and that among the four natural nucleobases, only guanine can significantly quench its fluorescence. Herein, we further show that the C≡N stretching frequency of 4CNI-NS is sensitive to the local environment, making it a useful site-specific infrared probe of oligonucleotides. In addition, we demonstrate that the fluorescence-quencher pair formed by 4CNI-NS and guanine can be used to quantitatively assess the binding affinity of a single-stranded DNA to the protein system of interest via fluorescence spectroscopy, among other applications. We believe that this fluorescence binding assay is especially useful as its potentiality allows high-throughput screening of DNA–protein interactions.

5.2: INTRODUCTION

Bio-macromolecules, including proteins and polynucleotides such as DNA and RNA, can adopt various complex structures that confer specific functions. Therefore, understanding the structure-function relationship of such bio-macromolecules has been the subject of numerous studies, many of which use infrared- and/or fluorescence-based spectroscopic techniques due to

their sensitivity and convenience. However, intrinsic infrared (IR) and fluorescence signals of bio-macromolecules, while very useful, often provide only limited information about the system in question. Therefore, the past two decades have seen significant efforts in the development of extrinsic IR and fluorescence probes that can be used to site-specifically interrogate the structure, dynamics, and function of bio-macromolecules (Ma et al., 2015; Serrano et al., 2012a). However, introducing an exogenous component into any bio-macromolecule will unavoidably perturb its structure (and consequently other properties). Therefore, many past studies have focused on identifying molecular moieties that not only afford the required spectroscopic attributes and can be easily incorporated into the bio-macromolecules of interest, but also minimally perturb their native structure. In this regard, the ideal extrinsic spectroscopic probes for protein (DNA) would be simple analogs of naturally-occurring amino acids (nucleosides). Indeed, a large body of work has been devoted to this idea, especially in the area of protein science. As a result, a large pool of unnatural amino acid-based (UAA-based) IR and fluorescence probes have been identified and utilized to study a wide range of problems in protein biochemistry and biophysics (Adhikary et al., 2017; Błasiak et al., 2017; Boxer, 2009; Ghosh et al., 2017; Haney et al., 2015; Hanoian et al., 2015). In comparison, however, little has been done in the development of minimally-perturbing, unnatural nucleoside-based spectroscopic probes. This is due, at least in part, to the fact that the fundamental building blocks of DNA involve only four nucleobases in comparison to the 20 canonical amino acids for protein, and specific base pairing in double-stranded DNA is achieved via well-defined hydrogen-bonding interactions. Herein, we aim to show that an unnatural nucleoside (UNS), 4-cyanoindole-2′-deoxyribonucleoside (4CNI-NS) (**Figure 5.1**), can be used as a site-specific IR and fluorescence probe of DNA structure and dynamics.

The indole ring serves as a structural scaffold to the natural nucleobases adenine and guanine. Therefore, it is possible to use indole-based UNSs as universal DNA bases. Indeed, 5-nitroindole-2′-deoxyribonucleoside (5NI-NS) (**Figure 5.1**) has long been used as an unnatural universal DNA base, as it can pair with all four natural DNA bases (Loakes and Brown, 1994). More recently, Passow and Harki (Passow and Harki, 2018) have shown that 4CNI-NS can also

be used as a universal DNA base. As demonstrated in previous studies (Hilaire et al., 2017b, 2017a; van Wilderen et al., 2018), in comparison to indole and other indole derivatives, 4-cyanoindole (4CNI) (**Figure 5.1**) is significantly more fluorescent, has a long fluorescence lifetime, a red-shifted absorption spectrum, and an emission spectrum in the blue region of the visible spectrum. These qualities give 4CNI-NS a distinct advantage over 5NI-NS as it can serve as a fluorescence reporter. Moreover, the study of Passow and Harki (Passow and Harki, 2018) has shown that guanine can effectively quench the fluorescence of 4CNI-NS when in close proximity. This suggests that 4CNI-NS and guanine constitute a useful fluorophore-quencher pair for investigating various DNA-related questions via fluorescence spectroscopy, such as DNA-protein binding interaction, DNA–DNA association and conformational distribution of single-stranded DNA (ssDNA) in solution. To validate this notion, herein we carry out a proof-of-principle experiment in which we use fluorescence intensity measurements to determine the binding constant of a 4CNI-NS-containing ssDNA to a transport protein, bovine serum albumin (BSA).

While the IR spectrum of a DNA/RNA molecule manifests its structure and dynamics, it is hardly interpretable in a site-specific manner, due to spectral overlapping and degeneracy. Therefore, in order to achieve site-specificity in IR measurement of DNA/RNA, an appropriate UNS-based vibrational probe is needed. To the best of our knowledge, there is only one such UNS, i.e., a uridine derivative (Schmitz et al., 2016), that can be used in this purpose, hence limiting the use of IR spectroscopy to gain site-specific structural and dynamical information of DNA. To overcome this limitation, we first examine the utility of the universal base, 5NI-NS, as a site-specific IR probe. We find that while the nitro group of 5NI-NS gives rise to an intense band at ca. 1530 cm^{-1}, its frequency is insensitive to a solvent, thus making it less useful as a vibrational probe. Several studies have shown that the C≡N stretching vibration of an alkyl or aryl nitrile is not only located at an uncongested region of the IR spectrum of bio-macromolecules, but is also sensitive to the local environment (Bagchi et al., 2012; Fafarman et al., 2006; Getahun et al., 2003; Waegele et al., 2009). Hence, several nitrile-containing UAAs have been employed to site-specifically interrogate various properties of the protein system in question via linear and/or

nonlinear IR methods (Ding et al., 2016; Ma et al., 2015). These previous studies lead us to believe that the C≡N stretching vibration of 4CNI-NS can also be used as a site-specific IR probe of DNA structure and dynamics. To corroborate this notion, we examine the dependence of the C≡N stretching vibrational frequency (vCN) of 4CNI-NS on a solvent. Our results show that vCN is sensitive to a solvent, similar to that observed for 4-cyanotryptohan (4CN-Trp) (van Wilderen et al., 2018). Therefore, this finding supports the idea that 4CNI-NS can be utilized to provide information about the local hydration and electrostatic environment of DNA in a site-specific manner.

5.3: EXPERIMENTAL SECTION

Sample Preparation. Synthesis of 5′-O-dimethoxytrityl protected 4CNI-NS phosphoramidite building block was prepared according to a known literature method (Passow and Harki, 2018). Oligo1 (5′-ACTTGGCC(4CNI-NS)NCCAATTTTG) was synthesized on an Applied Biosystems Expedite 8909 DNA synthesizer (Carlsbad, CA, USA) using a standard phosphoramidite method on 1 µmole scale as previously described (Weber et al., 2014). A mild cleavage condition (saturated ammonium hydroxide solution overnight at room temperature) was employed to minimize the hydrolysis of the nitrile group. Oligo1* (5′-ACTTGGCCGNCCAATTTTG) was purchased from Integrated DNA Technologies (Coralville, IA, USA).

Oligo2 (5′-4CNI-ACTTAACCACCATTTTT) and Oligo3 (5′-4CNI-ACTTAACCGCCA TTTTT) were generated by coupling 4-cyanoindole-3-acetic acid (4CNI-3AA) to 5′ amino modified oligonucleotides with 6-carbon linker which were purchased from Integrated DNA Technologies. Specifically, a solution of 4CNI-3AA (33 mg, 0.16 mmol) and sulfo-NHS (N-hydroxysuccinimide) (93 mg, 0.21 mmol) in anhydrous DMF (1 mL) was added a solution of DCC (44 mg, 0.21 mmol) in DMF (1 mL) at 0 °C. The mixture was stirred for 2 h under N_2 and poured into a solution of oligonucleotide (3.7 µmol) in phosphate buffer (0.1 M, pH 7.2, 20 mL) and stirred overnight. All oligonucleotides were purified by reversed-phase high-performance liquid chromatography using Agilent 1100 HPLC system (Santa Clara, CA, USA) to achieve >95% purity. Mass spectrometry

on a Shimadzu AXIMA Performance MALDI-TOF (Kyoto, Japan) was used to confirm product mass and purity. The concentrations of oligonucleotides and BSA (Sigma, St. Louis. MO, USA) were determined by UV/VIS (JASCO, Easton, MD, USA) using ε values of 7,790 M^{-1} cm^{-1} at 305 nm [7] and 43,824 M^{-1} cm^{-1} at 280 nm, respectively.

Fluorescence Measurement. All fluorescence spectra were collected on a Jobin Yvon Horiba Fluorolog 3.10 spectrofluorometer (Kyoto, Japan) at room temperature in a 1 cm quartz cuvette with a 1.0 nm resolution, 1 nm excitation/emission slit, an integration time of 1.0 nm/s, and an excitation wavelength of either 320 or 325 nm. All 4CNI-based samples were prepared by directly dissolving lyophilized solids in pure water (or THF), and the final concentration was 5.0 μM, except for that used in the BSA binding study, which was 1.0 μM. Fluorescence QY was determined using the following equation (Lakowicz, 1999):

$$QY_S = QY_R \frac{I_S A_R}{I_R A_S},$$ (5.3.1)

where I is the integrated fluorescence intensity, A is the optical density of the fluorophore at λ_{ex} (325 nm), and the subscripts S and R represent the sample and reference, respectively. In the current study, 4CNI was used as the reference (QYR = 0.78 in water) [12].

Isothermal Titration Calorimetry Measurement. ITC experiments were carried out on a MicroCal iTC200 (Malvern, United Kingdom), using the following instrument settings: 20 injections, initial delay 60 sec., spacing 180 sec., filter period 5 sec., injection volume 2 μL, measurement temperature of 25 °C, reference power of 6 μcal s^{-1}, and stirring speed of 1000 r.p.m. The BSA concentration in the syringe was 300 μM and the Oligo1* concentration in the cell was 30 μM.

FTIR Measurement. FTIR spectra were collected on a Nicolet Magna-IR 860 spectrometer (ThermoFisher Scientific, Waltham, MA, USA) using a home-made sample holder composed of two CaF2 windows and a 50 μm spacer. All samples were prepared by directly dissolving lyophilized solids in the specified solvents with a final concentration of ca. 10 mM.

5.4: RESULTS AND DISCUSSION

Fluorescence Study. The absorption spectrum of 4CNI-NS extends beyond 310 nm (**Figure 5.2**), indicating that its florescence can be selectively excited in the presence of aromatic amino acids. This feature is especially convenient for studies involving proteins, such as DNA–protein interactions. As shown (**Figure 5.2**), the fluorescence spectrum of 4CNI-NS peaks at ca. 412 nm in water, exhibiting a relatively large Stokes shift. Furthermore, the fluorescence quantum yield (QY) of 4CNI-NS, determined using 4CNI as a reference, is 0.85 ± 0.5 in water, which is in agreement with that measured by Passow and Harki (Passow and Harki, 2018). Since the fluorescence QY of 4CN-Trp is ca. 0.8 in water, this result indicates that covalently connecting a molecular group at either the 1 or 3 positions of the 4CNI ring does not significantly change its fluorescence property. In addition, similar to that observed for the 4CN-Trp (Hilaire et al., 2017b), in the hydrophobic solvent tetrahydrofuran (THF), the fluorescence QY of 4CNI-NS is decreased to 0.72, while the peak wavelength (λ_{em}) of its fluorescence spectrum is blue-shifted to 380 nm. This blue-shift, similar to that observed for indole, is a manifestation of the less stabilizing effect of a less polar solvent on the fluorophore's excited state, whose permanent dipole moment is different from the corresponding ground-state value (Lakowicz, 1999). These changes suggest that both λ_{em} and QY of 4CNI-NS fluorescence can potentially be useful as indictors of its local environment. However, given the fact that other nucleosides can quench the fluorescence of 4CNI-NS (see below), only λ_{em} is practically useful in this regard.

Interestingly, the study of Passow and Harki (Passow and Harki, 2018) demonstrated that the fluorescence of 4CNI-NS could be quenched by a guanine base. It has been shown that guanine can quench the fluorescence of various fluorescent dyes via the mechanism of electron

123

transfer (ET) (Seidel et al., 1996; Heinlein et al., 2003). Based on those previous findings, we believe that the same quenching mechanism is also at play for 4CNI-NS. Since the ET transfer rate exhibits an exponential distance dependence (Lakowicz, 1999), efficient fluorescence quenching via an ET mechanism can only occur when the corresponding fluorophore and quencher are sufficiently close or in Van der Waals contact (Lakowicz, 1999; Seidel et al., 1996; Thomas Heinlein et al., 2003). This property thus makes 4CNI-NS and guanine a very useful fluorophore-quencher pair that can be used to study, for example, DNA-protein interactions. To demonstrate this utility, we employ it to determine the binding constant of a ssDNA to BSA via fluorescence spectroscopy. In practice, DNA–protein associations are typically detected by techniques based on ultra-centrifugation, isothermal titration calorimetry (ITC) or surface plasmon resonance (SPR) (Jing and Bowser, 2011), which are time consuming and relatively low throughput. Therefore, devising a fluorescence-based assay would be quite advantageous, as it will provide a more convenient and potentially high-throughput means to explore specific DNA–protein binding interactions.

BSA is an abundant carrier protein in blood that has been shown to promiscuously bind with DNA and RNA oligonucleotides amongst other molecules such as fatty acids, small molecules, drugs, and peptides (Peters, 1985). To use fluorescence spectroscopy to probe the binding interaction between a ssDNA and BSA, we synthesized the following 18-base oligonucleotide (Oligo1): 5′-ACTTGGCC(4CNI-NS)CCAATTTTG. This sequence is designed with the consideration that a ssDNA molecule bound to a protein often adopts a more extended conformation in comparison to its free form (Cattan et al., 1969). Therefore, the fluorescence QY of Oligo1 is expected to increase due to the increase in the (average) separation distance between the 4CNI-NS fluorophore and the guanine quenchers upon binding to BSA. As shown (**Figure 5.3a**), the fluorescence spectra obtained under different solution conditions meets our expectation: (1) the fluorescence intensity of free Oligo1 is significantly smaller than that of 4CNI-NS, confirming the quenching effect of guanine; and (2) in the presence of BSA, the fluorescence

spectrum of Oligo1 not only is blue-shifted (ca. 8 nm) but also exhibits a larger intensity, which, combined, indicates protein binding.

To further validate the exclusive quenching effect of guanine toward the fluorescence of the 4CNI fluorophore in 4CNI-NS, we compared the fluorescence QYs of two other oligonucleotides. As indicated by their sequences, 5′-4CNI-ACTTAACCACCATTTTT (Oligo2) and 5′-4CNI-ACTTAACCGCCATTTTT (Oligo3), while each oligonucleotide has a 4CNI fluorophore appended at the 5′ end, only Oligo3 contains a guanine base. As shown (**Figure 5.3b**), under identical experimental conditions (i.e., concentration, excitation wavelength, solvent, temperature, and absorbance at λ_{ex}) the fluorescence intensity of Oligo2 is larger than that of Oligo3. Consistent with this finding, a more quantitative assessment revealed that the fluorescence QYs of Oligo2 and Oligo3 are 0.74 and 0.31, respectively. Therefore, taken together, these results support our proposal that 4CNI-NS and guanine constitute a fluorophore-quencher pair useful for studying various questions in DNA science, similar to those used in protein science (Goldberg et al., 2012; Mintzer et al., 2015; Taskent-Sezgin et al., 2010).

Finally, to demonstrate the utility of the proposed method, we employed fluorescence spectroscopy to determine the binding constant of Oligo1 to BSA. Specifically, we collected the fluorescence spectra of a series of solutions consisting of 1.0 μM Oligo1 and various concentrations of BSA ([BSA]) using a λ_{ex} of 320 nm. As indicated (**Figure 5.4**), the fluorescence intensity of the 4CNI-NS fluorophore in Oligo1 increases with increasing [BSA] and levels off at ca. 1 μM of BSA, indicating that Oligo1 has a strong affinity toward BSA. To provide a more quantitative assessment of this fluorescence binding curve, we analyzed it using a simple thermodynamic model that assumes that each BSA molecule can provide *n* identical, non-interacting binding sites. In other words, the effective concentration of BSA is scaled by *n*. It can be easily shown that:

$$[D_{\mathrm{B}}] = \frac{([D]_0 + n[\mathrm{BSA}]_0 + K_{\mathrm{d}}) - \sqrt{([D]_0 + n[\mathrm{BSA}]_0 + K_{\mathrm{d}})^2 - 4n[D]_0[\mathrm{BSA}]_0}}{2} \; ,$$

(**5.4.1**)

125

where $[D]_0$ and $[BSA]_0$ are the initial or total concentrations of Oligo1 and BSA, respectively; $[D_B]$ is the concentration of BSA-bound Oligo1; and K_d is the dissociation constant. Since the fluorescence signal contains contributions from the free and BSA-bound Oligo1 molecules, the binding curve (i.e., I versus $[BSA]_0$) is fit to the flowing equation:

$$I = I_F \times ([D]_0 - [D_B]) + I_B \times [D_B] \,, \tag{5.4.2}$$

where I_F and I_B are the fluorescence intensities of the free and BSA-bound Oligo1 molecules, respectively, which, along with K_d, were treated as fitting parameters. As shown (Figure 4), a single-binding-site model (i.e., n = 1) does not fit the experimental data well. Whereas the fluorescence binding curve can be fit reasonably well by a three-binding-site model (i.e., n = 3) with a K_d = 125 nM. While binding cooperativity is not considered in this simple model, the result is consistent with many studies showing that BSA can provide multiple binding sites for various ligands. In addition, the value of K_d is in the range of those determined for other DNA-protein systems, which are usually in the pM to nM range. Although BSA is not a specific DNA binder, this result indicates that it does have a relatively high affinity for ssDNA (Cattan et al., 1969; Peters, 1985).

To further validate this fluorescence binding assay, we used ITC to determine the BSA binding affinity of Oligo1*, whose sequence is identical to that of Oligo1 except that the 4CNI-NS base is replaced with guanine. As shown (**Figure 5.5**), the ITC measurements yielded a binding curve that is similar to that determined for Oligo1 via fluorescence spectroscopy. Therefore, this result not only corroborates the aforementioned fluorescence method but also demonstrates that replacing a guanine base in a DNA with 4CNI-NS will not significantly affect its interaction with proteins.

FTIR Study. Identifying UNSs that have a unique vibrational mode that can be used as a site-specific IR probe of polynucleotides would open up new avenues in the study of DNA/RNA

structure and dynamics using IR spectroscopy. Herein, we examine whether 5NI-NS and 4CNI-NS affords such utility. 5NI-NS is a widely-used universal base because of its ability to pair with all four nucleobases (A, T, C, G) through aromatic stacking. The asymmetric stretching frequency of the $-NO_2$ (nitro) group in nitrobenzene is around 1550 cm^{-1} (Clarkson and Smith, 2003), suggesting that this vibrational mode of 5NI-NS could be useful as a site-specific IR probe. To verify this notion, we measured the FTIR spectra of 5-nitro-indole (5NI), the functional group of 5NI-NS, in different solvents. As shown (**Figure 5.6a**), in the spectral region of 1500–1600 cm^{-1}, the FTIR spectrum of 5NI is rather complex and, perhaps more importantly, the asymmetric stretching frequency of the nitro group does not exhibit a simple dependence on the solvent. For example, the frequencies obtained in ethanol, a protic solvent with a dielectric constant of 24.5, and dimethyl ether, an aprotic solvent with a dielectric constant of 4.3, are nearly identical. In addition, polynucleotides containing A and G show intrinsic vibrational bands in this region of the spectrum (Seuvre and Mathlouthi, 1987; Wood, 2016). These factors suggest that 5NI-NS is unlikely to be useful as a site-specific IR probe of DNA structure and dynamics.

In comparison, the C≡N stretching vibrational band of 4CNI-NS is simpler and exhibits a more sensitive solvent dependence (**Figure 5.6b** and **Table 5.1**). For example, its frequency is ca. 2232 cm^{-1} in water, which is shifted to ca. 2226 cm^{-1} in THF. This change is similar to that observed for p-cyanophenylalanine (Getahun et al., 2003), a widely used site-specific IR probe of proteins, and 4CN-Trp (van Wilderen et al., 2018). Thus, these results, in conjunction with the fact that the C≡N stretching band is in an uncongested region of the IR spectrum of most natural bio-macromolecules (Ma et al., 2015; Serrano et al., 2012a), suggests that 4CNI-NS is a suitable site-specific IR probe of local hydration and electrostatic environment of DNA and RNA.

Zhang et al. (Zhang et al., 2016) have shown that the C≡N stretching frequency of another cyanoindole (i.e., 3-methyl-5-cyanoindole) is linearly correlated with an empirical solvent parameter $\sigma = \pi^* + \beta - \alpha$, where π^* (polarizability), β (hydrogen bond accepting ability), and α (hydrogen bond donating ability) are the Kamlet–Taft solvent parameters (Kamlet et al., 1983). This linear relationship suggests that both specific interactions, i.e., hydrogen-bonding

interactions (through α and β), and non-specific interactions (through π*) with the molecule work together to determine the C≡N stretching frequency of cyanoindoles. Because of the structural similarity between 4-cyanoindole and 5-cyanoindole, we expect that the C≡N stretching frequency of 4CNI-NS also shows a linear dependence on σ, as observed (**Figure 5.7**). Such a linear dependence is quite useful in practice, as it allows a more straightforward and quantitative interpretation of the result (according to the Kamlet–Taft treatment).

As discussed above, the fluorescence spectrum of BSA-bound Oligo1 is only modestly blue-shifted (i.e., by ca. 8 nm). This suggests that the local electrostatic environment of the 4CNI-NS in Oligo1 does not change significantly upon BSA binding. Consistent with this picture, the C≡N stretching frequency of 4CNI-NS in Oligo1 is only red-shifted by ca. 1.2 cm^{-1}, when bound to BSA (**Figure 5.8a**). Taken together, these results not only confirm the practical utility of the C≡N stretching mode of 4CNI-NS but also suggest that the interaction between Oligo1 and BSA is electrostatic in nature. If the binding was mainly controlled by hydrophobic forces, one would expect to observe a much larger shift in both the IR and fluorescence spectra. Indeed, in support of this notion, the fluorescence intensity of 4CNI-NS is significantly decreased when 2 M NaCl is added to the Oligo1-BSA solution in question (**Figure 5.8b**), due to salt-induced dissociation of Oligo1 from the protein through the charge screening effect. Finally, it is worth noting that the C≡N stretching frequency of individual 4CNI-NS in water is ca. 2232.3 cm^{-1}, whereas that of Oligo1 in water is ca. 2230.5 cm^{-1}. This red-shift is consistent with the aforementioned notion that Oligo1 can adopt an ensemble of compact conformations, leaving the 4CNI-NS base partially dehydrated.

5.5: CONCLUSIONS

We find that the C≡N stretching frequency of 4CNI-NS, a universal nucleobase, is dependent on the solvent. Because of its sensitivity to the environment and the fact that this vibrational band is located in an uncongested region of the IR spectrum of biological macromolecules, it can be used as a site-specific vibrational probe to assess the local hydration and electrostatic

environment of DNA and DNA–protein complexes. Furthermore, we devised a fluorescence assay, which relies on the quenching of 4CNI-NS fluorescence by guanine, for determining the binding constant of ssDNA–protein complexes. Since fluorescence measurement is easy, widely-available, low-cost, and can be performed in a high-throughput manner, we believe that this method will find wide application in the study of DNA–protein interactions. Moreover, it is our expectation that this fluorophore-quencher pair can find other novel applications. For example, it can be used (1) to study the kinetics of DNA–DNA interactions; (2) determine the conformational distribution of ssDNAs, similar to that done for peptides (Matthew J. Tucker et al., 2005a); (3) characterize the rate of intermolecular contact formation of ssDNAs; (4) study the thermodynamics and kinetics of DNA/RNA folding; and (5) interrogate DNA base flipping dynamics, similar to that done by 2-aminopurine (Holz et al., 1998).

Table 5.1: The center frequency (ω_0) and full-width at half maximum (FWHM) of the C≡N stretching band of 4-cyanoindole-2′-deoxyribonucleoside (4CNI-NS) in different solvents. Also listed for each solvent are its Kamlet–Taft parameters, taken from (Kamlet, et al., 1983).

Solvent	ω_0, cm^{-1}	FWHM, cm^{-1}	π^*	β	α	σ
Water (H$_2$O)	2232.3	11.1	1.09	0.47	1.51	0.05
Methanol (MeOH)	2230.8	9.6	0.6	0.62	0.93	0.29
Dimethyl sulfoxide (DMSO)	2223.7	7.4	1	0.76	0	1.76
Acetonitrile (ACN)	2226.8	10.1	0.75	0.31	0.19	0.87
Tetrahydrofuran (THF)	2225.8	7.17	0.58	0.55	0	1.13
Trifluoroethanol (TFE)	2235.4	22.6	0.73	0	1.51	−0.78
Chloroform (CHCl$_3$)	2228.2	11.2	0.53	0.1	0.2	0.43

5-nitroindole-2′-deoxyribonucleoside
5NI-NS

4-cyanoindole
4CNI

4-cyanoindole-2′-deoxyribonucleoside
4CNI-NS
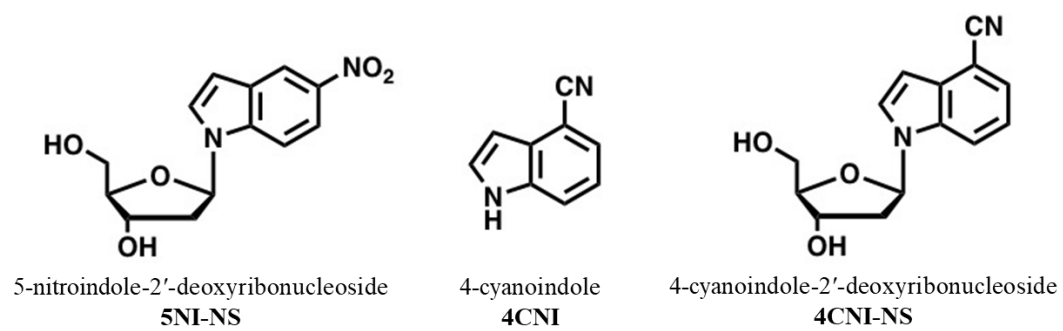
Figure 5.1: Structures of 5-nitroindole-2′-deoxyribonucleoside (5NI-NS), 4-cyanoindole (4CNI) and 4-cyanoindole-2′-deoxyribonucleoside (4CNI-NS).
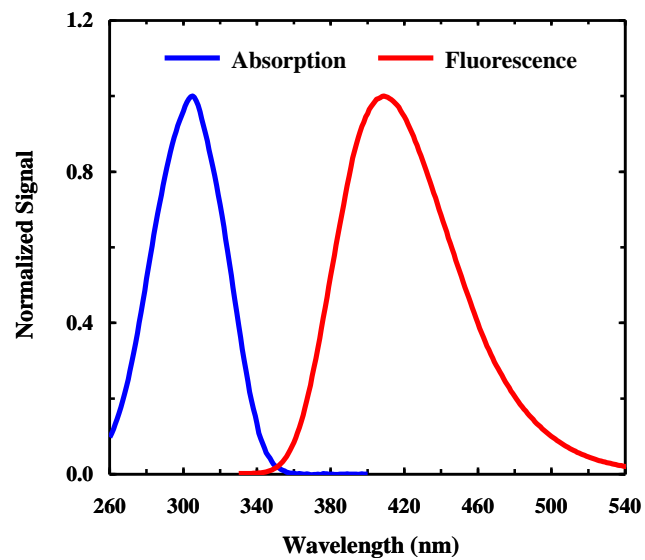
Figure 5.2: Normalized absorption and fluorescence spectra of 4-cyanoindole-2′-deoxyribonucleoside (4CNI-NS) (λ_{ex} = 320 nm) in water, as indicated.
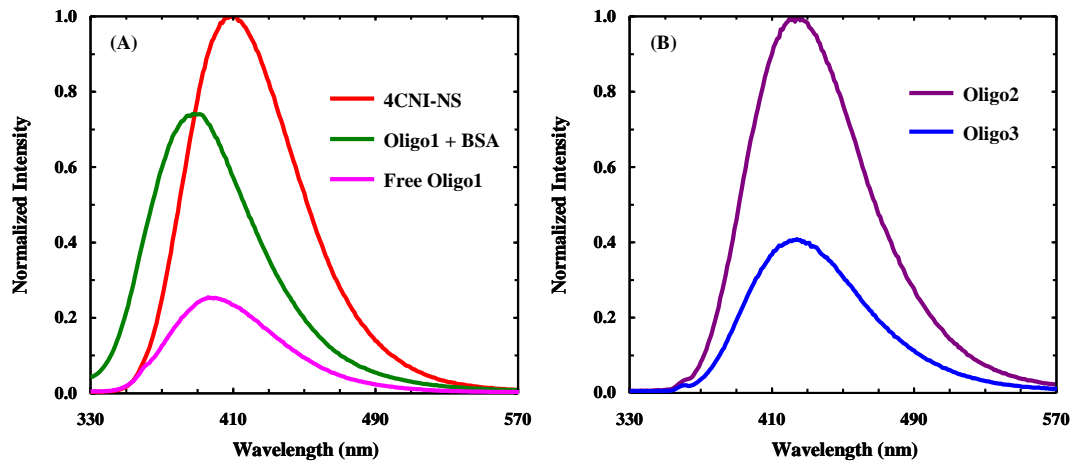
Figure 5.3: (**A**) Normalized fluorescence spectra of 4-cyanoindole-2′-deoxyribonucleoside (4CNI-NS), free Oligo1, and an Oligo1-BSA mixture, as indicated. (**B**) Normalized fluorescence spectra of Oligo2 and Oligo3, as indicated. These spectra were collected under the same experimental conditions (i.e., temperature, excitation/emission slit, and fluorophore absorbance at λ_{ex} = 320 nm).
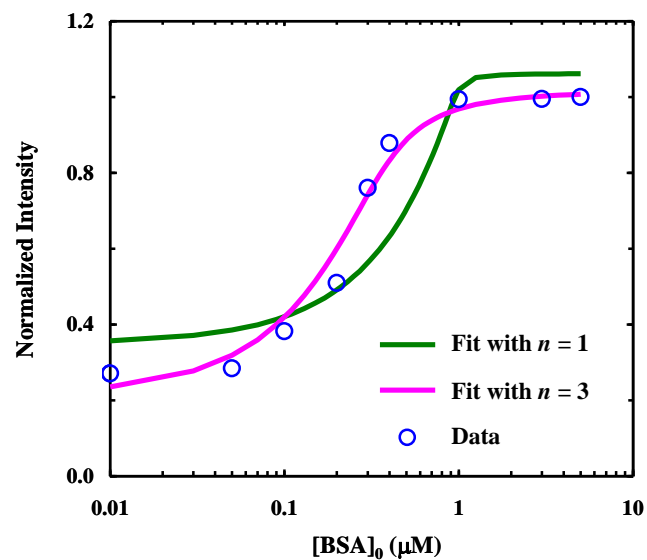
Figure 5.4: Normalized fluorescence intensity of Oligo1 (1.0 µM) as a function of the total concentration of BSA (open circles). The solid lines are best fits of these data to Equation (2) with different n values, as indicated. For $n = 3$, the fit yielded a K_d of 125 nM.
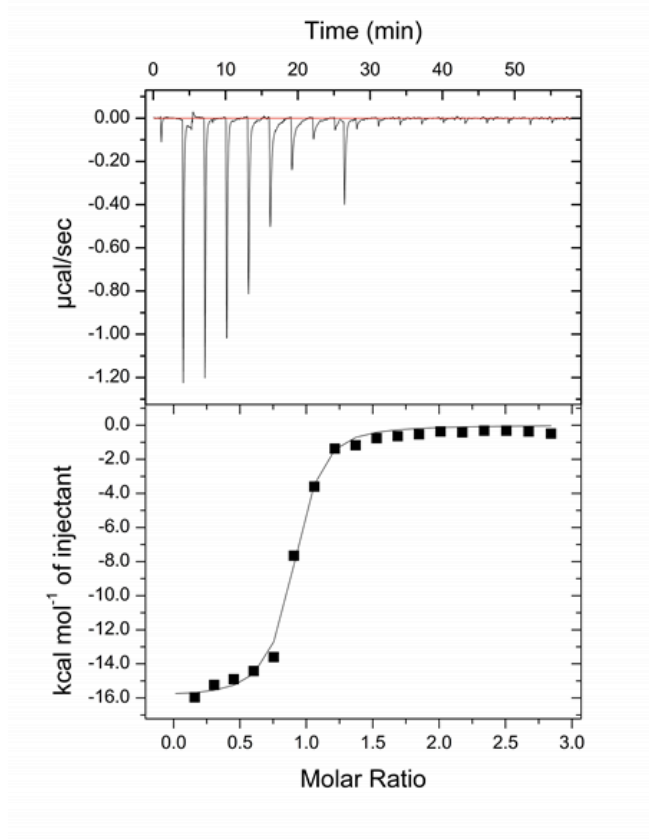
Figure 5.5: ITC heat signature (top panel) and the corresponding binding curve (bottom panel) when Oligo1* (30 µM) is titrated with BSA (300 µM). The solid line in the bottom panel is a guide to the eye.
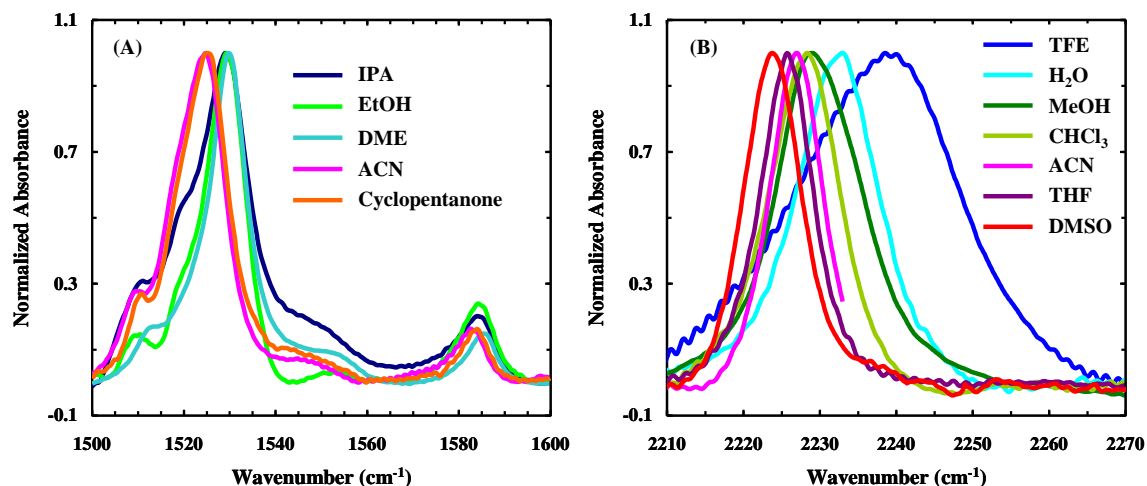
Figure 5.6: (**A**) FTIR spectra (in the region of the asymmetric stretching vibrational band of the nitro group) of 5-nitroindole (5NI) in isopropanol (IPA), ethanol (EtOH), dimethyl ether (DME), acetonitrile (ACN), and cyclopentanone, as indicated. (**B**) The C≡N stretching band of 4-cyanoindole-2′-deoxyribonucleoside (4CNI-NS) in different solvents, as indicated. Only part of the spectrum obtained in ACN is shown, due to the interference of the C≡N stretching band of the solvent at 2267 cm^{-1}. For easy comparison, the original spectra, which have a maximum absorbance of ca. 10 mOD, have been normalized.
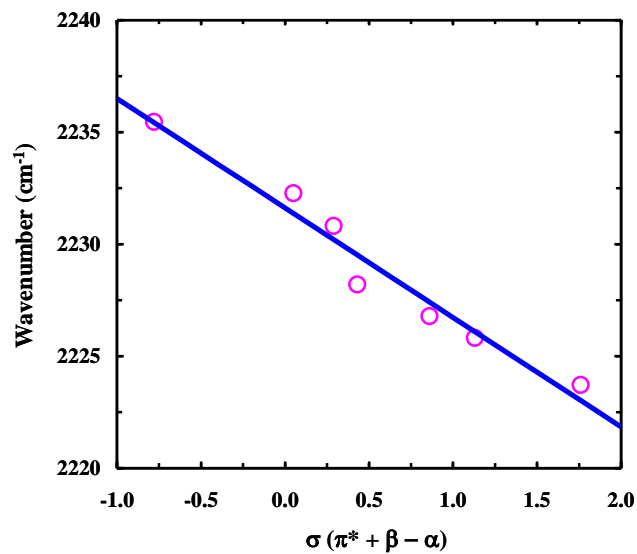
Figure 5.7: Center frequency of the C≡N stretching band of 4-cyanoindole-2′-deoxyribonucleoside (4CNI-NS) versus the solvent σ parameter. The solid line represents the linear regression of these data, yielding a slope of -4.9 cm^{-1} and an intercept of 2231.6 cm^{-1}.
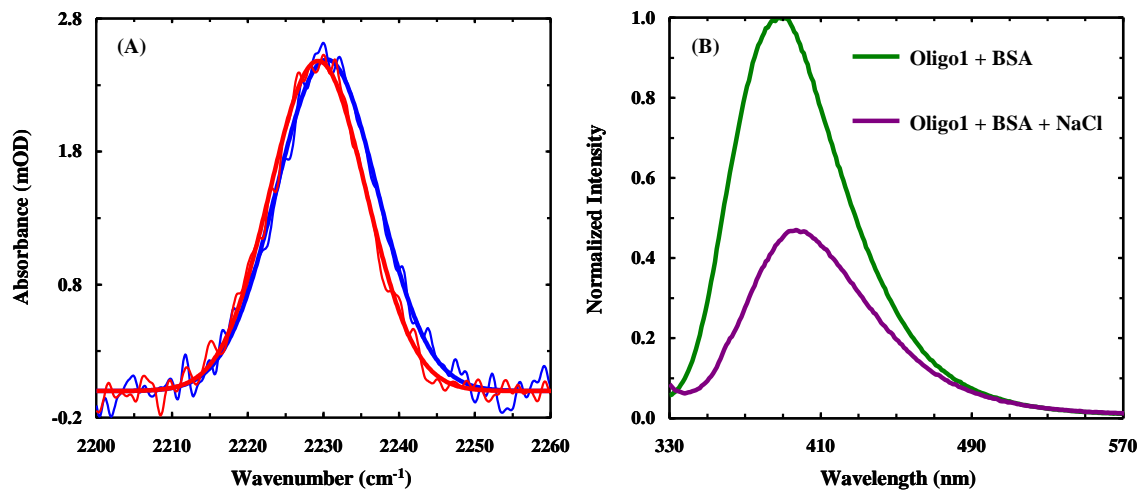
Figure 5.8: (**A**) Comparison of the C≡N stretching bands of Oligo1 obtained in the presence (red) and absence (blue) of BSA. (**B**) Normalized fluorescence spectra of an Oligo1-BSA mixture in the presence and absence of 2 M NaCl, as indicated.

6: Sensing pH via *p*-Cyanophenylalanine Fluorescence: Application to Determine Peptide pKₐ and Membrane-Penetration Kinetics

6.1: ABSTRACT

We expand the spectroscopic utility of a well-known infrared and fluorescence probe, *p*-cyanophenylalanine, by showing that it can also serve as a pH sensor. This new application is based on the notion that the fluorescence quantum yield of this unnatural amino acid, when placed at or near the N-terminal end of a polypeptide, depends on the protonation status of the N-terminal amino group of the peptide. Using this pH sensor, we are able to determine the N-terminal pK_a values of nine tripeptides and also the membrane penetration kinetics of a cell-penetrating peptide. Taken together, these examples demonstrate the applicability of using this unnatural amino acid fluorophore to study pH-dependent biological processes or events that accompany a pH change.

6.2: INTRODUCTION

Recently, p-cyanophenylalanine (Phe$_{CN}$) has emerged as a convenient and versatile site-specific fluorescence reporter for various biochemical and biophysical studies (Ma et al., 2015). The broad spectroscopic utility of this unnatural amino acid, which is a structural derivative of tyrosine (Tyr) or phenylalanine (Phe) (Meloni and Matsika, 2014), stems from the fact that its fluorescence quantum yield and lifetime are sensitive to the environment (Serrano et al., 2010), as well as its easy incorporation into peptides and proteins (Schultz et al., 2006). For example, dehydration leads to a significant decrease in the fluorescence intensity of Phe$_{CN}$, thus making it a useful probe of processes that involve exclusion of water, such as protein folding (Aprilakis et al., 2007; Serrano et al., 2012c), binding (Tang et al., 2009; Tucker et al., 2006), aggregation (Du et al., 2013; Marek et al., 2008, 2010), and interaction with membranes (Jia Tang et al., 2007; Liu

et al., 2009; Loll et al., 2014; Matthew J. Tucker et al., 2006). In addition, the fluorescence quantum yield of Phe_{CN} can be modulated by various metal ions (Pazos et al., 2013) and several amino acid sidechains (Glasscock et al., 2008; Goldberg et al., 2010; Matthew J. Tucker et al., 2005b; Taskent-Sezgin et al., 2009). In particular, its fluorescence can be quenched by a nearby tryptophan (Trp) residue via the mechanism of fluorescence resonance energy transfer (FRET) (Reif and Oostenbrink, 2014; Serrano et al., 2012b; Wissner et al., 2013). Since this FRET pair has a Förster radius of approximately 16 Å, it has become a very useful tool to probe protein conformational changes over a relatively short distance (Mintzer et al., 2015). Furthermore, Raleigh and coworkers have shown that the fluorescence intensity of Phe_{CN}, when placed at or near the N-terminus of a peptide, depends on the protonation status of the N-terminal amine group of the peptide (Taskent-Sezgin et al., 2010). However, the potential utility of this pH dependence of the Phe_{CN} fluorescence has not been demonstrated. Herein, we show that this unnatural amino acid can be used as a pH sensor to study pH-dependent biological processes or events that accompany a pH change. Specifically, we carry out two experiments to demonstrate the utility of Phe_{CN} fluorescence as a pH sensor; in the first one, we use it to determine the N-terminal pK_a values of a series of short peptides and, in the second one, we employ it to characterize the membrane penetration kinetics of a cell-penetrating peptide, the trans-activator of transcription (TAT) peptide derived from HIV-1 (Jones and Sayers, 2012; Shen et al., 2007; Tseng et al., 2002).

Several methods, including high voltage electrophoresis (Wallis, 1973), circular dichroism (CD) (Chakrabartty et al., 1993; Doig and Baldwin, 1995; Goto and Hagihara, 1992; Wilcox and Eisenberg, 1992) and nuclear magnetic resonance (NMR) spectroscopy (Brown et al., 1980; Zhu et al., 1995), have been used to determine the N-terminal pK_a values of unstructured or α-helical peptides (Pace et al., 2009). However, due to various limitations, for instance, the CD method relies on the peptide of interest to form a well-folded structure such as an α-helix, applying these methods in practice is not always feasible, straightforward or convenient. Since fluorescence measurements are easier and also accessible to most, if not all, biochemical and biophysical

researchers, it would be advantageous to devise a method that allows determination of N-terminal pK_a values of peptides and proteins via fluorescence spectroscopy.

Cell-penetrating peptides (CPPs) are short cationic peptides which can spontaneously translocate across cell membranes and, as such, are ideal vehicles to deliver exogenous cargos into cells (Zorko and Langel, 2005). In spite of many previous efforts, however, several aspects of the translocation actions of CPPs are not well understood or characterized. For example, measurements of the intrinsic membrane penetration kinetics of CPPs are often done by measuring the fluorescence signal of a dye molecule attached to the CPP of interest. However, the fluorescent dyes used in this type of studies are typically large in size and, consequently, may have a significant effect on the CPP's penetration rate. Here we show, using TAT as an example, that this concern can be alleviated by using Phe_{CN} fluorescence as a probe to follow the kinetics of CPP membrane penetration. Specifically, we replaced the N-terminal Tyr residue of TAT with Phe_{CN}, (the resultant peptide is hereinafter referred to as F_{CN}-TAT), which is expected to cause only a minimum perturbation to the peptide as Tyr and Phe_{CN} are similar in size, and we exploited the pH-dependence of the Phe_{CN} fluorescence to quantitatively assess the penetration kinetics of TAT across a model membrane (Swiecicki et al., 2014). Interestingly, we find that under our experimental conditions TAT first translocates across the membrane on a timescale of minutes and then causes membrane leakage on a timescale of hours.

6.3: EXPERIMENTAL SECTION

Peptide Synthesis and Sample Preparation. Peptides were synthesized on a PS3 automated peptide synthesizer (Protein Technologies, MA) using standard 9-fluorenylmethoxy-carbonyl (Fmoc) solid phase synthesis protocols and Fmoc-protected amino acids from either Bachem Americas (Torrence, CA) or AnaSpec (Fremont, CA). Before use, all peptide samples were further purified by reverse-phase high-performance liquid chromatography and verified by mass spectrometry. All peptide samples used in the pH titration measurements were prepared by dissolving lyophilized peptides into either acidic (25 mM H3PO4) or basic (25 mM NaOH)

Millipore water, and the final concentration of each sample (20 μM) was determined optically using the absorbance of PheCN at 280 nm and a molar extinction coefficient of 850 M^{-1} cm^{-1} (Tucker et al., 2006).

Fluorescence Measurement and pH Titration. Fluorescence spectra of the PheCN-containing peptides were measured at 25 oC on a Fluorolog 3.10 spectrofluorometer (Horiba, NJ) using a 1 cm quartz cuvette, a spectral resolution of 1 nm, an excitation wavelength of 275 nm, and an integration time 1 s/nm. During a specific pH titration experiment, the concentration of the peptide under consideration was maintained at 20 μM and the pH of the solution was varied between 2 and 12. This was achieved by mixing two 20 μM peptide stock solutions, one prepared in 25 mM H3PO4 and the other in 25 mM NaOH aqueous solution, in a cuvette at an appropriate volume ratio to achieve the desired pH, which was further measured using a pH meter (Orion/ThermoScientific, MA) Additionally, in each case a background spectrum, obtained with the buffer solution, was subtracted. Because the shape of the PheCN fluorescence spectrum is insensitive to the environment, only the peak intensity was used to generate the pH titration curves. To determine the N-terminal pK_a of each peptide, the corresponding fluorescence titration curve was fit to the following equation:

$$F(x) = \frac{(b_1 + m_1 x)(10^{-x}) + (b_2 + m_2 x)(10^{-pK_a})}{10^{-pK_a} + 10^{-x}} ,$$
(6.3.1)

where $F(x)$ is the fluorescence intensity at pH x, b_i and m_i are the intercept and slope of the linear base lines (i = 1 for acidic and i = 2 for basic).

Measurement of the Membrane Penetration Kinetics of F$_{CN}$-TAT. The 100 nm, large unilamellar vesicles (LUVs) used in the peptide penetration experiments were composed of either 100% DOPG or a mixture of DOPC, DOPG and cholesterol (3:1:1) and prepared following previously published procedures (Aurora et al., 1985) from stock solutions of the respective lipids (purchased from Avanti Polar Lipids Inc., AL). Briefly, the respective lipid solution in chloroform

was first dried under a flow of nitrogen, allowing a lipid film to form, which was followed by a 30-minute lyophilization to remove any remaining solvent. The resultant lipid film was then rehydrated with Millipore water, and the pH of the sample was adjusted to 3.5 - 4.0 using H_3PO_4 and NaOH. This sample was then subjected to seven rounds of slow vortexing, freezing and thawing to form LUVs. The resulting vesicle solution was then extruded 11 times through an extruder (Avanti Polar Lipids Inc., AL) equipped with a 100 nm membrane. After extrusion, the LUV solution was diluted to 100 µM (lipid concentration) with Millipore water, and the pH was adjusted between 10.0 - 10.5 with H_3PO_4 and NaOH. This process resulted in LUV's with acidic pH inside the vesicle and basic pH outside. All LUV solutions were stored at 4 °C and used within one week after preparation.

The membrane penetration kinetics of F_{CN}-TAT was initiated by manually adding an appropriate aliquot of a concentrated peptide solution (350 µM, pH 7.0) to a 2.0 mL LUV solution described above. The final peptide concentration was 1.0 µM, resulting in a 1:100 peptide to lipid ratio. The time-dependent fluorescence intensity of F_{CN}-TAT at 298 nm was collected at 25 °C using the time-based acquisition function of the Fluorolog 3.10 spectrofluorometer with time intervals of 2 seconds for the first 1,000 seconds, then every 10 seconds for the next 10,000 seconds and then 100 seconds for the remainder of the experiment. These data were then binned and the initial kinetic phase corresponding to the membrane penetration process was fit to a single-exponential function. To remove the contribution of water's Raman scattering to the fluorescence signal, most of the kinetic traces were collected using an excitation wavelength of 240 nm. In other cases, an excitation wavelength of 275 nm was used. The pH of the solution was recorded throughout the duration of the experiment.

The fluorescence quenching assay used to estimate the timescale of membrane leakage was modified from a liposome flux assay (Ma et al., 2009). Briefly, the LUVs were prepared using the abovementioned procedures, with the exception that a membrane impermeable fluorescent dye, pyranine (Alfa Aaser, MA), was added to the lipid solution. The concentration of pyranine

was 4 µM in phosphate buffer (pH 7.5, 30 mM sodium phosphate and 100 mM NaCl). After extrusion of the lipid solution, the dye molecules that were not encapsulated inside LUVs were separated from the LUVs by passing the mixture through a pD-10 column (GE Healthcare, NJ). The dye-encapsulated LUV solution was then diluted into the same phosphate buffer that also contains 25 µM p-xylene-bis-pyridinium bromide (DPX) (Sigma-Aldrich, MO), a pyranine fluorescence quencher, to yield a final lipid concentration of 100 µM. The pyranine fluorescence was excited at 417 nm (isosbestic point) and monitored at 515 nm with and without the presence of F_{CN}-TAT peptide.

6.4: RESULTS AND DISCUSSIONS

Effect of the N-terminal Amine Group on Phe$_{CN}$ Fluorescence. Previous studies have suggested that the fluorescence of a Phe$_{CN}$ residue, at or near the N-terminus of a peptide, can be significantly quenched by the deprotonated neutral form of the N-terminal amino group (Pazos et al., 2013; Taskent-Sezgin et al., 2010). To further validate this notion, we measured the PheCN fluorescence spectra of a tripeptide consisting of glycine (G) and Phe$_{CN}$ (F_{CN}) with a sequence of GF$_{CN}$G-CONH2, having either a free amino N-terminal end (the corresponding peptide is hereafter referred to as GF$_{CN}$G) or an acetylated N-terminus (the corresponding peptide is hereafter referred to as *GF$_{CN}$G). As shown (**Figure 6.1**), the Phe$_{CN}$ fluorescence intensity of GF$_{CN}$G shows a drastic decrease when the pH of the peptide solution is increased from 2.1 to 10.0. As the N-terminal pK_a of peptides is typically smaller than 9.0 and, thus, at pH 10.0 the N-terminus of GF$_{CN}$G is expected to be deprotonated, this result suggests that the N-terminal amino group is an effective quencher of the PheCN fluorescence in this case. In support of this conclusion, the PheCN fluorescence of *GF$_{CN}$G does not show such a pH dependence within the pH range studied (**Figure 6.1** inset). Taken together, these results indicate that a PheCN residue, when placed at the second position in a peptide sequence, can be used as a pH sensor.

To demonstrate its application in this regard, we first used Phe$_{CN}$ fluorescence to determine the N-terminal pK_a of GF$_{CN}$G. As shown (**Figure 6.2**), the intensity of Phe$_{CN}$

fluorescence of GF$_{CN}$G displays a sigmoidal dependence on pH, characteristic of an acid-base titration. As expected, the peptide *GF$_{CN}$G, whose N-terminus is acetylated, does not show such a transition. Further fitting the fluorescence titration curve of GF$_{CN}$G to a two-state model (i.e., eq. 5.3.1) yielded a pK$_a$ of 7.99, which is similar to the N-terminal pK$_a$ (7.85) of the amino group of an antimicrobial peptide [36], melittin, which has a Gly residue at the N-terminus. Interestingly, as indicated (**Figure 6.2**), swapping of the first two amino acids in GF$_{CN}$G (the resulting peptide is hereafter referred to as F$_{CN}$GG) leads to a decrease in not only the N-terminal pK$_a$ value (i.e., 6.97) but also the quantum yield of PheCN fluorescence. The latter is suggestive of a fluorescence quenching mechanism that is electron transfer in nature.

Determining N-terminal pK$_a$ Values of a series of Tripeptides. The results obtained with GF$_{CN}$G and F$_{CN}$GG indicate that the dynamic range of the PheCN as a pH sensor can be tuned by changing the identity of the first amino acid. To this end, we carried out acid-base titrations on eight additional peptides having the following sequence: NH$_2$-(XF$_{CN}$G)-CONH$_2$, where X represents a different amino acid in each case. Specifically, we chose several amino acids with varying sidechain properties, including positively charged (K), negatively charged (D), polar (N, S, G), or nonpolar (G, M, P, V, A) sidechains, to demonstrate the sensitivity of the fluorescence of Phe$_{CN}$. As shown (**Figure 6.3**), the corresponding pH titration curves of these peptides measured via Phe$_{CN}$ fluorescence intensity have similar sigmoidal shapes but different mid-points (i.e., pK$_a$ values). Indeed, the N-terminal pK$_a$ values of these peptides, determined by fitting the respective acid-base titrations curves to eq. 6.3.1, are spread over two units of pH, ranging from 6.7 to 8.6 (**Table 6.1**). In particular, the N-terminal pK$_a$ (7.99) of the AFCNG tripeptide agrees well with that (8.0) determined for two alanine-based pentapeptides (NH$_2$-AAEAA-Ac and NH$_2$-AAHAA-Ac [43]. Because of the scarcity of N-terminal pK$_a$ values of short peptides, however, making more comparisons with other studies is not possible. On the other hand, it is clear that the N-terminal pK$_a$ value of the XF$_{CN}$G peptide is, in most cases, smaller than that determined for an α-helical peptide with the same N-terminal residue (**Table 6.1**), suggesting that the peptide structure has a significant effect on the electronic property of the N-terminal amino group. In addition, as

145

indicated (**Figure 6.4**), except for the proline (P) and serine (S) variants, the N-terminal pK_a values of other peptides show a weak linear correlation with those (i.e., pK_2) of the corresponding free amino acids. While further confirmation is needed, the large deviation observed for proline and serine likely results from their interactions with the PheCN sidechain. Regardless of the exact mechanism that controls the N-terminal pK_a values of these tripeptides, the data in **Figure 6.3** indicates that by changing the N-terminal residue, the useful dynamic range of Phe_{CN} fluorescence can be extended over five units of pH, from approximately 5 to 10.

Using Phe_{CN} Fluorescence to Probe the Membrane-Penetrating Kinetics of TAT. In this example, we seek to demonstrate that Phe_{CN} can serve as a pH sensor to monitor the rate of membrane penetration of CPPs. Specifically, we employed a well-studied CPP, TAT (YGRKKRRQRRR-$CONH_2$) as our model peptide. As discussed above, the Phe_{CN} pH sensor was introduced via a Tyr to Phe_{CN} mutation. Previous studies have shown that the Tyr residue plays a minimal, if any, role in facilitating TAT translocation across membranes. Thus, we expect that the mutant peptide (i.e., F_{CN}-TAT) will exhibit similar, if not identical, membrane penetration kinetics as the wild type peptide. As expected (**Figure 6.5** inset), the Phe_{CN} fluorescence intensity of F_{CN}-TAT depends on pH. To utilize this dependence to report the membrane penetration kinetics of F_{CN}-TAT, we prepared a DOPG vesicle solution where the pH inside the vesicle was 3.5 and the pH outside the vesicle was 10.0. As shown (**Figure 6.5**), upon mixing this vesicle solution with a F_{CN}-TAT solution at pH 10.0, the Phe_{CN} fluorescence increases with time in two well separated kinetic phases. This increase in Phe_{CN} fluorescence is consistent with the idea that a certain fraction of the peptide molecules have translocated across the DOPG membranes and, thus, are experiencing a significant decrease in pH. However, repetitive measurements indicate that while the first or fast kinetic phase is well defined and reproducible, the second phase appears to be ill-defined and occurs on a much longer timescale. These differences prompt us to assign the fast phase to peptide internalization in vesicles (or membrane penetration) and the slow one to membrane leakage. This is based on the notion that, unlike the penetration process, a number of peptide molecules need to work together, for example, through the formation of transmembrane

pores (Ciobanasu et al., 2010; Katsumi Matsuzaki et al., 1996), to cause the membrane to leak; as a result, the leakage kinetics are intrinsically more stochastic like and statistically more sensitive to the experimental uncertainties (e.g., variations in the peptide and vesicle concentrations). To confirm this assignment, we measured the pH of the peptide-vesicle solution at different reaction times and found that after the completion of the fast kinetics phase (e.g., at 50 minutes) the pH of the system was practically unchanged (i.e., 10.0 to 9.9), whereas after 5 hours the pH of the solution was decreased to 7.4. On the other hand, adding the $GF_{CN}G$ tripeptide to the same DOPG vesicle solution did not cause any appreciable change in the pH value after 12 hours. Thus, taken together these results support the aforementioned assignment. Moreover, the notion that the slow kinetic phase reports on peptide-induced membrane leakage is further corroborated by the data obtained from a dye leakage experiment. As shown (**Figure 6.5**), under similar experimental conditions (i.e., same peptide and lipid concentrations) the fluorescence intensity of a dye (pyranine) that is initially encapsulated inside the DOPG vesicle shows a significant decrease (approximately 50% without counting photobleaching) over 10 hours, due to membrane leakage and consequently fluorescence quenching by a quencher (DPX) initially existing outside the vesicles. As indicated (**Figure 6.6**), the fast or peptide penetration kinetic phase can be fit by a single-exponential function with a time constant of 10.5 ± 2.2 minutes. This value is similar to those reported in the literature for TAT (Richard et al., 2003; Suzuki et al., 2002; Swiecicki et al., 2014), providing further supporting evidence for our assignment.

Using 240 nm photons to excite the PheCN fluorescence may lead to the undesirable heating effect. To verify that this is not the case, we also carried out fluorescence kinetic measurements using an excitation wavelength of 275 nm. As shown (**Figure 6.7**), the resultant penetration kinetics has a time constant of 12.8 ± 2.0 minutes, indicating that the photo-excitation induced heating affects minimally, if any, the kinetic results. Similarly, to better mimic eukaryotic cell membranes, we also conducted a penetration experiment using a ternary mixture of DOPC, DOPG, and cholesterol (Feigenson, 2009; Marsh, 2009). As indicated (**Figure 6.7**), the F_{CN}-TAT

penetration kinetics obtained with the mixed lipid membrane is comparable to those measured with DOPG vesicle, with a time constant of 11.9 ± 1.5 minutes.

6.5: CONCLUSIONS

The unnatural amino acid, p-cyanophenylalanine (Phe_{CN}), has unique infrared (IR) and fluorescent properties. As such, it has been used as both IR and fluorescence probes in a wide range of applications. Here, we further show that, for peptides having a Phe_{CN} residue at or near the N-terminus, the neutral form of the N-terminal amino group is much more effective than the protonated form in quenching the Phe_{CN} fluorescence. As a result, such peptides can be used as pH sensors. By examining the pH-dependence of the Phe_{CN} fluorescence of nine tripeptides (i.e., X-Phe_{CN}-Gly, where X represents either Asn, Met, Lys, Pro, Val, Asp, Ala, Gly or Ser), we are able to show that the dynamic range of such pH sensors can be tuned to cover several units of pH as the N-terminal pK_a of these peptides depends on the identity of the N-terminal amino acid. In addition, using TAT as an example, we demonstrate that the aforementioned pH-dependence of Phe_{CN} fluorescence can be exploited to measure penetration kinetics of cell-penetrating peptides across model membranes.

Table 6.1: The N-terminal pK_a value of the $XF_{CN}G$ tripeptide determined from the current experiment, the pK_2 value of the corresponding free amino acid X, and the N-terminal pK_a value (pK_a^H) of an α-helical peptide with X at the N-terminus. The uncertainty of the measured pK_a values is ±0.1.

X	pK_a	pK_a^H	pK_2
N	6.69	7.07	8.8
M	7.04	7.83	9.2
K	7.25	8.12	9.2
P	7.28	8.85	10.6
V	7.29	8.14	9.6
D	7.70	8.25	10.0
A	7.99	8.35	9.7
G	7.99	8.51	9.8
S	8.59	7.63	9.2
$F_{CN}GG$	6.97		9.1
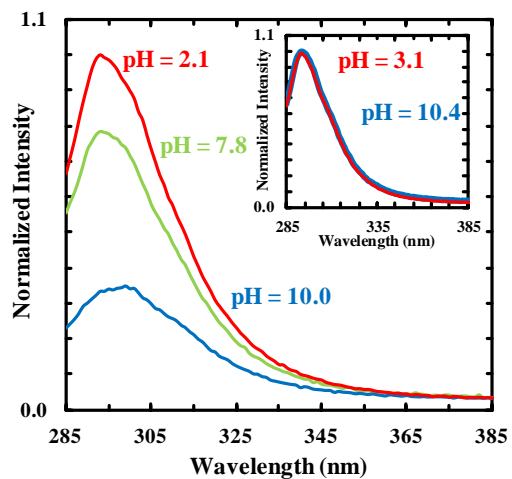
Figure 6.1: Normalized Phe$_{CN}$ fluorescence spectra of GF$_{CN}$G obtained at different pH values, as indicated. Shown in the inset are the normalized Phe$_{CN}$ fluorescence spectra of *GF$_{CN}$G (i.e., the N-terminal acetylated version of GF$_{CN}$G) collected under acidic and basic pH conditions.
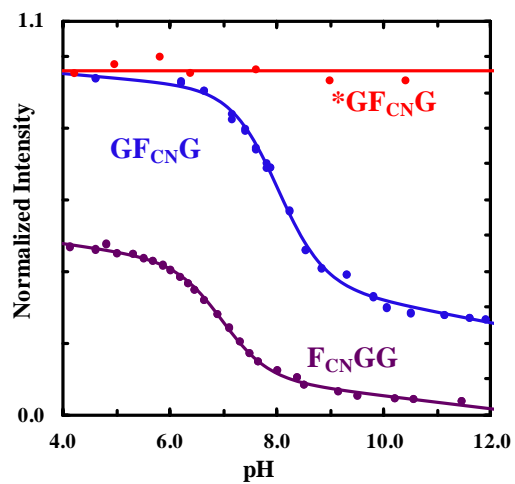
Figure 6.2: Normalized Phe$_{CN}$ fluorescence intensity versus pH of three tripeptides, as indicated. In each case, the smooth line represents the best fit of the corresponding fluorescence titration curve to Eq. 6.3.1 and the resulting pK_a value is listed in Table 6.1.
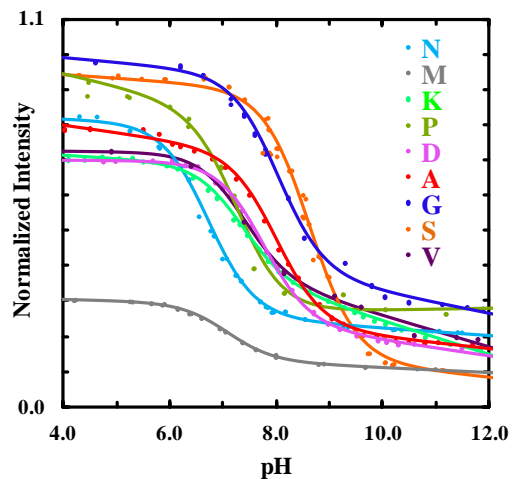
Figure 6.3: Normalized Phe_{CN} fluorescence intensity versus pH of $XF_{CN}G$ peptides, as indicated by the X amino acid. Smooth lines correspond to the best fits of these fluorescence titration curves to Eq. 1 and the resulting pK_a values are given in Table 6.1. In addition, the significantly decreased fluorescence intensity of $MF_{CN}G$ is due to the previously verified quenching effect of methionine sidechain toward Phe_{CN} fluorescence (Taskent-Sezgin et al., 2010).
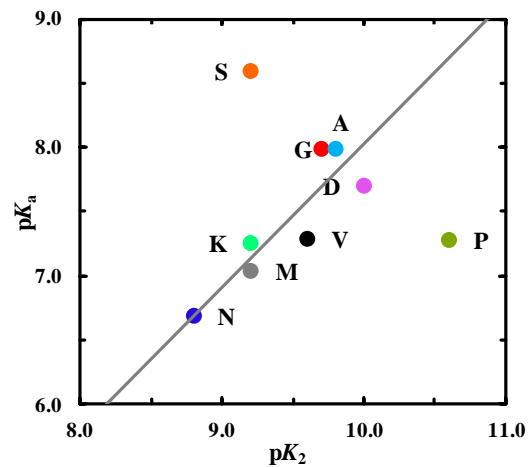
Figure 6.4: Comparison between the N-terminal pK_a of XF$_{CN}$G with the pK_2 of the corresponding free amino acid X. Except for proline and serine, a correlation (R^2 = 0.98) between the pK_a and pK_2 values exists for the amino acids studied.
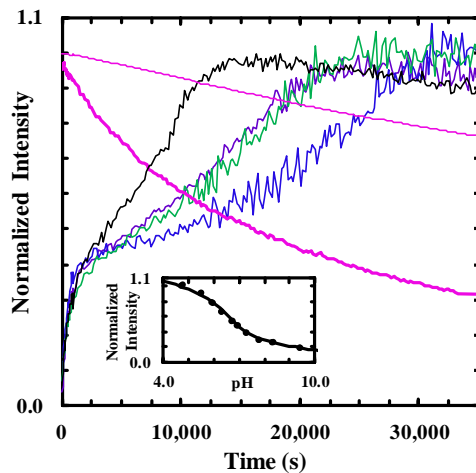
Figure 6.5: Four representative Phe_{CN} fluorescence kinetic traces (black, purple, green and blue) obtained upon mixing an F_{CN}-TAT solution with a DOPG LUV solution. The excitation wavelength was 240 nm. The pink lines are the time-dependent fluorescence intensities of the pyranine dye initially encapsulated inside the DOPG vesicles obtained in the presence (thick line) and absence (thin line) of the F_{CN}-TAT peptide and DPX quencher. In other words, the thin pink line measures the photobleaching rate of the pyranine dye, whereas the thick pink line reports the peptide-induced membrane leakage kinetics. Shown in the inset is the normalized Phe_{CN} fluorescence of F_{CN}-TAT as a function of pH.
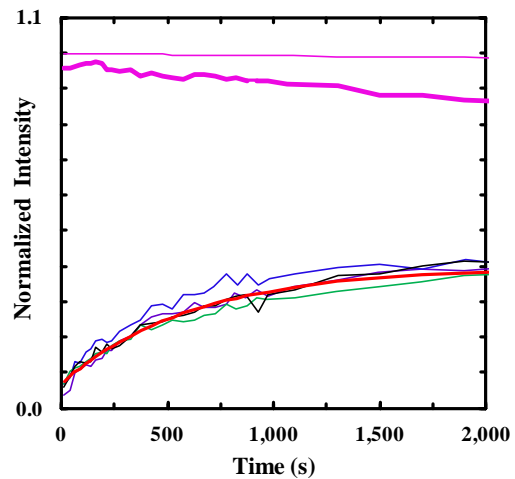
Figure 6.6: The first 2000-second portions of the same Phe_{CN} fluorescence kinetic traces in Figure 6.5. The red line represents the best fit of the averaged data of these four curves to a single-exponential function with a time constant of 10.5 ± 2.2 minutes.
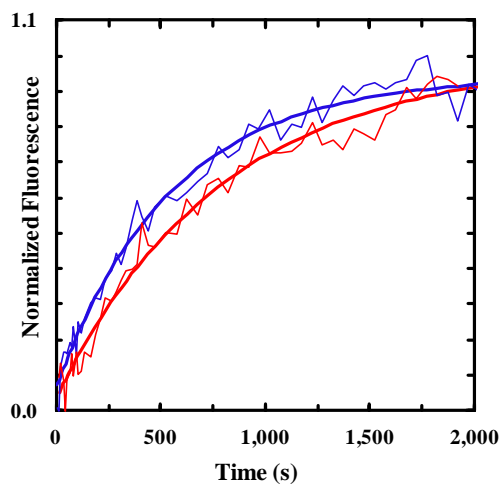
Figure 6.7: The thin red line represents the F_{CN}-TAT penetration kinetics obtained with DOPG LUVs and an excitation wavelength of 275 nm, and the thick and smooth red line represents the best fit of this kinetic trace to a single-exponential function with a time constant of 12.8 ± 2.0 minutes. The thin blue line represents the F_{CN}-TAT penetration kinetics obtained with the ternary lipid membrane described in the text, and the thick and smooth blue line represents the best fit of this kinetic trace to a single-exponential function with a time constant of 11.9 ± 1.5minutes.

7: Simple Method to Introduce an Ester Infrared Probe into Proteins

7.1: ABSTRACT

The ester carbonyl stretching vibration has recently been shown to be a sensitive and convenient infrared (IR) probe of protein electrostatics due to the linear dependence of its frequency on the local electric field. While an ester moiety can be easily incorporated into peptides via solid-phase synthesis, currently there is no method available to site-specifically incorporate it into a large protein. Herein, we show that it is possible to use a cysteine alkylation reaction to achieve this goal and demonstrate the feasibility of this simple method by successfully incorporating a methyl ester group ($-CH_2COOCH_3$) into a model peptide (YGGCGG), two amyloid-forming peptides derived from the insulin B chain and Aβ, and bovine serum albumin (BSA). IR results obtained with those peptide and protein systems further confirm the utility of this vibrational probe in monitoring, for example, the structural integrity of amyloid fibrils and ligand binding-induced changes in protein local hydration status.

7.2: INTRODUCTION

Infrared (IR) spectroscopy is a well-established and widely used technique for studying protein structure, dynamics and function (Ding et al., 2016; Gosavi and Korendovych, 2016). This is because a protein's backbone and sidechains can give rise to a multitude of vibrational transitions, some of which are sensitive to factors that depend on either environment, conformation, or both (Barth, 2007). For example, the amide I′ band arising from the backbone carbonyls has been extensively used to assess protein folding dynamics and mechanisms due to its sensitivity to protein secondary structures (Serrano et al., 2012). Despite the proven utility of various intrinsic vibrational modes of proteins, it is found that in many applications naturally occurring protein IR signals either cannot provide any insight or can only provide limited

information about the biochemical or biophysical process of interest. Therefore, the past decade has seen a continued effort to develop extrinsic IR probes that can circumvent this limitation (Barth, 2007; Ding et al., 2016; Gosavi and Korendovych, 2016; Ma et al., 2015; Serrano et al., 2012a). For instance, to improve the structural and spatial resolution of IR measurements, non-natural amino acids consisting of a unique vibrational probe, such as a -CN, -SCN, or -N_3 moiety, have been utilized in various studies, including those concerning protein folding (Brewer et al., 2007),[6] amyloid formation (Middleton et al., 2012), the structure and function of membrane proteins (Ghosh et al., 2011; Remorino et al., 2011), electrostatic field of enzymes (Choi and Cho, 2011; Levinson and Boxer, 2014; Liu et al., 2014; Stafford et al., 2012; Zimmermann et al., 2011),[10-14] and drug binding (Basom et al., 2016; Johnson et al., 2014).

Recently, Pazos *et al.* (Pazos et al., 2014) have shown that the ester carbonyl stretching vibration is a useful and convenient site-specific IR probe of protein electrostatics because (1) its frequency is linearly correlated with the local electrostatic field, even when the carbonyl group is engaged in hydrogen-bonding interactions; (2) its frequency is located in the spectral range of 1700 – 1800 cm^{-1}, wherein no intrinsic protein vibrations show up at neutral pH; and (3) it has a relatively high extension coefficient. In particular, Pazos *et al.* (Pazos et al., 2014) have tested and demonstrated the utility of two ester-containing non-natural amino acids, L-aspartic acid 4-methyl ester (hereafter referred to as D_M) and L-glutamic acid 5-methyl ester (hereafter referred to as E_M), by incorporating them into various model peptide systems via solid-phase peptide synthesis. In contrast, other non-natural amino acid-based IR probes, such as *p*-cyanophenylalanine (Fafarman and Boxer, 2010; Schultz et al., 2006) and *p*-azido-phenylalanine (Bazewicz et al., 2013; Ye et al., 2009), have been incorporated into large proteins *in vivo* via amber codon suppression (Miyake-Stoner et al., 2009). Currently, there are no *in vivo* methods available to site-specifically introduce either D_M or E_M into a protein of interest. Herein, we demonstrate a simple, robust method for selective incorporation of a methyl ester (hereafter referred to as ME), an IR probe analogous to D_M, into proteins.

As shown (**Figure 7.1**), alkylation through an S_N2 reaction between a free thiol (-SH) and a haloalkyl reagent is commonly used to covalently attach a desired functional (R) group, such as an alkyl, alkenyl, or aryl moiety, to the molecule of interest (Gunnoo and Madder, 2016). Because such alkylation reactions can be carried out under mild reaction conditions, they have found useful applications in protein post-translational modifications. For example, cysteine alkylation (Spicer and Davis, 2014) has been used to incorporate various spectroscopic probes (Jo et al., 2010; Peng et al., 2012). In the current study, we further expand the utility of this simple reaction by showing that it can be exploited to site-specifically label a peptide or protein with an ester IR probe. In addition, we verify the applicability of this IR probe by using it to assess the structural heterogeneity of amyloid fibrils and a ligand-binding induced local hydration status change in bovine serum albumin.

7.3: EXPERIMENTAL SECTION

Peptide Synthesis and Sample Preparation. All peptides were synthesized using standard 9-fluorenylmethoxycarbonyl (Fmoc) solid-phase synthesis protocol with Fmoc-protected amino acids from either Bachem Americas (Torrence, CA) or AnaSpec (Fremont, CA) on a CEM (Matthews, NC) Liberty Blue automated microwave peptide synthesizer. In each case, the C-terminus was capped by amidation. All synthesized peptides were cleaved from the corresponding resin using trifluoroacetic acid, and the resultant crude sample was purified by reverse-phase HPLC (Agilent Technologies 1260 Infinity) using a C18 column (Vydac). The mass of every peptide was verified by matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS). Peptide samples used in cysteine ligation reactions were prepared by dissolving lyophilized peptides into 50 mM sodium phosphate buffer (pH 5) with a final peptide concentration of 1-5 mM, determined optically using the absorbance of either tryptophan, tyrosine, or Lys(Nvoc) at 280 nm or 350 nm. Bovine Serum Albumin (BSA) was purchased from Sigma-Aldrich and used as received.

Ester Labeling via Cysteine Alkylation. For peptide labeling, an appropriate amount of lyophilized peptide solid was first dissolved in sodium phosphate buffer (50 mM, pH ~5) to reach a final peptide concentration of 1 mM. Then, an aliquot of a concentrated acetonitrile solution of methyl bromoacetate (100 mM) was added to 5 mL of this peptide solution to reach a 10:1 molar ratio between methyl bromoacetate and the peptide. The reaction mixture was left to slowly shake on a table top rocker for at least 8 hours at room temperature. Reaction progress was monitored by mass spectrometry until the majority of the starting material was converted to the desired product. The excess methyl bromoacetate from the synthesis was first removed by HPLC, the resultant product was then lyophilized and reconstituted into sodium phosphate buffer (50 mM, pH 7). The procedures used for BSA labeling are identical to those described above with the only differences being (1) the methyl bromoacetate stock solution was prepared in dimethyl sulfoxide (DMSO), (2) a PD-10 column (GE) was used to remove the excess methyl bromoacetate post reaction, and (3) the protein sample was then dialyzed overnight using a Slide-A-Lyzer (Thermo Scientific) dialysis cassette with a 10000 molecular weight cutoff against a sodium phosphate buffer (50 mM, pH 7) to remove any remaining methyl bromoacetate. For labeling reactions conducted in denaturing conditions, the peptide/protein sample was prepared in 8 M urea sodium phosphate buffer (50 mM, pH ~5) while keeping the other conditions the same. In addition, the removal of urea and unreacted methyl bromoacetate was achieved through three rounds of dialysis against a sodium phosphate buffer (50 mM, pH 7.0).

Trypsin Digestion of BSA and BSA-SME. Trypsin was purchased from Promega. The protein to be digested was first denatured by incubating it in a pH 8 buffer containing 8 M urea, 50 mM Tris-HCl and 5 mM dithiothreitol (DTT) at 37 °C for 1 hour. Then the protein sample was diluted with a pH 7.8 buffer containing 50 mM Tris-HCl and 2 mM $CaCl_2$ until the urea concentration reached 0.5 mM. Trypsin was then added to this solution at a ratio of 1:10 (trypsin: protein). The digestion process was allowed to proceed for 24 hours, and then the trypsin activity was halted by lowering the pH to 4 with acetic acid. The digested protein sample was then passed through a Pierce[®] C-18 spin column (Thermo) and subsequently analyzed by MALDI-MS.

FTIR and CD Measurements. FTIR spectra were collected on a Nicolet 6700 FTIR spectrometer using 1 cm^{-1} resolution and a CaF$_2$ sample cell with a Teflon spacer (52 μm). Sample concentrations were in the range of 1-5 mM. For all reported spectra, a solvent background has been subtracted. The oleic acid-bound BSA-SME samples were prepared using the protocol described by Spector *et al.* (Spector et al., 1969). Specifically, a 1 mM BSA-ME solution (in 50 mM deuterated phosphate buffer, pH 7.0) was first prepared, and then an appropriate aliquot of oleic acid in heptane was added to this solution to reach a final oleic acid concentration of 10 mM. The mixture was slowly shaken continuously and incubated at 37 °C for 24 hours in a water bath. The aqueous layer of this mixture was used for further FTIR measurements. CD spectra were collected on an AVIV 410 spectrometer using a 1 mm quartz cell. Sample concentration was ~40 μM.

Atomic Force Microscopy (AFM) Measurements. AFM experiments were performed at room temperature, using an atomic force microscope (Bruker Dimension Icon AFM, Billerica, MA). To ensure optimal surface coverage, the aggregated peptide sample under investigation was first diluted to ~1 mM (monomer concentration) with deionized water, and then a 20 μl of this diluted peptide sample was deposited onto the mica surface. This sample was allowed to settle on the mica surface for 1 minute before washing with 0.5 ml deionized water and drying under a stream of N$_2$ gas. AFM images were processed and analyzed using Gwyddion, a free program for visualizing and analyzing scanning probe microscopy data.

7.4: RESULTS AND DISCUSSION

Ester Incorporation into Short Peptides. To demonstrate the feasibility of selective incorporation of an ester moiety into a protein via the aforementioned cysteine alkylation reaction, we first tested this method using a short, C-terminally amidated peptide, YGGCGG, where the Tyr residue is introduced to help determine the peptide concentration. As indicated (**Figure 7.2**), the

alkylation reaction involves the use of an α-halocarbonyl electrophile, methyl bromoacetate, and is carried out under mild reaction conditions, which converts the thiol group (-SH) of the Cys residue in YGGCGG to $-SCH_3COOCH_3$ (hereafter referred to as -SME) with a high yield (approximately 90%). The identity of this cysteine alkylation product, YGGC*GG, where C* represents Cys-SME, was verified by mass spectrometry (**Figure 7.3**). This reaction also works well in the presence of 8 M urea, which may be needed in cases where the target cysteine residue is buried in the interior of the protein of interest.

As shown (**Figure 7.4**), comparison between the Fourier transform infrared (FTIR) spectra of YGGCGG and YGGC*GG obtained in D_2O also validates the successful incorporation of the ME group. This is because the YGGC*GG peptide, but not the YGGCGG peptide, gives rise to an IR band within the region of 1700 cm^{-1} to 1750 cm^{-1} (Pazos et al., 2014). As expected, this ester carbonyl stretching vibrational band is broad and peaked at about 1720 cm^{-1}, which is similar to that observed for a short peptide containing a D_M residue, due to hydrogen-bonding interactions with the solvent molecules (Pazos et al., 2014).

In the second test of the cysteine alkylation strategy, we sought to incorporate the ME IR probe into two peptides that have been shown to form amyloid aggregates. The first peptide corresponds to the 9-23 segment of the insulin B chain, which contains one cysteine residue (sequence: SHLVEALYLVCGERG, hereafter referred to as the IB peptide). The study by Ivanova *et al.* (Ivanova et al., 2009) indicated that the LVEALYL segment within the IB peptide can forms fibrils and may play an important role in the fibril formation of the full-length insulin B chain. The second peptide corresponds to a double mutant of $A\beta_{16-22}$ (sequence: KCVK*FAE, hereafter referred to as the Aβ peptide), with a cysteine at the 17th position and a lys-Nvoc (Fmoc-Lys(4,5-dimethoxy-2-nitro-benzyloxycarbonyl)-OH, K*) at the 19 position. Following the study of Measey *et al.* (Measey and Gai, 2012), the lys-Nvoc was used in this case to facilitate amyloid fibril formation. We found that the ME group can be incorporated into both peptides using the reaction conditions described in **Figure 2** and in both cases the yield of the resultant peptide product

(hereafter referred to as IB-SME and Aβ-SME, respectively) was approximately 90% and was verified by mass spectrometry (**Figures 7.5** and **7.6**).

As indicated (**Figure 7.7**), the FTIR spectra of IB-SME and Aβ-SME obtained in deuterated phosphate buffer (50 mM, pH 7.0) confirm that both peptides can aggregate as their amide I′ bands consist of a sharp feature at about 1625 cm^{-1}, due to formation of aggregates composing of tightly packed β-strands (Barth, 2007). In addition, both peptide samples give rise to an IR band characteristic of an ester carbonyl stretching vibration, further validating the successful incorporation of the ME moiety. Interestingly, the ester carbonyl stretching vibrational bands of these two aggregated samples show distinct differences; the one from IB-SME is broad and centered around 1725 cm^{-1}, whereas the one from Aβ-SME consists of two major and narrower peaks centered at 1722 and 1743 cm^{-1}. The study of Pazos *et al.* (Pazos et al., 2014) demonstrated that when the carbonyl group of an ester moiety is engaged in a hydrogen-bonding interaction with a protic solvent, its peak frequency is expected to be lower than 1730 cm^{-1}. In addition, the bandwidth of an IR transition in the condensed phase is typically a manifestation of the degree of heterogeneity of its underlying environment. Therefore, the above IR results suggest that under the current experimental conditions the aggregates formed by IB-SME are not well-structured, leaving a large fraction of the -SME groups exposed to D_2O, whereas those formed by Aβ-SME contain well-organized fibrils, leading to the burial of a large fraction of the -SME groups in a dehydrated environment and, hence, producing a sharp IR peak at 1743 cm^{-1}. To further corroborate these assessments, we also examined the structure and morphology of these aggregates using atomic force microscopy (AFM). Consistent with our expectations, the AFM images (**Figure 7.8**) clearly show that the Aβ-SME peptide forms distinct fibrils, whereas the IB-SME peptide produces mostly globular aggregates under our experimental conditions. Thus, taken together, these results highlight the potential utility of the ester carbonyl stretching vibration as a convenient means to probe the structural integrity of amyloid fibrils formed by peptides and proteins.

Ester Incorporation into Bovine Serum Albumin. To test whether the aforementioned chemical method can be used to incorporate a ME group into large proteins, we carried out the same cysteine alkylation reaction described above on bovine serum albumin (BSA), a 603-residue protein that has 35 cysteine residues but only one free cysteine residue (i.e., Cys34). Experimental results obtained by FTIR spectroscopy (**Figure 7.9**) confirm the incorporation of the ME group (hereafter the ester-labeled BSA is referred to as BSA-SME). In particular, the ratio between the integrated area of the amide I′ band and that of the ester carbonyl stretching band of the ester-labeled BSA is approximately 600:1, indicating that the protein is labeled with only one ME group. Using this information and the molar quantities of BSA (reactant) and BSA-SME (product), we estimated the ME labeling efficiency to be around 90%. Furthermore, since methyl bromoacetate can potentially react with other residues, such as Ser, Lys, Asp, Met, and His, we ruled out this possibility by means of trypsin digestion and mass spectrometry. As shown (**Figure 7.10**), comparison between the masses of trypsin-digested fragments of BSA and BSA-SME corroborates the exclusive addition of the ME group to Cys34. In addition, the circular dichroism (CD) spectra of the wild-type and ME-labeled BSA molecules indicate the reaction conditions used do not cause any significant changes in the structural integrity of the protein (**Figure 7.11**). In addition, we have successfully repeated the ester labeling reaction of BSA under denaturing conditions (i.e., in 8 M urea), further demonstrating the general applicability of this method to target a buried cysteine residue in the interior of a large protein.

Previous studies (Hamilton et al., 1991; Spector et al., 1969) have shown that BSA can interact with various ligands, such as fatty acids. Crystallographic studies showed that upon complexion with a fatty acids, such as oleic acid (Petitpas et al., 2001) or myristate (Petitpas et al., 2001), human serum albumin (HSA), which is very similar to BSA (Steinhardt et al., 1971) in sequence and structure, shows a substantial conformational change which, in particular, makes Cys34 more solvent accessible. In free BSA, Cys34 is located in a cavity approximately 9 Å from the protein surface, thus being protected from bulk water (Hull et al., 1975; Wang et al., 1995).[33,34] Since Cys34 in BSA-SME is labeled with an ester IR probe; we set to use IR

spectroscopy to verify whether BSA undergoes a similar binding-induced conformational change as HSA. As shown (**Figure 7.12**), the ester carbonyl stretching vibrational band of BSA-SME in buffer is peaked at ~1742 cm^{-1}. However, upon addition of oleic acid (OA), which is known to bind to BSA with a binding constant (K_d) of 3.96 μM (Spector et al., 1969), this band is shifted to ~1737 cm^{-1}. This red-shift indicates that interaction with OA causes the -SME group to be more solvent exposed, a result that is consistent with the crystallographic studies (Petitpas et al., 2001).[31,32] However, since the ester carbonyl stretching vibrational band of a completely hydrated -SME group is peaked at a frequency of less than 1730 cm^{-1} (**Figure 7.4**), this result suggests that Cys34 is only partially hydrated even in the OA-bound state of BSA.

7.5: CONCLUSIONS

In summary, we demonstrate a simple, post-translational method to site-specifically incorporate a methyl ester moiety into proteins via cysteine alkylation. Because the carbonyl stretching frequency of this ester exhibits a linear dependence on the local electric field and is sensitive to interactions with water, this method, which requires only mild reaction conditions and can be easily implemented, is expected to be useful in various applications. For example, in the current study, we show that this IR reporter can be used to characterize the structural heterogeneity of peptide aggregates and to site-specifically assess the local hydration status of a protein, as well as changes in this status due to ligand binding.
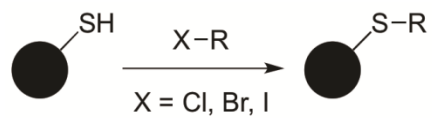
Figure 7.1: Generalized sulfhydryl alkylation reaction with haloalkyl reagents.
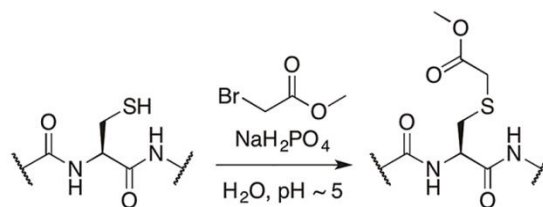
Figure 7.2: Methyl ester incorporation into peptides via cysteine alkylation.
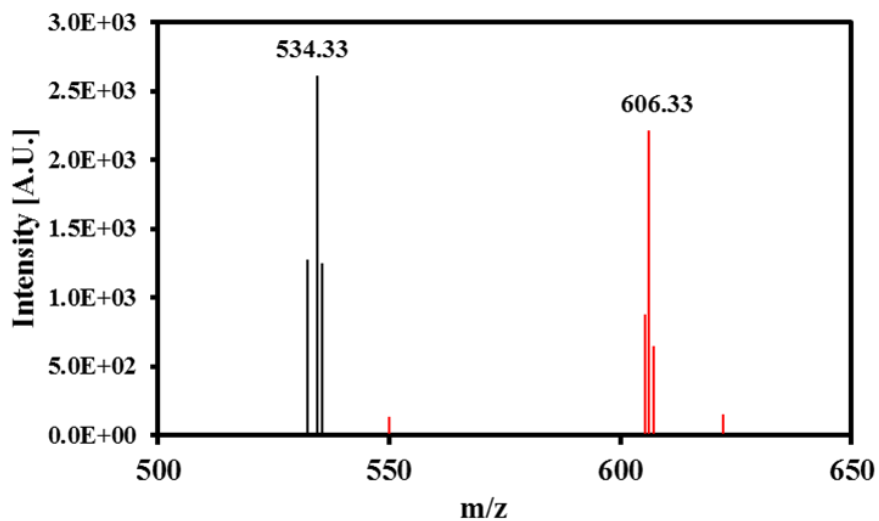
Figure 7.3: MADLI mass spectra for the unlabeled model peptide YGGCGG (black) and labeled peptide YGGC*GG (red). The expected mass for YGGCGG is 511.2 Da, and a value of 534.33 Da was observed, which has an adduct of 23 Da from a sodium molecule. The expected mass for YGGC*GG is 583.2 Da, and a value of 606.33 Da was observed, which has an adduct of 23 Da from a sodium molecule. This increase in 72 Da corresponds to the successful addition of the methyl ester (ME) moiety to YGGCGG.
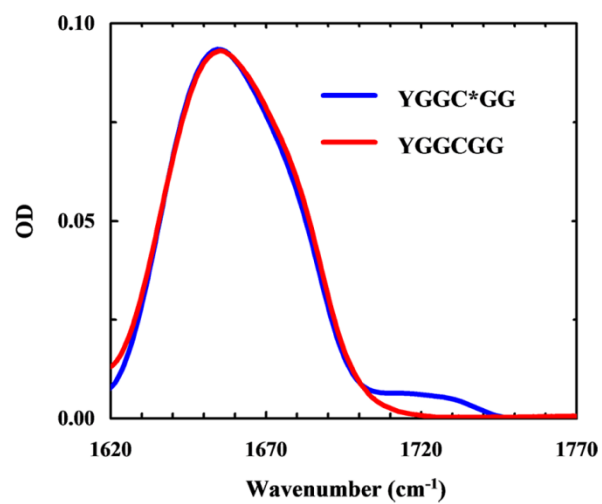
Figure 7.4: FTIR spectra of YGGCGG and YGGC*GG in the amide I′ region, as indicated. The peptide concentration was about 2.5 mM in D_2O.
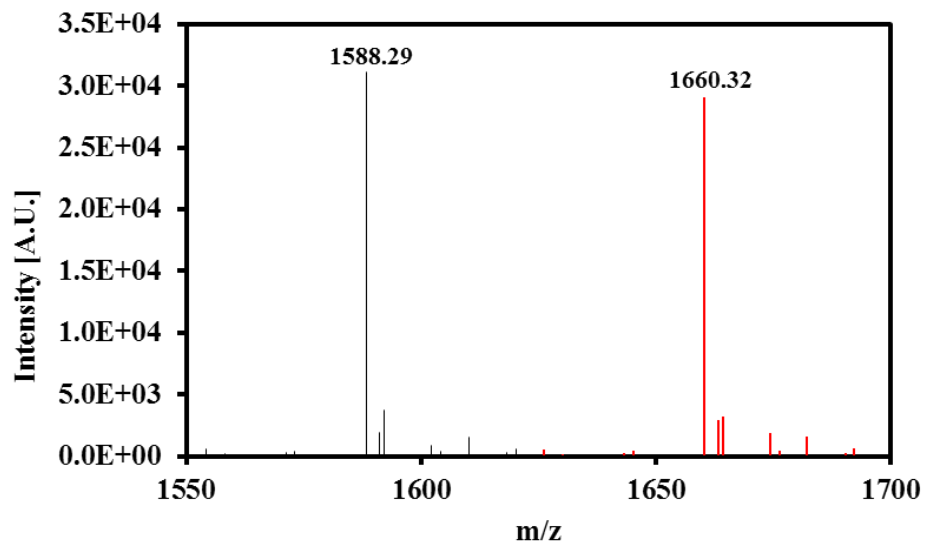
Figure 7.5: MADLI mass spectra for wildtype insulin B (IB) (black) and ME-labeled IB-SME (red). The expected mass for IB is 1588.85 D, and a value of 1588.29 Da was observed. The expected mass for IB-SME is 1660.85 Da and a value of is 1660.32 Da was observed. This increase in 72 Da corresponds to the successful addition of the ME moiety to IB.
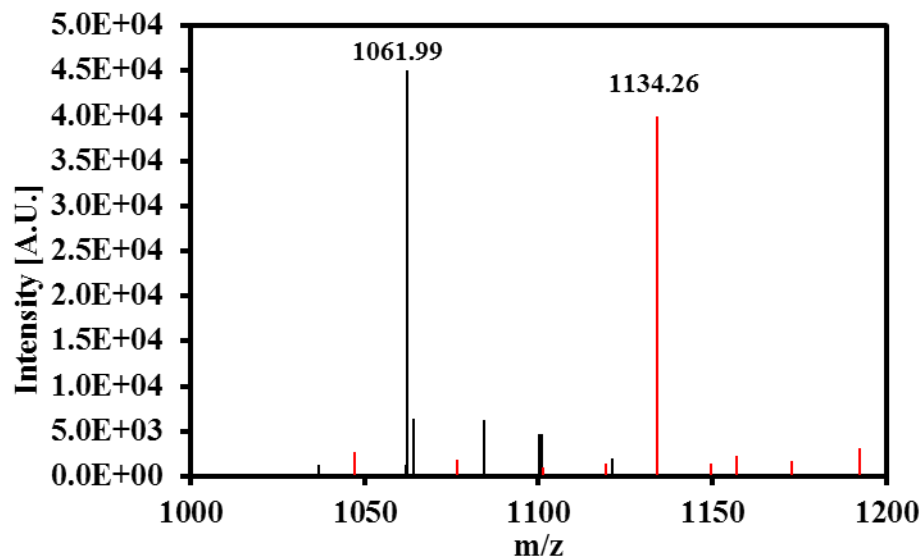
Figure 7.6: MADLI mass spectra for an Aβ peptide (black) and Aβ-SME (red). The expected mass for Aβ is 1062.17 Da, and a value of 1061.99 Da was observed. The expected mass for Aβ-SME is 1134.17 Da, and a value of 1134.26 Da was observed. This increase in 72 Da corresponds to the successful addition of the ME moiety to Aβ.
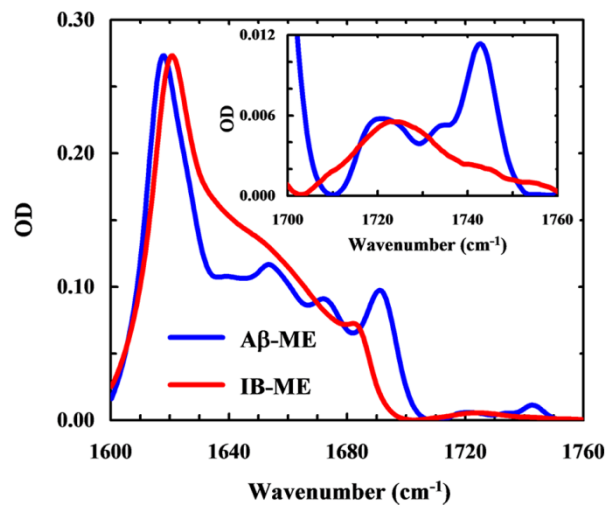
Figure 7.7: FTIR spectra of IB-ME and Aβ-ME in the amide I′ region, as indicated. The peptide concentration was about 4 mM in deuterated phosphate buffer (50 mM, pH 7.0). The spectrum for Aβ-ME has been scaled by a factor of 2.4 for easy comparison. Shown in the inset are the spectra in the frequency region of 1700 – 1760 cm^{-1}.
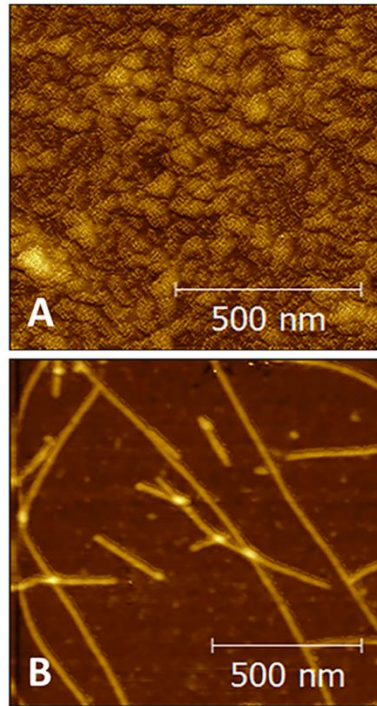
Figure 7.8: AFM images of the aggregates formed by IB-ME (A) and A□-ME (B) peptides. For both cases, the incubation time was 10 days.
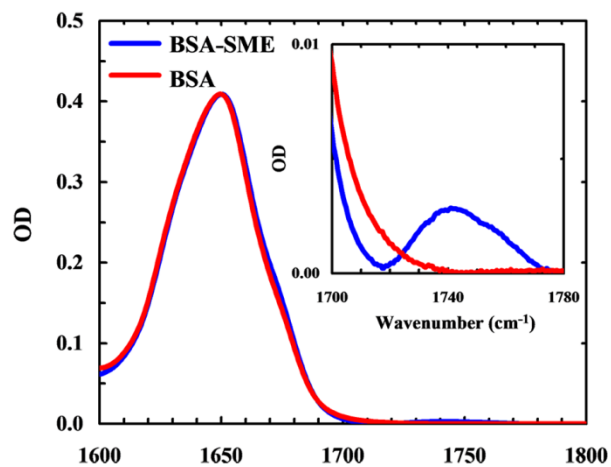
Figure 7.9: FTIR spectra of BSA and BSA-SME in the amide I′ region, as indicated. The protein concentration was about 1 mM in deuterated phosphate buffer (50 mM, pH 7.0). The spectrum of BSA-SME has been normalized against that of BSA for easy comparison. Shown in the inset are the spectra in the frequency region of 1700 – 1780 cm^{-1}.
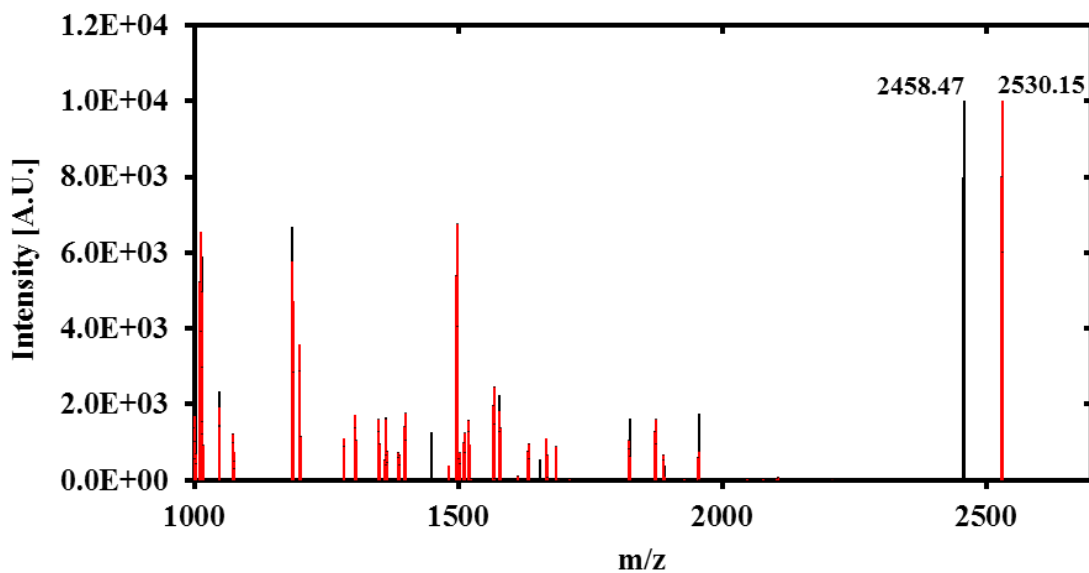
Figure 7.10: MADLI mass spectra for trypsin digested BSA (black) and BSA-SME (red). The expected mass for BSA fragment containing Cys34 is 2435.24 Da, and a value of 2458.47 Da was observed (Na^+ ion adduct). The expected mass for BSA-SME fragment containing Cys34 is 2507.24 Da, and a value of 2530.15 Da was observed (Na^+ ion adduct). This mass increase of 72 Da verifies the successful addition of a ME moiety to BSA at Cys34. There are no additional mass adducts that deviate from the expected masses of BSA fragments, indicating that ME is exclusively added to Cys34.
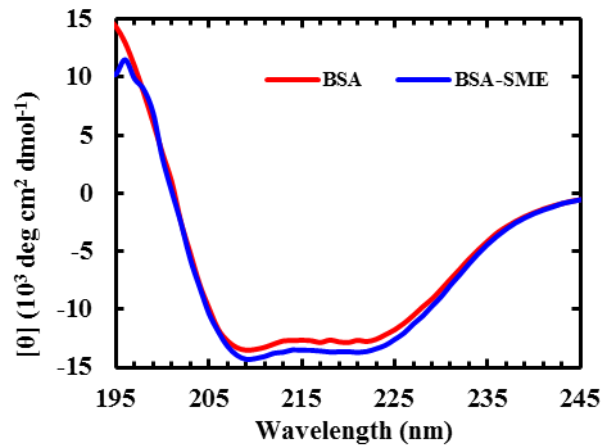
Figure 7.11: Circular Dichroism (CD) spectra of BSA (red) and BSA-SME (blue) which shows no major difference between the secondary structures of wild type and ME-labeled BSA, indicating that the reaction conditions used do not cause any significant changes in the structural integrity of the protein.
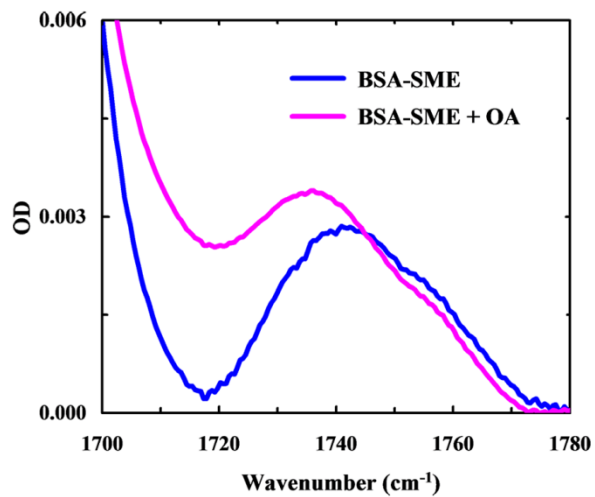
Figure 7.12: FTIR spectra of BSA-SME in the presence and absence of oleic acid (OA), as indicated. The protein concentration was about 1 mM in deuterated phosphate buffer (50 mM, pH 7.0).

8: Summary and Future Directions

Unnatural amino acids have emerged as robust probes for biological spectroscopy and microscopy due to their small size and wide-ranging diversity in chemical and spectroscopic features. These probes can be leveraged to observe proteins in their natural cellular environment via microscopy as well as be used to study the structure, dynamics, and function of proteins in vitro. In practice, unnatural amino acids with distinct spectral properties are used as site-specific probes within proteins for fluorescence and infrared (IR) spectroscopy applications. Characteristics of a useful probe for protein spectroscopy (Ma et al., 2015; Waegele et al., 2011) is one that has (1) large signal (brightness for fluorescence and a large extinction coefficient for IR) so that large amounts of sample is not required for measurements; (2) an absorption and emission that does not overlap with regions devoid of intrinsic protein absorption and emission; (3) is small in size, such that it is minimally perturbing to protein structure; (4) is incorporable into protein systems (chemically or biologically); and is sensitive to changes of a particular property of proteins and its environment, i.e., structure, solvent information, etc. With these features in mind, the goal of this thesis was to expand the experimental toolbox by developing unnatural amino acid-based probes and methods for biological studies.

The development of novel fluorophores designed to study biological processes, both in vitro and in vivo has been at the forefront of biological research (Chatterjee et al., 2013; Cohen, 2002; Hamada et al., 2005; Lepthien et al., 2008; Moroz et al., 2013; Smirnov et al., 1996; Summerer et al., 2006). Such fluorophores are used in cutting-edge fluorescent imaging methods. To date, there are hundreds of fluorophores with a multitude of distinct features. For example, the diversity of fluorescent probes can be witnessed in the Invitrogen Molecular Probes Handbook (available online: https://www.thermofisher.com/us/en/home/references/molecular-probes-the-handbook.html). However, to date, researchers continue to push the boundaries and develop new probes with new properties. In this regard, as discussed in Chapter 2, we sought to

178

develop an amino acid-based fluorophore that was as close in size to a natural amino acid as possible and also bright enough for microscopy applications. Previous studies pointed to the red-shifting of tryptophan absorbance and emission spectra by nitrile-derivatization (Talukder et al., 2015). We surveyed the absorption and emission properties of nitrile substituents at the 2, 3, 4, 5, 6, and 7 positions of indole and honed in on 4-cyanoindole as a lead candidate. We synthesized and characterized the photophysical properties 4-cyanotryptophan (4CN-Trp), a blue fluorescent amino acid. Specifically, 4CN-Trp, exhibits the following unique photophysical properties: it has an absorption spectrum peaked at ca. 325 nm, an emission spectrum peaked at ca. 420 nm, a large fluorescence QY (0.8 – 0.9), a long fluorescence lifetime (13.7 ns), is only a two atom modification of tryptophan, and has good photostability. Combined, these properties make 4CN-Trp an exceedingly useful fluorescent reporter for in vitro and in vivo spectroscopic and microscopic applications. We validate the application of 4CN-Trp for microscopy using standard epifluorescence microscopy and confocal microscopy. Specifically, we image cells undergoing apoptosis after being incubated with a 4CN-Trp labeled antimicrobial peptide.

We believe that the findings in this study are only the first step in exploring the potential utility of 4CN-Trp in biological research; we expect that it can be useful for a myriad of fluorescence-based applications. Especially in the cases where the size of the fluorophore is required to be small as in the case of signaling peptides, neuropeptides, membrane proteins, etc. An obvious extension for 4CN-Trp in imaging applications is using it as a fluorophore for two-photon absorption microscopy (Lippitz et al., 2002). Ideally, two-photon excitation of 4CN-Trp will circumvent the requirement for using cell-damaging UV-light via one-photon excitation. Our recent preliminary results suggest that two-photon excitation of 4CN-Trp is indeed possible because its two-photon cross-section was measured to be 1GM, which is decent for a molecule of its size and brighter than most endogenous molecules. As indicated **in Figure 8.1**, cells were incubated with either a cell penetrating peptide or a nuclear localizing peptide which was both labeled with 4CN-Trp and imaged using two-photon microscopy.

To further extend the utility of 4CN-Trp, in Chapter 3 we developed an improved method for the selective synthesis of L-4CN-Trp (and its Fmoc-protected version) that can be used for incorporation of this blue fluorescence amino acid into polypeptides via solid phase peptide synthesis. Furthermore, we demonstrate that this unnatural amino acid fluorophore is a viable fluorescence reporter to monitor the binding of peptides to cell membranes. Specifically, we devise and validate a new fluorescence method to assess peptide-membrane interactions, which is based on the application of a FRET pair (R_0 = 62 Å) consisting of 4CN-Trp (donor) and a universal membrane stain, DiO (acceptor). We anticipate that we can use this FRET pair to study peptide membrane dynamics using single-molecule measurements like fluorescence correlation spectroscopy (FCS). We envision that this approach in principle lays the foundation for ex vivo mapping of neuropeptide localization. Specifically, DiO can be used to stain all the neurons in a brain slice indiscriminately. Therefore, only the DiO labeled neurons that interact or are in proximity of neuropeptides labeled with 4CN-Trp will fluoresce through FRET excitation. Thus, this approach can be highly useful for studying neuropeptide dynamics and transport when coupled with techniques like opto/chemogenetics or electrophysiology.

In Chapter 4, we further extend the utility of 4CN-Trp as part of a dual FRET and PET pair. We demonstrate that endogenous tryptophan can serve as either a FRET donor or a PET quencher to 4CN-Trp. Specifically, we show that, depending on the excitation wavelength, tryptophan and 4-cyanotrptophan can interact with each other via the mechanism of either energy (λ_{ex}=270 nm) or electron transfer (λ_{ex}= 330 – 360 nm). As a proof of principle, we demonstrated the biological utility of this amino acid pair by applying it to study (1) the end-to-end collision rate of a short peptide, (2) the mode of interaction between a ligand and BSA, and (3) protease activity. Given the unique photophysical property of 4CN-Trp and the less-perturbating nature of this unnatural amino acid and tryptophan, we believe that this pair will find valuable applications in biochemistry and biophysics, especially in cases where using dye-based PET or FRET reporters is deemed inappropriate. We are particularly excited about the possibility of using this pair of amino acids in single-molecule fluorescence studies, such as PET-FCS (Doose et al.,

2009; Li et al., 2006; Lin et al., 2017), as well as using it to study the functional and/or conformational dynamics of Trp-rich proteins (Liu et al., 2014).

In addition to being a fluorescence reporter for protein-based studies, in Chapter 5 we adopt 4-cyanoindole, the fluorophore of 4CN-Trp, into an unnatural nucleoside for DNA/RNA based studies. We found the unnatural nucleoside, 4-cyanoindole-2′-deoxyribonucleoside (4CNI-NS), as a site-specific spectroscopic probe of DNA/RNA. A recent study revealed that 4CNI-NS is a universal nucleobase that maintains the high fluorescence quantum yield of 4-cyanoindole (Passow and Harki, 2018) and that among the four natural nucleobases, only guanine can significantly quench its fluorescence. We further show that similar to 4CN-Trp (van Wilderen, 2018), the C≡N stretching frequency of 4CNI-NS is sensitive to the local environment, making it a useful site-specific infrared probe of oligonucleotides. Also, we demonstrated that the fluorescence-quencher pair formed by 4CNI-NS and guanine could be used to quantitatively assess the binding affinity of a single-stranded DNA to the protein system of interest via fluorescence spectroscopy, among other applications. We believe that this fluorescence binding assay is especially useful as its potentiality allows high-throughput screening of DNA–protein interactions. Since 4CNI-NS and guanine constitute a useful fluorophore-quencher pair, we seek to use it to study the structural changes and dynamics of RNA structure and DNA superstructures. Moreover, the 4CNI-NS and guanine pair can be used strategically in DNA aptamer-based biosensor designed for specific analytes such as metals ions (Zhou et al., 2017) and neurotransmitters (Si, 2018). Binding of the analyte to the sensor will induce a conformational change in the DNA structure to either separate or bring together 4CNI-NS and guanine to induce an increase or decrease in fluorescence intensity respectively.

While 4CN-Trp has proved to be a remarkable and versatile unnatural amino acid-based fluorophore, further experimental and theoretical studies are needed to understand why 4-cyanoindole (and 4CN-Trp) possesses unique photophysical properties in comparison to indole (tryptophan) and its derivatives. To begin to tackle this question, Hilaire *et al.* (Hilaire et al., 2017a)surveyed the fluorescence decay kinetics as well as the absorption and emission spectra

of six cyanoindoles in different solvents. In particular, they found, among other results, that only 4-cyanoindole affords a relatively long fluorescence lifetime and consequently high quantum yield in most solvents. These results indicated that the 4 position on the indole sidechain uniquely enhances the fluorescent properties of the indole. To delve deeper into what properties dictate this, we surveyed the absorption and emission spectra of indole derivatives with various substituents at the carbon 4 position of the indole (**Figure 8.2**). We found that the absorption and emission of the substituted indoles of each have unique photophysical properties (**Figure 8.3**) and correlates positively with both the Hammett parameters (Abou-Hatab et al., 2017) and electrophilicity index (Parr et al., 1999) (**Figures 8.4**) of the substituent. These findings will provide not only a guide for choosing which 4-substituted tryptophan to use in practice but also data for computational modeling of the substitution effect at the 4-position on the electronic transitions of indole.

In our pursuit to develop the best possible tryptophan-based fluorophore, we search for candidates that have further red-shifted absorption and emission spectra with respect to 4CN-Trp. From our survey of substituted indoles, there are two stand-out candidates. Both 4-aldahydeindole and 4-nitroindole have absorption spectra that are significantly red-shifted with absorption λ_{max} of 360 and 375 nm and fluorescence λ_{max} 360 and 524 nm respectively. In our previous study (Chapter 2), we have observed when adding the amino acid backbone to the 3 position of 4-cyanoindole causes its absorbance and emission to red-shift by ca. 10 nm each. This is likely due to increasing the electron density on the pyrrole ring. We expect this to be the same when making the amino acid version of 4-aldahydeindole and 4-nitroindole.

In addition to red-shifted absorption and emission spectra, our preliminary results show that 4-aldahydeindole is slightly brighter than 4CN-Trp with a fluorescence QY is ca. 0.92. Therefore, it can be selectively excited in a spectral region which is separated from other intrinsic fluorescent amino acids (Trp, Tyr, Phe) as a cyan (blue/green) amino acid. This is exciting because while it is only slightly larger than 4-cyanoindole, its absorbance tails into the blue region of the spectrum enabling it to be more compatible with laser excitation commonly available in

microscope setups such as wide-field, multi-photon, and confocal fluorescence microscopy. Additionally, being that it is excitable by blue light, it reaps the benefit of avoiding UV light excitation which is generally harmful to cells. Taken together, these results indicate 4-aldahydeindole (4-formyltryptophan) is a promising lead small for a cyan fluorescent amino acid. To establish this, the synthesis of the amino acid and photophysical characterization thereof is forthcoming.

While 4-nitroindole's absorbance and fluorescence spectra are red-shifted the farthest, its QY is negligible. Therefore, as a first pass, it is unsuitable to be employed as a visibly fluorescent amino acid. This is likely due to non-radiative pathways of nitro-containing fluorophores such as excited state rotation or intramolecular charge transfer interaction with the indole nitrogen. In spite of this, due to its negligible QY, 4-nitroindole can theoretically be used as a quencher of blue fluorescent dyes (such as 4CN-Trp and Alexa350) through the FRET mechanism (**Figure 8.5**). This is advantageous because there are various fluorescent acceptors for blue fluorescent dyes (donors), but there are not many options as FRET-based fluorescent quenchers. Most blue dye fluorescence quenchers utilize the photoinduced electron transfer mechanism which requires short distance van der Waals contact. A FRET-based fluorescent quencher would allow for longer distance fluorescence quenching in scenarios where you would want to turn fluorescence on and off rather than having a change in color. For example, the R_0 for the 4CN-Trp-4-nitroindole and Alexa350-4-nitroindole pairs were calculated to be 53 and 48 Å respectively. We anticipate that this FRET-based blue dye fluorescence quencher can be used in many fluorescence spectroscopy and microscopy applications.

The previous chapters (2-5) we focused on the development and utility of 4CN-Trp as a versatile blue fluorescent amino acid which will no doubt be useful for many biological spectroscopy and microscopy applications. In the later chapters, we focus on extending the applications of two previously known unnatural amino acids. In Chapter 6 we focus on expanding the utility of *p*-cyanophenylalanine (Phe$_{CN}$), an unnatural amino acid that has unique IR and fluorescent properties (Mintzer et al., 2015; Serrano et al., 2012; Tucker et al., 2006). As such, it

has been used as both IR and fluorescence probes in a wide range of applications. We show that for peptides having a Phe$_{CN}$ residue at or near the N-terminus, the neutral form of the N-terminal amino group is much more effective than the protonated form in quenching the Phe$_{CN}$ fluorescence. As a result, such peptides can be used as pH sensors. By examining the pH-dependence of the Phe$_{CN}$ fluorescence of nine tripeptides (i.e., X-Phe$_{CN}$-Gly, where X represents either Asn, Met, Lys, Pro, Val, Asp, Ala, Gly or Ser), we are able to show that the dynamic range of such pH sensors can be tuned to cover several units of pH as the N-terminal pK_a of these peptides depends on the identity of the N-terminal amino acid. Also, using the cell penetrating peptide, TAT as an example, we demonstrated that the pH-dependence of Phe$_{CN}$ fluorescence could be exploited to measure penetration kinetics of cell-penetrating peptides across model membranes.

The second unnatural amino acid exploits an ester carbonyl stretching vibration which has recently been shown to be a sensitive and convenient infrared (IR) probe of protein electrostatics due to the linear dependence of its frequency on local electric field (Pazos et al., 2014a). While an ester moiety can be easily incorporated into peptides via solid-phase synthesis, currently there is no method available to site-specifically incorporate it into large proteins. Finally, in Chapter 7, we show that it is possible to use a cysteine alkylation reaction to achieve this goal and demonstrate the feasibility of this simple method by successfully incorporating a methyl ester group (-CH$_2$COOCH$_3$) into a model peptide, two amyloid-forming peptides derived from the insulin B chain and Aβ, and protein bovine serum albumin (BSA). We then carried out IR measurements with those peptide and protein systems to further confirm the utility of this vibrational probe in monitoring, for example, the structural integrity of amyloid fibrils and ligand binding-induced changes in protein local hydration status.

Building upon the method we used for installing a methyl ester moiety onto cysteine; we seek to incorporate 4CN-Trp and other Trp analog sidechains site-specifically into proteins. The first strategy to accomplish this is by cysteine alkylation. The second strategy would be via

184

posttranslational mutagenesis, a two-step reaction method devised by the Davis lab (Wright and Davis, 2017; Wright et al., 2016) which works at ambient conditions. The first step requires desulfurization of the cysteine to dehydroalanine which is then further reacted with a haloalkyl substrate containing the functional group of choice through free radical chemistry to form a natural amino acid, an unnatural amino acid or posttranslationally modified residue through a newly formed $C(sp^3) - C(sp^3)$ bond this is illustrated in **Figure 8.6** (Wright and Davis, 2017; Wright et al., 2016). The key advantage of this method is the desulfurization of the cysteine for the cases of the presence of a sulfur molecule can be undesirable such as side reactions or size of the sidechain.

The last strategy for incorporating unnatural amino acids into large proteins is via a cell-free biosynthetic method that involves chemically charging suppressor tRNAs with the amino acids of choice (Kuruma and Ueda, 2015; Noren et al., 1989). This circumvents the need for evolving tRNA synthase for unnatural amino acid genetic incorporation. This method has previously been used to incorporate 6 and 7-cyanotryptophan into proteins (Talukder et al., 2015a). To incorporate 4CN-Trp and other Trp analogs by this method, first, 4CN-Trp is coupled to a pdCpA dinucleotide in a multistep process (**Figure 8.7**). Then amber suppressor tRNAs are aminoacylated with the dinucleotide containing 4CN-Trp. After that, the misacylated suppressor tRNA is added to a mixture of the constituents required for translation, including DNA with the amber codon mutation at the desired position for 4CN-Trp incorporation. Together, these methods will expand the utility of unnatural Trp analogs in large proteins for various studies. To harness the full potential of 4CN-Trp, genetic incorporation of this unnatural amino acid is required. However, this is beyond the scope of our expertise, and therefore we hope others in the field would achieve this (Liu et al., 1997; Summerer et al., 2006).

A. **B.** **C.**

Buffer **4CNI-TAT** **4CNI-TAT-NLS**
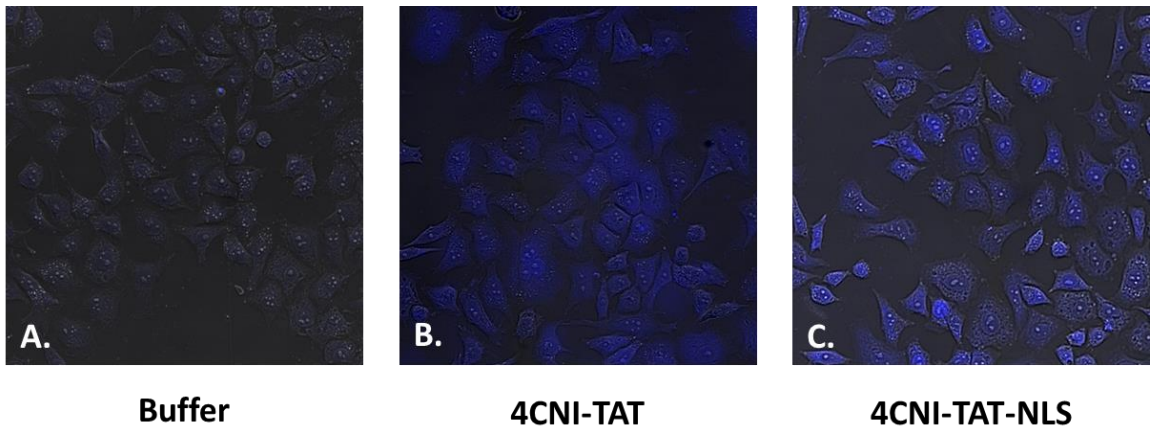
Figure 8.1: Two-photon images of HeLa cells incubated with A. PBS buffer, B. the peptide 4CNI-TAT for cytoplasm localization and C. the peptide 4CNI-TAT-NLS for nuclear localization.
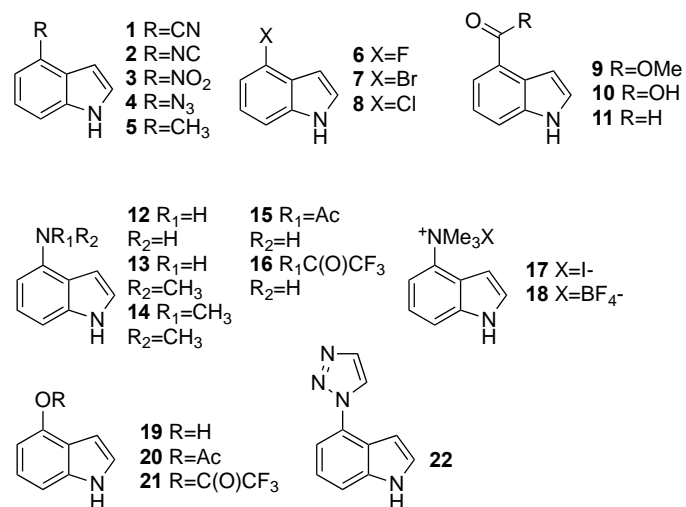
Figure 8.2: Structure of all 4-substituted indoles surveyed.
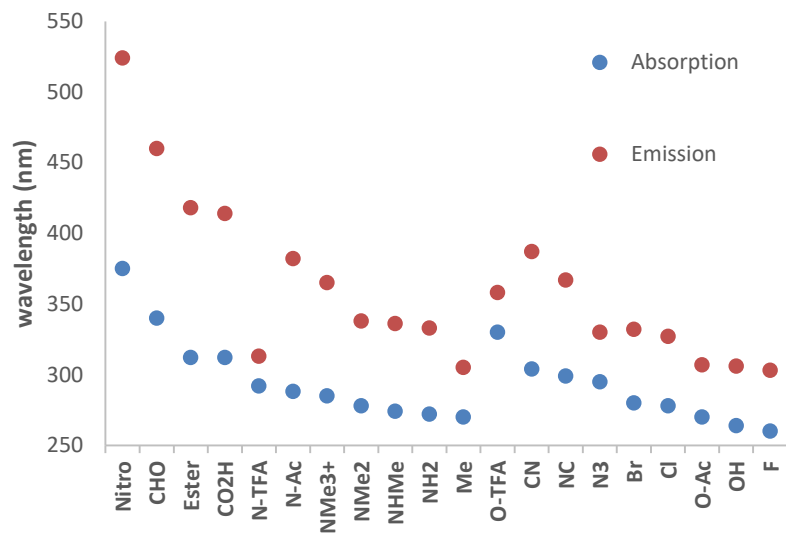
Figure 8.3: Absorption and emission max of the wavelength of all 4-Substituted indoles surveyed.
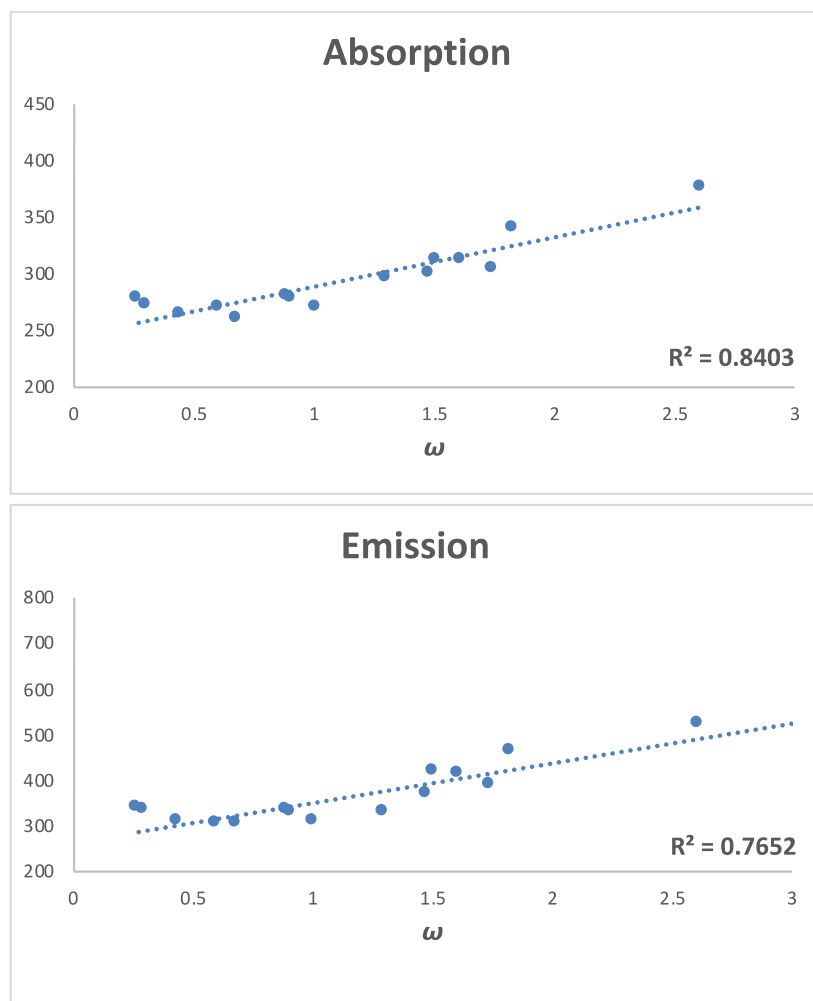
Figure 8.4: Correlation of absorption and emission max of the wavelength of all 4-Substituted indoles surveyed with the electrophilicity index (ω) of each substituent.
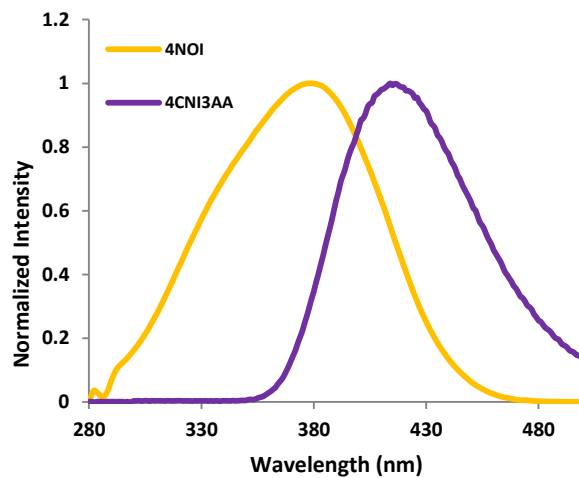
Figure 8.5: Normalized fluorescence spectrum of 4CNI3AA (purple) and absorption spectrum of 4-nitroindole (yellow) in methanol. R_0 of this FRET pair is 53 Å.
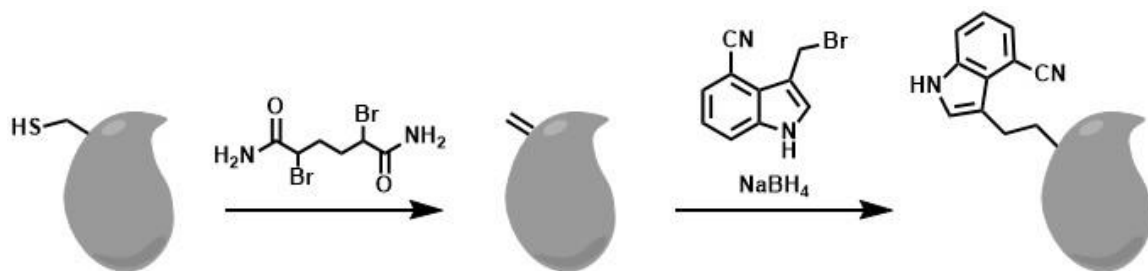
Figure 8.6: Site specific post-translational modification of cysteine to install a 4-cyanoindole functional group (and other indole derivatives) into proteins. This method is adopted from Davis and co-workers as a useful approach to generate unnatural amino acids and PTMs in proteins via cysteine. The first reaction de-sulfurizes cysteine into dehydroalanine using 2,5 dibromohexanediamide in pH 8 phosphate or sodium cacodylate buffer (urea can be used to unfold protein to access buried sites). After that, the reagents are purified out using a de-salting column. Lastly, dehydroalanine is treated with sodium borohydride and the alkyl halide of choice, 3-(bromomethyl)-1H-indole-4-carbonitrile to generate the 4CN-Trp unnatural amino acid at the site of a cysteine residue. This method is advantageous over cysteine alkylation due to the removal of the sulfur group, which produces closer to the natural amino acid structure.
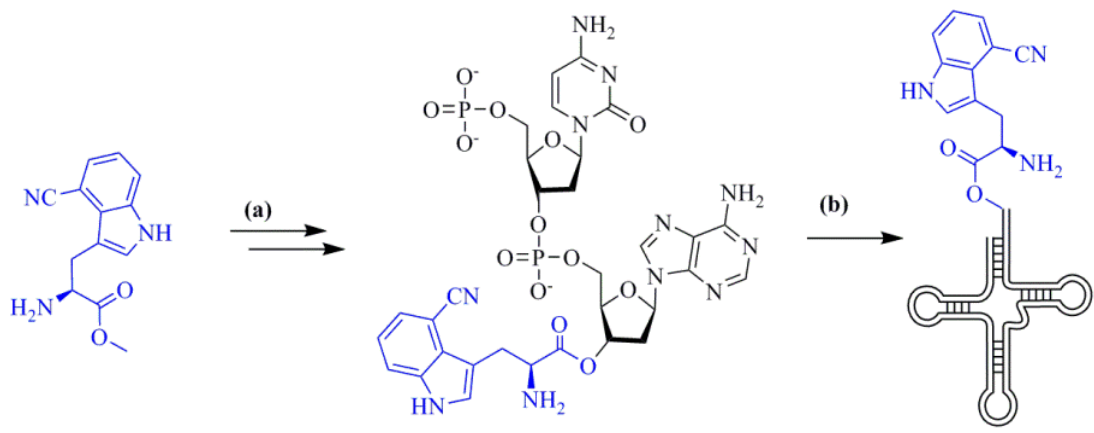
Figure 8.7: Cell-free expression method to incorporate unnatural amino acids such as 4CN-Trp into large proteins.

BIBLIOGRAPHY

Abou-Hatab, S., Spata, V.A., and Matsika, S. (2017). Substituent effects on the absorption and fluorescence properties of anthracene. J. Phys. Chem. A *121*, 1213–1222.

Acharya, K.R., and Lloyd, M.D. (2005). The advantages and limitations of protein crystal structures. Trends Pharmacol. Sci. *26*, 10–14.

Adams, P.D., Chen, Y., Ma, K., Zagorski, M.G., Sönnichsen, F.D., McLaughlin, M.L., and Barkley, M.D. (2002). Intramolecular quenching of tryptophan fluorescence by the peptide bond in cyclic hexapeptides. J. Am. Chem. Soc. *124*, 9278–9286.

Adhikary, R., Zimmermann, J., and Romesberg, F.E. (2017). Transparent window vibrational probes for the characterization of proteins with high structural and temporal resolution. Chem. Rev. *117*, 1927–1969.

Ahmed, I.A., and Gai, F. (2017). Simple method to introduce an ester infrared probe into proteins. Protein Sci. *26*, 375–381.

Ahmed, I.A., Acharyya, A., Eng, C.M., Rodgers, J.M., DeGrado, W.F., Jo, H., and Gai, F. (2019). 4-Cyanoindole-2′-deoxyribonucleoside as a dual fluorescence and infrared probe of DNA structure and dynamics. Molecules *24*, 602.

Andreev, O.A., Dupuy, A.D., Segala, M., Sandugu, S., Serra, D.A., Chichester, C.O., Engelman, D.M., and Reshetnyak, Y.K. (2007). Mechanism and uses of a membrane peptide that targets tumors and other acidic tissues in vivo. Proc. Natl. Acad. Sci. *104*, 7893–7898.

Aprilakis, K.N., Taskent, H., and Raleigh, D.P. (2007). Use of the novel fluorescent amino acid *p* -cyanophenylalanine offers a direct probe of hydrophobic core formation during the folding of the N-terminal domain of the ribosomal protein l9 and provides evidence for two-state folding. Biochemistry *46*, 12308–12313.

Aurora, T.S., Li, W., Cummins, H.Z., and Haines, T.H. (1985). Preparation and characterization of monodisperse unilamellar phospholipid vesicles with selected diameters of from 300 to 600 nm. Biochim. Biophys. Acta - Biomembr. *820*, 250–258.

Bagchi, S., Fried, S.D., and Boxer, S.G. (2012). A solvatochromic model calibrates nitriles' vibrational frequencies to electrostatic fields. J. Am. Chem. Soc. *134*, 10373–10376.

Barth, A. (2007). Infrared spectroscopy of proteins. Biochim. Biophys. Acta - Bioenerg. *1767*, 1073–1101.

Bartoccini, F., Bartolucci, S., Mari, M., and Piersanti, G. (2016). A simple, modular synthesis of C4-substituted tryptophan derivatives. Org. Biomol. Chem. *14*, 10095–10100.

Basom, E.J., Maj, M., Cho, M., and Thielges, M.C. (2016). Site-specific characterization of cytochrome p450cam conformations by infrared spectroscopy. Anal. Chem. *88*, 6598–6606.

Bazewicz, C.G., Liskov, M.T., Hines, K.J., and Brewer, S.H. (2013). Sensitive, site-specific, and stable vibrational probe of local protein environments: 4-Azidomethyl- l -phenylalanine. J. Phys. Chem. B *117*, 8987–8993.

Berezin, M.Y., and Achilefu, S. (2010). Fluorescence lifetime measurements and biological imaging. Chem. Rev. *110*, 2641–2684.

Bertuzzi, A., Mingrone, G., Gandolfi, A., Greco, A. V, Ringoir, S., and Vanholder, R. (1997). Binding of indole-3-acetic acid to human serum albumin and competition with L-tryptophan. Clin. Chim. Acta. *265*, 183–192.

Best, R.B., Zhu, X., Shim, J., Lopes, P.E.M., Mittal, J., Feig, M., and MacKerell, A.D. (2012). Optimization of the additive charmm all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ$_1$ and χ$_2$ dihedral angles. J. Chem. Theory Comput. *8*, 3257–3273.

Blaser, G., Sanderson, J.M., Batsanov, A.S., and Howard, J.A.K. (2008). The facile synthesis of a series of tryptophan derivatives. Tetrahedron Lett. *49*, 2795–2798.

Błasiak, B., Londergan, C.H., Webb, L.J., and Cho, M. (2017). Vibrational probes: From small molecule solvatochromism theory and experiments to applications in complex systems. Acc. Chem. Res. *50*, 968–976.

Boville, C.E., Romney, D.K., Almhjell, P.J., Sieben, M., and Arnold, F.H. (2018). Improved synthesis of 4-cyanotryptophan and other tryptophan analogues in aqueous solvent using variants of trpb from *Thermotoga maritima*. J. Org. Chem. *83*, 7447–7452.

Boxer, S.G. (2009). Stark Realities. J. Phys. Chem. B *113*, 2972–2983.

Brewer, S.H.,Song, B., Raleigh, D.P., and Dyer, R.B. (2007). Residue specific resolution of protein folding dynamics using isotope-edited infrared temperature jump spectroscopy. Biochemistry 46, 3279–3285.

Brown, L.R., Lauterwein, J., and Wüthrich, K. (1980). High-resolution ^{1}H-NMR studies of self-aggregation of melittin in aqueous solution. Biochim. Biophys. Acta - Protein Struct. *622*, 231–244.

Burton, D.R. (2002). Antibodies, viruses and vaccines. Nat. Rev. Immunol. *2*, 706–713.

Callis, P.R. (2014). Binding phenomena and fluorescence quenching. II: Photophysics of aromatic residues and dependence of fluorescence spectra on protein conformation. J. Mol. Struct. *1077*, 22–29.

Cattan, D., Bourqoin, D., and Joly, M. (1969). Change of conformation of DNA by association with proteins. European J. Biochem *8,* 541–546.

Chakrabartty, A., Doig, A.J., and Baldwin, R.L. (1993). Helix capping propensities in peptides parallel those in proteins. Proc. Natl. Acad. Sci. U. S. A. *90*, 11332–11336.

Chalker, J.M., Bernardes, G.J.L., Lin, Y.A., and Davis, B.G. (2009). Chemical modification of proteins at cysteine: Opportunities in chemistry and biology. Chem. -An Asian J. *4*, 630–640.

Chatterjee, A., Guo, J., Lee, H.S., and Schultz, P.G. (2013a). A genetically encoded fluorescent probe in mammalian cells. J. Am. Chem. Soc. *135*, 12540–12543.

Chatterjee, A., Sun, S.B., Furman, J.L., Xiao, H., and Schultz, P.G. (2013b). A versatile platform for single- and multiple-unnatural amino acid mutagenesis in *Escherichia coli*. Biochemistry *52*, 1828–1837.

Chattopadhyay, A., and Boxer, S.G. (1995). Vibrational stark effect spectroscopy. J. Am. Chem. Soc. *117*, 1449–1450.

Chen, Y., and Barkley, M.D. (1998). Toward understanding tryptophan fluorescence in proteins.

28, 9976–9982.

Chenault, H.K., Dahmer, J., and Whitesides, G.M. (1989). Kinetic resolution of unnatural and rarely occurring amino acids: enantioselective hydrolysis of N-acyl amino acids catalyzed by acylase I. J. Am. Chem. Soc. *111*, 6354–6364.

Cho, M. (2008). Coherent Two-Dimensional Optical Spectroscopy. Chem. Rev. *108*, 1331–1418.

Choi, J.-H., and Cho, M. (2011). Vibrational solvatochromism and electrochromism of infrared probe molecules containing C≡O, C≡N, C=O, or C−F vibrational chromophore. J. Chem. Phys. *134*, 154513.

Chudakov, D.M., Matz, M. V., Lukyanov, S., and Lukyanov, K.A. (2010). Fluorescent proteins and their applications in imaging living cells and tissues. Physiol. Rev. *90*, 1103–1163.

Ciobanasu, C., Siebrasse, J.P., and Kubitscheck, U. (2010). Cell-penetrating HIV1 TAT peptides can generate pores in model membranes. Biophys. J. *99*, 153.

Clark, P.I., and Lowe, G. (1977). Chemical mutations of papain. The preparation of ser 25- and gly 25-papain. J. Chem. Soc. Chem. Commun. *0*, 923.

Clarkson, J., and Ewen Smith, W. (2003). A DFT analysis of the vibrational spectra of nitrobenzene. J. Mol. Struct. *655*, 413–422.

Clegg, R. M. Fluorescence resonance energy transfer. (1996) In: Wang, X. F., Herman, B. editors. Fluorescence imaging spectroscopy and microscopy. John Wiley & Sons Inc.; New York, USA, 179–251.

Cohen, D.T., and Buchwald, S.L. (2015). Mild palladium-catalyzed cyanation of (hetero)aryl halides and triflates in aqueous media Org. Lett. *17*, 202–205.

Cohen, B.E., McAnaney, T.B., Park, E.S., Jan, Y.N., Boxer, S.G., and Jan, L.Y. (2002). Probing protein electrostatics with a synthetic fluorescent amino acid. Science *296*, 1700–1703.

Combs, C.A. (2010). Fluorescence microscopy: a concise guide to current imaging methods. Curr. Protoc. Neurosci. *Chapter 2*, Unit2.1.

Crankshaw, M.W., and Grant, G.A. (1996). Modification of cysteine. In Current Protocols in Protein Science, (Hoboken, NJ, USA: John Wiley & Sons, Inc.).

Crich, D. and Banerjee, A. (2007). Native chemical ligation at phenylalanine. J. Am. Chem. Soc. *129*, 10064–10065.

Dawson, P.E., Muir, T.W., Clark-Lewis, I., and Kent, S.B. (1994). Synthesis of proteins by native chemical ligation. Science *266*, 776–779.

Day, R.N., and Davidson, M.W. (2009). The fluorescent protein palette: tools for cellular imaging. Chem. Soc. Rev. *38*, 2887–2921.

Ding, B., Hilaire, M.R., and Gai, F. (2016). Infrared and fluorescence assessment of protein dynamics: from folding to function. J. Phys. Chem. B *120*, 5103–5113.

Doig, A.J., and Baldwin, R.L. (1995). N- and C-capping preferences for all 20 amino acids in α-helical peptides. Protein Sci. *4*, 1325–1336.

Doose, S., Neuweiler, H., and Sauer, M. (2009). Fluorescence quenching by photoinduced

electron transfer: a reporter for conformational dynamics of macromolecules. ChemPhysChem *10*, 1389–1398.

Du, D., Liu, H., and Ojha, B. (2013). Study protein folding and aggregation using nonnatural amino acid *p*-cyanophenylalanine as a sensitive optical probe. Methods Mol Biol. *1081*, 77–89.

Dumas, A., Lercher, L., Spicer, C.D., and Davis, B.G. (2015). Designing logical codon reassignment – Expanding the chemistry in biology. Chem. Sci. *6*, 50–69.

Duthaler, R.O. (1994). Recent developments in the stereoselective synthesis of α-aminoacids. Tetrahedron *50*, 1539–1650.

Englander, S.W., Mayne, L., and Krishna, M.M.G. (2007). Protein folding and misfolding: mechanism and principles. Q. Rev. Biophys. *40*, 1–41.

Fafarman, A.T., and Boxer, S.G. (2010). Nitrile bonds as infrared probes of electrostatics in ribonuclease S. J. Phys. Chem. B *114*, 13536–13544.

Fafarman, A.T., Webb, L.J., Chuang, J.I., and Boxer, S.G. (2006). Site-specific conversion of cysteine thiols into thiocyanate creates an IR probe for electric fields in proteins. J. Am. Chem. Soc. *128*, 13356–13357.

Fauvet, B., Butterfield, S.M., Fuks, J., Brik, A., and Lashuel, H.A. (2013). One-pot total chemical synthesis of human α-synuclein. Chem. Commun. *49*, 9254.

Feigenson, G.W. (2009). Phase diagrams and lipid domains in multicomponent lipid bilayer mixtures. Biochim. Biophys. Acta - Biomembr. *1788*, 47–52.

Fierz, B., Satzger, H., Root, C., Gilch, P., Zinth, W., and Kiefhaber, T. (2007). Loop formation in unfolded polypeptide chains on the picoseconds to microseconds time scale. Proc. Natl. Acad. Sci. U. S. A. *104*, 2163–2168.

Förster, T. (1948) Intermolecule energy migration and fluorescence. Annalen der Physik 2, 55–75.;English translation in Mielczarek, E. V., Greenbaum, E., Knox R. S., editors. Biological Physics. American Institute of Physics; New York, USA, 1993, 148–160.

Förster, T. (1965) Delocalized excitation and excitation transfer. In: Sinanoglu, O. editor. Modern Quantum Chemistry. Academic Press Inc.; New York, USA, 93–137.

Francis, D., Winn, M., Latham, J., Greaney, M.F., and Micklefield, J. (2017). An engineered tryptophan synthase opens new enzymatic pathways to β-methyltryptophan and derivatives. ChemBioChem *18*, 382–386.

Garcia-Mira, M.M. (2002). Experimental identification of downhill protein folding. Science *298*, 2191–2195.

Getahun, Z., Huang, C.Y., Wang, T., De León, B., DeGrado, W.F., and Gai, F. (2003). Using nitrile-derivatized amino acids as infrared probes of local environment. J. Am. Chem. Soc. *125*, 405–411.

Ghosh, A., Qiu, J., DeGrado, W.F., and Hochstrasser, R.M. (2011). Tidal surge in the M2 proton channel, sensed by 2D IR spectroscopy. Proc. Natl. Acad. Sci. U. S. A. *108*, 6115–6120.

Ghosh, A., Ostrander, J.S., and Zanni, M.T. (2017). Watching proteins wiggle: mapping structures with two-dimensional infrared spectroscopy. Chem. Rev. *117*, 10726–10759.

Giuliano, K.A., and Taylor, D.L. (1998). Fluorescent-protein biosensors: New tools for drug discovery. Trends Biotechnol. *16*, 135–140.

Glasscock, J.M., Zhu, Y., Chowdhury, P., Tang, J., and Gai, F. (2008). Using an amino acid fluorescence resonance energy transfer pair to probe protein unfolding: Application to the villin headpiece subdomain and the lysm domain. Biochemistry *47*, 11070–11076.

Goldberg, J.M., Batjargal, S., and Petersson, E.J. (2010). Thioamides as fluorescence quenching probes: Minimalist chromophores to monitor protein dynamics. J. Am. Chem. Soc. *132*, 14718–14720.

Goldberg, J.M., Speight, L.C., Fegley, M.W., and Petersson, E.J. (2012). Minimalist probes for studying protein dynamics: Thioamide quenching of selectively excitable fluorescent amino acids. J. Am. Chem. Soc. *134*, 6088–6091.

Gosavi, P.M., and Korendovych, I. V (2016). Minimalist IR and fluorescence probes of protein function. Curr. Opin. Chem. Biol. *34*, 103–109.

Goto, Y., and Hagihara, Y. (1992). Mechanism of the conformational transition of melittin. Biochemistry *31*, 732–738.

Greenfield, N.J. (2006). Using circular dichroism spectra to estimate protein secondary structure. Nat. Protoc. *1*, 2876–2890.

Gunnoo, S.B., and Madder, A. (2016). Chemical protein modification through cysteine. ChemBioChem *17*, 529–553.

Haase, C., Rohde, H., and Seitz, O. (2008). Native chemical ligation at valine. Angew. Chemie Int. Ed. *47*, 6807–6810.

Haj-Yahya, M., Fauvet, B., Herman-Bachinsky, Y., Hejjaoui, M., Bavikar, S.N., Karthikeyan, S. V., Ciechanover, A., Lashuel, H.A., and Brik, A. (2013). Synthetic polyubiquitinated α-Synuclein reveals important insights into the roles of the ubiquitin chain in regulating its pathophysiology. Proc. Natl. Acad. Sci. *110*, 17726–17731.

Hamada, H., Kameshima, N., Szymańska, A., Wegner, K., Łankiewicz, L., Shinohara, H., Taki, M., and Sisido, M. (2005). Position-specific incorporation of a highly photodurable and blue-laser excitable fluorescent amino acid into proteins for fluorescence sensing. Bioorg. Med. Chem. *13*, 3379–3384.

Hamilton, J.A., Era, S., Bhamidipati, S.P., and Reed, R.G. (1991). Locations of the three primary binding sites for long-chain fatty acids on bovine serum albumin. Proc. Natl. Acad. Sci. U. S. A. *88*, 2051–2054.

Haney, C.M., Wissner, R.F., and Petersson, E.J. (2015). Multiply labeling proteins for studies of folding and stability. Curr. Opin. Chem. Biol. *28*, 123–130.

Hanoian, P., Liu, C.T., Hammes-Schiffer, S., and Benkovic, S. (2015). Perspectives on electrostatics and conformational motions in enzyme catalysis. Acc. Chem. Res. *48*, 482–489.

Harpaz, Z., Siman, P., Kumar, K.S.A., and Brik, A. (2010). Protein synthesis assisted by native chemical ligation at leucine. ChemBioChem *11*, 1232–1235.

Hartl, F.U. (2017). Protein Misfolding Diseases. Annu. Rev. Biochem. *86*, 21–26.

Henriques, S.T., Melo, M.N., and Castanho, M.A.R.B. (2006). Cell-penetrating peptides and

antimicrobial peptides: how different are they? Biochem. J. *399*, 1–7.

Hilaire, M.R., Ahmed, I.A., Lin, C.-W., Jo, H., DeGrado, W.F., and Gai, F. (2017b). Blue fluorescent amino acid for biological spectroscopy and microscopy. Proc. Natl. Acad. Sci. U. S. A. *114*, 6005–6009.

Hilaire, M.R., Mukherjee, D., Troxler, T., and Gai, F. (2017a). Solvent dependence of cyanoindole fluorescence lifetime. Chem. Phys. Lett. *685*, 133–138.

Heinlein, T., Knemeyer, J.P., Piestert, O., andSauer, M. (2003). Photoinduced electron transfer between fluorescent dyes and guanosine residues in DNA-hairpins. J. Phys. Chem. B *107*, 7957–7964 .

Holz, B., Klimasauskas, S., Serva, S., and Weinhold, E. (1998). 2-Aminopurine as a fluorescent probe for DNA base flipping by methyltransferases. Nucleic Acids Res. *26*, 1076–1083.

Honig, M.G., and Hume, R.I. (1986). Fluorescent carbocyanine dyes allow living neurons of identified origin to be studied in long-term cultures. J. Cell Biol. *103*, 171–187.

Hoppmann, C., Lacey, V.K., Louie, G. V., Wei, J., Noel, J.P., and Wang, L. (2014). Genetically encoding photoswitchable click amino acids in *Escherichia coli* and mammalian cells. Angew. Chemie Int. Ed. *53*, 3932–3936.

Hoppmann, C., Maslennikov, I., Choe, S., and Wang, L. (2015). In situ formation of an azo bridge on proteins controllable by visible light. J. Am. Chem. Soc. *137*, 11218–11221.

Hull, H.H., Chang, R., and Kaplan, L.J. (1975). On the location of the sulfhydryl group in bovine plasma albumin. Biochim. Biophys. Acta *400*, 132–136.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J. Mol. Graph. *14*, 33–38, 27–28.

Hunt, N.T. (2009). 2D-IR spectroscopy: Ultrafast insights into biomolecule structure and function. Chem. Soc. Rev. *38*, 1837.

Hunt, J.F., Rath, P., Rothschild, K.J., and Engelman, D.M. (1997). Spontaneous, pH-dependent membrane insertion of a transbilayer α-helix. Biochemistry 36, 15177–15192.

Ivanova, M.I., Sievers, S.A., Sawaya, M.R., Wall, J.S., and Eisenberg, D. (2009). Molecular basis for insulin fibril assembly. Proc. Natl. Acad. Sci. *106*, 18990–18995.

Jacob, M.H., D'Souza, R.N., Schwarzlose, T., Wang, X., Huang, F., Haas, E., and Nau, W.M. (2018). Method-unifying view of loop-formation kinetics in peptide and protein folding. J. Phys. Chem. B *122*, 4445–4456.

Jing, M., and Bowser, M.T. (2011). Methods for measuring aptamer-protein equilibria: A review. Anal. Chim. Acta *686*, 9–18.

Jo, H., Culik, R.M., Korendovych, I. V., DeGrado, W.F., and Gai, F. (2010). Selective incorporation of nitrile-based infrared probes into proteins via cysteine alkylation. Biochemistry *49*, 10354–10356.

Johnson, M.N.R., Londergan, C.H., and Charkoudian, L.K. (2014). Probing the phosphopantetheine arm conformations of acyl carrier proteins using vibrational spectroscopy. J. Am. Chem. Soc. *136*, 11240–11243.

Jones, A.T., and Sayers, E.J. (2012). Cell entry of cell penetrating peptides: tales of tails wagging dogs. J. Control. Release *161*, 582–591.

Kamlet, M.J., Abboud, J.L.M., Abraham, M.H., and Taft, R.W. (1983). Linear solvation energy relationships. 23. A comprehensive collection of the solvatochromic parameters, π^*, α, and β, and some methods for simplifying the generalized solvatochromic equation. J. Org. Chem. *48*, 2877–2887.

Karolin, J., Fa, M., Wilczynska, M., Ny, T., and Johansson, L.B. (1998). Donor-donor energy migration for determining intramolecular distances in proteins: I. Application of a model to the latent plasminogen activator inhibitor-1 (PAI-1). Biophys. J. *74*, 11–21.

Kato, G.J., Piel, F.B., Reid, C.D., Gaston, M.H., Ohene-Frempong, K., Krishnamurti, L., Smith, W.R., Panepinto, J.A., Weatherall, D.J., Costa, F.F., et al. (2018). Sickle cell disease. Nat. Rev. Dis. Prim. *4*, 18010.

Matsuzaki, K., Yoneyama, S., Murase, O., and Miyajima, K. (1996). Transbilayer transport of ions and lipids coupled with mastoparan X translocation Biochemistry *35*, 8450–8456 .

Kim, P.S., and Baldwin, R.L. (1990). Intermediates in the folding reactions of small proteins. Annu. Rev. Biochem. *59*, 631–660.

Kim, C.H., Axup, J.Y., and Schultz, P.G. (2013). Protein conjugation with genetically encoded unnatural amino acids. Curr. Opin. Chem. Biol. *17*, 412–419.

Kulkarni, S.S., Sayers, J., Premdjee, B., and Payne, R.J. (2018). Rapid and efficient protein synthesis through expansion of the native chemical ligation concept. Nat. Rev. Chem. *2*, 0122.

Kumar, K.S.A., Bavikar, S.N., Spasser, L., Moyal, T., Ohayon, S., and Brik, A. (2011). Total chemical synthesis of a 304 amino acid k48-linked tetraubiquitin protein. Angew. Chemie Int. Ed. *50*, 6137–6141.

Kuruma, Y., and Ueda, T. (2015). The PURE system for the cell-free synthesis of membrane proteins. Nat. Protoc. *10*, 1328–1344.

Lakowicz, J.R. (1999). Principles of Fluorescence Spectroscopy (New York: Springer US).

Lakowicz, J.R., Zelent, B., Gryczynski, I., Kuśba, J., and Johnson, M.L. (1994). Distance-dependent fluorescence quenching of tryptophan by acrylamide. Photochem. Photobiol. *60*, 205–214.

Lang, K., and Chin, J.W. (2014). Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. Chem. Rev. *114*, 4764–4806.

Lavis, L.D., and Raines, R.T. (2014). Bright building blocks for chemical biology. ACS Chem. Biol. *9*, 855–866.

Lebleu, B. (2003). Cell-penetrating peptides. J. Biol. Chem. 278, 585–590.

Lee, D., Redfern, O., and Orengo, C. (2007). Predicting protein function from sequence and structure. Nat. Rev. Mol. Cell Biol. *8*, 995–1005.

Lepthien, S., Hoesl, M.G., Merkel, L., and Budisa, N. (2008). Azatryptophans endow proteins with intrinsic blue fluorescence. Proc. Natl. Acad. Sci. U. S. A. *105*, 16095–16100.

Levinson, N.M., and Boxer, S.G. (2014). A conserved water-mediated hydrogen bond network

defines bosutinib's kinase selectivity. Nat. Chem. Biol. *10*, 127–132.

Li, Q., and Seeger, S. (2006). Label-free detection of single protein molecules using deep uv fluorescence lifetime microscopy. Anal. Chem. *78*, 2732–2737.

Lichtman, J.W., and Conchello, J.-A. (2005). Fluorescence microscopy. Nat. Methods *2*, 910–919.

Lin, C.-W., Mensa, B., Barniol-Xicota, M., DeGrado, W.F., and Gai, F. (2017). Activation pH and gating dynamics of influenza A M2 proton channel revealed by single-molecule spectroscopy. Angew. Chemie Int. Ed. *56*, 5283–5287.

Lippitz, M., Erker, W., Decker, H., van Holde, K.E., and Basché, T. (2002). Two-photon excitation microscopy of tryptophan-containing proteins. Proc. Natl. Acad. Sci. U. S. A. *99*, 2772–2777.

Liu, C.T., Layfield, J.P., Stewart, R.J., French, J.B., Hanoian, P., Asbury, J.B., Hammes-Schiffer, S., and Benkovic, S.J. (2014a). Probing the electrostatics of active site microenvironments along the catalytic cycle for *Escherichia coli* dihydrofolate reductase. J. Am. Chem. Soc. *136*, 10349–10360.

Liu, D.R., Magliery, T.J., Pastrnak, M., and Schultz, P.G. (1997). Engineering a tRNA and aminoacyl-tRNA synthetase for the site-specific incorporation of unnatural amino acids into proteins in vivo. Proc. Natl. Acad. Sci. U. S. A. *94*, 10092–10097.

Liu, J., Strzalka, J., Tronin, A., Johansson, J.S., and Blasie, J.K. (2009). Mechanism of interaction between the general anesthetic halothane and a model ion channel protein, II: Fluorescence and vibrational spectroscopy using a cyanophenylalanine probe. Biophys. J. *96*, 4176–4187.

Liu, W., Brock, A., Chen, S., Chen, S., and Schultz, P.G. (2007). Genetic incorporation of unnatural amino acids into proteins in mammalian cells. Nat. Methods *4*, 239–244.

Liu, Z., Li, X., Zhong, F.W., Li, J., Wang, L., Shi, Y., and Zhong, D. (2014b). Quenching Dynamics of Ultraviolet-Light Perception by UVR8 Photoreceptor. J. Phys. Chem. Lett. *5*, 69–72.

Loakes, D., and Brown, D.M. (1994). 5-Nitroindole as an universal base analogue. Nucleic Acids Res. *22*, 4039–4043.

Loll, P.J., Upton, E.C., Nahoum, V., Economou, N.J., and Cocklin, S. (2014). The high resolution structure of tyrocidine A reveals an amphipathic dimer. Biochim. Biophys. Acta - Biomembr. *1838*, 1199–1207.

López-Otín, C., and Bond, J.S. (2008). Proteases: Multifunctional Enzymes in life and disease. J. Biol. Chem. *283*, 30433–30437.

Lotte, K., Plessow, R., and Brockhinke, A. (2004a). Static and time-resolved fluorescence investigations of tryptophan analogues – a solvent study. Photochem. Photobiol. Sci. *3*, 348–359.

Ma, C., Polishchuk, A.L., Ohigashi, Y., Stouffer, A.L., Schön, A., Magavern, E., Jing, X., Lear, J.D., Freire, E., Lamb, R.A., et al. (2009). Identification of the functional core of the influenza A virus A/M2 proton-selective ion channel. Proc. Natl. Acad. Sci. U. S. A. *106*, 12283–12288.

Ma, J., Pazos, I.M., Zhang, W., Culik, R.M., and Gai, F. (2015). Site-specific infrared probes of proteins. Annu. Rev. Phys. Chem. *66*, 357–377.

Mairbäurl, H., and Weber, R.E. (2012). Oxygen Transport by hemoglobin. In Comprehensive Physiology, (Hoboken, NJ, USA: John Wiley & Sons, Inc.).

Majorek, K.A., Porebski, P.J., Dayal, A., Zimmerman, M.D., Jablonska, K., Stewart, A.J., Chruszcz, M., and Minor, W. (2012). Structural and immunologic characterization of bovine, horse, and rabbit serum albumins. Mol. Immunol. *52*, 174–182.

Marek, P., Gupta, R., and Raleigh, D.P. (2008). The Fluorescent amino acid *p*-cyanophenylalanine provides an intrinsic probe of amyloid formation. ChemBioChem *9*, 1372–1374.

Marek, P., Mukherjee, S., Zanni, M.T., and Raleigh, D.P. (2010). Residue-Specific, Real-Time Characterization of lag-phase species and fibril growth during amyloid formation: a combined fluorescence and IR study of p-cyanophenylalanine analogs of islet amyloid polypeptide. J. Mol. Biol. *400*, 878–888.

Markiewicz, B.N., Mukherjee, D., Troxler, T., and Gai, F. (2016). Utility of 5-cyanotryptophan fluorescence as a sensitive probe of protein hydration. J. Phys. Chem. B *120*, 936–944.

Marsh, D. (2009). Cholesterol-induced fluid membrane domains: A compendium of lipid-raft ternary phase diagrams. Biochim. Biophys. Acta - Biomembr. *1788*, 2114–2123.

Martynov, V.I., Pakhomov, A.A., Popova, N. V, Deyev, I.E., and Petrenko, A.G. (2016). Synthetic Fluorophores for Visualizing Biomolecules in Living Systems. Acta Naturae *8*, 33–46.

Mayne, C.G., Saam, J., Schulten, K., Tajkhorshid, E., and Gumbart, J.C. (2013). Rapid parameterization of small molecules using the force field toolkit. J. Comput. Chem. *34*, 2757–2770.

Measey, T.J., and Gai, F. (2012). Light-Triggered Disassembly of Amyloid Fibrils. Langmuir *28*, 12588–12592.

Meloni, S.L., and Matsika, S. (2014). Theoretical studies of the excited states of p-cyanophenylalanine and comparisons with the natural amino acids phenylalanine and tyrosine. Theor. Chem. Acc. *133*, 1497.

Middleton, C.T., Marek, P., Cao, P., Chiu, C., Singh, S., Woys, A.M., de Pablo, J.J., Raleigh, D.P., and Zanni, M.T. (2012). Two-dimensional infrared spectroscopy reveals the complex behaviour of an amyloid fibril inhibitor. Nat. Chem. *4*, 355–360.

Mintzer, M.R., Troxler, T., and Gai, F. (2015). p-Cyanophenylalanine and selenomethionine constitute a useful fluorophore-quencher pair for short distance measurements: application to polyproline peptides. Phys. Chem. Chem. Phys. *17*, 7881–7887.

Miyake-Stoner, S.J., Miller, A.M., Hammill, J.T., Peeler, J.C., Hess, K.R., Mehl, R.A., and Brewer, S.H. (2009). Probing Protein Folding using site-specifically encoded unnatural amino acids as fret donors with tryptophan. Biochemistry *48*, 5953–5962.

Moore, J.E., and Ward, W.H. (1956). Cross-linking of bovine plasma albumin and wool keratin. J. Am. Chem. Soc. *78*, 2414–2418.

Moroz, Y.S., Binder, W., Nygren, P., Caputo, G.A., and Korendovych, I. V. (2013). Painting proteins blue: β-(1-azulenyl)-l-alanine as a probe for studying protein–protein interactions. Chem. Commun. *49*, 490–492.

Moult, J., and Melamud, E. (2000). From fold to function. Curr. Opin. Struct. Biol. *10*, 384–389.

Nájera, C., and Sansano, J.M. (2007). Catalytic Asymmetric Synthesis of α-Amino Acids. Chem.

Rev. *107*, 4584–4671.

Neuweiler, H., Schulz, A., Böhmer, M., Enderlein, J., and Sauer, M. (2003). Measurement of submicrosecond intramolecular contact formation in peptides at the single-molecule level J. Am. Chem. Soc. 125, 5324–5330.

Neuweiler, H., Doose, S., and Sauer, M. (2005). A microscopic view of miniprotein folding: enhanced folding efficiency through formation of an intermediate. Proc. Natl. Acad. Sci. U. S. A. *102*, 16650–16655.

Neuweiler, H., Löllmann, M., Doose, S., and Sauer, M. (2007). Dynamics of Unfolded Polypeptide Chains in Crowded Environment Studied by Fluorescence Correlation Spectroscopy. J. Mol. Biol. *365*, 856–869.

Noren, C.J., Anthony-Cahill, S.J., Griffith, M.C., and Schultz, P.G. (1989). A general method for site-specific incorporation of unnatural amino acids into proteins. Science *244*, 182–188.

Ong, I.L.H., and Yang, K.-L. (2017). Recent developments in protease activity assays and sensors. Analyst *142*, 1867–1881.

Pace, B.S., Ofori-Acquah, S.F., and Peterson, K.R. (2012). Sickle cell disease: genetics, cellular and molecular mechanisms, and therapies. Anemia *2012*, 143594.

Pace, C.N., Grimsley, G.R., and Scholtz, J.M. (2009). Protein ionizable groups: p*K* values and their contribution to protein stability and solubility. J. Biol. Chem. *284*, 13285–13289.

Palero, J.A., Boer, V.O., Vijverberg, J.C., Gerritsen, H.C., and Sterenborg, H.J.C.M. (2005). Short-wavelength two-photon excitation fluorescence microscopy of tryptophan with a photonic crystal fiber based light source. Opt. Express *13*, 5363.

Paredes-Gamero, E.J., Martins, M.N.C., Cappabianco, F.A.M., Ide, J.S., and Miranda, A. (2012). Characterization of dual effects induced by antimicrobial peptides: Regulated cell death or membrane disruption. Biochim. Biophys. Acta - Gen. Subj. *1820*, 1062–1072.

Passow, K.T., and Harki, D.A. (2018). 4-Cyanoindole-2′-deoxyribonucleoside (4CIN): A universal fluorescent nucleoside analogue. Org. Lett. *20*, 4310–4313.

Patil, P.C., and Luzzio, F.A. (2017). Unnatural amino acid derivatives through click chemistry: Synthesis of triazolylalanine analogues. Synlett *28*, 1729–1732.

Pazos, I.M., Roesch, R.M., and Gai, F. (2013). Quenching of p-cyanophenylalanine fluorescence by various anions. Chem. Phys. Lett. *563*, 93–96.

Pazos, I.M., Ghosh, A., Tucker, M.J., and Gai, F. (2014). Ester carbonyl vibration as a sensitive probe of protein local electric field. Angew. Chemie Int. Ed. *53*, 6080–6084.

Pedretti, A., Villa, L., and Vistoli, G. (2002). VEGA: a versatile program to convert, handle and visualize molecular structure on Windows-based PCs. J. Mol. Graph. Model. *21*, 47–49.

Peng, H., Chen, W., Cheng, Y., Hakuna, L., Strongin, R., and Wang, B. (2012). Thiol reactive probes and chemosensors. Sensors (Basel). *12*, 15907–15946.

Parr, R.G., Szentpály, L., Shubin Liu, S. (1999). Electrophilicity Index J. Am. Chem. Soc. 121, 1922–1924.

Peters, T. (1985). Serum Albumin. Adv. Protein Chem. *37*, 161–245.

Peters, F.B., Brock, A., Wang, J., and Schultz, P.G. (2009). Photocleavage of the polypeptide backbone by 2-nitrophenylalanine. Chem. Biol. *16*, 148–152.

Petitpas, I., Grüne, T., Bhattacharya, A.A., and Curry, S. (2001). Crystal structures of human serum albumin complexed with monounsaturated and polyunsaturated fatty acids. J. Mol. Biol. *314*, 955–960.

Petrich, J.W., Chang, M.C., McDonald, D.B., and Fleming, G.R. (1983). On the origin of nonexponential fluorescence decay in tryptophan and its derivatives. J. Am. Chem. Soc. *105*, 3824–3832.

Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kalé, L., and Schulten, K. (2005). Scalable molecular dynamics with NAMD. J. Comput. Chem. *26*, 1781–1802.

Rehms, A.A., and Callis, P.R. (1993). Two-photon fluorescence excitation spectra of aromatic amino acids. Chem. Phys. Lett. *208*, 276–282.

Reif, M.M., and Oostenbrink, C. (2014). Molecular dynamics simulation of configurational ensembles compatible with experimental FRET efficiency data through a restraint on instantaneous FRET efficiencies. J. Comput. Chem. *35*, 2319–2332.

Remorino, A., Korendovych, I. V., Wu, Y., DeGrado, W.F., and Hochstrasser, R.M. (2011). Residue-specific vibrational echoes yield 3d structures of a transmembrane helix dimer. Science *332*, 1206–1209.

Reshetnyak, Y.K., Andreev, O.A., Lehnert, U., and Engelman, D.M. (2006). Translocation of molecules into cells by pH-dependent insertion of a transmembrane helix. Proc. Natl. Acad. Sci. U. S. A. *103*, 6460–6465.

Reshetnyak, Y.K., Segala, M., Andreev, O.A., and Engelman, D.M. (2007). A monomeric membrane peptide that lives in three worlds: in solution, attached to, and inserted across lipid bilayers. Biophys. J. *93*, 2363–2372.

Reshetnyak, Y.K., Andreev, O.A., Segala, M., Markin, V.S., and Engelman, D.M. (2008). Energetics of peptide (pHLIP) binding to and folding across a lipid bilayer membrane. Proc. Natl. Acad. Sci. *105*, 15340–15345.

Riback, J.A., Bowman, M.A., Zmyslowski, A.M., Plaxco, K.W., Clark, P.L., and Sosnick, T.R. (2018b). Commonly-used FRET fluorophores promote collapse of an otherwise disordered protein. BioRxiv 376632.

Riback, J.A., Bowman, M.A., Zmyslowski, A., Knoverek, C.R., Jumper, J., Kaye, E.B., Freed, K.F., Clark, P.L., and Sosnick, T.R. (2018a). Response to Comment on "Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water". Science *361*, eaar7949.

Richard, J.P., Melikov, K., Vives, E., Ramos, C., Verbeure, B., Gait, M.J., Chernomordik, L. V., and

Rogers, J.M.G., Lippert, L.G., and Gai, F. (2010). Non-natural amino acid fluorophores for one- and two-step fluorescence resonance energy transfer applications. Anal. Biochem. *399*, 182–189.

Rogers, J.M.G., Polishchuk, A.L., Guo, L., Wang, J., DeGrado, W.F., and Gai, F. (2011). Photoinduced electron transfer and fluorophore motion as a probe of the conformational

dynamics of membrane proteins: application to the influenza AM2 proton channel. Langmuir *27*, 3815–3821.

Romney, D.K., Murciano-Calles, J., Wehrmüller, J.E., and Arnold, F.H. (2017). Unlocking reactivity of TrpB: A general biocatalytic platform for synthesis of tryptophan analogues. J. Am. Chem. Soc. *139*, 10769–10776.

Ross, J.B., Szabo, A.G., and Hogue, C.W. (1997). Enhancement of protein spectra with tryptophan analogs: Fluorescence spectroscopy of protein-protein and protein-nucleic acid interactions. Methods Enzymol. *278*, 151–190.

Rothman, S., Liebow, C., and Isenman, L. (2002). Conservation of digestive enzymes. Physiol. Rev. *82*, 1–18.

Roy, R., Hohng, S., and Ha, T. (2008). A practical guide to single-molecule FRET. Nat. Methods *5*, 507–516.

Royer, C.A. (2006). Probing protein folding and conformational transitions with fluorescence. Chem. Rev. *106*, 1769–1784.

Saibil, H. (2013). Chaperone machines for protein folding, unfolding and disaggregation. Nat. Rev. Mol. Cell Biol. *14*, 630–642.

Scheckel, C., and Aguzzi, A. (2018). Prions, prionoids and protein misfolding disorders. Nat. Rev. Genet. *19*, 405–418.

Schmitz, A.J., Hogle, D.G., Gai, X.S., Fenlon, E.E., Brewer, S.H., and Tucker, M.J. (2016). Two-dimensional infrared study of vibrational coupling between azide and nitrile reporters in a RNA nucleoside. J. Phys. Chem. B *120*, 9387–9394.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*, 671–675.

Schultz, K.C., Supekova, L., Ryu, Y., Xie, J., Perera, R., and Schultz, P.G. (2006). A genetically encoded infrared probe. J. Am. Chem. Soc. *128*, 13984–13985.

Seidel, C.A.M., Schulz, A., and Sauer, M.H.M. (1996). Nucleobase-specific quenching of fluorescent dyes. 1. nucleobase one-electron redox potentials and their correlation with static and dynamic quenching efficiencies. J. Phys. Chem. *100*, 5541–5553.

Serrano, A.L., Troxler, T., Tucker, M.J., and Gai, F. (2010). Photophysics of a fluorescent non-natural amino acid: p-Cyanophenylalanine. Chem. Phys. Lett. *487*, 303–306.

Serrano, A.L., Waegele, M.M., and Gai, F. (2012a). Spectroscopic studies of protein folding: linear and nonlinear methods. Protein Sci. *21*, 157–170.

Serrano, A.L., Bilsel, O., and Gai, F. (2012b). Native state conformational heterogeneity of hp35 revealed by time-resolved FRET. J. Phys. Chem. B *116*, 10631–10638.

Seuvre, A.M., and Mathlouthi, M. (1987). F.T.I.R. spectra of oligo- and poly-nucleotides. Carbohydr. Res. *169*, 83–103.

Shai, Y. (1999). Mechanism of the binding, insertion and destabilization of phospholipid bilayer membranes by alpha-helical antimicrobial and cell non-selective membrane-lytic peptides. Biochim. Biophys. Acta *1462*, 55–70.

Shaner, N.C., Steinbach, P.A., and Tsien, R.Y. (2005). A guide to choosing fluorescent proteins. Nat. Methods *2*, 905–909.

Shang, S., Tan, Z., Dong, S., and Danishefsky, S.J. (2011). An advance in proline ligation. J. Am. Chem. Soc. *133*, 10784–10786.

Shen, D., Liang, K., Ye, Y., Tetteh, E., and Achilefu, S. (2007). Modulation of nuclear internalization of Tat peptides by fluorescent dyes and receptor-avid peptides. FEBS Lett. *581*, 1793–1799.

Shi, Y. (2014). A Glimpse of structural biology through X-ray crystallography. Cell *159*, 995–1014.

Shieh, P., and Bertozzi, C.R. (2014). Design strategies for bioorthogonal smart probes. Org. Biomol. Chem. *12*, 9307.

Si, B., Song, E., Si, B., and Song, E. (2018). Recent advances in the detection of neurotransmitters. Chemosensors *6*, 1.

Siman, P., Karthikeyan, S.V., and Brik, A. (2012). Native chemical ligation at glutamine. Org. Lett. *14*, 1520–1523.

Smirnov, A.V., English, D.S., Rich, R.L., Lane, J., Teyton, L., Schwabacher, A.W., Luo, S., Thornburg, R.W., and Petrich, J.W. (1997). Photophysics and biological applications of 7-azaindole and its analogs. J. Phys. Chem. B *101*, 2758–2769 .

Smith, M.E.B., Schumacher, F.F., Ryan, C.P., Tedaldi, L.M., Papaioannou, D., Waksman, G., Caddick, S., and Baker, J.R. (2010). Protein modification, bioconjugation, and disulfide bridging using bromomaleimides. J. Am. Chem. Soc. *132*, 1960–1965.

Spector, A.A., John, K., and Fletcher, J.E. (1969). Binding of long-chain fatty acids to bovine serum albumin. J. Lipid Res. *10*, 56–67.

Speight, L.C., Muthusamy, A.K., Goldberg, J.M., Warner, J.B., Wissner, R.F., Willi, T.S., Woodman, B.F., Mehl, R.A., and Petersson, E.J. (2013). Efficient synthesis and in vivo incorporation of acridon-2-ylalanine, a fluorescent amino acid for lifetime and Förster resonance energy transfer/luminescence resonance energy transfer studies. J. Am. Chem. Soc. *135*, 18806–18814.

Spicer, C.D., and Davis, B.G. (2014). Selective chemical protein modification. Nat. Commun. *5*, 4740.

Stafford, A.J., Walker, D.M., and Webb, L.J. (2012). Electrostatic effects of mutations of ras glutamine 61 measured using vibrational spectroscopy of a thiocyanate probe. Biochemistry *51*, 2757–2767.

Steinhardt, J., Krijn, J., and Leidy, J.G. (1971). Differences between bovine and human serum albumins. Binding isotherms, optical rotatory dispersion, viscosity, hydrogen ion titration, and fluorescence effects. Biochemistry *10*, 4005–4015.

Stokes, G. (1852) On the Change of refrangibility of light Phil. Trans. R. Soc. Lond. *142*, 463–562.

Sudlow, G., Birkett, D.J., and Wade, D.N. (1975). The characterization of two specific drug binding sites on human serum albumin. Mol. Pharmacol. *11*, 824–832.

Summerer, D., Chen, S., Wu, N., Deiters, A., Chin, J.W., and Schultz, P.G. (2006). A genetically

encoded fluorescent amino acid.

Suzuki, K., Kobayashi, A., Kaneko, S., Takehira, K., Yoshihara, T., Ishida, H., Shiina, Y., Oishi, S., and Tobita, S. (2009). Reevaluation of absolute luminescence quantum yields of standard solutions using a spectrometer with an integrating sphere and a back-thinned CCD detector. Phys. Chem. Chem. Phys. *11*, 9850.

Suzuki, T., Futaki, S., Niwa, M., Tanaka, S., Ueda, K., and Sugiura, Y. (2002). Possible existence of common internalization mechanisms among arginine-rich peptides. J. Biol. Chem. *277*, 2437–2443.

Swiecicki, J.-M., Bartsch, A., Tailhades, J., Di Pisa, M., Heller, B., Chassaing, G., Mansuy, C., Burlina, F., and Lavielle, S. (2014). The Efficacies of Cell-penetrating peptides in accumulating in large unilamellar vesicles depend on their ability to form inverted micelles. ChemBioChem *15*, 884–891.

Talukder, P., Chen, S., Liu, C.T., Baldwin, E.A., Benkovic, S.J., and Hecht, S.M. (2014a). Tryptophan-based fluorophores for studying protein conformational changes. Bioorg. Med. Chem. *22*, 5924–5934.

Talukder, P., Chen, S., Arce, P.M., and Hecht, S.M. (2014b). Efficient asymmetric synthesis of tryptophan analogues having useful photophysical properties. Org. Lett. *16*, 556–559.

Talukder, P., Chen, S., Roy, B., Yakovchuk, P., Spiering, M.M., Alam, M.P., Madathil, M.M., Bhattacharya, C., Benkovic, S.J., and Hecht, S.M. (2015). Cyanotryptophans as novel fluorescent probes for studying protein conformational changes and DNA–protein interaction. Biochemistry *54*, 7457–7469.

Tang, J., Signarvic, R.S., DeGrado, W.F., and Gai, F. (2007). Role of helix nucleation in the kinetics of binding of Mastoparan X to phospholipid bilayers. Biochemistry 46, 13856–13863.

Tang, J., Yin, H., Qiu, J., Tucker, M.J., DeGrado, W.F., and Gai, F. (2009). Using two fluorescent probes to dissect the binding, insertion, and dimerization kinetics of a model membrane peptide. J. Am. Chem. Soc. *131*, 3816–3817.

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., and Carrico, I. (2009). Interpretation of *p*-cyanophenylalanine fluorescence in proteins in terms of solvent exposure and contribution of side-chain quenchers: A combined fluorescence, ir and molecular dynamics study. Biochemistry *48*, 9040–9046.

Taskent-Sezgin, H., Marek, P., Thomas, R., Goldberg, D., Chung, J., Carrico, I., and Raleigh, D.P. (2010). Modulation of *p*-cyanophenylalanine fluorescence by amino acid side chains and rational design of fluorescence probes of α-helix formation. Biochemistry *49*, 6290–6295.

Thielges, M.C., and Fayer, M.D. (2012). Protein dynamics studied with ultrafast two-dimensional infrared vibrational echo spectroscopy. Acc. Chem. Res. *45*, 1866–1874.

Thorn, K. (2016). A quick guide to light microscopy in cell biology. Mol. Biol. Cell *27*, 219–222.

Tompa, P. (2012). Intrinsically disordered proteins: a 10-year recap. Trends Biochem. Sci. *37*, 509–516.

Torbeev, V.Y., and Kent, S.B.H. (2007). Convergent chemical synthesis and crystal structure of a 203 amino acid "covalent dimer" HIV-1 protease enzyme molecule. Angew. Chemie Int. Ed. *46*, 1667–1670.

Tsai, Y.-C., Lin, C.-S., Tseng, T.-H., Lee, H., and Wang, Y.-J. (1992). Production and immobilization of d-aminoacylase of *Alcaligenes faecalis* DA1 for optical resolution of N-acyl-dl-amino acids. Enzyme Microb. Technol. *14*, 384–389.

Tseng, Y.-L., Liu, J.-J., and Hong, R.-L. (2002). Translocation of liposomes into cancer cells by cell-penetrating peptides penetratin and tat: a kinetic and efficacy study. Mol. Pharmacol. *62*, 864–872.

Tsien, R.Y. (1998). The green fluorescent protein. Annu. Rev. Biochem. *67*, 509–544.

Tucker, M.J., Oyola, R., and Gai, F. (2006). A novel fluorescent probe for protein binding and folding studies:p-cyano-phenylalanine. Biopolymers *83*, 571–576.

Tucker, M.J. Oyola, R., Gai, F. (2005). Conformational distribution of a 14-residue peptide in solution: A fluorescence resonance energy transfer study. J. Phys. Chem. B 109, 4788–4795.

Tucker, M.J., Tang, J., Gai, F. (2006). Probing the kinetics of membrane-mediated helix folding. J. Phys. Chem. B 110, 8105–8109.

Tyagi, S., and Lemke, E.A. (2013). Genetically encoded click chemistry for single-molecule fret of proteins. In Methods in Cell Biology.

Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., et al. (2010). CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J. Comput. Chem. *31*, 671–690.

Vivian, J.T., and Callis, P.R. (2001). Mechanisms of tryptophan fluorescence shifts in proteins. Biophys. J. *80*, 2093–2109.

Waegele, M.M., Tucker, M.J., and Gai, F. (2009). 5-Cyanotryptophan as an Infrared Probe of Local Hydration Status of Proteins. Chem. Phys. Lett. *478*, 249–253.

Waegele, M.M., Culik, R.M., and Gai, F. (2011). Site-specific spectroscopic reporters of the local electric field, hydration, structure, and dynamics of biomolecules. J. Phys. Chem. Lett. *2*, 2598–2609.

Wakamatsu, K., Okada, A., Miyazawa, T., Ohya, M., and Higashijima, T. (1992). Membrane-bound conformation of mastoparan-X, a G-protein-activating peptide. Biochemistry *31*, 5654–5660.

Wallis, M. (1973). The pKa values of the α-amino groups of peptides derived from the N-terminus of bovine growth hormone. Biochim. Biophys. Acta - Protein Struct. *310*, 388–397.

Wang, Q., and Wang, L. (2012). Genetic incorporation of unnatural amino acids into proteins in yeast. In Methods in Molecular Biology (Clifton, N.J.), 199–213.

Wang, H., Jing, M., and Li, Y. (2018). Lighting up the brain: Genetically encoded fluorescent sensors for imaging neurotransmitters and neuromodulators. Curr. Opin. Neurobiol. *50*, 171–178.

Wang, L., Brock, A., Herberich, B., and Schultz, P.G. (2001). Expanding the genetic code of Escherichia coli. Science *292*, 498–500.

Wang, L., Xie, J., and Schultz, P.G. (2006). EXPANDING THE GENETIC CODE. Annu. Rev. Biophys. Biomol. Struct. *35*, 225–249.

Wang, R., Sun, S., Bekos, E.J., and Bright, F. V (1995). Dynamics surrounding Cys-34 in native, chemically denatured, and silica-adsorbed bovine serum albumin. Anal. Chem. *67*, 149–159.

Wang, J., Xie, J., and Schultz, P.G. (2006). A Genetically Encoded Fluorescent Amino Acid. J. Am. Chem. Soc. 128, 8738–8739 .

Weber, R.J., Liang, S.I., Selden, N.S., Desai, T.A., and Gartner, Z.J. (2014). Efficient targeting of fatty-acid modified oligonucleotides to live cell membranes through stepwise assembly. Biomacromolecules *15*, 4621–4626.

Wilcox, W., and Eisenberg, D. (1992). Thermodynamics of melittin tetramerization determined by circular dichroism and implications for protein folding. Protein Sci. *1*, 641–653.

van Wilderen, L.J.G.W., Brunst, H., Gustmann, H., Wachtveitl, J., Broos, J., and Bredenbeck, J. (2018). Cyano-tryptophans as dual infrared and fluorescence spectroscopic labels to assess structural dynamics in proteins. Phys. Chem. Chem. Phys. *20*, 19906–19915.

Winn, M., Francis, D., and Micklefield, J. (2018). De novo biosynthesis of "non-natural" thaxtomin phytotoxins. Angew. Chemie Int. Ed. *57*, 6830–6833.

Wissner, R.F., Batjargal, S., Fadzen, C.M., and Petersson, E.J. (2013). Labeling proteins with fluorophore/thioamide förster resonant energy transfer pairs by combining unnatural amino acid mutagenesis and native chemical ligation. J. Am. Chem. Soc. *135*, 6529–6540.

Witter, A., and Tuppy, H. (1960). N-(4-Dimethylamino)-3,5-Dinitrophenyl)maleimide: A coloured sulfhydryl reagent: Isolation and investigation of cysteine-containing peptides from human and bovine serum albumin. Biochim. Biophys. Acta *45*, 429–442.

Wood, B.R. (2016). The importance of hydration and DNA conformation in interpreting infrared spectra of cells and tissues. Chem. Soc. Rev. *45*, 1980–1998.

Wright, T.H., and Davis, B.G. (2017). Post-translational mutagenesis for installation of natural and unnatural amino acid side chains into recombinant proteins. Nat. Protoc. *12*, 2243–2250.

Wright, T.H., Bower, B.J., Chalker, J.M., Bernardes, G.J.L., Wiewiora, R., Ng, W.-L., Raj, R., Faulkner, S., Vallee, M.R.J., Phanumartwiwath, A., et al. (2016). Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. Science (80-.). *354*, aag1465-aag1465.

Xie, J., and Schultz, P.G. (2006). A chemical toolkit for proteins — an expanded genetic code. Nat. Rev. Mol. Cell Biol. *7*, 775–782.

Yang, H., Yang, S., Kong, J., Dong, A., and Yu, S. (2015). Obtaining information about protein secondary structures in aqueous solution using Fourier transform IR spectroscopy. Nat. Protoc. *10*, 382–396.

Yang, T.T., Sinai, P., Green, G., Kitts, P.A., Chen, Y.T., Lybarger, L., Chervenak, R., Patterson, G.H., Piston, D.W., and Kain, S.R. (1998). Improved fluorescence and dual color detection with enhanced blue and green variants of the green fluorescent protein. J. Biol. Chem. *273*, 8212–8216.

Ye, S., Huber, T., Vogel, R., and Sakmar, T.P. (2009). FTIR analysis of GPCR activation using azido probes. Nat. Chem. Biol. *5*, 397–399.

Yokoyama, Y., Hikawa, H., Mitsuhashi, M., Uyama, A., and Murakami, Y. (1999). Syntheses

without protection: a three-step synthesis of optically active clavicipitic acid by utilizing biomimetic synthesis of 4-bromotryptophan. Tetrahedron Lett. *40*, 7803–7806.

Yokoyama, Y., Hikawa, H., Mitsuhashi, M., Uyama, A., Hiroki, Y., and Murakami, Y. (2004). Total synthesis without protection: Three-step synthesis of optically active clavicipitic acids by a biomimetic route. European J. Org. Chem. *2004*, 1244–1253.

Chen, Y., Bo Liu, B., Yu, H.T., and Barkley, M.D. (1996). The peptide bond quenches indole fluorescence J. Am. Chem. Soc. *118*, 9271–9278.

Yuan, T., Weljie, A.M., and Vogel, H.J. (1998). Tryptophan fluorescence quenching by methionine and selenomethionine residues of calmodulin: Orientation of peptide and protein binding. Biochemistry *37*, 3187–3195.

Zhang, K., Ahmed, I.A., Kratochvil, H., DeGrado, W.F., Gai, F., and Jo, H. (2019). Synthesis and application of the blue fluorescent amino acid l-4-cyanotryptophan to assess peptide-membrane interactions. Chem. Commun. DIO: 10.1039/C9CC01152H.

Zhang, W., Markiewicz, B.N., Doerksen, R.S., Smith, A.B., III, and Gai, F. (2016). C≡N Stretching vibration of 5-cyanotryptophan as an infrared probe of protein local environment: What determines its frequency? Phys. Chem. Chem. Phys. *18*, 7027.

Zhong, D. (2009). Hydration dynamics and coupled water-protein fluctuations probed by intrinsic tryptophan. (John Wiley & Sons, Ltd), pp. 83–149.

Zhou, W., Saran, R., and Liu, J. (2017). Metal Sensing by DNA. Chem Rev. 117, 8272–8325.

Zhu, L., Kemple, M.D., Yuan, P., and Prendergast, F.G. (1995). N-terminus and lysine side chain pKa values of melittin in aqueous solutions and micellar dispersions measured by ^{15}N NMR. Biochemistry *34*, 13196–13202.

Zimmermann, J., Thielges, M.C., Seo, Y.J., Dawson, P.E., and Romesberg, F.E. (2011). Cyano groups as probes of protein microenvironments and dynamics. Angew. Chemie Int. Ed. *50*, 8333–8337.

Zoonens, M., Reshetnyak, Y.K., and Engelman, D.M. (2008). Bilayer interactions of pHLIP, a peptide that can deliver drugs and target tumors. Biophys. J. *95*, 225–235.

Zorko, M., and Langel, U. (2005). Cell-penetrating peptides: Mechanism and kinetics of cargo delivery. Adv. Drug Deliv. Rev. *57*, 529–545.