ILLUMINATING COLLAGEN: EXPLORING TRIPLE HELIX FORMATION WITH FLUORESCENCE-BASED KINETICS

by

Rachel Lane Smith

A Signature Honors Project Presented to the

Honors College

East Carolina University

In Partial Fulfillment of the

Requirements for

Graduation with Honors

by

Rachel Lane Smith

Greenville, NC

May, 2024

Approved by:

Dr. William Allen

Department of Chemistry in the Thomas Harriot College of Arts and Sciences

Abstract:

Collagen mimicking peptides (CMPs) can be used to understand the properties of collagen, a vital protein in the human body. Synthesis of collagen chains with a naphthalimide fluorophore allows for monitoring triple helix folding kinetics via fluorescence spectroscopy. Here we address whether kinetic behavior of the CMP system changes when Pro-Hyp-Gly (POG) repeats are interrupted by a non-native fluorescent amino acid, and if the folding rate can be controlled. Results indicate that the $(POG)_7$ core can be interrupted by a fluorophore and still show first-order folding rates (k=0.001 s^{-1}). However, this is dependent on the location of the fluorophore.

Introduction:

Collagen, a main component of the extracellular matrix, is the most abundant protein in the human body. As such, the determination of its shape, structure and interactions with itself and other proteins has been researched extensively.[1] Collagen research is pertinent to applications in medicine, pharmaceuticals, biological synthetics for materials development, structural biology, and conjugation with fluorophores for diagnostics.[2] Collagen is a quaternary protein composed of three intertwined peptide chains where each individual chain forms a left handed polyproline-II type twist, and together, a triple helix (Fig. 1).[3] The individual peptide chains can be described by the general form $(Xaa\text{-}Yaa\text{-}Gly)_n$ in which Xaa and Yaa are most commonly proline and hydroxyproline. Gly is always found in the third position due to steric considerations.[1-2] This repetition gives rise to the name of "$(POG)_n$ repeats." The triple helix structure is largely driven by the cyclic shape of the proline amino acid side chains which cause conformational preorganization. Monomeric collagen exhibits a *cis* peptide bond between the residues; rotation about this bond to the *trans* form occurs during folding. This isomerization is rate limiting. This has been demonstrated via faster folding of CMPs that were synthesized to exist in an all-trans state.[4] Native collagen folds in the C → N direction with a first order rate constant, k, on the order of k=0.001 s^{-1}.[4] However, under certain conditions this folding direction can be reversed.[5,6]

Figure 1: Collagen triple helix structure

Collagen mimicking peptides (CMPs) are often used to study collagen folding. Building collagen models from ten for fewer $(POG)_n$ repeats allows for increased solubility and controlled studies.[2] However, it is important to note that the stability, measured by melting point (T_m),

generally increases with the size of the peptides. The smallest CMP known to form a stable triple helix at room temperature used n=7.[1] The primary sequence of CMPs used in this study are interrupted by a fluorophore but still sum to n=7.

Background:

Circular dichroism (CD) is a technique used to quantify differences between optically active chiral molecules. By using left and right circularly polarized light, a specific molecule will absorb more of one kind of light than the other. This difference is measured.[7] The right-handedness of a collagen triple helix makes it inherently chiral. Because of this CD can be, and routinely is, used to quantify the extent of folding collagen at varying temperatures. However, this method typically requires relatively high sample concentrations and long experimental acquisition times.[6]

Fluorescence spectroscopy uses a UV or visible light source of a specific wavelength to excite the electrons of a sample. As the electrons return to their ground state, light is released. This is recorded as an emission spectrum.[8] Fluorescence studies allow for the evaluation of dilute samples.[6] Previous studies show that the addition of a fluorophore to a collagen strand can allow for fluorescence testing to be used in place of CD to determine T_m values of triple helices. In this way, the intensity of the emission band(s) can be treated as an indirect measure of the abundance of triple helical structure.[6]

Through this research, naphthalimides acting as an amino acid analog will be attached to a collagen strand. Naphthalimides are common fluorescent tags because of their stability and intense emission of visible light.[9,10] These naphthalimides therefore can be used in temperature-dependent fluorescence studies of collagen helical structure. Previous studies have

demonstrated the ability for naphthalimides to be incorporated into a peptide via standard solid-phase peptide synthesis. They are suitable for Fmoc protection which is required for effective incorporation into peptide chains.[11] Additionally, naphthalimides have been shown to exhibit a self-quenching effect.[12] This allows for the visualization of naphthalimide aggregation via a decrease in fluorescence intensity. These compounds are also relatively non-polar, and the addition of a hydrophobic moiety on the end of a collagen stand has been proven to facilitate faster triple helix formation via the hydrophobic effect.[6] Here, non-polar surfaces nucleate together, providing an initiation point for further folding. We hypothesize that the location of the naphthalimide within the strand may dictate the prominence of this effect.

Purpose of the Study:

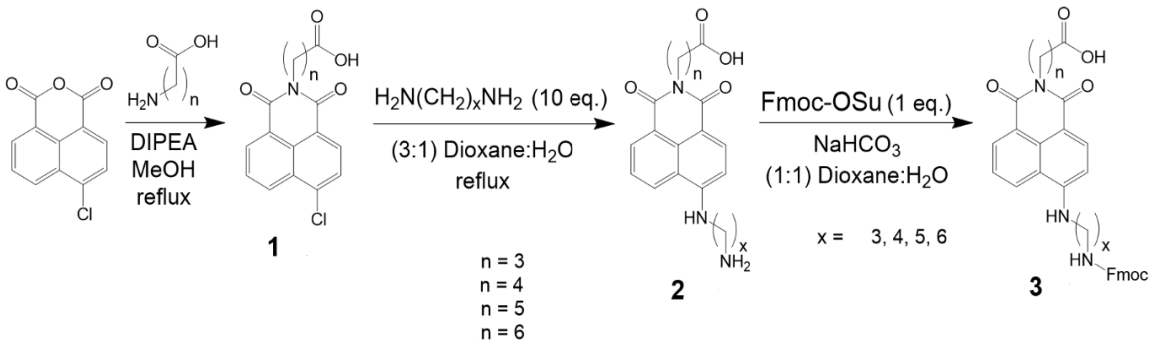

Figure 2: Reaction scheme for synthesis of Fmoc-protected naphthalimide amino acids.

The goal of this research is to synthesize collagen chains with naphthalimide linkers. Naphthalimides are prepared to include hydrocarbon head and tail groups of varying lengths (Figure 2), then the resultant artificial amino acids are incorporated into $(POG)_7$ CMPs. The location of the naphthalimide within the CMP is changed from the N-terminus, to after 2 POG repeats, to after 4 POG repeats. For ease of communication a unique naming convention has

been developed for these products. The naphthalimide products are described by the general form X-Nap-N-OH, where the first value is the number of carbons in the tail group group, x in Figure 2, and the second is the number of carbons in the head group, n in Figure 2. Here we use "Fmoc-6-Nap-6-OH" as the fluorophore in the CMPs. As the name suggests it is the Fmoc-protected naphthalimide product with 6 carbons in both the head and tail group. This naming of the naphthalimide is used to show its location as a non-native amino acid within the peptides.

$$Ac\text{-}(POG)_x\text{-}(6\text{-}Nap\text{-}6)\text{-}(POG)_y\text{-}G\text{-}K\text{-}K\text{-}NH_2 \quad (1)$$

Equation 1: Sample naming of CMPs.

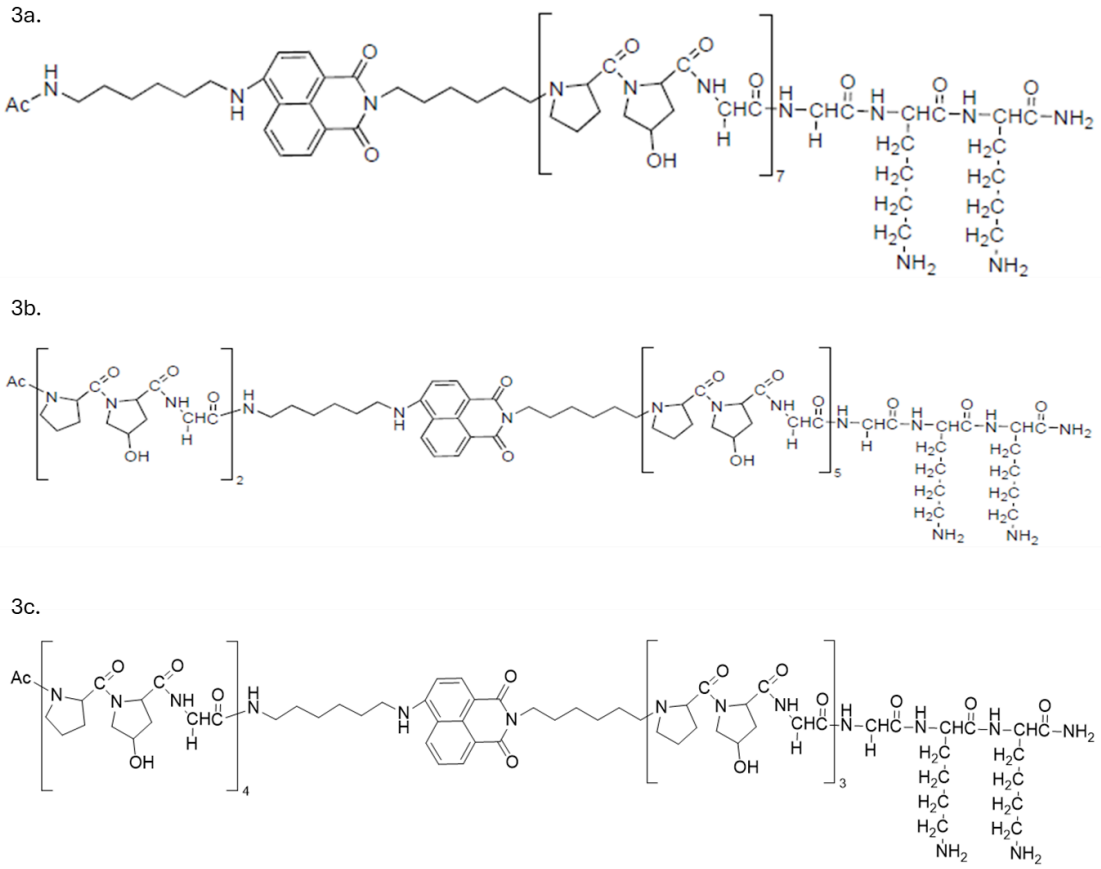

Figure 3: CMP 1-3 where the location of the fluorophore starts directly at the N-terminus and shifts right by two residues in each subsequent CMP.

Following synthesis, UV-visible and fluorescence spectroscopies were used to perform temperature dependent kinetic studies on the triple helical structure of these peptides. The observation of results that are similar to those from previous studies, specifically a decrease in fluorescence signal attributed to the unfolded monomeric CMPs, suggests that naphthalimide fluorescence is suitable for the quantification of triple helices.[6]

Research is thus aimed to answer if the kinetic behavior of the CMP system radically changes when POG repeats are interrupted by fluorescent reporter molecules, and if the folding rate(s) can be controlled by the the location of the fluorophore. This thesis will attempt to address these questions using a series of spectroscopic techniques.

Results:

Data acquisition began using diode-array UV-Vis spectroscopy because of its rapid scanning rate; this technique allows a full spectrum to be obtained in under 10 seconds. Those studies showed the folding is indeed slow enough to be manually monitored. However, after 20 minutes of data collection for any of the CMP systems, there was deviation from the first-order trend line. It became clear that the instrument lacked the sensitivity needed to detect small absolute changes in absorbance, prompting a switch to fluorescence spectroscopy. The increased sensitivity reduces noise in each recorded emission intensity. Additionally, using the integration of each peak instead of the highest absorbance value leads to more reliable data.

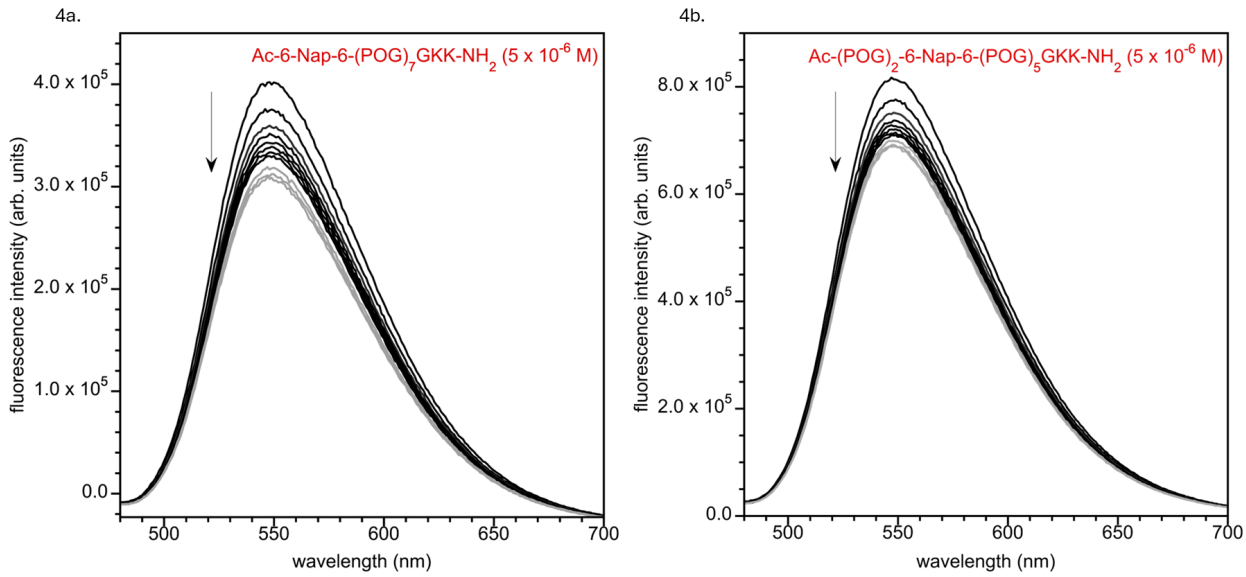

Figure 4a-b: Decreasing Fluorescence intensity with CMP folding.

As naphthalimides aggregate, the fluorescence is quenched. This is seen by the decreasing fluorescence intensity (Fig. 4) which suggests CMP folding. The trend for the CMP 2 system with an interrupting naphthalimide mirrors the CMP 1 control (Fig. 5). Fitting the data to first order kinetics revealed that CMPs 1 and 2 exhibit nearly equal *k* values (Fig. 6).

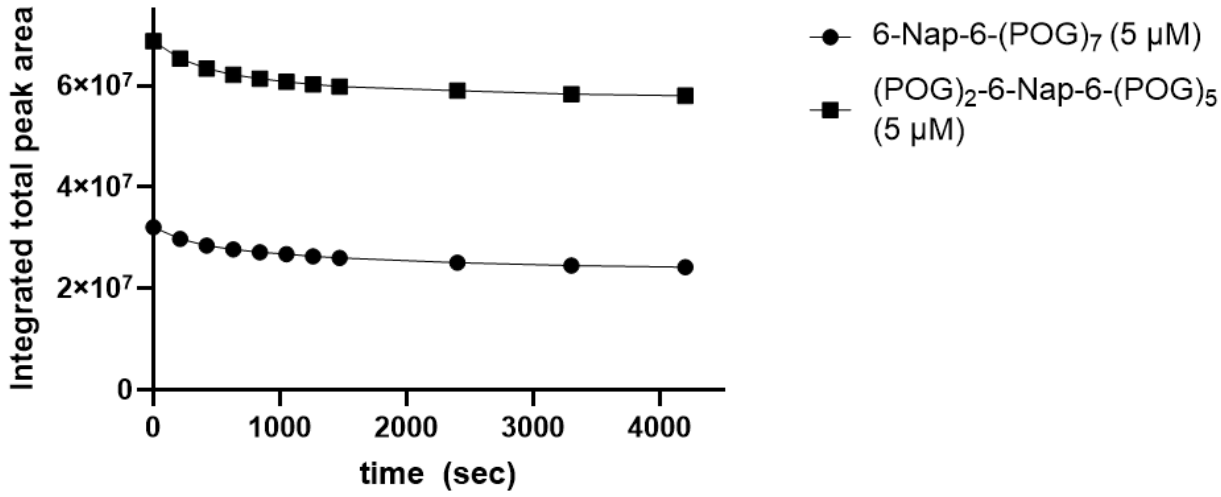

Figure 5: Integrated total peak area: CMP 1 and 2 (integrated from 500-650 nm).

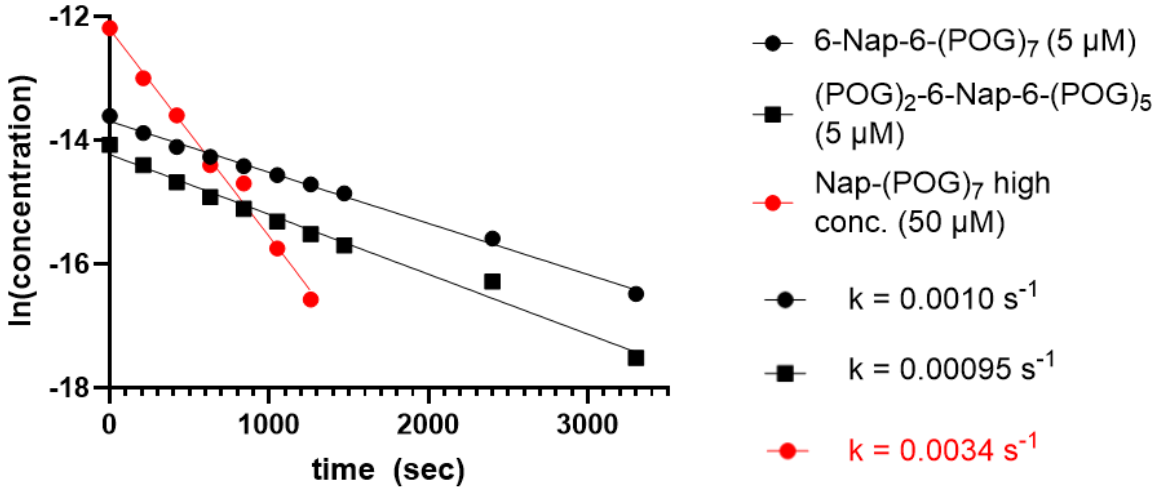

Figure 6: Fitting of integrated fluorescence intensity to first order kinetics.

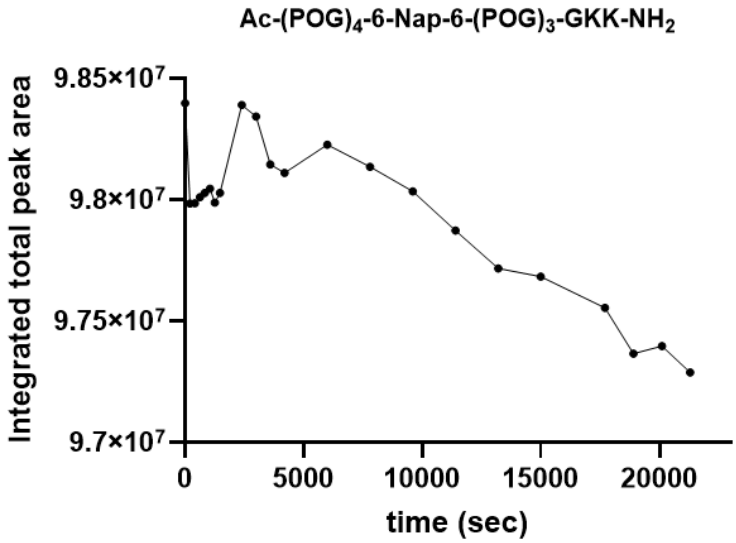

Figure 7: Integrated total peak area for CMP 3.

Integrated total peak area of CMP 3 does not mimic the exponential decay seen in CMP 1 and 2 (Fig.5). There is only a 1.1% decrease in total integrated peak area.

Table 1: *k* values for first order CMP systems

CMP	*k* value (s^{-1})
Ac-6-Nap-6-$(POG)_7$-GKK-NH_2	0.001
Ac-$(POG)_2$-6-Nap-6-$(POG)_5$-GKK-NH_2	0.00095
Ac-$(POG)_4$-6-Nap6-$(POG)_3$-GKK-NH_2	——

The *k* values for CMP 1 and 2 are effectively equivalent. No *k* value could be calculated for CMP 3.

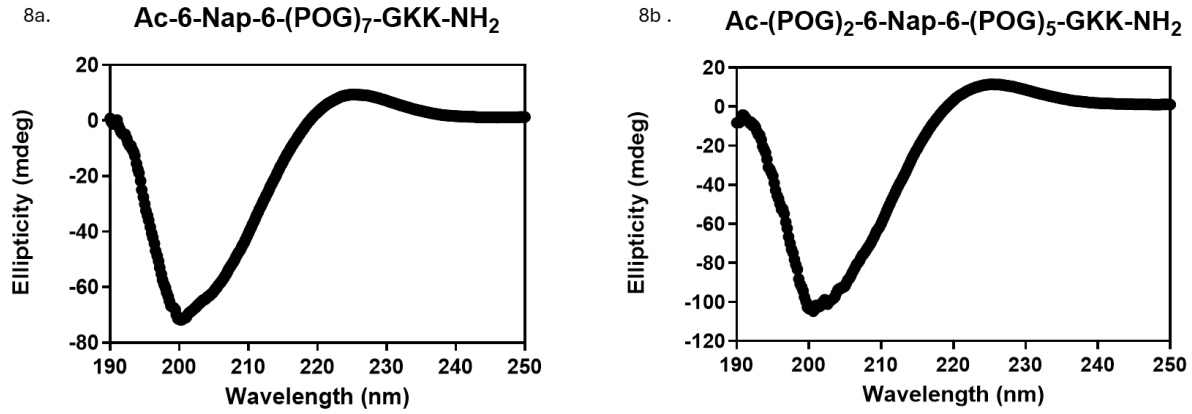

Figure 8: Ellipticity of CMP 1 and 2

CD spectra were then taken for confirmation of triple helical structure. A local maxima was located at ~ 225 nm for both CMP 1 and 2, consistent with the polyproline-II type structure found in collagens. Additionally the ellipticity at 225 nm is larger in the CMP 2 system than that in the CMP 1 system, suggesting that the former is more tightly ordered.

Discussion:

These results demonstrate that interruption of the $(POG)_7$ core by a fluorophore still allows for triple helix formation in a first order manner. To our knowledge this is the first example of a stable folded $(POG)_n$ system with fewer than seven consecutive repeats in which a

non-natural amino acid serves as the space unit. This is in contrast to previous studies using short proteinogenic sequences as spacers within collagen to yield robust triple helices.[12] However, there are limitations to folding based on the location of the fluorophore. It is thought that CMP 1 and 2 are folding in the native C → N direction. The $(POG)_5$ in CMP 2 appears to be long enough to initiate a triple helical folding stable enough to incorporate the fluorophore. However CMP 3 with only $(POG)_3$ before the fluorophore does not show the same effects. This lack of significant and patterned decrease in fluorescence intensity (Fig. 7) is likely indicative of a lack of folding. Three POG repeats does not appear to be stable enough to promote folding following a bulky non-native amino acid fluorophore.

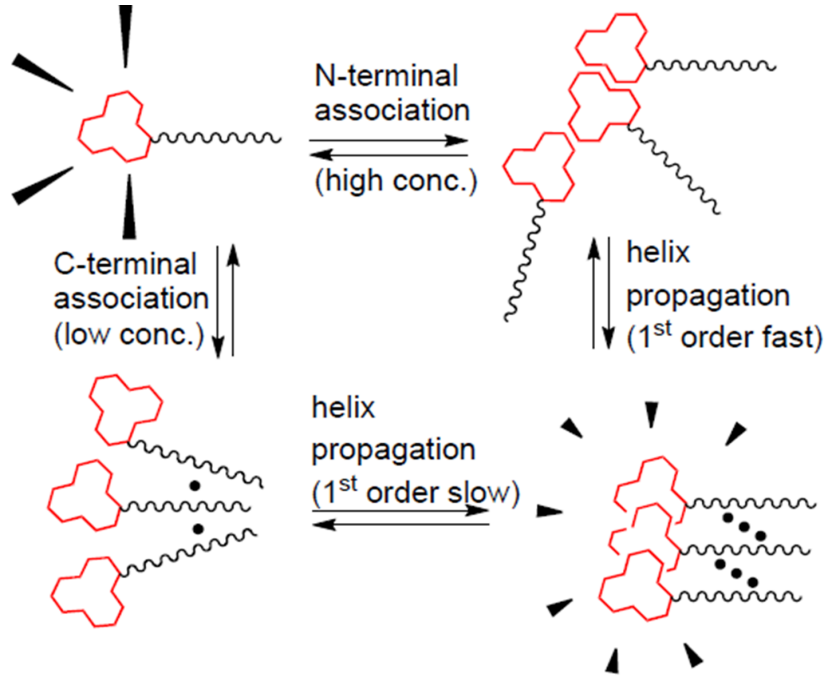

Figure 9: Schematic for strand aggregation effect on kinetic order

Additionally, folding rates were found to vary based on concentration. At increased concentrations the rate constant, *k*, is larger indicating faster folding (Fig. 6), yet the process remains first order. This may be indicative of N → C folding where the naphthalimide fluorophores are initially aggregating via the hydrophobic effect and then folding follows toward

the C terminus (Fig 8). This result is consistent with studies that used a highly hydrophobic fluorophore on the N-terminus.[5,6] Another possibility is that the triple helix *folding* of the $(POG)_7$ following the fluorophore is not actually faster, but a quicker decrease in fluorescence intensity is observed because the naphthalimides are aggregating first and thus being immediately quenched. This result may not be replicable with the other CMP systems because of the fluorophore location in the middle of the strand, however this could be investigated through future studies.

CD spectra showed a local maxima at ~ 225 nm (Fig. 8). This is used as evidence of a polyproline-II helical character. This suggests that the CMP systems are folding into a triple helix and supports the results found via fluorescence. The increased ellipticity seen in CMP 2 may be indicative of interactions between multiple triple helices. Previous research has shown that interruption of the POG repeats can promote fibril aggregation between strands.[13] A stable quaternary structure driven by the naphthalimide interruption could account for the higher ellipticity in the CMP 2 system.

Additional CD studies on the CMP 2 and 3 systems will provide more conclusive evidence of folding and perhaps prove that fluorescence spectroscopy alone is sufficient to detect such folding. Variable temperature measurements (VTM) will be performed to determine the melting temperature (*T*m) of the systems.

Conclusion:

Three new CMPs were synthesized with a naphthalimide fluorophore at varying locations. Circular dichroism was used to confirm the presence of a polyproline-II twist indicative of triple helix formation in both CMP 1 and 2. Fluorescence spectroscopy was used to monitor the kinetics of folding. It was found that, because native collagen folds in the C → N

direction, it is important to have enough POG repeats prior to the naphthalimide to promote folding past such a bulky non-native amino acid. Results indicate that when the naphthalimide is located either at the N-terminus or after two POG repeats toward the C-terminus (CMP 1 and 2), first order triple helix formation is observed (k=0.001 s^{-1} and k=0.00095 s^{-1}, respectively). When the naphthalimide is located even closer to the C-terminus, such as after four POG repeats towards the C-terminus in CMP 3, then there is not enough stability to promote folding through the non-native amino acid.

Being able to synthesize such short CMPs with an easy-to-monitor fluorophore is important for many applications. This lends itself to many more kinetic studies as well as pharmaceutical and biological applications. The capsule created by the naphthalimide fluorophores could be utilized as a drug delivery system, or the CMPs could be used as a molecular probe for biological systems. Not having to synthesize seven POG repeats on either side of the naphthalimide cuts down significantly on time and cost for any of these applications.

Experimental:

Synthesis of fluorophore: Fmoc-6-Nap-6-OH

7-(6-chloro-1,3-dioxo-1*H*-benzo[*de*]isoquinolin-2(3*H*)-yl)heptanoic acid (1)

A suspension of 7-aminoheptanoic acid (1.5 g, 10.3 mmol) and 4-chloro-1,8-naphthalic anhydride (2.0 g, 8.6 mmol) was prepared in methanol (25 mL). Following the addition of N,N-diisopropylethylamine (1.9 mL, 11.2 mmol), the mixture changed from pink to a deep orange color. An air condenser was affixed to the round bottom flask and the contents were heated to reflux and stirred (24 hours). The reaction mixture, a dark red-brown solution, was quenched with glacial acetic acid (1.5 mL) causing the product to precipitate. The yellow

precipitate was isolated via suction filtration and dried under vacuum (2.8 g, 89.3%). ^{1}H NMR (DMSO-d6) δ 1.46 (m, 4H), 1.68 (m, 2H), 1.76 (m, 2H), 2.38 (t, 2H), 4.18 (t, 2H), 7.86 (m, 2H), 8.51 (d, 1H), 8.61 (d, 1H), 8.67 (d, 1H).

7-(6-(6-aminohexylamino)-1,3-dioxo-1*H*-benzo[*de*]isoquinolin-2(3*H*)-yl)heptanoic acid (2)

Product **1** (517.5 mg, 1.44 mmol) was suspended in 3:1 dioxane:water (16 mL) with 1,6-diaminohexane (1.8 mL, 13.9 mmol). An air condenser was affixed to the round bottom flask and the solution was heated to reflux and stirred (24 hours) giving way to a thick yellow-gold solution. Following the addition of 5 % sodium hydroxide (25 mL) the solution became translucent and red. The mixture was washed with dichloromethane (3 x 10 mL) in a separatory funnel, and the organic layers were discarded. The aqueous layer was acidified with glacial acetic acid (10 mL) causing the product to precipitate. The yellow precipitate was isolated via suction filtration and dried under vacuum (480 g, 76%). ^{1}H NMR (DMSO-d6 and 1% TFA) δ 1.31 (m, 4H), 1.40 (m, 4H), 1.56 (m, 6H), 1.70 (m, 2H), 2.19 (t, 2H), 2.77 (m, 2H), 3.38 (t, 2H), 3.99 (t, 2H), 6.76 (d, 1H), 7.66 (t, 2H), 8.26 (d, 1H), 8.43 (d, 1H), 8.69 (d, 1H).

7-(6-(6-(((9*H*-fluoren-9-yl)methoxy)carbonylamino)hexylamino)-1,3-dioxo-1*H*-benzo[*de*]iso quinolin-2(3*H*)-yl)heptanoic acid (3)

Product **2** (340 mg, 0.770 mmol), sodium bicarbonate (251 mg, 2.99 mmol), and Fmoc-OSu (261 mg, 0.770 mmol) were suspended in 1:1 dioxane:water (50 mL) and stirred (24+ hours) at room temperature. The bright orange-yellow reaction mixture contained a lot of precipitate which was separated via suction filtration. The filtrate was then acidified with glacial acetic acid (2 mL) and stored overnight at 4 ℃. The yellow precipitate was then isolated via suction filtration and dried

under vacuum (335.5 g, 66%). ^{1}H NMR (DMSO-d6) δ 1.31 (brm, 6H), 1.49 (brm, 6H), 1.58 (brm, 2H), 1.69 (brm, 2H), 2.17 (t, 2H), 2.98 (m, 2H), ~3.4 (2H, obscured by H_2O), 3.99 (t, 2H), 4.19 (t, 1H), 4.28 (d, 2H), 6.77 (d, 1H) , 3.71 (t, 3H), 7.39 (t, 2H), 7.66 (m, 3H), 7.76 (brm, 1H), 7.86 (d, 3H), 8.25 (d, 1H), 8.42 (d, 1H), 8.96 (d, 1H).

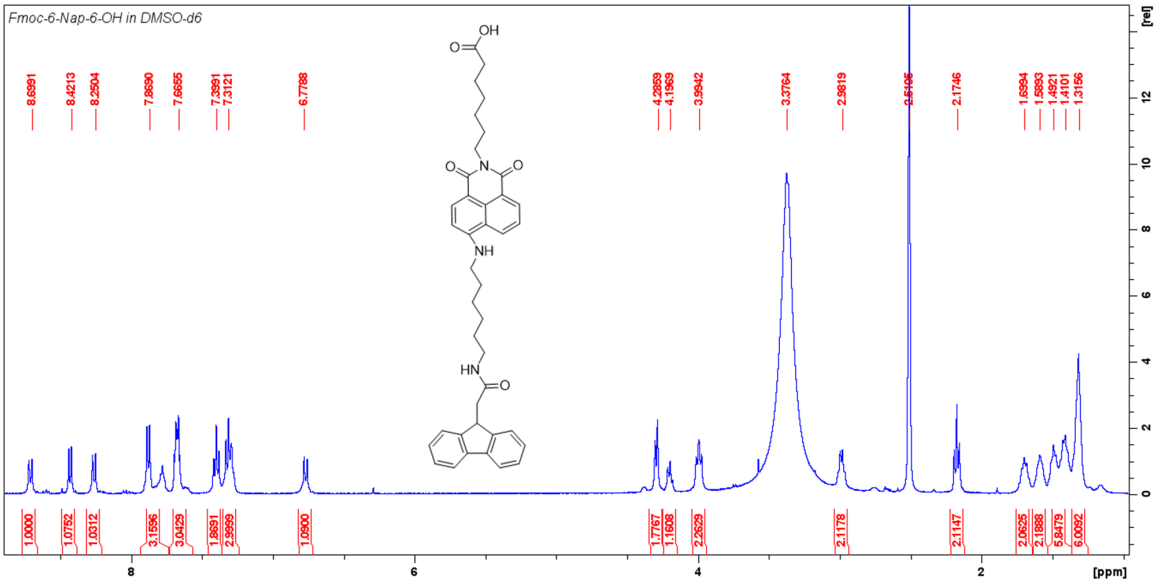

Figure 10: Structure of Product 3

Synthesis of CMPs:

$(POG)_7$-GKK-(resin)

Peptides were synthesized using standard Fmoc-based solid-phase peptide synthesis (SPPS) on Rink amide low-loading resin. For ease of later synthesis, a “pre-loaded” resin was made consisting of seven POG repeats tethered to glycine and two Boc protected lysine residues that would remain attached to the Rink Amide resin at the C-terminus. Fmoc-Pro-OH (0.135 g, 0.4 mmol), Fmoc-Hyp-OH (0.141 g, 0.4 mmol), Fmoc-Gly-OH (0.119 g, 0.4 mmol), Fmoc-Lys(Boc)-OH (0.187, 0.4), were weighed and added into respective vials each with HBTU (0.152 g, 0.4 mmol) based on a 0.2 mmol sale of Rink amide MBHA resin. The unloaded resin

(0.039 g, 0.025 mmol) was added to a reaction vessel and rinsed with DMF. To avoid deletions, each amino acid was "double coupled." Following SPPS, the peptide together with the resin was then isolated with suction filtration. The resin and peptide were washed sequentially with glacial acetic acid, CH_2Cl_2, and MeOH, transferred to a vial and dried under vacuum (572.57 mg, 98.2%).

Ac-(6-Nap-6)-$(POG)_7$-GKK-NH_2 (CMP 1)

SPPS was used to incorporate Product **3** into **CMP 1**. Product **3** (0.053 g, 0.08 mmol) was added to a vial with HBTU (0.031 g, 0.08 mmol) at a three-fold excess of the pre-loaded resin described above. Acetic Anhydride (2 mL) was used as the capping agent. The $(POG)_7$ pre-loaded resin (0.1058 g, 0.027 mmol) was added to a reaction vessel and rinsed with DMF. Following SPPS, the peptide together with the resin was then isolated with suction filtration. The resin and peptide were washed sequentially with glacial acetic acid, CH_2Cl_2, and MeOH, transferred to a vial and dried under vacuum. The peptide and resin were then suspended in a cleavage cocktail of triethylsilane (TES) (50 µL), 1,3-dimethoxybenzene (50 µL) and trifluoroacetic acid (TFA) (1.3 mL). The reaction mixture was capped and stirred (3 hours) at room temperature. The resin was then separated from the peptide via suction filtration. The CMP product, assumed to be in the filtrate, was pipetted dropwise into cold diethyl ether (25 mL) resulting in a bright yellow precipitate. The product was then isolated via suction filtration using a fine glass fritted funnel and washed with additional cold ether (72.79 mg, 95.2 %). The product was dissolved in (1:1) ACN:water and purified using preparative-scale reverse phase HPLC (RP-HPLC) using a C18 column (250×21 mm dimensions). The wavelengths for UV−vis detection were set to 214, 254, 280, and 355 nm. The run was performed with a 20% to 100%

gradient of mobile phase A (ACN with 0.1% TFA) and 80% to 0% mobile phase B (water with 0.1% TFA) over 30 minutes using a flow rate of 5 mL/min. Two yellow-colored, fluorescent fractions were collected at 17.5 minutes and 20 minutes, respectively. Analysis via triple quadrupole mass spectrometry (QqQ-MS) showed the two peaks were identical, so fractions were combined and lyophilized for further analysis via QqQ-MS which confirmed the presence of the fully formed peptide. Calc'd for CMP 1 + 2H+ = 1332.67; found = 1333.6. Calc'd for CMP 1 + 3H+ = 888.78; found = 889.7. Calc'd for CMP 1 + 4H+ = 666.84; found = 667.3.

$(POG)_5$-GKK-(resin)

A second "pre-loaded" resin was made consisting of only five POG repeats tethered to glycine and two Boc protected lysine residues that would remain attached to the Rink Amide resin. Fmoc-Pro-OH (0.135 g, 0.4 mmol), Fmoc-Hyp-OH (0.141 g, 0.4 mmol), Fmoc-Gly-OH (0.119 g, 0.4 mmol), Fmoc-Lys(Boc)-OH (0.187, 0.4), were weighed and added into respective vials each with HBTU (0.152 g, 0.4 mmol) based on a 0.2 mmol sale of Rink Amide MBHA resin. All Pro residues were "double coupled" as was the C-terminal Lys residue. The Rink Amide resin (0.039 g, 0.025 mmol) was added to a reaction vessel and rinsed with DMF. Following SPPS, the peptide together with the resin was then isolated with suction filtration. The resin and peptide were washed sequentially with glacial acetic acid, CH_2Cl_2, and MeOH, transferred to a vial and dried under vacuum (389.05 mg, 87.9%).

Ac-$(POG)_2$-(6-Nap-6)-$(POG)_5$-GKK-NH_2 (CMP 2)

SPPS using the $(POG)_5$ pre-loaded resin from above was used to incorporate Product **3** into **CMP 2**. Fmoc-Pro-OH (0.067 g, 0.2 mmol), Fmoc-Hyp-OH (0.071 g, 0.2 mmol), Fmoc-Gly-OH

(0.059 g, 0.2 mmol) were weighed and added into respective vials each with HBTU (0.076 g, 0.2 mmol) based on a 0.03 mmol sale of Rink Amide MBHA resin. All Pro residues were "double coupled." Product **3** (0.050 g, 0.076 mmol) was added to a vial with HBTU (0.076 g, 0.2 mmol) at a 2.5-fold excess of the Rink Amide MBHA resin. The $(POG)_5$ pre-loaded resin (0.109 g, 0.03 mmol) was added to a reaction vessel and rinsed with DMF. Following SPPS, the peptide together with the resin was then isolated with suction filtration. The resin and peptide were washed sequentially with glacial acetic acid, CH_2Cl_2, and MeOH, transferred to a vial and dried under vacuum. The peptide and resin were then suspended in a cleavage cocktail of TES (50 µL), 1,3-dimethoxybenzene (50 µL) and TFA (1.3 mL). The reaction mixture was capped and stirred (3 hours) at room temperature. The resin was then separated from the peptide via suction filtration. The CMP product, assumed to be in the filtrate, was pipetted dropwise into cold diethyl ether (25 mL) resulting in a bright yellow precipitate. The precipitate was then isolated via suction filtration using a fine glass fritted funnel and washed with additional cold ether (123.4 mg, 95.8 %). The product was dissolved in (1:1) ACN:water and purified using preparative-scale RP-HPLC using a C18 column (250×21 mm dimensions). The wavelength for UV−vis detection was set at 214, 254, 280, and 355 nm. The run was performed with a 20% to 100% gradient of mobile phase A (ACN with 0.1% TFA) and 80% to 0% mobile phase B (water with 0.1% TFA) over 30 minutes using a flow rate of 5 mL/min. Two yellow-colored, fluorescent fractions were collected at 16.5 minutes and 19 minutes, respectively. Analysis via QqQ-MS showed the two peaks were identical, so fractions were combined and lyophilized for further analysis via QqQ-MS which confirmed the presence of the fully formed peptide.
Calc'd for CMP 2 + 2H+ = 1332.67; found = 1333.8. Calc'd for CMP 2 + 3H+ = 888.78; found = 889.8. Calc'd for CMP 2 + 4H+ = 666.84; found = 667.5.

Ac-$(POG)_4$-(6-Nap-6)-$(POG)_3$-GKK-NH_2 (CMP 3)

SPPS was used to incorporate product **3** into **CMP 3**. Fmoc-Pro-OH (0.034 g, 0.1 mmol), Fmoc-Hyp-OH (0.036 g, 0.1 mmol), Fmoc-Gly-OH (0.030 g, 0.1 mmol), Fmoc-Lys(Boc)-OH (0.047, 0.1), were weighed and added into respective vials each with HBTU (0.038 g, 0.1 mmol) based on a 0.025 mmol sale of Rink Amide MBHA resin. Product **3** (0.033 g, 0.05 mmol) was added to a vial with HBTU (0.038 g, 0.1 mmol) at a two-fold excess of the Rink Amide MBHA resin. All Pro residues were "double coupled" as was the C-terminal Lys residue. The Rink Amide resin (0.039 g, 0.025 mmol) was added to a reaction vessel and rinsed with DMF. Following SPPS, the peptide together with the resin was then isolated with suction filtration. The resin and peptide were washed sequentially with glacial acetic acid, CH_2Cl_2, and MeOH, transferred to a vial and dried under vacuum. The peptide and resin were then suspended in a cleavage cocktail of TES (50 µL), 1,3-dimethoxybenzene (50 µL) and TFA (1.3 mL). The reaction mixture was capped and stirred (3 hours) at room temperature. The resin was then separated from the peptide via suction filtration. The CMP product, assumed to be in the filtrate, was pipetted dropwise into cold diethyl ether (25 mL) resulting in a bright yellow precipitate. The product was then isolated via suction filtration using a fine glass fritted funnel and washed with additional cold ether (178.4 mg, 88.1 %). The product was dissolved in 1:1 ACN:water and purified using preparative-scale RP-HPLC using a C18 column (250×21 mm dimensions). The wavelength for UV−vis detection was set at 214, 254, 280, and 355 nm. The run was performed with a 40% to 100% gradient of mobile phase A (ACN with 0.1% TFA) and 60% to 0% mobile phase B (water with 0.1% TFA) over 30 minutes using a flow rate of 5 mL/min. QqQ-MS was then used to confirm the presence of the fully formed peptide and the fraction was lyophilized.

Calc'd for CMP 3 + 2H+ = 1332.67; found = 1334.1. Calc'd for CMP 3 + 4H+ = 666.84; found = 666.9. Calc'd for CMP 3 + 5H+ + Na = 538.27; found = 537.4.

UV-vis procedure:

Synthesized CMPs were dissolved in PBS buffer (0.1 M) such that the concentration was $6.4 \times 10^{-5} M$. The solution was then heated to 80 ℃ for 10 minutes in a water bath such that all collagen strands were in a monomeric form. The solution was then flash cooled to 20 C in a bath of isopropanol and dry ice before being transferred to the cuvette. Temperature was monitored via an electronic thermometer probe. This procedure ensured that any changes in fluorescence intensity would only arise from folding and not from changes in temperature of the solvent. The transition from water bath to spectrophotometer typically took 90 seconds. The UV-Vis absorption range was from 280-580 nm. These UV-vis spectroscopy studies were performed in duplicate. A peak at 455 nm showed consistent decrease in intensity and a hypsochromic shift. The UV-Vis intensity could be equated to a percentage of CMP in the folded state such that: $\%\ folded\ = \frac{Abs_t - Abs_{min}}{Abs_t}$. This was then fit to first order kinetics where $y = ln(\%\ folded\ \times\ concentration)$ and x is time in seconds. Results of these studies indicated that UV-vis is not sensitive enough to monitor the small changes in absorbance that occur following ~15-20 minutes of folding, nonetheless fitting the first 10-15 minutes of data points produced a straight line.

Fluorescence procedure:

A similar procedure was used for fluorescence spectroscopy. Synthesized CMPs were dissolved in PBS buffer (0.1 M) such that the concentration was $5 \times 10^{-6} M$. The solution was then

heated to 80 ℃ for 10 minutes in a water bath such that all collagen strands were in a monomeric form. The solution was then flash cooled to 20 C in a bath of isopropanol and dry ice before being transferred to the cuvette. Temperature was monitored via an electronic thermometer probe. The transition from water bath to fluorometer typically took 90 seconds. Samples were excited at 455 nm, and fluorescence intensity was integrated from 500 - 650 to obtain values F. The fluorescence intensity, F, could be equated to a percentage of CMP in the folded state such that: $\% \, folded = \frac{F_t - F_{min}}{F_t}$. This was then fit to first order kinetics where $y = ln(\% \, folded \times concentration)$ and x is time in seconds. The slope of the best fit line was linear with $R^2 > 0.98$. Experiments were run in duplicate to obtain *k* values.

CD procedure:

Synthesized CMPs were dissolved in PBS buffer (0.1 M) such that the concentration was 0.2 mM. The solution was then heated to 80 ℃ for 20 minutes in a water bath so that all collagen strands were in a monomeric form. The vial was then left in the dark to slowly cool to room temperature (> 24 h). Spectra were recorded from 190 to 250 nm at 20.0 °C. A 1.00 mm pathlength cell was used. The presence of a local maximum at ~225 nm in CD spectra was taken as evidence for polyproline-II helical character (Fig. 8). The change in ellipticity at 225 nm was monitored.

References

1. Kubyshkin, V. Stabilization of the Triple Helix in Collagen Mimicking Peptides. *Org. Biomol. Chem.* **2019**, 17 (35), 8031–8047. https://doi.org/10.1039/C9OB01646E
2. Hulgan, S. A. H.; Hartgerink, J. D. Recent Advances in Collagen Mimetic Peptide Structure and Design. *Biomacromolecules.* **2022**, 23 (4), 1475–1489. https://doi.org/10.1021/acs.biomac.2c00028.
3. Clark, C. Emory chemists invent shape-shifting nanomaterial. **2021**. https://news.emory.edu/features/2020/12/esc-shape-shifting-nanomaterial/index.html
4. Bachmann A, Kiefhaber T, Boudko S, Engel J, Bächinger HP. Collagen triple-helix formation in all-trans chains proceeds by a nucleation/growth mechanism with a purely entropic barrier. *Proc Natl Acad Sci USA*. **2005** Sep 27;102(39):13897-902. doi: 10.1073/pnas.0505141102.
5. Frank, S.; Boudko, S.; Mizuno, K.; Schulthess, T.; Engel, J.; Bächinger, H. P. Collagen Triple Helix Formation Can Be Nucleated at Either End. *J. Biol. Chem.* **2003**, 278, 7747–7750, DOI: 10.1074/jbc.c200698200
6. Keever, J. M.; Banzon, P.D.; Hales, M.K.; Sargent, A. L.; Allen, W. E. Association between N-Terminal Pyrenes Stabilizes the Collagen Triple Helix. *J. Org. Chem.* **2023**, 88 (16) 11885–11894. DOI: 10.1021/acs.joc.3c01175
7. Andrews, S.S.; Tretton, J. Physical Principles of Circular Dichroism. *J. Chem. Educ*. **2020**, 97, (12) 4370–4376. https://doi.org/10.1021/acs.jchemed.0c01061
8. Raja, P. M. V.; Barron, A. R. Fluorescence Spectroscopy. LibreTexts; Rice University via OpenStax CNS. https://chem.libretexts.org/Bookshelves/Analytical_Chemistry/Physical_Methods_in_Ch

emistry_and_Nano_Science_(Barron)/01%3A_Elemental_Analysis/1.11%3A_Fluorescence_Spectroscopy

9. Zhou, J.; Liu, H.; Bing, J.; Liu, X.; Fu, H.; Shangguan, D. A guanidine derivative of naphthalimide with excited state deprotonation coupled intramolecular charge transfer properties and its application. *J. Mater. Chem. C*, **2013**, 29 (1), 4399–4506. DOI: 10.1039/c3tc30853g
10. Loving, G.; Imperiali, B. A Versatile Amino Acid Analogue of the Solvatochromic Fluorophore 4-N,N-Dimethylamino-1,8-naphthalimide: A Powerful Tool for the Study of Dynamic Protein Interactions. *J. Am. Chem. Soc.* **2008**,130 (41), 13630-13638. DOI: 10.1021/ja804754y
11. Yao, Z.-J.; Gao, Y.; Voigt, J. H.; Ford, H.; Burke, T. R. Synthesis of Fmoc-Protected 4-Carboxydifluoromethyl-L-Phenylalanine: A Phosphotyrosyl Mimetic of Potential Use for Signal Transduction Studies. *Tetrahedron* **1999**, 55 (10), 2865–2874. https://doi.org/10.1016/S0040-4020(99)00076-9.
12. W. Shen, J. Ge, S. He, R. Zhang, C. Zhao, Y. Fan, S. Yu, B. Liu, Q. Zhu, A Self-Quenching System Based on Bis-Naphthalimide: A Dual Two-Photon-Channel GSH Fluorescent Probe. *Chem. Asian J.* **2017**, 12, 1532. DOI:10.1002/asia.201700340
13. Hwang ES, Thiagarajan G, Parmar AS, Brodsky B. Interruptions in the collagen repeating tripeptide pattern can promote supramolecular association. *Protein Sci.* **2010** May;19 (5):1053-64. doi: 10.1002/pro.383.