Effects of pH and organic solvents on superfolder green fluorescent protein

by
Kali Bravo

A THESIS

submitted to

Oregon State University

Honors College

in partial fulfillment of
the requirements for the
degree of

Honors Baccalaureate of Science in Biochemistry and Molecular Biology
(Honors Scholar)

Presented November 12, 2019
Commencement June 2020

AN ABSTRACT OF THE THESIS OF

Kali Bravo for the degree of Honors Baccalaureate of Science in Biochemistry and Molecular Biology presented on November 12, 2019. Title: Effects of pH and organic solvents on superfolder green fluorescent protein.

Abstract approved:_____

Wei Kong

There is no shortage of challenges in modern science, including the task of obtaining high resolution images of biomolecular structures, such as proteins. Current atomic-resolution structure determination methods include x-ray crystallography and nuclear magnetic resonance (NMR), which present issues like protein crystallization and loss of resolution with increasing size. Dr. Wei Kong is working to overcome these challenges by developing a new atomic-resolution structure determination method called single-molecule serial electron diffraction imaging (SS-EDI). However, this method presents some challenges, because it will still expose samples to extreme conditions like high temperature, low pH, and different solvents. Understanding how proteins will react in these environments is important for determining conditions that will prevent proteins from denaturing or aggregating. Furthermore, observing a protein in native form gives insight into how the protein functions in biological conditions. Here, sfGFP is exposed to low pH and increasing concentrations of organic solvents to understand the level of stability the protein is able to maintain under the conditions of SS-EDI. Our findings suggest that sfGFP exhibits a conformational change

below pH 5.4, though the process is reversible with recovery of initial pH. Additionally, sfGFP appears to aggregate in the presence of acetonitrile and the conformational effects of methanol remain undetermined.

Key Words: superfolder green fluorescent protein, single-molecule serial electron diffraction imaging, fluorescence, organic solvents, pH

Corresponding e-mail address: bravok@oregonstate.edu

Effects of pH and organic solvents on superfolder green fluorescent protein

by
Kali Bravo

A THESIS

submitted to

Oregon State University

Honors College

in partial fulfillment of
the requirements for the
degree of

Honors Baccalaureate of Science in Biochemistry and Molecular Biology
(Honors Scholar)

Presented November 12, 2019
Commencement June 2020

Honors Baccalaureate of Science in Biochemistry and Molecular Biology project of Kali Bravo presented on November 12, 2019.

APPROVED:

_____

Wei Kong, Mentor, representing Department of Chemistry

_____

Richard Cooley, Committee Member, representing Department of Biochemistry and Biophysics

_____

Nathan Waugh, Committee Member, representing Department of Biochemistry and Biophysics

_____

Toni Doolen, Dean, Oregon State University Honors College

I understand that my project will become part of the permanent collection of Oregon State University, Honors College. My signature below authorizes release of my project to any reader upon request.

_____

Kali Bravo, Author

1.Introduction

Although advancements have been made in mapping the sequences of human genomes, analyzing the function of a molecule based on its structure remains a challenge. Proteins are one of four major classes of biomolecules that constitute the maintenance and metabolic processes of living organisms. Proteins perform endless critical functions in the human body including catalysis, formation of intracellular structures, and transport of nutrients. The manner by which an amino acid sequence folds into an appropriate tertiary structure is not consistent, but depends on the environment in which the protein is located [1]. In some cases (Figure 1), the protein will fold as intended but external factors cause the molecule to denature or aggregate [2]. Furthermore, a protein's structure or environment can lead to aggregation [1], which has been found as a common mechanism in diseases like Alzheimer's and Parkinson's diseases [3]. This illustrates that the function of a protein is dependent on proper folding. Therefore, accurate, high-resolution determination of protein structure is crucial to understanding processes in the human body.

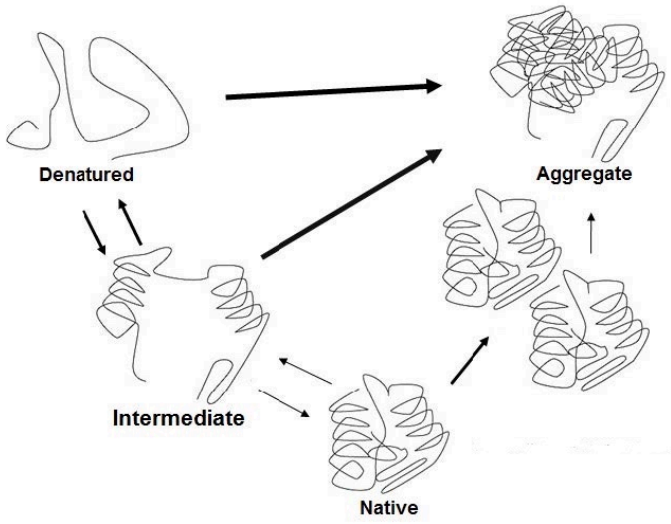


Figure 1. Possible protein conformations. Above is an illustration of an ambiguous protein folded correctly in its native form and the different ways it can become non-functional from the native state.

Currently, there are two dominant atomic-resolution methods by which to observe protein structure; x-ray crystallography and nuclear magnetic resonance (NMR). Although these methods have significantly expanded our knowledge of protein structure and folding, there are limitations to both methods that decrease resolution and in turn, the reliability of the resulting atomic structure.

X-ray crystallography utilizes crystallized proteins to produce a high-resolution diffraction pattern that can be analyzed to determine the protein structure [4]. However, fewer than half of all known proteins can form a crystal [5] and when it is possible, it can take years to discover conditions that allow for crystallization. Furthermore, the protein structure determined by x-ray crystallography is based on the protein in a lattice structure, which is not necessarily reflective of the protein structure in solution [6]. On the other hand, NMR methods analyze protein structure in solution but face different challenges. NMR has low sensitivity and thus is

limited in sample size, with the upper limit being approximately 35 kDa [7]. With proteins ranging from 50 - 2,000 amino acids [8], corresponding to a mass range of 5.5 kDa to 220 kDa, this leaves a large portion of proteins that are unable to be analyzed with NMR.

Dr. Wei Kong is working with her research group to eliminate the limitations of these protein structure determination methods by developing a technique called single-molecule serial electron diffraction imaging (SS-EDI). SS-EDI can be applied to proteins without requiring crystallization and is not limited by molecular size [5]. In SS-EDI, samples are prepared through electrospray ionization (ESI) to obtain sample ions, then cooled by embedding the sample ions into superfluid helium droplets. A polarized laser is used to create an induced electric field that orients the molecules in the same direction, at which point the diffraction data is collected. This process, shown in Figure 2, can be repeated as many as twenty times per second. Diffraction patterns from different molecular orientations, induced by a change in laser polarization, is used to produce a complete 3-D model of the sample's molecular structure [5].

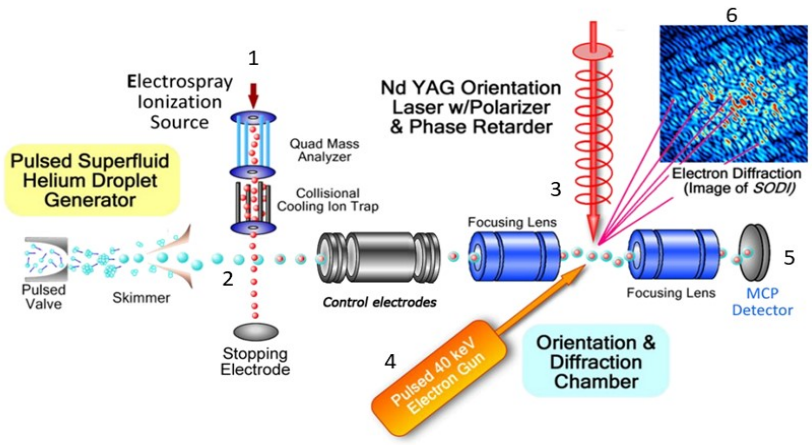


Figure 2. Mechanism of SS-EDI. (1) Sample is prepared via ESI, (2) sample is embedded into superfluid helium droplets, (3) laser induced magnetic field to align molecules, (4) diffraction of sample to obtain (6) a diffraction pattern of the structure.

Although SS-EDI overcomes many of the limitations of NMR and X-ray crystallography, there are limitations within this process itself. Proteins in solution must be able to maintain

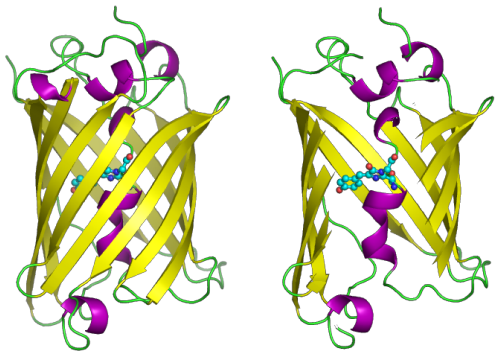
native form under the conditions of ESI, which include high temperature and low pH, for the duration of the spray process (5 -50 μs). While the spray process is relatively short, there is still potential for the protein sample to undergo conformational change (e.g. denaturation or aggregation). If this occurs, the final diffraction image will not necessarily reflect the native structure of the protein and thus loses its validity. Because of this, it is critical for us to first explore the conditions created by ESI under which the sample can remain folded and functional to ensure the final image is as representative as possible of the protein's native structure.

One protein in particular, green fluorescent protein (GFP), has presented itself as an ideal sample to use for ESI because of its stability and other unique properties. First discovered in 1961 in *Aequorea Victoria* jellyfish [9], GFP has since become a commonly used tool in medicine and molecular biology. GFP's chromophore is internally synthesized by a reaction of three amino acid residues and can be autocatalyzed. This means GFP does not require other cofactors to form the chromophore, and requires only oxygen as a substrate [10]. Although its three-dimensional structure is made up of 11 β-sheets, it is often referred to as a β-barrel that also contains a central α-helical structure that links the sheets. The barrel-like structure, shown in Figure 3 [11], is thought to protect the chromophore from being quenched of oxygen as well as attack by ions, and is credited for GFP's stability [12]. In the last 50 years since its discovery, many mutants of GFP have been created to learn more about the protein and ways in which it can be utilized. In the following experiments, we will use superfolder green fluorescent protein (sfGFP), which made its debut in 2005 [13]. Initial experiments observed that sfGFP showed greater resistance to chemical denaturants and improved folding kinetics.

Previous experiments of GFP mutants have found GFP fluorescence to be resistant to heat, high pressure, and alkaline pH [14]. Here, I have examined the stability and fluorescence of

sfGFP under varying conditions including organic solvents and pH. Results found in these experiments, in combination with previous experiments, will serve to provide the Kong group with conditions that sfGFP can withstand for the ESI step in SS-EDI. Ultimately, knowing how these conditions affect the fluorescence of sfGFP can be used to later investigate the structure of the protein in these conditions.

Figure 3. Structure of GFP variant, sfGFP. GFP's structure is composed of 11 β-sheets (yellow) and a central α-helix (purple), with the chromophore enclosed in the barrel-like structure.



2. Methods

2.1 Materials. Columns, pipettes, and experimental reagents were provided by the Mehl lab in the Biochemistry and Biophysics department at Oregon State University. A Beckman DU-64 spectrophotometer was used for all Bradford assays. Fluorescence measurements were performed using a fluorimeter built by Dr. Colin Johnson of the Biochemistry and Biophysics department at Oregon State University. Optical filters used in the fluorimeter were purchased from Thor Labs. A ROSS Glass Combination Micro pH Electrode was used to collect pH measurements, and pH calibration buffers were supplied by Michael Freitag in the Biochemistry and Biophysics department at Oregon State University. A quartz micro cuvette was purchased from Hellma Analytics.

2.2 Protein Expression. A plasmid containing the gene that encodes for sfGFP was transformed into a DH10b strain of *E. coli.* Starter cultures were grown in non-inducing media for 19 hours at 250 rpm and 37˚C in 10 mL culture tubes. These starter cultures were used to inoculate autoinduction media [15] for expression, then grown under the same conditions for 25 hours. Cells were isolated and stored at -80˚C until the following day, then resuspended in lysis buffer (30 mM HEPES, 300 mM NaCl, pH 7.4) and lysed in a Microfluidics M110P microfluidizer. Following lysis, cobalt resin was washed twice with lysis buffer and incubated with sample supernatant for a short period. The resulting cobalt resin sample was poured into an elution column and washed three times with wash buffer (30 mM HEPES, 300 mM NaCl, 5 mM imidazole, pH 7.4). A PD-10 desalting column was used to transfer the protein sample into imidazole-free lysis buffer for use in all subsequent experiments.

2.3 Protein Quantification. The Bio-Rad Bradford protein assay method was used to estimate protein concentrations. All protein samples were diluted to 50 μM in buffer and stored long-term at 4˚C.

2.4 sfGFP in organic solvents. Emission spectra of sfGFP were measured at varying dilutions in several organic solutions including 99.9% methanol (MeOH), 99.9% acetonitrile (AcN), a 1:1 v/v solution of MeOH and deionized (DI) water, and a 1:1 v/v solution of AcN and DI water. Each solution was prepared to a total volume of 100 μL, beginning with a stock solution of 50 μM sfGFP in buffer. Each experiment was performed in organic solvent concentrations spanning 100% v/v to 10% v/v in incremental dilutions of 10% v/v. A separate dilution series was performed using each of lysis buffer, MeOH, AcN, MeOH:H_2O, and AcN:H_2O. Increasing volumes of organic solution directly correlated with decreasing sfGFP concentrations (50 μM to

5 μM) in increments of 5 μM. To obtain emission spectra for each sample, 75 μL of sample were transferred to a microcuvette and placed in the fluorimeter to collect fluorescence intensity. Excitation wavelength was set at 488 nm—the major excitation wavelength for sfGFP—and the chamber was programmed to remain at 25˚C throughout the entirety of the reading (350-600 nm). Fluorescence readings for each sample were performed in duplicate. Between readings, the sample cuvette was emptied via vacuum aspiration and rinsed with DI water. This procedure was repeated for each of the five solutions, with two natural and two technical replicates for each addition.

2.5 pH Probe Calibration. The pH probe was calibrated using buffers at pH 4.01, pH 7.00, and pH 10.01 prior to all experiments. All reagents were stored in glass storage containers. Formic acid at concentration 19.1 M ± 0.2 M was diluted until linear pH decrease could be consistently observed. All samples were prepared in 1.7 mL Eppendorf microcentrifuge tubes. Beginning with 400 μL of buffer, small increments of acid were added until pH was well below the lower end of the buffer range (pH 6.8). Solution pH was measured following each acid addition. This process was subsequently repeated in MQ water. Following addition of acid to the buffer solution, NaOH was added stepwise to the sample until initial pH returned.

2.6 sfGFP in acid. The procedure for probe calibration was repeated with sfGFP present in the buffer solution, beginning with 400 μL of 50 μM sfGFP in buffer. Small increments of acid were added, increasing the effective [FA], until visible green fluorescence could not be observed. With each addition, pH was measured using the pH probe. The fluorescence of the sample was subsequently transferred to a cuvette and measured in the fluorimeter. After the measurement was collected, the sample was returned to the microcentrifuge tube for the next addition of acid.

2.7 sfGFP fluorescence recovery with base. When loss of qualitative fluorescence was achieved, the solution appeared translucent with no visible green tint. At this point, increasing volumes of NaOH were added until initial pH of sfGFP in buffer was returned. With each addition of base, the solution pH and fluorescence intensity were measured.

2.8 Concentration and pH calculations. The following procedure was used to calculate the expected pH, given a known concentration of acid. First, the concentration of acid after it is added to solution (C_2) was calculated using the following equation:

$$(Step\ 1)\ C_1V_1 = C_2V_2 \implies C_2 = \frac{C_1V_1}{V_2}$$

where C_1 is the concentration of the acid stock solution (1.912 x 10^{-4} M), V_1 is the volume of the acid stock added to a given volume of water, and V_2 is the volume of water plus the volume of acid. Solving for C_2, which will be later represented as [HA], we can use the following equation to calculate pH:

$$(Step\ 2)\ pH = -\log[H_3O^+]$$

Although the concentration of acid is known, [H_3O^+] represents the concentration of hydronium ions present at equilibrium, which is unknown. The next step is to incorporate the acid dissociation constant (K_a) equation (Step 3) to solve for an expression for [H_3O^+]. The equation for K_a is based on the reaction equation of a given solution at equilibrium; a generic example is below. Values for K_a of any given acid are known, and the K_a for formic acid is 1.8 x 10^{-4}.

$$HA + H_2O \leftrightarrows A^- + H_3O^+$$

In this reaction, we assume the concentration of water remains constant and thus can simplify the equation.

$$(Step\ 3)\ K_a = \frac{[H_3O^+][A^-]}{[HA][H_2O]} \Longrightarrow [H_3O^+] = \frac{K_a[HA]}{[A^-]}$$

Removing $[H_2O]$ from Step 3, we can solve for $[H_3O^+]$ and incorporate this expression into Step 1;

$$(Step\ 4)\ pH = -log\frac{K_a[HA]}{[A^-]}$$

The values of $[H_3O^+]$ and $[A^-]$ are the concentration of acid and conjugate base ions in solution at equilibrium. Expressions for these two values can be found using the chemical reaction equation and an ICE (initial, change, equilibrium) table (Table 1).

Chemical equation: $HA \rightleftharpoons H_3O^+ + A^-$			
	$[HA]$	$[H_3O^+]$	$[A^-]$
Initial concentration	$[HA]$	0	0
Change	$-x$	$+x$	$+x$
Equilibrium concentration	$[HA] - x$	x	x

Table 1. ICE table for the dissociation of a weak acid. ICE tables are commonly used in chemistry courses to help visualize how a weak acid dissociates in water. The expressions for "equilibrium concentration" are the variables used in the K_a equation (Step 2). This can then be used to solve for x (Step 4) and find the numerical values for equilibrium concentration of a given system with known [HA].

Substitute the equilibrium expressions in the original K_a equation, and when the equation is simplified we get a second order polynomial (Step 5):

$$(Step\ 5)\ \ K_a = \frac{(x)(x)}{[HA] - x} \Rightarrow 0 = x^2 + K_a x - K_a[HA]$$

The quadratic formula can then be used to find an expression for "x":

$$(Step\ 6)\ [H_3O^+] = [A^-] = x = \frac{-K_a \pm \sqrt{{K_a}^2 - 4(K_a)[HA]}}{2}$$

Ultimately, we want to be able to calculate the pH of a solution with a known added acid concentration [HA]. Thus, we can solve for "x" and substitute the resulting value into the equation for pH from Step 4.

$$(Step\ 7)\ pH = -\log\left(\frac{K_a([HA] - x)}{x}\right)$$

2.9 Error propagation. To calculate the error present in the calculation of expected pH, the uncertainty of each measurement must first be calculated and correctly incorporated into each calculation where the value is used. The fractional uncertainty in Step 1 is found by:

$$\frac{\delta C_2}{C_2} = \sqrt{\left(\left(\frac{\delta C_1}{C_1}\right)^2 + \left(\frac{\delta V_1}{V_1}\right)^2 + \left(\frac{\delta V_2}{V_2}\right)^2\right)}$$

Recall that C_2 = [HA], then the fractional uncertainty of Step 4 is:

$$\frac{\delta x}{x} = \frac{1}{2}\left(\frac{\delta[HA]}{[HA]}\right)$$

If we say $g = ([HA] - x)$, then the uncertainty of g is:

$$\delta g = \sqrt{\left(\frac{\delta[HA]}{[HA]} * [HA]\right)^2 + \left(\frac{\delta x}{x} * x\right)^2}$$

Thus, the theoretical uncertainty in the calculated pH is given by:

$$\delta pH = -\sqrt{\left(\left(\frac{\delta g}{g}\right)^2 + \left(\frac{\delta x}{x}\right)^2\right)}$$

3. Results

3.1 Protein expression and quantification. High yields during protein expression were made possible by a strong arabinose promoter within the pBad plasmid, the facile growth of the *E. coli* DH10b strain, and the intrinsic intracellular stability of the final sfGFP molecule. As the pBad plasmid also contains an ampicillin resistance gene, growth of *E. coli* in the presence of ampicillin ensured that only cells capable of producing sfGFP via plasmid uptake were able to reproduce. Protein was expressed and purified multiple times throughout the project as needed, with final yields ranging from 80 μM to nearly 400 μM.

3.2 sfGFP in methanol. Maximum values from emission spectra (at 513 nm) of solutions of sfGFP and methanol are plotted in Figure 4, with sfGFP in buffer referenced as a baseline. Figure 4 shows initial fluorescence of sfGFP in the presence of MeOH was greater than in the presence of buffer from 50-40 μM. The slope of the MeOH series was $3.3 \times 10^4\ \mu M^{-1}$ (R=0.9677) and the slope of the buffer series was $2.5 \times 10^4\ \mu M^{-1}$ (R=0.9809). It can also be observed in Figure 4 that there is a consistent loss of fluorescence intensity of the addition series with MeOH:H_2O (R=0.9853) that mirrors the trend of fluorescence of sfGFP in buffer. However,

the series in MeOH differs from this trend and maintains greater fluorescence intensity than in buffer. This suggests that the additions of MeOH had less of an effect on the protein relative to MeOH:H_2O, which is reflected in the fluorescence of protein under these conditions.

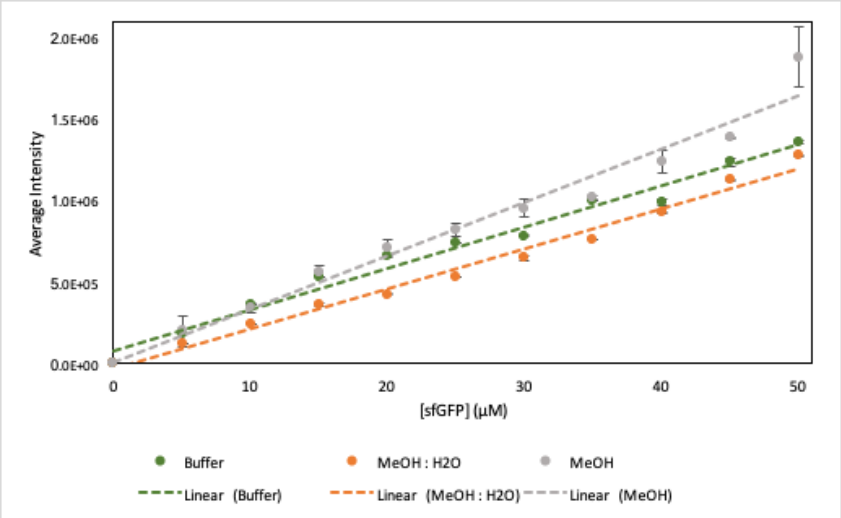


Figure 4. Peaks of emission spectra of sfGFP with methanol solution additions. The intensity value from each emission spectra at 513 nm is plotted above, with the measurements in buffer as a reference point. All measurements were taken at 25˚C, excitation at 488 nm. Decreasing [sfGFP] corresponds to increasing [MeOH].

3.3 sfGFP in acetonitrile. Figure 5 shows the trend in fluorescence of sfGFP as a function of sfGFP concentration in the presence of buffer, AcN, or AcN:H_2O. For sfGFP in AcN, there is no decrease in fluorescence from 50-35 μM, where it is approximately equal to fluorescence of sfGFP in buffer. The fluorescence then decreases rapidly at 20 μM, relative to the buffer series, with an inflection point at 15 μM. This trend is distinct from that of sfGFP in buffer or

AcN:H_2O, both of which show a consistent loss of fluorescence. Furthermore, the addition series in AcN:H_2O retains greater fluorescence intensity than the buffer series across all measurements.

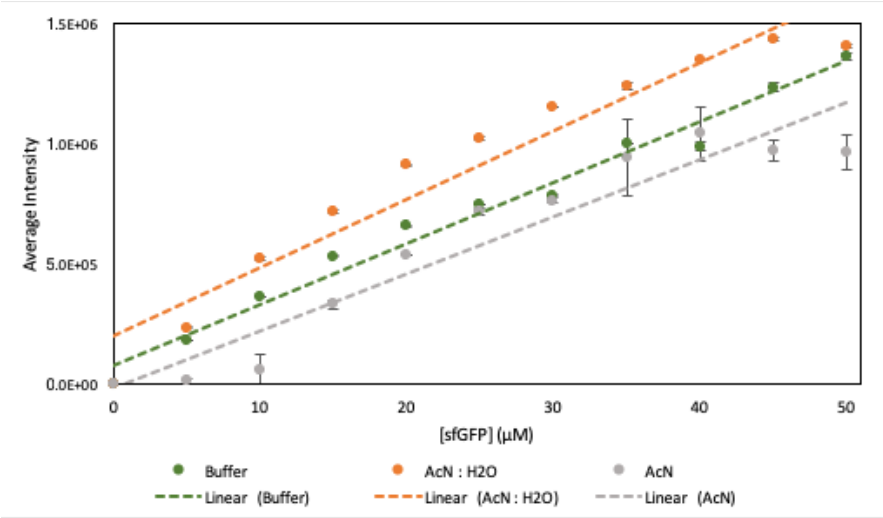
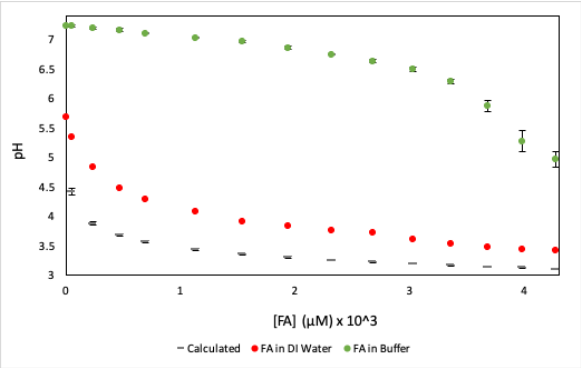


Figure 5. Peaks of emission spectra of sfGFP with acetonitrile solution additions. The intensity value from each emission spectra at 513 nm is plotted above, with the measurements in buffer as a reference point. All measurements were taken at 25˚C, excitation at 488 nm. Decreasing [sfGFP] (µM) corresponds to increasing [AcN].

3.4 pH Probe Calibration. Figure 6a shows 14 additions of acid in lysis buffer and deionized (DI) water. Calculated and measured pH values in water are significantly different. The discrepancy between these measured and calculated pH values is much smaller in Figure 6b, which shows the same procedure repeated with MQ water instead of DI water.

(a) 

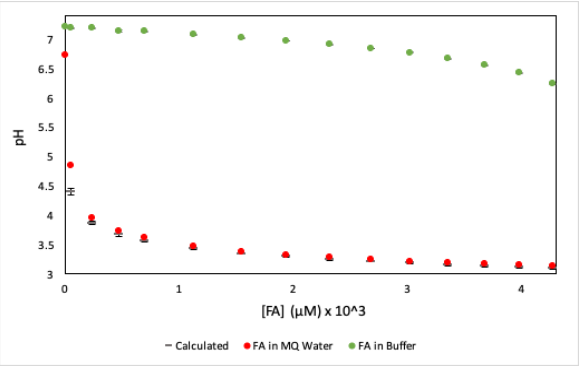
(b) 

Figure 6. Effects of acid in buffer and water (a) Calculated (expected) and measured pH values in DI water and buffer. (b) Acid in MQ water and buffer. Dilution of lysis buffer with 1.912 x 10^{-4} M FA compared with MQ water diluted with FA. Expected (calculated) pH values in MQ water are also plotted, with error bars on experimental values.

Subsequent to the addition of acid (Figure 6b), increasing volumes of NaOH were added until pH 7.0 was reached (Figure 7). Figure 7a shows the stable decrease of pH in buffer, and the recovery of pH with base follows the same trend. The pH in water drops rapidly from pH 7 to pH 4 after just two additions of acid. Then the decrease in pH is stable until the solution is at pH 3.

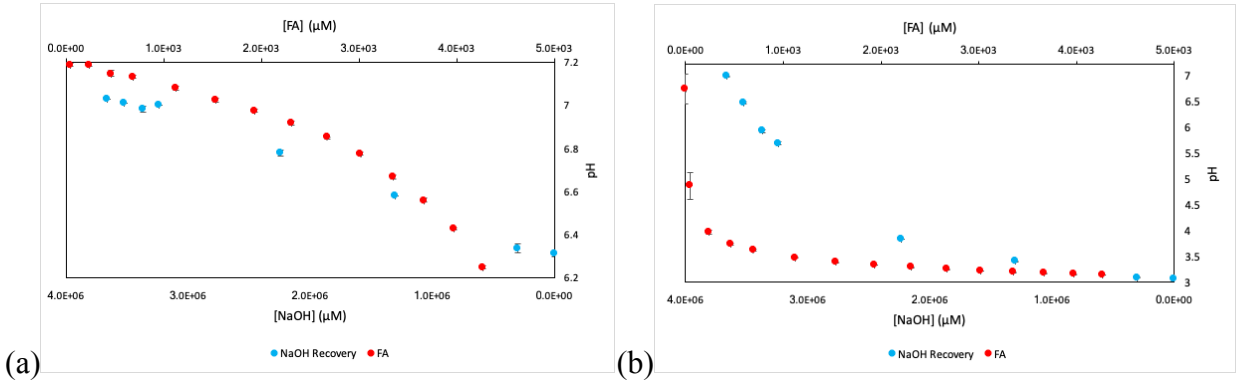


Figure 7. Effects of FA and NaOH on pH and recovery. (a) pH trends of lysis buffer diluted with formic acid, followed by additions of NaOH until neutral pH is recovered. Note that increasing [FA] correlates to decreasing pH, and increasing [NaOH] correlates to increasing pH. (b) pH trend of MQ water diluted with increasing volumes of FA, followed by additions of NaOH until neutral pH is recovered.

3.5 sfGFP in acid. Figure 8 shows the results of increasing concentrations of formic acid on sfGFP in solution. Figure 8a shows a one-third decrease in final fluorescence of sfGFP in buffer as well as [sfGFP]. In contrast, Figure 8b shows complete loss of sfGFP fluorescence in the presence of formic acid. The fluorescence decrease in Figure 8b is steady from 50 µM – 47 µM sfGFP, after which it decreases in large steps until measurable fluorescence is nearly zero.

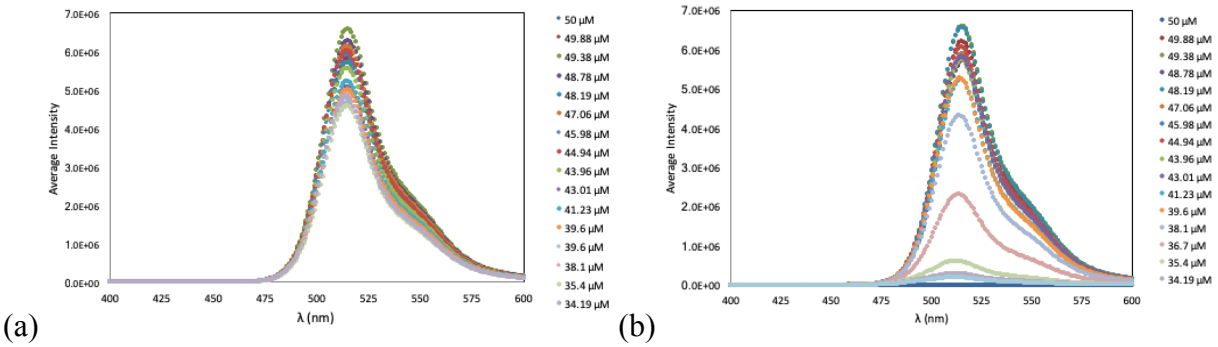


Figure 8. Emission spectra of sfGFP in buffer or formic acid. (a) Emission spectra of sfGFP diluted with buffer shows a steady decrease in fluorescence as [sfGFP] decreases. (b) Emission spectra of sfGFP diluted with increasing volumes of FA show a much more dramatic decrease in fluorescence.

3.6 sfGFP fluorescence recovery with base. Upon loss of fluorescence, NaOH was added until initial pH was recovered. However, visible green fluorescence was recovered at an earlier point. Emission spectra were collected from each sample, with natural and technical replicates (Figure 9). The final concentration of sfGFP was 16.2 µM ± 0.5 µM (32% of the initial concentration), and the final fluorescence intensity in buffer was ~2.25 x 10^6 (39% of initial intensity). These trends, of decrease in fluorescence and concentration, behave similarly.

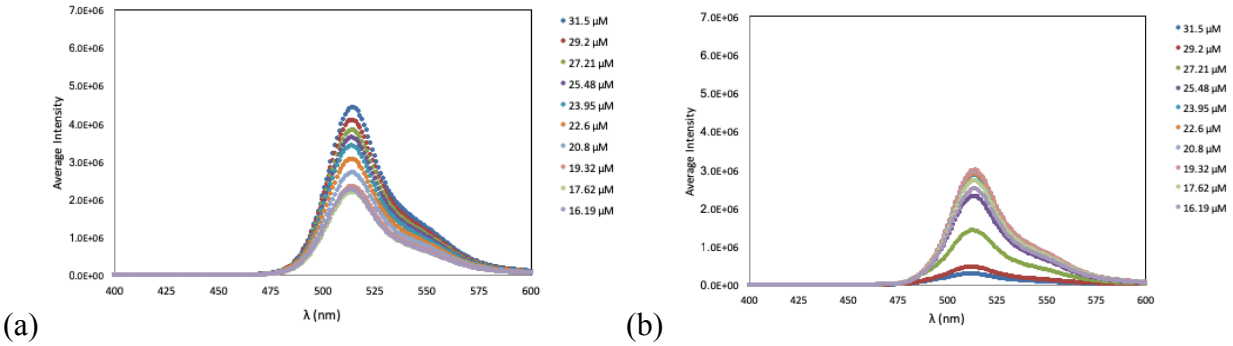


Figure 9. Emission spectra of sfGFP in buffer/formic acid diluted with sodium hydroxide. (a) Emission spectra of sfGFP in buffer shows a similar steady decrease in fluorescence as Figure 4a. (b) Emission spectra of sfGFP with FA and increasing volumes of NaOH shows an increase in fluorescence intensity, then a slight decrease as the [sfGFP] continues to decrease.

Fluorescence intensity maxima from sfGFP emission spectra (Figure 4, Figure 5) at 513 nm were used to plot intensity values compared to measured pH values (Figure 10). Fluorescence and pH curves (Figure 10) were normalized to measurements of sfGFP in buffer. The intensity readings exhibit some degree of variance until sfGFP drops to approximately 41 µM ± 0.5 µM, after which the readings steadily decrease in correlation with pH. Upon additions of base, both pH and fluorescence steadily increase. From approximately 21 µM – 16 µM, sfGFP intensity decreases due to the effects of dilution of the protein. These intensity values are consistent with those of sfGFP in buffer (Figure S1). The pH of sfGFP in buffer remained at 7.14 ± 0.01.

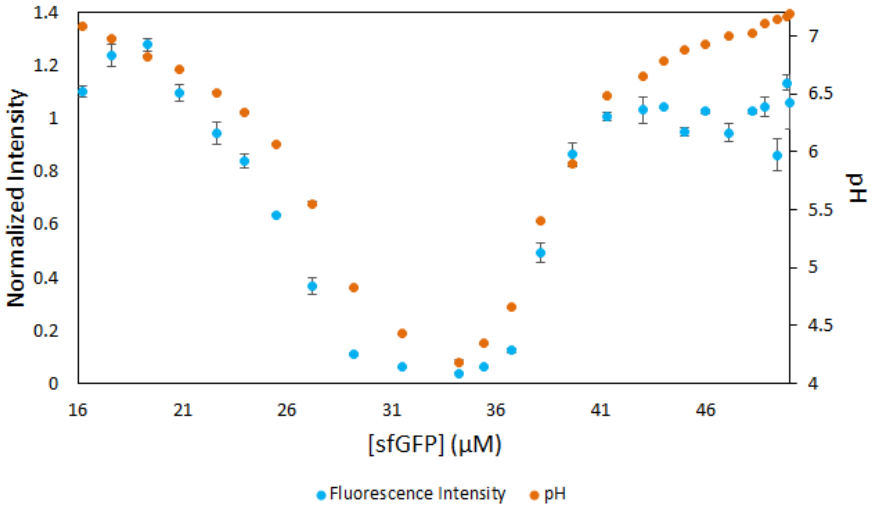


Figure 10. Fluorescence loss and recovery of sfGFP with acid and base. Fluorescence intensity peaks at 513 nm following visible loss of fluorescence of sfGFP by dilution with FA, and visible recovery of fluorescence following additions of NaOH. Figure 10 also shows the pH trend following the loss and recovery of fluorescence of sfGFP.

4. Discussion

4.1 sfGFP in methanol. Experiments with 1:1 v/v MeOH:H_2O resulted in lower fluorescence intensity than with buffer, suggesting sfGFP in methanol may cause some degree of conformational change that results in a decreased yield of fluorescence. This result that was not surprising because sfGFP should theoretically be more stable in buffer than organic solvent. A buffer system is one that ideally will resist any change in pH and thus keep the protein in a stable, native form. The data from sfGFP in MeOH and sfGFP in MeOH:H_2O (Figure 4) show a higher yield of fluorescence in MeOH than MeOH:H_2O. Protein in the presence of MeOH:H_2O follows a consistent trend of loss of fluorescence intensity when compared to MeOH, noting the outlier at 50 µM. sfGFP maintains approximately 50% fluorescence through 80% v/v in both MeOH and MeOH:H_2O, suggesting that methanol may cause conformational changes in the

protein that result in a loss of fluorescence at higher concentrations of methanol. The hypothesized mechanism for the phenomena is based on methanol's size and hydrogen bonding abilities. Because methanol is a very small molecule relative to the protein, and has hydrogen bonding sites, it will have a higher affinity than the buffer to bind to the surface of sfGFP at hydrogen bonding locations. As mentioned previously, hydrogen bonding is a vital part of protein structure that constitutes the protein's tertiary structure. Thus, when hydrogen bonds are disrupted, the protein will become denatured upon losing its tertiary structure. However, the reasons behind these observations of sfGFP in the presence of methanol solutions are not immediately clear and should be further studied using SS-EDI to investigate any conformational changes resulting from these conditions.

4.2 sfGFP in acetonitrile. The results found in the AcN additions (Figure 5) can be advantageous for future experiments concerned with the fluorescence of sfGFP. It was found that sfGFP in 1:1 % v/v of AcN:H_2O displayed a higher yield of fluorescence intensity relative to sfGFP in buffer throughout all samples. However, sfGFP in the presence of AcN observed no decrease in fluorescence until 20 μM sfGFP. After this point, the trend of fluorescence follows a sigmoidal function until measured fluorescence is lost. This occurrence can likely be attributed to aggregation, a phenomenon in which the protein's environment is unfavorable and causes the macromolecules to accumulate together. In fact, the effects of high concentrations of methanol and acetonitrile were readily apparent in previous experiments (Figure S2). Phase separation was observed at high concentrations of AcN, thus the fluorescence measurements in Figure 5 may not necessarily be representative of the fluorescence of sfGFP in solution. If phase separation occurs at higher [AcN], where a loss of fluorescence was observed, we can speculate that the solution is

not homogeneous and thus the protein may not be in the light path of the fluorimeter for measurement. Further experiments to test the structure of sfGFP in 1:1 % v/v AcN:H_2O will be necessary to reach a conclusion, but it appears the acetonitrile solution increases the fluorescence of the protein, which is more favorable for biological experiments that use sfGFP as a visible marker.

4.3 pH Probe Calibration. Once a large margin of error was identified in measurements, I stopped all experiments and made two changes before repeating previous experiments. First, I changed my storage method. Initially, all reagents were stored in 5 mL polypropylene Eppendorf tubes. Although formic acid is a weak acid, some error could be attributed to the acid interacting with the plastic container and the solution would thus be altered from the original. To minimize any potential effects of solution reacting with the storage vessel, I replaced all stock solutions and stored them in glass containers. Second, I changed my water source from DI water to MQ water. The amount of hydrogen ions (protons) present in solution is measured by a pH electrode. While DI water contains minimal ions, there are still ions and other uncharged molecules that may interfere with this reading and alter the pH measurement. These effects were significant, as noted by the difference between the two plots in Figure 6. Therefore, I chose to use MQ water for my remaining experiments to minimize error.

4.4 sfGFP in acid. Experiments with formic acid resulted in a complete loss of fluorescence compared to ~30% loss of fluorescence in experiments with buffer. Loss of fluorescence is first observed qualitatively when the sample transforms from luminescent green to translucent. However, even after fluorescence is not visible to the human eye, the fluorimeter is able to

collect quantitative fluorescence. For this reason, pH measurements were used in tandem with intensity measurements to ensure the protein was well below its stable pH range in addition to losing 100% fluorescence. According to Figure 10, sfGFP can maintain 50% fluorescence down to pH 5.4, as reflected in recent literature that states sfGFP fluorescence is stable as low as pH 5.5 [16].

4.5 sfGFP fluorescence recovery with base. Upon reaching pH 4.13 with sfGFP in formic acid, recovery of initial pH with sodium hydroxide resulted in 100% recovery of sfGFP fluorescence. This suggests the loss of fluorescence in the protein is a reversible conformational change of the β-barrel, but the chromophore structure remains intact. The preservation of chromophore structure can likely be attributed to the tight hydrogen bond network interaction with the β-barrel [17]. More extensive experiments can be conducted to confirm the effects of reversible denaturation on the structure of sfGFP with SS-EDI, now that we have confirmed 100% fluorescence recovery is possible after complete loss of fluorescence. For example, a sample of sfGFP could be measured in SS-EDI at different stages of these additions to observe the structure in different pH environments. If the hypothesis is correct, the protein's β-barrel undergoes a conformational change that results in the loss of fluorescence, but the structure remains largely intact. This would be reflected in the results of SS-EDI for sfGFP in pH 7.0, pH 5.5, and pH 4.5.

It would also be beneficial to explore the effects of pH on sfGFP in acetonitrile compared to buffer, taking into account that 1:1 % v/v AcN:H_2O resulted in an increased fluorescence intensity of the protein. I hypothesize that the protein in acetonitrile would maintain a higher

fluorescence as pH decreases, but the protein would have difficulty refolding due to the increased number of foreign molecules in solution.

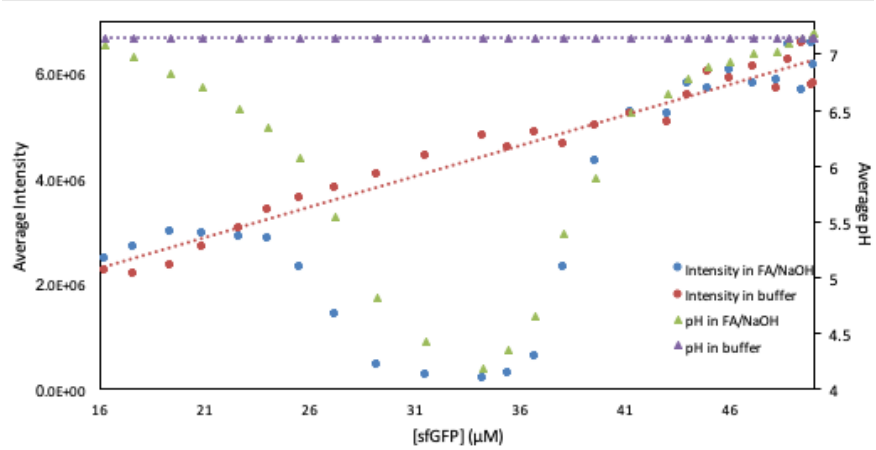
5. Conclusion

In the case of organic solvents, it appears that AcN at lower concentrations results in an increased yield of fluorescence, though higher concentrations may overwhelm the protein and cause aggregation that results in phase separation. Still unknown is the effect of methanol and acetonitrile on the conformation of sfGFP, and whether these effects are severe enough to denature to protein at high concentrations. Additionally, it has been observed that loss of fluorescence via acid is fully recoverable as low as pH 4.0, thus the effect of pH on the structure of the protein is reversible. However, it remains unknown if the effect of pH on structure is reversible when sfGFP is placed in pH lower than pH 4.0. Finally, previous experiments [18] have shown that sfGFP loses fluorescence at 80˚C, but fluorescence is fully recovered when the sample is returned to 25˚C. Further experiments should investigate if fluorescence is fully recoverable when sfGFP is subjected to temperatures higher than 80˚C and if these measurements are time-dependent.

Ultimately, we have gained a better understanding of the behavior of sfGFP in different environments of organic solvents, high temperature, and changing pH, but the mechanism behind these observations remains unknown. The next step is to resolve these unknown mechanisms, with SS-EDI and any other methods that may uncover more information about these processes and the stable structure of sfGFP.

6. Supplemental Figures

Figure S1. Fluorescence intensity of sfGFP in buffer and in changing pH environment. Included in Figure S1 is the non-normalized data from Figure 10 to compare the fluorescence of sfGFP in native conditions (buffer) to sfGFP in changing pH.



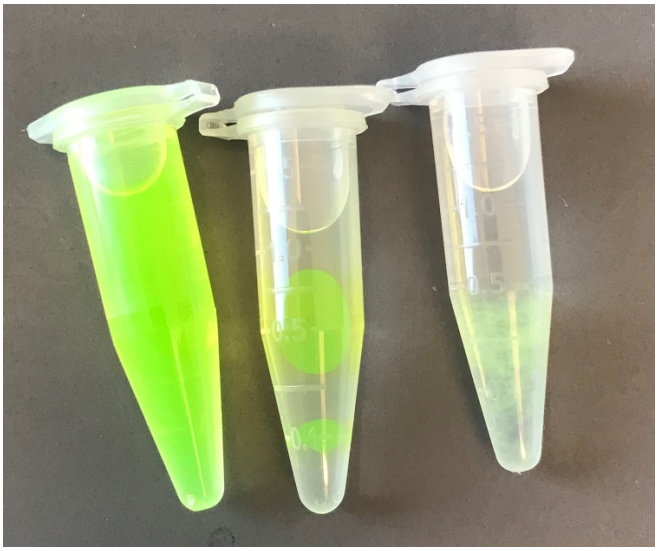


Figure S2. Effect of acetonitrile and methanol on sfGFP. The protein is shown in buffer (left), acetonitrile (center), and methanol (right). sfGFP in acetonitrile (center) visibly gathers in two large aggregates. sfGFP in methanol (right) appears to also aggregate, though on a much smaller scale. This effect could also be attributed to denaturation.

7. References

1. Dobson, C.M., *Protein folding and misfolding.* Nature, 2003. **426**(6968): p. 884.
2. Lin, S.-Y., *Salmon calcitonin: conformational changes and stabilizer effects.* 2015.
3. Ross, C.A. and M.A. Poirier, *Protein aggregation and neurodegenerative disease.* Nature medicine, 2004. **10**(7s): p. S10.
4. Stallforth, P. and J. Clardy, *X-ray crystallography: One size fits most.* Nature, 2013. **495**(7442): p. 456.
5. Beckman, J., et al., *Apparatus and method for determining molecular structure*. 2016, Google Patents.
6. Acharya, K.R. and M.D. Lloyd, *The advantages and limitations of protein crystal structures.* Trends in pharmacological sciences, 2005. **26**(1): p. 10-14.
7. Elipe, M.V.S., *Advantages and disadvantages of nuclear magnetic resonance spectroscopy as a hyphenated technique.* Analytica Chimica Acta, 2003. **497**(1-2): p. 1-25.
8. Alberts, B., et al., *The shape and structure of proteins*, in *Molecular Biology of the Cell. 4th edition*. 2002, Garland Science.
9. Shimomura, O., F.H. Johnson, and Y. Saiga, *Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea.* Journal of cellular and comparative physiology, 1962. **59**(3): p. 223-239.
10. Scott, D.J., et al., *A novel ultra-stable, monomeric green fluorescent protein for direct volumetric imaging of whole organs using clarity.* Scientific reports, 2018. **8**(1): p. 667.

11. contributors, W. *Green fluorescent protein*. 2019 19 October 2019 [cited 2019 21 October]; Available from: https://en.wikipedia.org/w/index.php?title=Green_ fluorescent_protein&oldid=922099268.

12. Zimmer, M., *Green fluorescent protein (GFP): applications, structure, and related photophysical behavior.* Chemical reviews, 2002. **102**(3): p. 759-782.

13. Pédelacq, J.-D., et al., *Engineering and characterization of a superfolder green fluorescent protein.* Nature biotechnology, 2006. **24**(1): p. 79.

14. Ehrmann, M.A., C.H. Scheyhing, and R.F. Vogel, *In vitro stability and expression of green fluorescent protein under high pressure conditions.* Letters in applied microbiology, 2001. **32**(4): p. 230-234.

15. Studier, F.W., *Stable expression clones and auto-induction for protein production in E. coli*, in *Structural Genomics*. 2014, Springer. p. 17-32.

16. Roberts, T.M., et al., *Identification and Characterisation of a pH-stable GFP.* Scientific reports, 2016. **6**: p. 28166.

17. Craggs, T.D., *Green fluorescent protein: structure, folding and chromophore maturation.* Chemical Society Reviews, 2009. **38**(10): p. 2865-2875.

18. Ho, L., *Stability of Super Folder Green Fluorescent Protein Under Varying Organic Solvent and Temperature Conditions.* 2019.

8. Acknowledgments

I would like to take this opportunity to thank Dr. Ryan Mehl and Mehl Lab members for their assistance and allowing me access to materials for this project. Additionally, I would like to thank Dr. Colin Johnson for granting me access to the fluorimeter he built. Both the Mehl and Johnson labs have been welcoming and willing to help me throughout this process, and this project would not have been possible without them.

I would also like to thank a few individuals who have been instrumental to my Honors thesis.

Lylan Ho, my undergraduate research partner, joined me on this project about the time I realized how much this project involved. If I had continued this project by myself, I would likely have not had time to investigate most of what is included in my thesis. It was a joy to watch Lylan's thesis process and defense while continuing my own work on this project.

Rick Cooley, one of my committee members, was nothing short of happy to participate as a member of my committee. I did not spend much time working with Rick, but I am beyond grateful that he was willing to invest time in me and my academic career.

Nathan Waugh, one of my committee members, advised me throughout this project and taught me what it is to be a researcher. Nathan was present for all my excitements but also frustrations regarding this project. Most importantly to me, he always empowered me to learn, rather than giving me simple answers. I am grateful for his encouragement and the ways that he challenged me, and I look forward to seeing all that Nathan will accomplish in the future.

Wei Kong, my research mentor, taught me what it is to be a scientist that is hard-working, passionate, and kind. Wei has become a role model to me and I am so grateful that she shared her work with me and allowed me the opportunity to take this project and grow it into one that her graduate students can continue and publish in the future. Working with Wei has made me a better student, a better scientist, and a better person.