

THE PENNSYLVANIA STATE UNIVERSITY  
SCHREYER HONORS COLLEGE

DEPARTMENT OF CHEMISTRY

SYNTHESIS OF FLUOROGENIC PROBES FOR CLIP-TAG  
PROTEIN LABELING BASED ON GFP AND RFP FLUOROPHORES  
& OPTIMIZATION OF RFP MIMIC FLUOROPHORE SYNTHESIS

SOJUNG KIM  
SPRING 2020

A thesis  
submitted in partial fulfillment  
of the requirements  
for a baccalaureate degree  
in Chemistry  
with honors in Chemistry

Reviewed and approved\* by the following:

Xin Zhang  
Assistant Professor of Chemistry  
Thesis Supervisor

Pshemak Maslak  
Associate Professor of Chemistry  
Honors Adviser

\* Electronic approvals are on file.

## ABSTRACT

Abnormal protein aggregation has been linked to multiple neurodegenerative diseases; however, elucidating the biological roles of these protein aggregates has remained challenging due to a lack of available techniques for visualizing protein aggregates in living cells. This work reports progress towards the development and synthesis of CLIP-tag probes using Green Fluorescent Protein and Red Fluorescent Protein based fluorophores to enable simultaneous labeling of insoluble protein aggregates and soluble protein oligomers in living cells. Five different GFP fluorophore based CLIP-tag probes were synthesized to test a series of different linker moieties in the probe (C2, C4, C6, cyclohexane, and proline). Though purification of these probes via preparatory HPLC proved challenging, the synthesis of these GFP based CLIP-tag probes will enable quantification of their fluorogenic properties and selection of the probe that exhibits the most desirable characteristics for use in *in vivo* protein aggregation imaging. To enable future design and synthesis of RFP fluorophore based fluorogenic probes, the synthesis of the desired RFP fluorophore was optimized. Anhydrous conditions with 0.1 eq. loading of  $\text{BF}_3 \cdot \text{Et}_2\text{O}$  as a Lewis acid catalyst improved yield of the key enamine condensation from approximately 25% to 55%, with improved product to starting material ratio to also increase ease of purification. These optimized conditions will allow for more efficient future large-scale synthesis of the RFP fluorophore to enable synthesis of a series of RFP based CLIP-tag probes. Ultimately, developing such tools to enable the study of protein aggregation in living cellular environments may help elucidate how such protein aggregation causes biological effects, which is of particular interest in the study of neurodegenerative diseases.

**TABLE OF CONTENTS**

LIST OF FIGURES .....	iii
LIST OF TABLES .....	v
ACKNOWLEDGEMENTS .....	vi
Chapter 1 Background .....	1
Fluorogenic Probes for Visualization of Abnormal Protein Aggregation .....	1
Synthesis of RFP Mimic Fluorophore .....	5
Chapter 2 Synthesis of GFP-Benzylcytosine Fluorogenic Probes for CLIP-tag .....	7
Synthesis of O6-Benzylcytosine .....	7
Synthesis of Linker Molecules.....	11
Synthesis of GFP Fluorophore.....	13
Synthesis of Final GFP-Linker-Benzylcytosine Probes .....	15
Purification of Final GFP-Linker-Benzylcytosine Probes.....	17
Chapter 3 Optimization of RFP Mimic Fluorophore Synthesis.....	21
Chapter 4 Conclusions & Future Directions .....	34
Appendix A Experimental .....	37
Appendix B Supplemental Information .....	47
<sup>1</sup> H NMR Data.....	48
HPLC Traces.....	65
BIBLIOGRAPHY.....	72

**LIST OF FIGURES**

Figure 1. General Structure of GFP Based CLIP-tag Probes .....	4
Figure 2. GFP and RFP Based Fluorophores.....	5
Figure 3. Synthetic Scheme for O6-Benzylcytosine.....	8
Figure 4. HPLC Trace for O6-Benzylcytosine .....	10
Figure 5. Synthetic Scheme for Linker Molecules .....	12
Figure 6. Synthetic Scheme for GFP Fluorophore .....	13
Figure 7. Synthetic Scheme for GFP-Cyclohexane-BC .....	15
Figure 8. Mechanism of Peptide Coupling .....	16
Figure 9. Final GFP-Linker-BC Probes.....	17
Figure 10. HPLC Purification of GFP-Pro-BC.....	19
Figure 11. Synthetic Scheme for RFP Fluorophore.....	21
Figure 12. Proposed Mechanism of RFP Synthesis.....	23
Figure 13. Annotated <sup>1</sup> H NMR of GFP/RFP Mixture .....	26
Figure 14. TLC Analysis of ZnCl <sub>2</sub> , AlCl <sub>3</sub> , and TiCl <sub>4</sub> Conditions.....	30
Figure 15. TLC Analysis of Zn(OAc) <sub>2</sub> Conditions.....	31
Figure 16. TLC Analysis of MgCl <sub>2</sub> Conditions .....	32
Figure 17. RFP-Benzylcytosine Probe .....	33
Figure 18. Annotated <sup>1</sup> H NMR of Protected O6-Benzylcytosine (6).....	48
Figure 19. Annotated <sup>1</sup> H NMR of O6-Benzylcytosine (7). .....	49
Figure 20. Annotated <sup>1</sup> H NMR of C4 Linker (9).....	50
Figure 21. Annotated <sup>1</sup> H NMR of C6 Linker (11).....	51
Figure 22. Annotated <sup>1</sup> H NMR of Cyclohexane Linker (15). .....	52
Figure 23. Annotated <sup>1</sup> H NMR of Protected GFP Fluorophore (22).....	53
Figure 24. Annotated <sup>1</sup> H NMR of Deprotected GFP Fluorophore (23). .....	54

Figure 25. Annotated $^1\text{H}$ NMR of GFP-Cyclohexane-Benzylcytosine (25). .....	55
Figure 26. Annotated $^1\text{H}$ NMR of GFP-C2-Benzylcytosine (26).....	56
Figure 27. Annotated $^1\text{H}$ NMR of GFP-C4-Benzylcytosine (27).....	57
Figure 28. Annotated $^1\text{H}$ NMR of GFP-C6-Benzylcytosine (28).....	58
Figure 29. Annotated $^1\text{H}$ NMR of GFP-Pro-Benzylcytosine (29).....	59
Figure 30. Annotated $^1\text{H}$ NMR of Protected RFP Fluorophore (31).....	60
Figure 31. Annotated $^1\text{H}$ NMR of RFP-Benzylcytosine (32).....	61
Figure 32. $^1\text{H}$ NMR RFP Screen: Piperidine in Pyridine. ....	62
Figure 33. $^1\text{H}$ NMR RFP Screen: 0.5 eq. $\text{ZnCl}_2$ with wet solvent.....	63
Figure 34. $^1\text{H}$ NMR RFP Screen: 0.1 eq. $\text{BF}_3 \cdot \text{Et}_2\text{O}$ . ....	64
Figure 35. Reverse Phase HPLC Trace for O6-Benzylcytosine Purification (7). ....	65
Figure 36. Reverse Phase HPLC Traces for Protected GFP/RFP Purification (22/31). ....	66
Figure 37. Reverse Phase HPLC Trace for GFP-Cyclohexane-Benzylcytosine Purification (25).67	
Figure 38. Reverse Phase HPLC Traces for GFP-C2-Benzylcytosine Purification (26). ....	68
Figure 39. Reverse Phase HPLC Traces for GFP-C4-Benzylcytosine Purification (27). ....	69
Figure 40. Reverse Phase HPLC Traces for GFP-C6-Benzylcytosine Purification (28). ....	70
Figure 41. Reverse Phase HPLC Traces for GFP-Pro-Benzylcytosine Purification (29). ....	71

**LIST OF TABLES**

Table 1. RFP Synthesis Condition Screen.....	24
----------------------------------------------	----

## ACKNOWLEDGEMENTS

I would like to thank Prof. Xin Zhang for supervising my thesis research and supporting my learning, scientific growth, and research endeavors throughout this project. I would also like to thank Kwan Ho Jung, who served as my graduate student mentor, and the whole Zhang lab for their willingness to answer questions and provide support on a day to day basis.

I would also like to acknowledge the Millennium Scholars Program, both staff and students, for supporting me and guiding me throughout my undergraduate career, as well as the Eberly College of Science for giving me multiple opportunities to travel to conferences to present my research and study abroad to enhance my scientific systems thinking. Finally, I would also like to thank Prof. Tom Mallouk, who was my first undergraduate research supervisor for my freshman and sophomore years. Though my research with the Mallouk lab is not detailed in this thesis, the scientific skills learned during my early years of training continue to drive my research career.

This material is based upon work supported by the Burroughs Welcome Fund Career Award at the Scientific Interface (X.Z.), Paul Berg Early Career Professorship (X.Z.), Lloyd and Dottie Huck Early Career Award (X.Z.), Alfred P. Sloan Research Fellowship (X.Z.), Pew Biomedical Scholars Program (X.Z.), and National Institute of Health R35 GM133484 (X.Z.). Any opinions, findings, and conclusions or recommendations expressed in this thesis are those of the author, and do not necessarily reflect the views of the National Institute of Health or other awarding agencies.

## Chapter 1

### Background

#### **Fluorogenic Probes for Visualization of Abnormal Protein Aggregation**

Abnormal protein aggregation involving the assembly of misfolded proteins into oligomeric and higher order structures has been linked to multiple neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and various prion diseases.<sup>1,2</sup> However, elucidating the biological mechanisms and roles of these abnormal protein aggregates in neurodegenerative diseases has remained challenging because of the lack of available techniques for detecting and visualizing these protein aggregates in living cells.<sup>1</sup> Though protein aggregation can be more easily studied under controlled conditions in a test tube, studying protein aggregation in the complex cellular environment of a living cell is essential to understanding its role in neurodegenerative diseases due to the complex interactions of proteins with other molecules that affect aggregation (e.g. chaperonins, proteases).<sup>3</sup>

To this end, several fluorescence imaging techniques have been developed to study protein aggregation in live cells, including many methods to genetically fuse fluorescent proteins to proteins of interest in order to enable fluorescent visualization of aggregation behavior.<sup>1,3</sup> The ability to detect and visualize these aggregated proteins enables studies to elucidate protein aggregation behavior, dynamics, and operation in the complex environment of living cells.<sup>4</sup> One drawback of this technique, however, is that these fluorescent proteins are always fluorescent, making it difficult to distinguish folded proteins from oligomers and aggregates.<sup>1</sup> To overcome

this limitation, more recent techniques have focused on the development of fluorogenic probes that only turn-on fluorescence when proteins aggregate.

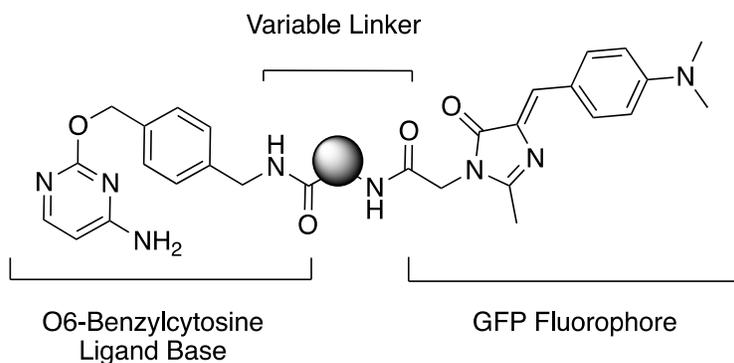
For example, Zhang et al. have developed AggTag, an imaging technique that enables the visualization and detection of protein aggregates (both soluble oligomers and insoluble aggregates) by ensuring that fluorescence only turns on when proteins aggregate.<sup>5</sup> In this technique, a protein tag called Halo-tag is fused to the protein of interest. This protein tag can be labeled with a Halo-tag ligand, and this ligand can be modified to contain a fluorophore. For AggTag, a GFP (green fluorescent protein) chromophore was used because the fluorophore itself is a molecular rotor. These fluorophores do not normally fluoresce when freely dissolved in dilute solution because excitation energy is dissipated via twisted-intramolecular charge transfer, wherein free bond rotation in the structures allows for energy dissipation via conformational isomerization. This bond rotation acts as a non-radiative relaxation pathway that quenches fluorescence. When proteins aggregate, however, the microenvironment around the fluorophore becomes rigid, inhibiting free rotation and locking the fluorophore into one conformation. This removes the non-radiative twisted-intramolecular charge transfer energy dissipation and turns on fluorescence.<sup>6</sup> Thus, the protein of interest can be bio-orthogonally conjugated with a fluorogenic probe that only turns on fluorescence if and when the protein aggregates.<sup>1,5</sup> This “turn-on” feature allows for protein visualization without having to wash away extra, un-bound fluorescent probes, which makes real-time analysis possible. Fluorogenicity also improves the signal to noise ratio of images collected.<sup>7</sup>

More recently, Zhang et al. also developed a fluorogenic AggTag method based on Halo- and SNAP-tag to simultaneously detect the aggregation of two different proteins in live cells. The orthogonal fluorescence of the two different fluorogenic probes, which also contain differently colored fluorophores, enables simultaneous detection of two different proteins, instead of one

single protein at a time.<sup>1</sup> By expanding the toolkit of fluorogenic probes available for the scientific community to use, this work enables further study of protein aggregation and how such aggregation may be causatively implicated in various neurodegenerative diseases.

To build off this development, the overall objective of this research is to expand the family of fluorogenic probes to include CLIP-tag probes. CLIP-tag is an alternative protein label that can bind O<sup>6</sup>-benzylcytosine as a ligand. Moreover, because CLIP-tag possesses orthogonal substrate specificity when compared to the previously developed SNAP-tag platform, SNAP and CLIP fusion proteins can also be used to simultaneously label different proteins in living cells for fluorescent visualization of aggregation.<sup>8</sup> Specifically, this thesis details work towards the development of a system of CLIP-tag probes using GFP and RFP (red fluorescent protein) based fluorophores to enable simultaneous labeling of insoluble protein aggregates and soluble protein oligomers in live cells. This thesis work expands upon the developed systems because the aim is to develop a fluorogenic labeling method that can detect insoluble aggregates and soluble oligomers *with distinct fluorescence signals*, rather than simply based on how diffuse or punctate fluorescent signals appear.

The general molecular structure of this family of fluorogenic probes consists of a fluorophore conjugated via peptide bond to some linker structure, which is then connected via peptide bond to the CLIP-tag ligand that is recognized by the protein, in this case, O<sup>6</sup>-benzylcytosine (Fig. 1).



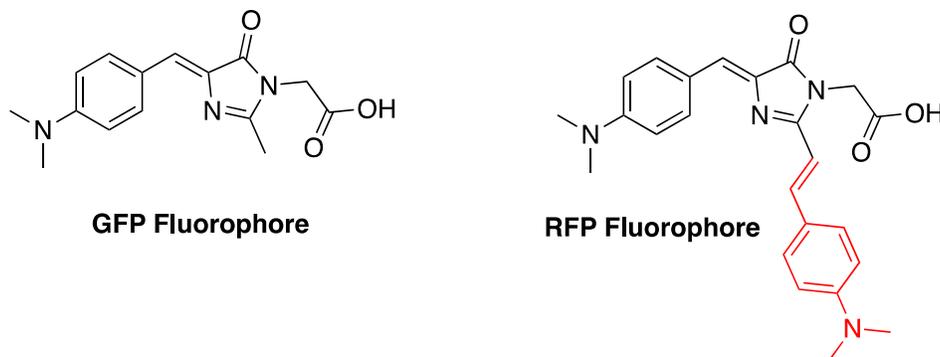
**Figure 1. General Structure of GFP Based CLIP-tag Probes**

To develop this system, a series of fluorogenic probes bearing the GFP fluorophore were synthesized using five different linkers to attach the fluorophore to the O6-benzylcytosine ligand base. Because the fluorogenic properties of the fluorophore, and hence the overall probe, depend on the ability of the fluorophore to experience dramatic changes microenvironment upon protein aggregation, the length and rigidity of the molecular structure linking the fluorophore to the ligand that binds the protein tag significantly affects fluorogenic properties of the probe. Synthesis of these probes will allow for determination of the optimal linker structure for probes with the best fluorogenic properties (e.g. largest increase in fluorescence upon aggregation, long fluorescence lifetime).

Ultimately, synthesis of these fluorogenic CLIP-tag probes will enable further investigation of their fluorescent properties, selection of the most desirable linker structure, and finally application of CLIP-tag based fluorogenic probes in live cell imaging to study abnormal protein aggregation. Introducing an additional orthogonal tag system to the SNAP-tag and Halo-tag platforms will enable more complex biological studies of multiple simultaneously labeled proteins of interest in live cells.<sup>1</sup>

## Synthesis of RFP Mimic Fluorophore

For simultaneous labeling of different proteins of interest, it is not only necessary to have orthogonal protein labeling platforms, but also necessary to use two different fluorophores that excite at different distinguishable frequencies and produce visually distinct colors upon fluorescence.<sup>1</sup> The fluorophores should ideally absorb different wavelengths of light to allow excitation and visualization of one protein at a time if desired, and must also produce visually distinct colors of fluorescence to allow researchers to visually distinguish and identify the proteins being studied. Hence, in addition to fluorogenic probes containing the GFP fluorophore (yellow fluorescence, excitation at 480 nm), recent work has been done to develop fluorogenic probes containing an RFP based fluorophore with red fluorescence (excitation at 540 nm).<sup>9,10</sup> The general structures of both fluorophores are shown in Fig. 2.



**Figure 2. GFP and RFP Based Fluorophores**

The RFP fluorophore is a derivative of the GFP fluorophore with the addition of a dimethylamino phenyl moiety to the original amide heterocycle via a C=C double bond (highlighted in red in Fig. 2). The addition of this moiety increases conjugation of the whole molecule and decreases the HOMO-LUMO energy gap. This decrease in the HOMO-LUMO

energy gap explains why the RFP fluorophore absorbs light at longer wavelengths (lower energy), and also why the fluorophore emits at longer wavelengths that are red-shifted.

Synthesis of the RFP fluorophore has been previously reported to proceed in one step from the *boc*-protected form of the GFP fluorophore via a Lewis acid catalyzed enamine condensation with dimethylamino benzaldehyde.<sup>10</sup> Using the reported procedures, the original goal of this thesis was to synthesize five RFP based fluorogenic CLIP-tag probes in addition to the five GFP based probes in order to develop fluorogenic probes of two different colors. However, in the midst of the synthetic efforts, it was realized that the current synthetic conditions for the RFP fluorophore needed significant optimization for the desired RFP derivative with the dimethylamino group. Not only was conversion of starting material to product very low, but the precedented conditions also simultaneously degraded some of the desired product, and similar  $R_f$  values between starting material and product precluded efficient purification.

Therefore, the second portion of this thesis focused on screening conditions for the synthesis of the RFP fluorophore from the GFP starting material. A higher yielding synthetic route to the RFP fluorophore would enable more efficient, large-scale synthesis of this desirable fluorophore, ultimately enabling straightforward and accessible synthesis of RFP based fluorogenic probes for future use not only with CLIP-tag probes, but also any compatible fluorogenic probe platform.

## Chapter 2

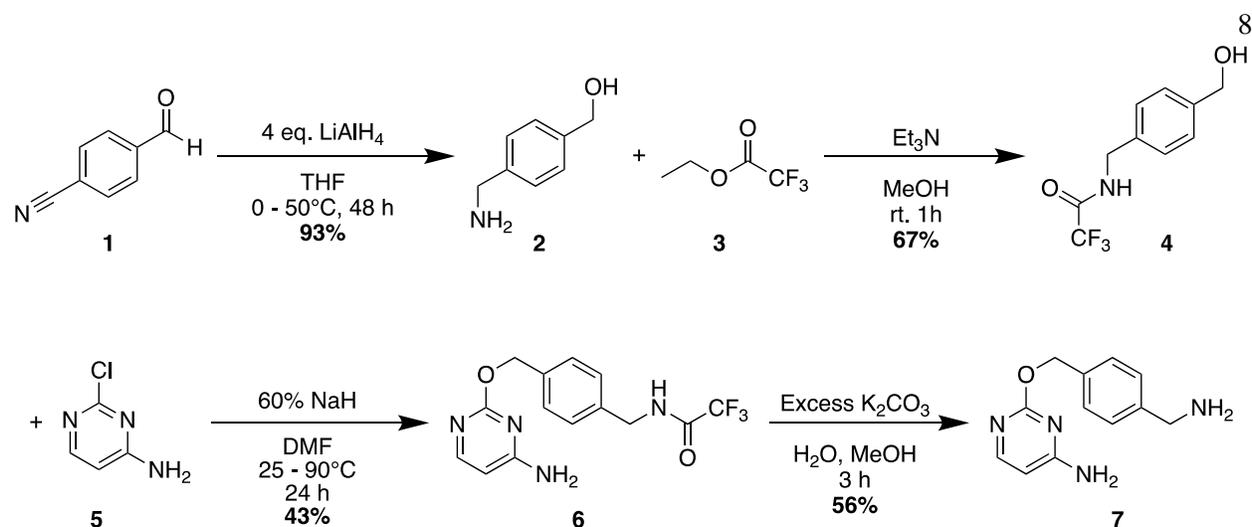
### Synthesis of GFP-Benzylcytosine Fluorogenic Probes for CLIP-tag

To enable fluorogenic labeling of the CLIP-tag protein, the initial goal of this work was to synthesize five different fluorogenic probes with different linker moieties attaching the GFP based fluorophore and the O6-benzylcytosine ligand. Because the fluorogenic properties of these probes arise from how significantly the fluorophores turn on fluorescence upon changes in their microenvironment upon aggregation, the length and rigidity of the linker significantly affects fluorogenic properties by dictating how the fluorophore is held close to/far from the protein of interest. To this end, a series of GFP based CLIP-tag probes were synthesized using 2, 4, and 6-carbon linkers, a cyclohexane linker, and a proline linker.

Ultimately, the fluorogenic properties of these probes will be evaluated, and the probe with the best fluorogenic properties (e.g. largest fluorescence “turn-on” effect, highest quantum yield, etc.) will be selected as the optimized probe for GFP fluorophore based CLIP-tag labeling.

### Synthesis of O6-Benzylcytosine

O6-Benzylcytosine derivatives bearing a fluorescent chemical probe can be used to label CLIP fusion (CLIP) proteins in cells because O6-benzylcytosine can act as a ligand that tags the CLIP protein.<sup>11</sup> Therefore, O6-benzylcytosine was synthesized in order to attach various fluorescent probes to this ligand base, allowing for fluorescent tagging of desired proteins. O6-Benzylcytosine was synthesized in four steps following literature precedent (Fig. 3).<sup>11</sup>



**Figure 3. Synthetic Scheme for O6-Benzylcytosine**

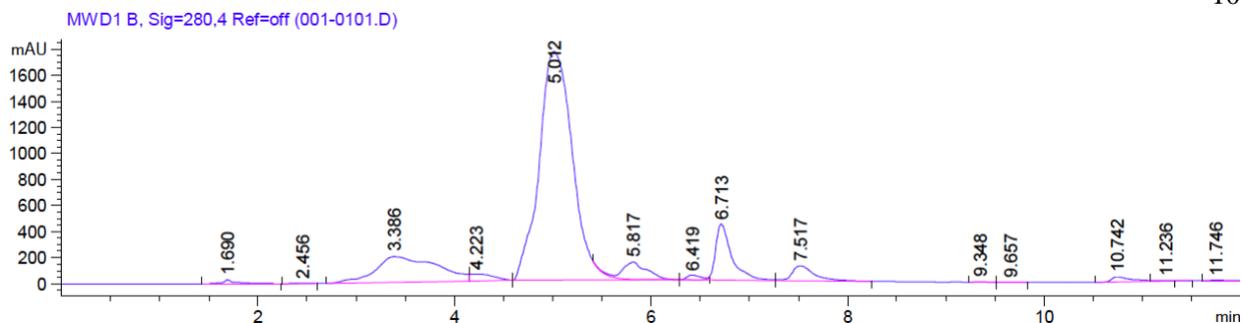
The commercially available starting material formylbenzonitrile (**1**) was reduced to 4-hydroxymethyl benzylamine (**2**) using lithium aluminum hydride to reduce the aldehyde and nitrile functionalities. The primary amine of benzylamine **2** was protected as an amide to afford benzylamide **4** in order to prevent the primary amine from acting as a nucleophile in the next synthetic step. The addition of sodium hydride to benzylamide **4** deprotonates the hydroxyl group, and the generated alkoxide then attacks the 4-amino-2-chloropyrimidine (**5**) in a nucleophilic aromatic substitution reaction to replace the aromatic chlorine substituent and afford protected O6-benzylcytosine **6**. Detailed experimental procedures and  $^1\text{H}$  NMR characterization can be found in appendices A and B, respectively.

The yield for this nucleophilic aromatic substitution was acceptable for larger scale synthesis, but not high (43%). In a departure from the reported synthesis of O6-benzylcytosine, nucleophilic aromatic substitution of chloropyrimidine **5** was also run using 1-methylpyrrolidine in the hopes that the resulting positively charged, quaternary amine substituent would be a much more facile leaving group than chloride. Although this would add one step to the synthesis ( $\text{S}_{\text{N}}\text{Ar}$

with 1-methylpyrrolidine to install a better leaving group, followed by the desired  $S_NAr$  to build complexity), this option was investigated to see whether the overall conversion of benzylamide **4** to protected O6-benzylcytosine **6** would be significantly improved. Unfortunately, the  $S_NAr$  reaction with 1-methylpyrrolidine did not proceed to near-quantitative levels, making the addition of this step unhelpful in terms of overall yield.

The amide functionality of protected O6-benzylcytosine was then cleaved to afford the deprotected primary amine in the desired O6-benzylcytosine molecule **7**. The original literature procedure reports deprotection using 5 mL of methylamine 33% by weight in ethanol.<sup>11</sup> To mimic these conditions without the same exact reagent, deprotection was tested under three different conditions: 2M  $NH_3$  in methanol, 40% methylamine by weight in water, and excess  $K_2CO_3$  in water and methanol. Ultimately, the deprotection conditions using potassium carbonate proved the fastest, highest yielding, and simplest to work up.

The most significant challenge in synthesizing large quantities of O6-benzylcytosine was purification. The molecule is quite polar because of the two primary amine functional groups on **7**, preventing efficient organic-aqueous extraction and purification via silica gel column chromatography. When spotted onto silica coated TLC plates, the desired product spot stays at the baseline (visualized with ninhydrin stain) along with other byproducts and undesired materials. Thus, the crude product **7** from the deprotection was purified via preparatory HPLC on a reverse-phase column. Using a solvent system that gradually changed from 94.5% water, 4.5% acetonitrile, and 1% trifluoroacetic acid to 67.5% water, 31.5% acetonitrile, and 1% trifluoroacetic acid over the course of 12 minutes, purified O6-benzylcytosine was successfully obtained as the major component of the crude mixture (eluted at 5.012 min, Fig. 4).

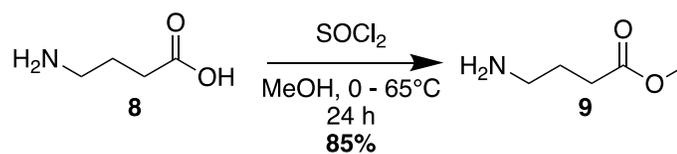
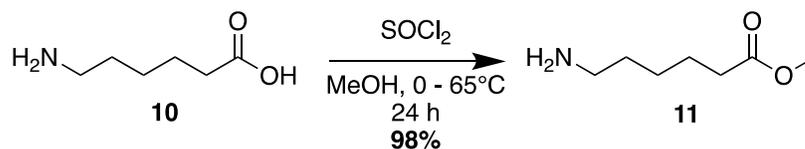
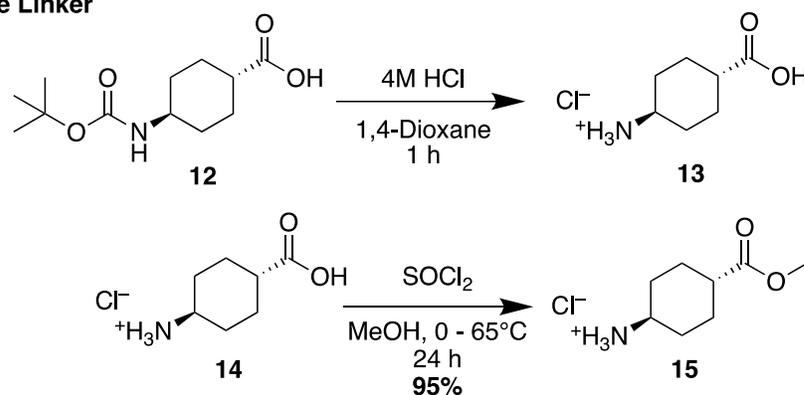


**Figure 4. HPLC Trace for O6-Benzylcytosine**

However, this solution to the purification problem was far from ideal — only about 30-40 mg of crude material in 0.9 mL of solvent could be injected into the HPLC during one run, and the few milligrams of purified product was then obtained in approximately 20 mL of water and acetonitrile. Because O6-benzylcytosine cannot be extracted from water using organic solvents, the HPLC eluents were then frozen using liquid nitrogen and lyophilized for several days to remove the solvent, affording O6-benzylcytosine in very small quantities as a white, flossy solid. Though 970 mg of O6-benzylcytosine was successfully purified in this manner, this purification was extraordinarily time consuming and inefficient. For future synthesis of additional O6-benzylcytosine, purification using aluminum instead of silica gel column chromatography may prove a more efficient method of purification than HPLC. Aluminum is a basic medium that should not protonate the primary amine moieties of O6-benzylcytosine, hopefully enabling the product to move through the solid phase of the column instead of being protonated and bound to acidic silica gel. For the purposes of this project, however, the purified O6-benzylcytosine was sufficient to continue with the fluorogenic probe synthesis.

## Synthesis of Linker Molecules

The utility of these fluorogenic probes depends on their ability to detect changes in their surrounding microenvironments and respond with a change in fluorescence. For these GFP based fluorogenic probes, the GFP fluorophore increases fluorescence intensity whenever the surrounding environment becomes rigid, preventing energy losses through rotational motion of the molecule. When the labeled proteins aggregate, the environment around the fluorophore becomes more rigid and fluorescence intensity increases. Thus, the connectivity between the protein tag ligand (O6-benzylcytosine) and the fluorophore strongly influences how strongly the fluorophore feels changes in the protein environment. Both the length and rigidity of the linker connecting the fluorophore and O6-benzylcytosine can dictate how closely the probe is held to the protein active site and, therefore, how much the environment around the probe changes. To investigate these effects, linkers of five different lengths and rigidities were used to synthesize GFP-benzylcytosine probes. Screening of this series of molecules would identify the linker that yields a fluorogenic probe with the most desirable properties. The linkers examined in this study were 2-carbon, 4-carbon, 6-carbon, cyclohexane, and proline linkers. As the proline and 2-carbon (glycine) linkers are commercially available, only the other three linkers were synthesized (Fig. 5).

**4-Carbon Linker****6-Carbon Linker****Cyclohexane Linker**

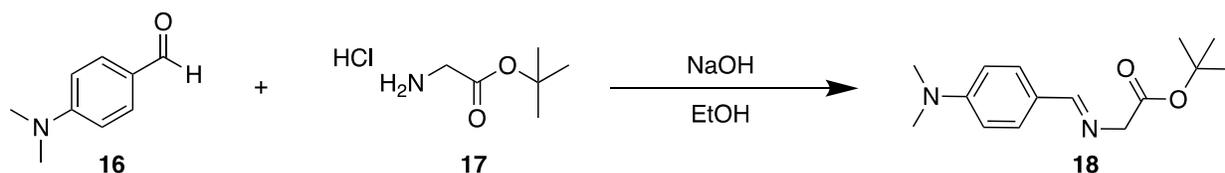
**Figure 5. Synthetic Scheme for Linker Molecules**

The synthesis of these linkers was straightforward even on gram scale, and the linkers did not require extensive purification other than the removal of solvent and volatile reagents and byproducts under vacuum.

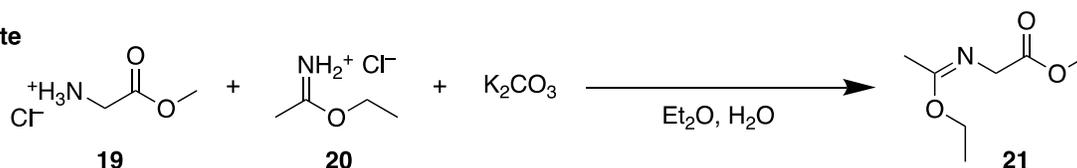
## Synthesis of GFP Fluorophore

The GFP fluorophore was synthesized in four steps (Fig. 6).

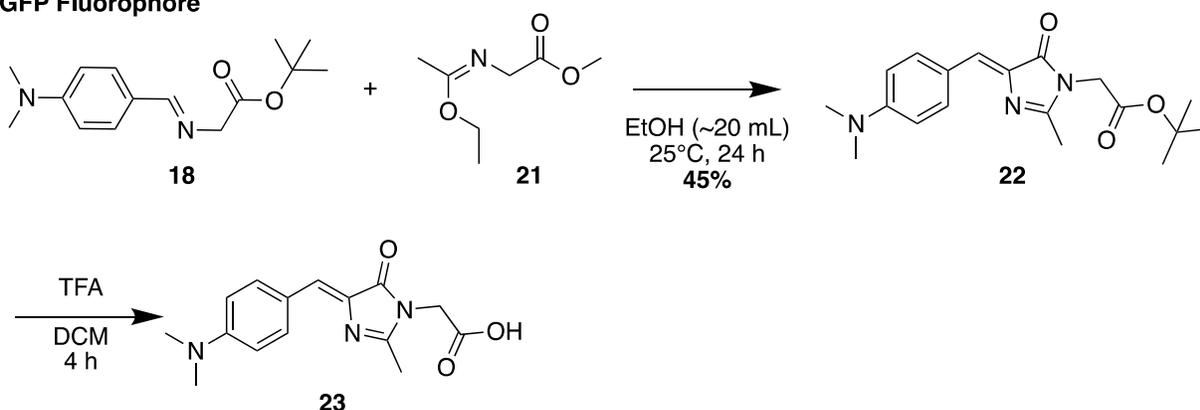
### Schiff Base



### Imidate



### GFP Fluorophore



**Figure 6. Synthetic Scheme for GFP Fluorophore**

The Schiff base **18** was synthesized via nucleophilic acyl substitution of 4-dimethylamino benzaldehyde (**16**) with glycine *t*-butyl ester hydrochloride (**17**) to convert the aldehyde functional group to an imine.

The synthesis of the desired imidate **21** was achieved using a biphasic reaction, which was vigorously stirred to enable the reaction to occur at the water/ether interface (Fig. 4). Although the reaction was worked up by collecting and drying the ether layers and removing most of ether under vacuum, the imidate product was not further purified or characterized due to instability of the

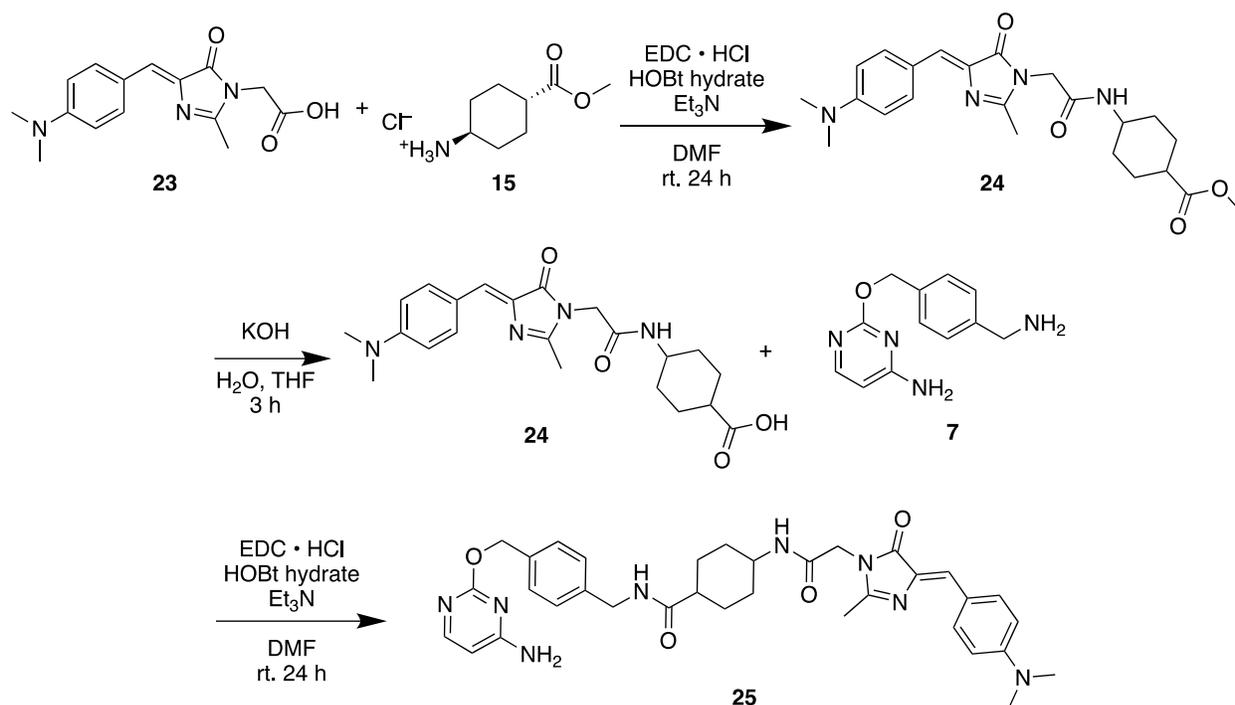
product, and the reaction mixture was not analyzed by TLC. Once most of the ether was evaporated, with a small amount remaining to avoid decomposition of the product, the freshly prepared imidate solution was added directly to the Schiff base reaction mixture.

The unique procedure of this step accommodates the mechanism of the desired reaction and allows for an expedient, efficient work-up to avoid loss of product to decomposition. Potassium carbonate and glycine methyl ester hydrochloride (**19**) are added to the ether/water first so the base can deprotonate **19** and make it a strong nucleophile. Deprotonated **19** then partitions into ether while inorganic salt byproducts remain in the aqueous phase. Addition of **20** then allows for rapid formation of the imidate as **20** is already in the protonated, highly electrophilic form. The amine of **19** attacks the imine of **20**, resulting in the expulsion of ammonia and the formation of the imidate. Because the active forms of **19** and **20** reside in different phases (**19** in ether, **20** in water), stirring the biphasic mixture vigorously is required because the reaction only takes place at the solvent interface. Once the reaction occurs, the neutral product partitions into the ether, allowing for easy separation from unwanted byproduct salts and isolation into a usable form.

The synthesis of the desired protected GFP fluorophore **22** proceeded smoothly from the Schiff base and freshly prepared imidate in 45% yield. The *boc*-protected fluorophore **22** was used as the starting material for synthesis of the RFP mimic fluorophore, and also deprotected under acidic conditions to cleave the *boc* ester and yield the GFP fluorophore **23** to be used in the syntheses of GFP-benzylcytosine based fluorogenic probes. Detailed experimental procedures and <sup>1</sup>H NMR characterization can be found in appendices A and B.

### Synthesis of Final GFP-Linker-Benzylcytosine Probes

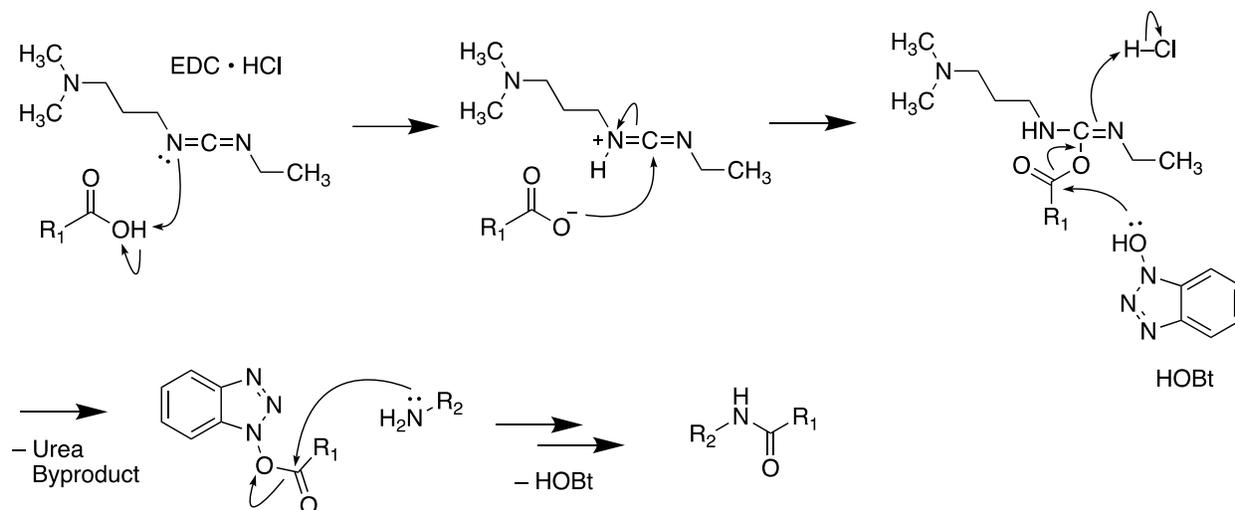
With each of the three pieces in hand (GFP fluorophore, linker molecules, and O6-benzylcytosine), the final fluorogenic probes were synthesized through a series of peptide coupling and carboxylic acid methyl ester deprotection reactions. The synthetic scheme for the GFP-Cyclohexane-BC probe is shown below in Fig. 7 as a representative scheme.



**Figure 7. Synthetic Scheme for GFP-Cyclohexane-BC**

The GFP fluorophore was coupled to the linker molecules via peptide coupling to form a new amide bond. The mechanism of peptide coupling under the conditions used for this synthesis is shown below in Fig. 6. The EDC · HCl first activates the carboxylic acid by deprotonating the acid, then the carboxylate group re-attacks EDC · HCl to form an activated intermediate. HOBT then attacks the activated carbonyl to generate a HOBT ester and release urea byproducts. This

HOBt ester is now primed nucleophilic acyl substitution via attack by the amine, ultimately yielding the coupled final product with a new amide bond.

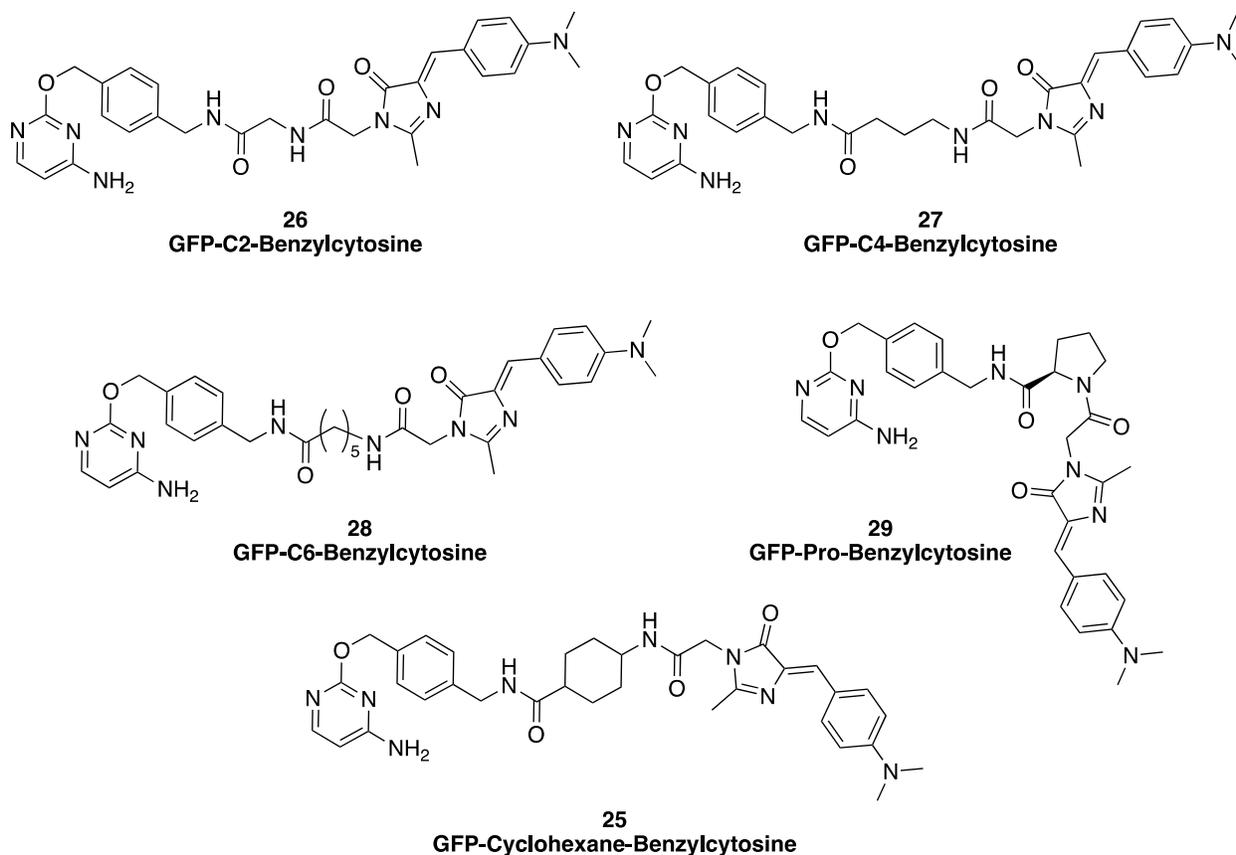


**Figure 8. Mechanism of Peptide Coupling**

After the initial peptide coupling, the GFP-linker methyl ester intermediate was deprotected under basic conditions to refurbish the carboxylic acid functionality for the next peptide coupling reaction. The deprotection reaction mixture was quenched with pH 4.2 acetic acid buffer to protonate the carboxylic acid without protonating any of the basic nitrogenous functional groups and enable extraction of the product with organic solvent. However, much of the product stayed in the aqueous phase even with careful pH control. This may be due to the introduction of DMF into the extraction as the reaction solvent, which probably pulled the product into the aqueous phase. Much of the product was lost in the aqueous phase at this stage despite multiple extractions, so overall yields were significantly diminished.

The deprotected GFP-linker intermediates were then finally coupled to O6-benzylcytosine via peptide coupling to form an amide bond between the carboxylic acid of the GFP-linker intermediate and the primary benzylamine of O6-benzylcytosine. This general synthetic route was

used to synthesize five GFP-linker-Benzylcytosine probes using C2, C4, C6, cyclohexane, and proline linkers (Fig. 9).



**Figure 9. Final GFP-Linker-BC Probes**

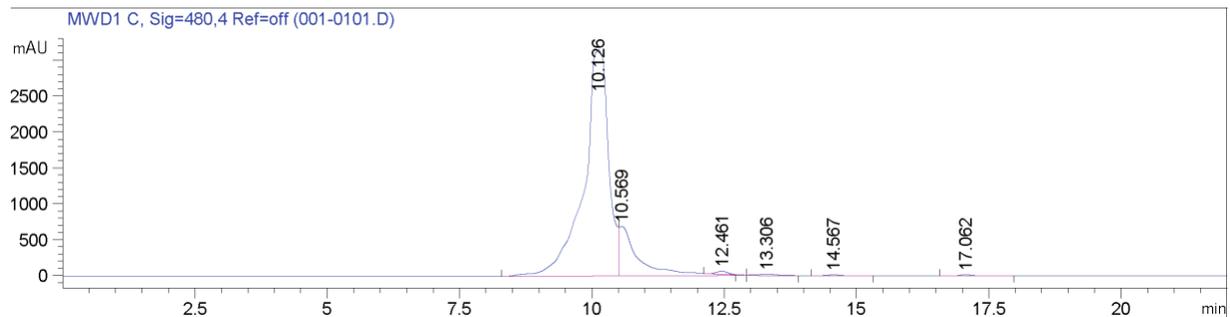
### Purification of Final GFP-Linker-Benzylcytosine Probes

Due to the remaining primary amine functionality in the final GFP-linker-BC probes, the small scale of the final peptide coupling reactions (typically under 20 mg scale), and the fact that the peptide couplings were run in dimethylformamide, doing an aqueous work-up and silica gel column purification proved difficult. Therefore, the final product probes were purified via reverse-phase HPLC using gradient mixtures of solvent A (94.5% water, 4.5% acetonitrile, 1% trifluoroacetic acid) and solvent B (94.5% acetonitrile, 4.5% water, 1% trifluoroacetic acid).

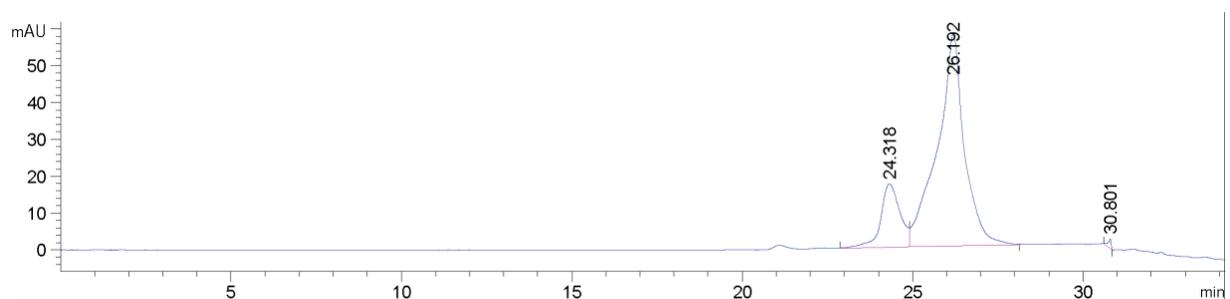
However, due to the small scale of these reactions and the close polarity of the desired product and byproducts, identification of appropriate solvent conditions for purification proved challenging—running multiple test gradients resulted in the gradual erosion of material, and the collected samples were obtained in dilute aqueous solution.

The crude GFP-proline-BC probe **29** was used as a reference to identify improved HPLC purification conditions for the probes. Purification of probe **29** using the established HPLC purification protocols from the purification of protected GFP and RFP fluorophores (2 min 100% A to separate polar organic solvents, gradient to 100% B over 22 min total) resulted in poor separation, with two distinct shoulders visible in the HPLC trace monitored at 480 nm (Fig. 10A). Slowing the gradient to extend over 30 minutes instead of 20 successfully separated a small impurity that eluted slightly before the desired product (Fig. 10B), but increasing the flow rate to decrease peak width and try and improve resolution actually resulted in poorer separation of the two peaks (Fig. 10C). After several other trials, using a mixed gradient and step solvent program was found to result in the best separation (Fig. 10D): 2 min 100% A to allow polar, organic solvents to elute, gradient to 24 min 22% B, hold at 22% B until 28 min to allow the impurity to elute, gradient to 31 min 100% B to then push the desired product off the column.

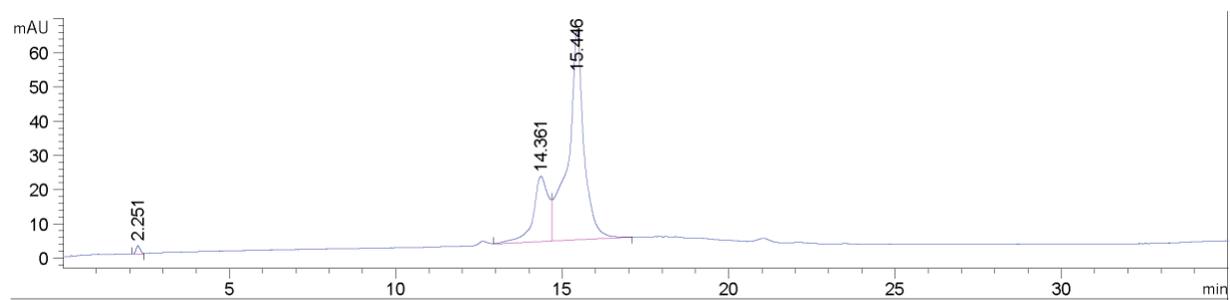
## A) Gradient from 0 - 100% Solvent B over 20 minutes at 20 mL/min



## B) Gradient from 0 - 100% Solvent B over 30 minutes at 20 mL/min



## C) Gradient from 0 - 100% Solvent B over 30 minutes at 30 mL/min



## D) Gradient purification with 4 min. step at 22% Solvent B

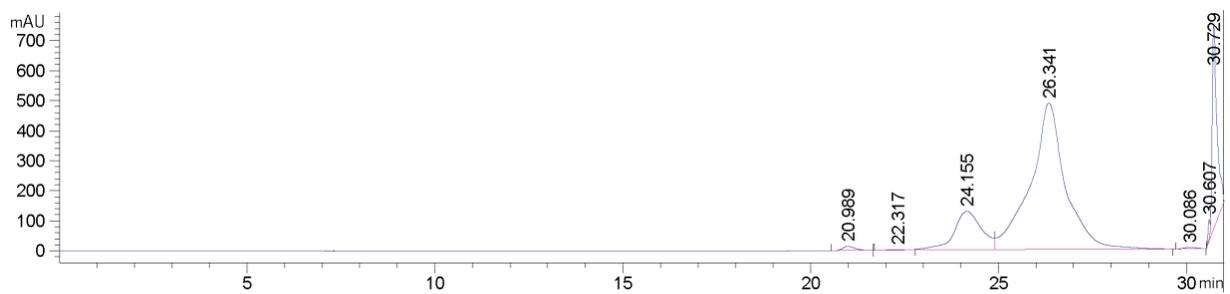


Figure 10. HPLC Purification of GFP-Pro-BC

Using similar small-scale HPLC runs to test purification conditions for probes **25** to **29**, the final probes were purified using three different HPLC solvent programs (for final conditions and HPLC traces, see Appendix B). However, due to the small quantity of product synthesized, not all probes could be completely purified as testing each new HPLC purification condition resulted in loss of product. The purification of all probes was sufficient to allow for structural characterization and confirmation by  $^1\text{H}$  NMR. Ultimately, three probes were effectively purified: GFP-Cyclohexane-BC **25**, GFP-C2-BC **26**, and GFP-C4-BC **27**. The other probes, GFP-C6-BC **28** and GFP-Pro-BC **29**, were mostly purified.

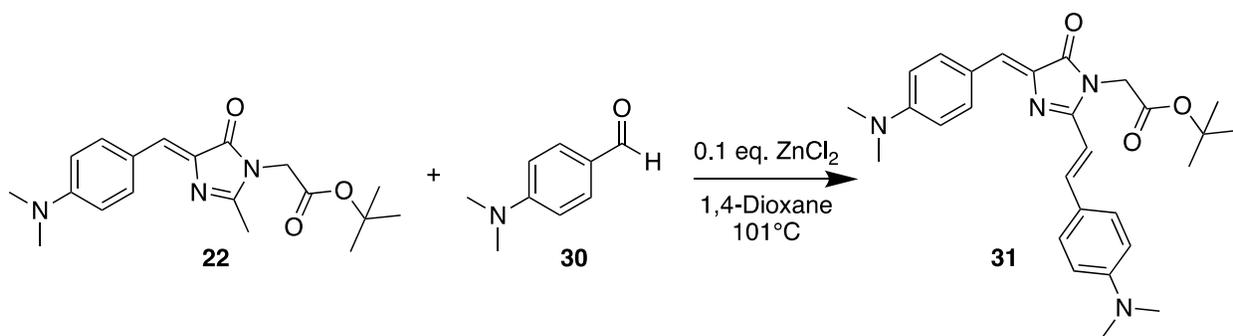
With these five synthesized probes in hand, future directions will include scaling up these syntheses to enable more facile purification of the final probes and yield slightly larger quantities of product. Once that is accomplished, these five fluorogenic probes can be screened for their fluorogenic properties using CLIP protein under conditions that induce protein aggregation. Quantification of their fluorogenic properties will allow for the selection of the best probe for future applications in the labeling and imaging of aggregating CLIP tagged proteins of interest in live cells.

## Chapter 3

### Optimization of RFP Mimic Fluorophore Synthesis

Increasing conjugation of the GFP fluorophore by adding a benzaldehyde derivative changes the color of the molecule's fluorescence from yellow to red. By increasing conjugation of the pi system, the HOMO-LUMO energy gap of the molecular orbitals decreases and the wavelength of light absorbed by the molecule increases from 480 nm for GFP to 540 nm for the RFP mimic fluorophore. Synthesizing a series of fluorogenic probes using the RFP fluorophore attached to the O6-benzylcytosine ligand via different linkers would enable dual-color imaging of CLIP tagged proteins that aggregate (both GFP and RFP fluorophores could be used).

The protected RFP fluorophore **31** can be synthesized from the protected GFP fluorophore **22** in one synthetic step through a Lewis acid catalyzed enamine condensation with dimethylamino benzaldehyde (**30**) (Fig. 11).

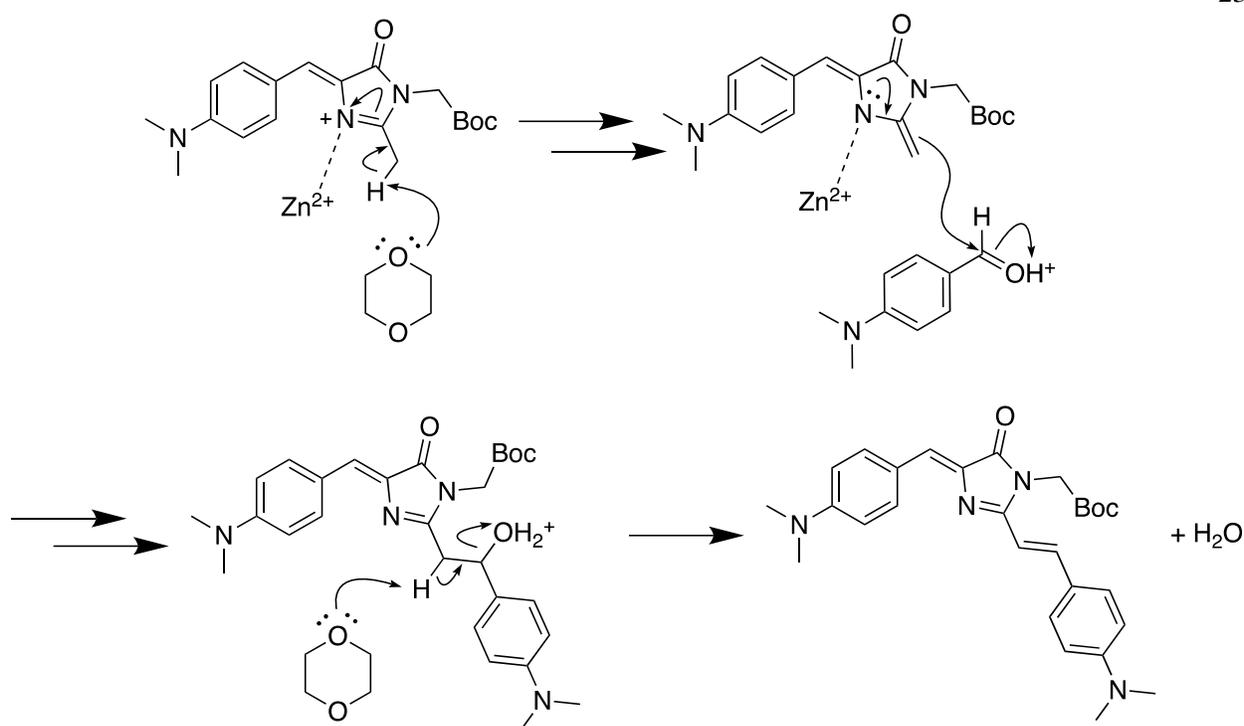


**Figure 11. Synthetic Scheme for RFP Fluorophore**

These conditions have been frequently used to synthesize the RFP fluorophore.<sup>10</sup> However, product conversion under these conditions is typically only low to moderate, resulting in low yields of the desired RFP fluorophore (~20%). Additionally, the  $\text{Zn}^{2+}$  Lewis acid catalyst can also deprotect the *boc*-protected carboxylic acid moiety of the fluorophores;<sup>12</sup> this side reactivity eats

up both the starting material and the product, significantly reducing yields. Particularly problematic is the fact that the *boc* ester carbonyl in the product is more electron rich than the carbonyl in the starting material due to additional electron donation from the dimethylamino phenyl moiety. This added electron richness means the Lewis acid catalyst may coordinate more readily to the product, and selectively catalyze deprotection of the product over the starting material, which would decrease RFP:GFP ratio. Moreover, the GFP and RFP fluorophores have extremely similar polarities and  $R_f$  values, making them difficult or even impossible to cleanly separate via silica gel column chromatography. Multiple trials with several different solvent systems (ethyl acetate & hexanes, DCM & methanol, etc.) failed to separate the RFP product from the GFP starting material both on a manual silica gel column and a Biotage auto-column. Recrystallization of a GFP/RFP mixture also resulted in some recrystallization of leftover GFP starting material, but the desired RFP product and a significant amount of GFP remained in solution. Thus, the only viable purification method for the protected RFP fluorophore was, again, preparatory reverse-phase HPLC, which was time-consuming and highly inefficient when trying to synthesize large, gram-scale quantities of this fluorophore (see Fig. 32, Appendix B). Overall, low product conversion combined with purification difficulty makes the synthesis of the desired RFP fluorophore highly inefficient, precluding the facile synthesis of a series of RFP based fluorogenic probes.

Thus, optimization of the RFP synthesis was undertaken by screening different conditions for the key enamine condensation. Initially, different conditions for this reaction were selected using mechanistic understanding of the reaction. The proposed reaction mechanism for the current RFP synthesis is shown below in Fig. 12.



**Figure 12. Proposed Mechanism of RFP Synthesis**

Protected RFP **31** is synthesized from protected GFP **22** via a Lewis acid catalyzed enamine condensation with 4-dimethylamino benzaldehyde (**30**). The Lewis acid, Zn<sup>2+</sup> under precedented conditions, coordinates to the most basic nitrogen in the protected GFP structure. This coordination activates the protected GFP by making the methyl hydrogens acidic, as the C–H  $\sigma$ -bonds hyperconjugate into the adjacent electron deficient  $\pi$  system. This activation enables deprotonation of the acidic methyl group by the solvent or another electron donor in solution, generating a nucleophile that then attacks the activated aldehyde carbonyl of the benzaldehyde starting material. Because the precedented reaction conditions are acidic, the aldehyde carbonyl is likely protonated or also coordinated to the Lewis acid, which facilitates the nucleophilic attack by GFP during the condensation reaction. The resultant benzylic hydroxyl group is likely also protonated under the acidic conditions, allowing for facile elimination of water to form the final double bond. Formation

of this double bond is favorable both because water is a good leaving group, and also because installation of this double bond dramatically increases conjugation over the entire molecule.

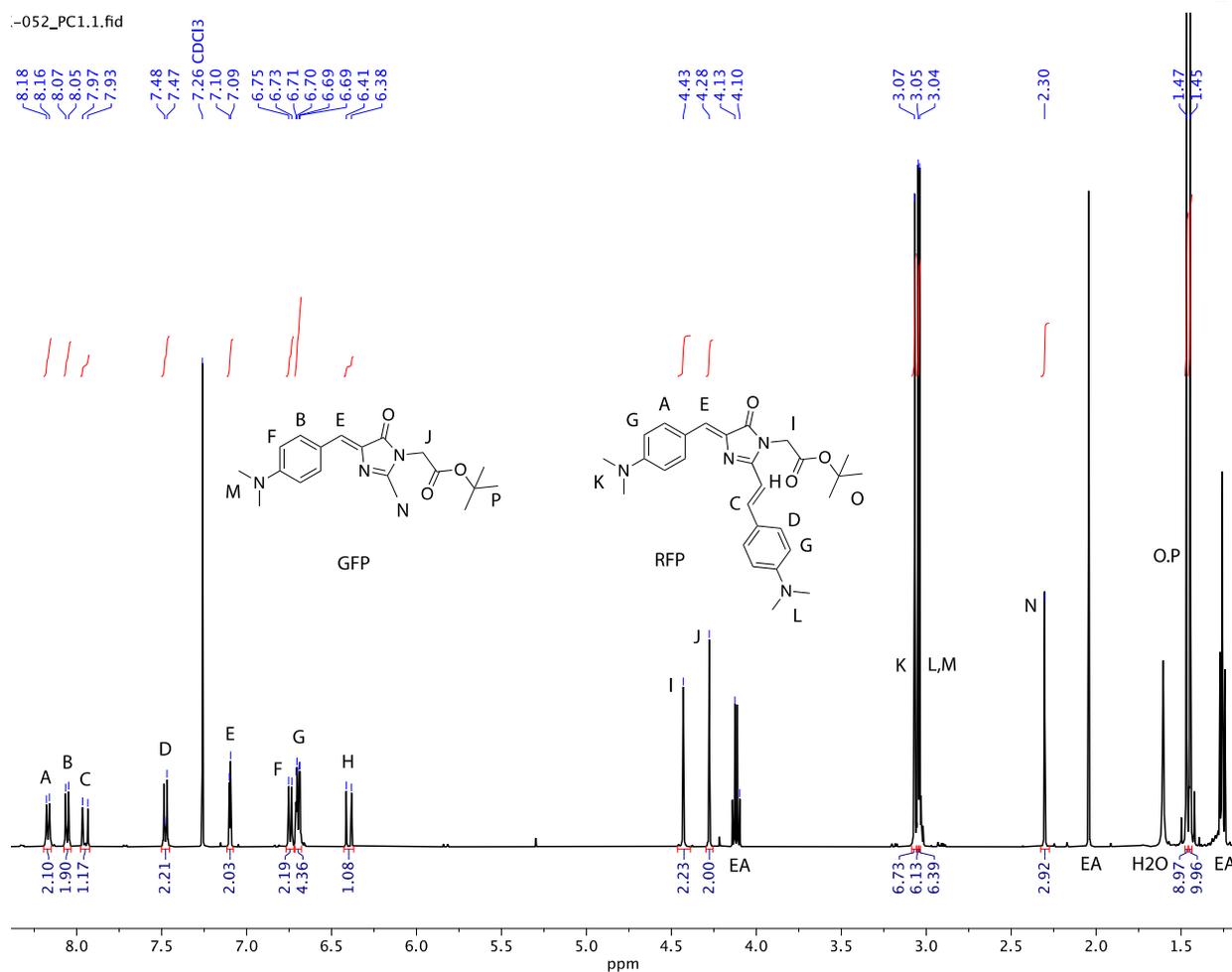
Using this mechanistic understanding of the key transformation, different reaction conditions were tested for the RFP synthesis. These conditions were screened using small scale (~50 mg starting material) reactions that were purified via manual silica gel column chromatography to obtain a GFP/RFP mixture. The ratio of GFP:RFP was calculated from the  $^1\text{H}$  NMR spectrum by comparing the integration of key characteristic signals, and this ratio was also used to calculate the NMR yield of RFP for each condition. All results of this screen for optimized reaction conditions are shown in Table 1. A dash (-) in Table 1 indicates that these results are from conditions that were tested, but for which the final  $^1\text{H}$  NMR results have not yet been obtained.

**Table 1. RFP Synthesis Condition Screen**

Conditions	RFP:GFP Ratio	NMR Yield (%)
0.1 eq. $\text{ZnCl}_2$ , 101°C, not anhydrous	1 : 0.9	26.3
Piperidine in pyridine, 115°C	0 : 1	N/A
0.5 eq. $\text{ZnCl}_2$ , 101°C, not anhydrous	1 : 0.46	16.0
0.1 eq. $\text{ZnCl}_2$ , 101°C, anhydrous solvent	-	-
0.1 eq. $\text{BF}_3 \cdot \text{Et}_2\text{O}$ , 101°C, anhydrous solvent	1 : 0.245	54.4
0.1 eq. $\text{AlCl}_3$ , 101°C, anhydrous solvent	-	-
0.1 eq. $\text{TiCl}_4$ , 101°C, anhydrous solvent	-	-
0.1 eq. $\text{MgCl}_2$ , 101°C, anhydrous solvent	-	-
0.1 eq. $\text{Zn}(\text{OAc})_2$ , 101°C, anhydrous solvent	-	-

The solvent for all the screened acidic conditions was kept as 1,4-dioxane. Although some literature procedures also use THF as a solvent,<sup>13</sup> the boiling point of THF is much lower than 1,4-dioxane (66°C vs. 101°C) which limits the ability to heat the reaction. Though the RFP:GFP ratio and yield of RFP was not calculated for this preliminary reaction, running this condensation at 70°C instead of 101°C using the original ZnCl<sub>2</sub> conditions resulted in less than ~30% conversion of GFP to RFP (qualitative TLC analysis). Therefore, the solvent was maintained as 1,4-dioxane to enable sufficient heating to push the reaction forward as much as possible.

The first screened condition was the original, precedent conditions using 0.1 eq. ZnCl<sub>2</sub> as the Lewis acid catalyst under not strictly anhydrous conditions (using wet 1,4-dioxane). The <sup>1</sup>H NMR of the purified GFP/RFP mixture is shown in Figure 13.



**Figure 13. Annotated  $^1\text{H}$  NMR of GFP/RFP Mixture**

By comparing the  $^1\text{H}$  NMR spectrum of the mixture to  $^1\text{H}$  NMR spectra of both purified GFP and RFP fluorophores, it is possible to definitively assign every signal to either the GFP starting material or the RFP product. Then, using the ratio of peak integrations for equivalent protons, the ratio of RFP:GFP was calculated. Specifically, the signals for the methylene protons between the *boc*-protected carboxylic acid and the amide nitrogen were selected because of their easy identification and integration (no overlapping signals or splitting). The singlet for these methylene protons shows up at 4.43 ppm for protected RFP and 4.28 ppm for protected GFP, and both signals should integrate to 2H if the molecules are in a 1:1 ratio. The ratio of RFP:GFP was

calculated by dividing the integration value of the characteristic GFP methylene signal by the integration value for the same protons on the RFP molecule. This ratio was then used to back-calculate the yield of RFP from the mass of the GFP/RFP mixture. For all  $^1\text{H}$  NMR of the GFP/RFP mixtures from the screened conditions, see Appendix B.

Under the original reaction conditions, the RFP:GFP ratio is 1:0.9 and the RFP yield is only 26.3%. Typically, similar kinds of condensation reactions can also be run under basic conditions (e.g. Aldol condensation), wherein a strong base deprotonates a slightly acidic position to generate an enolate nucleophile that then attacks a carbonyl. Thus, a basic condition was also tested following literature precedent using pyridine as a basic solvent and a small amount of piperidine as a proton donor.<sup>13</sup> Under these conditions, however, there was no observable product formation at all—the  $^1\text{H}$  NMR of the product mixture was essentially pure starting material (Fig. 28, Appendix B). Thus, attempting to run this condensation under basic conditions was ruled out as a viable option. The addition of an acid catalyst seemed necessary to activate both the GFP (enabling deprotonation) and aldehyde (facilitating nucleophilic attack) starting materials.

Three different routes were considered for the optimization of this enamine condensation under acidic conditions: 1) changing equivalents of the catalyst, 2) changing the Lewis acidity of the catalyst, 3) changing the counterion of the Lewis acid catalyst. Increasing the equivalents of  $\text{ZnCl}_2$  from 0.1 to 0.5 indeed improved the conversion of GFP to RFP, but the overall yield of the RFP fluorophore was actually lower than with 0.1 eq. of  $\text{ZnCl}_2$ . This result is most probably due to the fact that  $\text{ZnCl}_2$  can both facilitate the conversion of GFP to RFP but also deprotect the *boc* ester, resulting in degradation of the desired product over time. The  $\text{Zn}^{2+}$  cation can also activate the *boc* ester carbonyl, and with trace water in solution, water can act as a nucleophile to cleave the *boc* protecting group and reveal the carboxylic acid functionality. The deprotected product

appears as an insoluble black gel-residue that coats the inside of the reaction vessel, and adsorbs strongly to silica gel due to its high polarity. To avoid product degradation due to deprotection of the carboxylic acid functional group, the equivalents of the Lewis acid catalyst were maintained at 0.1 eq.

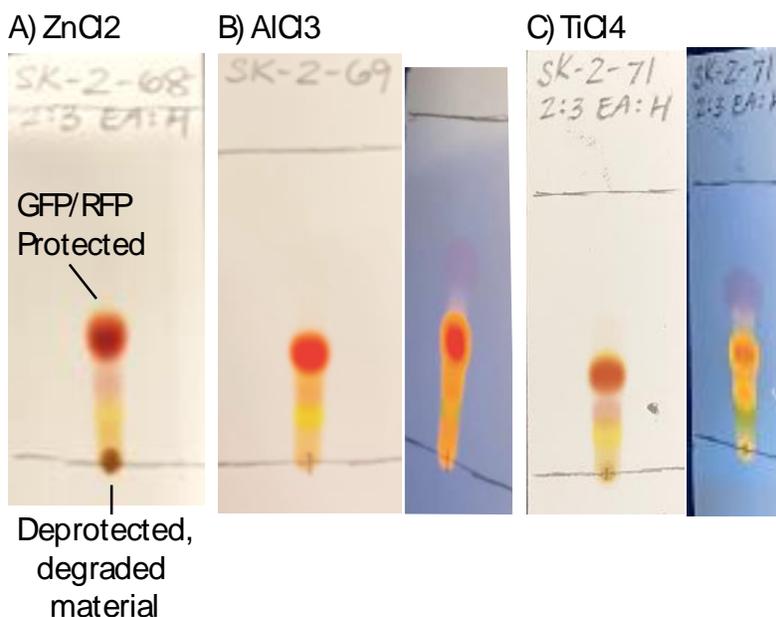
Next, the reaction was run under the same conditions, but using 1,4-dioxane that had been dried over activated molecular sieves for several days. Although this reaction may not be incredibly water sensitive, water could potentially act as a nucleophile to cleave the *boc* protecting group, add into the aldehyde instead of the protected GFP starting material, or it could also coordinate to the Lewis acid catalyst and decrease catalytic efficiency. Running the same reaction under anhydrous conditions indeed resulted in less visible black residue in the reaction mixture from deprotected product, though it afforded a similar mass of purified GFP/RFP mixture after the initial column purification. The purified mixture looked qualitatively similar to the product formed with non-anhydrous solvent; however, quantitative analysis by  $^1\text{H}$  NMR has not yet been performed. Ultimately, anhydrous 1,4-dioxane was chosen as the preferred solvent system because 1) the reaction mixture appeared to have less visible deprotected product (based on appearance of reaction mixture and TLC analysis), and 2) using an anhydrous solvent enhances the reproducibility of this reaction by eliminating an otherwise uncontrolled variable and more generally adheres to good synthetic practice.

Next, different Lewis acid catalysts were screened to investigate whether the acidity of the catalyst had a significant impact on the reaction. As  $\text{ZnCl}_2$  is  $2+$  on the catalytic metal center, different catalysts ranging from  $+1$  to  $+4$  charge were screened. Because Lewis acid coordination to the nitrogen and the subsequent removal of electron density from the  $\pi$  system is mechanistically required for deprotonation of the terminal methyl to generate the nucleophile, it was hypothesized

that a more cationic/acidic Lewis acid would result in higher GFP to RFP conversion and yield because it would more strongly activate the starting material and lower the energy barrier for generation of the nucleophile.

Surprisingly, however, using 0.1 eq.  $\text{BF}_3 \cdot \text{Et}_2\text{O}$  resulted in significantly higher GFP to RFP conversion (1 : 0.245) and RFP yield (54.4%). This result directly contradicts the initial hypothesis that more Lewis acidic catalysts would increase conversion and yield. This unexpected result may be due to the fact that  $\text{BF}_3 \cdot \text{Et}_2\text{O}$  is not a strong enough Lewis acid catalyst to significantly activate the *boc* ester and enable cleavage of the *boc* group (deprotection of carboxylic acid functionality). By being just acidic enough to catalyze the enamine condensation without also catalyzing deprotection of the carboxylic acid,  $\text{BF}_3 \cdot \text{Et}_2\text{O}$  may lead to higher yields and RFP:GFP ratios by ensuring that the product RFP is not degraded over the course of the reaction. To test this hypothesis, reactions with more Lewis acidic catalysts,  $\text{AlCl}_3$  and  $\text{TiCl}_4$ , were also examined. If  $\text{BF}_3 \cdot \text{Et}_2\text{O}$  leads to increased conversion and yield because it is a weaker Lewis acid that does not catalyze *boc* cleavage as effectively as  $\text{ZnCl}_2$ , stronger Lewis acid catalysts like  $\text{Al}^{3+}$  and  $\text{Ti}^{4+}$  should have the opposite effect – enhancing deprotection of the product and reducing overall yields and conversion. Regrettably, though reactions with  $\text{AlCl}_3$  and  $\text{TiCl}_4$  were run and purified, quantitative  $^1\text{H}$  NMR analysis has not yet been accomplished. However, from visual, qualitative observations of the reaction mixtures and TLC plates (Fig. 14), it appears that neither  $\text{AlCl}_3$  or  $\text{TiCl}_4$  significantly promoted deprotection of the product. After refluxing the reaction with  $\text{AlCl}_3$  overnight, no significant amount of black gel-like residue (deprotected product) was visible in the reaction vial. TLC analysis typically reveals the deprotected product as a brownish-red baseline spot due to the high polarity of the deprotected carboxylic acid. This brown baseline spot is clearly visible in Fig. 14A, the conditions with 0.1 eq.  $\text{ZnCl}_2$  in anhydrous 1,4-dioxane, but much

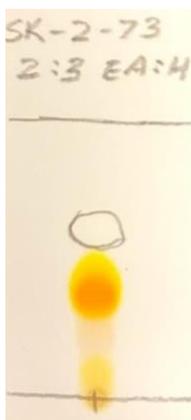
fainter in Fig. 14B, the conditions with 0.1 eq.  $\text{AlCl}_3$  in anhydrous solvent.  $\text{TiCl}_4$  proved to be a difficult Lewis acid to handle due to its high reactivity with atmospheric water and the lack of an available glass micro-syringe to properly measure out the required volume. Within 20 minutes of addition to the reaction, the reaction mixture turned viscous and changed to a dark black-brown color. Within 1 hour, the reaction was stopped because TLC analysis revealed that several new byproducts were rapidly forming while the color and fluorescence of the main starting material/product spot indicated that not much conversion to product had occurred (more orange-yellow than red, Fig. 14C).



**Figure 14. TLC Analysis of  $\text{ZnCl}_2$ ,  $\text{AlCl}_3$ , and  $\text{TiCl}_4$  Conditions**

From qualitative observations,  $\text{TiCl}_4$  appeared to be a poor Lewis acid catalyst for this reaction, potentially because the extremely high Lewis acidity and reactivity promoted various other unwanted side reactions. A definitive decision on  $\text{AlCl}_3$  cannot be made without quantitation of the GFP:RFP ratio and overall yield of RFP fluorophore. Observation of a lower RFP:GFP ratio and lower yields would support the proposed hypothesis.

$\text{Zn}(\text{OAc})_2$  was also screened as a  $\text{Zn}^{2+}$ -based Lewis acid catalyst with a more basic anion. The mechanistic justification for this experiment was that the more basic anion (acetate vs. chloride) may more effectively deprotonate the starting material, initiating the enamine condensation more efficiently. Unfortunately, when  $\text{Zn}(\text{OAc})_2$  was used as the catalyst, the reaction was extremely inefficient and sluggish. Though the NMR yield has not yet been determined, significant amounts of remaining starting material was clearly visible via TLC analysis (Fig. 15), despite holding all other reaction conditions constant.



**Figure 15. TLC Analysis of  $\text{Zn}(\text{OAc})_2$  Conditions**

This result may be explained by the fact that the acetate ligand is slightly more basic than  $\text{Cl}^-$  and also bidentate; thus, the acetate ligand might coordinate more strongly to  $\text{Zn}^{2+}$  and prevent the cation from acting as efficiently as an acid catalyst. Alternatively, the slightly basic nature of the acetate anion may negatively affect the acidic conditions required for the desired reaction to occur (the piperidine/pyridine conditions showed that the reaction does not occur under basic conditions).

Curious as to whether zinc in particular yields better results than an alternative  $2+$  Lewis acid catalyst, a reaction was also run with  $\text{MgCl}_2$  as the Lewis acid catalyst. Though the reaction

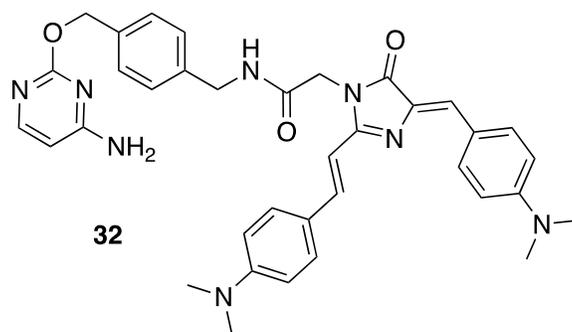
did proceed, a significant amount of remaining starting material (more so than the comparable  $\text{ZnCl}_2$  results) was clearly visible via TLC analysis (Fig. 16).



**Figure 16. TLC Analysis of  $\text{MgCl}_2$  Conditions**

As  $\text{Mg}^{2+}$  is a smaller, harder Lewis acid, it may be less able to activate the starting materials – for example, if  $\text{Zn}^{2+}$  coordinates to both the GFP precursor and the carbonyl of the benzaldehyde to activate both simultaneously, the smaller  $\text{Mg}^{2+}$  cation may be a poorer catalyst because the smaller size makes it less able to coordinate the two large starting materials at the same time.

Though not enough pure protected RFP fluorophore was obtained to synthesize a series of RFP based CLIP-tag probes (similar to the GFP probes detailed in Chapter 2), a single RFP based CLIP-tag probe was synthesized without any linker moiety connecting the fluorophore to the ligand. The protected RFP fluorophore **31** was deprotected to reveal the carboxylic acid functionality using trifluoroacetic acid, then the deprotected fluorophore was coupled to O6-benzylcytosine under peptide coupling conditions to afford an RFP-benzylcytosine fluorogenic probe **32** (Fig. 17, for  $^1\text{H}$  NMR spectrum see Fig. 31, Appendix B).



**Figure 17. RFP-Benzylcytosine Probe**

Though the initial goal of this work was to synthesize a series of RFP fluorophore based CLIP-tag probes, large-scale synthesis and purification of the RFP fluorophore proved to be too inefficient and difficult for efficient synthesis of the desired probes. Therefore, an optimization campaign for the synthesis of the RFP fluorophore was undertaken. Overall, without having quantitative NMR yields and ratios to compare all of the experimental conditions, it is difficult to definitively choose optimum reaction conditions and rationalize why certain reagents perform better than others. At this point,  $\text{BF}_3 \cdot \text{Et}_2\text{O}$  appears to be the best Lewis acid catalyst that affords the highest RFP:GFP ratios and product yields. Catalyst loading should be kept low, at 0.1 eq, and anhydrous solvent should be used. The acidic reaction conditions appear to be necessary for notable reaction progression. Significant progress was made towards optimization of the RFP fluorophore synthesis, and future selection of optimized reaction conditions will enable more efficient synthesis of this highly desirable fluorophore on larger scales for applications in synthesizing new RFP based fluorogenic probes.

## Chapter 4

### Conclusions & Future Directions

This thesis reports progress toward the development of a dual-color system of fluorogenic probes to detect protein aggregation in live cells based on GFP and RFP mimic fluorophores and a CLIP-tag protein system. Such fluorogenic CLIP-tag probes are desirable because they would represent the addition of an orthogonal tag system to the SNAP-tag and Halo-tag platforms, enabling more complex biological studies of multiple simultaneously labeled proteins of interest in live cells. Developing tools to enable the study of protein aggregation in living cellular environments may ultimately help elucidate how such protein aggregation causes biological effects, which is of particular interest in the study of neurodegenerative diseases.

Five GFP fluorophore based probes were synthesized by conjugating the GFP fluorophore to O6-benzylcytosine (CLIP-tag ligand) via a linker moiety. The structure and the rigidity of the linker (C2, C4, C6, cyclohexane, or proline) can significantly affect the fluorogenicity of these probes, hence a series of molecules was synthesized to enable further quantification of desirable fluorogenic properties and selection of the best probe. Though all five probes were successfully synthesized and characterized via  $^1\text{H}$  NMR, purification of the probes proved difficult, even on a reverse-phase preparatory HPLC system. Three probes, GFP-Cyclohexane-BC **25**, GFP-C2-BC **26**, and GFP-C4-BC **27**, were successfully purified. The other two probes, GFP-C6-BC **28** and GFP-Pro-BC **29**, were mostly purified; however, there was not enough material to successfully identify a better HPLC purification gradient. Hence, future work will include scaling up the syntheses of these last two probes to allow for further testing of different purification gradients. Once the proper purification protocol is identified for each probe, and enough material is purified, the fluorogenic properties of each probe will be tested to quantify desirable properties such as

“turn-on” intensity. Ultimately, the probe with the best fluorogenic properties (biggest fold difference in fluorescence intensity upon protein aggregation) will be selected for use in live cell imaging to study protein aggregation.

To enable dual-color imaging with the CLIP-tag system, it is necessary to have at least two different families of CLIP-tag probes with different colored fluorophores. Though the ultimate goal of this project is to develop and synthesize an RFP fluorophore based CLIP-tag probe, the original synthesis of the RFP fluorophore proved to be too inefficient and challenging to purify for large-scale, preparative synthesis that would be necessary for the synthesis of multiple different RFP based probes. Therefore, the synthesis of the desired RFP fluorophore was optimized, focusing specifically on optimizing the enamine condensation that converts protected GFP fluorophore to protected RFP fluorophore. The original conditions of 0.1 eq.  $\text{ZnCl}_2$  in 1,4-dioxane yielded only 26.3% desired product, and afforded a nearly 1:1 mixture of the GFP and RFP that proved challenging to separate on a standard silica gel column. After screening different conditions that changed catalyst loading, anhydrous solvents, and different Lewis acids, the best conditions found so far are 0.1 eq.  $\text{BF}_3 \cdot \text{Et}_2\text{O}$  in anhydrous 1,4-dioxane, which yields 54.4% desired product and a roughly 4:1 ratio of RFP to GFP. The current hypothesis to explain these results is that  $\text{BF}_3 \cdot \text{Et}_2\text{O}$  acts as a weaker Lewis acid catalyst that can catalyze the enamine condensation without also catalyzing the undesirable deprotection of the RFP product to give degraded carboxylic acid byproducts. However, quantification of yields from the other screened conditions (particularly the  $\text{AlCl}_3$  experiment) must still be completed, and these results could either support or negate the proposed hypothesis. Still, the discovery of the new, optimized reaction conditions marks a significant improvement in the RFP fluorophore synthesis, with nearly doubled yields and a 4x larger ratio of desired product to remaining starting material, which makes purification more

straightforward. Further work includes obtaining  $^1\text{H}$  NMR data for the purified products of all remaining screened conditions to either select an even better synthetic condition, or help understand why the  $\text{BF}_3 \cdot \text{Et}_2\text{O}$  shows such a marked improvement in synthetic yield. Ultimately, optimization of the RFP fluorophore synthesis will enable more efficient, large-scale synthesis of this desirable fluorophore to subsequently enable design and synthesis of RFP based fluorogenic CLIP-tag probes.

## Appendix A

### Experimental

#### General Methods

All commercially available compounds and solvents were purchased and used without further purification unless otherwise noted. Reactions were not run under strictly anhydrous or inert atmospheric conditions unless otherwise noted. All HPLC purification was performed on an Agilent 1260 Infinity HPLC (Model No. A11) with gradient mixtures of Solvent A (94.5% H<sub>2</sub>O, 4.5% acetonitrile, 1% trifluoroacetic acid) and Solvent B (94.5% acetonitrile, 4.5% water, 1% trifluoroacetic acid) with flow rate 20 mL/min. All <sup>1</sup>H NMR spectra were run on a 500 MHz Bruker AV-III-HD 500 spectrometer.

#### 4-Hydroxymethyl benzylamine (2).

LiAlH<sub>4</sub> (5.79 g, 0.152 mol) was added slowly to 200 mL of anhydrous THF at 0°C. 4-Formylbenzotrile (5.0 g, 38.13 mmol) was dissolved in 40 mL anhydrous tetrahydrofuran to yield a yellow solution. The 4-formylbenzotrile solution was then added dropwise to the solution of LiAlH<sub>4</sub> with stirring at 0°C. After addition was complete, the reaction mixture was brought to reflux under Ar at 50°C for 48 h. The color of the reaction mixture changes from gray to dark green then pea green as the reaction progresses, and the formation of chunky dark green precipitate was also observed. After 48 h, the reaction was quenched by slowly adding 60 mL of H<sub>2</sub>O at 0°C until the color changed from green to white. 20 mL of 10% NaOH<sub>(aq)</sub> was added during the work up, the reaction mixture was filtered, the precipitate was washed with EtOAc, and the yellow filtrate was collected. This filtrate was concentrated, filtered through a pad of Celite, and lyophilized to yield 4.8565 g of crude product as a tacky, sticky, cream-colored solid (92.9%). This material was used without further purification or characterization.

**2,2,2-Trifluoro-N-(4-(hydroxymethyl)benzyl)acetamide (4).**

Crude **2** (5.23 g, 38.13 mmol) was dissolved in 40mL MeOH to afford a yellow solution. Triethylamine (5.31 mL, 38.13 mmol) was added to the solution with stirring. Ethyl trifluoroacetate (5.90 mL, 49.56 mmol) was added dropwise to the stirring reaction mixture. The reaction was allowed to stir at room temperature for 24 h, then the solvent was removed under vacuum to yield a honey colored gel. This residue was dissolved in DCM, dry-loaded onto 15g of silica, then purified by silica column chromatography (gradient hexanes:ethyl acetate from 2:1 to 1:1). The desired product has an  $R_f \sim 0.5$  in 1:1 Hex:EA. The pure fractions were collected and the solvent was removed under vacuum to yield 5.55g of purified product as a white solid (67% yield). Other than TLC analysis, this product was not further characterized.

***N*-(4-((4-aminopyrimidin-2-yl)oxy)methyl)benzyl)-2,2,2-trifluoroacetamide (6).**

**4** (1.591 g, 6.82 mmol) was dissolved in 25mL anhydrous DMF to generate a light yellow solution. 60% NaH dispersion in mineral oil (1.09 g, 27.29 mmol) was added slowly at 0°C, and the reaction mixture changed from yellow to dark teal, dark forest green, pea green, then light yellow green. The mixture was stirred for 30 min at rt. After 30 min, 4-amino-2-chloropyrimidine (1.79 g, 13.81 mmol) was added and the reaction turned a dark blackish-green color. The reaction was refluxed under Ar at 90°C for 24 h until the mixture became an orange-red color. The reaction was quenched with 100 mL H<sub>2</sub>O and extracted with ethyl acetate (5 x 100 mL). The organic extract was evaporated under vacuum and azeotroped with toluene to remove remaining DMF, yielding a sticky orange solid. This solid was purified via flash silica column chromatography (1:20 MeOH:DCM), and 970 mg of purified product was obtained as a light yellow, crystalline solid (43%). <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>, 500 MHz)  $\delta$  10.05 (t, 1H), 8.63 (br s, 2H), 7.98 (d, 1H), 7.41 (m, 4H), 6.33 (d, 1H), 5.43 (s, 2H), 4.40 (d, 2H).

**2-(4-(Aminomethyl)benzyloxy)4-aminopyrimidine (O6-benzylcytosine, 7).**

**6** (970 mg, 2.97 mmol) and  $K_2CO_3$  (2.46 g, 17.84 mmol) were dissolved in 100 mL of 1:24  $H_2O:MeOH$ . The reaction mixture was refluxed under Ar at  $65^\circ C$  for 3 h, then filtered over Celite to remove remaining  $K_2CO_3$ . The filtrate was evaporated under vacuum, redissolved in MeOH to precipitate remaining salts, and filtered again to remove remaining salt byproducts. This crude product solution was purified by preparatory reverse-phase HPLC (2 min at 100% A, gradient to 30% B by 8 min, then gradient to 100% B by 12 min) to afford 380 mg purified benzylcytosine as a fluffy white solid (55.6%).  $^1H$  NMR ( $DMSO-d_6$ , 500 MHz)  $\delta$  8.54 (br s, 2H), 8.35 (br s, 2H), 7.99 (d, 1H), 7.52 (m, 4H), 6.35 (d, 1H), 5.43 (s, 2H), 4.06 (s, 2H).

**4-Aminobutanoic acid methyl ester hydrochloride (C4 linker, 9).**

4-Amino-n-butyric acid (412.5 mg, 4 mmol) was dissolved in 10 mL anhydrous MeOH, and the solution was cooled to  $0^\circ C$ . Thionyl chloride (1 mL, 12 mmol) was added dropwise to the stirring solution at  $0^\circ C$ . The reaction was refluxed at  $65^\circ C$  under Ar for 24 h until the reaction turned into a clear yellow solution. The solvent was removed under vacuum, and the solid was further dried under high vacuum to obtain 400 mg of creamy white solid (85%). The product was used without further purification.  $^1H$  NMR ( $DMSO-d_6$ , 500 MHz)  $\delta$  8.16 (br s, 3H), 3.59 (s, 3H), 2.78 (m, 2H), 2.44 (t, 2H), 1.81 (m, 2H).

**6-Aminohexanoic acid methyl ester hydrochloride (C6 linker, 11).**

6-Amino-n-hexanoic acid (524.7 mg, 4 mmol) was dissolved in 10 mL anhydrous MeOH, and the solution was cooled to  $0^\circ C$ . Thionyl chloride (1 mL, 12 mmol) was added dropwise to the stirring solution at  $0^\circ C$ . The reaction was refluxed at  $65^\circ C$  under Ar for 24 h until the reaction turned into a clear yellow solution. The solvent was removed under vacuum, and the solid was further dried under high vacuum to obtain 570 mg of creamy white solid (98%). The product was used without further purification.  $^1H$  NMR

(DMSO-*d*<sub>6</sub>, 500 MHz)  $\delta$  7.95 (br s, 3H), 3.58 (s, 3H), 2.74 (m, 2H), 2.30 (t, 2H), 1.52 (m, 4H), 1.30 (p, 2H).

#### ***Trans*-4-aminocyclohexylcarboxylic acid hydrochloride (13)**

*Trans*-4-(*boc*-amino)cyclohexane carboxylic acid (973 mg, 4 mmol) was added to 6 mL 4M HCl in 1,4-dioxane. An additional 3 mL 1,4-dioxane was added to help mixing efficiency as the product salt precipitates. The reaction mixture was stirred at rt for 1 h, then the mixture was dried under vacuum and carried forth to the next step without purification or characterization.

#### ***Trans*-4-aminocyclohexylcarboxylic acid methyl ester HCl (cyclohexane linker, 15)**

X7 (719 mg, 4 mmol) was dissolved in 10 mL anhydrous MeOH and cooled to 0°C. Thionyl chloride (1 mL, 12 mmol) was added dropwise to the reaction at 0°C, and the mixture was refluxed under Ar overnight at 65°C until it turned into a clear yellow solution. The reaction mixture was dried under vacuum to afford 734 mg of a creamy white solid (95%). <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>, 500 MHz)  $\delta$  8.18 (br s, 3H), 3.58 (s, 3H), 2.90 (m, 1H), 2.25 (m, 1H), 1.95 (m, 4H), 1.36 (m, 4H).

#### **Schiff base (18).**

4-Dimethylaminobenzaldehyde (2 g, 13.4 mmol), glycine *tert*-butyl ester hydrochloride (4.49 g, 26.8 mmol), and NaOH (1.07 g, 26.8 mmol) were dissolved in 20 mL EtOH to yield an opaque, bright yellow mixture. The reaction mixture was stirred at rt under Ar for 24 h until the color became a creamier yellow. This reaction was not worked-up, and the product was not isolated or characterized.

#### **Imidate (21).**

Glycine methyl ester hydrochloride (10 g, 80 mmol) and K<sub>2</sub>CO<sub>3</sub> (10 g, 80 mmol) were suspended in Et<sub>2</sub>O (100 mL) and H<sub>2</sub>O (50 mL) to yield a biphasic mixture. Ethyl acetamide hydrochloride (10 g, 80

mmol) was added to the mixture, which was stirred extremely vigorously for 6 min. The Et<sub>2</sub>O was separated and collected, fresh Et<sub>2</sub>O (100 mL) was added to the reaction, and the mixture was stirred for an additional 6 min before removing the Et<sub>2</sub>O. Combined Et<sub>2</sub>O layers were dried over anhydrous Na<sub>2</sub>SO<sub>4</sub> (s), and the solution was dried under vacuum without heat to yield a lightly yellow tinted liquid. The product was only dried until no ether could be detected by smell, as drying to completion causes decomposition of the product. The product was not further characterized or purified. The requisite mass was immediately measured and added to the next reaction.

### **Protected GFP Fluorophore (22).**

The concentrated solution of imidate **21** in Et<sub>2</sub>O (2.13 g, 13.4 mmol) was added directly to the reaction mixture of Schiff base **18**. The reaction mixture immediately becomes bright yellow. The reaction was stirred at rt for 24 h until the color turns yellow-orange. H<sub>2</sub>O (20 mL) was added to the reaction mixture, the ethanol was removed under vacuum, and the crude mixture in H<sub>2</sub>O was extracted with ethyl acetate (5 x 100 mL) to afford the crude product as a vibrant orange gel. The crude material was dry-loaded onto silica and purified via flash silica column chromatography in 2:1 Hex:EA to afford 2.03 g of a bright yellow-orange solid (45% yield). <sup>1</sup>H NMR (CDCl<sub>3</sub>, 500 MHz) δ 8.06 (d, 2H), 7.10 (s, 1H), 6.70 (d, 2H), 4.28 (s, 2H), 3.05 (s, 6H), 2.30 (s, 3H), 1.47 (s, 9H).

### **GFP Fluorophore (23).**

Protected GFP fluorophore (796 mg, 2.32 mmol) was dissolved in DCM (5.5 mL). Trifluoroacetic acid (5.5 mL, 69.6 mmol) was added to the stirring solution dropwise, and the color changed from yellow-orange to dark blackish-yellow. The reaction was stirred at rt for 4 h, then azeotroped with toluene to remove excess trifluoroacetic acid. The crude material was dissolved in ethyl acetate (50 mL) and washed with pH 4.2 acetate buffer (3 x 30 mL) to remove trace trifluoroacetic acid. The crude product was then dried under

vacuum and used without further purification.  $^1\text{H}$  NMR (DMSO-*d*<sub>6</sub>, 500 MHz)  $\delta$  8.05 (d, 2H), 6.92 (s, 1H), 6.75 (d, 2H), 4.37 (s, 2H), 3.00 (s, 6H), 2.27 (s, 3H).

### **General Procedure for GFP-Linker Intermediates.**

#### ***GFP-Linker Methyl Esters***

GFP fluorophore (0.50 mmol), the linker molecule (1.00 mmol), EDC • HCl (1.00 mmol), and 1-hydroxybenzotriazole hydrate (1.00 mmol) were dissolved in 4 mL anhydrous DMF. Triethylamine (3.00 mmol) was added to the solution, and the reaction mixture was stirred under Ar at rt for 24 h. The reaction mixture was quenched with H<sub>2</sub>O (5 mL), extracted with DCM (5 x 20 mL), and dried under vacuum to afford the coupled GFP-linker intermediate. This product was not characterized or purified.

#### ***GFP-Linker Carboxylic Acids***

The coupled GFP-linker methyl ester intermediate (0.5 mmol) and KOH (2.5 mmol) were dissolved in 6 mL of 1:1 THF:H<sub>2</sub>O. The reaction mixture was stirred vigorously for 4 h, then quenched with pH 4.2 acetate buffer (20 mL), and extracted with DCM (5 x 50 mL). The organic layers were dried over MgSO<sub>4</sub> (s) and the solvent was removed under vacuum to afford the GFP-linker intermediates with deprotected carboxylic acid functionality. This product was not characterized or purified.

### **General Procedure for GFP-Linker-Benzylcytosine coupling.**

The GFP-linker carboxylic acid intermediate (0.088 mmol), O6-benzylcytosine (0.10 mmol), EDC • HCl (0.176 mmol), and hydroxybenzotriazole hydrate (0.176 mmol) were dissolved in 1 mL anhydrous DMF. Triethylamine (0.528 mmol) was added to the stirring solution. The reaction mixture was then left to stir under Ar at rt for 24 h. The final GFP-linker-benzylcytosine products were purified via preparatory reverse phase HPLC.

**GFP-Cyclohexane-Benzylcytosine Probe (25).**

Purified via reverse-phase HPLC (2 min 100% A, gradient to 100% B over 22 min total). <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>, 500 MHz)  $\delta$  8.84 (br s, 2H), 8.30 (t, 1H), 8.16 (d, 1H), 8.06 (d, 2H), 7.99 (d, 1H), 7.45 (d, 2H), 7.26 (d, 2H), 6.88 (s, 1H), 6.76 (d, 2H), 6.36 (d, 1H), 5.43 (s, 2H), 4.26 (d, 2H), 4.20 (s, 2H), 3.49 (m, 1H), 3.00 (s, 6H), 2.24 (s, 3H), 2.13 (m, 1H), 1.40 (m, 2H), 1.19 (m, 2H).

**GFP-C2-Benzylcytosine Probe (26).**

Purified via reverse-phase HPLC (2 min 100% A, gradient to 19% B by 21 min, hold at 19% B until 25 mi, gradient to 100% B by 30 min). <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>, 500 MHz)  $\delta$  8.73 (br s, 2H), 8.57 (t, 1H), 8.45 (t, 1H), 8.06 (d, 2H), 7.98 (d, 1H), 7.45 (d, 2H), 7.30 (d, 2H), 6.88 (s, 1H), 6.75 (d, 2H), 6.34 (d, 1H), 5.42 (s, 2H), 4.30 (s, 3H), 3.78 (m, 2H), 3.05 (s, 2H), 3.01 (s, 6H), 2.25 (d, 2H).

**GFP-C4-Benzylcytosine Probe (27).**

Purified via reverse-phase HPLC (2 min 100% A, gradient to 100% B over 22 min total). <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>, 500 MHz)  $\delta$  8.83 (br s, 2H), 8.37 (m, 1H), 8.28 (m, 1H), 8.06 (d, 2H), 7.98 (d, 1H), 7.45 (d, 2H), 7.28 (d, 2H), 6.84 (m, 2H), 6.76 (m, 1H), 6.36 (d, 1H), 5.42 (s, 2H), 4.40 (m, 1H), 4.26 (d, 2H), 4.23 (m, 1H), 3.08 (s, 3H), 3.01 (s, 6H), 2.25 (m, 2H), 2.16 (m, 2H), 1.67 (m, 2H).

**GFP-C6-Benzylcytosine Probe (28).**

Purified via reverse-phase HPLC (2 min 100% A, gradient to 100% B over 22 min total). Product was not completely purified using this gradient program, but further purification using different HPLC gradients resulted in product loss. <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>, 500 MHz)  $\delta$  8.82 (br s, 2H), 8.34 (m, 1H), 8.22 (m, 1H), 8.06 (d, 2H), 7.97 (d, 1H), 7.45 (m, 2H), 7.29 (m, 2H), 6.89 (s, 1H), 6.75 (m, 2H), 6.36 (d, 1H), 5.43 (s, 2H), 4.22 (s, 3H), 3.01 (s, 6H), 2.19 (m, 2H), 2.12 (m, 2H), 1.49 (m, 4H), 1.42 (m, 4H), 1.26 (m, 2H).

**GFP-Proline-Benzylcytosine Probe (29).**

Purified via reverse-phase HPLC (2 min 100% A, gradient to 22% B by 24 min, hold at 22% B until 28 min, gradient to 100% B by 31 min). Product was not completely purified using this gradient program. <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>, 500 MHz)  $\delta$  8.74 (br s, 2H), 8.44 (t, 1H), 8.08 (d, 2H), 7.96 (m, 1H), 7.42 (m, 2H), 7.26 (m, 2H), 6.94 (s, 1H), 6.75 (d, 2H), 6.33 (d, 1H), 5.40 (s, 2H), 4.55 (m, 2H), 4.30 (m, 3H), 3.63 (m, 2H), 3.02 (s, 6H), 2.19 (m, 2H), 2.10 (m, 2H), 1.96 (m, 2H), 1.85 (m, 2H).

**General Procedure for *Boc*-protected RFP Fluorophore (31) – Acidic Conditions**

*Boc*-protected GFP fluorophore (80 mg, 0.233 mmol), 4-dimethylamino benzaldehyde (69.5 mg, 0.466 mmol), and the Lewis acid catalyst (0.0233 mmol) were added to 1,4-dioxane dried over molecular sieves (1 mL). The reaction was heated in a sealed pressure vial at 101°C and stirred for 24 h under Ar. The color changed from light orange-yellow to dark ruby-red. The reaction was then quenched with H<sub>2</sub>O (5 mL) and extracted with DCM (5 x 10 mL). The combined organic layers were dried over Na<sub>2</sub>SO<sub>4</sub> (s) and the solvent was removed under vacuum to afford the crude product as a sticky red residue. The crude product was then purified via flash silica column chromatography in 2:1 Hex:EA to afford the purified product which was still a mixture of GFP and RFP fluorophores. The final yield for each condition was calculated from the purified mass and NMR ratio of GFP:RFP.

**General Procedure for *Boc*-protected RFP Fluorophore (31) – Basic Conditions**

*Boc*-protected GFP fluorophore (50 mg, 0.146 mmol), 4-dimethylamino benzaldehyde (108.9 mg, 0.730 mmol), and piperidine (6  $\mu$ L) were dissolved in pyridine (1.5 mL) to afford a light yellow-orange solution. The reaction mixture was heated in a tightly sealed pressure vial at 115°C, and stirred for 24 h under Ar. The reaction turned a red-orange color after 24 h. The reaction mixture was quenched with H<sub>2</sub>O (5 mL) and 10% CuSO<sub>4</sub> (aq) (5 mL), and the product was extracted with DCM (5 x 10 mL). The combined organic layers were dried over Na<sub>2</sub>SO<sub>4</sub> (s) and the solvent was removed under vacuum to afford 194.7 mg

of crude product as a red-orange residue. The crude product was purified via silica gel column chromatography in 2:1 hexanes:ethyl acetate to obtain a GFP/RFP mixture (54.4% NMR yield).

### ***Boc*-protected RFP fluorophore (31).**

To obtain pure RFP fluorophore (*boc*-protected), the GFP/RFP mixtures obtained after purification on a manual silica gel column were re-purified via reverse-phase HPLC (2 min 100% A, gradient to 100% B over 22 min total). <sup>1</sup>H NMR (CDCl<sub>3</sub>, 500 MHz)  $\delta$  8.16 (d, 2H), 7.96 (d, 1H), 7.48 (d, 2H), 7.10 (s, 1H), 6.75 (d, 2H), 6.70 (d, 2H), 6.40 (d, 1H), 4.44 (s, 2H), 3.07 (s, 3H), 3.04 (s, 3H), 1.45 (s, 9H).

### **RFP fluorophore.**

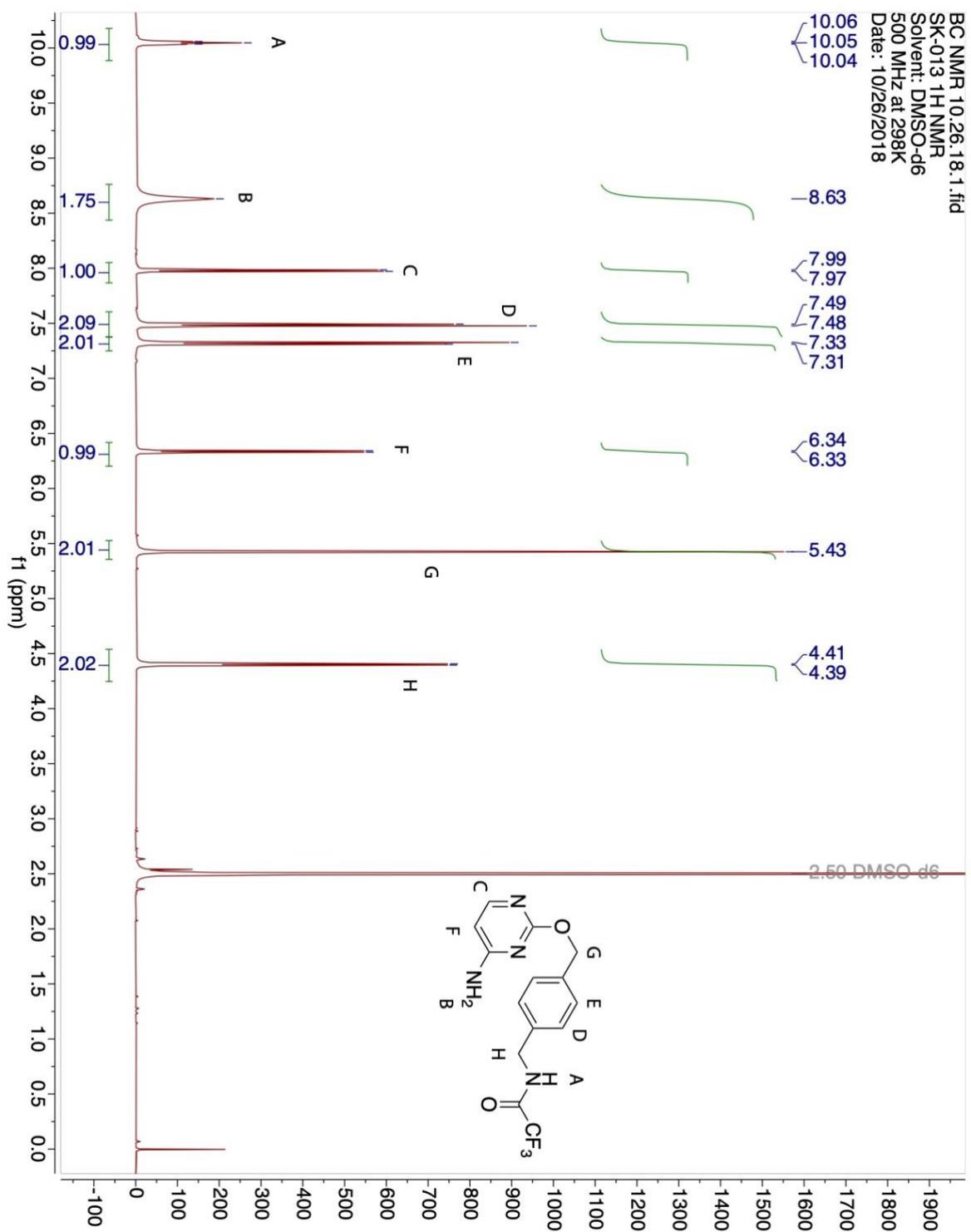
*Boc*-protected RFP fluorophore (70 mg, 0.148 mmol) was dissolved in DCM (0.5 mL), and trifluoroacetic acid (0.5 mL, 6.48 mmol) was added dropwise to the stirring solution. With the addition of acid, the color of the reaction mixture changed from red to dark blackish-yellow. The reaction mixture was stirred at rt for 3 h, then azeotroped with toluene and dried under vacuum to afford the final product as a dark purple-black residue. This product was not further purified or characterized.

### **RFP-Benzylcytosine Probe (32).**

Deprotected RFP fluorophore (83.29 mg, 0.199 mmol), O<sup>6</sup>-benzylcytosine (45.8 mg, 0.199 mmol), EDC • HCl (57.23 mg, 0.299 mmol), and hydroxybenzotriazole hydrate (45.72 mg, 0.299 mmol) were dissolved in 2 mL anhydrous DMF to afford a dark red-purple solution. Triethylamine (83.2  $\mu$ L, 0.60 mmol) was added to the solution, and the reaction was left to stir under Ar at rt for 24 h. The final product was purified via reverse-phase HPLC (2 min 100% A, gradient to 100% B over 22 min total) to afford the final product as a dark red solid. The final purified mass was not calculated as only a portion of the reaction mixture was purified via HPLC. <sup>1</sup>H NMR (DMSO-*d*<sub>6</sub>, 500 MHz)  $\delta$  8.66 (br s, 2H), 8.31 (m, 2H), 8.16 (d,

2H), 7.95 (d, 1H), 7.82 (d, 1H), 7.57 (d, 2H), 7.37 (m, 2H), 7.31 (m, 2H), 6.87 (s, 1H), 6.81 (m, 2H), 6.76 (m, 2H), 6.73 (d, 1H), 6.33 (d, 1H), 5.38 (s, 2H), 4.5 (s, 2H), 4.33 (d, 2H), 3.05 (s, 12H).

**Appendix B**  
**Supplemental Information**

$^1\text{H}$  NMR DataFigure 18. Annotated  $^1\text{H}$  NMR of Protected O6-Benzylcytosine (6).

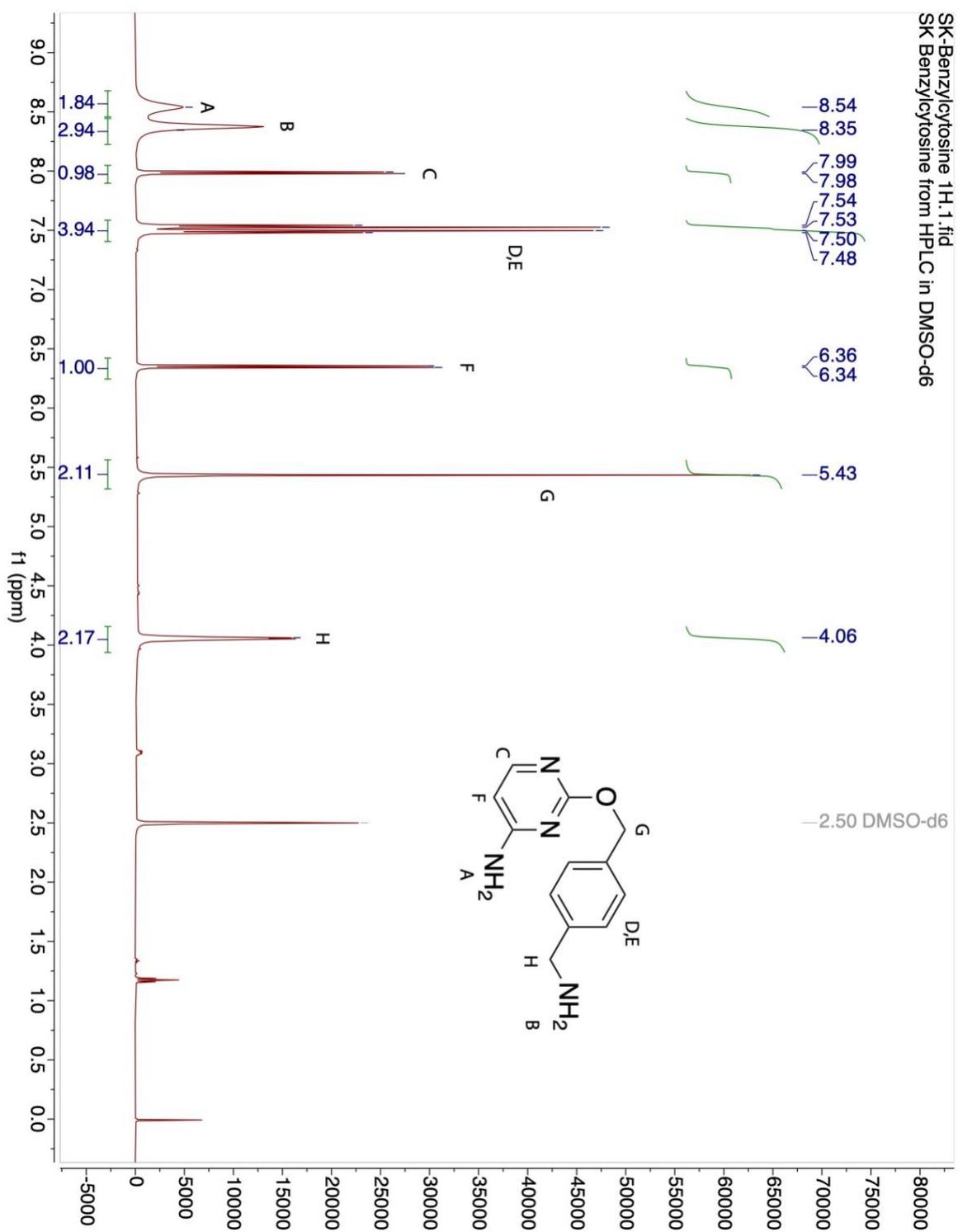
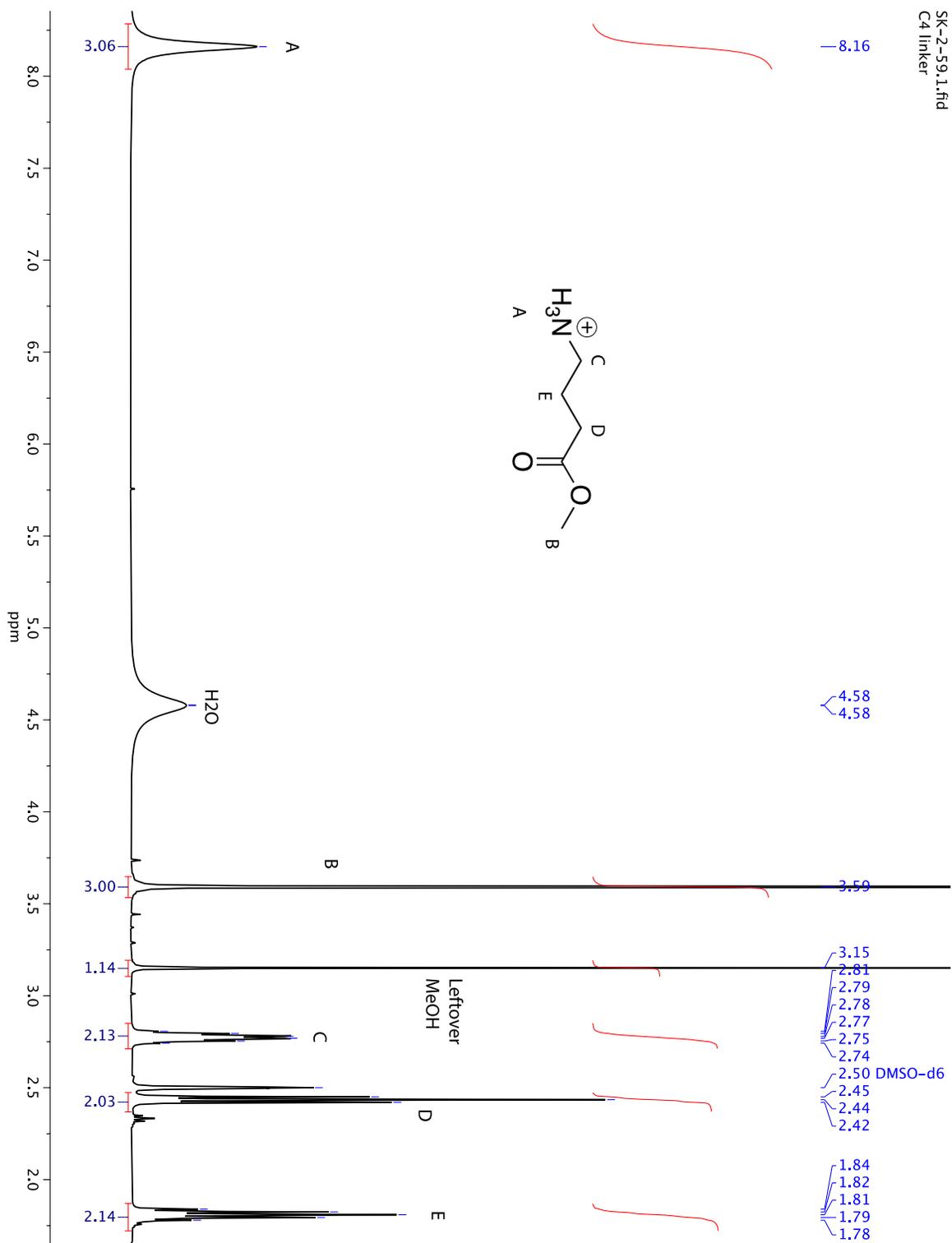
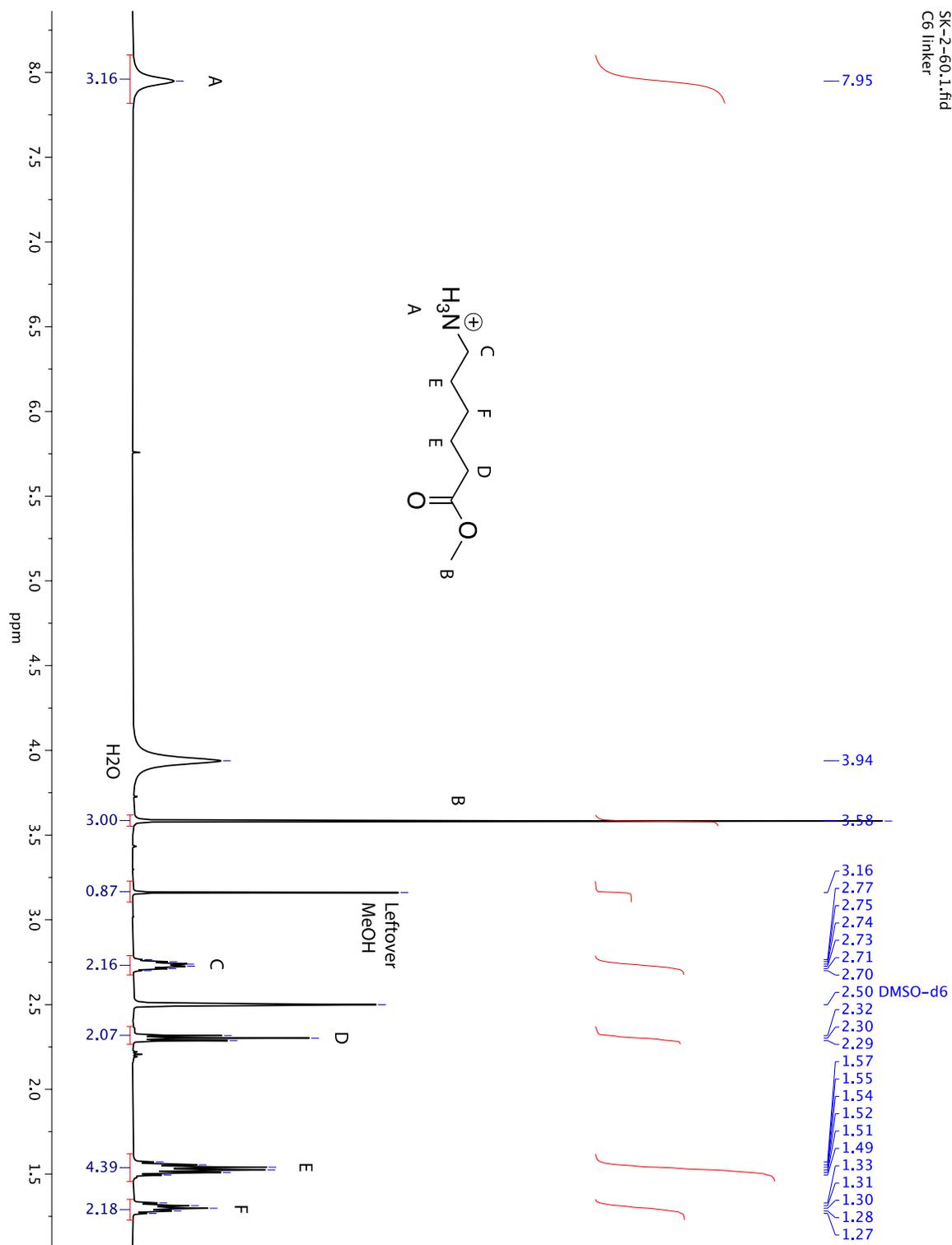
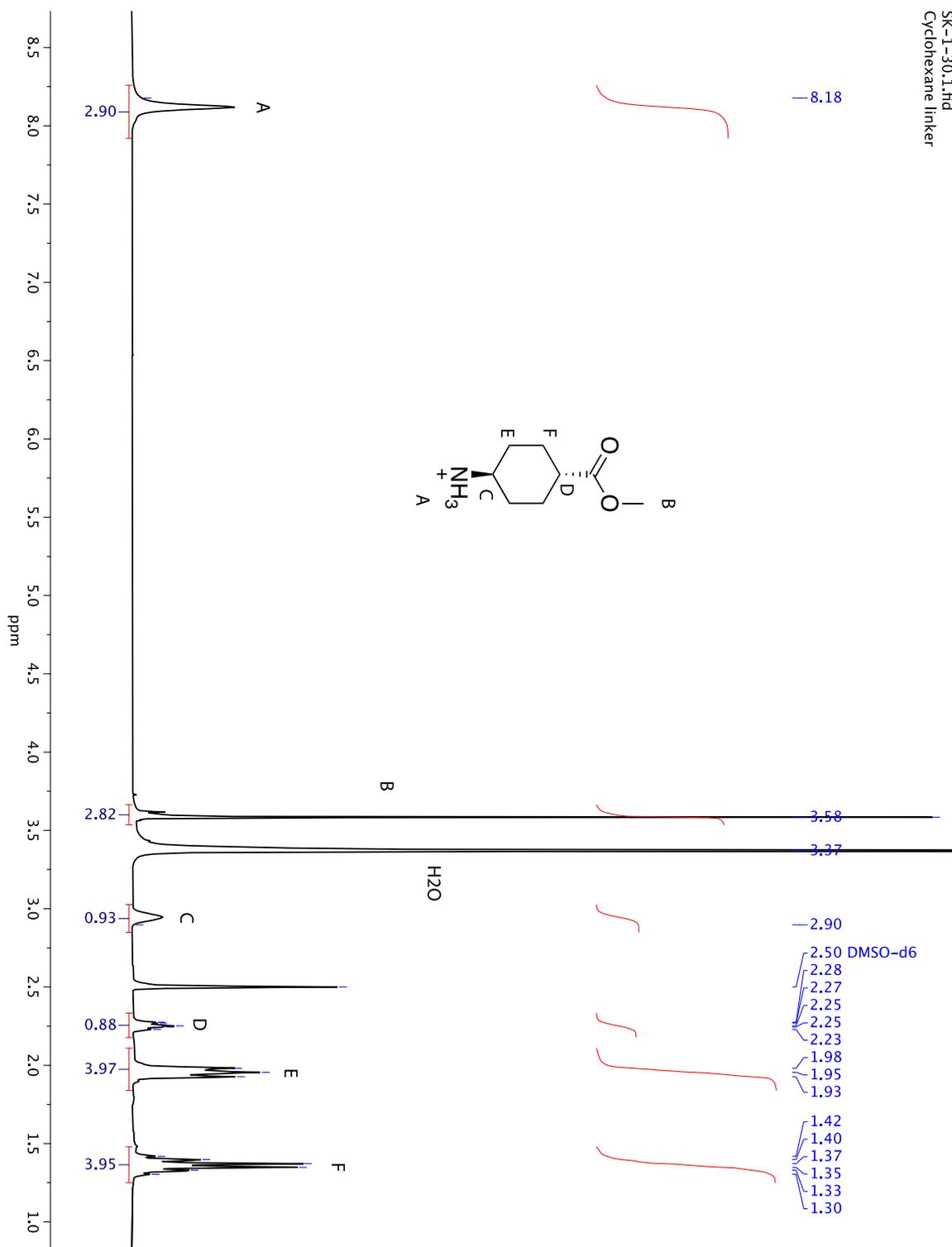


Figure 19. Annotated <sup>1</sup>H NMR of O6-Benzylcytosine (7).

SK-2-59\_1.fid  
C4 linkerFigure 20. Annotated  $^1\text{H}$  NMR of C4 Linker (9).

SK-2-60\_1.fid  
C6 linkerFigure 21. Annotated  $^1\text{H}$  NMR of C6 Linker (11).

SK-1-30.1.fid  
Cyclohexane linkerFigure 22. Annotated <sup>1</sup>H NMR of Cyclohexane Linker (15).

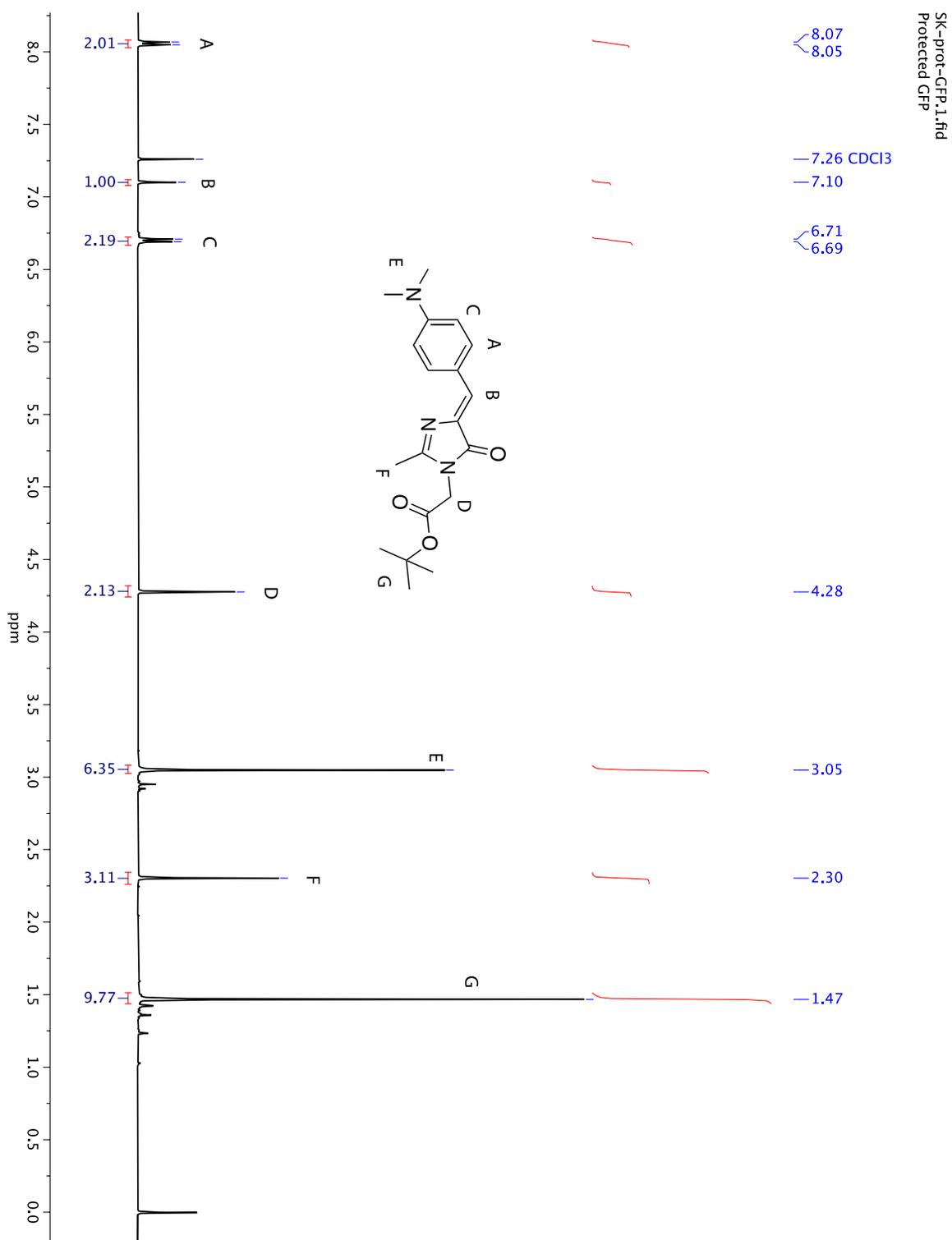


Figure 23. Annotated <sup>1</sup>H NMR of Protected GFP Fluorophore (22).

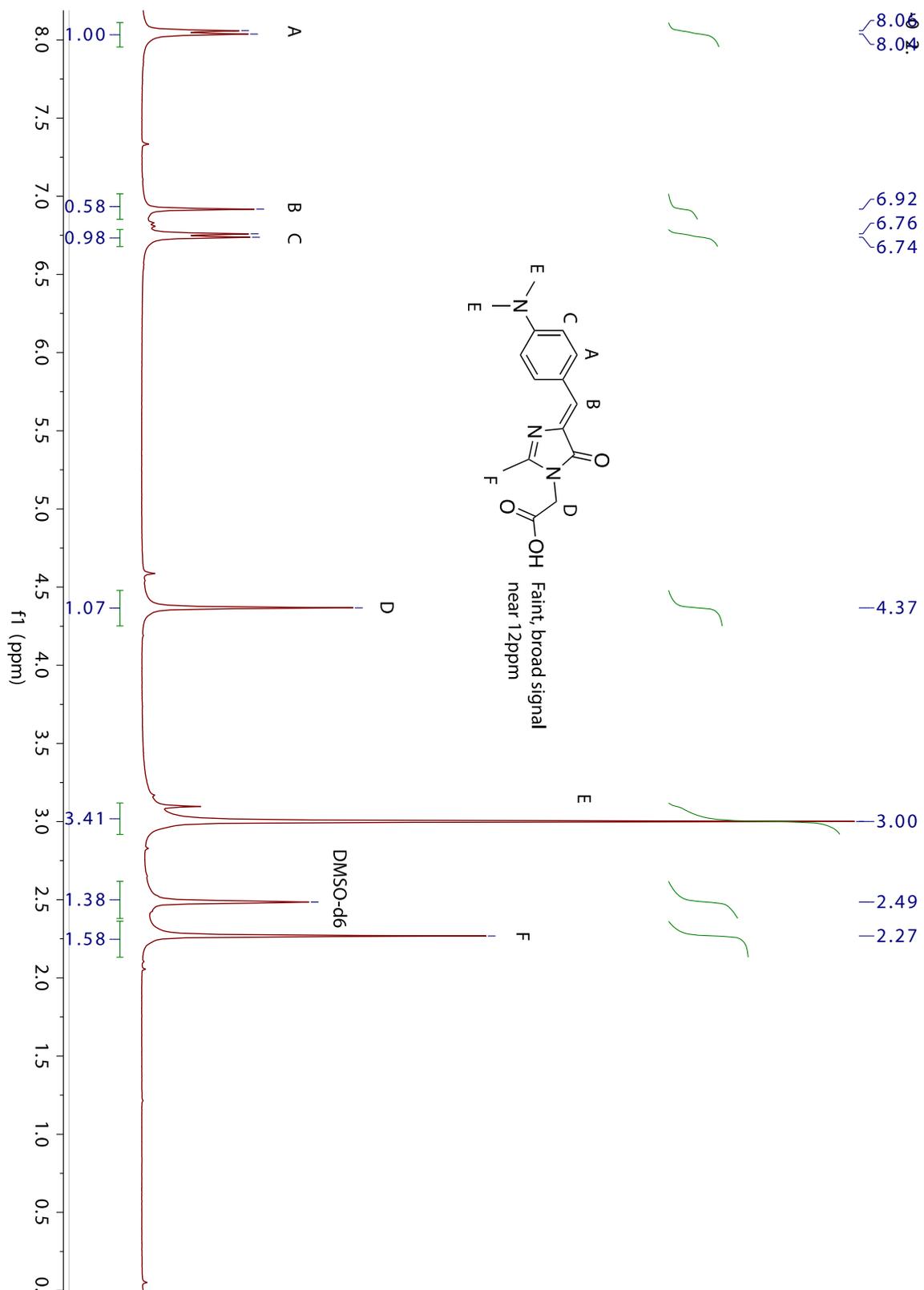


Figure 24. Annotated  $^1\text{H}$  NMR of Deprotected GFP Fluorophore (23).

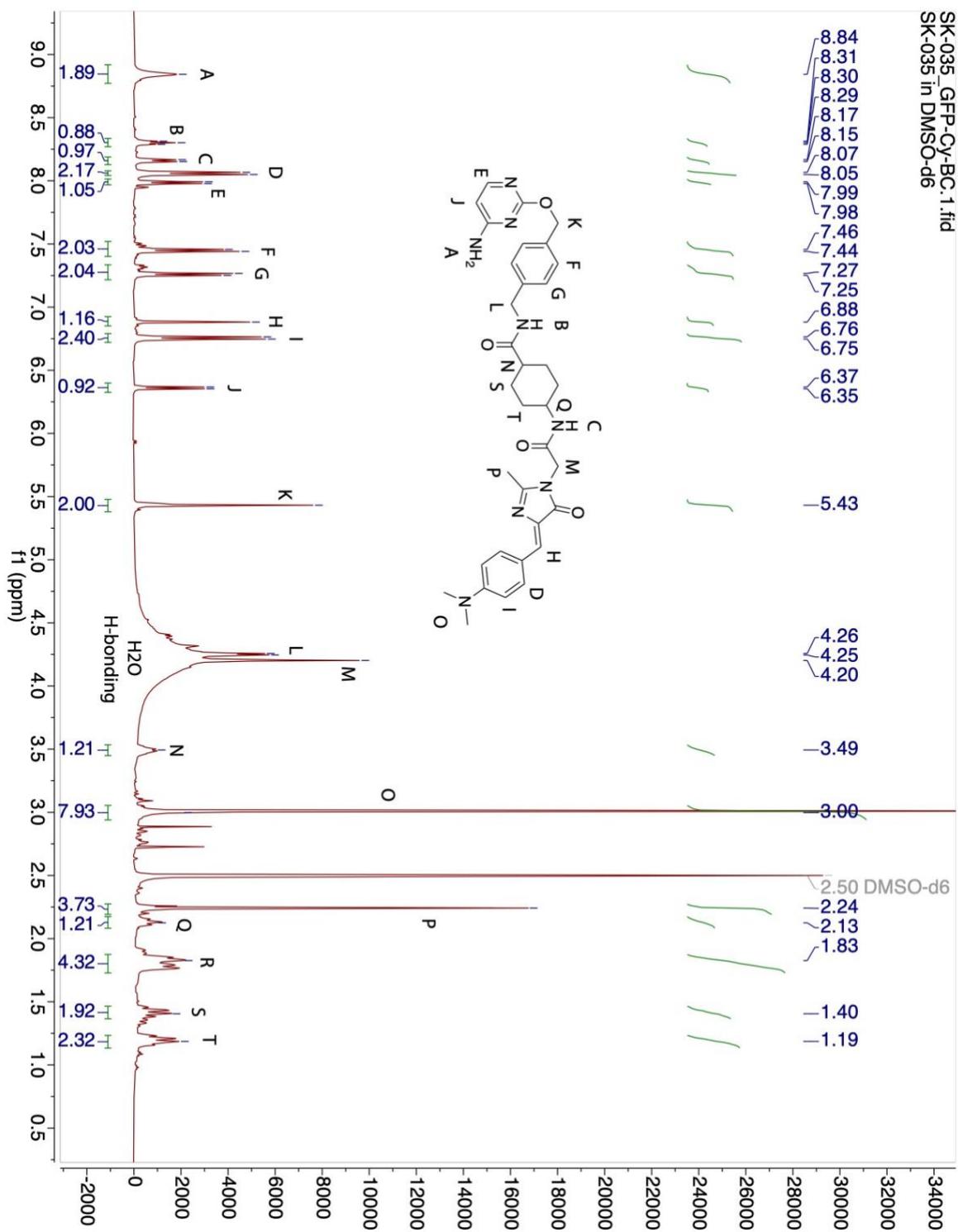


Figure 25. Annotated  $^1\text{H}$  NMR of GFP-Cyclohexane-Benzylcytosine (25).

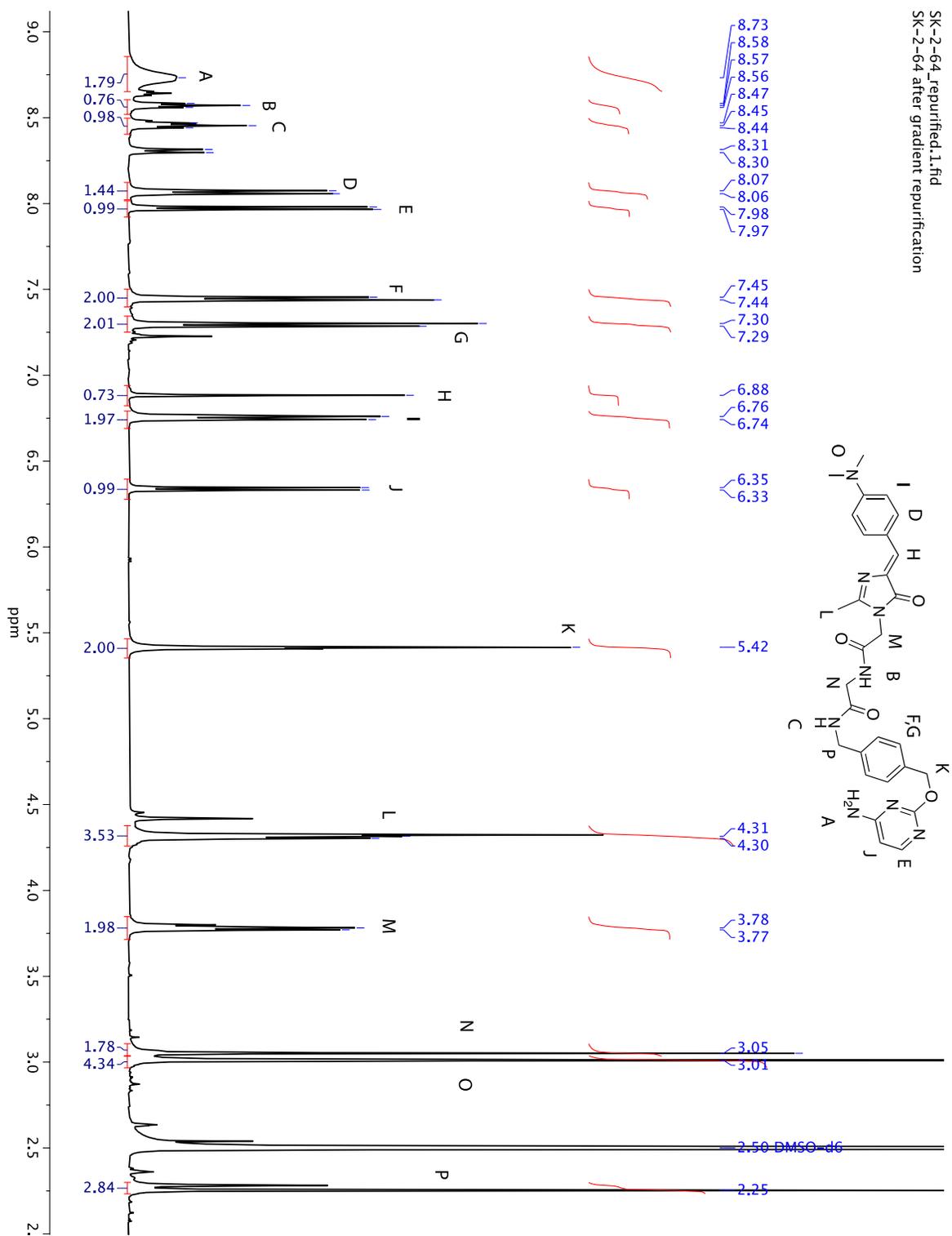


Figure 26. Annotated  $^1\text{H}$  NMR of GFP-C2-Benzylcytosine (26).

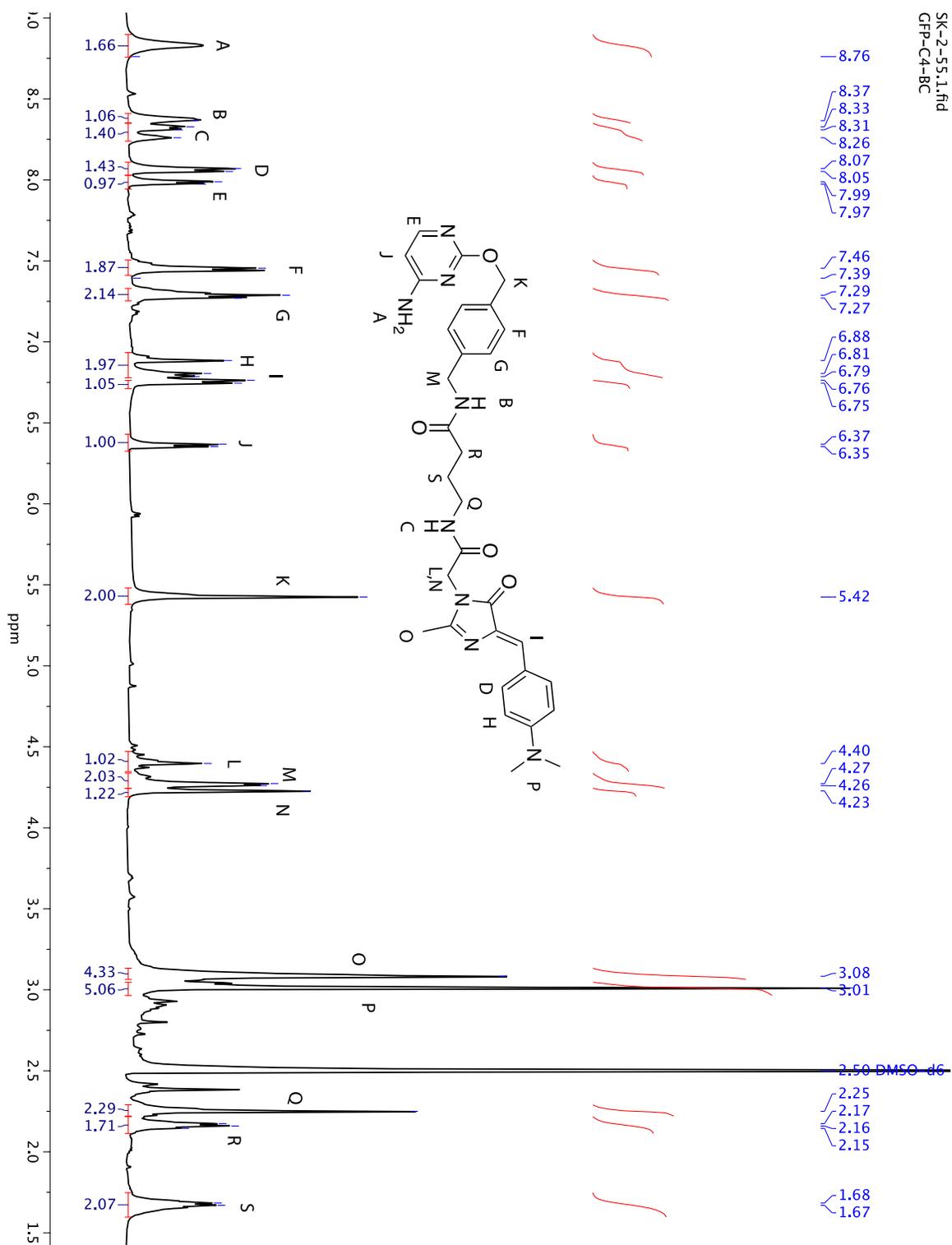


Figure 27. Annotated <sup>1</sup>H NMR of GFP-C4-Benzylcytosine (27).

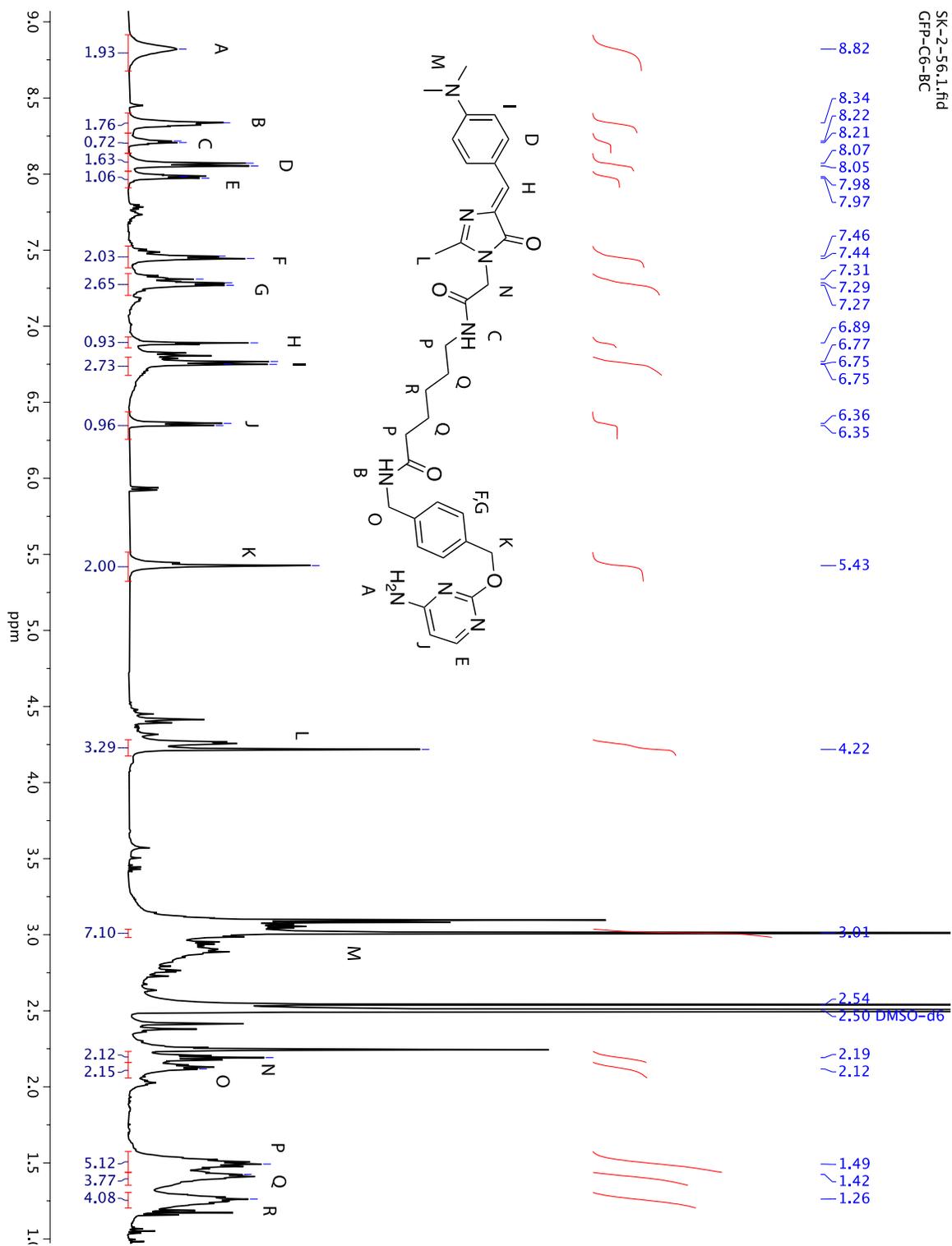


Figure 28. Annotated  $^1\text{H}$  NMR of GFP-C6-Benzylcytosine (28).

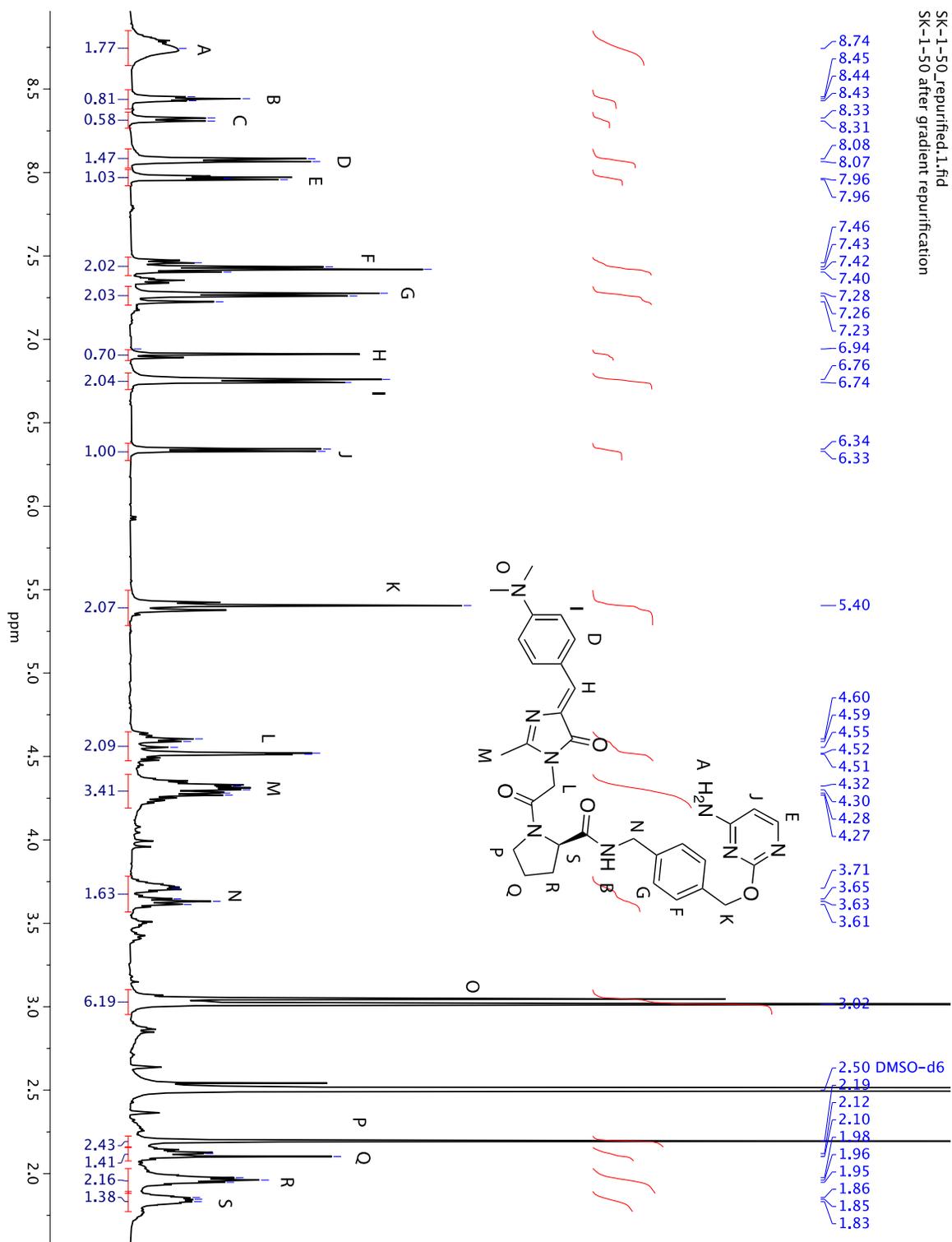


Figure 29. Annotated  $^1\text{H}$  NMR of GFP-Pro-Benzylcytosine (29).

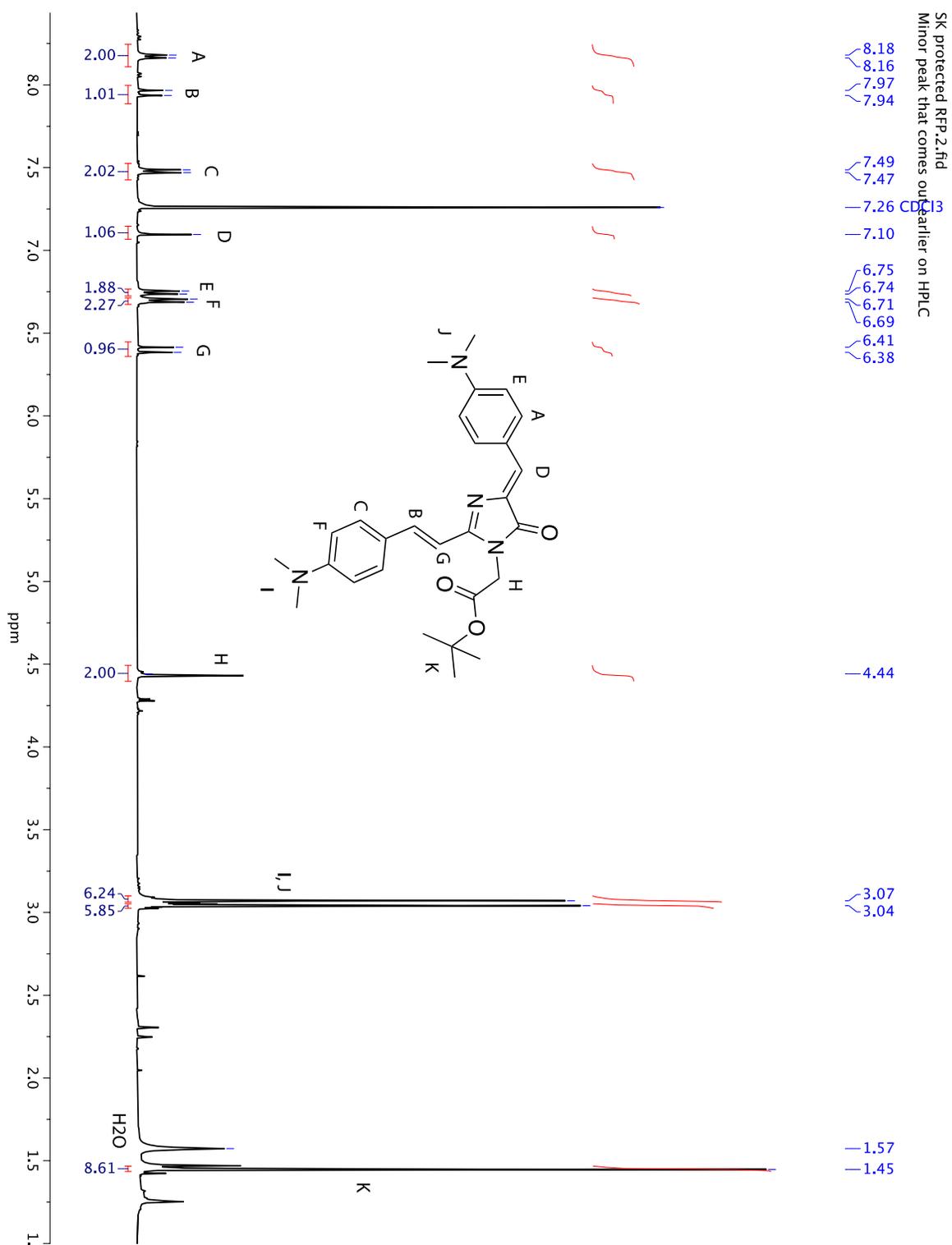


Figure 30. Annotated <sup>1</sup>H NMR of Protected RFP Fluorophore (31).

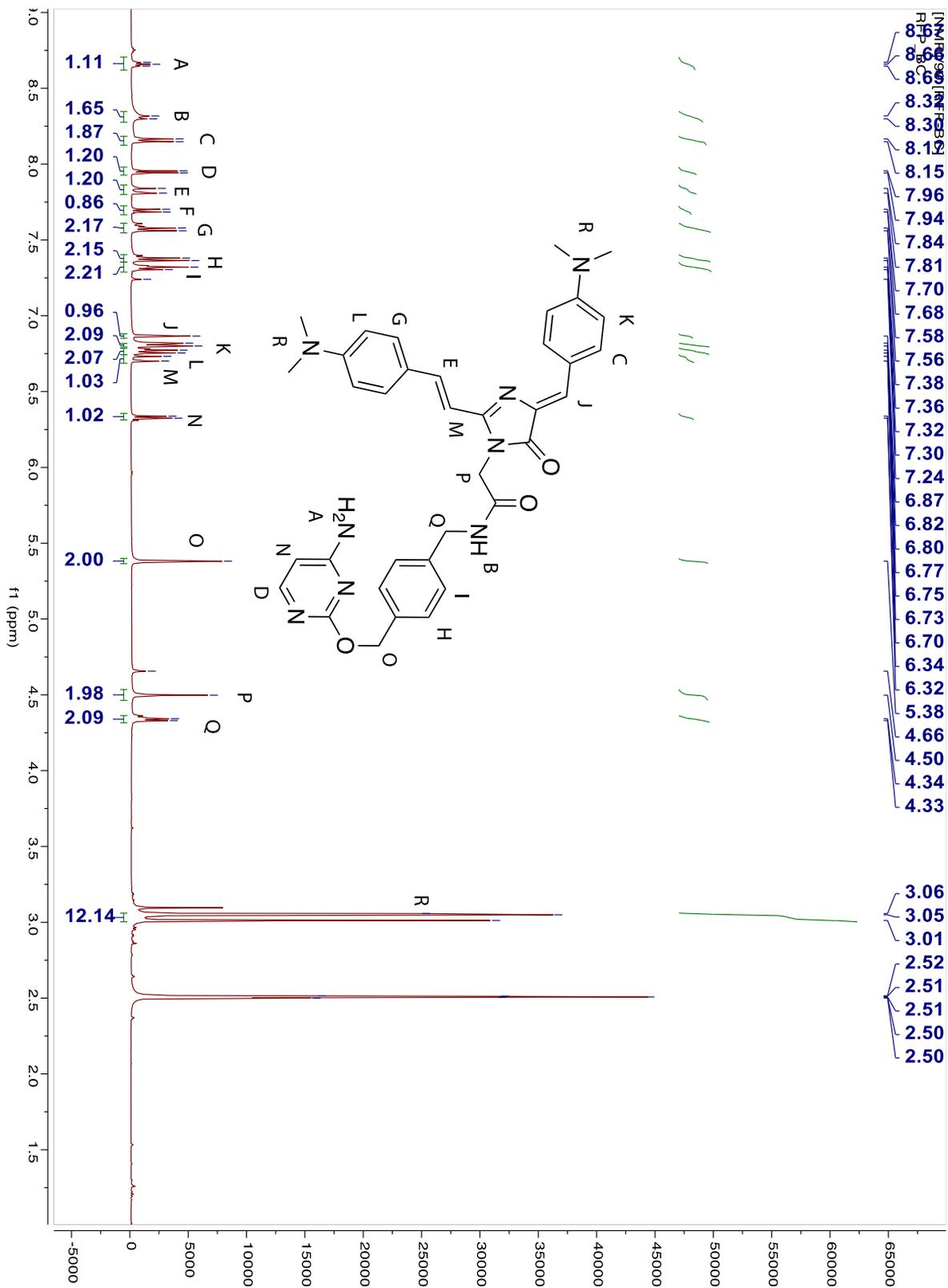


Figure 31. Annotated  $^1\text{H}$  NMR of RFP-Benzylcytosine (32).

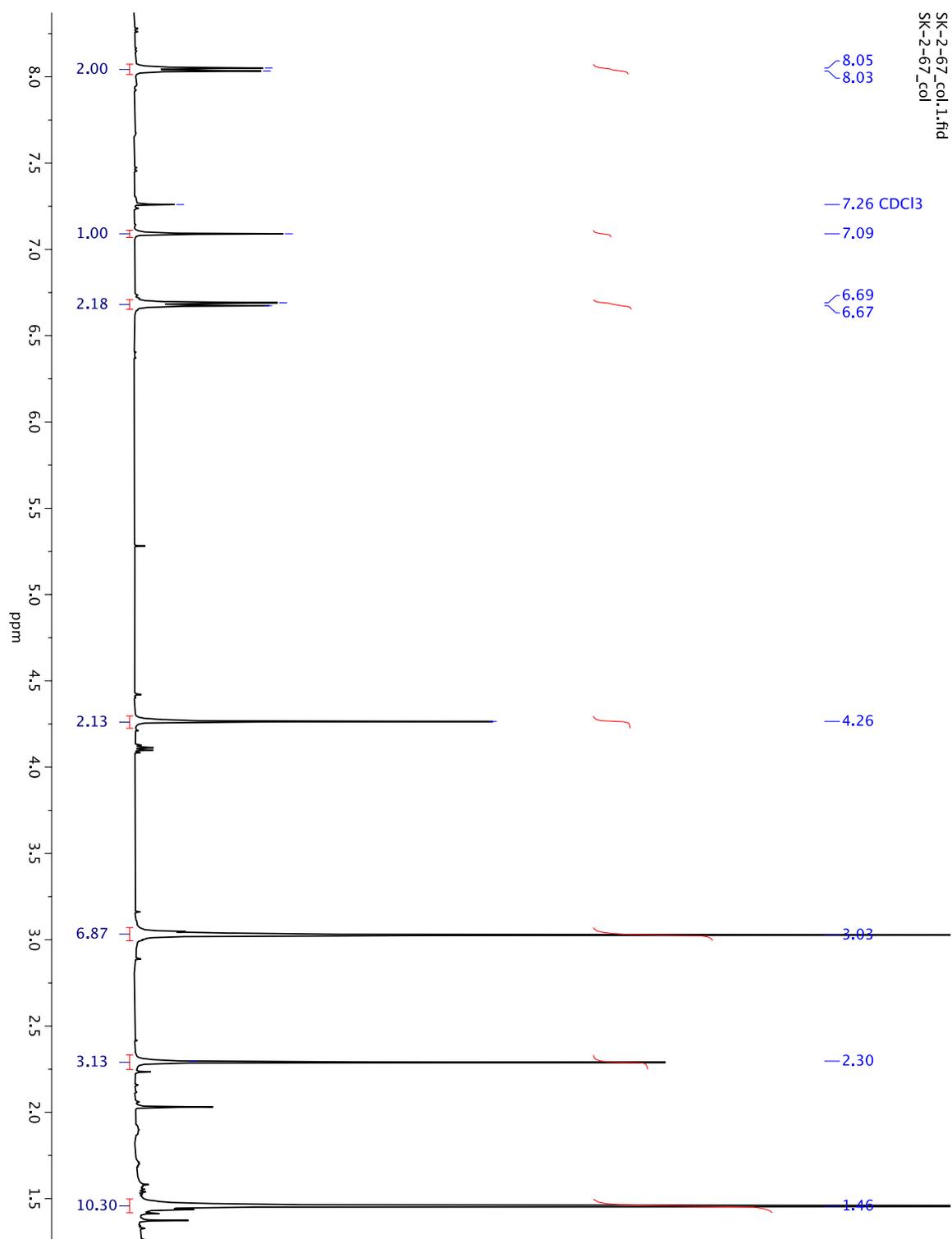


Figure 32.  $^1\text{H}$  NMR RFP Screen: Piperidine in Pyridine.

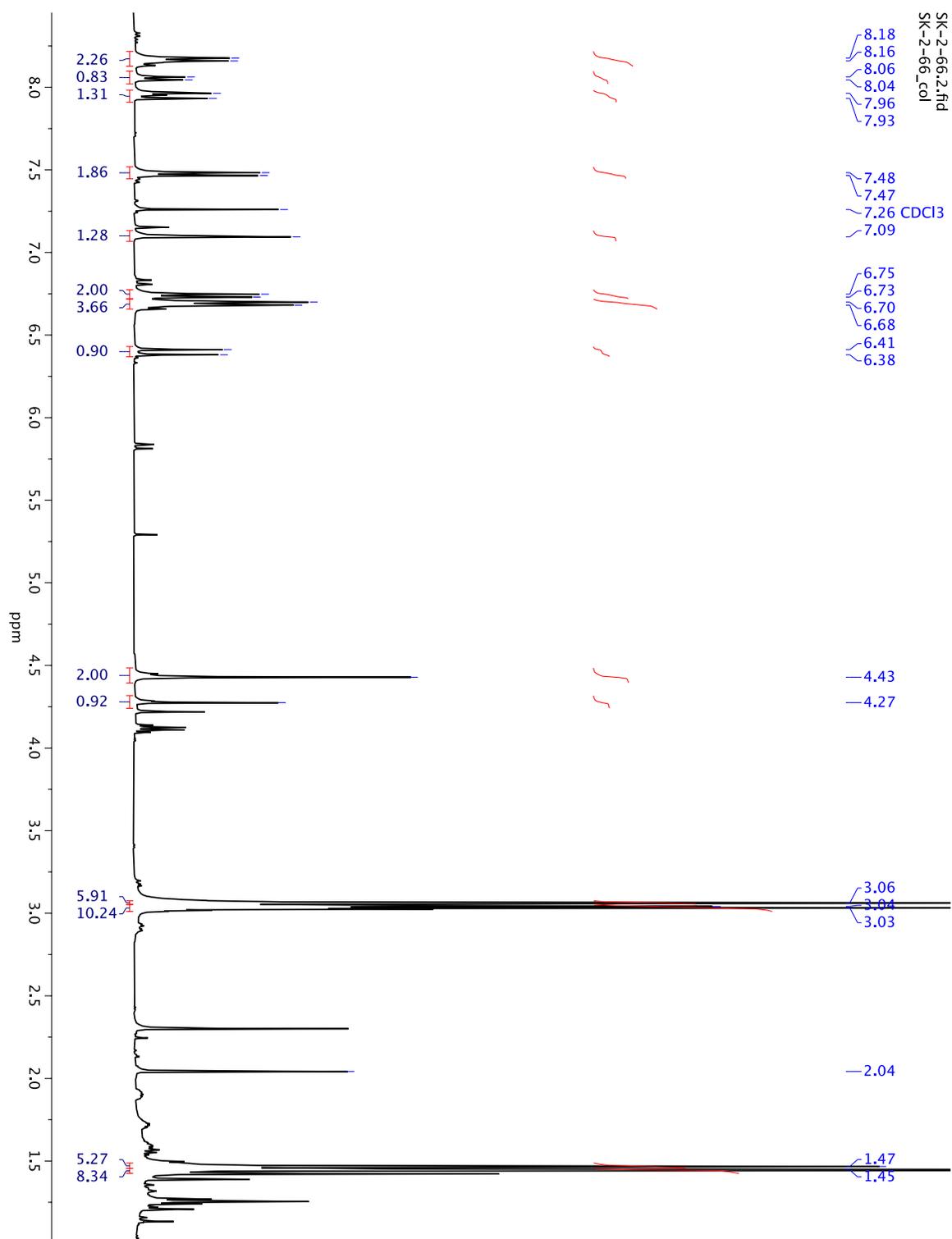


Figure 33. <sup>1</sup>H NMR RFP Screen: 0.5 eq. ZnCl<sub>2</sub> with wet solvent.

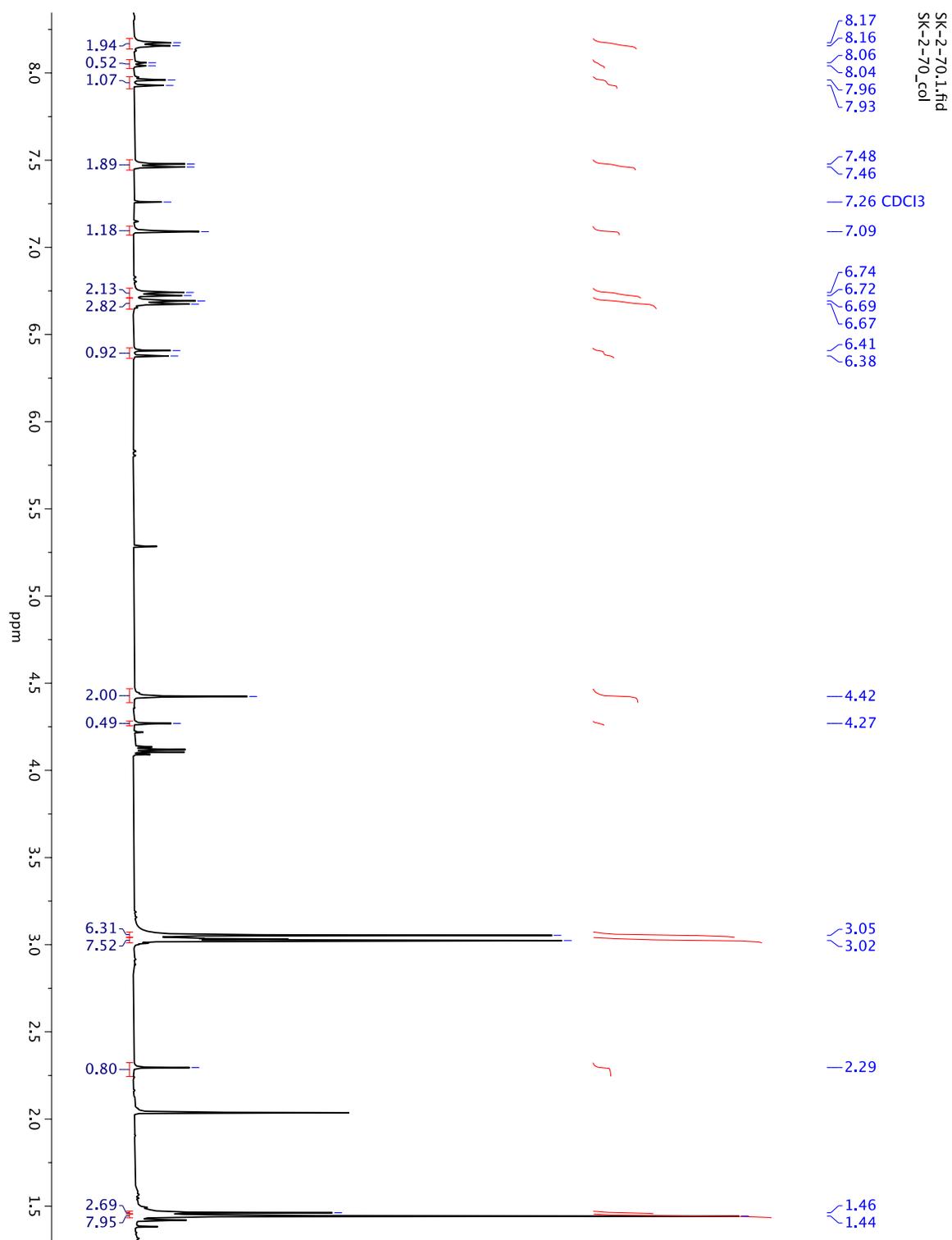


Figure 34.  $^1\text{H}$  NMR RFP Screen: 0.1 eq.  $\text{BF}_3 \cdot \text{Et}_2\text{O}$ .















**BIBLIOGRAPHY**

- (1) Jung, K. H.; Kim, S. F.; Liu, Y.; Zhang, X. A Fluorogenic AggTag Method Based on Halo- and SNAP-Tags to Simultaneously Detect Aggregation of Two Proteins in Live Cells. *ChemBioChem* **2019**, *16802*, 1–11.
- (2) Ross, C. A.; Poirier, M. A. Protein Aggregation and Neurodegenerative Disease. *Nat. Med.* **2004**, *10* (7), S10.
- (3) Ami, D.; Natalello, A.; Lotti, M.; Doglia, S. M. Why and How Protein Aggregation Has to Be Studied in Vivo. *Microb. Cell Fact.* **2013**, *12* (1), 2–5.
- (4) Jing, C.; Cornish, V. W. Chemical Tags for Labeling Proteins inside Living Cells. *Acc. Chem. Res.* **2011**, *44* (9), 784–792.
- (5) Liu, Y.; Wolstenholme, C. H.; Carter, G. C.; Liu, H.; Hu, H.; Grainger, L. S.; Miao, K.; Fares, M.; Hoelzel, C. A.; Yennawar, H. P.; et al. Modulation of Fluorescent Protein Chromophores to Detect Protein Aggregation with Turn-On Fluorescence. *J. Am. Chem. Soc.* **2018**, *140* (24), 7381–7384.
- (6) Hsu, Y. H.; Chen, Y. A.; Tseng, H. W.; Zhang, Z.; Shen, J. Y.; Chuang, W. T.; Lin, T. C.; Lee, C. S.; Hung, W. Y.; Hong, B. C.; et al. Locked Ortho - And Para -Core Chromophores of Green Fluorescent Protein; Dramatic Emission Enhancement via Structural Constraint. *J. Am. Chem. Soc.* **2014**, *136* (33), 11805–11812.
- (7) Komatsu, T.; Johnsson, K.; Okuno, H.; Bito, H.; Inoue, T.; Nagano, T.; Urano, Y. Real-Time Measurements of Protein Dynamics Using Fluorescence Activation-Coupled Protein Labeling Method. *J. Am. Chem. Soc.* **2011**, *133* (17), 6745–6751.
- (8) Gautier, A.; Juillerat, A.; Heinis, C.; Corrêa, I. R.; Kindermann, M.; Beaufils, F.; Johnsson, K. An Engineered Protein Tag for Multiprotein Labeling in Living Cells. *Cell*

- Press Chem. Biol.* **2008**, *15* (2), 128–136.
- (9) Jung, K. H.; Fares, M.; Grainger, L. S.; Wolstenholme, C. H.; Hou, A.; Liu, Y.; Zhang, X. A SNAP-Tag Fluorogenic Probe Mimicking the Chromophore of the Red Fluorescent Protein Kaede. *Org. Biomol. Chem.* **2018**.
- (10) Chuang, W. T.; Chen, B. S.; Chen, K. Y.; Hsieh, C. C.; Chou, P. T. Fluorescent Protein Red Kaede Chromophore; One-Step, High-Yield Synthesis and Potential Application for Solar Cells. *Chem. Commun.* **2009**, No. 45, 6982–6984.
- (11) Gautier, A.; Juillerat, A.; Heinis, C.; Corrêa, I. R.; Kindermann, M.; Beaufils, F.; Johnsson, K. An Engineered Protein Tag for Multiprotein Labeling in Living Cells. *Chem. Biol.* **2008**, *15* (2), 128–136.
- (12) Kaul, R.; Brouillette, Y.; Sajjadi, Z.; Hansford, K. A.; Lubell, W. D. Selective Tert-Butyl Ester Deprotection in the Presence of Acid Labile Protecting Groups with Use of ZnBr<sub>2</sub>. *J. Org. Chem.* **2004**, *69* (18), 6131–6133.
- (13) Chen, C.; Baranov, M. S.; Zhu, L.; Baleeva, N. S.; Smirnov, A. Y.; Zaitseva, S. O.; Yampolsky, I. V.; Solntsev, K. M.; Fang, C. Designing Redder and Brighter Fluorophores by Synergistic Tuning of Ground and Excited States. *Chem. Commun.* **2019**, *55* (17), 2537–2540.

## ACADEMIC VITA

# Sojung F. Kim

---

### EDUCATION

**The Pennsylvania State University**

The Schreyer Honors College | The Millennium Scholars Program

Bachelor of Science in Chemistry

**University Park, PA**

Class of 2020

**State College Area High School** (Class of 2016)

**State College, PA**

### ACADEMIC HONORS & AWARDS

*John & Elizabeth Holmes Teas Scholarship*

June 2019

- Full tuition scholarship awarded to one senior chemistry major at Penn State for outstanding academic achievement, ability, and passion for chemistry

*2019 Barry M. Goldwater Scholarship*

April 2019

*2019 Astronaut Scholarship Nominee*

March 2019

*Evan Pugh Senior Scholar Award*

March 2018

- Academic recognition for students with senior standing in the upper 0.5% of their class

*The Millennium Scholars Program*

June 2016 – May 2020

- Competitive merit-based scholarship program funded by the Howard Hughes Medical Institute designed to prepare students for the pursuit of doctoral degrees and foster diversity in STEM disciplines

*Schreyer Honors College*

June 2016 – May 2020

*Dean's List*

Fall 2016 – May 2020

### RESEARCH EXPERIENCES

*2019 UC Berkeley Amgen Scholar – Synthetic Organic Chemistry*

Summer 2019

*Prof. Richmond Sarpong* – University of California, Berkeley, Berkeley, CA

- Established synthetic progress towards the total synthesis of a *Daphniphyllum* alkaloid
- Developed and completed eight-step synthesis of a late-stage intermediate to enable tests of the key radical cyclization strategy
- Conducted air-free, moisture-free reactions using a Schlenk line & glovebox, purified products by silica gel column chromatography and preparative TLC, characterized compounds by LC-MS,  $^1\text{H}$  &  $^{13}\text{C}$  NMR, various 2D NMR techniques (COSY, HMQC, HMBC), and variable temperature NMR

*Undergraduate Researcher – Synthetic Organic Chemistry & Biochemistry*

Aug. 2018 – May 2020

*Prof. Xin Zhang* – The Pennsylvania State University, University Park, PA

- Synthesized fluorogenic probes to enable fluorescent SNAP- and CLIP-tag protein labeling and visualization for applications in studying protein aggregation in live cells
- Synthesized five red fluorescent and green fluorescent derivatives of the CLIP-tag system to investigate effects of linker length and rigidity on fluorogenic properties
- Optimized synthetic conditions, purified products by silica gel column chromatography and preparative HPLC, characterized by  $^1\text{H}$  &  $^{13}\text{C}$  NMR
- Collaborated with biochemists to use synthesized probes in protein visualization experiments

Princeton SURP-DC Undergraduate Researcher – Transition Metal Catalysis

Summer 2018

*Prof. Brad P. Carrow* – Princeton University, Princeton, New Jersey

- Participated in the Princeton Summer Undergraduate Research Program for Diversity in Chemistry
- Studied synthesis of a novel three-coordinate palladium(II) complex with a fluoride ligand for use in catalyzing aryl C–F bond formation through reductive elimination
- Tested a variety of synthetic routes under inert Schlenk line and glovebox conditions, characterized species using  $^{31}\text{P}$  &  $^{19}\text{F}$  NMR
- Investigated the energetic feasibility of reductive elimination from organopalladium complexes with different ligands using Density Functional Theory calculations

VICHF-IREU Exchange Researcher – Synthetic Organic Chemistry

Summer 2017

*Prof. Kenichiro Itami* – Institute of Transformative Bio-Molecules, Nagoya University, Nagoya, Japan

- Selected as one of five research fellows for the Virtual Institute for C–H Functionalization International Research Experience for Undergraduates (VICHF-IREU) by the Center for Selective C–H Functionalization (CCHF), sponsored by the NSF Science Across Virtual Networks Initiative
- Based on independent literature reviews, designed, synthesized, and characterized ( $^1\text{H}$  NMR, LC-MS, X-ray crystallography) over 40 derivatives of DL1, a small molecule that inhibits strigolactone receptor proteins
- Tested the inhibition efficacy of derivatives through a competitive binding assay
- Enabled deeper understanding of the structure-activity relationship of the inhibitor and discovered a novel molecule 8x more potent than DL1

Undergraduate Researcher – Nanoscale Inorganic Materials

Aug. 2016 – May 2019

*Prof. Thomas E. Mallouk* – The Pennsylvania State University, University Park, PA

- Investigated complexation of antibiotic drug KKL-40 with cyclodextrins (CD); in collaboration with Profs. Ken Keiler and Jim Adair, tested encapsulation of drug-CD complexes into nanoparticles for increased solubility and targeted drug delivery purposes
  - Synthesized drug molecule, screened complexation conditions and quantified complexation using UV/Vis, characterized complexes with  $^1\text{H}$  NMR
  - Studied reaction thermodynamics using Schrödinger Maestro modeling
- Pillared layered perovskite materials with bulky cations to investigate material rigidity and increase surface area for heterogeneous catalysis applications
  - Synthesized and acid exchanged  $\text{HCa}_2\text{Nb}_3\text{O}_{10}$
  - Intercalated and flocculated with varying amounts of different cations such as Keggin ion
  - Characterized interlayer spacing and crystalline structure using powder X-ray diffraction, scanning electron microscopy, and energy dispersive X-ray spectroscopy
- Synthesized and tested a single-atom iridium catalyst for oxygen evolution experiments
- Studied chemisorption of iridium species on perovskite nanosheets using X-ray diffraction, transmission electron microscopy, and isothermal titration calorimetry

## RESEARCH PUBLICATIONS

1. Jung, K.H.; **Kim, S.F.**; Liu, Y.; Zhang, X. A fluorogenic Aggtag method based on Halo- and SNAP-tag to simultaneously detect the aggregation of two proteins in live cells. *ChemBioChem*. **2019**, *20*, 1078-1087. Part of the ChemBioTalents Special Issue.
2. Yoshimura, M.; **Kim, S.F.**; Takise, R.; Kusano, S.; Yagi, A.; Itami, K.; Hagihara, S. Development of potent inhibitors for strigolactone receptor DWARF 14 (*in review*).
3. Alumasa, J.N.; Miller, A.R.; Miller, S.A.; Lewis, A.E.; **Kim, S.F.**; Upon, A.; Mdluli, K.; Keiler, K.C. The Requirement for Ribosomal Protein L12 During Elongation May Be Distinct from that Observed for Ribosome Rescue (*in preparation*).

## SELECTED RESEARCH PRESENTATIONS

- Progress Toward the Total Synthesis of a *Daphniphyllum* Alkaloid via a Nitrogen-Centered Radical Cascade Approach. **S.F. Kim**, B.A. Wright, R. Sarpong. 2019 UC Berkeley Amgen Scholars Program Symposium (Oral Presentation & Poster), Berkeley, CA, Aug. 14, 2019.
- Synthesis of a Sterically Hindered Three-Coordinate Pd(II) Complex for C–F Reductive Elimination. **S.F. Kim**, L. Wang, L. Chen, B.P. Carrow. Penn State 2019 Undergraduate Exhibition, April 17, 2019.
- Synthesis of a Sterically Hindered Three-Coordinate Pd(II) Complex for C–F Reductive Elimination. **S.F. Kim**, L. Wang, L. Chen, B.P. Carrow. Division of Inorganic Chemistry, 257th American Chemical Society National Meeting, Orlando, FL, March 31-April 4, 2019.
- Synthesis of a Sterically Hindered Three-Coordinate Pd(II) Complex for C–F Reductive Elimination. **S.F. Kim**, L. Wang, L. Chen, B.P. Carrow. Fall 2018 Penn State Eberly College of Science Undergraduate Research Poster Exhibition, Sept. 25, 2018.
- Synthesis of a Sterically Hindered Three-Coordinate Pd(II) Complex for C–F Reductive Elimination. **S.F. Kim**, L. Wang, L. Chen, B.P. Carrow. 8th Annual Undergraduate Research Poster Symposium of the Central Pennsylvania American Chemical Society Local Section, Penn State University, Sept. 4, 2018.
- Synthesis of a Sterically Hindered Three-Coordinate Pd(II) Complex for C–F Reductive Elimination. **S.F. Kim**, L. Wang, L. Chen, B.P. Carrow. Princeton University Chemistry Undergraduate Research Poster Symposium, Aug. 9, 2018.
- Synthesis of a Novel Small Molecule Inhibitor of D14-type Strigolactone Receptors. **S.F. Kim**, R. Takise, M. Yoshimura, S. Hagihara, K. Itami. Penn State 2018 Undergraduate Exhibition, April 18, 2018.
- Synthesis of a Novel Small Molecule Inhibitor of D14-type Strigolactone Receptors. **S.F. Kim**, R. Takise, M. Yoshimura, S. Hagihara, K. Itami. Division of Organic Chemistry, 255th American Chemical Society National Meeting, New Orleans, LA, March 18-22, 2018.
- Synthesis of a Novel Small Molecule Inhibitor of D14-type Strigolactone Receptors. **S.F. Kim**, R. Takise, M. Yoshimura, S. Hagihara, K. Itami. Annual Biomedical Research Conference for Minority Students (ABRCMS), Phoenix, AZ, Nov. 1-4, 2017.
  - Received an outstanding poster presentation award
  - Abstract selected as an exemplary abstract from 2017 Student Awardees, featured as a sample for future attendees
  - Invited as a panelist for the ABRCMS “Writing a Compelling Abstract” webinar series
- Synthesis of a Novel Small Molecule Inhibitor of D14-type Strigolactone Receptors. **S.F. Kim**, R. Takise, M. Yoshimura, S. Hagihara, K. Itami. Penn State Eberly College of Science Undergraduate Research Poster Session, Oct. 5, 2017.

## FELLOWSHIPS & GRANTS

<i>Penn State Undergraduate Conference Travel Grant (\$1200, \$1400)</i>	Mar. 2019, Nov. 2017
<i>PSU Office of Science Engagement Undergraduate Research Grant (\$500)</i>	Nov. 2018
<i>Schreyer International Study &amp; Engagement Grant (\$1100)</i>	May 2018
<i>Penn State Student Engagement Network Grant (\$1500)</i>	April 2018
<i>Eberly College of Science Global Experiences Scholarship (\$1200)</i>	Feb. 2018
<i>DAAD RISE Germany Fellowship (declined to accept Princeton SURP)</i>	Feb. 2018

## STUDY ABROAD EXPERIENCE

*Biology of Eco-Health (BIOL 498A) – Study Abroad in Tanzania* May 2018 – June 2018

- Camped, backpacked, and traveled throughout rural Tanzania for three weeks (Arusha, Karatu, Loliondo, Moshi, Ngorongoro, Serengeti, Yaeda Valley)
- Studied human-environment interactions, the development of zoonotic diseases, conservation of natural resources, and other aspects of regional ecology and culture
- Contributed to long-term data sets on land use and ground cover change
- Gained real-world perspective on global, bio-health challenges and how historical, cultural, political, and traditional factors contribute to human-environment conflicts and interactions

## LEADERSHIP ACTIVITIES

*Search Committee Member for Dean of the Eberly College of Science* Spring 2020

- Invited to serve on this search committee as the representative for the Eberly College of Science undergraduate student body
- Attended search committee meetings and communicated with a hired search firm to establish a diverse candidate pool, reviewed and discussed candidate applications, helped interview select candidates, and made final recommendations to the Provost

*Science LionPride – The Pennsylvania State University* Sept. 2016 – May 2020

- Selective student ambassador group that represents the Eberly College of Science to prospective students, parents, and alumni, while also engaging in community service, science outreach, and fundraising for pediatric cancer research
- *Science Outreach Chair* (2017 - 2018): organized community outreach events to engage with younger students and community members about science and research

*Nittany Chemical Society – The Pennsylvania State University* Sept. 2016 – May 2020

- Student chapter of the American Chemical Society
- Organized and participated in chemistry themed outreach events for younger students and adult learners in the Central Pennsylvania community
- *Treasurer* (2018 –2019): organized funding for outreach events and conference travel

## TEACHING EXPERIENCES

*Teaching Assistant for Structural Analysis of Organic Compounds (Chem 430)* Spring 2019

- Hold 4-5 weekly office hours to teach ~44 upper-class chemistry majors about spectroscopic techniques like UV/Vis spectroscopy, IR spectroscopy, mass spectrometry, <sup>1</sup>H NMR, <sup>13</sup>C NMR, and various 2D NMR techniques
- Run workshops to review chemistry fundamentals
- Grade homework problem sets and help professor design course materials

*Grader for Transition Metal Chemistry (Chem 412)* Fall 2019

- Paid job grading weekly problem sets, quizzes, and exams for a senior level chemistry course

*Physics Tutor for Calculus-based Mechanics, Electricity & Magnetism* Jan. 2018 – Jan. 2019

- Paid job within Millennium Scholars Program tutoring students during weekly physics review and homework sessions

*Grader for Honors General Chemistry and Organic Chemistry* Aug. 2017 – Dec. 2018

- Paid job grading weekly problem sets, quizzes, and exams

## COMMUNITY SERVICE & OUTREACH WORK

*Exploration-U 2019* Nov. 12, 2019

- Worked in a team of four Ph.D. students to develop and run a hands-on science outreach activity focused on teaching elementary & middle school students about water quality (pH, turbidity, and nitrate levels)

*Haunted-U Chemistry Demo Show* Oct. 2016, 2017, 2018, 2019

- Worked with Science-U as part of the Nittany Chemical Society to inspire elementary school age students to engage with chemistry by organizing and acting in a Halloween chemistry demo show with interactive activities

*THON Explorer – The Magic of Science* Feb. 2018, 2019

- Organized and led an interactive science demo show for 15 pediatric cancer patients and their families during THON, explaining “magic” with science (lava lamps, slime, elephant toothpaste, card tricks, homemade ice cream, etc.)

*Think Outside the Beaker* Sept. 2017, Oct. 2018

- Helped organize and lead an interactive polymer chemistry themed workshop (2017) and a workshop on quantum chemistry and electrochemistry (2018) for middle school students

*Penn State Scholars Day* Sept. 2018, 2019

- Selected by the PSU Chemistry Department to serve as a chemistry representative during this recruitment event for potential Schreyer Honors College students

*Women in STEM Exposition – Penn State Society of Women Engineers* April 15, 2018

- Ran a hands-on demonstration about everyday polymer properties using “diaper polymer” for elementary and middle school girl scouts interested in STEM

*Volunteer Debate Coach at State College Area High School* Nov. 2016 – May 2019

- Taught students debate skills, including how to argue effectively and work under time constraints, and encouraged logical, critical thinking and research oriented persuasion
- Traveled with debate team to tournaments as a judge

## PROFESSIONAL ORGANIZATIONS

*Phi Beta Kappa Honor Society* Nov. 2019 – Present

*American Chemical Society* Oct. 2017 – Present

*Korean-American Scientists and Engineers Association* Sept. 2016 – Present