

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Interfacing chemistry and biology for new approaches to site-selective peptide modification and mycobacteria-specific fluorogenic probes

Permalink

<https://escholarship.org/uc/item/9p20w7s5>

Author

Keyser, Samantha Grace Lyons

Publication Date

2018

Peer reviewed|Thesis/dissertation

Interfacing chemistry and biology for new approaches to site-selective peptide modification and mycobacteria-specific fluorogenic probes

By

Samantha Grace Lyons Keyser

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Carolyn R. Bertozzi, Co-Chair

Professor David E. Wemmer, Co-Chair

Professor Matthew B. Francis

Professor Susan Marqusee

Spring 2018

Interfacing chemistry and biology for new approaches to site-selective peptide modification and mycobacteria-specific fluorogenic probes

© 2018 Samantha Grace Lyons Keyser

Abstract

Interfacing chemistry and biology for new approaches to site-selective peptide modification and mycobacteria-specific metabolic probes

by

Samantha Grace Lyons Keyser

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor Carolyn R. Bertozzi, Co-Chair

Professor David E. Wemmer, Co-Chair

Selective labeling of biological matter enables not only basic research but also improved diagnostic platforms and therapeutics. Bioorthogonal chemistry provides a set of tools for researchers to study biomolecules of interest by inserting selectively reactive chemical partners into living systems. In the case of proteins, genetic manipulation can allow the introduction of uniquely reactive peptide sequences for small-molecule or enzymatic labeling. For molecules that cannot be genetically manipulated, metabolic probes provide an attractive alternative; by exploiting gaps in substrate specificity, a biomolecule can be modified with a bioorthogonal functional group without the need to extensively engineer the living system. In some cases, these biomolecules are organism- or cell type-specific, allowing them to be used as diagnostics for cancer or bacterial infection.

In Chapter 1 of this dissertation, I provide an overview of commonly used methods for selective protein modification as well as for selective labeling of the bacterial cell wall. In Chapter 2, I describe the computation-guided rational design of a cysteine- and lysine-containing 11-residue peptide sequence that reacts with 2-cyanobenzothiazole (CBT) derivatives. Our data show that the cysteine residue reversibly reacts with the nitrile group on the CBT moiety to form an intermediate thioimidate, which undergoes irreversible $S \rightarrow N$ transfer to the lysine residue, yielding an amidine-linked product. The concepts outlined herein lay a foundation for future development of peptide tags in the context of site-selective modification of lysine residues within engineered microenvironments. In Chapter 3, I outline our ongoing efforts to optimize this motif for protein labeling. In Chapter 4, I describe the synthesis and characterization of several mycobacteria-selective environment-sensitive fluorogenic probes as well as their potential applications in studies of host-pathogen interactions and diagnostics.

To my parents and my cats

(and the dog)

Interfacing chemistry and biology for new approaches to site-selective peptide modification and mycobacteria-specific fluorogenic probes

Table of Contents

List of Figures..... v

List of Schemes..... ix

List of Tables..... x

List of Abbreviations..... xi

Acknowledgments.....xiv

Chapter 1. Chemical biology approaches to selective labeling of cellular components

Introduction..... 2

 Bioorthogonal chemistry..... 2

 Aldehydes/ketones and hydrazines/hydrazides/alkoxyamines..... 2

 Azides and phosphines/alkynes/cycloalkynes..... 4

 Tetrazines and trans-cyclooctenes/norbornenes/cyclopropenes..... 5

 Other transition metal-mediated reactions..... 6

 Methods for selective protein labeling..... 6

 Protein fusion tags..... 6

 Amino acid-specific modifications..... 7

 Engineered peptide tags..... 10

 Summary and outlook for site-selective protein labeling..... 14

 Metabolic labeling of bacterial cells..... 15

 The bacterial cell envelope..... 15

| | |
|---|----|
| Metabolic probes of the bacterial cell envelope..... | 17 |
| Fluorogenic probes..... | 18 |
| Summary and outlook for metabolic probes of bacterial cell envelope components..... | 19 |
| References..... | 20 |

Chapter 2. Computation-guided rational design of a peptide motif that reacts with cyanobenzothiazoles via internal cysteine-lysine relay

| | |
|-----------------------------|----|
| Introduction..... | 29 |
| Results and Discussion..... | 32 |
| Conclusions..... | 48 |
| Materials and Methods..... | 49 |
| Supplementary Figures..... | 62 |
| References..... | 81 |

Chapter 3. Toward a method for site-selective protein modification based on a peptide motif that reacts with cyanobenzothiazoles via internal cysteine-lysine relay

| | |
|-----------------------------|-----|
| Introduction..... | 85 |
| Results and Discussion..... | 87 |
| Conclusions..... | 96 |
| Materials and Methods..... | 97 |
| Supplementary Figures..... | 115 |
| References..... | 123 |

Chapter 4. Synthesis of solvatochromic probes to label the mycobacterial cell wall

| | |
|-----------------------------|-----|
| Introduction..... | 127 |
| Results and Discussion..... | 132 |
| Conclusions..... | 156 |
| Materials and Methods..... | 157 |
| References..... | 170 |

Appendix

| | |
|---|-----|
| LC-MS time-course spectra..... | 176 |
| ¹ H and ¹³ C NMR spectra..... | 203 |

List of Figures

- Figure 1.1.** Aldehydes and ketones can be reacted with hydrazines, hydrazides, and alkoxyamines to attach probes to biomolecules. 3
- Figure 1.2.** Azides can be reacted with triarylphosphines (the Staudinger ligation), terminal alkynes (copper-catalyzed azide-alkyne cycloaddition, CuAAC), and cyclooctynes (strain-promoted AAC, SPAAC) to introduce fluorophores or affinity tags to a biomolecule. 4
- Figure 1.3.** Tetrazines react rapidly with *trans*-cyclooctenes, cyclopropenes, and norbornenes via an inverse electron-demand Diels-Alder reaction followed by a retro-Diels Alder extrusion of nitrogen gas to form a stable adduct. 5
- Figure 1.4.** Fusion of fluorescent proteins such as GFP or self-labeling enzymes such as SNAP-tag to a protein of interest enables fluorescence imaging of its localization in cells and whole organisms or rapid attachment of affinity handles. 7
- Figure 1.5.** An unnatural amino acid can be selectively incorporated into a protein of interest after introducing an amber stop codon (UAG) and an orthogonal tRNA/RNA synthetase pair. 8
- Figure 1.6.** Affinity-based probes enable labeling of proteins with binding pockets. 10
- Figure 1.7.** The tetracysteine motif reacts with FIAsh-EDT₂, a diarsenical probe that fluoresces upon binding to the tagged protein of interest. 12
- Figure 1.8.** Peptide tags can be irreversibly modified by small-molecule probes. 13
- Figure 1.9.** Differences between the gram-positive, the gram-negative, and the mycobacterial cell envelopes. 16
- Figure 1.10.** Many trehalose-based probes of the mycobacterial cell wall have been reported. 18
- Figure 1.11.** Examples of fluorogenic reactions. 19
- Figure 2.1.** Chemoselective reactions involving aminothiols. 31
- Figure 2.2.** Rational design of a peptide that covalently binds a CBT derivative. 33
- Figure 2.3.** Assay for peptides that are covalently modified by CBTNAc and representative results. 36

| | |
|--|-------|
| Figure 2.4. Results of the ¹³ CBTNAc binding assay and cysteine treatment for CBTag 1.0 when conducted in the presence of 5 mM glutathione (reduced). | 37 |
| Figure 2.5. Results of binding experiment and cysteine competition assays for CBTag 1.0 (K9R) and CBTag 1.0 (C7S). | 38 |
| Figure 2.6. Results of chloroacetate labeling of thiols under various conditions. | 40 |
| Figure 2.7. Tandem mass spectrometry (MS/MS) on CBTag 1.0-CBTNAc demonstrates that modification is on the lysine residue. | 41 |
| Figure 2.8. ¹³ C NMR confirms that the final CBTag 1.0-CBTNAc adduct is not the tetrahedral intermediate. | 42 |
| Figure 2.9. LC-MS-based time-course data for the reaction between CBTag 1.0 and CBTNAc. | 44 |
| Figure 2.10. ¹³ CBTNAc binding assay and cysteine treatment results for a 1:1 mixture of CBTag 1.0 (K9R) and CBTag 1.0 (C7S). | 45 |
| Figure 2.11. Proposed mechanism of reaction between CBTag 1.0 and CBTNAc. | 47 |
| Figure 2.12. Lowest energy conformation of each potential final product (Maestro, Macromodel). | 48 |
| Figure 2.S1. ¹³ CBTNAc binding assay and cysteine competition MALDI-TOF-MS data for the eighteen peptides synthesized by SPPS. | 65-68 |
| Figure 2.S2. Overlay of LC-MS chromatographs for 1:4 CBTag 1.0/ ¹³ CBTNAc over the time-course experiment and extracted mass ions. | 69 |
| Figure 2.S3. Cysteine treatment identifies peaks that contain thioimidate adducts. | 70 |
| Figure 2.S4. Extraction of relevant mass ions from the LC-MS spectrum of the reaction between CBTag 1.0 (K9R) and ¹³ CBTNAc shows that the peptide is only modified by one equivalent of ¹³ CBTNAc at the final time point of the assay. | 71 |
| Figure 2.S5. Cysteine treatment causes the minute 9.3 peak corresponding to the singly modified CBTag 1.0 (K9R) to disappear. | 72 |
| Figure 2.S6. The LC-MS spectrum and the corresponding extracted mass ions at the final time point show that CBTag 1.0 (C7S) is not modified by ¹³ CBTNAc. | 73 |

| | |
|---|-------|
| Figure 3.1. Previous approaches to utilizing CBT derivatives in protein labeling experiments. | 86 |
| Figure 3.2. MALDI-TOF MS shows that CBTag 1.0 only minimally reacts with a peptide bearing a C-terminal benzyl thioester. | 87 |
| Figure 3.3. MALDI-TOF MS shows that CBTag 1.0 is mono- and di-acetylated by <i>S</i> -(4-nitrophenyl)-thioacetate. | 88 |
| Figure 3.4. CBTag 1.0 can react with some CBT probes other than CBTNAc. | 89 |
| Figure 3.5. Sites within sfGFP where the CBTag 1.0 peptide was introduced are highlighted. | 90 |
| Figure 3.6. sGFP construct design and expression workflow. | 91 |
| Figure 3.7. MALDI-TOF MS spectra for each of the sfGFP constructs show that batch purification results in monomer and dimer of the constructs in different ratios. | 92 |
| Figure 3.8. MALDI-TOF MS data for reaction of a representative sfGFP construct with different CBT probes are inconclusive. | 93 |
| Figure 3.9. DIVERSE workflow. | 94 |
| Figure 3.10. MALDI-TOF MS analysis shows that none of the DIVERSE hit peptides formed more of a stable adduct than CBTag 1.0. | 95-96 |
| Figure 3.S1. Cysteine treatment identifies peaks that contain cysteine adducts. | 115 |
| Figure 3.S2. The LC-MS spectrum and the corresponding extracted mass ions at the final time point of cysteine treatment show that very little CBTag 1.0 forms a stable amide bond with the NCL peptide. | 116 |
| Figure 3.S3. The LC-MS spectrum and the corresponding extracted mass ions at the final time point show that CBTag 1.0 is mono- and di-acetylated by <i>S</i> -(4-nitrophenyl)thioacetate. | 117 |
| Figure 3.S4. The LC-MS spectrum and the corresponding extracted mass ions at the final time point show that CBTag 1.0 (K9R) is monoacetylated by <i>S</i> -(4-nitrophenyl)thioacetate. | 118 |
| Figure 3.S5. The LC-MS spectrum and the corresponding extracted mass ions at the final time point show that CBTag 1.0 (C7S) is not acetylated by <i>S</i> -(4-nitrophenyl)thioacetate. | 119 |

| | |
|---|-----|
| Figure 3.S6. Plasmid with the C-terminal His ₆ -tagged sfGFP. | 120 |
| Figure 4.1. An overview of the mycobacterial trehalose metabolic pathways and select drugs that inhibit them. | 129 |
| Figure 4.2. Solvatochromic trehalose probes label the mycobacterial mycomembrane. | 132 |
| Figure 4.3. Excitation and emission spectra of 3-hydroxychromone and Nile-red trehalose dyes. | 138 |
| Figure 4.4. Fluorescence analysis of Msmeg cells labeled with 3-hydroxychromone or Nile-red trehalose dyes using various excitation and emission filter sets. | 141 |
| Figure 4.5. Wash steps may be necessary to remove excess unincorporated dyes. | 143 |
| Figure 4.6. Solvatochromic trehalose dyes label Msmeg more strongly than solvatochromic glucose dyes, with the exception of MeNR-Tre. | 145 |
| Figure 4.7. Solvatochromic dyes use the same pathway as free trehalose. | 146 |
| Figure 4.8. Dye-Tre incorporation is dependent on metabolic activity. | 147 |
| Figure 4.9. DMN-Tre and EtNR-Tre compete for incorporation into Msmeg. | 148 |
| Figure 4.10. Dye-Tre incorporation does not primarily occur through cytosolic trehalose metabolic pathways. | 149 |
| Figure 4.11. Specific labeling of organisms in the Corynebacterineae suborder with solvatochromic trehalose conjugates. | 151 |
| Figure 4.12. Selective imaging of organisms in the Corynebacterineae suborder with solvatochromic trehalose conjugates. | 152 |
| Figure 4.13. Solvatochromic trehalose probes can report on drug susceptibility. | 154 |
| Figure 4.14. 3-HC-2-Tre, MeNR-Tre, and EtNR-Tre label <i>M. marinum</i> more brightly than DMN-Tre at 10x lower concentrations and can be used to visualize pre-labeled <i>M. marinum</i> in zebrafish. | 155 |

List of Schemes

| | |
|--|-----|
| Scheme 4.1. Synthesis of 6-Br-Ac7-Tre from trehalose. | 133 |
| Scheme 4.2. Synthesis of 3-hydroxychromones. | 134 |
| Scheme 4.3. Synthesis of 3-HC-2. | 134 |
| Scheme 4.4. Synthesis of 3-HC-3. | 134 |
| Scheme 4.5. Synthesis of 2-hydroxy Nile red analogs. | 135 |
| Scheme 4.6. Synthesis of the 6-dye-trehalose (dye-Tre) probes. | 135 |
| Scheme 4.7. Synthesis of methyl 6-iodo-2,3,4-tri- <i>O</i> -acetyl-6-deoxy-glucoside (Me 6-I-Ac3-Glc). | 136 |
| Scheme 4.8. Synthesis of the methyl 6-dye-glucoside (dye-Glc) control compounds. | 136 |

List of Tables

| | |
|--|---------|
| Table 2.S1. List of sequences modeled using PEP-FOLD. | 62-64 |
| Table 2.S2. Calculations for 1:1 CBTag 1.0/13CBTNAc LC-MS time-course assay (two replicates). | 74-76 |
| Table 2.S3. Calculations for 1:4 CBTag 1.0/13CBTNAc LC-MS time-course assay (two replicates). | 77-78 |
| Table 2.S4. Calculations for 4:1 CBTag 1.0/13CBTNAc LC-MS time-course assay (two replicates). | 79-80 |
| Table 3.1. Percentage of sfGFP in the active (fluorescent) form for each construct. ... | 91 |
| Table 3.2. Primer sequences used to insert CBTag 1.0 at different sfGFP locations. | 100 |
| Table 3.3. Sequences of primers used for DIVERSE selections. | 104 |
| Table 3.4. Full list of hit peptides determined from sequencing data. | 108 |
| Table 3.S1. Sequences of the sfGFP constructs. | 121-122 |
| Table 4.1. The EtNR-Tre wavelength of maximum absorbance shifts significantly as the ratio of dioxane to water changes. | 139 |
| Table 4.2. For each probe, the ideal excitation/emission filter set (commercially available, not necessarily common) and the acceptable excitation/emission filter set (common) is outlined. | 142 |

List of Abbreviations

| | |
|-------------|---|
| 13CBTNAc | <i>N</i> -(2-(cyano- ¹³ C)-benzo[<i>d</i>]thiazol-6-yl)acetamide |
| Ac | acetyl functional group |
| AFB | acid-fast bacteria |
| Ag85 | antigen 85 |
| a.i. | arbitrary intensity |
| Ala (A) | alanine |
| ANOVA | analysis of variance |
| Arg (R) | arginine |
| Asn (N) | asparagine |
| Asp (D) | aspartic acid |
| a.u. | arbitrary units |
| AU | absorbance units |
| azidoCBTNAc | <i>N</i> -(2-azidoacetyl)-6-amino-2-cyanobenzothiazole |
| Bs | <i>Bacillus subtilis</i> |
| CBT | cyanobenzothiazole |
| CBTNAc | <i>N</i> -(2-(cyanobenzo[<i>d</i>]thiazol-6-yl)acetamide |
| Cg | <i>Corynebacterium glutamicum</i> |
| COMU | (1-cyano-2-ethoxy-2-oxoethylideneaminoxy)dimethylamino-morpholino-carbenium hexafluorophosphate |
| CuAAC | copper(I)-catalyzed azide-alkyne cycloaddition |
| Cys (C) | cysteine |
| DBCO | dibenzocyclooctyne |
| DCM | dichloromethane (methylene chloride) |
| DIPEA | diisopropylethylamine |
| DIVERSE | Directed <i>in Vitro</i> Evolution of Reactive peptide tags via Sequential Enrichment |
| DMF | <i>N,N</i> -dimethylformamide |
| DMN | 4-(dimethylamino)-1,8-naphthalimide |
| DMSO | dimethylsulfoxide |
| DNA | deoxyribonucleic acid |
| DTT | dithiothreitol |
| Ec | <i>Escherichia coli</i> |
| EDC•HCl | 1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide hydrochloride |
| EDT | ethanedithiol |
| EDTA | ethylenediaminetetraacetic acid |
| ESI | electrospray ionization |
| Et | ethyl group |
| EtOAc | ethyl acetate |
| EtOH | ethanol |
| EtNR | 2-hydroxy Nile red, 9-(diethylamino)-2-hydroxy-5H-benzo[<i>a</i>]phenoxazin-5-one |
| FITC | fluorescein isothiocyanate |

| | |
|--------------|---|
| Fl | fluorescein |
| FIAsh | fluorescein arsenical hairpin binder |
| Fmoc-miniPEG | 9-fluorenylmethoxycarbonyl-8-amino-3,6-dioxaoctanoic acid |
| FPLC | fast protein liquid chromatography |
| GFP | green fluorescent protein |
| Glc | glucose |
| Gln (Q) | glutamine |
| Glu (E) | glutamic acid |
| Gly (G) | glycine |
| HATU | 1-[bis(dimethylamino)methylene]-1 <i>H</i> -1,2,3-triazolo[4,5- <i>b</i>]pyridinium 3-oxide hexafluorophosphate |
| HCTU | 2-(6-Chloro-1- <i>H</i> -benzotriazole-1-yl)-1,1,3,3-tetramethylamminium hexafluorophosphate |
| His (H) | histidine |
| HPLC | high-performance liquid chromatography |
| HRMS | high-resolution mass spectrometry |
| Ile (I) | isoleucine |
| INH | isoniazid |
| IPTG | Isopropyl β -D-1-thiogalactopyranoside |
| LB | lysogeny broth |
| LC-MS | liquid chromatography and mass spectrometry |
| Leu (L) | leucine |
| Lys (K) | lysine |
| mAU | milli-absorbance units |
| MALDI-TOF MS | matrix-assisted laser desorption/ionization-time of flight mass spectrometry |
| Me | methyl group |
| MeCN | acetonitrile |
| MeOH | methanol |
| MeNR | 9-(dimethylamino)-2-hydroxy-5 <i>H</i> -benzo[<i>a</i>]phenoxazin-5-one |
| Met (M) | methionine |
| MFI | mean fluorescence intensity |
| MS | mass spectrometry |
| MS/MS | tandem mass spectrometry |
| Msmeg | <i>Mycobacterium smegmatis</i> |
| Mtb | <i>Mycobacterium tuberculosis</i> |
| NADH | nicotinamide adenine dinucleotide |
| NaOMe | sodium methoxide |
| NBS | <i>N</i> -bromosuccinimide |
| NCL | native chemical ligation |
| NHS | <i>N</i> -hydroxysuccinimide |
| Ni-NTA | nickel-nitrilotriacetic acid |
| NMP | <i>N</i> -methylpyrrolidone |
| NMR | nuclear magnetic resonance |

| | |
|-------------------|---|
| OD ₆₀₀ | optical density at 600 nm |
| PBS | phosphate-buffered saline |
| PCR | polymerase chain reaction |
| PDB | Protein Data Bank |
| PEG | polyethylene glycol |
| PET | positron emission tomography |
| Phe (F) | phenylalanine |
| Pro (P) | proline |
| qPCR | quantitative polymerase chain reaction |
| RFP | red fluorescent protein |
| RNA | ribonucleic acid |
| RP-HPLC | reversed-phase high-performance liquid chromatography |
| Sa | <i>Staphylococcus aureus</i> |
| SDS | sodium dodecyl sulfate |
| SEM | standard error of the mean |
| Ser (S) | serine |
| sfGFP | superfolder green fluorescent protein |
| SOC | Super Optimal broth with Catabolite repression |
| SPAAC | strain-promoted azide-alkyne cycloaddition |
| SPPS | solid-phase peptide synthesis |
| TAMRA | carboxytetramethylrhodamine |
| TB | tuberculosis |
| TCEP | tris(2-carboxyethyl)phosphine |
| TDM | trehalose dimycolate |
| TFA | trifluoroacetic acid |
| THF | tetrahydrofuran |
| Thr (T) | threonine |
| TLC | thin layer chromatography |
| TMM | trehalose monomycolate |
| TMR | tetramethylrhodamine |
| Tre | trehalose |
| Trp (W) | tryptophan |
| Tyr (Y) | tyrosine |
| UPLC | ultra-performance liquid chromatography |
| UV | ultraviolet |
| Val (V) | valine |
| WT | wildtype |

Acknowledgments

As I sit here typing this Acknowledgments section two hours before the filing deadline, I'd like to thank all the people who got me to this point.

As is tradition, I want to start by saying how grateful I am to Carolyn for having me in her lab and letting me follow her to Stanford while remaining a Berkeley student. Her laissez-faire style of leadership has given me the opportunity to grow as an independent researcher, but she has always had my back when I needed it.

I would also like to thank the other members of my dissertation committee, Matt Francis, Susan Marqusee, and David Wemmer. Matt Francis has been extremely supportive and encouraging over the years, and I cannot express how much that has meant to me. As my original outside member changed institutions, I deeply appreciate Susan Marqusee's willingness to join my committee after only a brief introductory conversation at a Chemical Biology in the Bay Area Day three years ago. Finally, David Wemmer's agreement to serve as co-chair for the remaining UC Berkeley students in Carolyn's lab allowed me to remain a UC Berkeley student this past year even though Carolyn finally had to give up her academic senate appointment. While not a member of my committee, I'm also grateful to Dr. P. J. Utz for his support.

None of this would have been possible without the Bertozzi Lab administrative support team. Asia Avelino in particular was essential to daily life. Meridee Mannino, Sia Kruschke, Maryann Koopman, Cheryl McVaugh, Olga Martinez, and Karen Carkhuff were also responsible for everything running as smoothly as it does.

I would especially like to thank all of my friends and labmates in graduate school. The following are listed in no particular order:

My undergraduate Ashley Utz has easily been the best thing about moving to Stanford. Ashley, thank you for being so wonderful. I'd also like to thank the other two undergraduates who have worked with me, Aidan Pezacki and Teresia Chen; they've also been fantastic.

Peyton Shieh was my Bertozzi Lab rotation mentor, and his brilliance, good nature, and mentorship made joining the lab an easy decision. I also have him to thank for coming up with the original idea for using environment-sensitive molecules to selectively label mycobacteria.

Peter Robinson has been an excellent friend and labmate to have, and his sense of humor coupled with his gift for improv has kept my spirits up throughout grad school. I deeply admire how he came in one day, said "I think I want to start a company," and then made it happen.

David Spiciarich was one of the first people in the Bertozzi Group I got to know well and a constant source of such uplifting messages as “Every corpse on Mount Everest was once an extremely motivated person.” He has always been willing to lend a hand whenever I’ve needed it, and his encouragement has gotten me through the last stretch of writing this dissertation.

Mireille Kamariza has been an excellent friend and colleague. The work in Chapter 4 could not have been done without her biological expertise. Like David, Mireille has also been a constant source of encouragement and understanding. Words cannot express how much I appreciate her.

Ioannis Mountziaris has been a fantastic person to talk science and life with at all times of the day and night. Perhaps too much so, as we discovered when my undergraduate’s parents ended up calling campus security to check on her late one evening when she first started working for me and hadn’t been answering her phone.

I’ve always enjoyed my conversations with Douglas Fox, a brilliant scientist with a long and storied past. He is also to thank for helping to whip my first and last chapters into shape.

Frances Rodríguez-Rivera was also a fantastic mentor and friend, a brilliant scientist, and a fun person to have in my office. Room 817 forever! Speaking of, Matt Zhou was also a great member of the office, even if we did feel that a bell was necessary at times to stop him sneaking up on us.

I don’t always get to know new people in the lab well now that I’m a senior graduate student, but I’m glad that Steven Banik ended up in Frances’ chair. I’ve really enjoyed chatting with Steven about science, reminiscing about Harvard life, and listening to his Banikdoes (term coined by Corleone Delaveris, also in our office and great fun to hang out with, even if the baked goods he brings in consistently contain things I’m allergic to).

Mason Appel was always willing to teach me new biological techniques even though they were totally unrelated to his projects. I’m so grateful for his friendship over the years.

To my yearmates in the Bertozzi Group, Ian Blong, Jason Tsai, and Fred Tomlin, we may not be as close as other years, but I always felt like we were in things together. I couldn’t have asked for a better group to go through graduate school with.

To Green Ahn and Wilson Sinclair, thanks for letting me talk you into joining the lab and the tuberculosis project!

Thanks go to Gaby de Almeida and Julia McKechnie for laying the groundwork for the CBT project.

To all of the Bertozzians I haven't mentioned by name (I could go on all day, but I really need to upload this), thank you for making the lab such a great environment over the years.

To Alisha Contractor and Jackie Blake-Hedges, my two dearest Berkeley friends and enablers extraordinaire – thank you for agreeing that spontaneously adopting two kittens was a great idea. I might not have done it without you, and I definitely wouldn't have been at Petco that fateful weekend if we hadn't been meeting nearby for lunch. I'd also like to thank the Chemical Biology Program, which contained a wonderful group of people and was a wonderful experience.

To Paige Engelbrektsson, one of my best friends – hopefully now that this dissertation is in I can finally catch up with you. The same goes to my close college friends and roommates, Laura Fontanills, Eleanor Martin, Suniti Thapa, Abbie Modaff, and Belinda Vuong.

I'd also like to thank a few people in particular for setting me on the path to science: Dr. Brian Kennedy, who told me “you're applying to these summer research opportunities” my junior year of high school; Dr. Christopher Klug, who accepted me into his lab that summer; Professor J. Houston Miller, Professor Jörn Piel (and Dr. Agus Uria), and Dr. Jay Schneekloth, who hosted me in their labs during different summers in college; and Professor Jon Clardy (and Dr. Mo Seyedsayamdost) for letting me work in his lab as an undergraduate. I'd like to especially thank Jon for his support and understanding over the years. Everyone in the Harvard Chemistry Department (special mention to Dr. Gregg Tucci) and the UC Berkeley Department of Chemistry as well as the people I've interacted with at Stanford have been fantastic about working with me to get things done.

Finally, I'd like to thank my family. My grandfather and his work in chemistry have been an inspiration ever since I can remember, even if my research ended up being in a very different field. I also appreciate the steady interest and encouragement from my grandmother, aunts and uncles, and cousins. Last but foremost, I am deeply grateful to my parents (and my pets) for their love and support over the years. I would not be here now without them.

Chapter 1

Chemical biology approaches to selective labeling of cells and their components

Chapter 1. Chemical biology approaches to selective labeling of cells and their components

INTRODUCTION

The last two decades have seen a rapid rise in methods for modifying biological materials. The importance of controlling the location and multiplicity of such modifications is increasingly apparent in applications ranging from basic research to therapeutics.¹⁻¹⁰ Site-selective labeling techniques allow researchers to tag a biomolecule of interest in living cells and image its presence, localization, and interacting partners in real time while minimizing or eliminating perturbations introduced by the modification. This level of control enables pharmaceutical companies to create homogenous populations of antibody-drug conjugates, improving efficacy and off-target effects while allaying concerns of regulators.^{3,9-10} Selective labeling techniques can even be used to identify aberrant or heterologous cells and cell types, e.g., in point-of-care diagnostic platforms.¹¹

Many approaches to selective modification of cellular components exist, but they generally rely on one of three methods: genetic engineering, affinity-based labeling, and metabolic probe incorporation and/or activation. This chapter focuses on current approaches to site-selective protein modification and metabolic probes for fluorescence-based analyses of bacterial cell wall components. However, the efficacy of many of these methods depends on the complementary concept of bioorthogonal chemistry, and so our discussion will begin there.

BIOORTHOGONAL CHEMISTRY

Bioorthogonal chemistry is a relatively new term that refers to any researcher-orchestrated reaction that can take place inside a living system with minimal perturbation to native processes. Bioorthogonal reactions must proceed at a “reasonable” rate relative to the phenomena being studied, in aqueous conditions at physiological pH and temperature without being toxic to the host cell. Fulfilling all of these requirements is non-trivial, yet more than twenty unique transformations have been reported. Bioorthogonal reactions are especially useful for applications in which the labeling process cannot tolerate large or charged functional groups. In the first step, labeling can be accomplished using a probe functionalized with a small bioorthogonal reaction handle such as an azide, then the desired modification (e.g., a fluorophore, affinity tag) can be functionalized with the handle’s reaction partner and introduced in a bioorthogonal reaction. Below is an overview of commonly used bioorthogonal reagents.

Aldehydes/ketones and hydrazines/hydrazides/alkoxyamines

Aldehydes and ketones were some of the first functional groups used for conjugation of probes to biomolecules.¹²⁻¹⁵ When reacted with hydrazines, hydrazides, or alkoxyamines

under physiological conditions, they release a molecule of water to form hydrazones and oximes, respectively (Figure 1.1). However, because these reactions are reversible, the products are prone to some degree of hydrolysis.¹⁶ Recent developments have focused on creating more stable conjugates of the aldehyde or ketone targets. A ring-forming Pictet-Spengler ligation, in which an aminoxy-tryptamine¹⁷ or hydrazino-tryptamine¹⁸ reagent reacts with an aldehyde to produce a stable C-C bond, has made this class of reactions significantly more appealing to the pharmaceutical industry, where hydrolytic stability can determine the viability of biologics such as antibody-drug conjugates.¹⁹⁻²⁰

Aldehydes and ketones, although common in intracellular small molecule metabolites, are not typically present on cell surface macromolecules. Thus, they are functionally bioorthogonal under specific circumstances.

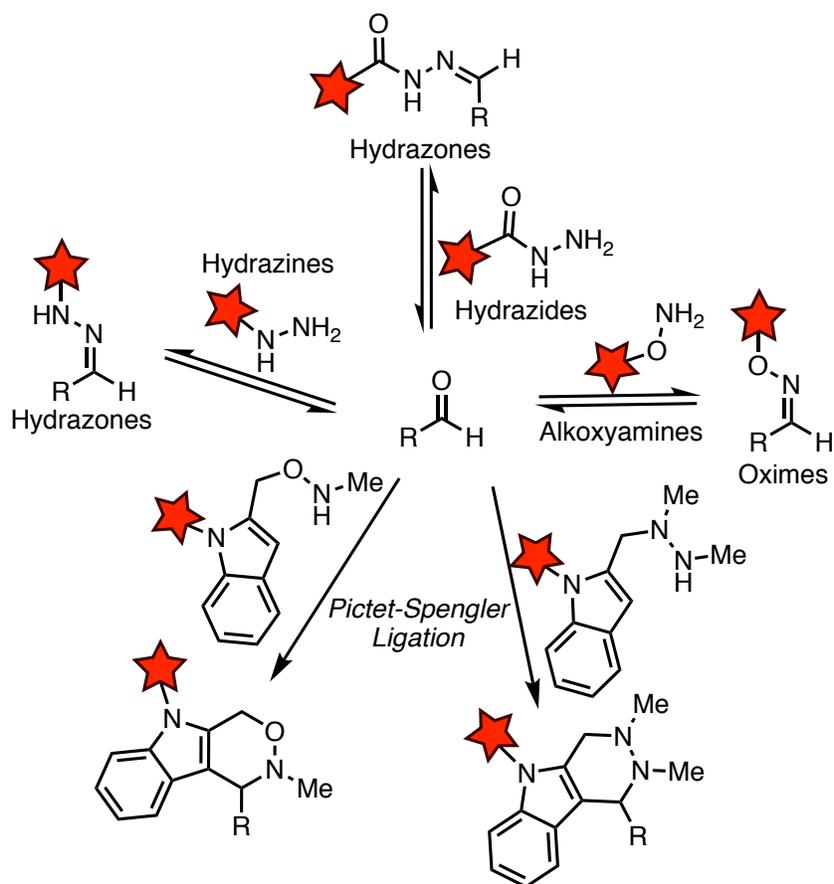


Figure 1.1. Aldehydes and ketones can be reacted with hydrazines, hydrazides, and alkoxyamines to attach probes to biomolecules. The R group represents a biomolecule, while the red stars represent probes such as fluorophores or affinity tags. The top three reactions are reversible under certain physiological conditions, but the bottom two variations on the Pictet-Spengler ligation are stable.

Azides and phosphines/alkynes/cycloalkynes

Azides are a popular bioorthogonal reaction handle due to their relatively small size, stability, and reaction versatility (Figure 1.2). With the exception of aryl azides, which can be rapidly reduced by thiols,²¹ azide reagents tend to be bioinert.²² When reacted with triarylphosphines, azides undergo a Staudinger ligation, releasing a molecule of nitrogen gas.²³⁻²⁴ Azides and alkynes in the presence of copper catalyst perform a 1,3-cycloaddition;²⁵⁻²⁶ with cycloalkynes, the reaction is termed a strain-promoted 1,3-cycloaddition.²⁷ Each of these reactions has advantages and disadvantages. The Staudinger ligation is highly biocompatible and the product is stable, but the reaction is slow and the phosphine reaction partner's selectivity is limited by its vulnerability to oxidation. The copper-catalyzed azide-alkyne cycloaddition (CuAAC) reaction is fast, but the copper catalyst can be cytotoxic due to its role in generation of reactive oxygen species.²⁸ However, the advent of new bulky ligands to stabilize the copper(I) ion and minimize perturbations to the native biological processes has allowed CuAAC reactions to be used inside cells.²⁹⁻³⁰ In cases where the alkyne partner is used in the metabolic incorporation step, the development of azides that are themselves copper chelators reduces both toxicity and the number of necessary components from three (azide, alkyne, copper reagent) to two (alkyne, azide-copper complex). Finally, strain-promoted azide-alkyne cycloaddition (SPAAC) removes the requirement for a copper catalyst and still proceeds at reasonable rates. Some cyclooctyne reagents, however, are susceptible to attack by thiols such as reduced cysteine residues.³¹

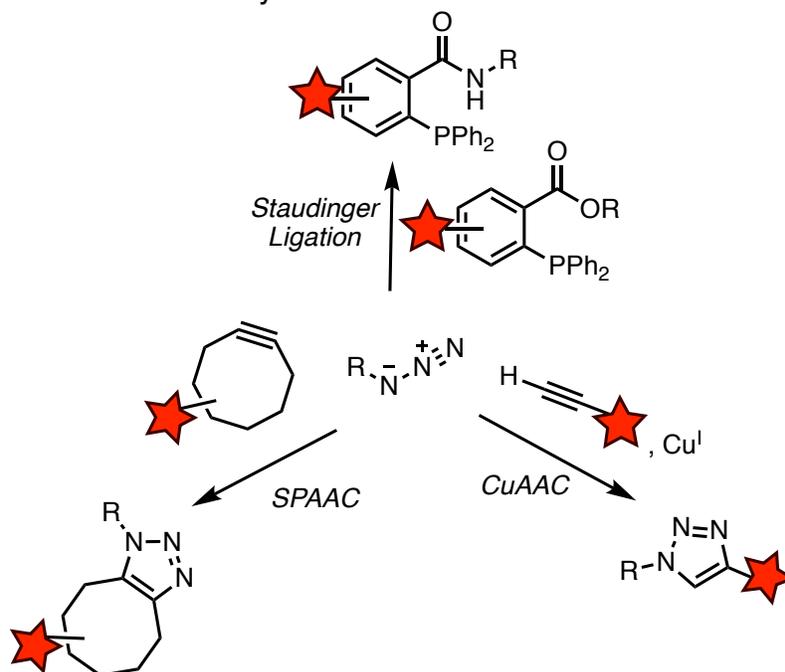


Figure 1.2. Azides can be reacted with triarylphosphines (the Staudinger ligation), terminal alkynes (copper-catalyzed azide-alkyne cycloaddition, CuAAC), and cyclooctynes (strain-promoted AAC, SPAAC) to introduce fluorophores or affinity tags (red stars) to a biomolecule (R group).

Cyclooctynes have also been reacted with nitrones (strain-promoted alkyne-nitrone cycloaddition, or SPANC), nitrile oxides, and diazo compounds with second-order rate constants comparable to or exceeding that of SPAAC reactions.³² However, nitrones, nitrile oxides, and diazo groups tend to be less hydrolytically stable, and so they often must be formed *in situ* close to the conjugation site.

Tetrazines and *trans*-cyclooctenes/norbornenes/cyclopropenes

The tetrazine ligation (TL) rapidly rose in popularity with the discovery that electron-deficient tetrazines and *trans*-cyclooctenes (TCOs) undergo an extremely rapid inverse electron-demand Diels-Alder reaction followed by an irreversible retro-Diels Alder elimination of nitrogen gas (Figure 1.3).³³ TL is the fastest bioorthogonal reaction to date with a second-order rate constant in 9:1 methanol/water of $2000 \text{ M}^{-1}\text{s}^{-1}$. The reaction between a TCO-functionalized protein and a tetrazine probe reaches 100% conversion in 5 minutes at reagent concentrations of $15 \text{ }\mu\text{M}$ and $30 \text{ }\mu\text{M}$, respectively.

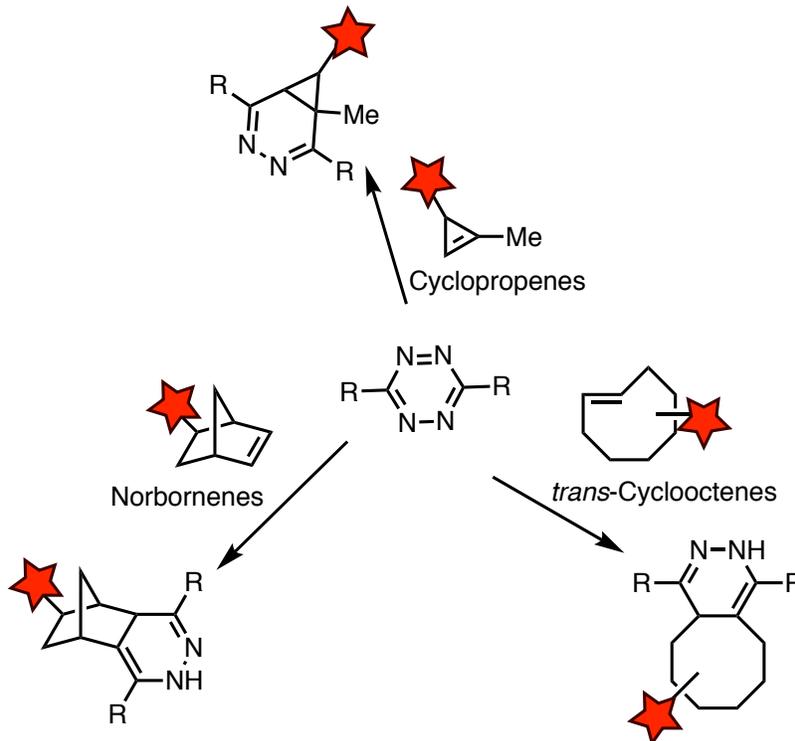


Figure 1.3. Tetrazines react rapidly with *trans*-cyclooctenes, cyclopropenes, and norbornenes via an inverse electron-demand Diels-Alder reaction followed by a retro-Diels Alder extrusion of nitrogen gas to form a stable adduct. Red stars represent affinity handles or fluorophores, R represents attachment to a biomolecule.

The major disadvantage of TCOs is that they can undergo isomerization over time to the unreactive *cis*-cyclooctenes. Alternatives to TCOs include cyclopropenes³⁴⁻³⁵ and norbornenes,³⁶ both of which are more stable in solution. In addition, like the azide group, the cyclopropene group is small relative to TCOs and norbornenes, making it an attractive

handle. However, the second-order rate constants for the reaction of tetrazines with these compounds are several orders of magnitude lower than that for reaction with TCOs.

Other transition metal-mediated reactions

The final class of bioorthogonal reactions that will be described in this section includes reactions that involve transition metals such as palladium and ruthenium. Olefin cross-metathesis, Sonogashira reactions, Mizoroki-Heck reactions, and Suzuki-Miyaura reactions have all been adapted to functionalize biomolecules, including cell-surface proteins.³⁷ These reactions tend not to be ideal for live-cell imaging, however, as the transition metal reagents and their reaction partners (e.g., boronic acids) can be toxic to the host. Moreover, when reacting a small molecule with a biomacromolecule (as opposed to with another small molecule) superstoichiometric amounts of transition metal “catalyst” are required.³⁸ This is likely due to non-specific coordination of the metal by endogenous functional groups such as thiols (e.g., reduced cysteines) and amines (e.g., lysines). This is a problem not only for *in vivo* and *in cellulo* applications but also for therapeutic processes. For example, the Food and Drug Administration (FDA) recommends that drugs delivered parenterally should expose patients to less than 10 $\mu\text{g/day}$ of palladium and ruthenium.³⁹ For context, the equivalent values for lead, arsenic, and copper are 5 $\mu\text{g/day}$ and 15 $\mu\text{g/day}$, and 340 $\mu\text{g/day}$ respectively. Thus, if a reaction mediated by (superstoichiometric amounts of) a palladium or ruthenium complex is used to conjugate a small molecule to an antibody, the resulting product would need to be very well-purified in order to be safe for use in humans.

The bioorthogonal reactions described above are essential components of many of the site-selective protein labeling and metabolic incorporation experiments described in the following sections.

METHODS FOR SELECTIVE PROTEIN LABELING

Protein fusion tags

The concept of modifying specific proteins to enable imaging in living cells began to gain traction in the 1990s, when Chalfie, Tsien, and others focused on fully characterizing green fluorescent protein (GFP).⁴⁰⁻⁴² Their work made possible not only the generation of genetically encodable fluorescent proteins covering the color spectrum but also the use of GFP as a tag for protein expression and localization (Figure 1.4A).⁴³ Since then, many researchers have genetically engineered proteins of interest so that they are fused with GFP and thus can be visualized by fluorescence microscopy in cells as well as whole organisms. More recently, researchers have designed the SNAP-tag (Figure 1.4B),⁴⁴⁻⁴⁵ CLIP-tag,⁴⁶ and HaloTag,⁴⁷ similar protein fusion-based labeling techniques that rely on the fusion protein’s ability to “self-label”, or covalently react with a substrate that has been

functionalized with a probe (e.g., a fluorophore, biotin). These tags naturally provide a greater range of labeling options, as they are not themselves fluorescent, and their substrates can readily be modified to meet the needs of the researcher. However, all of these protein fusion tags are relatively large: the molecular weight of GFP is 27 kDa, the molecular weight of HaloTag is 33 kDa, and the molecular weight of the SNAP- and CLIP-tags is 19.4 kDa. When the protein of interest is of a comparable or lesser size, the insertion of one of these protein tags can significantly impact native protein folding and dynamics. Thus, in the years since GFP proved its power as a technological platform, much effort has been dedicated to improving on it by decreasing the size of tag necessary for site-selective modification of proteins.

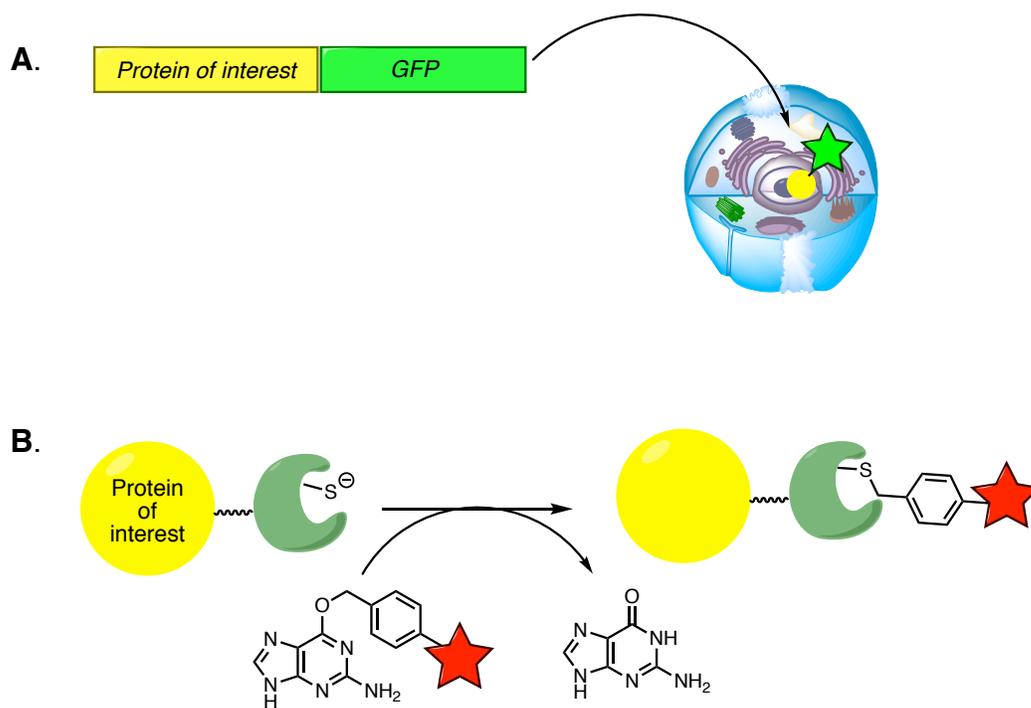


Figure 1.4. Fusion of fluorescent proteins such as GFP (**A**) or self-labeling enzymes such as SNAP-tag (**B**) to a protein of interest enables fluorescence imaging of its localization in cells and whole organisms or rapid attachment of affinity handles.

Amino acid-specific modifications

Long before the advent of protein fusion tags, scientists took advantage of unique side chain reactivity to label different amino acids. For example, maleimides will selectively react with cysteine thiols, and *N*-hydroxysuccinimidyl esters are used to selectively modify the terminal amine of lysine residues. But unless the protein of interest only contains one example of the target residue, it is difficult to predict which residues will be modified and what the final stoichiometry of small molecule modification to protein will be. Heterogeneous mixtures of products are also common. Several approaches to site-selective amino acid functionalization are described below.

Unnatural amino acid incorporation

The smallest possible tag that can be engineered into a protein of interest is an unnatural amino acid (UAA).⁴⁸⁻⁵¹ In order to incorporate these UAAs, the following conditions must be met: a codon must be available for repurposing, and an orthogonal tRNA/RNA synthetase (RS) pair that recognize only the desired UAA must be engineered into the system (Figure 1.5). The codon most often used for this purpose is the amber stop codon (UAG), which is the least used codon in *Escherichia coli*, making up less than 10% of its stop codons. Orthogonal tRNA/RNA synthetase pairs are generally engineered by choosing a tRNA/RS pair from a different organism than the target host and subjecting them to selection methods until they specifically recognize the desired UAA. The most commonly chosen sets are tRNA^{Tyr}-TyrRS and tRNA^{Amber}-PylRS, which incorporate tyrosines and pyrrolysines. Sets from different organisms have been implemented in *E. coli*, yeast, and mammalian cells to introduce unnatural amino acids with a bioorthogonal reaction handle or a built-in fluorescent moiety.⁵⁰⁻⁵¹

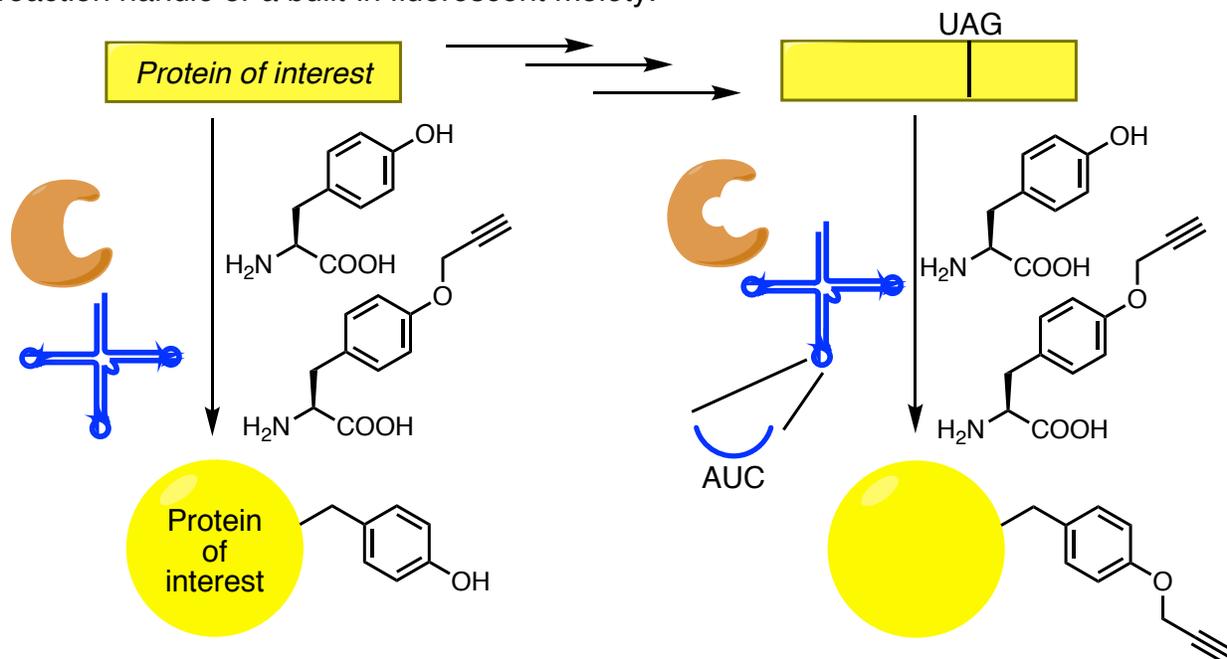


Figure 1.5. An unnatural amino acid can be selectively incorporated into a protein of interest after introducing an amber stop codon (UAG) and an orthogonal tRNA (blue)/RNA synthetase (brown) pair.

There are many disadvantages to relying on UAA incorporation, however. First, the engineering steps are non-trivial, especially in mouse, rat, and human cells, in which the amber stop codon accounts for 20-23% of all stop codons. Church and coworkers have demonstrated that it is possible to replace all amber stop codons in *E. coli* with other stop codons,⁵²⁻⁵³ but this is not (yet) a feasible approach in mammalian cells. When UAAs are successfully incorporated, efficiency is often low. Percent incorporation in the single-digits is not uncommon in mammalian systems that have not been extensively optimized, e.g.,

by varying promoter regions or engineering a release factor to increase translational read-through of the amber stop codon.⁵⁴

Chemo- and regioselective modification of native amino acid residues

The reactivity of an amino acid side chain can vary greatly depending on the surrounding microenvironment, and researchers have taken advantage of this fact in several ways. Most commonly, reagents are targeted at the N-terminal amine, cysteine residues with unusually low pKa values, and in a recent leap forward, unusually reactive lysine residues.⁵⁵ Site-selective modification of a single residue can be achieved, but usually the stoichiometry of the labeling reagent must be carefully controlled. Additionally, while these methods can be used to label the most reactive residue on a particular protein, they are not usually protein-specific, making them more applicable to *in vitro* labeling of purified biomacromolecules. The main advantage of these and the following methods is that they require no genetic engineering prior to use.

A more specific approach to modification of proteins is affinity-based labeling (Figure 1.6A).⁵⁶ This approach often relies on modifying protein's ligand to include a fluorophore, biotin moiety, or bioorthogonal chemical handle for later analysis as well as a group that can be used for covalent modification of the protein of interest. Many examples of these probes react irreversibly with an enzymatic active-site residue such as cysteine, making these compounds more suited for applications such as proteomics, in which post-labeling enzyme activity is not the focus of analysis.

Ligand-directed chemistry is a subset of this type of labeling reaction (Figure 1.6B).⁵⁷ These probes have the ligand attached to a leaving group (e.g., a tosyl group) such that attack by a nucleophilic residue external to the active site results in simultaneous modification by the desired probe and release of the ligand, freeing the protein's ligand-binding site for its normal activity. These probes are ideal for studying proteins that have a ligand-binding site and a ligand amenable to derivatization, but they often result in a heterogeneous mixture of products because they are not targeted at one particular nucleophilic residue.

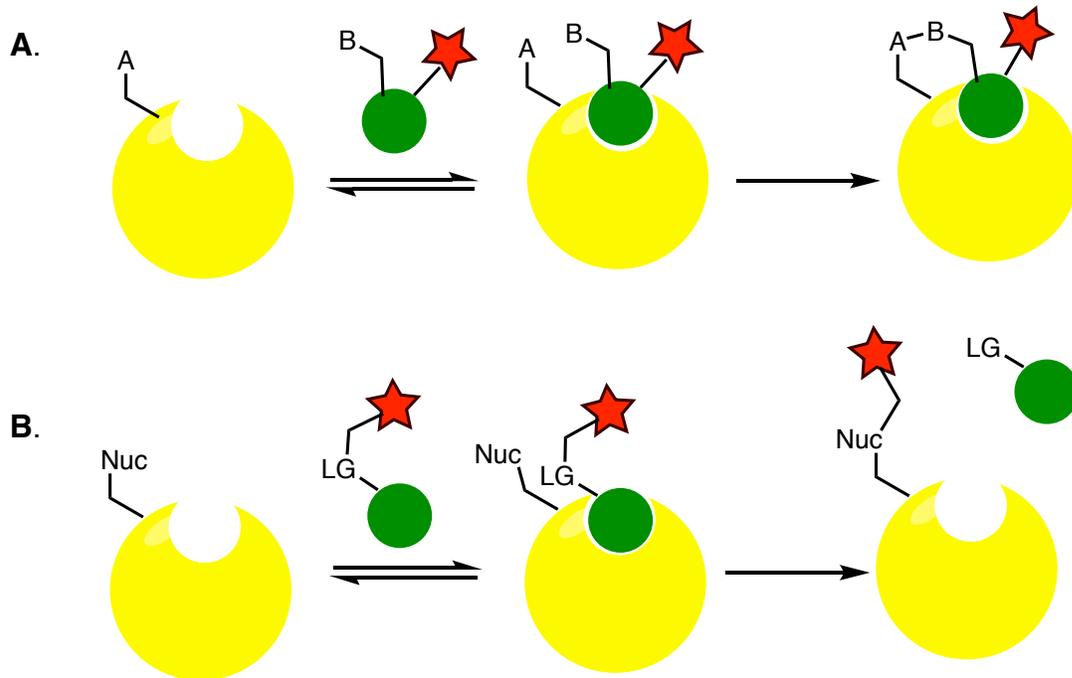


Figure 1.6. Affinity-based probes enable labeling of proteins with binding pockets. **A)** Traditional affinity-based systems consist of a reactive group (B) and a probe (red star) attached to the ligand (green circle) of the protein of interest (yellow). Covalent reaction between the reactive group (B) and a nearby residue (A) trap the ligand and thus the probe at the protein of interest. **B)** Ligand-directed probes consist of a ligand (green circle) and a label (red star) with a linker that contains a leaving group (LG). When the ligand binds the protein, a nearby nucleophilic residue forms a bond with the probe, displacing the ligand. The ligand is free to diffuse away, and the labeled protein's normal activity can resume.

Engineered peptide tags

Peptide tags consisting solely of canonical amino acids are a compromise between protein fusion tags and amino acid-specific modifications. They are small relative to protein fusion tags, and with the shortest reported peptide tag only four residues in length,⁵⁸ perturbation to the protein of interest can be minimal. Peptide tags require far fewer engineering steps than unnatural amino acid incorporation, and they can more reliably provide a unique microenvironment for site-selective labeling of a single type of protein in a cell than can most amino acid-specific modifications. Unlike affinity-based and ligand-directed probes, which require the protein of interest to have a ligand-binding site, peptide tags can be engineered into most any protein as long as their placement is chosen to maximize probe accessibility and minimize perturbations to the protein's native structure and dynamics. Finally, as with certain protein fusions and amino acid-specific modifications, the small molecule partner of the peptide tag can provide easy access to

a range of different modifications. Below, a few different classes of peptide tags are described.

Metal and small molecule-chelating peptide tags

Some of the best-known peptide tags chelate metals or small molecules. Polyhistidine (His₆) tags are perhaps the most widely used due to their well-established application in affinity purification of proteins.⁵⁹ Proteins with an N- or C-terminal His₆ tag are selectively retained on nickel-charged resins, allowing their separation from other proteins. As this interaction is reversible under certain conditions, the His₆-tagged protein can then be cleanly eluted from the column and used in further applications. Some researchers have even labeled His₆ tagged proteins in mammalian cells, though permanent conjugation required a special nickel ligand that could be photolytically activated for covalent modification of the target protein.⁶⁰ Off-target binding to proteins containing polyhistidine sequences can also become an issue in live-cell labeling.⁶¹

Another class of small molecule-binding peptide tags was introduced by Tsien and coworkers in 1998. In the original work,⁶² the authors described the rational design of a peptide sequence (Cys-Cys-Pro-Gly-Cys-Cys) containing four cysteine residues oriented such that they can rapidly react with FIAsh-EDT₂, a fluorescein-based diarsenical probe that becomes fluorescent on reaction with the peptide (Figure 1.7). The resulting adduct is stable despite its reliance on chelation of the arsenic groups rather than covalent modification. Despite the fact that their tag consists entirely of canonical amino acids, selective ligation is possible due to the reversibility of off-target interactions with naturally occurring Cys-Cys sequences. Unfortunately, minimizing background generally requires the addition of ethanedithiol (EDT), which is cytotoxic. However, the early success of this peptide tag led to variations in the dyes and sequence used for modification. ReAsH, a red-shifted resorufin-based version of FIAsh, also undergoes reaction with tetracysteine motifs.⁶³ RhoBo, a rhodamine-derived bisboronic acid, performs an analogous reaction with a tetraserine motif, though with more background due to the higher incidence of Ser-Ser-X-X-Ser-Ser sequences in native proteins.⁶⁴ Finally, because these probes are membrane-permeable, it has been possible for researchers to use these tetracysteine and tetraserine motifs to label proteins inside living cells.⁶⁵

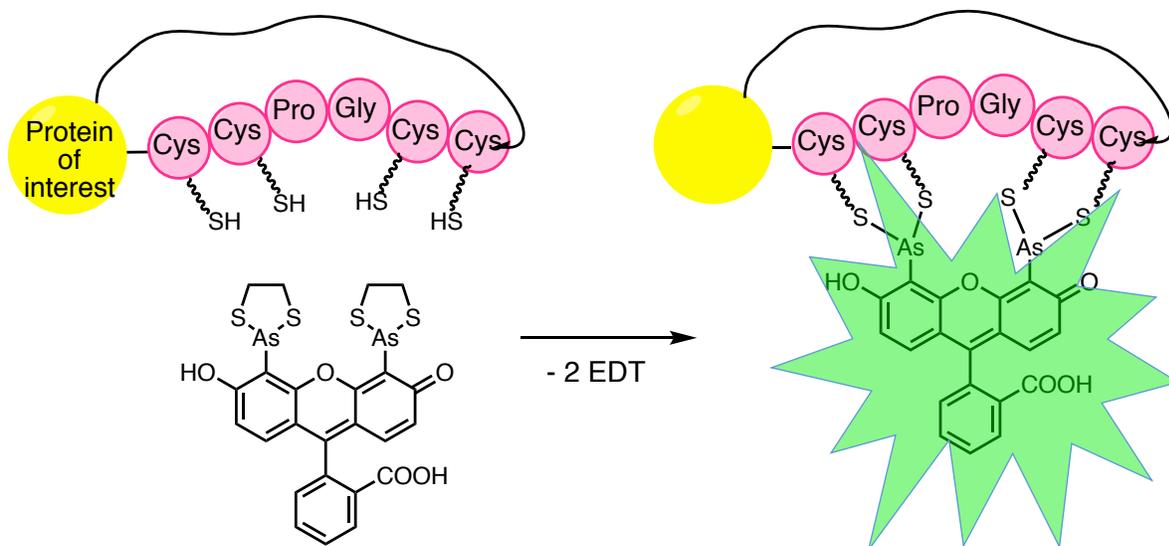


Figure 1.7. The tetracysteine motif reacts with FIAsh-EDT₂, a diarsenical probe that fluoresces upon binding to the tagged protein of interest.

Peptide tags that undergo irreversible covalent modification

Peptide tags that can be irreversibly ligated to chemical probes are of particular interest in the field of protein labeling. While reversible modifications have been cleverly utilized by researchers, stable modifications are essential for certain applications. For example, because antibody-drug conjugates often require the use of highly cytotoxic drugs to achieve therapeutic efficacy, the modification method must ensure that these toxic compounds remain attached to the antibody until they are delivered to the target cells.^{9-10,19-20}

Recently, the reaction of perfluoroaromatic probes with the ‘ π -clamp’ peptide tag was reported by Pentelute and coworkers (Figure 1.8A).⁵⁸ This tag consists of only four residues, Phe-Cys-Pro-Phe, earning it the distinction of being the shortest example of a small molecule-ligating peptide tag to date. While endogenous cysteines do not react with perfluoroaromatic probes in water,⁶⁶ the reactive form of this peptide sandwiches the probe between the two phenylalanine residues, positioning it for a nucleophilic aromatic substitution reaction with the cysteine residue. The authors demonstrated that the tag can be used to label the N- and C-termini of proteins, including antibodies. However, it has not been successfully inserted into the middle of a protein sequence, limiting its use in the study of proteins with inaccessible termini, such as certain transmembrane proteins. Various groups have reported longer peptide tags that ligate other compounds; one such method uses a cyanobenzothiazole derivative as the small molecule reaction partner and will be described in more detail in Chapters 2 and 3.⁶⁷

Enzyme-mediated labeling is a different approach to irreversible modification of peptide tags.⁶⁸ Enzymes such as biotin ligase (Figure 1.8B),⁶⁹ lipoic acid ligase,⁷⁰ sortase,⁷¹ and

transglutaminase⁷² have been used to introduce functional group handles to their substrate peptides. Subsequently, these handles can be reacted with functionalized fluorophores, biotin, drugs, and other compounds of interest. In a similar vein, formylglycine-generating enzyme (FGE) can be used to convert the cysteine residue in its consensus sequence to an aldehyde-containing formylglycine residue, which can then be reacted with any of the aldehyde-specific bioorthogonal reagents described earlier in this chapter.⁷³ Enzyme-mediated labeling allows rapid and site-selective labeling of proteins but when applied in living cells, it requires more engineering steps than simple metal chelating- or small molecule-ligating peptide tags: in addition to introducing the peptide tag itself, researchers must optimize enzyme expression to ensure adequate labeling of the target protein.

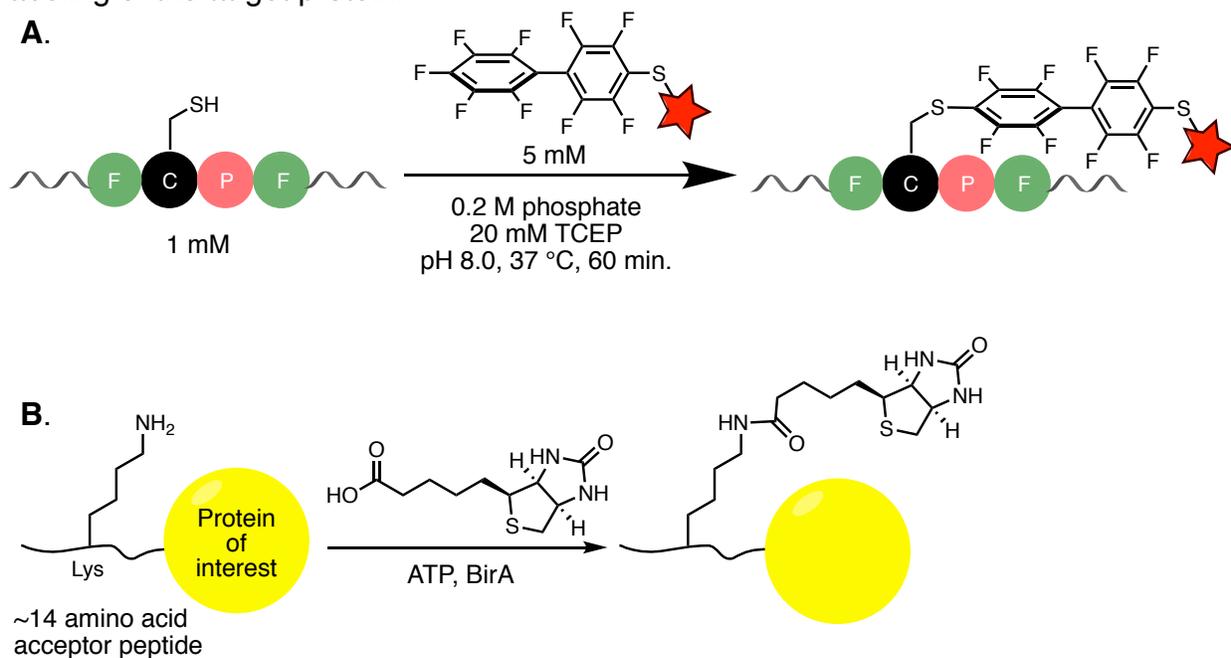


Figure 1.8. Peptide tags can be irreversibly modified by small-molecule probes. **A)** The shortest reported peptide tag has a cysteine residue that reacts with a perfluoroaromatic reagent to form a stable adduct. **B)** Biotin ligase (BirA) attaches biotin to a ~14 amino acid acceptor peptide in the presence of ATP.

Generating novel peptide tags

When researchers aim to develop new non-enzymatic peptide tag-based protein labeling methods, there are several approaches they may take. One approach is rational design, in which the researcher attempts to create logically a peptide with a secondary structure and microenvironment conducive to reaction with a desired small molecule. This approach has many drawbacks, most notably the fact that the success of rational design is directly correlated to how well the system in question has been characterized. A computation-guided approach to rational design of a small molecule-ligating peptide is presented in Chapter 2.

Another way to develop new peptide tags is via selection. Phage display allows researchers to introduce a library of peptide sequences into cell-surface proteins and select for the ones that react with the small molecule of interest.⁷⁴⁻⁷⁵ Phage that meet the desired criteria are used to produce more phage, and the selection is repeated with increasingly stringent conditions. After the final round of selection, the candidate peptides are identified by DNA sequencing. Yeast display, bacterial display, and mammalian display are similar *in vitro* selection methods. The major advantage of these methods is also in some sense the major disadvantage: the peptide is inserted into a native cell-surface protein. While this significantly accelerates the identification of peptide sequences that will react with the probe on insertion into a protein, established workflows often have built-in mechanisms to prevent denaturation of the protein used for display which can add additional unwanted selection conditions to the experiment. For example, phage display libraries might have cysteine residues placed at the beginning and end of the peptide insertion site to lock the peptide in a loop conformation via a disulfide bond, thus decreasing the potential for denaturation of the cell-surface protein.⁷⁴ For reactions that rely on the presence of reduced cysteines, this can be problematic. More generally, inserting the library of peptides into a cell-surface protein selects for peptides that are ideal for cell-surface labeling, not necessarily cytosolic protein labeling.

Alternatively, cell-free *in vitro* protein selection methods such as ribosome display and mRNA display can be used to select for appropriate small molecule-ligating peptides.⁷⁶ In these methods, a DNA library is transcribed into mRNA and translated, but instead of releasing the candidate peptide as would normally occur, the system is engineered such that the mRNA remains associated with the peptide. After selection for peptides that react with a small molecule of interest, the mRNA corresponding to the candidate probes can be reverse transcribed and the resulting DNA PCR-amplified for the next round or for cloning and sequencing. Decoupling the selection process from cells allows for much larger library sizes and avoids the introduction of unwanted selection bias. However, as the peptides do not have to be inserted into a protein, there is no guarantee that candidate sequences will still react with the small molecule when incorporated into the tertiary structure of a protein. In Chapter 3, the results of an mRNA display selection for a cyanobenzothiazole-ligating peptide is described.

Summary and outlook for site-selective protein labeling

Each of the protein-labeling methods described in this section has unique advantages and disadvantages, but researchers are still chasing a perfect *in vivo* labeling method that minimally perturbs the natural localization and function of the protein of interest, permits spatiotemporally controlled imaging, allows rapid and specific labeling of any cell-surface or cytosolic protein at not only the termini but also within the sequence, provides a wide range of potential modifications, and requires minimal engineering. While this dissertation does not claim to introduce such a method, Chapter 2 describes the rational design of a

peptide motif that ligates cyanobenzothiazoles via an internal cysteine-lysine transfer, while Chapter 3 highlights ongoing efforts to develop this motif into an effective protein-labeling method.

METABOLIC LABELING OF BACTERIAL CELLS

Although the field of protein labeling has seen incredible advances in the last few decades, the dynamics of cellular components such as lipids and glycans cannot easily be studied by techniques based on genetic manipulation or ligand specificity.⁷⁷ In these cases, researchers have turned instead to the concept of metabolic labeling.⁶ Metabolic probes are structurally similar to the native lipid or glycan, permitting them to be processed by endogenous cellular machinery, but also contain one or more uniquely reactive functional groups to distinguish them in later analyses. For example, compounds that are enriched for stable isotopes of certain atoms (e.g., ¹³C) can be used to trace the fate of a metabolite using Nuclear Magnetic Resonance (NMR) or mass spectrometry-based tools, whereas compounds that contain a radioactive isotope (e.g., ¹⁸F) can be followed using techniques such as positron emission tomography (PET). Isotope- or radioactivity-based compounds of this kind are “ideal” metabolic labeling probes because they are chemically distinguishable from natural substrates by analytic methods but not (for the most part) by the cellular machinery.

While metabolic labeling methods have been applied in cells and organisms of all complexities, including living humans, this section will focus on their use in fluorescence imaging of select bacterial cell wall components.

The bacterial cell envelope

Gram-positive, gram-negative, and acid-fast bacteria are classified as such due to differences in their cell envelopes (Figure 1.9).⁷⁸⁻⁸¹ Gram-positive bacteria have a cytoplasmic lipid membrane surrounded by a thick peptidoglycan layer consisting of long chains of alternating *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) sugar monomers cross-linked extensively via short chains of L- and D-amino acids. These two layers are separated by periplasmic space. Gram-positive bacteria also have teichoic acid-based components to these layers. In contrast, gram-negative bacteria consist of inner and outer membranes separated by a periplasm that contains a thin, only partially cross-linked peptidoglycan layer. They also present lipopolysaccharides on their cell surface. Acid-fast bacteria (AFB) describe mycobacteria and other subsets of the Actinobacteria phylum. Close to the cytoplasm, the mycobacterial cell envelope resembles that of gram-positive bacteria, containing a cytoplasmic lipid membrane, periplasmic space, and a thick peptidoglycan layer. The AFB cell wall diverges radically thereafter in its structure, with an arabinogalactan layer that is covalently linked to both

the peptidoglycan layer and a layer of long-chain mycolic acids. These mycolic acids form the inner leaflet of the unconventional outer membrane bilayer, also known as the mycomembrane. The outer leaflet is thought to consist of trehalose mycolates and phospholipids that non-covalently interact with the inner leaflet. Acid-fast bacteria also contain assorted lipoglycans such as lipoarabinomannan throughout their cell envelope. Gram-positive, gram-negative, and acid-fast bacteria can all have capsular and extracellular polysaccharides on the outermost part of their cell envelopes.

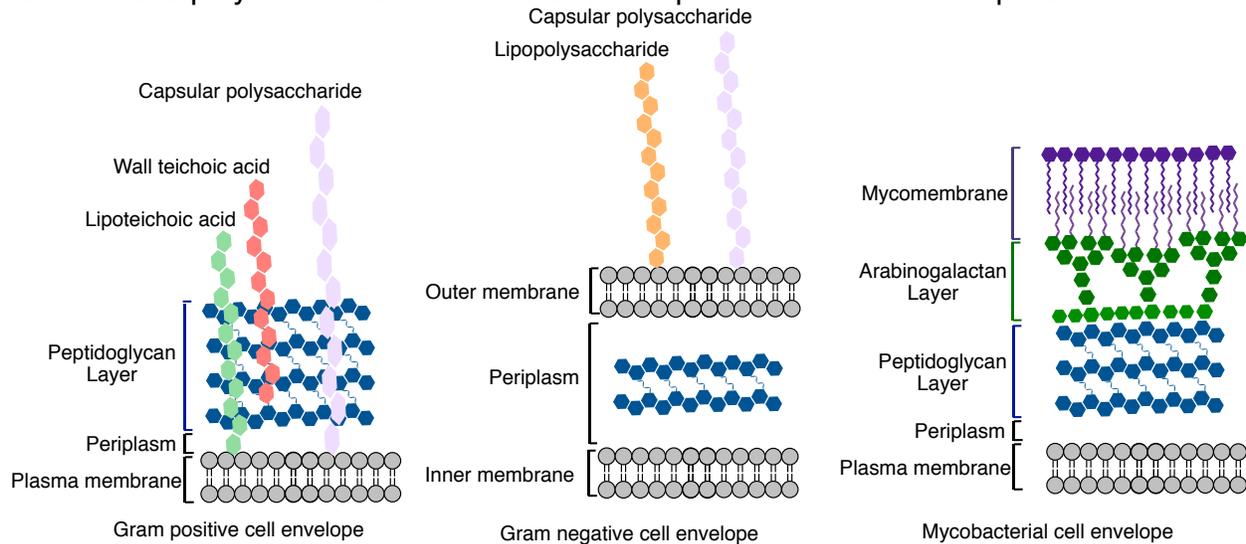


Figure 1.9. Differences between the gram-positive, the gram-negative, and the mycobacterial cell envelopes. Gram-positive bacteria have a thick peptidoglycan layer and teichoic acid-based biomolecules. Gram-negative bacteria have a thinner peptidoglycan layer, an outer membrane, and extracellular lipopolysaccharides. Mycobacteria have an arabinogalactan layer and a mycomembrane layer that serve as a pseudo-outer membrane.

Functionally, these unique aspects of each class of cell envelope defend the organism from outside stresses, including environmental insults, host defenses, and small molecule drugs. The Gram staining method relies on the thick peptidoglycan layer of the gram-positive bacteria to retain a dye that the thinner peptidoglycan layer of gram-negative bacteria cannot under the conditions of the test. Similarly, small molecules that can easily permeate one cell envelope may not be able to pass the barriers erected by other cell envelopes. Glycopeptide-based drugs such as vancomycin can act on gram-positive bacteria because they are not blocked by an outer membrane barrier. Additionally, antibiotics that kill one class of bacteria may be less effective against another class; aminoglycosides disproportionately affect gram-negative bacteria because they work by disrupting interactions between the lipopolysaccharides that stabilize the outer membrane, allowing other antibiotics and large molecules to penetrate the cell wall.⁸²

Metabolic probes of the bacterial cell envelope

Researchers have developed a range of metabolic probes to take advantage of the unique structures and biochemistries of bacterial cell envelopes (Figure 1.10). For example, D-amino acids are rare in metazoans but are ubiquitous in bacteria where D-alanine and D-glutamic acid are key components of peptidoglycan.⁸³ Notably, the D-Ala-D-Ala motif at the end of the short peptide sequences appended to GlcNAc-MurNAc chains is the binding site of vancomycin. It is also the recognition site for DD-transpeptidase, the enzyme that cross-links strands of the peptidoglycan layer and is the target of β -lactam antibiotics. Crucially, bacteria are able to incorporate exogenous D-amino acids. All of these qualities have made fluorescent D-alanine derivatives as well as azide- and alkyne-functionalized D-alanine derivatives powerful tools for studying peptidoglycan dynamics in living bacteria.⁸⁴⁻⁸⁹

Another class of probes relies on functionalization of carbohydrates. While vertebrates commonly use nine particular monosaccharides, bacteria use many more.⁹⁰ Additionally, monosaccharides and oligosaccharides can be metabolized in a species-dependent manner. For example, trehalose, which consists of two D-glucose monomers linked via an α,α -1,1 glycosidic bond, is used and produced by many organisms. However, only mycobacteria and certain other members of the Actinobacteria phylum are thought to incorporate lipidated trehalose into their cell walls.¹¹ Trehalose monomycolate (TMM) and trehalose dimycolate (TDM) are crucial components of the mycomembrane. Taking advantage of the bacteria's ability to scavenge extracellular trehalose as well as the high tolerance of essential biosynthetic machinery for large modifications at certain locations on trehalose, researchers have used probes ranging from azidotrehalose molecules and alkyne-functionalized TMM analogs to fluorophore- and fluorophore/quencher pair-based trehalose derivatives to study cell wall dynamics in mycobacteria and to selectively label pathogens such as *Mycobacterium tuberculosis*, the causative agent of tuberculosis.^{11, 91-}

97

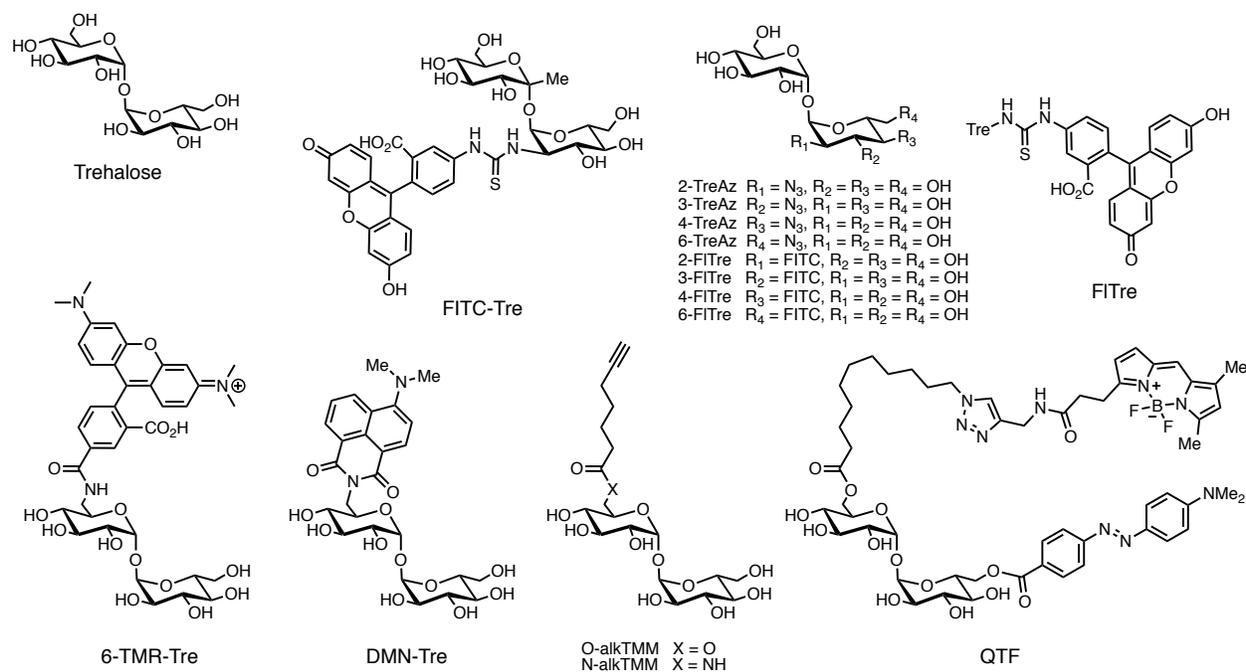
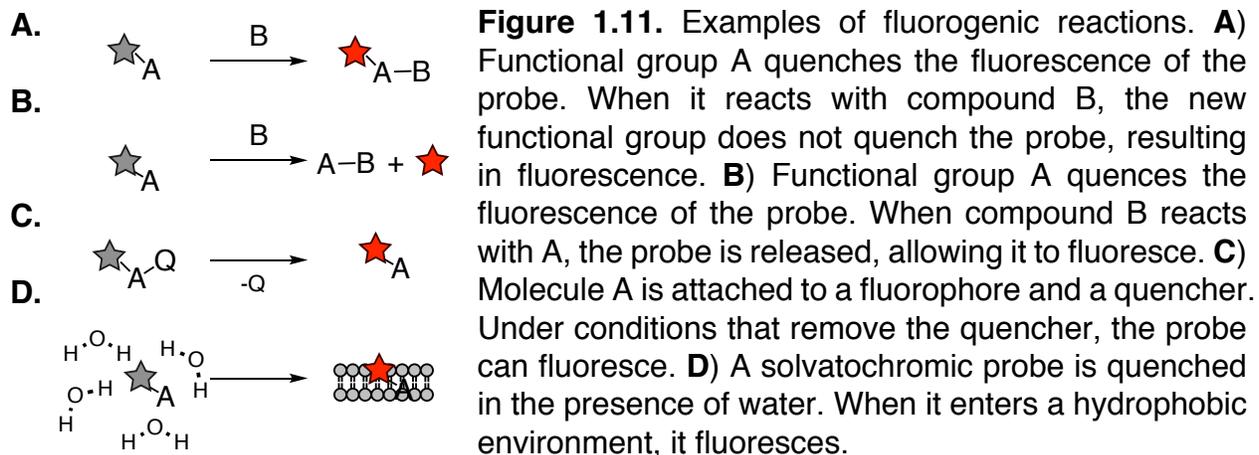


Figure 1.10. Many trehalose-based probes of the mycobacterial cell wall have been reported. The trehalose-processing biosynthetic machinery tolerates large functional groups at certain locations on the sugar scaffold.

Fluorogenic probes

While the methods that rely on installation of a bioorthogonal functional handle can be used with any applicable reaction partner, perhaps the most useful are the fluorogenic probes, also known as “turn-on” or “smart” probes.⁹⁸⁻⁹⁹ Samples incubated with standard fluorescent probes tend to suffer from high background unless thoroughly washed, which is not always possible, especially in *in vivo* experiments. In contrast, fluorogenic probes only become fluorescent under certain conditions which are present solely or predominantly in the desired context (Figure 1.11). Many examples of non-fluorescent bioorthogonal reagents that react to make a fluorescent product have been reported, and the state of this field has been extensively reviewed.⁹⁹

Finally, another type of turn-on probe is based on the concept of solvatochromism. Solvato(fluoro)chromic molecules change in color or emission intensity depending on their microenvironment. A subset of these molecules are fluorescent only in hydrophobic environments, meaning their background in aqueous solutions is much lower than that of standard fluorophores. This characteristic makes them highly attractive for *in vitro*, *ex vivo*, and *in vivo* experiments.¹⁰⁰⁻¹⁰² A use of metabolic probes based on a solvatochromic scaffold is presented in Chapter 4.



Summary and outlook for metabolic probes of bacterial cell envelope components

The demand for new metabolic probes is increasing as methods for *in vivo* imaging become more powerful and refined. Probes that target components of the bacterial cell envelope can allow researchers to study native cell envelope dynamics both *in vitro* and during the course of cell-cell interactions (e.g., host-pathogen interactions). Insights from these experiments will be essential to the development of new antibiotics, which are a critical need in this age of rising bacterial drug resistance.

REFERENCES

1. Jing, C.; Cornish, V. W. Chemical Tags for Labeling Proteins Inside Living Cells. *Acc. Chem. Res.* **2011**, *44*, 784-792 and references cited therein.
2. Spicer, C. D.; Davis, B. G. Selective chemical protein modification. *Nat. Commun.* **2014**, *5*:4740 and references cited therein.
3. Cal, P. M. S. D.; Bernardes, G. J. L.; Gois, P. M. P. Cysteine-Selective Reactions for Antibody Conjugation. *Angew. Chem. Int. Ed.* **2014**, *53*, 10585-10587 and references cited therein.
4. Stephanopoulos, N.; Francis, M. B. Choosing an effective protein bioconjugation strategy. *Nat. Chem. Biol.* **2011**, *7*, 876-884 and references cited therein.
5. Lotze, J.; Reinhardt, U.; Seitz, O.; Beck-Sickinger, A. G. Peptide-tags for site-specific protein labelling *in vitro* and *in vivo*. *Mol. BioSyst.* **2016**, *12*, 1731-1745 and references cited therein.
6. Xie, R.; Hong, S.; Chen, X. Cell-selective metabolic labeling of biomolecules with bioorthogonal functionalities. *Curr. Opin. Chem. Biol.* **2013**, *17*, 747-752 and references cited therein.
7. Krall, N.; da Cruz, F. P.; Boutureira, O.; Bernardes, G. J. L. Site-selective protein-modification chemistry for basic biology and drug development. *Nat. Chem.* **2016**, *8*, 103-113 and references cited therein.
8. McKay, C. S.; Finn, M. G. Click Chemistry in Complex Mixtures: Bioorthogonal Bioconjugation. *Chem. Biol.* **2014**, *21*, 1075-1101.
9. Strop, P.; Delaria, K.; Foletti, D.; Witt, J. M.; Hasa-Moreno, A.; Poulsen, K.; Casas, M. G.; Dorywalska, M.; Farias, S.; Pios, A.; Lui, V.; Dushin, R.; Zhou, D.; Navaratnam, T.; Tran, T.-T.; Sutton, J.; Lindquist, K. C.; Han, B.; Liu, S.-H.; Shelton, D. L.; Pons, J.; Rajpal, A. Site-specific conjugation improves therapeutic index of antibody drug conjugates with high drug loading. *Nat. Biotechnol.* **2015**, *33*, 694-696.
10. Sun, X.*; Ponte, J. F.*; Yoder, N. C.; Laleau, R.; Coccia, J.; Lanieri, L.; Qiu, Q.; Wu, R.; Hong, E.; Bogalhas, M.; Wang, L.; Dong, L.; Setiady, Y.; Maloney, E. K.; Ab, O.; Zhang, X.; Pinkas, J.; Keating, T. A.; Chari, R.; Erickson, H. K.; Lambert, J. M. Effects of Drug-Antibody Ratio on Pharmacokinetics, Biodistribution, Efficacy, and Tolerability of Antibody-Maytansinoid Conjugates. *Bioconjugate Chem.* **2017**, *28*, 1371-1381.
11. Kamariza, M.*; Shieh, P.*; Ealand, C. S.; Peters, J. S.; Chu, B.; Rodríguez-Rivera, F. P.; Babu Sait, M. R.; Treuren, W. V.; Martinson, N.; Kalscheuer, R.; Kana, B. D.; Bertozzi, C. R. Rapid detection of *Mycobacterium tuberculosis* in sputum with a solvatochromic trehalose probe. *Sci. Transl. Med.* **2018**, *10*, eaam6310.
12. Van Lenten, L.; Ashwell, G. Studies on the Chemical and Enzymatic Modification of Glycoproteins. *J. Biol. Chem.* **1971**, *246*, 1889-1894.

13. O'Shannessy, D. J.; Dobersen, M. J.; Quarles, R. H. A novel procedure for labeling immunoglobulins by conjugation to oligosaccharide moieties. *Immunol. Lett.* **1984**, *8*, 273-277.
14. Cornish, V. W.; Hahn, K. M.; Schultz, P. G. Site-specific protein modification using a ketone handle. *J. Am. Chem. Soc.* **1996**, *118*, 8150–8151.
15. Shao, J.; Tam, J. P. Unprotected Peptides as Building Blocks for the Synthesis of Peptide Dendrimers with Oxime, Hydrazone, and Thiazolidine Linkages. *J. Am. Chem. Soc.* **1995**, *117*, 3893-3899.
16. Jencks, W. P. Studies on the Mechanism of Oxime and Semicarbazone Formation. *J. Am. Chem. Soc.* **1959**, *81*, 475-481.
17. Agarwal, P.; van der Weijden, J.; Sletten, E. M.; Rabuka, D.; Bertozzi, C. R. A Pictet-Spengler ligation for protein chemical modification. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 46–51.
18. Agarwal, P.; Kudirka, R.; Albers, A. E.; Barfield, R. M.; de Hart, G. W.; Drake, P. M.; Jones, L. C.; and Rabuka, D. Hydrazino-Pictet-Spengler ligation as a biocompatible method for the generation of stable protein conjugates. *Bioconjugate Chem.* **2013**, *24*, 846–851.
19. Wu, A. M.; Senter, P. D. Arming antibodies: Prospects and challenges for immunoconjugates. *Nat. Biotechnol.* **2005**, *23*, 1137-1146.
20. Tsuchikama, K.; An, Z. Antibody-drug conjugates: recent advances in conjugation and linker chemistries. *Protein Cell* **2018**, *9*, 33-46.
21. Staros, J. V.; Bayley, H.; Standring, D. N.; Knowles, J. R. Reduction of aryl azides by thiols: Implications for the use of photoaffinity reagents. *Biochem. Biophys. Res. Commun.* **1978**, *80*, 568-572.
22. Griffin, R. J. The medicinal chemistry of the azido group. *Prog. Med. Chem.* **1994**, *31*, 121-232.
23. Saxon, E.; Bertozzi, C. R. Cell surface engineering by a modified Staudinger reaction. *Science* **2000**, *287*, 2007–2010.
24. Lin, F. L.; Hoyt, H. M.; van Halbeek, H.; Bergman, R. G.; Bertozzi, C. R. Mechanistic Investigation of the Staudinger Ligation. *J. Am. Chem. Soc.* **2005**, *127*, 2686-2695.
25. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective "Ligation" of Azides and Terminal Alkynes. *Angew. Chem. Int. Ed.* **2002**, *41*, 2596-2599.
26. Wang, Q.; Chan, T. R.; Hilgraf, R.; Fokin, V. V.; Sharpless, K. B.; Finn, M. G. Bioconjugation by Copper(I)-Catalyzed Azide-Alkyne [3 + 2] Cycloaddition. *J. Am. Chem. Soc.* **2003**, *125*, 3192-3193.

27. Baskin, J. M.; Prescher, J. A.; Laughlin, S. T.; Agard, N. J.; Chang, P. V.; Miller, I. A.; Lo, A.; Codelli, J. A.; Bertozzi, C. R. Copper-free click chemistry for dynamic *in vivo* imaging. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 16793-16797.
28. Wolbers, F.; ter Braak, P.; Le Gac, S.; Lutge, R.; Andersson, H.; Vermes, I.; van den Berg, A. Viability study of HL60 cells in contact with commonly used microchip materials. *Electrophoresis* **2006**, *27*, 5073.
29. Uttamapinant, C.; Tangpeerachaikul, A.; Grecian, S.; Clarke, S.; Singh, U.; Slade, P.; Gee, K. R.; Ting, A. Y. Fast, Cell-Compatible Click Chemistry with Copper-Chelating Azides for Biomolecular Labeling. *Angew. Chem. Int. Ed.* **2012**, *51*, 5852-5856.
30. Su, Y.; Li, L.; Wang, H.; Wang, X.; Zhang, Z. All-in-One azides: empowered click reaction for *in vivo* labeling and imaging of biomolecules. *Chem. Commun.* **2016**, *52*, 2185-2188.
31. van Geel, R.; Pruijn, G. J. M.; van Delft, F. L.; Boelens, W. C. Preventing Thiol-Yne Addition Improves the Specificity of Strain-Promoted Azide-Alkyne Cycloaddition. *Bioconjugate Chem.* **2012**, *23*, 392-398.
32. Patterson, D. M.; Nazarova, L. A.; Prescher, J. A. Finding the Right (Bioorthogonal) Chemistry. *ACS Chem. Biol.* **2014**, *9*, 592-605 and references cited therein.
33. Blackman, M. L.; Royzen, M.; and Fox, J. M. Tetrazine ligation: Fast bioconjugation based on inverse-electron-demand Diels-Alder reactivity. *J. Am. Chem. Soc.* **2008**, *130*, 13518-13519.
34. Yang, J.; Seckute, J.; Cole, C. M.; Devaraj, N. K. Live-cell imaging of cyclopropene tags with fluorogenic tetrazine cycloadditions. *Angew. Chem. Int. Ed.* **2012**, *51*, 7476-7479.
35. Patterson, D. M.; Nazarova, L. A.; Xie, B.; Kamber, D. N.; Prescher, J. A. Functionalized cyclopropenes as bioorthogonal chemical reporters. *J. Am. Chem. Soc.* **2012**, *134*, 18638-18643.
36. Devaraj, N. K.; Weissleder, R.; Hilderbrand, S. A. Tetrazine-Based Cycloadditions: Application to Pretargeted Live Cell Imaging. *Bioconjugate Chem.* **2008**, *19*, 2297-2299.
37. Yang, M.; Li, J.; Chen, P. R. Transition metal-mediated bioorthogonal protein chemistry in living cells. *Chem. Soc. Rev.* **2014**, *43*, 6511-6526 and references cited therein.
38. Vinogradova, E. V.; Zhang, C.; Spokoiny, A. M.; Pentelute, B. L.; Buchwald, S. L. Organometallic palladium reagents for cysteine bioconjugation. *Nat. Chem.* **2015**, *526*, 687-691.
39. *Q3D Elemental Impurities: Guidance for Industry*; U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Biologics Evaluation and Research (CBER); 2015.

<https://www.fda.gov/downloads/drugs/guidances/ucm371025.pdf> (accessed May 8, 2018).

40. Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C. Green fluorescent protein as a marker for gene expression. *Science* **1994**, *263*, 802-805.
41. Heim, R.; Cubitt, A. B.; Tsien, R. Y. Improved Green Fluorescence. *Nature* **1995**, *373*, 663-664.
42. Ward, W. W. Biochemical and Physical Properties of Green Fluorescent Protein. In *Green Fluorescent Protein: Properties, Applications, and Protocols*; Chalfie, M.; Kain, S., Eds.; Wiley-Liss, Inc.: New York, 1998, Ch. 3, pp. 45-75.
43. Chudakov, D. M.; Matz, V. M.; Lukyanov, S.; Lukyanov, K. A. Fluorescent Proteins and their applications in imaging living cells and tissues. *Physiol. Rev.* **2010**, *90*, 1103-1163.
44. Juillerat, A.; Gronemeyer, T.; Keppler, A.; Gendreizig, S.; Pick, H.; Vogel, H.; Johnsson, K. Directed Evolution of O⁶-Alkylguanine-DNA Alkyltransferase for Efficient Labeling of Fusion Proteins with Small Molecules in Vivo. *Cell Chem. Biol.* **2003**, *10*, 313-317.
45. Mollwitz, B.; Brunk, E.; Schmitt, S.; Pojer, F.; Bannwarth, M.; Schiltz, M.; Rothlisberger, U.; Johnsson, K. Directed Evolution of the Suicide Protein O⁶-Alkylguanine-DNA Alkyltransferase for Increased Reactivity Results in an Alkylated Protein with Exceptional Stability. *Biochemistry* **2012**, *51*, 986–994.
46. Gautier, A.; Juillerat, A.; Heinis, C.; Corrêa, I. R. Jr.; Kindermann, M.; Beaufils, F.; Johnsson, K. An engineered protein tag for multiprotein labeling in living cells. *Chemistry and Biology* **2008**, *15*, 128–136.
47. Los, G. V.; Encell, L. P.; McDougal, M. G.; Hartzell, D. D.; Karassina, N.; Zimprich, C.; Wood, M. G.; Learish, R.; Ohana, R. F.; Urh, M.; Simpson, D.; Mendez, J.; Zimmerman, K.; Otto, P.; Vidugiris, G.; Zhu, J.; Darzins, A.; Klaubert, D. H.; Bülleit, R. F.; Wood, K. V. HaloTag: a novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol.* **2008**, *3*, 373-382.
48. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. Expanding the Genetic Code of *Escherichia coli*. *Science* **2001**, *292*, 498-500.
49. Chin, J. W. Expanding and Reprogramming the Genetic Code of Cells and Animals. *Annu. Rev. Biochem.* **2014**, *83*, 379-408 and references cited therein.
50. Davis, L.; Chin, J. W. Designer proteins: applications of genetic code expansion in cell biology. *Nat. Rev. Mol. Cell Biol.* **2012**, *13*, 168–182 and references cited therein.
51. Liu, C. C.; Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **2010**, *79*, 413–444 and references cited therein.
52. Isaacs, F. J.; Carr, P. A.; Wang, H. H.; Lajoie, M. J.; Sterling, B.; Kraal, L.; Tolonen, A. C.; Gianoulis, T. A.; Goodman, D. B.; Reppas, N. B.; Emig, C. J.; Bang, D.; Hwang, S. J.; Jewett, M. C.; Jacobson, J. M.; Church, G. M. Precise Manipulation of

- Chromosomes in Vivo Enables Genome-Wide Codon Replacement. *Science* **2011**, *333*, 348-353.
53. Lajoie, M. J.; Rovner, A. J.; Goodman, D. B.; Aerni, H.-R.; Haimovich, A. D.; Kuznetsov, G.; Mercer, J. A.; Wang, H. H.; Carr, P. A.; Mosberg, J. A.; Rohland, N.; Schultz, P. G.; Jacobson, J. M.; Rinehart, J.; Church, G. M.; Isaacs, F. J. Genomically Recoded Organisms Expand Biological Functions. *Science* **2013**, *342*, 357-360.
54. Schmied, W. H.; Elsässer, S. J.; Uttamapinant, C.; Chin, J. W. Efficient Multisite Unnatural Amino Acid Incorporation in Mammalian Cells via Optimized Pyrrolysyl tRNA Synthetase/tRNA Expression and Engineered eRF1. *J. Am. Chem. Soc.* **2014**, *136*, 15577-15583.
55. Matos, M. J.; Oliveira, B. L.; Martínez-Sáez, N.; Guerreiro, A.; Cal, P. M. S. D.; Bertoldo, J.; Maneiro, M.; Perkins, E.; Howard, J.; Deery, M. J.; Chalker, J. M.; Corzana, F.; Jiménez-Osés, G.; Bernardes, G. J. L. Chemo- and Regioselective Lysine Modification on Native Proteins. *J. Am. Chem. Soc.* **2018**, *140*, 4004-4017.
56. Ranjitkar, P.; Perera, G. K.; Swaney, D.; Hari, S. B.; Larson, E. T.; Krishnamurty, R.; Merritt, E. A.; Villén, J.; Maly, D. J. Affinity-Based Probes Based on Type II Kinase Inhibitors. *J. Am. Chem. Soc.* **2012**, *134*, 19017-19025.
57. Tsukiji, S.; Miyagawa, M.; Takaoka, Y.; Tamura, T.; Hamachi, I. Ligand-directed tosyl chemistry for protein labeling *in vivo*. *Nat. Chem. Biol.* **2009**, *5*, 341-343.
58. Zhang, C.; Welborn, M.; Zhu, T.; Yang, N. J.; Santos, M. S.; Van Voorhis, T.; Pentelute, B. L. π -Clamp-mediated cysteine conjugation. *Nat. Chem.* **2015**, *8*, 120-128.
59. Hochuli, E.; Bannwarth, W.; Döbeli, H.; Gentz, R.; Stüber, D. Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent. *Bio/Technology* **1988**, *6*, 1321-1325.
60. Lai, Y.-T.; Chang, Y.-Y.; Hu, L.; Yang, Y.; Chao, A.; Du, Z.-Y.; Tanner, J. A.; Chye, M.-L.; Qian, C.; Ng, K.-M.; Li, H.; Sun, H. Rapid labeling of intracellular His-tagged proteins in living cells. *Proc. Natl. Acad. Sci.* **2015**, *112*, 2948-2953.
61. Andersen, K. R.; Leksa, N. C.; Schwartz, T. U. Optimized *E. coli* expression strain LOBSTR eliminates common contaminants from His-tag purification. *Proteins* **2013**, *81*, 1857-1861.
62. Griffin, B. A.; Adams, S. R.; Tsien, R. Y. Specific Covalent Labeling of Recombinant Protein Molecules Inside Live Cells. *Science* **1998**, *281*, 269-272.
63. Adams, S. R.; Campbell, R. E.; Gross, L. A.; Martin, B. R.; Walkup, G. K.; Yao, Y.; Llopis, J.; Tsien, R. Y. New Biarsenical Ligands and Tetracysteine Motifs for Protein Labeling in Vitro and in Vivo: Synthesis and Biological Applications. *J. Am. Chem. Soc.* **2002**, *24*, 6063-6076.
64. Halo, T. L.; Appelbaum, J.; Hobert, E. M.; Balkin, D. M.; Schepartz, A. Selective Recognition of Protein Tetraserine Motifs with a Cell-Permeable, Pro-fluorescent Bis-boronic Acid. *J. Am. Chem. Soc.* **2009**, *131*, 438-439.

65. Hoffmann, C.; Gaietta, G.; Zürn, A.; Adams, S. R.; Terrillon, S.; Ellisman, M. H.; Tsien, R. Y.; Lohse, M. J. Fluorescent labeling of tetracysteine-tagged proteins in intact cells. *Nat. Protoc.* **2010**, *5*, 1666-1677.
66. Zhang, C.; Spokoiny, A. M.; Zou, Y.; Simon, M. D.; Pentelute, B. L. Enzymatic 'click' ligation: selective cysteine modification in polypeptides enabled by promiscuous glutathione S-transferase. *Angew. Chem. Int. Ed.* **2013**, *52*, 14001-14005.
67. Ramil, C. P.; An, P.; Yu, Z.; Lin, Q. Sequence-specific 2-cyanobenzothiazole ligation. *J. Am. Chem. Soc.* **2016**, *138*, 5499-5502.
68. Rashidian, M.; Dozier, J. K.; Distefano, M. D. Enzymatic Labeling of Proteins: Techniques and Approaches. *Bioconjugate Chem.* **2013**, *24*, 1277-1294.
69. Slavoff, S. A.; Chen, I.; Choi, Y.-A.; Ting, A. Y. Expanding the Substrate Tolerance of Biotin Ligase through Exploration of Enzymes from Diverse Species. *J. Am. Chem. Soc.* **2008**, *130*, 1160-1162.
70. Fernández-Suárez, M.; Baruah, H.; Martínez-Hernández, L.; Baskin, J. M.; Bertozzi, C. R.; Ting, A. Y. Re-directing lipoic acid ligase for cell surface protein labeling with small-molecule probes. *Nat. Biotechnol.* **2007**, *25*, 1483-1487.
71. Mao, H.; Hart, S. A.; Schink, A.; Pollok, B. A. Sortase-Mediated Protein Ligation: A New Method for Protein Engineering. *J. Am. Chem. Soc.* **2004**, *126*, 2670-2671.
72. Steffen, W.; Ko, F. C.; Patel, J.; Lyamichev, V.; Albert, T. J.; Benz, J.; Rudolph, M. G.; Bergmann, F.; Streidl, T.; Kratzsch, P.; Boenitz-Dulat, M.; Oelschlaegel, T.; Schraeml, M. Discovery of a microbial transglutaminase enabling highly site-specific labeling of proteins. *J. Biol. Chem.* **2017**, *292*, 15622-15635.
73. Appel, M. J.; Bertozzi, C. R. Formylglycine, a Post-Translationally Generated Residue with Unique Catalytic Capabilities and Biotechnology Applications. *ACS Chem. Biol.* **2015**, *10*, 72-84.
74. Sidhu, S. S.; Weiss, G. A.; Wells, J. A. High Copy Display of Large Proteins on Phage for Functional Selections. *J. Mol. Biol.* **2000**, *296*, 487-495.
75. Lamboy, J. A.; Arter, J. A.; Knopp, K. A.; Der, D.; Overstreet, C. M.; Palermo, E.; Urakami, H.; Yu, T.-B.; Tezgel, O.; Tew, G.; Guan, Z.; Kuroda, K.; Weiss, G. A. Phage Wrapping with Cationic Polymers Eliminates Non-specific Binding between M13 Phage and High pI Target Proteins. *J. Am. Chem. Soc.* **2009**, *131*, 16454-16460.
76. Kawakami, T.; Ogawa, K.; Goshima, N.; Natsume, T. DIVERSE System: De Novo Creation of Peptide Tags for Non-enzymatic Covalent Labeling by In Vitro Evolution for Protein Imaging Inside Living Cells. *Chem. Biol.* **2015**, *22*, 1671-1679.
77. Chang, P. V.; Bertozzi, C. R. Imaging beyond the proteome. *Chem. Commun.* **2012**, *48*, 8864-8879.
78. Silhavy, T. J.; Kahne, D.; Walker, S. The Bacterial Cell Envelope. *Cold Spring Harb. Perspect. Biol.* **2010**, *2*, a000414.

79. Whitfield, C.; Szymanski, C. M.; Aebi, M. Eubacteria. In *Essentials of Glycobiology* [Online], 3rd Ed.; Varki A, Cummings RD, Esko JD, et al., Eds.; Cold Spring Harbor Laboratory Press: New York, 2017. Chapter 21, https://www.ncbi.nlm.nih.gov/books/NBK453062/#_Ch21_s1_ (accessed May 8, 2018).
80. Brennan, P. J.; Nikaido, H. The Envelope of Mycobacteria. *Annu. Rev. Biochem.* **1995**, *64*, 29–63.
81. Jankute, M.; Cox, J. A. G.; Harrison, J.; Besra, G. S. Assembly of the Mycobacterial Cell Wall. *Annu. Rev. Microbiol.* **2015**, *69*, 405–423.
82. Magalhães M.L., Blanchard J.S. Aminoglycosides: Mechanisms of Action and Resistance. In *Antimicrobial Drug Resistance*; Mayers, D. L., Ed.; Infectious Disease. Humana Press, 2009.
83. Martínez-Rodríguez, S.; Martínez-Gómez, A. I.; Rodríguez-Vico, F.; Clemente-Jiménez, J. M.; Las Heras-Vázquez, F. J. Natural Occurrence and Industrial Applications of D-Amino Acids: An Overview. *Chem. Biodivers.* **2010**, *7*, 1531-1548.
84. Liang, H.; DeMeester, K. E.; Hou, C.-W.; Parent, M. A.; Caplan, J. L.; Grimes, C. L. Metabolic labelling of the carbohydrate core in bacterial peptidoglycan and its applications. *Nat. Commun.* **2017**, *8*, 15015.
85. Kuru, E.; Tekkam, S.; Hall, E.; Brun, YV; Van Nieuwenhze, M. S. Synthesis of fluorescent D-amino acids and their use for probing peptidoglycan synthesis and bacterial growth in situ. *Nat. Protoc.* **2015**, *10*, 33-52.
86. Siegrist, M. S.; Whiteside, S.; Jewett, J. C.; Aditham, A.; Cava, F.; Bertozzi, C. R. D-Amino Acid Chemical Reporters Reveal Peptidoglycan Dynamics of an Intracellular Pathogen. *ACS Chem. Biol.* **2013**, *8*, 500-505.
87. MacKenzie, D. A.; Sherratt, A. R.; Chigrinova, M.; Kell, A. J.; Pezacki, J. P. Bioorthogonal labelling of living bacteria using unnatural amino acids containing nitrones and a nitron derivative of vancomycin. *Chem. Commun.* **2015**, *51*, 12501-12504.
88. Siegrist, M. S.; Swarts, B. M.; Fox, D. M.; Lim, S. A.; Bertozzi, C. R. Illumination of growth, division and secretion by metabolic labeling of the bacterial cell surface. *FEMS Microbiol. Rev.* **2015**, *39*, 184-202.
89. Shieh, P.; Siegrist, M. S.; Cullen, A. J.; Bertozzi, C. R. Imaging bacterial peptidoglycan with near-infrared fluorogenic azide probes. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 5456-5461.
90. Herget, S.; Toukach, P. V.; Ranzinger, R.; Hull, W. E.; Knirel, Y. A.; von der Lieth, C.-W. Statistical analysis of the Bacterial Carbohydrate Structure Data Base (BCSDB): Characteristics and diversity of bacterial carbohydrates in comparison with mammalian glycans. *BMC Struct. Biol.* **2008**, *8*:35.

91. Swarts, B. M.; Holsclaw, C. M.; Jewett, J. C.; Alber, M.; Fox, D. M.; Siegrist, M. S.; Leary, J. A.; Kalscheuer, R.; Bertozzi, C. R. Probing the Mycobacterial Trehalome with Bioorthogonal Chemistry. *J. Am. Chem. Soc.* **2012**, *134*, 16123-16126.
92. Foley, H. N.; Stewart, J. A.; Kavunja, H. W.; Rundell, S. R.; Swarts, B. M. Bioorthogonal Chemical Reporters for Selective In Situ Probing of Mycomembrane Components in Mycobacteria. *Angew. Chem. Int. Ed.* **2016**, *55*, 2053–2057.
93. Rundell, S. R.; Wagar, Z. L.; Meints, L. M.; Olson, C. D.; O'Neill, M. K.; Piligian, B. F.; Poston, A. W.; Hood, R. J.; Woodruff, P. J.; Swarts, B. M. Deoxyfluoro-D-trehalose (FDTre) analogues as potential PET probes for imaging mycobacterial infection. *Org. Biomol. Chem.* **2016**, *14*, 8598-8609.
94. Backus, K. M.; Boshoff, H. I.; Barry, C. S.; Boutureira, O.; Patel, M. K.; D'Hooge, F.; Lee, S. S.; Via, L. E.; Tahlan, K.; Barry, C. E. III.; Davis, B. G. Uptake of unnatural trehalose analogs as a reporter for *Mycobacterium tuberculosis*. *Nat. Chem. Biol.* **2011**, *7*, 228-235.
95. Rodríguez-Rivera, F. P.; Zhou, X.; Theriot, J. A.; Bertozzi, C. R. Visualization of mycobacterial membrane dynamics in live cells. *J. Am. Chem. Soc.* **2017**, *139*, 3488-3495.
96. Rodríguez-Rivera, F. P.; Zhou, X.; Theriot, J. A.; Bertozzi, C. R. Acute modulation of mycobacterial cell envelope biogenesis by front-line TB drugs. *Angew. Chem. Int. Ed.* **2018**, *57*, 5267-5272.
97. Hodges, H. L.; Brown, R. A.; Crooks, J. A.; Weibel, D. B.; Kiessling, L. L. Imaging mycobacterial growth and division with a fluorogenic probe. *Proc. Natl. Acad. Sci.* **2018**. 201720996; published ahead of print April 27, 2018. <https://doi.org/10.1073/pnas.1720996115>.
98. Shieh, P.; Dien, V. T.; Beahm, B. J.; Castellano, J. M.; Wyss-Coray, T.; Bertozzi, C. R. CalFluors: A Universal Motif for Fluorogenic Azide Probes across the Visible Spectrum. *J. Am. Chem. Soc.* **2015**, *137*, 7145-7151.
99. Shieh, P.; Bertozzi, C. R. Design strategies for bioorthogonal smart probes. *Org. Biomol. Chem.* **2014**, *12*, 9307-9320.
100. Klymchenko, A. S.; Mély, Y. Fluorescent Environment-Sensitive Dyes as Reporters of Biomolecular Interactions. *Prog. Mol. Biol. Transl. Sci.* **2013**, *113*, 35-58
101. Loving, G. S.; Sainlos, M.; Imperiali, B. Monitoring Protein Interactions and Dynamics with Solvatochromic Fluorophores. *Trends Biotechnol.* **2010**, *28*, 73-83.
102. Klymchenko, A. S. Solvatochromic and Fluorogenic Dyes as Environment-Sensitive Probes: Design and Biological Applications. *Acc. Chem. Res.* **2017**, *50*, 366-375.

Chapter 2

Computation-guided rational design of a peptide motif that reacts with cyanobenzothiazoles via internal cysteine-lysine relay

Chapter 2. Computation-guided rational design of a peptide motif that reacts with cyanobenzothiazoles via internal cysteine-lysine relay

INTRODUCTION

Site-selective protein modification is an important tool for imaging targets *in vitro* and *in vivo*, creating new therapeutics, and installing functionality on proteins to probe activity.¹⁻⁴ However, the limited number of proteinogenic amino acids and their natural repetition in proteins limits them as targets for site-selective ligation. Easy access to different modifications, site selectivity, and the potential for a large substrate scope makes covalent modification of engineered peptide tags by small molecules an attractive and widely used platform for biotherapeutic and basic research.⁵ Much like an enzyme active site, these peptide tags are engineered to produce a microenvironment that enables preferential ligation to a peptide sequence of interest, thus allowing for selective and specific labeling of a protein. For application in biological systems, any reaction between the peptide tag and the small molecule should proceed in aqueous conditions, at physiological pH and temperature, and, if intended for *in cellulo* or *in vivo* studies, without deleterious effects.

Previously, Tsien and coworkers described a foundational approach to rational design of a peptide sequence that can undergo stable modification by a small molecule.⁶⁻⁷ In their sequence (Cys-Cys-Pro-Gly-Cys-Cys), four cysteine residues are primed to rapidly react with FIAsh-EDT₂, a fluorescein-based diarsenical probe. The engineered peptide tag consists entirely of proteinogenic amino acids, yet selective ligation is achievable –any off-target interactions with naturally occurring peptide sequences can be reversed with addition of ethanedithiol (EDT). As no unnatural amino acids or changes to the endogenous translational machinery are necessary, perturbations to the cell are minimized. The early success of this rationally designed peptide for protein imaging inspired an ongoing wave of research into novel peptide tags.

Recently, Pentelute and colleagues introduced the shortest small molecule-binding peptide to date, the ‘ π -clamp’ (Phe-Cys-Pro-Phe).⁸ Using a library selection method to identify peptides that are subject to arylation reactions in water, they found that this four-residue tag provided a favorable microenvironment for the cysteine residue to displace a fluorine substituent on a perfluoroaromatic probe via a nucleophilic aromatic substitution. As they had previously shown that endogenous cysteines do not react with perfluoroaryl moieties in water,⁹ the discovery of π -clamp indicates the feasibility of designing other short natural amino acid sequences that elicit emergent functionality.

Both the FIAsh tag and the π -clamp exploit the tunability of cysteine’s reactivity when developing new peptide conjugation methods. In addition, cysteine is the most reactive of the twenty canonical amino acids at physiological pH and appears in proteins at a relatively low frequency, which makes it a prime target for ligation chemistry.^{2-3,5} However,

cysteine plays fundamental structural and chemical roles in proteins, so modifying or removing endogenous cysteine residues from a protein of interest to accommodate an engineered peptide tag is not always feasible. Thus, engineered peptide tags that enable site-selective modification of different amino acid residues without requiring exogenous enzymes or translational machinery remain an unmet need.

Another amino acid that is nucleophilic under certain conditions is lysine. Cravatt and coworkers recently reported the results of a global profiling experiment on lysine reactivity in the human proteome.¹⁰ They found that of the more than 9,000 lysines they quantified in the human proteome, a few hundred had heightened reactivity, and over one hundred could be targeted by amine-reactive electrophilic fragments. As lysine is a common surface-exposed residue but non-nucleophilic at physiological pH due to protonation of the side chain amine, these data suggest peptide tags can form a microenvironment conducive to lysine modification. Further supporting this hypothesis is the recent work of Bernardes and coworkers, in which the site-selective modification of the single most reactive lysine on a given protein by sulfonyl acrylate reagents was reported.¹¹ Moreover, the authors were able to computationally predict which lysine would be modified in each of the five proteins they analyzed.

In designing a new small molecule-binding peptide tag, we took inspiration from the enhanced amine reactivity seen in native chemical ligation (NCL) between an N-terminal cysteine and a C-terminal thioester (Figure 2.1A).¹² During NCL, the nucleophilic N-terminal cysteine thiol undergoes reversible transthioesterification with the C-terminal thioester. Then, the acyl group is irreversibly transferred to the N-terminal amine, resulting in a new amide bond between the two original peptides. The transthioesterification step increases the effective concentration of the amine, greatly enhancing the rate of amide bond formation. Transglutaminases use a similar concept, forming a thioester bond between an enzyme cysteine and a substrate glutamine which can then be attacked by a second peptide's N-terminal amine or a lysine side chain to form an isopeptide bond.¹³⁻¹⁴ In the case of NCL, two distinct nucleophiles on the same residue are key to allowing amide bond formation to proceed at a reasonable rate. We hypothesized that an engineered peptide could position a cysteine residue and a lysine residue in such a way as to attain an analogous functionality to NCL, but without the requirement of an N-terminal cysteine.

Additionally, we were drawn to the reaction between 2-cyanobenzothiazole derivatives and 1,2-aminothiols. In the most widely used application, firefly luciferin is synthesized by reacting 6-hydroxy-2-cyanobenzothiazole with D-cysteine in the laboratory and in nature (Figure 2.1B).¹⁵⁻¹⁶ The cysteine thiol and the nitrile group react to form a thioimidate, which is then attacked by the lysine amine to form a tetrahedral intermediate. In contrast to NCL, in which the thiol "hands off" the acyl group to the amine, the tetrahedral intermediate eliminates a molecule of ammonia to reach the final product. Rao and coworkers found that the reaction of a 6-amino-2-cyanobenzothiazole (CBT) derivative with cysteine proceeds with a second order rate constant of $9 \text{ M}^{-1} \text{ s}^{-1}$ *in vitro* at room temperature and

physiological pH.¹⁷ This feature as well as the reversibility of the reaction between CBT and endogenous thiols allowed them to develop a method to selectively label engineered N-terminal cysteines *in vivo*. Other groups have also devised clever ways to utilize this reaction.¹⁸⁻²⁰ Most recently, Lin and coworkers reported a genetically encodable CBT derivative-binding peptide tag discovered through a phage display experiment.²¹ As with previous examples of engineered peptide tags, this too is a cysteine conjugation tag, and substituting adjacent residues with lysine did not appear to change the mode or mechanism of binding.

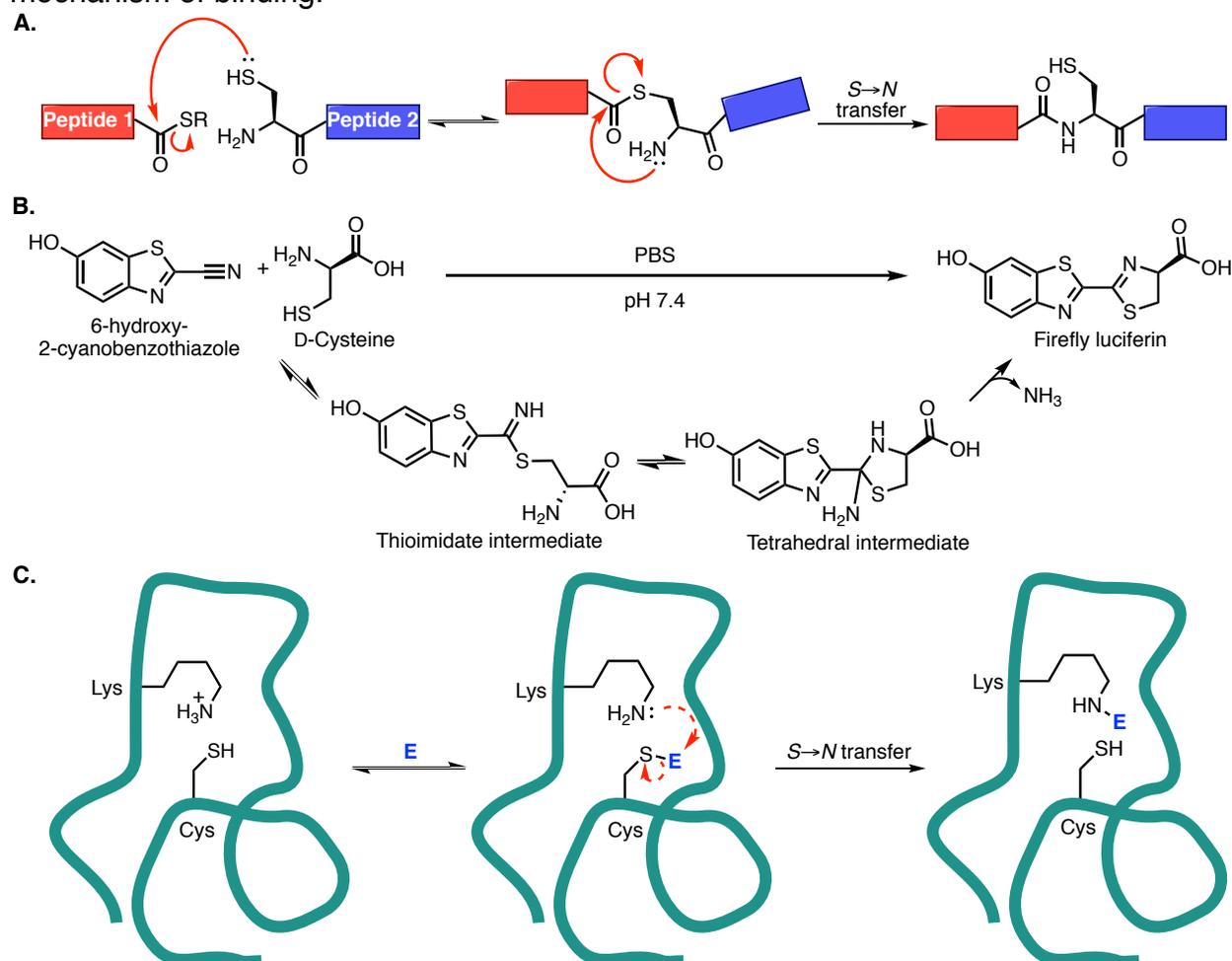


Figure 2.1. Chemoselective reactions involving aminothiols. **A.** Mechanism of native chemical ligation. Transthioesterification occurs when an N-terminal cysteine on one peptide attacks a C-terminal thioester on another. An $S \rightarrow N$ acyl transfer occurs, irreversibly forming a new amide bond. **B.** 6-Hydroxy-2-cyanobenzothiazole and D-cysteine react to form firefly luciferin. The cysteine thiol attacks the nitrile group to form a thioimidate intermediate, then the amino group attacks the thioimidate to form a tetrahedral intermediate. Finally, irreversible loss of ammonia drives the reaction to completion. **C.** A proposed peptide (teal line) that uses cysteine and lysine residues to accomplish a reaction analogous to the native chemical ligation. The cysteine attacks the electrophile and then transfers it to the lysine residue in an irreversible reaction.

With both NCL and the CBT reaction in mind, we focused on developing a short, genetically encodable peptide tag that selectively and irreversibly ligates the nitrile group of a CBT derivative under physiological conditions. To do so, we designed *in silico* a peptide with a lysine and a cysteine residue oriented such that a selective reaction with the nitrile group of an *N*-acylated 6-amino-2-cyanobenzothiazole would occur (Figure 2.1C). Herein, we show that our first-generation engineered peptide creates a microenvironment for a “cysteine hand-off” mechanism analogous to NCL, resulting in an irreversible, covalent ligation.

RESULTS AND DISCUSSION

Initially, we considered selecting candidate peptide sequences using the phage display method reported previously by Weiss, Wells, and coworkers.²²⁻²⁴ However, this method requires a disulfide bond in the peptide library to restrict the conformational space during the selection process, which complicates library generation and screening for our reaction of interest as our intended peptide would require a reduced cysteine for reaction with CBT. Considering the knowledge that phage display can lead to unanticipated results²⁵ and the nontrivial nature of designing selection conditions that would favor a final product with the CBT analog attached to the lysine rather than the cysteine, we alternatively pursued a rational design *in silico*. This approach yielded several short peptide sequences (ten to eleven amino acids long) that efficiently bind cyanobenzothiazole derivatives under various conditions.

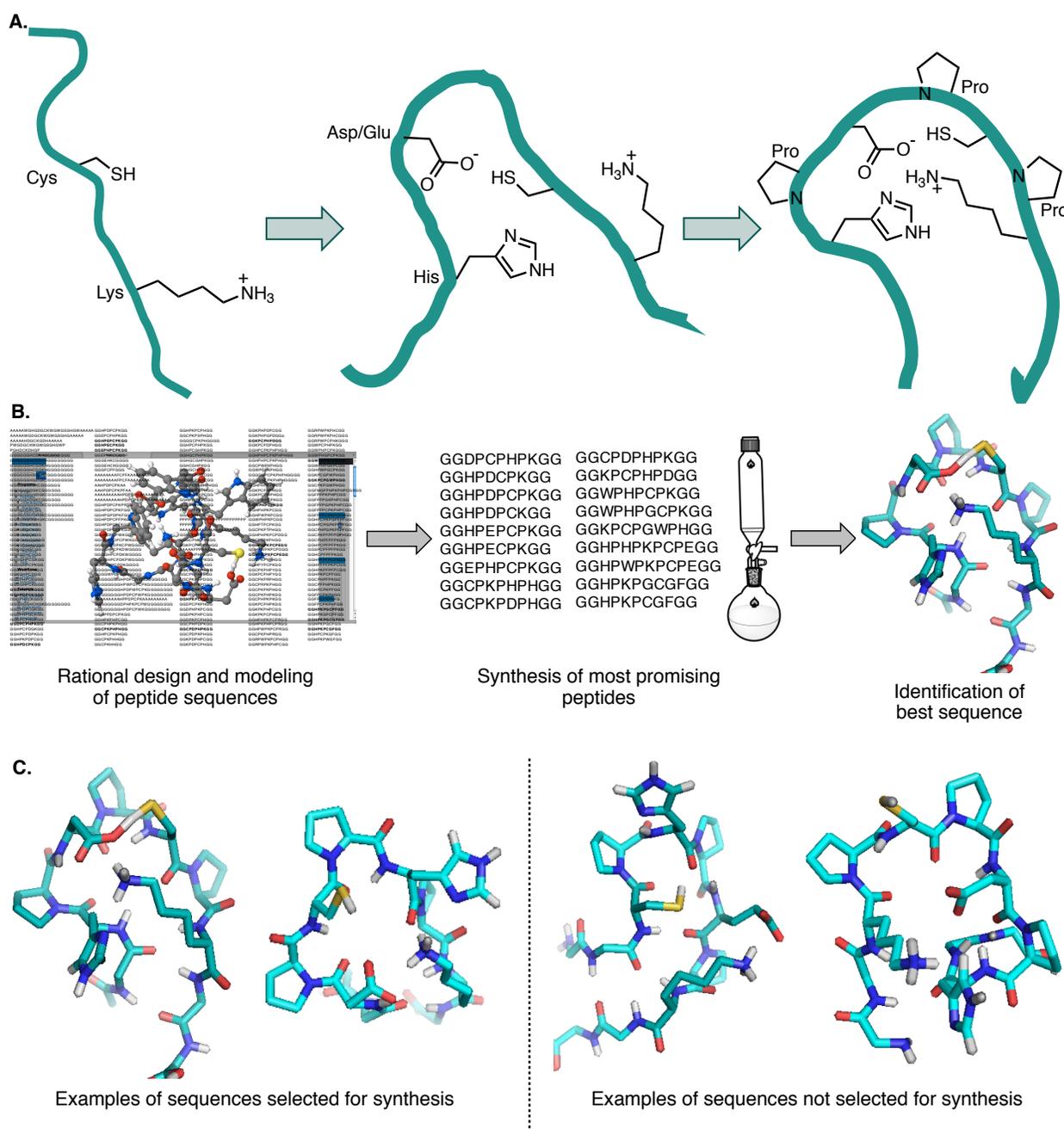


Figure 2.2. Rational design of a peptide that covalently binds a CBT derivative. **A.** The principles behind generating peptide sequences for computational modeling. Lysine and cysteine residues were interspersed with histidine and aspartic or glutamic acid residues to increase nucleophilicity. Prolines were added to lock the sequence (teal) into a favorable conformation for reaction with a CBT derivative. **B.** Selection of sequences to test for reaction with a CBT derivative. 218 sequences were modeled using the program PEP-FOLD (see Supplementary Table 2.S1 for a full list).²⁶⁻²⁷ Of these peptides, eighteen that best fit our selection criteria were synthesized by solid-phase peptide synthesis (SPPS) and tested in a binding assay. **C.** Examples demonstrating how peptides were

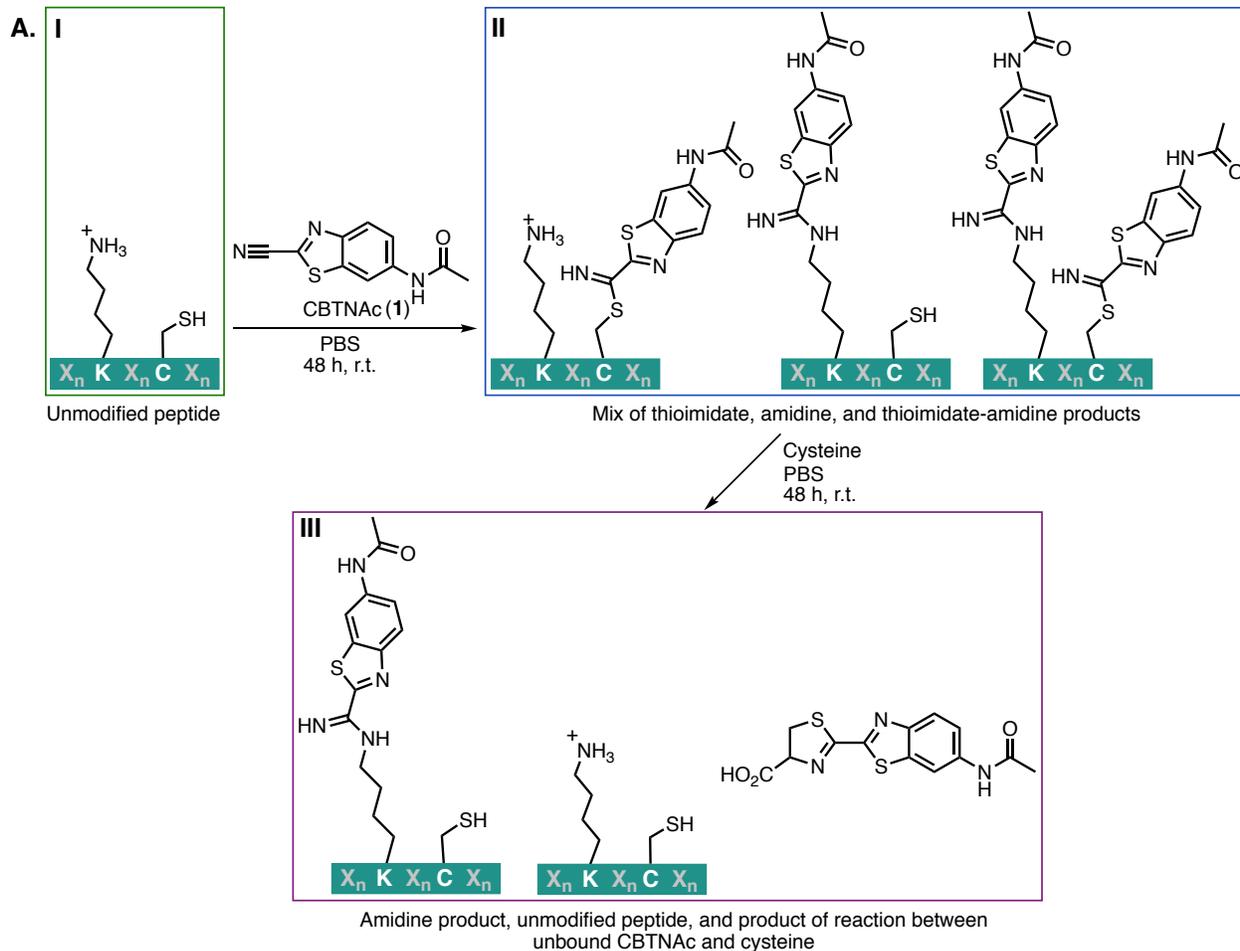
selected for SPPS. The peptides on the left were two of the best models, whereas the two peptides on the right were not chosen for SPPS.

For the *in silico* approach, the online computational tool PEP-FOLD²⁶⁻²⁷ was used to model short peptides containing lysine, cysteine, histidine, aspartic or glutamic acid, and proline residues. Inspired by enzyme-resident catalytic triads, we hypothesized that histidine and aspartic acid would increase the nucleophilicity of cysteine and lysine residues if proline residues were incorporated to help lock the peptide into the desired conformation (Figure 2.2A). In addition, the protonated imidazole of the histidine residue could also catalyze the reaction between the cysteine thiol and the nitrile group of a CBT derivative.²⁸⁻²⁹ Of the 218 sequences modeled (Supplementary Table 2.S1), eighteen were selected for synthesis via solid phase peptide synthesis (SPPS) and further analysis (Figure 2.2B, C). In Figure 2.2C, the two peptides on the left are representative of the best models: the lysine, histidine, aspartic acid, and cysteine residues are all in relatively close proximity. In particular, the histidine side chain appeared poised to help deprotonate the lysine residue. The two peptides on the right do not fit our criteria: either the histidine residue was not in a position to deprotonate the cysteine or the lysine residue or the cysteine and lysine residues appeared too far apart to feasibly work together in forming covalent bonds with a CBT derivative.

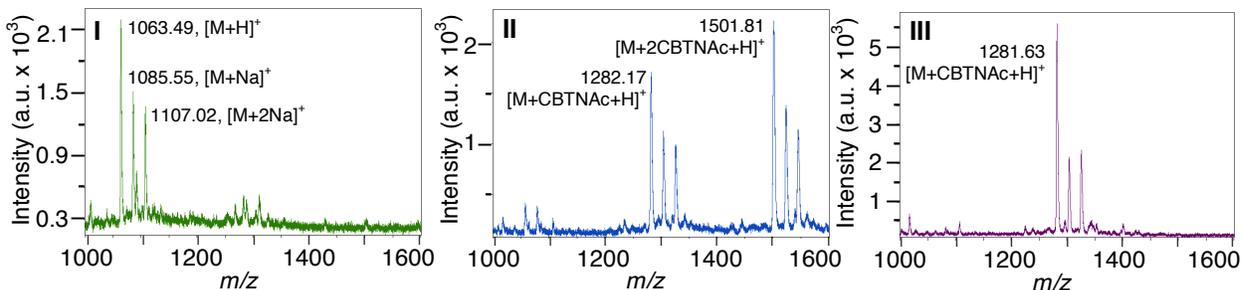
Based on MALDI-TOF MS, many of the peptides reacted with *N*-(2-(cyano-¹³C)-benzo[*d*]thiazol-6-yl)acetamide (CBTNAc, compound **1**) in phosphate-buffered saline (PBS) at room temperature over 48 hours even in the absence of reducing agents such as dithiothreitol (DTT) or tris(2-carboxyethyl)phosphine (TCEP) (assay workflow and representative data shown in Figure 3; data for all eighteen peptides, peptides **4-21**, shown in Supplementary Figure 2.S1). Several but not all of the peptides also reacted with two equivalents of the CBT derivative. We hypothesized that if only one equivalent of CBTNAc was bound, it could be attached to either the lysine or cysteine residue, while if two were bound, they were attached to both the lysine and cysteine residues. In a follow-up experiment, free cysteine was added to the peptide-CBTNAc adducts to trap unbound CBTNAc as a luciferin analog. As the reaction of the peptide's cysteine residue with the CBTNAc nitrile group is reversible, any CBTNAc bound to the cysteine and not transferred to the lysine residue should be in equilibrium with free peptide and CBTNAc. When free cysteine is added, it reacts irreversibly with unbound CBTNAc to form the luciferin analog. Depletion of the unbound CBTNAc in this manner will then drive the peptide-CBTNAc adduct toward free CBTNAc and an unmodified cysteine residue on the peptide by Le Châtelier's Principle. Indeed, addition of free cysteine resulted in significant loss of CBTNAc from the peptides, with only one equivalent remaining on some (Figure 2.3B, Supplementary Figure 2.S1) and no modification remaining on others (Figure 2.3C, Supplementary Figure 2.S1).

Placing the cysteine and the lysine residues adjacent to one another resulted in minimal reaction with CBTNAc, but placing a proline in between them was not in itself sufficient to ensure significant formation of a stable adduct (Supplementary Figure 2.S1F, 2.S1O),

while placing additional residues in between the cysteine and the lysine resulted in formation of just as much or more of the stable CBTNAc adduct as some peptides with cysteine and lysine close together (e.g., Supplementary Figure 2.S1C, 2.S1J, and 2.S1Q compared to 2.S1F, 2.S1O). Exchanging aspartic acid and glutamic acid altered the reactivity of a given peptide (Supplementary Figure 2.1A, 2.1E), as did replacing an aspartic or glutamic acid residue with or adding a residue such as histidine (Supplementary Figure 2.S1H, 2.S1O), tryptophan (Supplementary Figure 2.S1L, 2.S1M, 2.S1N, 2.S1P), or phenylalanine (Supplementary Figure 2.S1Q, 2.S1R).



B. Peptide with extensive modification of Lys: GGHPDPCPKGG



C. Peptide with minimal modification of Lys: GGDHPCKGG

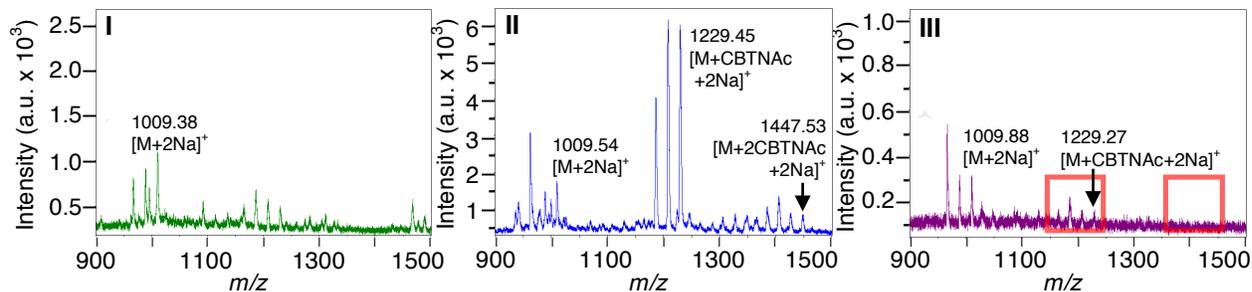


Figure 2.3. Assay for peptides that are covalently modified by CBTNAC (**1**) and representative results. **A.** Overview of assay used to determine degree of irreversible covalent modification of candidate peptides. Candidate N-terminal acetylated, C-terminal amidated peptides were incubated with CBTNAC in PBS for 48 h at room temperature, then an aliquot of the reaction mixture was analyzed by MALDI-TOF MS. The remaining solution was incubated for 48 h at room temperature with free cysteine, with or without DTT or TCEP reducing agent, and the reaction mixture was again analyzed by MALDI-TOF MS. **B.** Representative MALDI-TOF MS spectra at each stage of the assay for a peptide with a lysine residue extensively modified by CBTNAC (peptide **4**). The first spectrum (I) shows crude peptide alone, the second spectrum (II) was taken after incubation with CBTNAC and desalting, and the third spectrum (III) followed incubation with free cysteine. **C.** Representative MALDI-TOF MS spectra at each stage of the assay for a peptide with a lysine residue not significantly modified by CBTNAC (peptide **5**). The samples analyzed in each spectrum were prepared the same way as in B. Red boxes indicate the expected locations of masses corresponding to modified peptide.

Of the eighteen peptides tested, we proceeded with characterization of the peptide modified to the greatest extent by CBTNAC, CBTag 1.0 (Ac-Gly-Gly-His-Pro-Asp-Pro-Cys-Pro-Lys-Gly-Gly-NH₂, **4**). Performing the assay in the presence of 5 mM glutathione to mimic physiological thiol conditions still resulted in a significant amount of the post-cysteine treatment CBTag 1.0-CBTNAC adduct, though less of it than when conducting the assay in the absence of glutathione (Figure 2.4). These results indicated that some competition between endogenous thiols and the CBTag 1.0 cysteine is probable, but, promisingly, a stable bond between CBTag 1.0 and CBTNAC can still form.

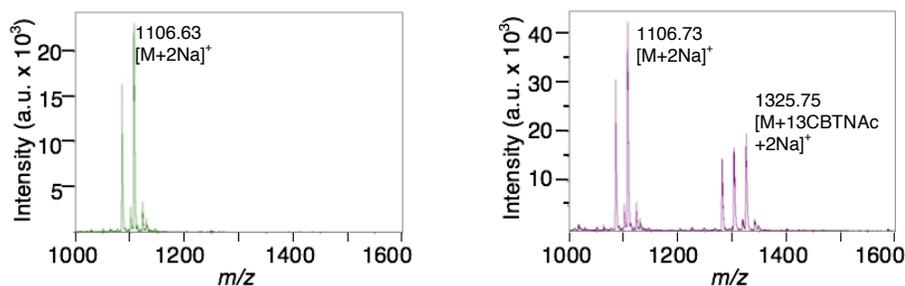
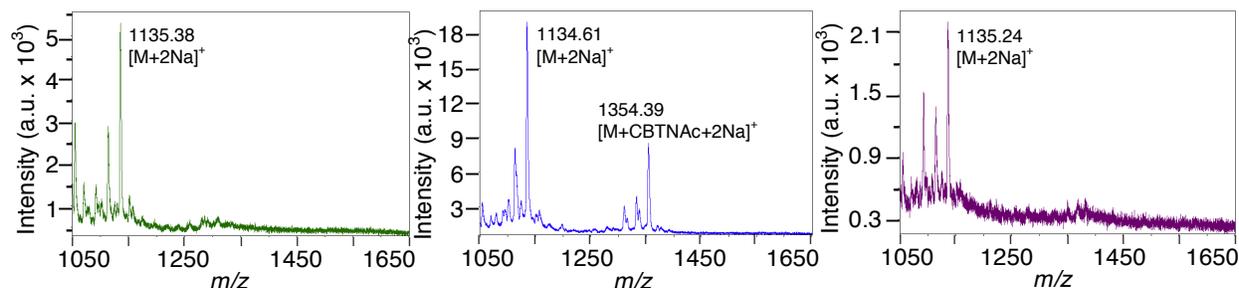


Figure 2.4. Results of the ¹³CBTNAC binding assay and cysteine treatment for CBTag 1.0 when conducted in the presence of 5 mM glutathione (reduced). Left (green spectrum), negative control after treatment with cysteine and DTT. Right (purple spectrum), CBTag 1.0 treated with 1.1 equiv. of ¹³CBTNAC then with 10 equiv. cysteine and 15 equiv. DTT relative to ¹³CBTNAC. CBTag 1.0-¹³CBTNAC adduct is observed, though less has formed than in the absence of glutathione.

To determine the necessity of the cysteine and the lysine residues, two modified peptides were synthesized: one with the lysine residue converted to arginine to preserve the positive charge while decreasing nucleophilicity further (**22**), and one with the cysteine residue converted to serine to decrease nucleophilicity (**23**). These derivatives were subjected to the conditions described in the previous paragraph. The lysine to arginine mutant reacted with one equivalent of CBTNAc before cysteine treatment, likely due to the cysteine residue in the peptide (Figure 2.5A). Treatment with free cysteine abrogated binding, resulting in free peptide, as expected. The cysteine to serine mutant did not bind any equivalents of CBTNAc, indicating that the cysteine residue is necessary for activity (Figure 2.5B). Thus, both cysteine and lysine are necessary for irreversible covalent bond formation with CBTNAc. Combined with the earlier observation that not all of the synthesized peptides irreversibly reacted with CBTNAc, these data led us to propose that cysteine and lysine are necessary but not sufficient for CBTNAc binding, indicating that conformation of the peptide is important for the intended reactivity.

A.



B.

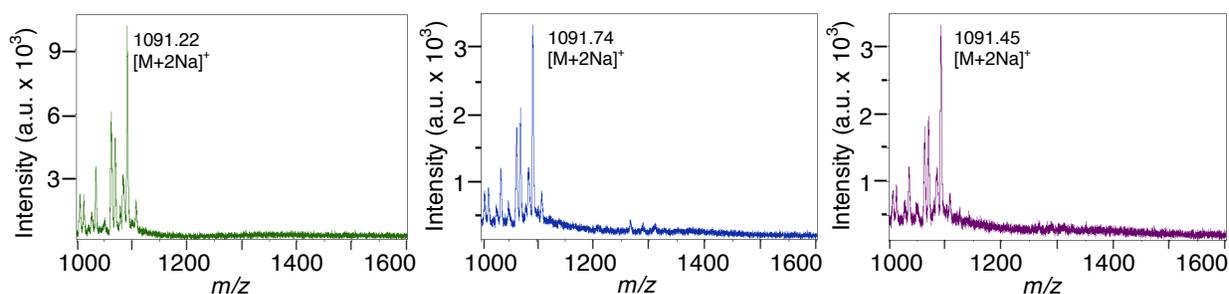
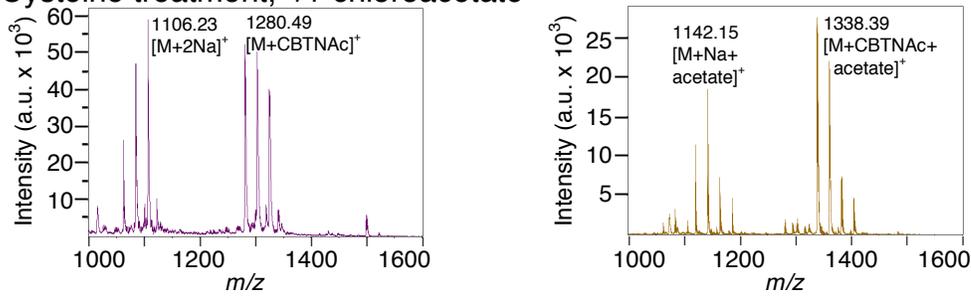


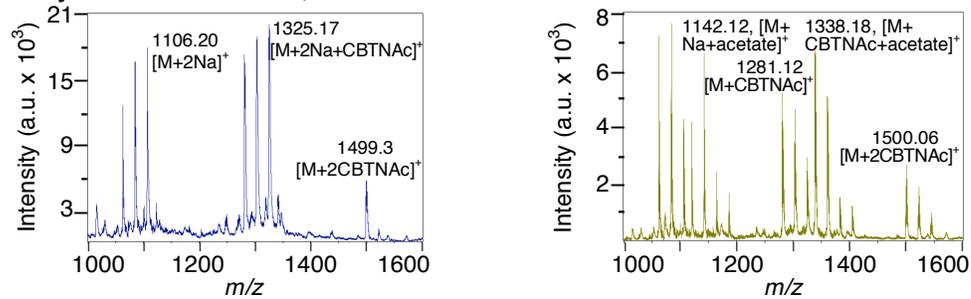
Figure 2.5. Results of binding experiment and cysteine competition assays for CBTAg 1.0 (K9R) (**A.**) and CBTAg 1.0 (C7S) (**B.**). Spectra are analogous to those in Figure 3. The first spectrum in each row corresponds to crude peptide, the second to the results of the binding assay, and the third to the results of the cysteine competition. In **A.**, CBTAg 1.0 (K9R) is partially modified by one equivalent of CBTNAc, and that equivalent is removed by cysteine treatment, suggesting that, as expected, the modification is on the cysteine residue. In **B.**, CBTAg 1.0 (C7S) is not at all modified by CBTNAc, implying that the cysteine residue is necessary for lysine modification.

To further establish that the remaining equivalent of CBTNAc was not attached to the cysteine residue, the peptide-CBTNAc adduct was reacted with chloroacetate after treatment with free cysteine. Chloroacetate specifically reacts with thiols, not amines, and thus should only label the peptide-CBTNAc adduct if the CBTNAc is not bound to the cysteine residue (i.e., if the CBTNAc is bound to lysine). Indeed, a clean mass shift indicating reaction with chloroacetate was observed for the CBTag 1.0-CBTNAc adduct after cysteine treatment (Figure 2.6A). Moreover, when CBTag 1.0 was incubated with CBTNAc but not treated with free cysteine, chloroacetate did not react with any peptide bound to two equivalents of CBTNAc. If the CBTag 1.0 peptide was pretreated with chloroacetate, no binding of CBTNAc occurred, though this could be due to the addition of a negative charge causing a conformational shift (Figure 2.6B, C). As a control, we confirmed that the cysteine to serine mutant did not react with chloroacetate, while the lysine to arginine mutant did (Figure 2.6D).

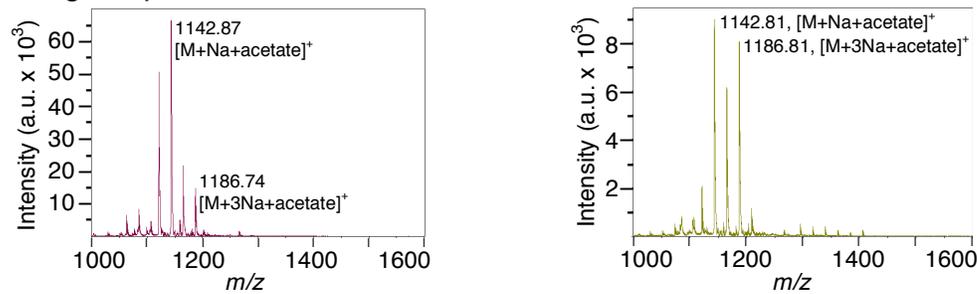
A. Cysteine treatment, +/- chloroacetate



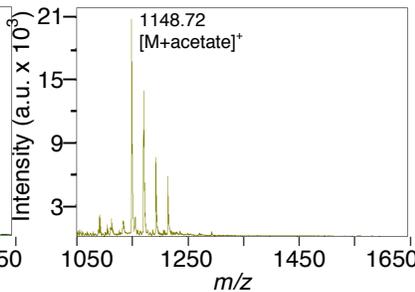
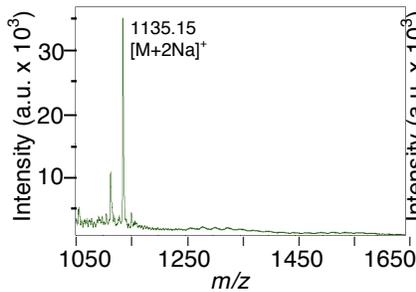
B. No cysteine treatment, +/- chloroacetate



C. CBTag 1.0 pretreated with chloroacetate, +/- CBTNAC



D. CBTag 1.0 (K9R) +/- chloroacetate



CBTag 1.0 (C7S) +/- chloroacetate

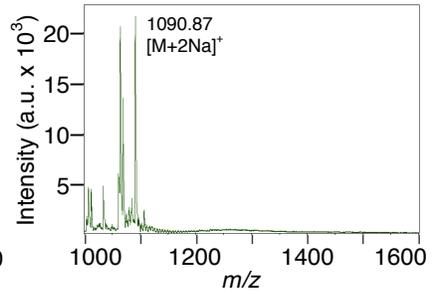


Figure 2.6. Results of chloroacetate labeling of thiols under various conditions. **A.** The left MALDI-TOF MS spectrum is after cysteine treatment of CBTag 1.0 that had been reacted with CBTNAC. The right spectrum are the results of treating the sample in the left spectrum with chloroacetate. The peaks shift by 57 m/z , indicating labeling of a thiol and suggesting CBTNAC is attached to the lysine residue. **B.** The left MALDI-TOF MS spectrum is after CBTag 1.0 was reacted with CBTNAC but before cysteine treatment.

Chloroacetate treatment of this sample gave the spectrum on the right, where one population of the singly modified peptide was labeled (lysine adduct) and another was not (cysteine adduct). The peptide modified by two equivalents of CBTNAc was not labeled by chloroacetate. **C.** When CBTag 1.0 was pretreated with chloroacetate (left), no CBTNAc reacted with the resulting peptide (right). **D.** As a control to confirm chloroacetate thiol specificity, CBTag 1.0 (K9R) (top left) and CBTag 1.0 (C7S) (bottom left) were treated with chloroacetate. Only CBTag 1.0 (K9R) (top right) was labeled, confirming that lysine residues are not modified by chloroacetate.

Additionally, tandem mass spectrometry (MS/MS) on CBTag 1.0 following cysteine treatment confirmed that the remaining CBTNAc was indeed bound to lysine (Figure 2.7). Finally, ^{13}C NMR spectra acquired after reaction between CBTNAc with a ^{13}C -labeled nitrile group (13CBTNAc, compound **3**) and CBTag 1.0 did not contain a peak corresponding to a stable tetrahedral intermediate but did contain peaks that match predicted values for the α -carbon of a thioimidate or amidine bond (CBTag 1.0-13CBTNAc amidine adduct **24**, Figure 2.8).

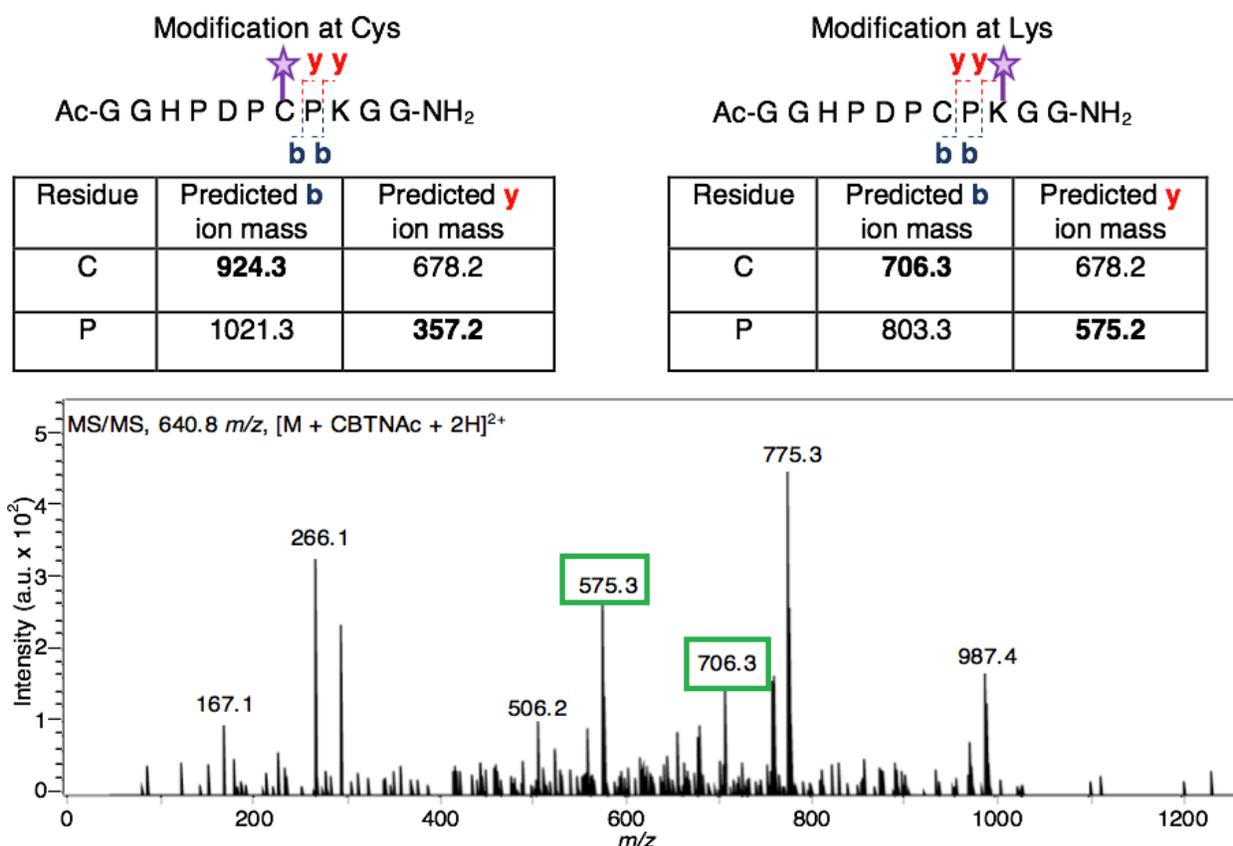


Figure 2.7. Tandem mass spectrometry (MS/MS) on CBTag 1.0-CBTNAc demonstrates that modification is on the lysine residue. Top, predicted masses for MS/MS b and y product ions if the CBTNAc (purple star) is on the cysteine residue (left) or the lysine

residue (right). Bottom, tandem MS/MS on the $[M + \text{CBTNAc} + 2\text{H}]^{2+}$ ion shows product ions matching those predicted for lysine modification (green boxes).

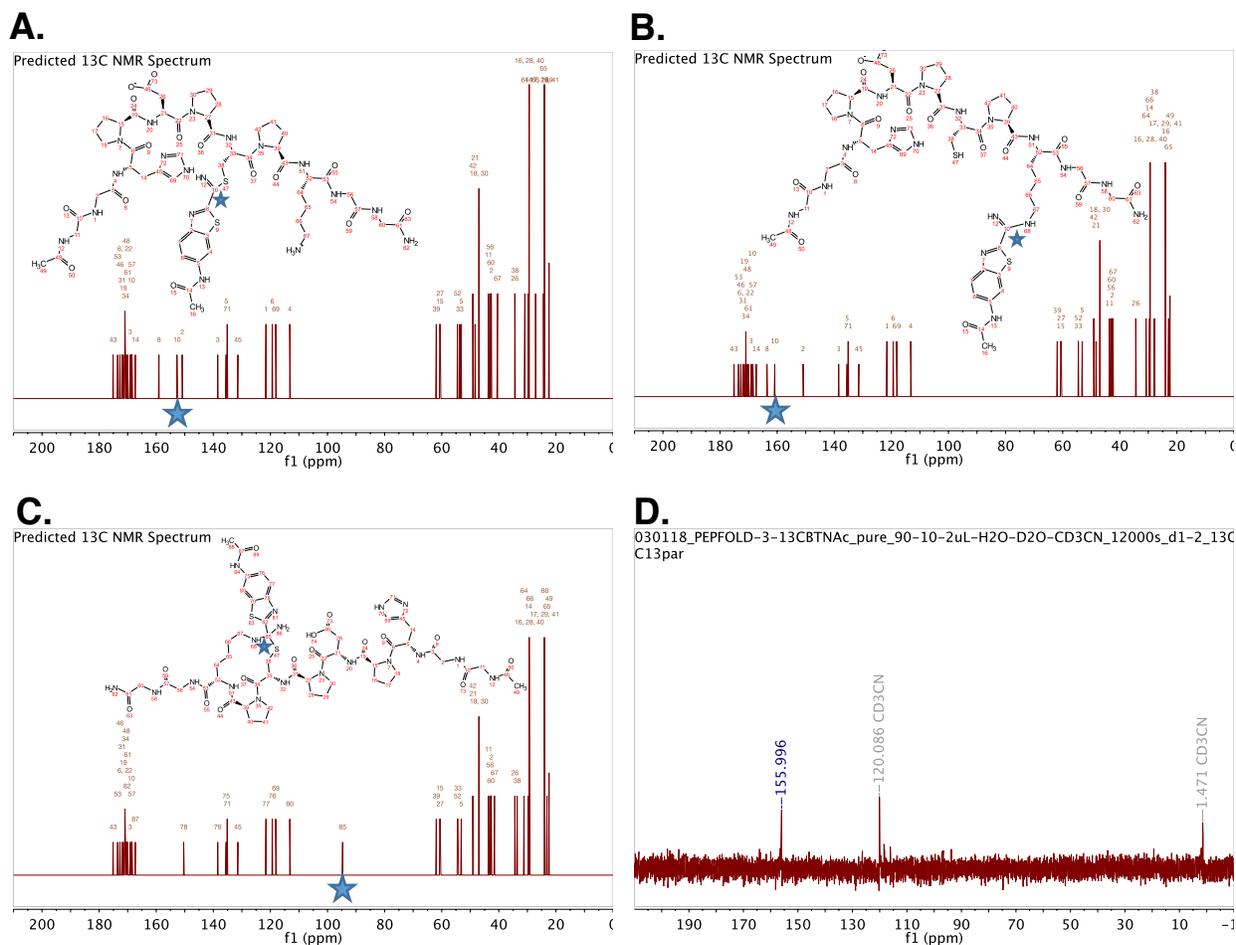


Figure 2.8. ^{13}C NMR confirms that the final CBTNAc adduct is not the tetrahedral intermediate. **A.-C.**, Predicted spectra (MestReNova) for the cysteine adduct, the lysine adduct, and the tetrahedral intermediate. α -Carbon and corresponding peak are labeled with a blue star. **D.** ^{13}C NMR spectrum of (dilute) purified CBTNAc shows that the peak corresponding to the ^{13}C -labelled α -carbon is at 156 ppm (referenced to acetonitrile- d_3 methyl peak), which is similar to the predicted values for the lysine and cysteine adducts but significantly different from the predicted value for the tetrahedral intermediate.

In order to quantitate relative amounts of starting materials, intermediates, and products over time, we developed an LC-MS method to follow the course of the reaction. As expected, the lysine adduct, the cysteine adduct, and the doubly modified peptide were resolvable using our method, though higher concentrations of the peptide than were used in the MALDI-TOF assay were necessary in order to clearly observe a 210 nm signal suitable for integration (Supplementary Figure 2.S2-2.S3). In order to deconvolute the

data, TCEP was included in the injected samples to remove peaks corresponding to disulfide dimers. Comparing peak areas in the chromatographs over time, we found that at the first time point (1.75-2.8 h), the majority of the CBTNAc attached to CBTAG 1.0 at the final time point (40-45 h) was already covalently bound to either a cysteine or a lysine residue; however, only 7-10% of this peptide-bound CBTNAc population was attached to lysine, suggesting the thioimidate intermediate forms relatively quickly but transfer to the lysine is slow (Figure 2.9, Supplementary Table 2.S2-2.S4). By the end of the experiment, approximately 35-41% of peptide-bound CBTNAc was on a lysine residue whether the ratio of CBTAG 1.0 to CBTNAc was 1 to 1 (22.5 mM each), 4 to 1 (22.5 mM and 5.63 mM), or 1 to 4 (5.63 mM and 22.5 mM), while the percentage of total CBTAG 1.0 that had a lysine residue modified by CBTNAc ranged from 17 to 38% depending on the conditions (Figure 2.9, Supplementary Table 2.S2-2.S4). The low variation in rate of formation of lysine-bound CBTNAc despite significant changes in concentration of the peptide also provide evidence that transfer of the CBTNAc is intramolecular, not intermolecular. These numbers matched the approximate yields of 16-36% we obtained when we isolated the amidine product after cysteine treatment. However, somewhat unexpectedly based on the MALDI-TOF MS results, conversion of the free peptide to peptide modified on the cysteine, the lysine, or both ranged from 38 to 80% - under the conditions of this assay, full conversion to a bound state was not observed even when CBTNAc was present in excess (Figure 2.9, Supplementary Table 2.S2-2.S4). We attribute this to reversion of some portion of the cysteine monoadduct to the free peptide and unreacted CBTNAc on passage through the column (no peaks corresponding to hydrolysis of CBTNAc were observed), unforeseen effects of including TCEP in the reaction, aggregation of the peptide or the CBTNAc at higher concentrations, or differences between running the reaction in a test tube and an LC-MS vial. Again, as a confirmation, the lysine to arginine mutant did form a peak corresponding to cysteine modification over time, while the cysteine to serine mutant spectrum did not change (Supplementary Figure 2.S4-2.S6).

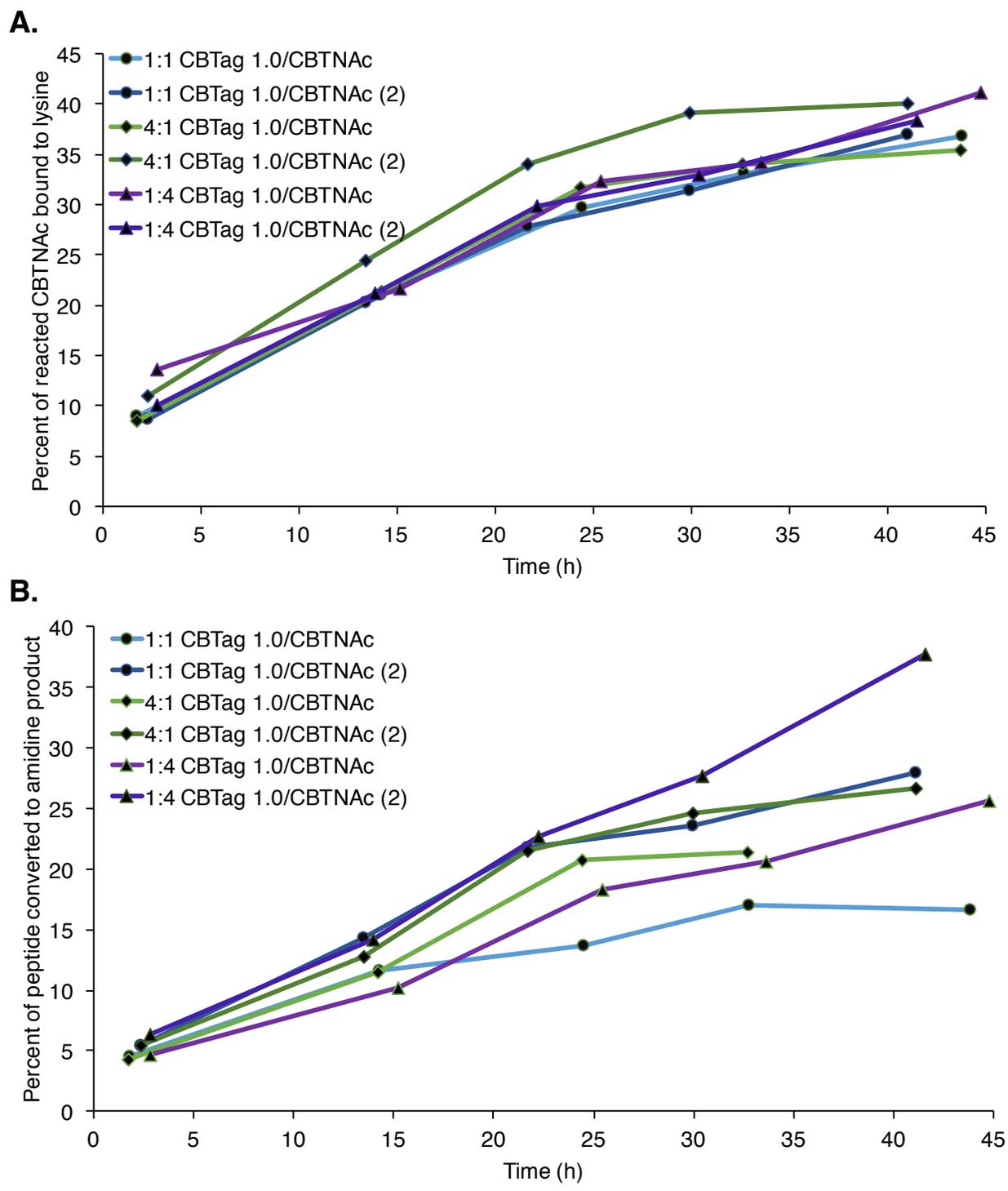


Figure 2.9. LC-MS-based time-course data for the reaction between CBTag 1.0 and CBTNac. **A)** Percent of the CBTNac bound to a peptide that is ligated through an amidine bond over time. **B)** Percent of the peptide that was converted to the amidine product over time. Blue lines and circular markers represent two replicates of the reaction between CBTag 1.0 (22.5 mM) and CBTNac (22.5 mM). Green lines and diamond markers

represent two replicates of the reaction between CBTag 1.0 (22.5 mM) and CBTNAc (5.63 mM). The last time point of the first replicate is not shown because an instrument error during injection resulted in overall lower-than-expected signal. Purple lines and triangular markers represent two replicates of the reaction between CBTag 1.0 (5.63 mM) and CBTNAc (22.5 mM). Spreadsheets containing calculations, LC-MS spectra, and LC-MS integration results are in Supplementary Table 2.S2-2.S4 and Appendix 1.

As an additional confirmation that transfer from cysteine to lysine is intramolecular and not intermolecular, a 1:1 mixture of CBTag 1.0 (K9R) (**22**) and CBTag 1.0 (C7S) (**23**) was subjected to our initial MALDI-TOF MS assay conditions (Figure 2.10). Since CBTag 1.0 (K9R) has a cysteine but not a lysine and reacts with one equivalent of CBTNAc prior to cysteine treatment, while CBTag 1.0 (C7S) has a lysine but not a cysteine and does not react with CBTNAc at all, any formation of a CBTag (C7S)-CBTNAc adduct would have to be a result of intermolecular cysteine-lysine relay. While a mass corresponding to formation of the CBTag 1.0 (K9R)-CBTNAc adduct was observed prior to cysteine treatment, no mass corresponding to the CBTag 1.0 (C7S)-CBTNAc adduct appeared. Moreover, treatment with cysteine reversed the CBTag 1.0 (K9R)-CBTNAc adduct, resulting in a MALDI-TOF-MS spectrum that showed only the unreacted peptides. These data and the LC-MS data indicating that rate of formation of the CBTag 1.0-CBTNAc lysine adduct is independent of peptide concentration suggest an intramolecular cysteine-lysine transfer mechanism.

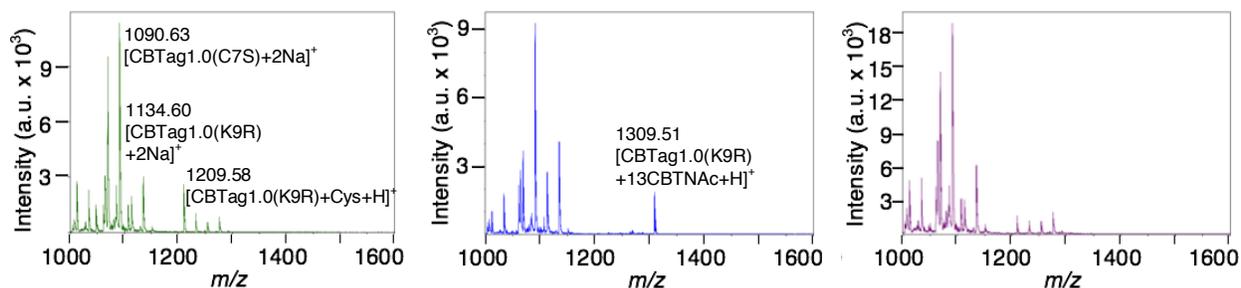


Figure 2.10. ^{13}C CBTNAc binding assay and cysteine treatment results for a 1:1 mixture of CBTag 1.0 (K9R) and CBTag 1.0 (C7S). The binding assay and cysteine competition were as follows: 40 μL of a 10% DMF in PBS solution of 1:1 CBTag 1.0 (K9R)/CBTag 1.0 (C7S) (0.9 mM total peptide) with or without ^{13}C CBTNAc (1 mM) was incubated 48 h at r.t. Then, the negative control (without ^{13}C CBTNAc) and half (20 μL) of the sample incubated with ^{13}C CBTNAc were each treated with an equal volume (20 μL) of a solution of cysteine and DTT (10 and 15 mM, respectively) in PBS. The other half of the sample incubated with ^{13}C CBTNAc was treated with an equal volume of a solution of DTT (15 mM) in PBS. After 48 h at r.t., MALDI-TOF MS was used to analyze the negative control (left column, green spectra), the sample incubated with ^{13}C CBTNAc then DTT alone (middle column, blue spectra), and the sample incubated with ^{13}C CBTNAc then cysteine and DTT (right column, purple spectra). The CBTag 1.0 (K9R)- ^{13}C CBTNAc adduct disappears on

cysteine treatment; no CBTAg 1.0 (C7S)-13CBTNAc adduct is apparent under any of these conditions.

Together, our results support the hypothesis that the nucleophilic cysteine thiol rapidly reacts with the CBTNAc to form a thioimidate adduct, which is slowly transferred to the lysine, creating a stable amidine bond (Figure 2.11). To further rationalize why the tetrahedral intermediate proceeds to the amidine product instead of the product of the condensation reaction (as in luciferin), we used Maestro software to build both products and the integrated program MacroModel to search for the lowest energy conformations. The optimized conformation of the amidine product had a significantly lower potential energy than that of the product of the condensation reaction, even taking into account the released ammonia molecule (-488 kcal/mol and -456 kcal/mol, respectively; a difference of 22 kcal/mol) (Figure 2.12). These results are not unexpected, as the entropic gain from the release of ammonia is unlikely to compensate for the entropic loss associated with the increased rigidity of an already large macrocycle, especially in the absence of conditions favoring the loss of ammonia. Additionally, the transition state for the amidine product is likely lower in energy relative to that of the condensation product. In both transition states, a partial positive charge builds up on a nitrogen atom, but in the transition state for the amidine product, the partial negative charge is on the cysteine sulfur atom, whereas in the transition state for the condensation product, the partial negative charge is on a nitrogen atom. Sulfur atoms are more polarizable than nitrogen atoms and are better able to stabilize a negative charge; thus, the amidine transition state is stabilized more than the condensation reaction transition state. In sum, both the transition state and the product energies favor the amidine product.

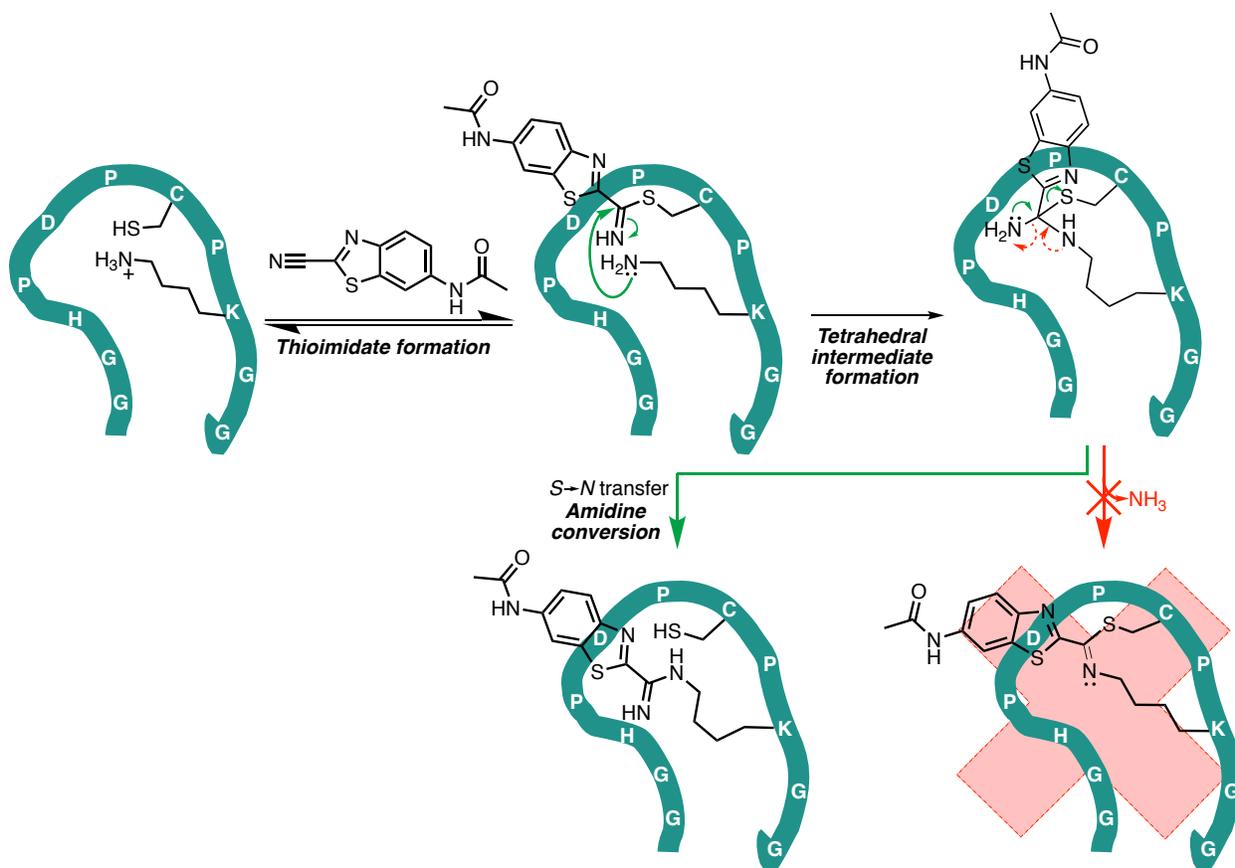


Figure 2.11. Proposed mechanism of reaction between CBTag 1.0 and CBTNac. First, the nucleophilic thiol on the cysteine residue reversibly forms a thioimidate bond with CBTNac, likely catalyzed by the imidazole of the histidine residue.²⁸⁻²⁹ Then, the lysine amine attacks the thioimidate, resulting in a tetrahedral intermediate. Unlike the reaction to make firefly luciferin, a condensation step that releases ammonia does not proceed (red arrows). Instead, the end result is a stable amidine bond between the CBTNac and the lysine residue in a “cysteine hand-off” mechanism analogous to that of NCL.

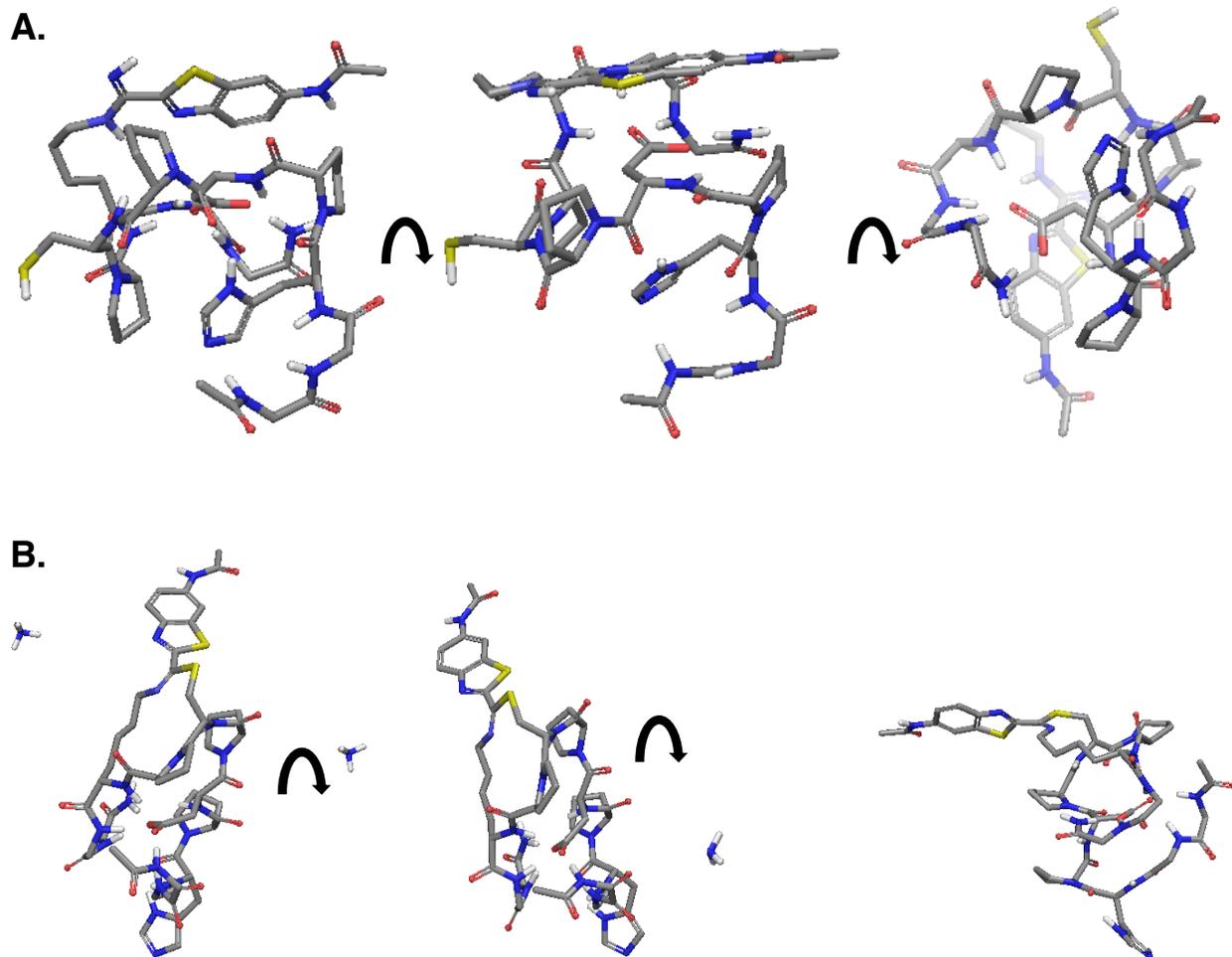


Figure 2.12. Lowest energy conformation of each potential final product (Maestro, Macromodel). **A.** Rotated views of the amidine product (observed). The energy of this conformer was found to be -488 kcal/mol. **B.** Rotated views of the condensation reaction product, including ammonia (not observed). The energy of this conformer was found to be -456 kcal/mol.

CONCLUSIONS

In conclusion, we have designed and characterized a peptide that forms an irreversible amidine bond between a uniquely reactive lysine residue and a CBT derivative. We have demonstrated that similar to NCL, this mechanism occurs through transfer of the electrophilic group from a cysteine thiol to a lysine amine. These results are a promising first step toward a novel peptide tag for site-specific lysine functionalization, though significant optimization remains to be done before this system can be implemented as a practical tool. We are currently exploring mRNA display-based methods³⁰ to select for an improved version of CBTag 1.0 as well as computational methods to engineer a peptide sequence that strongly favors the reactive conformation of the cysteine and lysine residues. Regardless, in using concepts that invoke themes from enzymology to bioorthogonal chemistry, we provide here a potential template for rational design of

peptide tags with engineered “hot” lysines that can undergo site-selective covalent modification.

MATERIALS AND METHODS

General. Reactions were performed in flame- or oven-dried glassware under an inert nitrogen atmosphere unless otherwise noted. Anhydrous solvents were either purchased or obtained by passing solvent through an activated alumina column via a Pure Process Technology Glass Contour Solvent Purification System. All reagents and solvents were used as received unless otherwise noted. Water was passed through a Milli-Q filtration system prior to use. Where noted, samples were concentrated *in vacuo* at 40 °C using a BÜCHI Rotavapor R-114 equipped with a BÜCHI B-480 heating bath and a Welch Self-Cleaning Dry Vacuum System (Model 2025) or an IKA RV 10 basic rotary evaporator equipped with an IKA HB 10 basic heating bath and a Welch Self-Cleaning Dry Vacuum System (Model 2025). If necessary, compounds were then further dried under high vacuum using an Edwards RV8 Two Stage Rotary Vane Pump or by lyophilization in a LABCONCO FreeZone 4.5Plus.

Thin layer chromatography was performed using SiliCycle SiliaPlate glass-backed silica gel plates containing a fluorescent indicator (Fisher Scientific 50964470). Plates were visualized using a UVGL-25 Compact UV Lamp, 254/365 nm, 4 W (P/N 95-0021-12). For flash column chromatography, the stationary phase was SiliCycle SiliaFlash P60 or Fisher Silica Gel Sorbent (230-400 Mesh, Grade 60) silica gel. For purifications involving preparative or semi-preparative Reversed-Phase High Performance Liquid Chromatography (RP-HPLC), the following conditions were used: the instrument consisted of an Agilent Technologies ProStar 325 UV-Vis detector, two PrepStar Solvent Delivery Modules, and a 440-LC Fraction Collector; the column was either a Varian Microsorb 100Å C18, 8 µm, 21.4 x 250 mm Dynamax preparative column (R0080220C8) equipped with a Microsorb 100Å C18, 8 µm guard column (R0080220G8) or an Agilent ZORBAX Eclipse XDB 80Å C18, 5 µm, 9.4 x 250 mm semi-preparative column (990967-202). Solvent A was 0.1% TFA in Milli-Q water; solvent B was 0.1% TFA in acetonitrile (MeCN). The UV-Vis detector was used to monitor wavelengths at 210 and 254 nm. All pure compounds and stock solutions were stored at -20 °C.

Liquid Chromatography-Mass Spectrometry (LC-MS) experiments were performed on an Agilent Technologies 1260 Infinity attached to a 6120 Quadrupole MS and a Peak Scientific NM32LA nitrogen generator. An Agilent InfinityLab Poroshell 120 EC-C18, 2.7 µm, 4.6 x 50 mm analytical LC column (part number 699975-902) was used. As with the HPLC, solvent A was 0.1% TFA in Milli-Q water; solvent B was 0.1% TFA in MeCN. Wavelengths of 210 nm, 254 nm, and 280 nm were monitored using the diode array detector. For ESI-MS, the 105-2000 *m/z* range was monitored in both positive and negative ion mode. Data was processed and analyzed using LC/MSD ChemStation (Agilent Technologies, Rev. B.04.03[16]). High resolution mass spectrometry (HRMS) data were acquired by ESI-LC/MS on a Waters Acquity UPLC and Thermo Exactive

Orbitrap mass spectrometer by Dr. Theresa McLaughlin at the Stanford University Mass Spectrometry facility.

MALDI-TOF MS spectra were obtained on either an Applied Biosystems Voyager-DE PRO BioSpectrometry Workstation using an Applied Biosystems Sample Plate SS, Numbers & Circles (P/N V700666) and the Applied Biosystems Voyager Instrument Control Panel v5.10 software (QB3/Chemistry Mass Spectrometry Facility, UC Berkeley) or a Bruker microFlex MALDI-TOF (S/N 256969.00028) using a microScout Target MSP 96 target polished steel BC plate (P/N 8280800) and the Bruker Daltonics FlexControl software (Vincent Coates Foundation Mass Spectrometry Laboratory, Stanford University). In all cases, the matrix was α -cyano-4-hydroxycinnamic acid (CHCA) dissolved in 50% MeCN in 0.1% TFA. Plates were spotted with a mixture of 0.5 μ L sample and 0.5 μ L matrix. For the Voyager, parameters were as follows: reflectron mode, positive ion detection; pulsed nitrogen laser (337 nm, 20 Hz); accelerating voltage 20 kV, grid voltage 94%, guide wire voltage 0.05%, delay time 270 nsec; mass range 200-2500 Da, low mass gate 200 Da; spectrum acquisition – manual control, 1500 shots/spectrum; manual laser intensity ~1500-2500. For the Bruker microFlex, the parameters were as follows: reflectron mode, positive ion detection; pulsed nitrogen laser (337 nm, 3 ns pulse width, pulse energy 150 μ J, 60 Hz); detector gain (reflector) 4.0x, 1810 V; sample rate 2.00 GS/s; laser power percentage was adjusted on a sample-by-sample basis to obtain an arbitrary intensity of 10^3 - 10^4 and a flat baseline; mass range 0-4000 Da; spectrum acquisition – random walk partial sample mode, results were the sum of at least 3 x 100 laser shots; calibration of the system was performed using the Bruker Peptide Calibration Standard mixture (part number 8206195). Spectra were analyzed using the open-source software mMass (version 5.5.0)³¹⁻³³ after data had been exported as an ASCII file from Voyager (Applied Biosystems) or FlexAnalysis (Bruker Daltonics).

Nuclear magnetic resonance (NMR) spectra were obtained on a Bruker AVIII HD 400 MHz NMR spectrometer (UC Berkeley), a Varian Inova 500 (Stanford University), or an Agilent VNMRs 800 MHz NMR spectrometer (Stanford University). The 800 MHz NMR spectra were obtained by Dr. Corey Liu and Dr. Stephen Lynch of the Stanford University Department of Chemistry NMR Facility. Spectra were processed using MestReNova v10.0.2-15465 (Mestrelab Research S.L., 2015).

PEP-FOLD screening. PEP-FOLD²⁶⁻²⁷ was used for initial *in silico* screening of peptide sequences (Supplementary Table 2.S1). MacPyMOL Molecular Graphics System v1.4 (Schrödinger, LLC, New York, NY, 2010) was used to visualize output.

Binding assay (adapted from reference 22). N-terminal acetylated, C-terminal amidated peptides (36 μ L of a 1 mM stock solution, pH adjusted to approximately 7) were incubated with CBTNAc (0.4 or 4 μ L of a 100 mM stock solution in DMF) or DMF alone (negative control) in PBS for 48 h at room temperature. An aliquot (0.5 μ L) of each reaction mixture was analyzed by MALDI-TOF MS either without prior purification or after desalting with a μ -C18 ZipTip[®].

Cysteine competition assay. After MALDI-TOF MS analysis, half (20 μ L) of the binding assay solution was incubated for 48 h at room temperature after adding an equal volume of cysteine (10 mM or 10 equiv. relative to CBTNAC, unless cysteine would interfere with a later assay, in which case 1 mM or 1.1 equiv. was used; pH adjusted to approximately 7) with or without DTT or TCEP reducing agent (10-15 mM or 10-15 equiv., unless the reducing agent would interfere with a later assay, in which case 1 mM or 1.1 equiv. was used; pH adjusted to approximately 7). The remainder of the binding assay solution was incubated with an equal volume of PBS or an equal volume of DTT or TCEP reducing agent (10-15 mM or 10-15 equiv., unless the reducing agent would interfere with a later assay, in which case 1 mM or 1.1 equiv. was used; pH adjusted to approximately 7). The reaction mixtures were again analyzed by MALDI-TOF MS either without prior purification or after desalting with a μ -C18 ZipTip[®].

Binding assay in the presence of glutathione. CBTtag 1.0 (18 μ L of a 1 mM stock solution in PBS) was incubated with reduced glutathione (0.2 μ L of a 500 mM stock solution in PBS) and CBTNAC (2 μ L of a 10 mM stock solution in DMF) or DMF alone (negative control) in PBS for 48 h at room temperature. After completion, an equal volume (20 μ L) of a solution of cysteine and DTT (10 and 15 mM, respectively) in PBS was added, and the resulting solution was incubated a further 48 h at room temperature. An aliquot (0.5 μ L) of each reaction mixture was analyzed by MALDI-TOF MS without prior purification.

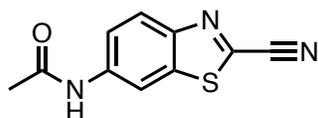
Chloroacetate labeling of cysteine residues. Peptide solution (0.9 mM; 0.45 mM if after treatment with a slight excess of cysteine) was incubated with an equal volume of 30 or 500 mM chloroacetate (pH adjusted to approximately 7) overnight at room temperature. The reaction mixtures were analyzed by MALDI-TOF MS without prior purification.

Tandem LC-MS/MS. After the cysteine competition assay (using 1.1 equiv. cysteine), samples were diluted 1:9 with PBS, then 1:4 with water to give final concentrations (where applicable) of 10 μ M peptide, 10 μ M CBTNAC, 10 μ M cysteine, 15 μ M DTT. Dr. Theresa McLaughlin of the Stanford University Mass Spectrometry facility analyzed the samples by ESI-MS on an Agilent 1260 HPLC connected to a Bruker MicroTOF-QII mass spectrometer and equipped with an Agilent Pursuit 5 diphenyl 2.1 x 150 mm LC column held at 60 °C. The injection volume was 10 μ L, the flow rate was 0.3 mL/min. and data was collected in autoMS2 mode with a mass range of 50-2000 Da. Solvent A was 0.1% formic acid (FA) in water, solvent B was 0.1% FA in MeCN, and the gradient was as follows: 5% B for 2 minutes, 5-95% B over 5 minutes, 95% B for 1 minute, 95-5% B over 0.1 minutes, and 5% B for 6.9 minutes. Predicted fragment ions were calculated using the MS/MS Fragment Ion Calculator in the Proteomics Toolkit provided by the Institute for Systems Biology (<http://db.systemsbiology.net/proteomicsToolkit/FragIonServlet.html>).

LC-MS time-course experiment. Samples were prepared as follows: In an LC-MS vial equipped with a 300- μ L insert, a solution of 22.5 or 5.63 mM peptide (CBTag 1.0, CBTag 1.0 (K9R), or CBTag 1.0 (C7S); adjusted to pH 7 in PBS), 2 mM TCEP (adjusted to pH 7 in water), and 22.5 or 5.63 mM 13CBTNAc (**3**) was prepared in 5-6% DMF in PBS (total volume of 10 μ L) and the starting time noted. Ratios of 1:1, 1:4, and 4:1 peptide/13CBTNAc were assessed. An LC-MS sequence was prepared such that each sample was injected five times over approximately 45 hours, with an injection volume of 1 μ L each time. The LC-MS gradient was: 5% solvent B for 2 minutes, 5-75% solvent B over 15 minutes, 75-100% solvent B over 1 minute, 100% solvent B for 6 minutes, 100-5% solvent B over 1 minute, and 5% solvent B for 3 minutes, where solvent A was 0.1% TFA in water and solvent B was 0.1% TFA in acetonitrile. Identification of peaks was performed by comparison to the spectra of pure compounds, determining whether predicted masses were present, and checking whether addition of an equal volume of 200 mM cysteine (adjusted to pH 7 in water) caused the disappearance of the peak over time. The relevant peaks in the 210 nm and 254 nm chromatographs for two replicates were integrated manually and used to calculate degree of 13CBTNAc modification over time. See Supplementary Figure 2.S2-2.S6, Supplementary Table 2.S2-2.S4, and Appendix 1 for details.

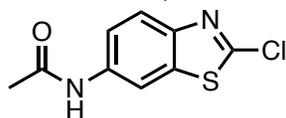
Computational determination of lowest energy conformer for amidine and condensation reaction products. For conformational searching and molecular dynamics calculations, Maestro (Version 10.7.015, MMshare Version 3.5.015, Release 2016-3, Platform Darwin-x86_64, Schrödinger, LLC, New York, NY, 2016) and the integrated program Macromodel (Schrödinger, LLC, New York, NY, 2016) were used. In general, standard parameters were applied. For calculating potential energy, an OPLS3 force field with water as the solvent was used. Structures were minimized using the Polak-Ribiere Conjugate Gradient (PRCG) method until a gradient below the convergence threshold of 0.05 kJ/ \AA -mol was reached, with a maximum of 2500 iterations. For conformational searching, a mixed torsional/low-mode sampling method (multi-ligand) was used, with 1000 steps maximum, 100 steps per rotatable bond, and the energy window for saving structures set to 5.02 kcal/mol. For molecular dynamics simulations, a stochastic dynamics method was used, the simulation temperature was 300.0 K, the time step was 1.5 fs, the equilibration time was 1.0 ps, and the simulation time was 10 ns. Each structure was subjected to five rounds of conformational searching, with molecular dynamics simulated after rounds 2, 3, and 4. MacPyMOL Molecular Graphics System v1.4 (Schrödinger, LLC, New York, NY, 2010) was used to save images of the lowest energy conformer for the amidine product and the condensation reaction product (Figure 2.12).

Synthesis and characterization (small molecules, peptides)



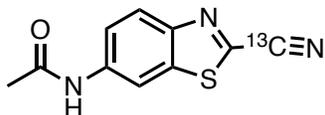
***N*-(2-cyanobenzo[*d*]thiazol-6-yl)acetamide (CBTNAc, 1).** Synthetic procedures were adapted from reference 22; NMR spectra (methanol-*d*₄ and chloroform-*d*) were reported in references 22 and 29. 6-Amino-2-cyanobenzothiazole (12 mg, 0.069 mmol) was suspended in anhydrous DCM (0.60 mL). Pyridine (3 drops) and acetyl chloride (0.020 mL, 0.28 mmol, 4.8 equiv.) were added, and the reaction was allowed to stir 1 h at room temperature. The resulting clear light yellow-orange solution was concentrated *in vacuo* then resuspended in ethyl acetate (5.0 mL) and aqueous saturated sodium bicarbonate (5.0 mL). The ethyl acetate layer was separated and concentrated *in vacuo* to a yellow-white solid. The crude product was purified by silica gel chromatography (1:5 ethyl acetate/hexanes to 100% ethyl acetate). By TLC (1:1 ethyl acetate/hexanes), the product was visible under shortwave UV light as a dark bright blue spot at *R*_f 0.145. Product-containing fractions were combined and concentrated *in vacuo* then lyophilized to yield 7.87 mg (52.5%) of an off-white powder. LC-MS indicated ~96% purity.

¹H NMR (500 MHz, Acetone-*d*₆) δ 9.68 (br s, 1H), 8.81 (d, *J* = 2.0 Hz, 1H), 8.12 (d, *J* = 9.0 Hz, 1H), 7.72 (dd, *J* = 9.0, 2.0 Hz, 1H), 2.15 (s, 3H). ¹³C{¹H} NMR (126 MHz, Acetone-*d*₆) δ 169.6, 149.0, 141.2, 137.9, 135.6, 125.8, 121.3, 114.1, 111.4, 24.4. HRMS (ESI) *m/z*: [M + H]⁺ Calcd for C₁₀H₈N₃OS 218.0383; Found 218.0381.



***N*-(2-chlorobenzo[*d*]thiazol-6-yl)acetamide (ClBTNAc, 2).** Synthetic procedures were adapted from reference 33. 6-amino-2-chlorobenzothiazole (0.66 g, 3.57 mmol) was dissolved in acetic anhydride (6.5 mL, 68.8 mmol, 19.2 equiv.). Pyridine (1.5 mL, 19.3 mmol, 5.2 equiv.) was added dropwise, then the light yellow-brown solution was slowly raised to boiling point (approx. 140 °C) in an oil bath. After heating for 5 minutes, the dark orange-brown solution was cooled, poured into water (20 mL), and extracted three times with 20-mL aliquots of ethyl acetate. The combined bright yellow ethyl acetate layers were washed twice with 20-mL aliquots of brine, dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo* to an orange-brown oil. The crude product was purified by silica gel chromatography, eluting with 2.5 column volumes of 40% ethyl acetate in hexanes then 5 column volumes of 50% ethyl acetate in hexanes. By TLC (1:1 ethyl acetate/hexanes) visualized with shortwave UV, three major spots eluted. By NMR peaks and mass spectrometry, these were identified as starting material (*R*_f 0.818), product (*R*_f 0.515), and starting material with the chlorine atom replaced by a hydrogen atom (*R*_f 0.318). Removal of the solvent *in vacuo* followed by drying under high vacuum overnight yielded recovered starting material as a pale orange-yellow powder (0.186 g, ~28.2% unreacted) and product as a pale yellow-tan powder (0.364 g, 45% yield, 76.8% yield based on recovered starting material). LC-MS indicated 96-98% purity.

^1H NMR (400 MHz, Methanol- d_4) δ 8.32 (d, $J = 2.2$ Hz, 1H), 7.75 (d, $J = 8.8$ Hz, 1H), 7.45 (dd, $J = 8.8, 2.2$ Hz, 1H), 2.15 (s, 3H). $^{13}\text{C}\{^1\text{H}\}$ NMR (101 MHz, Methanol- d_4) δ 171.7, 153.3, 148.3, 138.3, 137.9, 123.4, 120.6, 112.9, 23.9. HRMS (ESI) m/z : $[\text{M} + \text{H}]^+$ Calcd for $\text{C}_9\text{H}_8\text{CN}_2\text{OS}$ 227.0040; Found 227.0038.



N-(2-(cyano- ^{13}C)-benzo[*d*]thiazol-6-yl)acetamide (**13CBTNAc**, **3**). Synthetic procedures were adapted from reference 34; NMR spectra (chloroform-*d*) of the compound (not enriched for ^{13}C) were reported in references 22 and 35. [**CAUTION:** Potassium cyanide is extremely hazardous; compounds dissolved in DMSO are more easily absorbed by skin. Use appropriate safety measures.] Potassium cyanide- ^{13}C purchased from Cambridge Isotope Laboratories (48.3 mg, 0.71 mmol, 1.19 equiv.) in anhydrous DMSO (12.5 mL) was heated at 130 °C in an oil bath until it was mostly dissolved into a clear solution (15 minutes). CIBTNAc (**2**) (139.4 mg, 0.615 mmol) was added, and the resulting dark orange reaction mixture was stirred 2 h at 130 °C, becoming bright red over time. The solution was allowed to stand at room temperature overnight, turning a deep red-orange on cooling. The next day, the reaction mixture was poured into a separatory funnel and 75 mL of 0.2 M monopotassium phosphate (KH_2PO_4 , pH 4-5) was added. This aqueous solution was extracted once with 125 mL of diethyl ether and twice with 50-mL aliquots of diethyl ether. The combined neon yellow ether extracts were washed twice with 25-mL aliquots of water then dried over anhydrous sodium sulfate. The solution was filtered, silica gel (1.1 g) was added, and the solvent was removed *in vacuo*. The dry residue was added to a column of silica gel (100 mL) pre-packed with petroleum ether, then product was eluted using 0-50% ethyl acetate in petroleum ether followed by 50-100% ethyl acetate in hexanes. Pure fractions were identified by TLC (1:1 ethyl acetate/hexanes), where the product was visible as a dark blue spot (R_f 0.26) under short- and longwave UV light. The impurity (R_f 0.31) eluted just before the product. Concentration of the product-containing fractions *in vacuo* followed by drying under high vacuum overnight yielded 74 mg (55.1%) of product as a white to pale yellow solid. Product was ~97% pure by LC-MS.

^1H NMR (500 MHz, Acetone- d_6) δ 9.68 (br s, 1H), 8.81 (d, $J = 2.1$ Hz, 1H), 8.12 (d, $J = 9.0$ Hz, 1H), 7.72 (dd, $J = 8.9, 2.1$ Hz, 1H), 2.16 (s, 3H). $^{13}\text{C}\{^1\text{H}\}$ NMR (126 MHz, Acetone- d_6) δ 169.6, 149.0, 141.1, 137.9, 135.1, 125.8, 121.3, 114.1, 111.4, 24.4. HRMS (ESI) m/z : $[\text{M} + \text{H}]^+$ Calcd for $\text{C}_9^{13}\text{CH}_8\text{N}_3\text{OS}$ 219.0416; Found 219.0414.

General peptide synthesis conditions. Peptides were made by solid-phase peptide synthesis (SPPS) either manually in a peptide synthesis vessel or using an AAPPTec Apex 396 automated peptide synthesizer equipped with a 96-well reactor. Solvents were HPLC grade or peptide synthesis grade; neither solvents nor reagents were anhydrous. Rink Amide AM resin (100-200 mesh, 0.41 mmol/g loading) from EMD Millipore (855130) was used as the solid phase unless otherwise noted. All coupling, deprotection, and cleavage steps were performed at ambient temperature.

General procedure for manual synthesis. A 50 mL peptide synthesis vessel (VWR 80071-382) was equipped with a house nitrogen and a vacuum line. Rink Amide AM resin (0.05 g, 0.02 mmol) was added and allowed to swell for 10 minutes in either dichloromethane (DCM) or *N,N*-dimethylformamide (DMF) without nitrogen agitation. The resin was drained and dried with several cycles of vacuum and nitrogen, then the resin Fmoc protecting group was cleaved with 20% piperidine in DMF or *N*-methyl-2-pyrrolidone (NMP) (5-10 mL) for 15-20 minutes with nitrogen agitation, rinsing the sides every 5-10 minutes to ensure all beads were deprotected. Beads were drained and dried with several rounds of vacuum and nitrogen, then a Kaiser test³⁶ was performed on a small amount of resin to ascertain degree of completion. If deprotection was complete, the first amino acid was coupled with monomer (0.2 mmol, 10 equiv.), coupling reagent (0.2 mmol, 10 equiv. of HATU for Fmoc-His(Trt) coupling, COMU for all others), and *N,N*-diisopropylethylamine (DIPEA; 15 equiv. for Fmoc-Cys(Trt) coupling, 20 equiv. for all others) in NMP (1.5 mL) or 1:1 DCM/NMP (1.5 mL; Fmoc-Cys(Trt) coupling only) with nitrogen agitation. Directly before coupling, the DIPEA was added to the monomer/coupling reagent solution and the vial gently inverted for 30-60 seconds before addition to resin. Again, sides of the vessel were rinsed with minimal DCM or DMF 5-10 minutes into the reaction. After 15-30 minutes, the resin was drained and washed thoroughly with DCM, then the first coupling reaction was repeated. The Kaiser test was used to confirm completion, then any remaining free amines on the resin were capped with a solution of 10% acetic anhydride and 5% DIPEA in DMF or NMP, agitating with nitrogen for 15-20 minutes. After draining and washing the resin, the Fmoc group on the first amino acid was removed with 20% piperidine in DMF or NMP as previously. Subsequent couplings were performed only once unless the amino acid was β -branched or the Kaiser test did not show completion. After the coupling and deprotection of the final amino acid, the N-terminus was acetylated using a solution of 10% acetic anhydride and 5% DIPEA in DMF or NMP, agitating with nitrogen for 15-20 minutes. In preparation for deprotection and cleavage, the resin was thoroughly washed with DCM and left under vacuum until dry (15-30 minutes). A clean roundbottom flask was attached to the synthesis vessel and the cleavage cocktail Reagent L (for 0.02 mmol resin: 0.96 g dithiothreitol, 0.04 mL triisopropylsilane, 0.1 mL ddH₂O, and trifluoroacetic acid (TFA) to a final volume of 2 mL) was added to cleave and deprotect the peptide. The resin was allowed to sit 1.5 h without nitrogen agitation; the mixture was manually stirred every 30 minutes with a glass stir rod. After draining the cleavage solution into the clean roundbottom flask, the solvent was removed *in vacuo* and the resulting viscous cloudy pale yellow to pale white solution was pipetted into ice-cold ether (45 mL). The flask was rinsed with minimal TFA and the resulting solution was also added to the ether. The suspension was centrifuged in a Sorvall Legend RT at 3700 x g and 4°C for 8-10 minutes. The ether was decanted carefully, and the remaining pale yellow to off-white precipitate was washed twice more with ice-cold ether (45 mL). After the final centrifugation, the ether was carefully decanted and the precipitate allowed to dry 15-30 minutes in a hood to remove any remaining ether. Finally, the precipitate was resuspended in minimal ddH₂O and lyophilized in a pre-weighed vial using a LABCONCO FreeZone 4.5Plus. Crude peptide was stored at -20 °C until purification.

General procedure for automated peptide synthesis. Rink Amide AM resin (0.05 g, 0.02 mmol) was added to a well in the AAPPTec Apex 396 automated peptide synthesizer. The instrument was programmed to complete the following steps: NMP (1 mL) was added and the resin was mixed for 3 minutes at 600 rpm. After draining the solvent, this step was repeated. To deprotect the first Fmoc group, 20% piperidine in DMF (1 mL) was added and the resin mixed for 3 minutes at 600 rpm. After draining, another aliquot of 20% piperidine in DMF (1 mL) was added and deprotection was allowed to proceed for 20 minutes at 600 rpm. After the deprotection solution was drained, the resin was washed three times with NMP (1 mL), MeOH (1 mL), and NMP (1 mL) at 600 rpm for 3 minutes each. Then, 0.4 M protected monomer in NMP (0.6 mL), 0.4 M HATU or HCTU in NMP (0.5 mL), and 2 M DIPEA in NMP (0.25 mL) were transferred to the well and mixed for 30 minutes at 600 rpm. After draining the solution, the coupling step was repeated, then the resin was washed twice with NMP (1 mL) for 3 minutes at 600 rpm. The deprotection and coupling steps were repeated until the final amino acid was deprotected, then the N-terminus of the peptide was acetylated by twice adding 11.63% acetic anhydride in DMF (1.3 mL) and 2 M DIPEA in DMF (0.25 mL) and mixing for 30 minutes at 600 rpm. Cleavage and precipitation of crude peptide from ether were performed as described above for manual peptide synthesis.

RP-HPLC purification of peptides. Purification was performed by semi-preparative (3 mL/minute flow rate) or preparative RP-HPLC (20 mL/minute flow rate). The general gradient was as follows: 0% solvent B for 2 minutes, 0-33% solvent B over 18 minutes, 33-100% solvent B over 2 minutes, 100% solvent B for 3 minutes. Pure fractions were combined and concentrated *in vacuo*, taken up in water, and lyophilized.

CBTag 1.0, Ac-GGHPDPCPKGG-NH₂ (**4**): Automated synthesis followed by preparative RP-HPLC purification yielded 1.351-4.837 mg (6.2-17%) of a white fluffy solid. Manual synthesis followed by preparative RP-HPLC purification yielded 15.171 mg (68.3%) of a white fluffy solid. Product was 95% pure by LC-MS. ¹H NMR (800 MHz, D₂O) δ 8.60 – 8.25 (m, 8H), 7.52 (s, 1H), 7.29 (s, 1H), 7.09 (s, 1H), 4.41 (s, 2H), 4.30 – 4.25 (m, 1H), 4.01 – 3.66 (m, 15H), 3.56 (q, J = 6.9 Hz, 2H), 3.25 – 2.71 (m, 9H), 2.56 (dd, J = 16.1, 7.7 Hz, 1H), 2.26 (tt, J = 12.6, 6.2 Hz, 3H), 2.06 – 1.72 (m, 17H), 1.67 (s, 3H), 1.53 – 1.39 (m, 3H). ¹³C{¹H} NMR (126 MHz, D₂O) δ 175.2, 174.9, 174.5, 174.3, 174.1, 173.9, 172.5, 172.0, 171.1, 170.5, 170.1, 169.7, 163.3, 163.0, 162.7, 133.7, 128.4, 117.8, 117.5, 115.2, 60.7, 60.6, 54.1, 53.8, 50.7, 48.5, 48.2, 48.0, 42.8, 42.4, 42.3, 39.6, 39.5, 35.4, 30.3, 29.6, 29.5, 26.5, 26.1, 24.9, 24.8, 24.6, 22.2, 21.9. HRMS (ESI) *m/z*: [M + 2H]²⁺ Calcd for C₄₄H₆₉N₁₅O₁₄S 531.7429; Found 531.7436.

PEP-FOLD-4, Ac-GGDPHPCKGG-NH₂ (**5**): Automated synthesis followed by semi-preparative RP-HPLC purification yielded 1.977 mg (7.5%) of a white fluffy solid. Product was 95% pure by LC-MS. HRMS (ESI) *m/z*: [M + 2H]²⁺ Calcd for C₃₉H₆₂N₁₄O₁₃S₂ 483.2165; Found 483.2170.

PEP-FOLD-1, Ac-GGDPCPHPKGG-NH₂ (**6**): Automated synthesis (0.06 mmol scale) followed by preparative RP-HPLC purification yielded 0.919 mg (1.4%) of a white fluffy

solid. Product was 95% pure by LC-MS. HRMS (ESI) m/z : $[M + 2H]^{2+}$ Calcd for $C_{44}H_{69}N_{15}O_{14}S$ 531.7429; Found 531.7428.

PEP-FOLD-2, Ac-GGHPDCPKGG-NH₂ (**7**): Automated synthesis (0.06 mmol scale) combined with manual synthesis (0.02 mmol scale) followed by preparative RP-HPLC purification yielded 1.503 mg (1.9%) of a white fluffy solid. Product was 93.5% pure by LC-MS. As the ratio of labeled to unlabeled peptide after cysteine treatment was not competitive with CBTag 1.0 (Figure 2.S1B), no further purification was performed. HRMS (ESI) m/z : $[M + 2H]^{2+}$ Calcd for $C_{39}H_{62}N_{14}O_{13}S$ 483.2165; Found 483.2162.

PEP-FOLD-5, Ac-GGHPEPCPKGG-NH₂ (**8**): Automated synthesis (0.06 mmol scale) followed by preparative RP-HPLC purification yielded 1.774 mg (2.7%) of a white fluffy solid. Product was 95% pure by LC-MS. HRMS (ESI) m/z : $[M + 2H]^{2+}$ Calcd for $C_{44}H_{69}N_{15}O_{14}S$ 538.7507; Found 538.7510.

PEP-FOLD-6, Ac-GGHPECPKGG-NH₂ (**9**): Automated synthesis (0.06 mmol scale) followed by preparative RP-HPLC purification yielded 33.860 mg (57.6%) of a white fluffy solid. Product was 92% pure by LC-MS. As the ratio of labeled to unlabeled peptide after cysteine treatment was not competitive with CBTag 1.0 (Figure 2.S1F), no further purification was performed. HRMS (ESI) m/z : $[M + 2H]^{2+}$ Calcd for $C_{40}H_{64}N_{14}O_{12}S$ 490.2244; Found 490.2246.

PEP-FOLD-7, Ac-GGEPHPCPKGG-NH₂ (**10**): Automated synthesis (0.06 mmol scale) followed by preparative RP-HPLC purification yielded 2.959 mg (4.6%) of a white fluffy solid. Product was 90.9% pure by LC-MS. As the ratio of labeled to unlabeled peptide after cysteine treatment was not competitive with CBTag 1.0 (Figure 2.S1G), no further purification was performed. HRMS (ESI) m/z : $[M + 2H]^{2+}$ Calcd for $C_{45}H_{71}N_{15}O_{14}S$ 538.7507; Found 538.7509.

PEP-FOLD-8, Ac-GGCPKPHPHGG-NH₂ (**11**): Manual synthesis followed by preparative RP-HPLC purification yielded 0.896 mg (4.0%) of a white fluffy solid. Product was 94.3% pure by LC-MS. As the ratio of labeled to unlabeled peptide after cysteine treatment was not competitive with CBTag 1.0 (Figure 2.S1H), no further purification was performed. HRMS (ESI) m/z : $[M + 2H]^{2+}$ Calcd for $C_{46}H_{71}N_{17}O_{12}S$ 542.7589; Found 542.7595.

PEP-FOLD-9, Ac-GGCPKDPHGG-NH₂ (**12**): Manual synthesis followed by preparative RP-HPLC purification yielded 1.537 mg (6.9%) of a white fluffy solid. Product was 95% pure by LC-MS. HRMS (ESI) m/z : $[M + 2H]^{2+}$ Calcd for $C_{44}H_{69}N_{15}O_{14}S$ 531.7429; Found 531.7434.

PEP-FOLD-11, Ac-GGCPDPHPKGG-NH₂ (**13**): Manual synthesis followed by preparative RP-HPLC purification yielded 5.329 mg (24.3%) of a white fluffy solid. Product was 83% pure by LC-MS; major impurity was due to deletion of two glycine residues, which was inseparable from the desired product and accounted for 11.2% of the final

compound mixture. As missing glycines should not significantly affect reactivity, this compound was used under the assumption that it was effectively 94.2% “pure” though the MALDI-TOF spectra would have to be interpreted carefully. As the ratio of labeled to unlabeled peptide after cysteine treatment was not competitive with CBTag 1.0 (Figure 2.S1J), no further purification was attempted. HRMS (ESI) m/z : $[M + 2H]^{2+}$ Calcd for $C_{44}H_{69}N_{15}O_{14}S$ 531.7429; Found 531.7435.

PEP-FOLD-12, Ac-GGKPCPHPDGG-NH₂ (**14**): Manual synthesis followed by preparative RP-HPLC purification yielded 5.153 mg (23.7%) of a white fluffy solid. Product was 96.9% pure by LC-MS. HRMS (ESI) m/z : $[M + 2H]^{2+}$ Calcd for $C_{44}H_{69}N_{15}O_{14}S$ 531.7429; Found 531.7432.

PEP-FOLD-13, Ac-GGWPHPCPKGG-NH₂ (**15**): Manual synthesis followed by preparative RP-HPLC purification yielded 5.141 mg (21.9%) of a white fluffy solid. Product was 95% pure by LC-MS. HRMS (ESI) m/z : $[M + 2H]^{2+}$ Calcd for $C_{51}H_{74}N_{16}O_{12}S$ 567.2691; Found 567.2696.

PEP-FOLD-14, Ac-GGWPHPGCPKGG-NH₂ (**16**): Manual synthesis followed by preparative RP-HPLC purification yielded 9.628 mg (39.1%) of a white fluffy solid. Product was 95% pure by LC-MS. HRMS (ESI) m/z : $[M + 2H]^{2+}$ Calcd for $C_{52}H_{77}N_{17}O_{12}S$ 595.7798; Found 595.7802.

PEP-FOLD-15, Ac-GGKPCPGWPHGG-NH₂ (**17**): Automated synthesis followed by preparative RP-HPLC purification yielded 4.893 mg (20.1%) of a white fluffy solid. Product was 95% pure by LC-MS. HRMS (ESI) m/z : $[M + 2H]^{2+}$ Calcd for $C_{52}H_{77}N_{17}O_{12}S$ 595.7798; Found 595.7805.

PEP-FOLD-16, Ac-GGHPHPKPCPEGG-NH₂ (**18**): Automated synthesis followed by preparative RP-HPLC purification yielded 20.41 mg (76.0%) of a white fluffy solid. Product was 91% pure by LC-MS. As the ratio of labeled to unlabeled peptide after cysteine treatment was not competitive with CBTag 1.0 (Figure 2.S1O), no further purification was performed. HRMS (ESI) m/z : $[M + 2H]^{2+}$ Calcd for $C_{56}H_{85}N_{19}O_{16}S$ 655.8066; Found 655.8071.

PEP-FOLD-17, Ac-GGHPWPKPCPEGG-NH₂ (**19**): Automated synthesis followed by preparative RP-HPLC purification yielded 7.203 mg (25.8%) of a white fluffy solid. Product was 95.6% pure by LC-MS. HRMS (ESI) m/z : $[M + 2H]^{2+}$ Calcd for $C_{61}H_{88}N_{18}O_{16}S$ 680.3168; Found 680.3176.

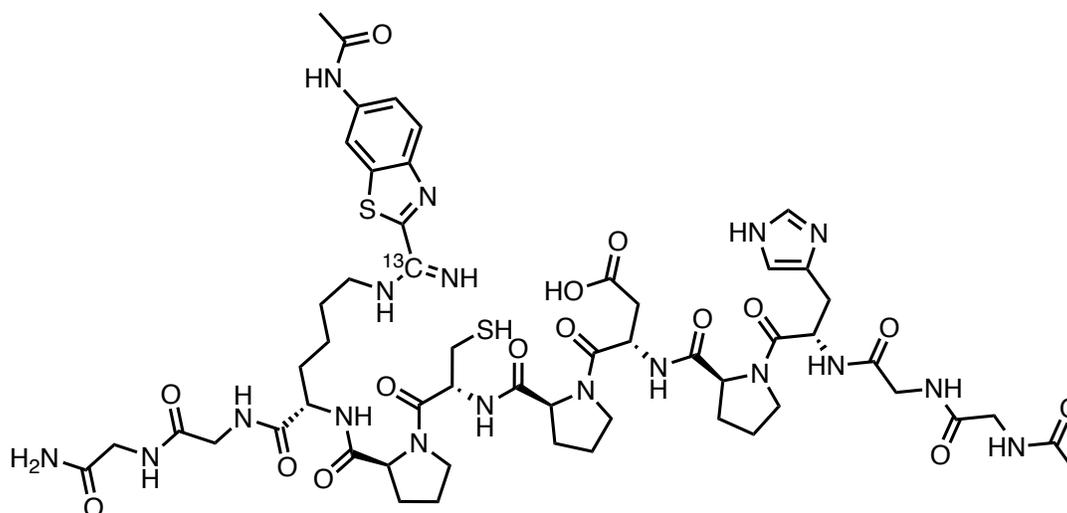
PEP-FOLD-18, Ac-GGHPKPGCGFGG-NH₂ (**20**): Automated synthesis followed by preparative RP-HPLC purification yielded 0.955 mg (4.2%) of a white fluffy solid. Product was 92.5% pure by LC-MS. As the ratio of labeled to unlabeled peptide after cysteine treatment was not competitive with CBTag 1.0 (Figure 2.S1Q), no further purification was

performed. HRMS (ESI) m/z : $[M + 2H]^{2+}$ Calcd for $C_{48}H_{72}N_{16}O_{12}S$ 556.2587; Found 556.2594.

PEP-FOLD-19, Ac-GGHPKPCGFGG-NH₂ (**21**): Automated synthesis followed by preparative RP-HPLC purification yielded 1.109 mg (5.1%) of a white fluffy solid. Product was 90.8% pure by LC-MS; major impurity was due to deletion of one glycine residue, which was inseparable from the desired product and accounted for 5.3% of the final compound mixture. As missing glycines should not significantly affect reactivity, this compound was used under the assumption that it was effectively 96.1% “pure” though the MALDI-TOF spectra would have to be interpreted carefully. As the ratio of labeled to unlabeled peptide after cysteine treatment was not competitive with CBTag 1.0 (Figure 2.S1R), no further purification was attempted. HRMS (ESI) m/z : $[M + 2H]^{2+}$ Calcd for $C_{46}H_{69}N_{15}O_{12}S$ 527.7480; Found 527.7488.

CBTag 1.0 (K9R), Ac-GGHPDPCPRGG-NH₂ (**22**): Manual synthesis followed by preparative RP-HPLC purification yielded 13.894 mg of a white fluffy solid containing two peaks by LC-MS (987 and 1090 m/z). These peaks were assumed to correspond to the $[M + H]^+$ for a cysteine deletion and the product, respectively, based on the masses, the successful reactions of chloroacetate and CBTNAc with the product peptide, and the lack of an observed reaction between the impurity and CBTNAc or chloroacetate. By integration of the LC-MS trace, approximately 45.8% (6.365 mg, 27.9% yield) was desired product. A portion of this sample (8.186 mg) was repurified using the following preparative RP-HPLC method: 10% solvent B for 10 minutes. Combining the >95% pure fractions yielded 1.34 mg (16.4% of 8.186 mg injected mixture; 10% overall yield assuming 16.4%, or 2.274 mg, of the 13.184 mg obtained after the first purification would yield pure product) of a white solid after lyophilization. Product was 97% pure by LC-MS. HRMS (ESI) m/z : $[M + 2H]^{2+}$ Calcd for $C_{44}H_{69}N_{17}O_{14}S$ 545.7460; Found 545.7454.

CBTag 1.0 (C7S), Ac-GGHPDPSPKGG-NH₂ (**23**): Manual synthesis followed by preparative RP-HPLC purification yielded 12.777 mg (58.4%) of a white fluffy solid. Product was 95% pure by LC-MS. HRMS (ESI) m/z : $[M + 2H]^{2+}$ Calcd for $C_{44}H_{69}N_{15}O_{15}$ 523.7543; Found 523.7537.



CBTag 1.0-13CBTNac (amidine) (24): Two methods were used to obtain product.

Method 1: In a 2 mL Eppendorf tube, 10 mM CBTag 1.0 (10.383 mg) and 5 mM 13CBTNac (1.067 mg) in approximately 986 μ L of 5% DMF in PBS was adjusted to approximately pH 7 and allowed to rotate on a Barnstead/Thermolyne Labquake Rotisserie Shaker (Model #C415110) for 40 h at room temperature. At that time, an equal volume of cysteine (10 mM) and DTT (10 mM) adjusted to approximately pH 7 was added to reach final concentrations of 5 mM CBTag 1.0, 2.5 mM 13CBTNac, 5 mM cysteine, and 5 mM DTT in 1.972 mL of 2.5% DMF in PBS. The tube was allowed to rotate a further 48 h at room temperature. The reaction mixture was then injected directly onto a preparative HPLC column and purified using the following method (20 mL/min. flow rate): 0% solvent B for 2 minutes, 0-33% solvent B over 18 minutes, 33-100% solvent B over 2 minutes, 100% solvent B for 3 minutes. Pure fractions were combined and concentrated *in vacuo*, taken up in water, and lyophilized to yield 0.993 mg of a white powder. Assuming 13CBTNac was the limiting reagent and each peptide would only bind one equivalent of 13CBTNac, the maximum theoretical yield would be 6.260 mg. Based on this, the percent yield was roughly 15.9%.

Method 2: In a 2 mL Eppendorf tube, 5 mM CBTag 1.0 (5.220 mg) and 5 mM 13CBTNac (1.073 mg) in 1 mL of 5% DMF in PBS was adjusted to approximately pH 7 and allowed to rotate on a Barnstead/Thermolyne Labquake Rotisserie Shaker (Model #C415110) for 48 h at room temperature. At that time, an equal volume of cysteine (10 mM) and DTT (200 mM) adjusted to approximately pH 7 was added to reach final concentrations of 2.5 mM CBTag 1.0, 2.5 mM 13CBTNac, 5 mM cysteine, and 100 mM DTT in 2 mL of 2.5% DMF in PBS. The tube was allowed to rotate a further 48 h at room temperature. The reaction mixture was then injected directly onto a preparative HPLC column and purified using the following method (20 mL/min. flow rate): 0% solvent B for 2 minutes, 0-33% solvent B over 18 minutes, 33-100% solvent B over 2 minutes, 100% solvent B for 3 minutes. Pure fractions were combined and concentrated *in vacuo*, taken up in water, and lyophilized to yield 1.549 mg of a white powder. Assuming each peptide would only bind one equivalent of 13CBTNac, the maximum theoretical yield would be 6.293 mg. Based on this, the percent yield was roughly 35.8%. Final product was >95% pure by LC-

MS. ^1H NMR (800 MHz, D_2O) δ 10.12 (br s, 1H), 8.57 – 8.16 (m, 7H), 7.61 (d, $J = 8.8$ Hz, 1H), 7.48 (s, 1H), 7.28 (s, 1H), 7.05 (s, 1H), 4.50 – 4.26 (m, 8H), 4.00 – 3.51 (m, 17H), 3.22 (dd, $J = 15.5, 5.3$ Hz, 1H), 3.06 (ddd, $J = 51.4, 14.4, 7.5$ Hz, 2H), 2.88 – 2.64 (m, 8H), 2.51 (dd, $J = 16.1, 8.0$ Hz, 1H), 2.21 (q, $J = 9.3, 8.3$ Hz, 4H), 2.03 (s, 2H), 2.01 – 1.74 (m, 12H), 1.64 – 1.46 (m, 3H), 1.24 (d, $J = 6.6$ Hz, 3H). $^{13}\text{C}\{^1\text{H}\}$ NMR (126 MHz, D_2O) δ 156.0 [run at low concentration to identify ^{13}C -enriched atoms]. HRMS (ESI) m/z : $[\text{2M} + 2\text{H}]^{2+}$ Calcd for $\text{C}_{106} (^{13}\text{C})_2 \text{H}_{148}\text{N}_{36}\text{O}_{30}\text{S}_4$ 1279.5051; Found 1279.5043.

SUPPLEMENTARY FIGURES

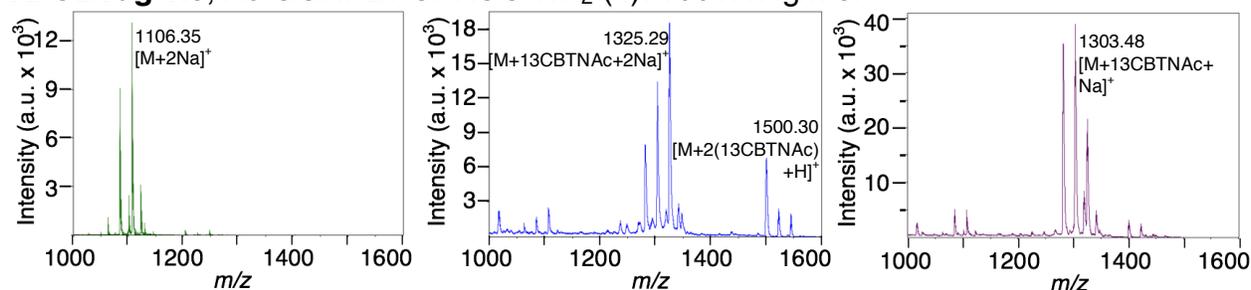
Table 2.S1. List of sequences modeled using PEP-FOLD.²⁵⁻²⁶ Bolded peptides in blue-shaded cells were subsequently made by SPPS.

| | |
|------------------------------|--------------------------|
| AAAAAWGHGDGCKWGWGSGHGWA AAAA | AAAAAWGDGCKWGWGSGHGAAAAA |
| AAAAAHDGCKGDHAAAAA | PWGDGCKWGWGSGHGWP |
| PGHDCKDHGP | GGGGGGHCDKHGGGGGGGG |
| GGGGGGHCDDKHGGGGGGGG | GGGGGGHCGDKHGGGGGGGG |
| GGGGGGHCDGKHGGGGGGGG | GGGGGGHSDGKCGGGGGGG |
| GGGGGGHGDGKCGGGGGGG | GGGGHSDGKCGGGGGGG |
| GGGGHGDGKCGGGGGGG | GGGGGGGHDGKCGGGGGGG |
| GGGGGHDGKCGGGGGGG | GGGGGGGDHGKCGGGGGGG |
| GGHGDGKCGG | GGGGGGHGHGKCGGGGGGG |
| GGGGGGGGHGHKCGGGGGGG | GGHCDGKGGG |
| GGHGDGCKGG | GGDGHGCKGG |
| GGDGHGCGKGG | GGKGDGHGCGG |
| GGKGDHGCGG | GGKGDHCGGG |
| GGDGHPCGKGG | GGDGHPCPKGG |
| GGDPHPCPKGG | GGKGHPDGG |
| GGDPHPCKGGG | GGDGHPCKGGG |
| GGDPHPCKGGG | GGHPDPCKGGG |
| DGDHPHPCKGGS | GGGGGGHCHGKGGGGGGGG |
| GGDPHPGCKGG | GGDPHPGCKGG |
| GGDPCPHGKGG | GGDPCPHPKGG |
| GGHPCPDGKGG | GGHPCDPKGG |
| GGHPKPDPCGG | GGHPDPCPKGG |
| GGHPDPCPKGG | GGEPCHPKGG |
| GGHPEPCPKGG | GGHPEPCPKGG |
| GGEPHPCPKGG | GGEPHKCGGG |
| GGGEHKCGGGG | GGGEHCKGGGG |
| GGGFCPFGGG | AAAAAAAFCPFAAAAAAAA |
| AAAAAAAFPCFAAAAAAAA | AAHPDPCPKAA |
| AAHPDPCPKPFAA | AAAAAAAHDPDPCPKAAAAAAA |
| AAAAAAAHDPDPCPKPFAAAAAAAA | GGHPDPCPKPFGG |
| GGHPDPCPKPFGG | GGHPGDPKFGG |
| GGHPDPCPFKGG | GGHPDPCPFKGG |
| GGHPDPCPYKGG | GGHPDPCPWKGG |
| GGGGHPDPCPWKGGGG | GGGGHPDPKCPWGGGG |
| GGGGHPCPKDWGGGG | GGGGHPCPKDWGGGG |
| GGGGHPCPKPWGGGG | GGGGHWCPDKGGGG |
| GGGGHCPDKPWGGGG | GGGGWPCPKHGGGG |
| GGGGGGGGHPDPCWKGGGGGGGG | GGGGGGGGHPWPDPCKGGGGGGGG |
| GGGGGGGGHPDPWPCPKGGGGGGGG | GGGGGGGGHPDPCWPKGGGGGGGG |
| AAAAAAAHPPDPCPKAAAAAAA | GGGGGGGGDHPKPCPWGGGGGGGG |
| GGGGGGGGHPDPCPWKGGGGGGGG | GGHPPDPCPKGG |
| GGHPKPHPCGG | GGCPHPKPHGG |

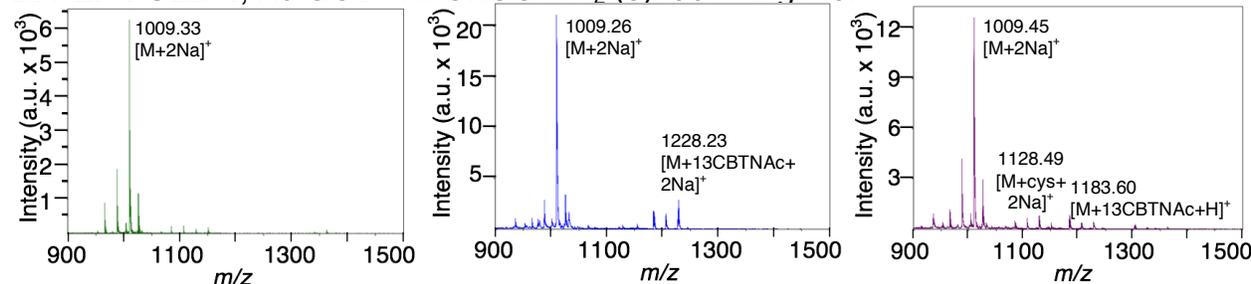
| | |
|--------------------------------------|----------------------|
| GGCPKPHPHGG | GGHPCPKPHGG |
| GGCPKPHHGG | GGCPKHGG |
| GGHPKPCPHGG | GGCPKPSPHGG |
| GGGGCPKPHGGGG | GGHPCPHPKGG |
| GGHPCGHPKGG | GGHGCPHPKGG |
| GGHGCGHPKGG | GGHCGHPKGG |
| GGHGCHPKGG | GGHPKPHGCGG |
| GGHPKWPCGG | GGCPKVPHPHGG |
| GGCPKWPHGG | GGHPWPKPCGG |
| GGHPWPKPCGG | GGCPKPHPWGG |
| PPPPPPPPPP | GGAPKPHCWGG |
| PPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP | AAAAAAAAAAPAAAAAAAAA |
| GGKCPHPPPHGG | GGGCPHKPPHGG |
| GGKPCPHPPHGG | AAAAAPPAAAAA |
| GGCPHHPKGG | GGCPHPWKGG |
| GGCPKPEPHGG | GGCPKDPHGG |
| GGCPKHPEGG | GGCPKHPDGG |
| GGCPHPKPDGG | GGHPDKPCGG |
| GGHPCPKPDGG | GGHPKPCPDGG |
| GGDPCPKPHGG | GGDHPKPCGG |
| GGDPKPHPCGG | GGDPKPCPHGG |
| GGCPHPDPKGG | GGCPDHPKGG |
| GGCPDKPHGG | GGKDPHPCGG |
| GGKPDPCPHGG | GGKPHDPCGG |
| GGKPHGPDGG | GGKPCHPDGG |
| GGKPCDPHGG | GGHPCPKPHPHGG |
| GGHPHPCPKPHPHGG | GGHPHPCPKPHGG |
| GGCPWPKPHGG | GGWPCPKPHGG |
| GGGGHPCPKPHPHGGGG | GGGGHPCPKPHPHGGGG |
| GGHPHPCPKGG | GGKPCPHPHGG |
| GGKPCWPHGG | GGHPWPCPKGG |
| GGWPHPCPKGG | GGHPHPKPCGG |
| GGHPWPKPCGG | GGWPHKPCGG |
| GGHPTPCPKGG | GGCPKTPHGG |
| GGHPHPKPCPDGG | GGCPTPKPHGG |
| GGHPHPKPCPEGG | GGHPWPKPCPDGG |
| GGHPWPKPCPEGG | GGCPKYPHGG |
| GGHPYPCPKGG | GGCPKHPYGG |
| GGKPCHPYGG | GGCPKPRPHGG |
| GGCPRKPHGG | GGCPKPKPHGG |
| GGCPKHPKGG | GGCPHPKPKGG |
| GGKPCHPKGG | GGKPKHPCGG |
| GGKPKPCPHGG | GGHPHPKPCPRGG |
| GGHPKPHPCPRGG | GGHPWPKPCPRGG |
| GGCPKHPRGG | GGRPWPKPCPHGG |
| GGRPWPKHPCGG | GGRPWPKPHCGG |

| | |
|----------------|---------------------|
| GGRPWPCKPHCGGG | GGRPWPCPHKGGG |
| GGWPCPHPKGG | GGWPHPCPKGGG |
| GGWPHGPCPKGG | GGWPHPGCPKGG |
| GGWPHPGKPCGG | GGHPWPGCPKGG |
| GGKPCGPWPHGG | GGKPCPGWPHGG |
| GGCPKPGWPHGG | GGFMGFPHPKPHPCGSGG |
| GGFFPKPHPCGG | GGFFPGKPHPCGG |
| GGFFPGPKPHPCGG | GGFFPCPKPHPCGG |
| GGFFPCPKPCPHGG | GGHPCPKPGPFPGG |
| GGHPCPGKPFPGG | GGCPHPGKPFPGG |
| GGCPKPFPPDPHGG | GGCPHPFPFKGG |
| GGHPCPFPPKGG | GGSPCFPPFKGG |
| GGHCPFPFKGG | GGHPFPFCPKGG |
| GGHPFPFKPCGG | GGHPKPFPPCGG |
| GGHPKPCPFGG | GGCPHPFKGG |
| GGCPFPHPKGG | GGHPCPKPFGG |
| GGFPCPKPHGG | GGFPCPKPHFG |
| GGFPCPKPHFGG | GGHPKPGCPFGG |
| GGHPKGPCPFGG | GGHPKPGCGFGG |
| GGHPKPGCFGG | GGHPKPCGFGG |
| GGHPCPKGFGG | GGHPKPWGFGG |

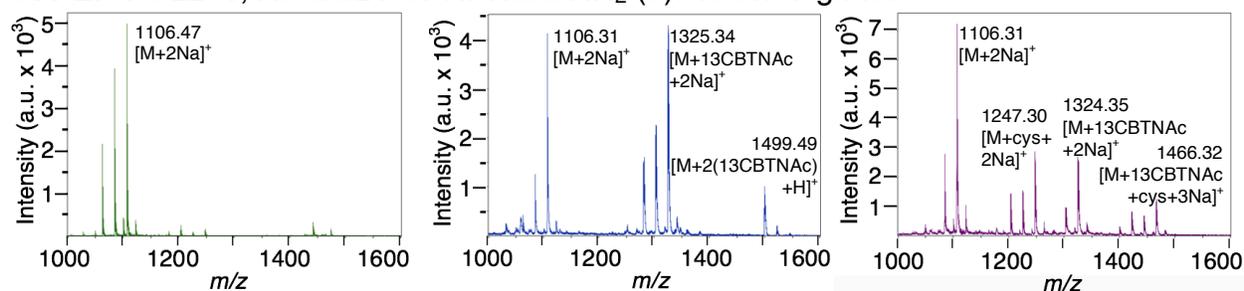
A. CBTa_g 1.0, Ac-GGHPDPCPKGG-NH₂ (4): 1061.47 g/mol



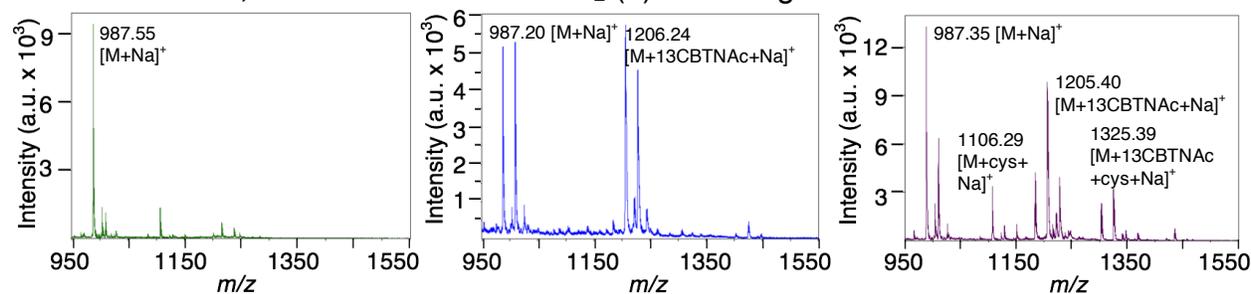
B. PEP-FOLD-4, Ac-GGDHPCKGG-NH₂ (5): 964.42 g/mol



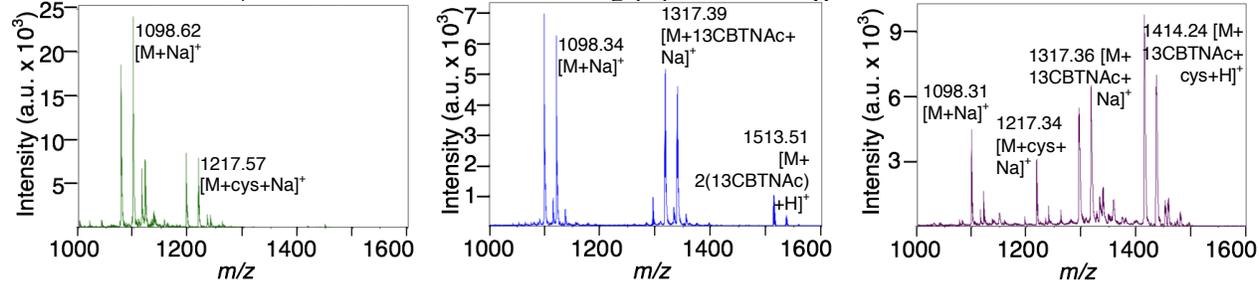
C. PEP-FOLD-1, Ac-GGDPCPHPKGG-NH₂ (6): 1061.47 g/mol



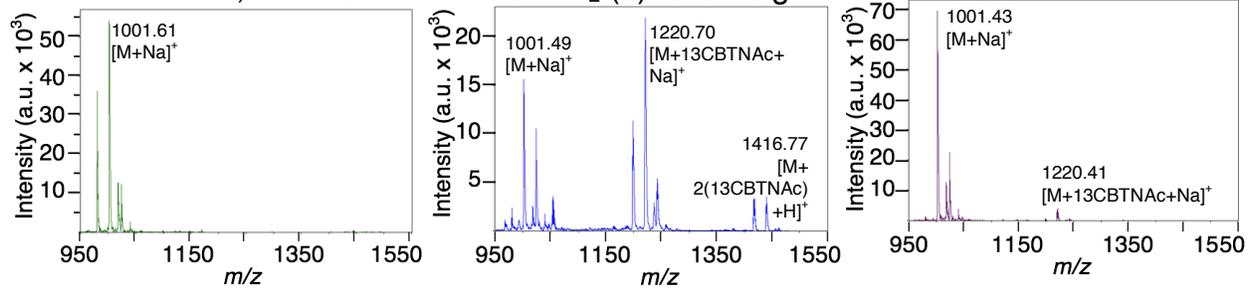
D. PEP-FOLD-2, Ac-GGHPDCPKGG-NH₂ (7): 964.42 g/mol



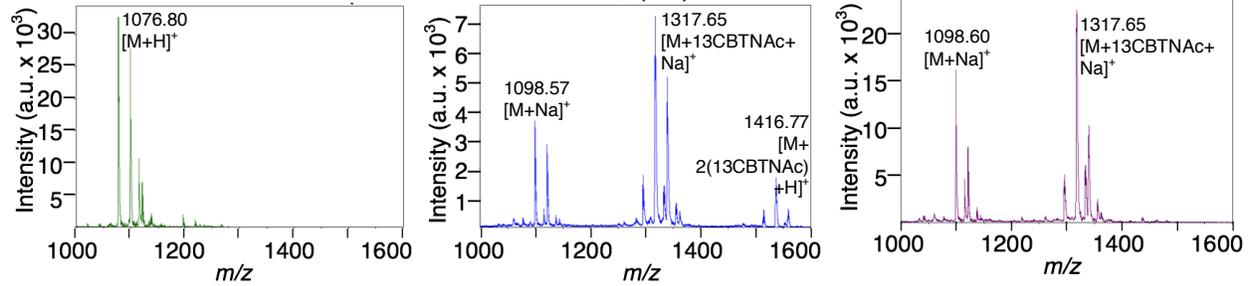
E. PEP-FOLD-5, Ac-GGHPEPCPKGG-NH₂ (8): 1075.49 g/mol



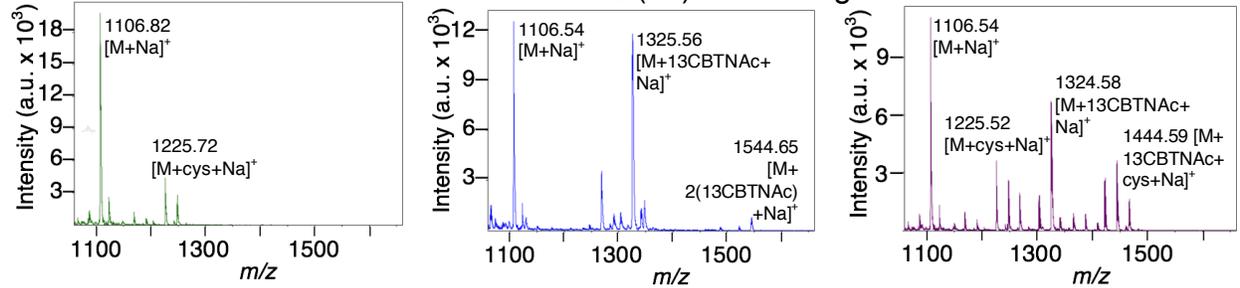
F. PEP-FOLD-6, Ac-GGHPECPKGG-NH₂ (9): 978.43 g/mol



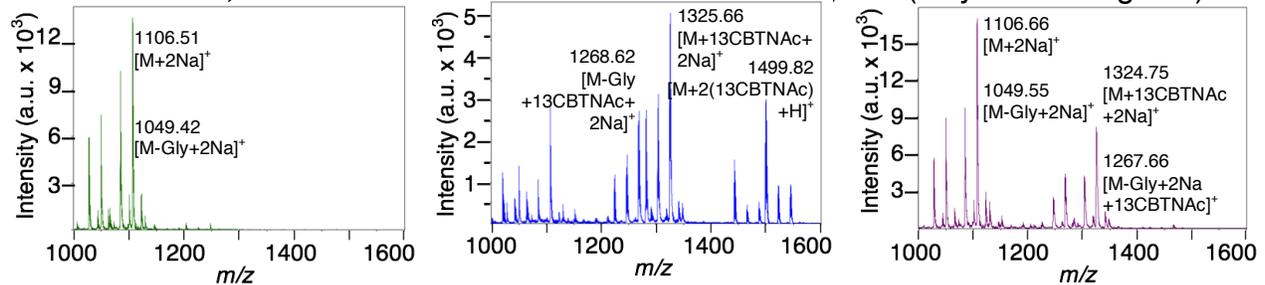
G. PEP-FOLD-7, Ac-GGEPHPCPKGG-NH₂ (10): 1075.49 g/mol



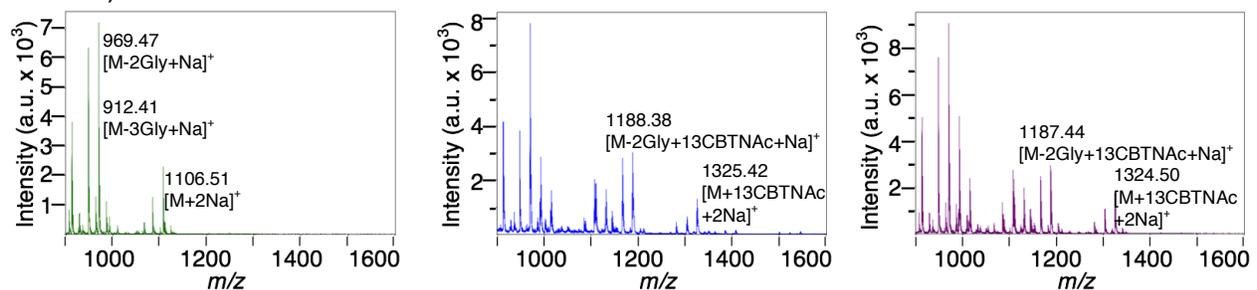
H. PEP-FOLD-8, Ac-GGCPKPHPHGG-NH₂ (11): 1083.50 g/mol



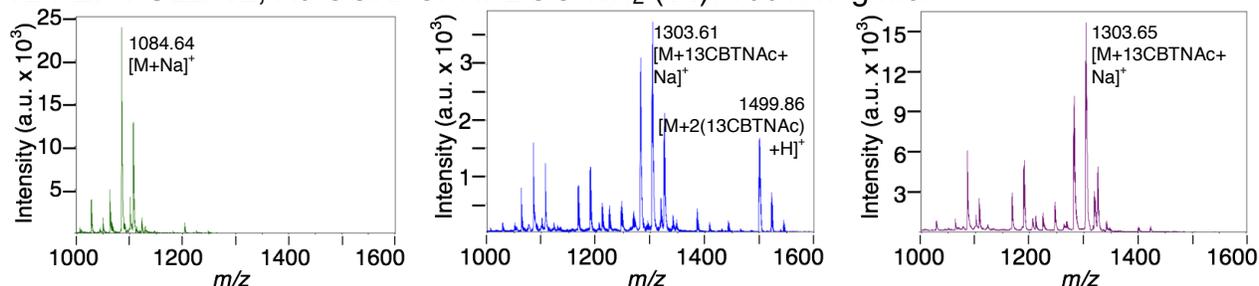
I. PEP-FOLD-9, Ac-GGCPKPDPHGG-NH₂ (12): 1061.47 g/mol (-Gly: 1004.45 g/mol)



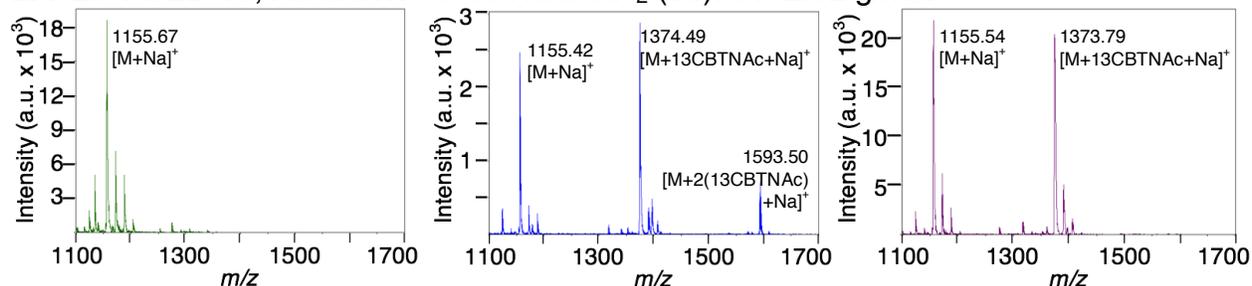
J. PEP-FOLD-11, Ac-GGCPDHPKGG-NH₂ (13): 1061.47 g/mol (-2Gly: 947.43; -3Gly: 890.41)



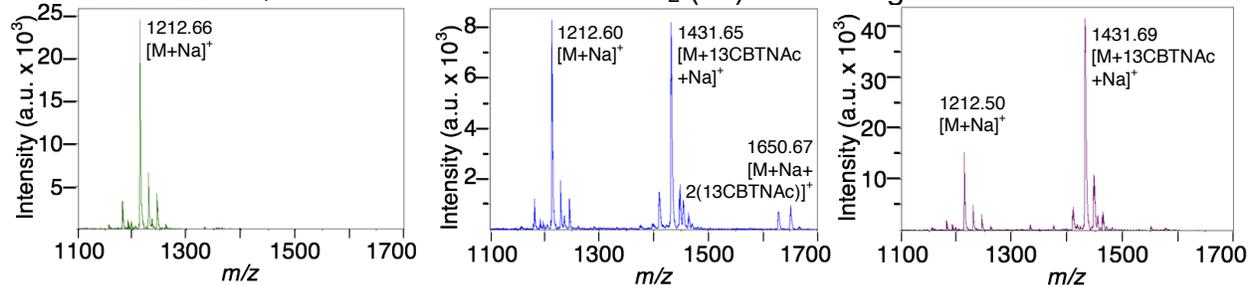
K. PEP-FOLD-12, Ac-GGKPCPHPDGG-NH₂ (14): 1061.47 g/mol



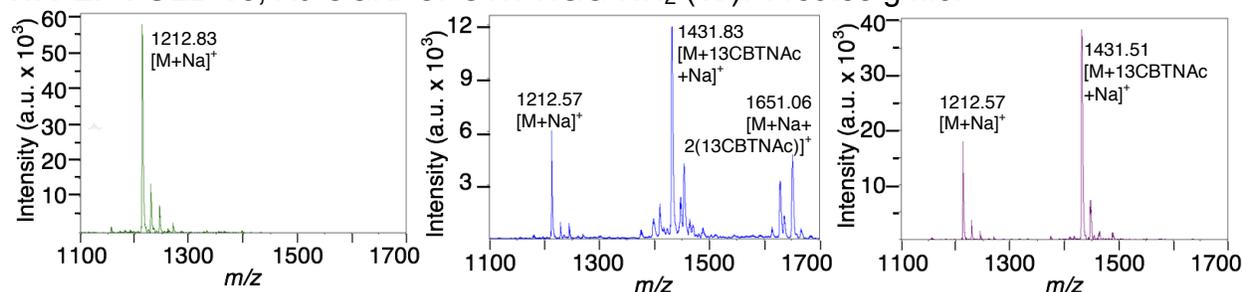
L. PEP-FOLD-13, Ac-GGWPHPCPKGG-NH₂ (15): 1132.52 g/mol



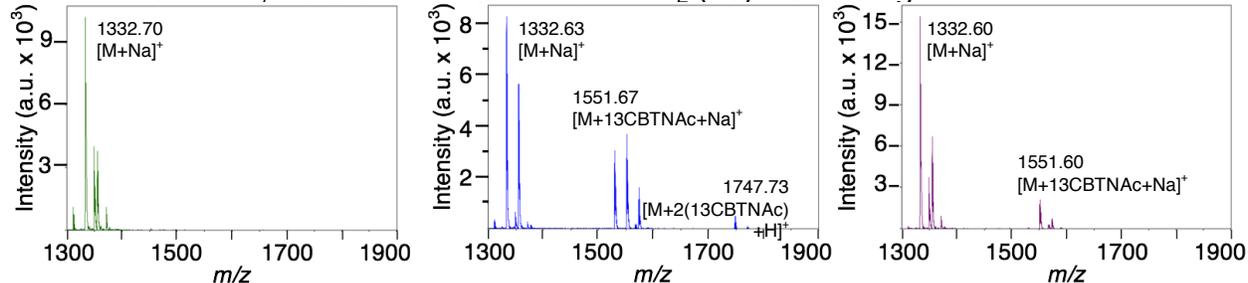
M. PEP-FOLD-14, Ac-GGWPHPGCPKGG-NH₂ (16): 1189.55 g/mol



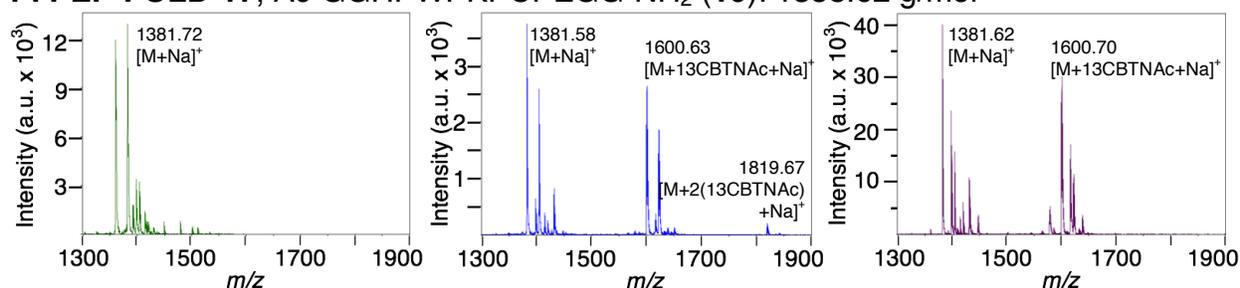
N. PEP-FOLD-15, Ac-GGKPCPGWPHGG-NH₂ (17): 1189.55 g/mol



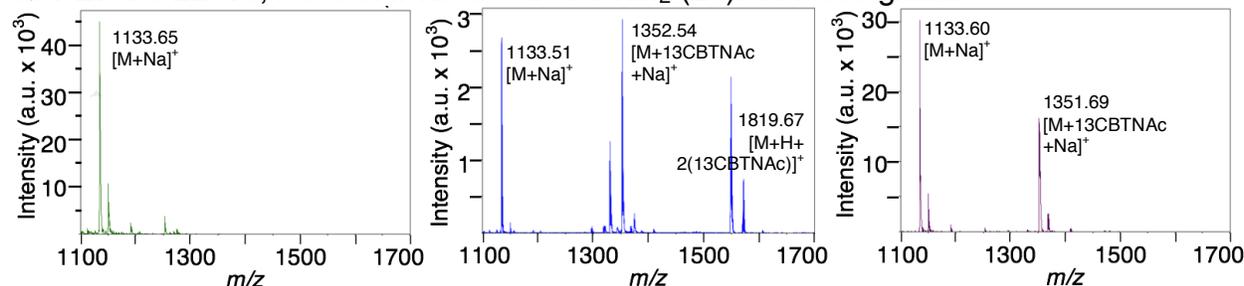
O. PEP-FOLD-16, Ac-GGHPHPKPCPEGG-NH₂ (18): 1309.60 g/mol



P. PEP-FOLD-17, Ac-GGHPWPKPCPEGG-NH₂ (19): 1358.62 g/mol



Q. PEP-FOLD-18, Ac-GGHPKPGCGFGG-NH₂ (20): 1110.50 g/mol



R. PEP-FOLD-19, Ac-GGHPKPGCGFGG-NH₂ (21): 1053.48 g/mol (-Gly: 996.46 g/mol)

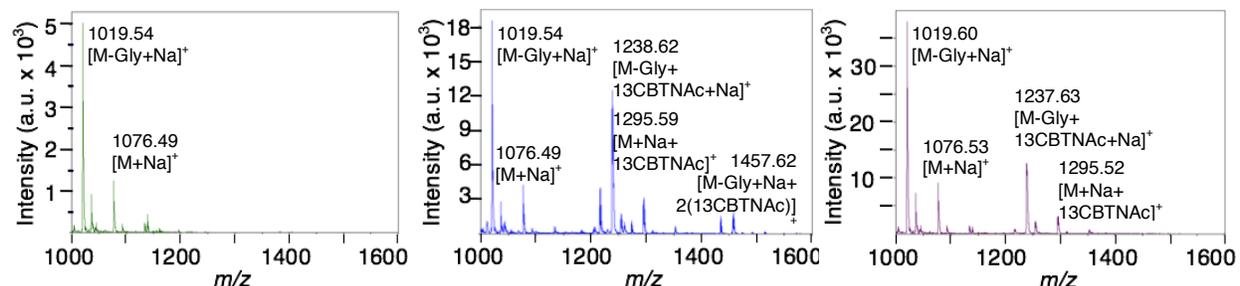


Figure 2.S1. 13CBTNac binding assay and cysteine competition MALDI-TOF-MS data for the eighteen peptides synthesized by SPPS. All peptides were re-purified by preparative RP-HPLC immediately prior to this experiment and all data in this figure were obtained in parallel in order to minimize sources of error. The binding assay and cysteine competition were as follows: 40 μ L of a 10% DMF in PBS solution of peptide (0.9 mM) with or without 13CBTNac (1 mM) was incubated 48 h at r.t. Then, the negative control (without 13CBTNac) and half (20 μ L) of the sample incubated with 13CBTNac were each treated with an equal volume (20 μ L) of a solution of cysteine and DTT (10 and 15 mM, respectively) in PBS. The other half of the sample incubated with 13CBTNac was treated with an equal volume of a solution of DTT (15 mM) in PBS. After 48 h at r.t., MALDI-TOF MS was used to analyze the negative control (left column, green spectra), the sample incubated with 13CBTNac then DTT alone (middle column, blue spectra), and the sample incubated with 13CBTNac then cysteine and DTT (right column, purple spectra). Labels are in the format **Name of peptide**, Sequence (**Compound #**): Calculated exact mass (identity of major impurity and calculated exact mass, where applicable). **A)** CBTag 1.0 (**4**), analogous to Figure 2.3B. **B)** PEP-FOLD-4, analogous to Figure 2.3C (**5**). **C-R)** Other peptides tested. Peptides in **I**, **J**, and **R** have additional peaks corresponding to glycine deletions, but the residues that confer reactivity (e.g., Cys, Lys) were not affected.

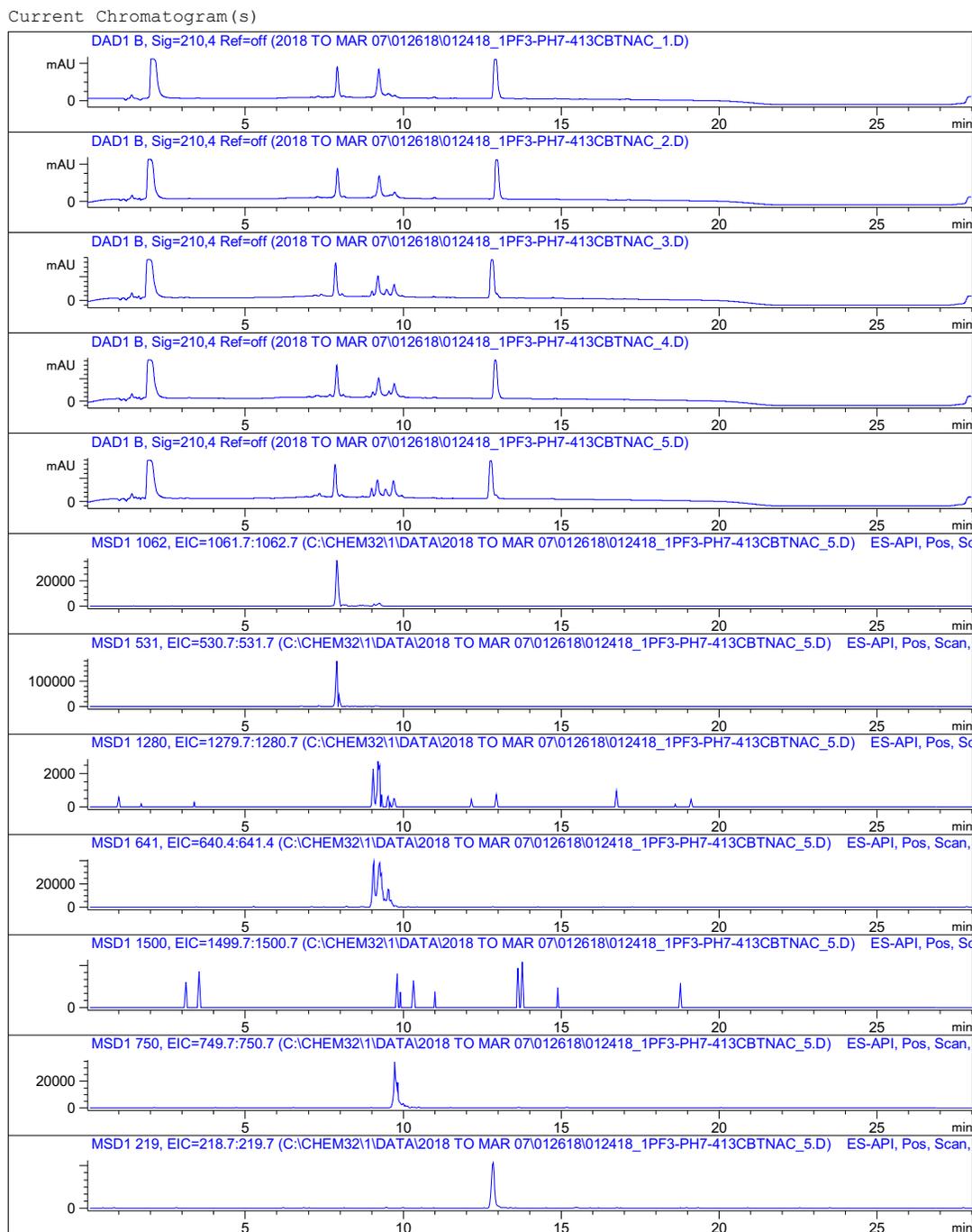


Figure 2.S2. Overlay of LC-MS chromatographs for 1:4 CBTag 1.0/13CBTNAC over the time-course experiment and extracted mass ions. Identities of extracted mass ions: 1062 m/z, minute 7.9 – [CBTag 1.0 + H]⁺; 531 m/z, minute 7.9 – [CBTag 1.0 + 2H]²⁺; 1280 m/z, minutes 9 and 9.2 – [CBTag 1.0-13CBTNAC + H]⁺; 641 m/z, minutes 9 and 9.2 – [CBTag 1.0-13CBTNAC + 2H]²⁺; 1500 m/z, insignificant signal – [CBTag 1.0-2(13CBTNAC) + H]⁺; 750 m/z, minute 9.7 – [CBTag 1.0-2(13CBTNAC) + 2H]²⁺; 219 m/z, minute 12.9 – [13CBTNAC + H]⁺

Print of window 38: Current Chromatogram(s)
 Data File : C:\CHEM32\1\DATA\2018 TO MAR 07\010518\010518_1MMPEPFOLD-3-1MMCBTNAC-2MMTCEP_6.D
 Sample Name : 010518_lmMpepfold-3-1mMcbtnac-2mMTCEP_6

=====

| | |
|------------------------------------|-------------------------------|
| Acq. Operator : sglk | Seq. Line : 99 |
| Acq. Instrument : Instrument 1 | Location : Vial 23 |
| Injection Date : 1/7/18 4:19:26 PM | Inj : 1 |
| | Inj Volume : 15.000 µl |
| | Actual Inj Volume : 10.000 µl |

Different Inj Volume from Sequence !
 Acq. Method : D:\CHEM32\1\METHODS\AP_C18_INSERT.M
 Last changed : 11/10/17 3:07:57 PM by Mason
 Analysis Method : Z:\DESKTOP\BERTOZZI LAB\LCMS INFO\LCMS METHODS TO 112917\SGLK_BERTOZZIBP_WITHINSERT.M
 Last changed : 3/8/18 8:51:08 AM
 (modified after loading)
 Additional Info : Peak(s) manually integrated

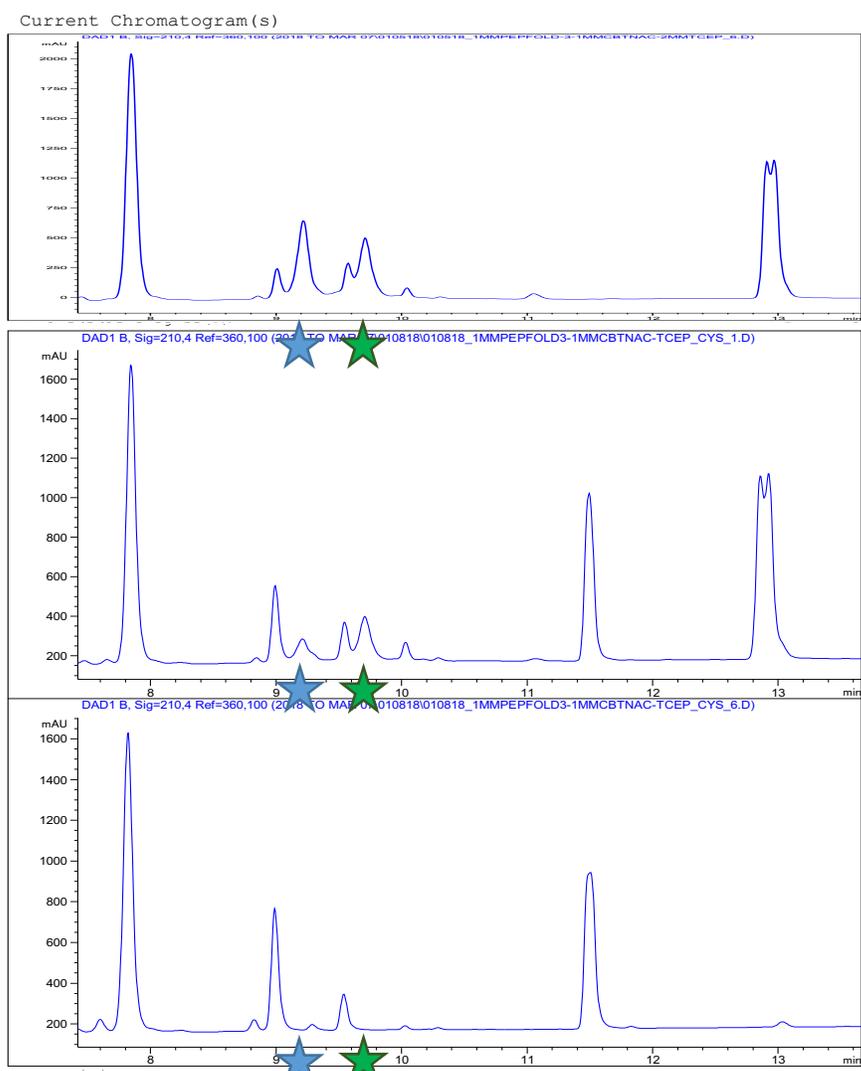


Figure 2.S3. Cysteine treatment identifies peaks that contain thioimide adducts. Top spectrum - before cysteine treatment; middle - early time point during cysteine treatment; bottom - end of the time-course assay. Based on these results and the mass spectrum in Figure 2.S2, the blue star (minute 9.2) indicates the cysteine adduct, while the green star (minute 9.7) indicates the doubly modified peptide; the peak at minute 9 is the lysine adduct.

Print of window 38: Current Chromatogram(s)
Data File : C:\CHEM32\...TA\2018 TO MAR 07\010518\010518_1MMPEPFOLD-3R-1MM13CBTNAC-2MMTCEP.D
Sample Name : 010518_1mPepfold-3R-1mMcbtnac-2mMTCEP_1
=====

| | |
|------------------------------------|-------------------------------|
| Acq. Operator : sglk | Seq. Line : 21 |
| Acq. Instrument : Instrument 1 | Location : Vial 31 |
| Injection Date : 1/6/18 2:26:29 AM | Inj : 1 |
| | Inj Volume : 15.000 µl |
| | Actual Inj Volume : 10.000 µl |

Different Inj Volume from Sequence !
Acq. Method : D:\CHEM32\1\METHODS\AP_C18_INSERT.M
Last changed : 11/10/17 3:07:57 PM by Mason
Analysis Method : Z:\DESKTOP\BERTOZZI LAB\LCMS INFO\LCMS METHODS TO 112917\SLK_BERTOZZIBP_ WITHINSERT.M
Last changed : 3/8/18 8:51:08 AM
(modified after loading)
Additional Info : Peak(s) manually integrated

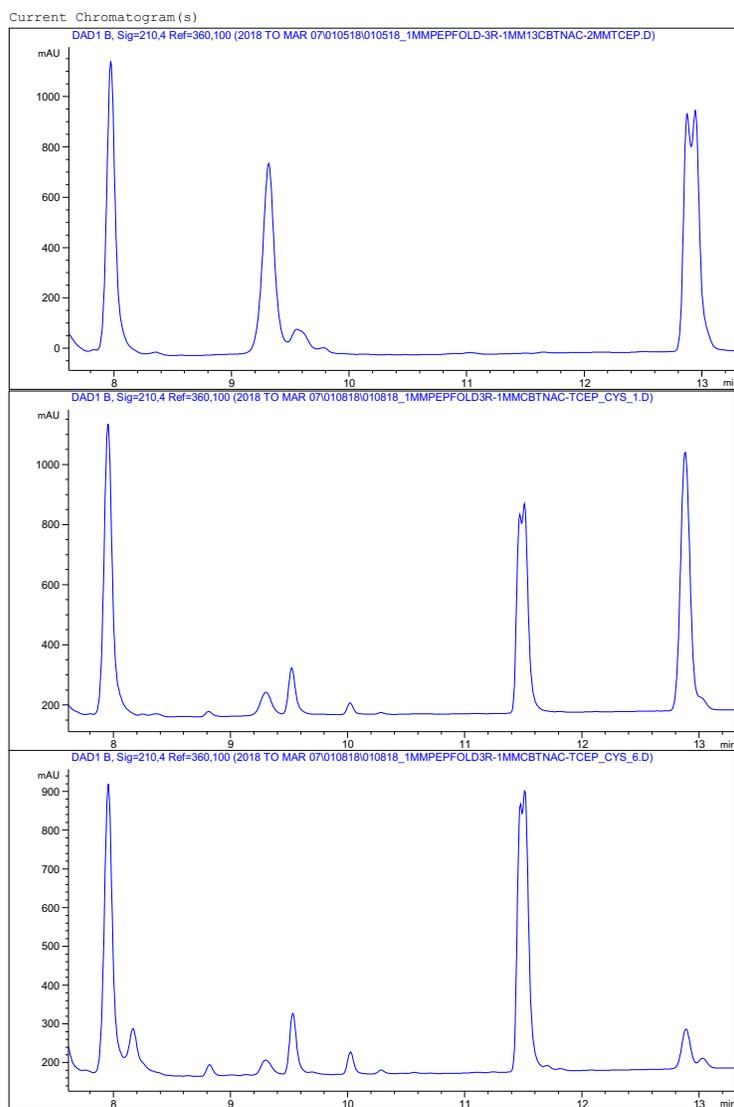


Figure 2.S5. Cysteine treatment causes the minute 9.3 peak corresponding to the singly modified CBTtag 1.0 (K9R) to disappear.

Table 2.S2. Calculations for 1:1 CBTAg 1.0/13CBTNAC LC-MS time-course assay (two replicates). Spectra and integration results used herein are included in Appendix 1.

| | | Time 0 = 3:16 - | | Time 0 = 3:16 - | | Time 0 = 3:16 - | |
|--|---------------------------|---|---------|---|-----------------------------------|---|---|
| | | 0:05:30 = | | 3:10:30 | | 3:10:30 | |
| | | 26-Jan | | 26-Jan | | 26-Jan | |
| 210 nm | 012418_PFB-PHZ-13CBTNAC_1 | Time of injection in file (h) | 1.80 | CBTAg 1.0 (min 7.9) | CBTAg 1.0-Lys-13CBTNAC (min 8.99) | CBTAg 1.0-Cys-13CBTNAC (min 9.19) | CBTAg 1.0-Cys-13CBTNAC-Lys-13CBTNAC (min 9.7) |
| | 012418_PFB-PHZ-13CBTNAC_2 | 4:58:31 p.m. | 1.80 | 19072.5 | 875.5 | 20975.8 | 13CBTNAC (min 9.7) |
| | 012418_PFB-PHZ-13CBTNAC_3 | 5:25:45 a.m. | 14.25 | 21742.2 | 2265.4 | 20082 | 13CBTNAC (min 9.7) |
| | 012418_PFB-PHZ-13CBTNAC_4 | 3:39:38 p.m. | 24.48 | 31604.91 | 2944.3 | 13766.1 | 13CBTNAC (min 9.7) |
| | 012418_PFB-PHZ-13CBTNAC_5 | 11:53:39 p.m. | 32.72 | 28513.8 | 3432 | 13322.2 | 13CBTNAC (min 9.7) |
| | | 11:01:48 a.m. | 43.85 | 25995.5 | 2775.6 | 8668.3 | 13CBTNAC (min 9.7) |
| 26-Jan | | Time 0 = 3:16 - | 3:10:30 | CBTAg 1.0-Lys-13CBTNAC (min 8.99) | CBTAg 1.0-Cys-13CBTNAC (min 9.19) | CBTAg 1.0-Cys-13CBTNAC-Lys-13CBTNAC (min 9.7) | CBTAg 1.0-Cys-13CBTNAC-Lys-13CBTNAC (min 9.7) |
| | | 0:05:30 = | 3:10:30 | 13CBTNAC (min 8.99) | 13CBTNAC (min 9.19) | 13CBTNAC (min 9.7) | 13CBTNAC (min 9.7) |
| 254 nm | 012418_PFB-PHZ-13CBTNAC_1 | Time of injection in file (h) | 1.80 | CBTAg 1.0 (min 7.9) | CBTAg 1.0-Lys-13CBTNAC (min 8.99) | CBTAg 1.0-Cys-13CBTNAC (min 9.19) | CBTAg 1.0-Cys-13CBTNAC-Lys-13CBTNAC (min 9.7) |
| | 012418_PFB-PHZ-13CBTNAC_2 | 4:58:31 p.m. | 14.25 | 138.3 | 136.8 | 4066.1 | 13CBTNAC (min 9.7) |
| | 012418_PFB-PHZ-13CBTNAC_3 | 5:25:45 a.m. | 24.48 | 365.8 | 579.3 | 3527.3 | 13CBTNAC (min 9.7) |
| | 012418_PFB-PHZ-13CBTNAC_4 | 3:39:38 p.m. | 32.72 | 579.3 | 561 | 2843.1 | 13CBTNAC (min 9.7) |
| | 012418_PFB-PHZ-13CBTNAC_5 | 11:53:39 p.m. | 43.85 | 561 | 2454 | 2454 | 13CBTNAC (min 9.7) |
| | | 11:01:48 a.m. | 43.85 | 445.7 | 445.7 | 1481.1 | 13CBTNAC (min 9.7) |
| Assumption: area under modified peptide peaks is proportional to number of modifications | | | | | | | |
| 210 nm | | | | | | | |
| Lysine-modified: | | Any modification: Sum | | All peptide: Sum | | Percent of total peptide that has any modification (Any) | |
| Sum CBTAg 1.0-Lys-13CBTNAC and 1/2(CBTAg 1.0-Cys-13CBTNAC-Lys-13CBTNAC) | | Sum CBTAg 1.0-Lys-13CBTNAC, CBTAg 1.0-Cys-13CBTNAC, CBTAg 1.0-Cys-13CBTNAC-Lys-13CBTNAC | | Sum CBTAg 1.0-Lys-13CBTNAC, CBTAg 1.0-Cys-13CBTNAC, CBTAg 1.0-Cys-13CBTNAC-Lys-13CBTNAC | | Percent of total peptide bound to a lysine (Lysine-modified/Any modification/All peptide) | |
| 429.8 | | 23942.9 | | 41969.6 | | 54.5563932 | |
| 1153.35 | | 5530.9 | | 4735.1 | | 54.0868821 | |
| 1651.9 | | 7180.25 | | 52551.26 | | 39.85898206 | |
| 1842.3 | | 8663.65 | | 50399.65 | | 43.42460712 | |
| 1414.35 | | 6936.95 | | 41600.75 | | 37.5119439 | |
| 254 nm | | | | | | | |
| Lysine-modified: | | Any modification: Sum | | Percent of total peptide that has any modification (Any) | | Percent of total peptide bound to a lysine (Lysine-modified/Any modification/All peptide) | |
| Sum CBTAg 1.0-Lys-13CBTNAC and 1/2(CBTAg 1.0-Cys-13CBTNAC-Lys-13CBTNAC) | | Sum CBTAg 1.0-Lys-13CBTNAC, CBTAg 1.0-Cys-13CBTNAC, CBTAg 1.0-Cys-13CBTNAC-Lys-13CBTNAC | | Sum CBTAg 1.0-Lys-13CBTNAC, CBTAg 1.0-Cys-13CBTNAC, CBTAg 1.0-Cys-13CBTNAC-Lys-13CBTNAC | | Percent of total peptide bound to a lysine (Lysine-modified/Any modification/All peptide) | |
| 429.8 | | 23942.9 | | 41969.6 | | 54.5563932 | |
| 1153.35 | | 5530.9 | | 4735.1 | | 54.0868821 | |
| 1651.9 | | 7180.25 | | 52551.26 | | 39.85898206 | |
| 1842.3 | | 8663.65 | | 50399.65 | | 43.42460712 | |
| 1414.35 | | 6936.95 | | 41600.75 | | 37.5119439 | |

2-Feb Time 0 = 5:02 -
0:05:30 = 4:56:30

*a lot of these peaks were saturated using 280 nm as control

| 210 nm | Actual time point (h) | CBTag 1.0 (min 7.9) | CBTag 1.0-Lys-13CBTNAC (min 8.99) | CBTag 1.0-Cys-13CBTNAC (min 9.19) | CBTag 1.0-Cys-13CBTNAC 210 nm control | CBTag 1.0-Cys-13CBTNAC-Lys-13CBTNAC (min 9.7) |
|---------------------------|-----------------------|---------------------|-----------------------------------|-----------------------------------|---------------------------------------|---|
| 020218_PFG-PH7-13CBTNAC_1 | 6:47:32 p.m. | 2.33 | 14758.6 | 20164.3 | 21027.50305 | 2894.9 |
| 020218_PFG-PH7-13CBTNAC_2 | 5:55:51 a.m. | 13.47 | 15407.4 | 24026.3 | 24579.91136 | 11859.8 |
| 020218_PFG-PH7-13CBTNAC_3 | 2:09:56 p.m. | 21.71 | 11620.8 | 16445.3 | 18346.71955 | 15010.7 |
| 020218_PFG-PH7-13CBTNAC_4 | 10:23:55 p.m. | 29.94 | 11865.8 | 17331.6 | 18916.25472 | 16845.4 |
| 020218_PFG-PH7-13CBTNAC_5 | 9:32:20 a.m. | 41.08 | 8382.1 | 12603.9 | 10621.54062 | 13433.9 |

280 nm control for saturation CBTag 1.0-Cys-280 nm nm/13CBTNAC 280

| | | | |
|---------|---------|-------------|---------|
| 4321.66 | 4321.66 | 2.390602787 | 1807.77 |
| 4781.66 | 4781.66 | 2.172444969 | 2201.05 |
| 4214.79 | 4214.79 | 1.896478179 | 2222.43 |
| 3953.2 | 3953.2 | 3.885437962 | 1017.44 |
| 1891.59 | 1891.59 | 2.604276294 | 726.34 |

2-Feb Time 0 = 5:02 -
0:05:30 = 4:56:30

*a lot of these peaks were saturated

| 254 nm | Actual time point (h) | CBTag 1.0 (min 7.9) | CBTag 1.0-Lys-13CBTNAC (min 8.99) | CBTag 1.0-Cys-13CBTNAC (min 9.19) | CBTag 1.0-Cys-13CBTNAC-Lys-13CBTNAC (min 9.7) |
|---------------------------|-----------------------|---------------------|-----------------------------------|-----------------------------------|---|
| 020218_PFG-PH7-13CBTNAC_1 | 6:47:32 p.m. | 2.33 | 167.6 | 6653.1 | 1016 |
| 020218_PFG-PH7-13CBTNAC_2 | 5:55:51 a.m. | 13.47 | 449.1 | 6981.6 | 3533.3 |
| 020218_PFG-PH7-13CBTNAC_3 | 2:09:56 p.m. | 21.71 | 668.3 | 5847.8 | 5163.1 |
| 020218_PFG-PH7-13CBTNAC_4 | 10:23:55 p.m. | 29.94 | 900.5 | 5806.4 | 6531.1 |
| 020218_PFG-PH7-13CBTNAC_5 | 9:32:20 a.m. | 41.08 | 765.4 | 3172 | 5281 |

Assumption: area under modified peptide peaks is proportional to number of modifications
210 nm

| Lysine-modified: | | Any modification: Sum | | All peptide (adjusted for saturation): Sum | | Percent of total peptide that has any 13CBTNAC modification (Any modification - 1/2/CBTAG 1.0-Cys-13CBTNAC-Lys-peptide) | |
|---|--|--|--|--|--|---|--|
| Sum CBTAG 1.0-Lys-13CBTNAC and 1/2/CBTAG 1.0-Cys-13CBTNAC-Lys-13CBTNAC) | 13CBTNAC, CBTAG 1.0-Lys-13CBTNAC, CBTAG 1.0-Cys-13CBTNAC, CBTAG 1.0-Cys-13CBTNAC-Lys-13CBTNAC) | 13CBTNAC, CBTAG 1.0-Cys-13CBTNAC, CBTAG 1.0-Cys-13CBTNAC-Lys-13CBTNAC) | 13CBTNAC, CBTAG 1.0-Cys-13CBTNAC, CBTAG 1.0-Cys-13CBTNAC-Lys-13CBTNAC) |
| 2115.15 | 23726.9 | 37038.05 | 8.914565325 | 5.710748811 | 60.15286982 | 67.28532815 | 71.27780346 |
| 7662.6 | 37618.8 | 47096.3 | 20.36907078 | 16.27006792 | 69.79456782 | 71.27780346 | 72.93359575 |
| 10406.45 | 34357.1 | 38472.55 | 30.28908144 | 27.04902586 | 69.79456782 | 71.27780346 | 72.93359575 |
| 12114.9 | 37889.2 | 41312.3 | 31.99143367 | 29.32516466 | 71.27780346 | 72.93359575 | 72.93359575 |
| 9982.65 | 29303.5 | 30968.65 | 34.06640845 | 32.23469541 | 72.93359575 | 72.93359575 | 72.93359575 |

| Lysine-modified: | | Any modification (adjusted for saturation): | | All peptide (adjusted for saturation): Sum | | Percent of total peptide that has any 13CBTNAC modification (adjusted for saturation) (Any modification - 1/2/CBTAG 1.0-Cys-13CBTNAC-Lys-peptide) | |
|---|--|--|--|--|--|---|--|
| Sum CBTAG 1.0-Lys-13CBTNAC and 1/2/CBTAG 1.0-Cys-13CBTNAC-Lys-13CBTNAC) | 13CBTNAC, CBTAG 1.0-Cys-13CBTNAC, CBTAG 1.0-Cys-13CBTNAC-Lys-13CBTNAC) | 13CBTNAC, CBTAG 1.0-Cys-13CBTNAC, CBTAG 1.0-Cys-13CBTNAC-Lys-13CBTNAC) |
| 2115.15 | 24590.10305 | 39348.70305 | 8.601631296 | 5.375399533 | 58.81427152 | 60.15286982 | 60.05342144 |
| 7662.6 | 38172.41136 | 53579.81136 | 20.07366981 | 14.30128215 | 60.17660485 | 60.05342144 | 60.46641367 |
| 10406.45 | 36258.51955 | 47879.31955 | 28.70070297 | 21.73474915 | 60.05342144 | 60.46641367 | 60.46641367 |
| 12114.9 | 39453.85472 | 51319.65472 | 30.70650533 | 23.60674495 | 60.46641367 | 60.46641367 | 60.46641367 |
| 9982.65 | 27321.14062 | 35703.24062 | 36.53818901 | 27.960067 | 57.70958116 | 57.70958116 | 57.70958116 |

| Lysine-modified: | | Any modification: Sum | | Percent of total peptide that has any 13CBTNAC bound to a lysine (Lysine-modified/Any modification) | |
|---|--|--|--|---|--|
| Sum CBTAG 1.0-Lys-13CBTNAC and 1/2/CBTAG 1.0-Cys-13CBTNAC-Lys-13CBTNAC) | 13CBTNAC, CBTAG 1.0-Lys-13CBTNAC, CBTAG 1.0-Cys-13CBTNAC, CBTAG 1.0-Cys-13CBTNAC-Lys-13CBTNAC) | 13CBTNAC, CBTAG 1.0-Cys-13CBTNAC, CBTAG 1.0-Cys-13CBTNAC-Lys-13CBTNAC) | 13CBTNAC, CBTAG 1.0-Cys-13CBTNAC, CBTAG 1.0-Cys-13CBTNAC-Lys-13CBTNAC) | 13CBTNAC, CBTAG 1.0-Cys-13CBTNAC, CBTAG 1.0-Cys-13CBTNAC-Lys-13CBTNAC) | 13CBTNAC, CBTAG 1.0-Cys-13CBTNAC, CBTAG 1.0-Cys-13CBTNAC-Lys-13CBTNAC) |
| 675.6 | 7836.7 | 8.620975666 | 20.20932142 | 27.82596411 | 31.47038828 |
| 2215.75 | 10964 | 8.620975666 | 20.20932142 | 27.82596411 | 31.47038828 |
| 3249.85 | 11679.2 | 27.82596411 | 31.47038828 | 36.94675866 | 36.94675866 |
| 4166.05 | 13238 | 31.47038828 | 36.94675866 | 36.94675866 | 36.94675866 |
| 3405.9 | 9218.4 | 36.94675866 | 36.94675866 | 36.94675866 | 36.94675866 |

| 254 nm | Time of injection in file | Actual time point (h) | CBTag 1.0 (min 7.9) | CBTag 1.0-Lys- 13CBTNAC (min 8.99) | CBTag 1.0-Cys- 13CBTNAC (min 9.19) | CBTag 1.0-Cys- 13CBTNAC-Lys- 13CBTNAC (min 9.7) | CBTag 1.0-Cys- 13CBTNAC (min 12.9) |
|-----------------------------|---------------------------|-----------------------|---------------------|---------------------------------------|---------------------------------------|---|---------------------------------------|
| 020218_IPF3-PH7-413CBTNAC_1 | 7:45:39 p.m. | 2.82 | 4179.1 | 967.7 | 9649.2 | 1156.3 | 14850.1 |
| 020218_IPF3-PH7-413CBTNAC_2 | 6:54:01 a.m. | 13.96 | 3660.3 | 461.7 | 9230.7 | 3393.8 | 14882.8 |
| 020218_IPF3-PH7-413CBTNAC_3 | 3:08:04 p.m. | 22.19 | 3616.8 | 660.2 | 6993.7 | 4871.4 | 945.4 |
| 020218_IPF3-PH7-413CBTNAC_4 | 11:22:05 p.m. | 30.42 | 3166.8 | 877 | 7026.9 | 6074.1 | 15707.4 |
| 020218_IPF3-PH7-413CBTNAC_5 | 10:30:27 p.m. | 41.56 | 2262.7 | 1027.3 | 4835.6 | 6570.6 | 11867 |
| 254 nm | | | | | | | |
| 020218_IPF3-PH7-413CBTNAC_1 | 7:45:39 p.m. | 2.82 | 945.85 | 1117.32 | 14774.15 | 6.402060355 | 71.71343191 |
| 020218_IPF3-PH7-413CBTNAC_2 | 6:54:01 a.m. | 13.96 | 2128.6 | 13026.2 | 14774.15 | 16.34091293 | 75.62984367 |
| 020218_IPF3-PH7-413CBTNAC_3 | 3:08:04 p.m. | 22.19 | 3115.9 | 12545.3 | 13726.4 | 24.83719002 | 73.65077515 |
| 020218_IPF3-PH7-413CBTNAC_4 | 11:22:05 p.m. | 30.42 | 3914.05 | 13978 | 14107.75 | 28.00150236 | 77.55276355 |
| 020218_IPF3-PH7-413CBTNAC_5 | 10:30:27 p.m. | 41.56 | 4312.6 | 12433.5 | 11410.9 | 34.68532593 | 80.17071397 |
| | | | 9148.2 | 11410.9 | 11410.9 | 37.79368849 | 80.17071397 |
| | | | 199.1 | 1984 | 10.03528226 | 21.1944508 | |
| | | | 556.1 | 2623.8 | 29.81467701 | 32.88174987 | |
| | | | 841.4 | 2822.1 | 38.33184587 | | |
| | | | 1032.75 | 3140.8 | | | |
| | | | 987.85 | 2577.1 | | | |

REFERENCES

1. Jing, C.; Cornish, V. W. Chemical Tags for Labeling Proteins Inside Living Cells. *Acc. Chem. Res.* **2011**, *44*, 784-792 and references cited therein.
2. Spicer, C. D.; Davis, B. G. Selective chemical protein modification. *Nat. Commun.* **2014**, *5*:4740 and references cited therein.
3. Cal, P. M. S. D.; Bernardes, G. J. L.; Gois, P. M. P. Cysteine-Selective Reactions for Antibody Conjugation. *Angew. Chem. Int. Ed.* **2014**, *53*, 10585-10587 and references cited therein.
4. Stephanopoulos, N.; Francis, M. B. Choosing an effective protein bioconjugation strategy. *Nat. Chem. Biol.* **2011**, *7*, 876-884 and references cited therein.
5. Lotze, J.; Reinhardt, U.; Seitz, O.; Beck-Sickinger, A. G. Peptide-tags for site-specific protein labelling *in vitro* and *in vivo*. *Mol. BioSyst.* **2016**, *12*, 1731-1745 and references cited therein.
6. Griffin, B. A.; Adams, S. R.; Tsien, R. Y. Specific Covalent Labeling of Recombinant Protein Molecules Inside Live Cells. *Science* **1998**, *281*, 269-272.
7. Adams, S. R.; Campbell, R. E.; Gross, L. A.; Martin, B. R.; Walkup, G. K.; Yao, Y.; Llopis, J.; Tsien, R. Y. New Biarsenical Ligands and Tetracysteine Motifs for Protein Labeling in Vitro and in Vivo: Synthesis and Biological Applications. *J. Am. Chem. Soc.* **2002**, *124*, 6063-6076.
8. Zhang, C.; Welborn, M.; Zhu, T.; Yang, N. J.; Santos, M. S.; Van Voorhis, T.; Pentelute, B. L. π -Clamp-mediated cysteine conjugation. *Nat. Chem.* **2015**, *8*, 120-128.
9. Zhang, C.; Spokoiny, A. M.; Zou, Y.; Simon, M. D.; Pentelute, B. L. Enzymatic 'click' ligation: selective cysteine modification in polypeptides enabled by promiscuous glutathione S-transferase. *Angew. Chem. Int. Ed.* **2013**, *52*, 14001-14005.
10. Hacker, S. M.; Backus, K. M.; Lazear, M. R.; Forli, S.; Correia, B. E.; Cravatt, B. F. Global profiling of lysine reactivity and ligandability in the human proteome. *Nat. Chem.* **2017**, *9*, 1181-1190.
11. Matos, M. J.; Oliveira, B. L.; Martínez-Sáez, N.; Guerreiro, A.; Cal, P. M. S. D.; Bertoldo, J.; Maneiro, M.; Perkins, E.; Howard, J.; Deery, M. J.; Chalker, J. M.; Corzana, F.; Jiménez-Osés, G.; Bernardes, G. J. L. Chemo- and Regioselective Lysine Modification on Native Proteins. *J. Am. Chem. Soc.* **2018**, *140*, 4004-4017.
12. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. Synthesis of proteins by native chemical ligation. *Science* **1994**, *266*, 776-778.
13. Burke, H. M.; McSweeney, L.; Scanlan, E. M. Exploring chemoselective S-to-N acyl transfer reactions in synthesis and chemical biology. *Nat. Commun.* **2017**, *8*:15655 and references cited therein.

14. Steffen, W.; Ko, F. C.; Patel, J.; Lyamichev, V.; Albert, T. J.; Benz, J.; Rudolph, M. G.; Bergmann, F.; Streidl, T.; Kratzsch, P.; Boenitz-Dulat, M.; Oelschlaegel, T.; Schraeml, M. Discovery of a microbial transglutaminase enabling highly site-specific labeling of proteins. *J. Biol. Chem.* **2017**, *292*, 15622-15635.
15. White, E. H.; McCapra, F.; Field, G. F. Structure and Synthesis of Firefly Luciferin. *J. Am. Chem. Soc.* **1963**, *85*, 337-343.
16. Gomi, K.; Kajiyama, N. Oxyluciferin, a Luminescence Product of Firefly Luciferase, Is Enzymatically Regenerated into Luciferin. *J. Biol. Chem.* **2001**, *276*, 36508-36513.
17. Ren, H.; Xiao, F.; Zhan, K.; Kim, Y.-P.; Xie, H.; Xia, Z.; Rao, J. Biocompatible Condensation Reaction for the Labeling of Terminal Cysteine Residues on Proteins. *Angew. Chem. Int. Ed.* **2009**, *48*, 9658-9662.
18. Nguyen, D. P.; Elliott, T.; Holt, M.; Muir, T. M.; Chin, J. W. Genetically Encoded 1,2-Aminothiols Facilitate Rapid and Site-Selective Protein Labeling via a Bio-orthogonal Cyanobenzothiazole Condensation. *J. Am. Chem. Soc.* **2011**, *133*, 11418-11421.
19. Van de Bittner, G. C.; Bertozzi, C. R.; Chang, C. J. Strategy for Dual-Analyte Luciferin Imaging: *In Vivo* Bioluminescence Detection of Hydrogen Peroxide and Caspase Activity in a Murine Model of Acute Inflammation. *J. Am. Chem. Soc.* **2013**, *135*, 1783-1795.
20. Jeon, J.; Shen, B.; Xiong, L.; Miao, Z.; Lee, K. H.; Rao, J.; Chin, F. Efficient Method for Site-Specific ¹⁸F-Labeling of Biomolecules Using the Rapid Condensation Reaction between 2-Cyanobenzothiazole and Cysteine. *Bioconjugate Chem.* **2012**, *23*, 1902-1908.
21. Ramil, C. P.; An, P.; Yu, Z.; Lin, Q. Sequence-specific 2-cyanobenzothiazole ligation. *J. Am. Chem. Soc.* **2016**, *138*, 5499-5502.
22. de Almeida, G. Improving Tools for Bioorthogonal Chemistry. Ph.D. Dissertation [Online], University of California, Berkeley, Berkeley, CA, Spring 2014. http://digitalassets.lib.berkeley.edu/etd/ucb/text/deAlmeida_berkeley_0028E_14273.pdf (accessed May 29, 2017).
23. Sidhu, S. S.; Weiss, G. A.; Wells, J. A. High Copy Display of Large Proteins on Phage for Functional Selections. *J. Mol. Biol.* **2000**, *296*, 487-495.
24. Lamboy, J. A.; Arter, J. A.; Knopp, K. A.; Der, D.; Overstreet, C. M.; Palermo, E.; Urakami, H.; Yu, T.-B.; Tezgel, O.; Tew, G.; Guan, Z.; Kuroda, K.; Weiss, G. A. Phage Wrapping with Cationic Polymers Eliminates Non-specific Binding between M13 Phage and High pI Target Proteins. *J. Am. Chem. Soc.* **2009**, *131*, 16454-16460.
25. Eldridge, G. M.; Weiss, G. A. Hydrazide Reaction Peptide Tags for Site-Specific Protein Labeling. *Bioconjugate Chem.* **2011**, *22*, 2143-2153.
26. Thévenet, P.; Shen, Y.; Maupetit, J.; Guyon, F.; Derreumaux, P.; Tufféry, P. PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic Acids Res.* **2012**, *40*, W288-293.

27. Shen, Y.; Maupetit, J.; Derreumaux, P.; Tufféry, P. Improved PEP-FOLD approach for peptide and miniprotein structure prediction. *J. Chem. Theor. Comput.* **2014**, *10*, 4745-4758.
28. Sluyterman, L. A. Æ.; Wijdenes, J. Benzoylamidoacetonitrile as an inhibitor of papain. *Biochim. Biophys. Acta* **1973**, *302*, 95-101.
29. Brisson, J.-R.; Carey, P. R.; Storer, A. C. Benzoylamidoacetonitrile Is Bound as a Thioimidate in the Active Site of Papain. *J. Biol. Chem.* **1986**, *261*, 9087-9089.
30. Kawakami, T.; Ogawa, K.; Goshima, N.; Natsume, T. DIVERSE System: De Novo Creation of Peptide Tags for Non-enzymatic Covalent Labeling by In Vitro Evolution for Protein Imaging Inside Living Cells. *Chem. Biol.* **2015**, *22*, 1671-1679.
31. Strohal, M.; Hassman, M.; Košata, B.; Kодиček, M. mMass Data Miner: an Open Source Alternative for Mass Spectrometric Data Analysis. *Rapid Commun. Mass Spec.* **2008**, *22*, 905-908.
32. Strohal, M.; Kavan, D.; Novák, P.; Volný, M.; Havlíček, V. mMass 3: A Cross-Platform Software Environment for Precise Analysis of Mass Spectrometric Data. *Anal. Chem.* **2010**, *82*, 4648-4651.
33. Niedermeyer, T. H. J.; Strohal, M. mMass as a Software Tool for the Annotation of Cyclic Peptide Tandem Mass Spectra. *PLoS ONE* **2012**, *7*, e44913.
34. White, E. H.; Wörther, H.; Seliger, H. H.; McElroy, W. D. Amino Analogs of Firefly Luciferin and Biological Activity Thereof. *J. Am. Chem. Soc.* **1966**, *88*, 2015-2019.
35. Sharma, D. K.; Adams, S. T., Jr.; Liebmann, K. L.; Miller, S. C. Rapid Access to a Broad Range of 6'-Substituted Firefly Luciferin Analogues Reveals Surprising Emitters and Inhibitors. *Org. Lett.* **2017**, *19*, 5836-5839.
36. Kaiser, E.; Colecott, R. L.; Bossinger, C. D.; Cook, P. I. Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides. *Anal. Biochem.* **1970**, *34*, 595-598.

Chapter 3

Toward a method for site-selective protein modification based on a peptide motif that reacts with cyanobenzothiazoles via internal cysteine-lysine relay

Chapter 3. Toward a method for site-selective protein modification based on a peptide motif that reacts with cyanobenzothiazoles via internal cysteine-lysine relay

INTRODUCTION

Site-specific protein labeling is central to the field of chemical biology. One common approach to attaining site-selectivity is to modify a protein to include a peptide sequence or unnatural amino acid to serve as a target for functionalization. In the past two decades, scientists have developed various genetically encodable protein tagging methods, including green fluorescent protein (GFP) fusions, unnatural amino acid incorporation, small molecule-binding peptide tags, metal-chelating peptide tags, and enzyme-substrate peptide tags, all of which were introduced in Chapter 1.¹⁻¹⁰ Researchers have used these methods to image targets *in vitro* and *in vivo*, create new therapeutics, and install unnatural functionality on proteins to probe their activity; however, there remains a need for an *in vivo* labeling method that minimally perturbs the natural localization and function of the protein of interest, allows spatiotemporally controlled imaging, avoids off-target labeling, and requires minimal engineering. Additionally, techniques to label proteins at internal sites as opposed to the N- or C-terminus are currently limited yet would have important applications in studying proteins with less accessible termini, e.g., certain transmembrane proteins. Finally, if the reaction is to be used *in vivo*, it should be applicable on a reasonably short timescale and must be bioorthogonal, i.e., it must proceed in aqueous conditions at physiological pH and temperature without being toxic to the host cell.

In the previous chapter, we reported the rational design of CBTag 1.0, an 11-mer peptide motif that reacts with a 2-cyanobenzothiazole (CBT) moiety via internal cysteine-lysine transfer. The reaction between CBT derivatives and 1,2-aminothiols has been adapted into a bioorthogonal protein labeling method by several different groups. Rao and coworkers provided the initial report on its utility *in vitro*, demonstrating that an engineered N-terminal cysteine as well as an unnatural C-terminal 1,2-aminothiol can rapidly ligate the CBT functional group in a condensation reaction under physiological conditions (Figure 3.1A).¹¹ Soon after, Chin and colleagues demonstrated that unnatural amino acid incorporation can be used to introduce a 1,2-aminothiol into a protein for reaction with CBT (Figure 3.1B).¹² Unfortunately, they discovered that a protected version of the 1,2-aminothiol was necessary, as the free 1,2-aminothiol formed nonproductive adducts with endogenous metabolites such as pyruvate. In order to deprotect the 1,2-aminothiol, the authors used 200 mM *O*-methylhydroxylamine in 6 M guanidinium chloride at pH 4 for four hours. These harsh conditions make this method acceptable for *in vitro* functionalization of proteins but not for studies in living cells.

Recently, Lin and coworkers used phage display to select for a genetically encodable peptide tag that reacts with a biotin-PEG₄-CBT derivative.¹³ Unlike the previous studies, this peptide tag does not ligate CBT via a condensation with 1,2-aminothiols; instead,

reaction occurs with a pair of cysteines surrounded by a sequence that provides a favorable microenvironment for the formation of a thioimide bond (Figure 3.1C). The researchers were able to use this sequence to label the outer loop of a transmembrane protein by first incubating with their biotin-PEG₄-CBT derivative, then treating the cells with a streptavidin-fluorophore conjugate. Streptavidin has extremely high affinity for biotin, but it is large (52.8 kDa), making this system poorly suited for certain applications. Unfortunately, the authors were constrained to this two-step labeling method because when they attempted to replace the biotin group with a fluorophore, the reaction no longer occurred, suggesting that their peptide's reactivity is dependent on the nature of the substituents on the CBT core.

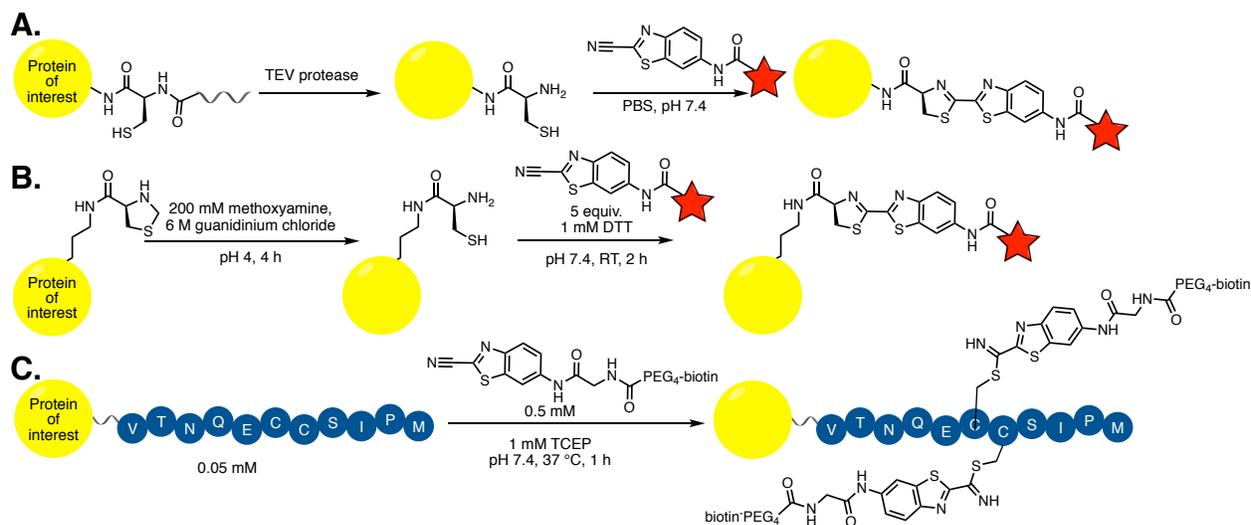


Figure 3.1. Previous approaches to utilizing CBT derivatives in protein labeling experiments. **A)** A protein was engineered such that an N-terminal cysteine was exposed after Tobacco Etch Virus (TEV) protease treatment, and that cysteine was then reacted with a CBT derivative. **B)** An unnatural amino acid containing a protected 1,2-aminothiols was incorporated into a protein. After harsh deprotection conditions, the 1,2-aminothiols could be reacted with a CBT derivative. **C)** A peptide tag selected through phage display reacts with one or two equivalents of a biotin-functionalized CBT.

In this chapter, we describe efforts toward further characterization and improvement of our peptide motif from Chapter 2. We began by assessing the degree to which the peptide reacts with thioesters because we had established that the mechanism of covalent bond formation between *N*-(2-(cyano-¹³C)-benzo[*d*]thiazol-6-yl)acetamide (CBTNAc, compound **1**) and CBTag 1.0 is similar to that of native chemical ligation (NCL), in which an N-terminal cysteine thiol undergoes a transthioesterification reaction with a thioester then is displaced by the N-terminal amine to form a new amide bond. We next synthesized different fluorophore- and azide-functionalized CBT probes and characterized their reactivity with CBTag 1.0. We also created six superfolder GFP (sfGFP) mutants with CBTag 1.0 inserted at one of the termini or loop regions in order to determine whether CBTag 1.0 still reacts with CBT derivatives after incorporation into a protein. Finally, we used the knowledge gained from CBTag 1.0 to begin an mRNA display-based approach

to finding an optimized peptide motif for the ligation of CBT via internal cysteine-lysine transfer.

RESULTS AND DISCUSSION

Since we demonstrated in Chapter 2 that the mechanism of covalent bond formation between CBTNAc and CBTtag 1.0 resembles that of NCL, we tested CBTtag 1.0 for thioester reactivity. We first synthesized a 5-residue peptide with the sequence Ala-Leu-Pro-Phe-Gly, an N-terminal acetyl group, and a C-terminal benzyl thioester. The sequence was chosen for simplicity and ease of synthesis; none of the residue side chains were β -branched, nor did they require protecting groups. Additionally, the glycine residue was placed at the C-terminus to minimize potential unfavorable interactions between CBTtag 1.0 and a bulky side chain on the NCL peptide. We then assessed reaction with CBTtag 1.0 using the assays described in Chapter 2, with this peptide in place of CBTNAc. By MALDI-TOF MS, CBTtag 1.0 reacted with at least one equivalent of the thioester in a significant amount before cysteine treatment, but cysteine treatment left only a minor peak corresponding to what was assumed to be the lysine adduct (Figure 3.2). These results were confirmed using a modified version of our established LC-MS method; only approximately 2.4% of the area under CBTtag 1.0-containing peaks corresponded to the lysine adduct after cysteine treatment (Supplementary Figure 3.S1, 3.S2).

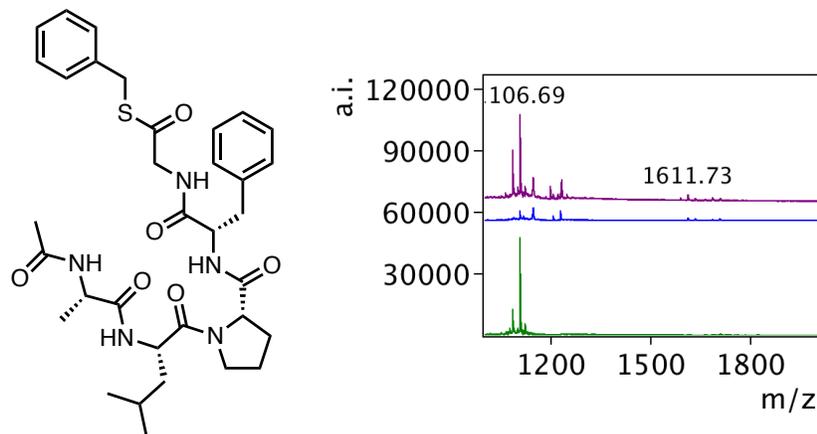


Figure 3.2. MALDI-TOF MS shows that CBTtag 1.0 only minimally reacts with a peptide bearing a C-terminal benzyl thioester (NCL peptide, structure on left). Green (bottom) trace: CBTtag 1.0 alone; blue (middle) trace: CBTtag 1.0 treated with 1.1 equiv. of the NCL peptide for 48 h, then with 10 equiv. DTT for 48 h; purple (top) trace: CBTtag 1.0 treated with 1.1 equiv. of the NCL peptide for 48 h, then with 10 equiv. cysteine and 10 equiv. DTT for 48 h. [CBTtag 1.0 + 2Na]⁺ 1106.69 m/z; [CBTtag 1.0 + NCL peptide - benzyl mercaptan + Na]⁺ 1611.73 m/z. Arbitrary intensity (a.i.).

Believing the low conversion to product may have been due to the bulk of the peptide and its corresponding effect on the conformation of CBTag 1.0 after reaction with the cysteine residue, we next chose to look at the reaction of CBTag 1.0 with *S*-(4-nitrophenyl) thioacetate. Not only is *S*-(4-nitrophenyl)thioacetate a commercially available activated thioester, but also the acetyl modification is small enough to minimize perturbation to the conformation of CBTag 1.0. In initial experiments, MALDI-TOF MS showed peaks corresponding to the mono- and di-acetylated CBTag 1.0 (Figure 3.3). We proceeded to analyze the reaction using our LC-MS method. The data confirmed the formation of the mono- and di-acetylated CBTag 1.0; it also showed that the lysine to arginine mutant was mono-acetylated and the cysteine to serine mutant was unmodified (Supplementary Figure 3.S3-3.S5). However, we observed only one peak corresponding to monoacetylated CBTag 1.0 rather than one corresponding to modification of the cysteine residue and one corresponding to modification of the lysine residue, perhaps because reaction with the cysteine occurs rapidly enough to prevent a signal originating from a monoacetylated lysine adduct. Chloroacetate treatment also gave inconclusive results, complicating analysis of conversion rates. Regardless, the lack of modification of the cysteine to serine mutant indicates that the lysine alone is not able to react with the small molecule thioester, providing indirect evidence that the mechanism of formation of the di-acetylated peptide requires an *S*→*N* acyl transfer.

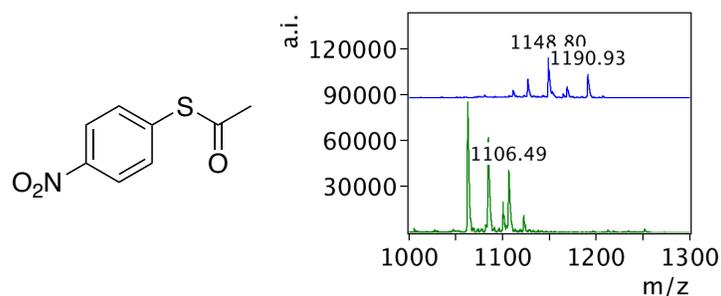


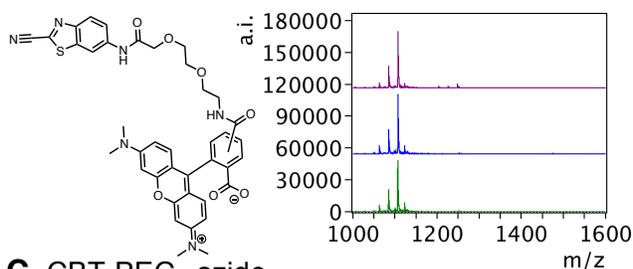
Figure 3.3. MALDI-TOF MS shows that CBTag 1.0 is mono- and di-acetylated by *S*-(4-nitrophenyl)-thioacetate (structure on left). Green (bottom) trace: CBTag 1.0 alone. Blue (top) trace: CBTag 1.0 treated with 1.1 equiv. *S*-(4-nitrophenyl)-thioacetate. [CBTag 1.0 + 2Na]⁺ 1106.49 m/z; [CBTag 1.0 + acetyl + 2Na]⁺ 1148.80 m/z; [CBTag 1.0 + 2(acetyl) + 2Na]⁺ Arbitrary intensity (a.i.).

Although we did observe labeling of CBTag 1.0 with the activated thioester, we were pleased in hindsight that only minimal reaction with the NCL peptide occurs. A recent study from James et al. reported that proximal cysteine residues mediate non-enzymatic *N*-acetylation of lysine residues by acetyl coenzyme A (acetyl-CoA).¹⁴ Acetyl-CoA is an essential co-factor for enzymatic lysine acetylation and has been reported at concentrations of 20-600 μ M in *Escherichia coli* cells.¹⁵ If the lysine residue in CBTag 1.0 had been rapidly and extensively modified by the NCL peptide, it may also have been subject to acetylation by endogenous acetyl-CoA in cells, preventing the desired CBT ligation. We decided to move ahead with development of new CBT probes and assess

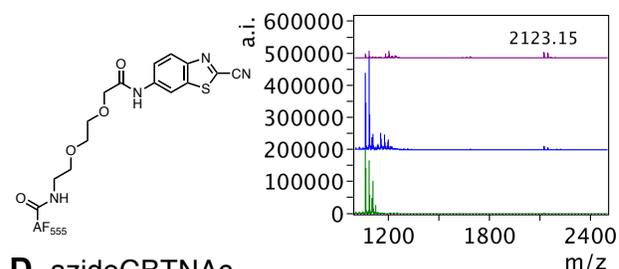
the level of acetyl-CoA side reactivity only after we had determined whether CBTag 1.0 would still react with CBT derivatives once incorporated into a protein.

With future protein labeling experiments in mind, our next goal was to synthesize and characterize a CBT derivative that had a probe or a functional group by which a probe could be attached. We first tried attaching 5(6)-carboxytetramethylrhodamine (5(6)-TAMRA) and AlexaFluor 555 to CBT with a mini-PEG linker. These dyes were chosen on the basis that they could be used in future fluorescence resonance energy transfer (FRET) experiments with engineered GFP constructs. However, the TAMRA conjugate aggregated to a large extent in aqueous solution and did not appreciably bind to CBTag 1.0 (Figure 3.4A). The second dye, Alexa Fluor 555, was chosen for its high water solubility and similar fluorescence properties to TAMRA. While this CBT-fluorophore conjugate was highly water soluble, it too did not appear to bind strongly to CBTag 1.0 by MALDI-TOF MS, though a small amount of peptide-CBT-fluorophore adduct did appear (Figure 3.4B). Our attempts to rationalize this behavior was complicated by the fact that the structure of Alexa Fluor 555 is not published. We hypothesized that the charges on Alexa Fluor 555 (commercially available as a tris(triethylammonium) salt,¹⁶ implying three negative charges) interfere in the reaction with the peptide, either by repelling certain residues (e.g., the aspartic acid) or by interacting with key residues (e.g., the lysine residue) and blocking the desired reaction with the nitrile group of CBT. These results led us to redirect our focus toward addition of an azide handle for later derivatization via well-established bioorthogonal reactions such as Cu^I-catalyzed azide-alkyne cycloaddition (CuAAC), strain-promoted azide-alkyne cycloaddition (SPAAC), and the Staudinger ligation, described in Chapter 1. The resulting compounds, CBT-PEG₄-azide and *N*-(2-azidoacetyl)-CBT (azidoCBTNAC), were confirmed to bind CBTag 1.0, though the probes did not label as extensively as CBTNAC (Figure 3.4C, D).

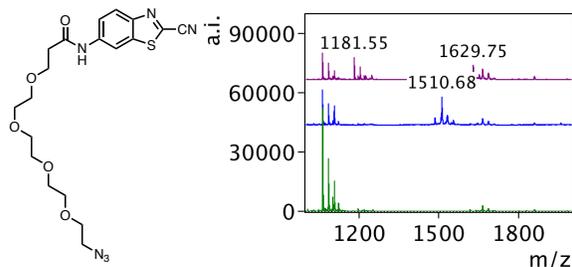
A. CBT-miniPEG-TAMRA



B. CBT-miniPEG-Alexa Fluor 555



C. CBT-PEG₄-azide



D. azidoCBTNAC

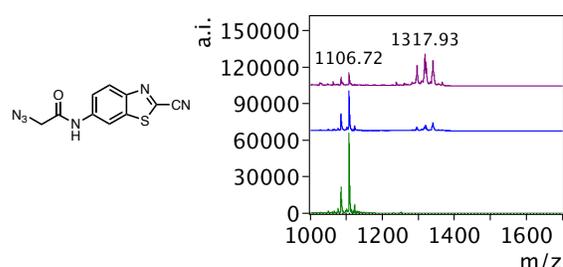
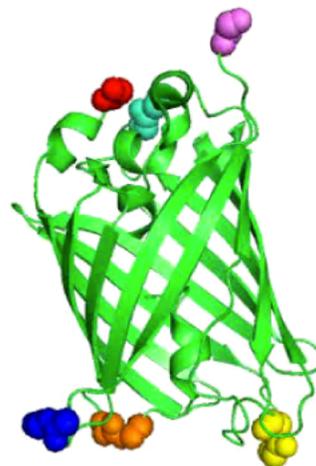


Figure 3.4. CBTag 1.0 can react with some CBT probes other than CBTNAC. Structures of probes (where known) are to the left of their corresponding spectra. In all MALDI-TOF

MS spectra, green (bottom) trace is CBTag 1.0 alone; blue (middle) is CBTag 1.0 treated with 1.1 equiv. probe for 48 h, then 10 equiv. DTT for 48 h; purple (top) is CBTag 1.0 treated with 1.1 equiv. probe for 48 h, then 10 equiv. cysteine and 10 equiv. DTT for 48 h. Arbitrary intensity (a.i.). **A)** CBTag 1.0 does not react with CBT-miniPEG-TAMRA. **B)** CBTag 1.0 only minimally reacts with CBT-miniPEG-Alexa Fluor 555, but the modification is retained through cysteine treatment. [CBTag 1.0-CBT-miniPEG-Alexa Fluor 555 + H]⁺ 2123.15 m/z. **C)** CBTag 1.0 can form a stable adduct with CBT-PEG₄-azide. [CBTag 1.0-CBT-PEG₄-azide + H]⁺ 1510.68 m/z; [CBTag 1.0-Cys + H]⁺ 1181.55 m/z; [CBTag 1.0-CBT-PEG₄-azide-Cys + H]⁺ 1629.75 m/z. Incomplete reduction by DTT led to formation of cys adducts where noted. **D)** CBTag 1.0 forms a stable adduct with azidoCBTNAc. [CBTag 1.0 + 2Na]⁺ 1106.72 m/z; [CBTag 1.0-azidoCBTNAc + H]⁺ 1317.93 m/z.

With these probes in hand, we designed and expressed His₆-tagged sfGFP constructs containing CBTag 1.0 at the N-terminus, C-terminus, or an internal site (see Materials and Methods section for details). We also expressed wildtype sfGFP with a C-terminal His₆ tag for control experiments. Aside from the termini, the CBTag 1.0 peptide was inserted into four random coil regions of the GFP crystal structure (PDB 2B3P) to minimize perturbation of the protein's native conformation (Figure 3.5). Once the sfGFP constructs were expressed and purified (Figure 3.6), we calculated the percent mature (i.e., fluorescent) sfGFP (Table 3.1). The wildtype sample was 54% mature sfGFP. We were pleased to find that for five out of six of the mutant sfGFP constructs, between 48% and 52% of the sfGFP was in the mature form, indicating that our inserts did not significantly inhibit folding of sfGFP. For the final mutant construct, 37% of the sfGFP was in the mature form, which though not ideal was enough for our purposes.

Figure 3.5. Sites within sfGFP where the CBTag 1.0 peptide was introduced are highlighted. Red – N-terminal (sfGFP4); Orange – Loop 3-interhelix (sfGFP8); Yellow – Loop 5-6 (sfGFP7); Green-Cyan – Loop 9-10 (sfGFP6); Blue – Loop 10-11 (sfGFP5); Purple – C-terminal insert (sfGFP9). Wildtype sfGFP was also expressed. Crystal structure: PDB 2B3P.



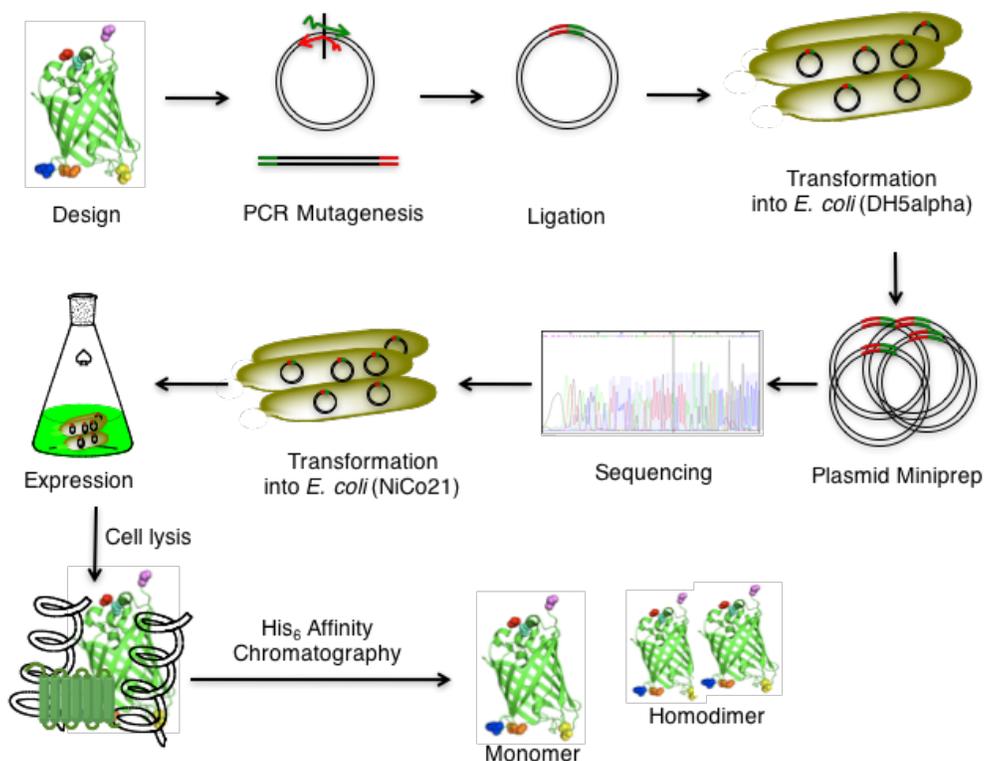


Figure 3.6. sGFP construct design and expression workflow. Site-directed mutagenesis was used to clone CBTtag 1.0 into His₆-tagged sfGFP. Protein was expressed in *E. coli* and purified by batch Ni-NTA affinity chromatography.

Table 3.1. Percentage of sfGFP in the active (fluorescent) form for each construct. For calculation, see Materials and Methods.

| sfGFP Construct | % Mature sfGFP |
|------------------------------|----------------|
| Wildtype | 54 |
| N-terminal CBTtag 1.0 | 49 |
| Loop 3-ih CBTtag 1.0 insert | 52 |
| Loop 6-7 CBTtag 1.0 insert | 49 |
| Loop 9-10 CBTtag 1.0 insert | 52 |
| Loop 10-11 CBTtag 1.0 insert | 37 |
| C-terminal CBTtag 1.0 | 48 |

We then began to characterize the degree to which various CBT probes modified the tagged proteins. Unfortunately, it was discovered that our MALDI-TOF MS assay was not as useful for analysis of intact proteins, as the wider isotopic envelope at higher masses made it difficult to ascertain whether a comparatively small modification such as CBTNAc

was present (Figure 3.7, 3.8A, B). The dye-CBT conjugate results were also inconclusive, likely because the lack of a wash or other purification step before MALDI-TOF MS analysis allowed the proteins to retain multiple dye molecules via nonspecific hydrophobic or charge-charge interactions. For example, it was difficult to determine whether the mass of the CBT-miniPEG-Alexa Fluor 555-treated sfGFP shifted further than that of Alexa Fluor 555 carboxylic acid-treated sfGFP (Figure 3.8C). We considered subjecting our proteins to digestion and using other MS techniques to determine whether the probes were selectively ligated to the lysine residue of CBTtag 1.0, but since we had previously established that formation of the lysine-CBTNac amidine bond is fairly slow (Chapter 2), we decided to attempt to optimize the CBTtag 1.0 peptide instead.

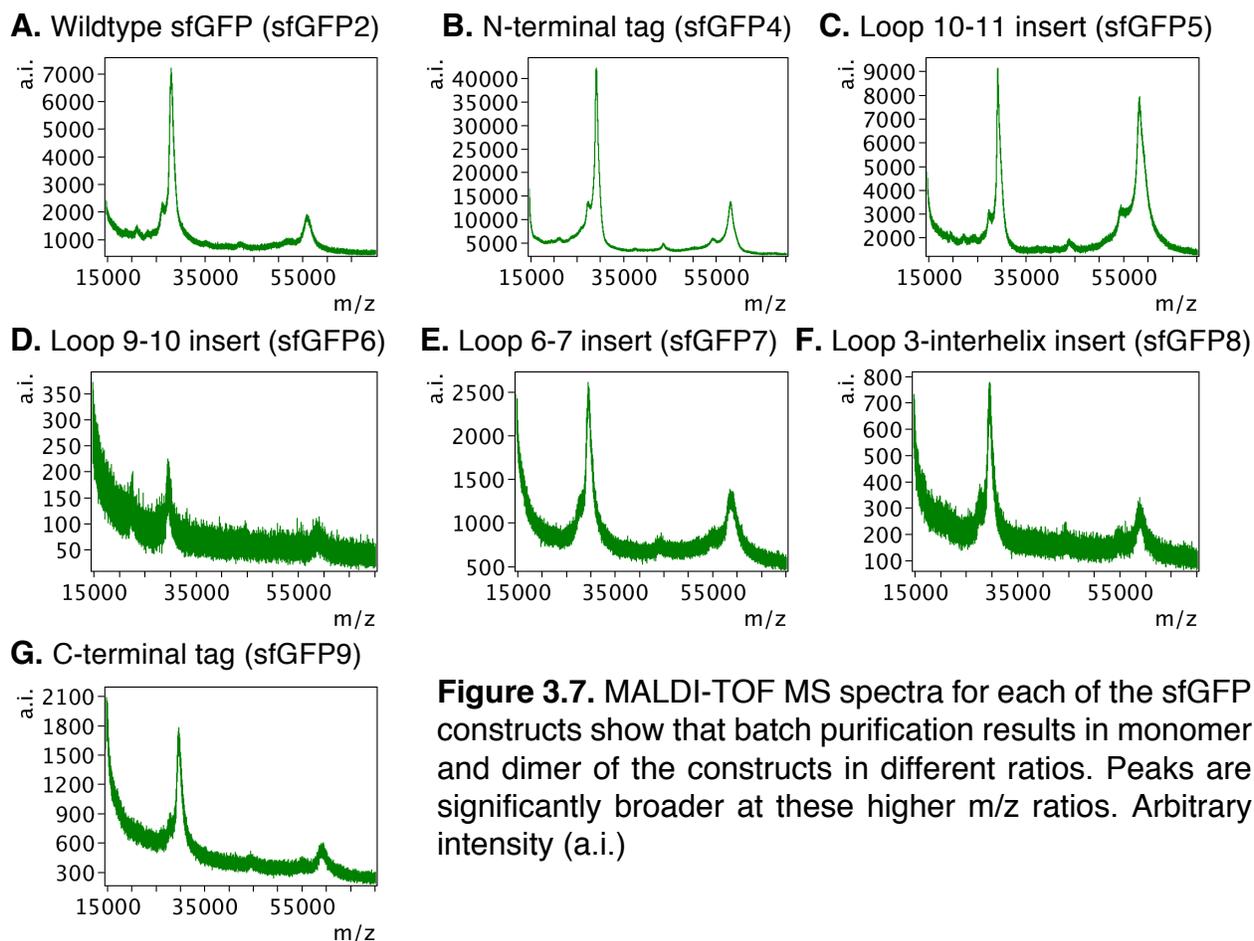


Figure 3.7. MALDI-TOF MS spectra for each of the sfGFP constructs show that batch purification results in monomer and dimer of the constructs in different ratios. Peaks are significantly broader at these higher m/z ratios. Arbitrary intensity (a.i.)

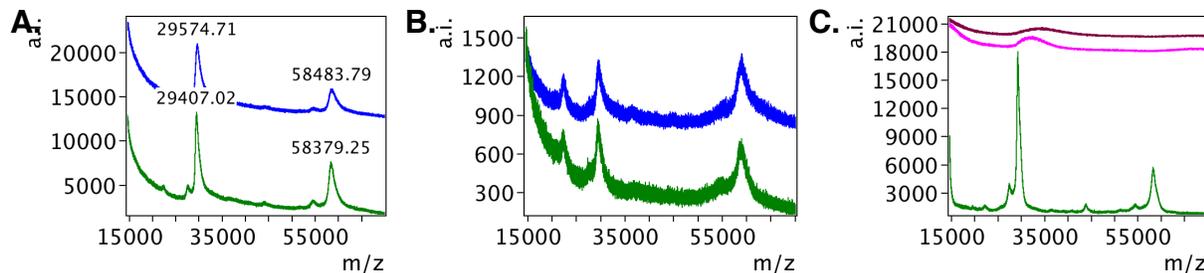


Figure 3.8. MALDI-TOF MS data for reaction of a representative sfGFP construct with different CBT probes are inconclusive. sfGFP6 (loop 9-10 CBTag 1.0 insert; green trace) was incubated for 48 h with 1.1 equiv. of **A**) CBTNAc (blue), **B**) CBT-PEG₄-azide (blue), **C**) CBT-miniPEG-Alexa Fluor 555 (maroon, top) or Alexa Fluor 555 carboxylic acid (pink, middle). Arbitrary intensity (a.i.).

To improve the properties of our peptide motif, we chose to adapt the Directed *in Vitro* Evolution of Reactive peptide tags via Sequential Enrichment (DIVERSE) method reported by Kawakami and coworkers to our needs (Figure 3.9).¹⁷ We began with a library of DNA-encoding primers on either side of an insert which would become a sequence of 13-15 amino acids after translation. Some of these inserts were fully randomized while others were designed to contain Cys-Lys, Lys-Cys, Cys-X-Lys, or Lys-X-Cys motifs. We chose this positioning of the cysteine and lysine residues based on the knowledge gained during the course of the CBTag 1.0 design, described in detail in Chapter 2. This DNA library was PCR-amplified then transcribed, translated, and reverse transcribed using a cell-free release factor-free *in vitro* system and reverse transcriptase in combination with a puromycin-modified oligonucleotide that kept the translated peptide attached to its cDNA. Before proceeding to the first round of selection, the solution containing the library of peptide-cDNA conjugates was divided equally. One half was left untouched; the other was pretreated with tris(2-carboxyethyl)phosphine (TCEP) to ensure any cysteines were reduced. These untreated and TCEP-treated conjugates were separately subjected to the positive selection conditions, i.e., incubation with CBT-functionalized magnetic beads. After washing away non-covalently bound peptides, the cDNA of the remaining peptides was eluted and used in the next round of selection, which also included a negative selection in which unfunctionalized beads were used to remove peptides that were binding the bead instead of reacting with CBT. The untreated and TCEP-treated conjugates were subjected to 13 rounds of selection in parallel, then blunt end TOPO cloning was applied to the crude PCR products to enable expression in *E. coli*. 200 colonies were picked for Sanger sequencing, 100 each from the untreated and TCEP-treated candidate pools, and the peptide sequences were inferred from the DNA sequence (Materials and Methods, Table 3.4).

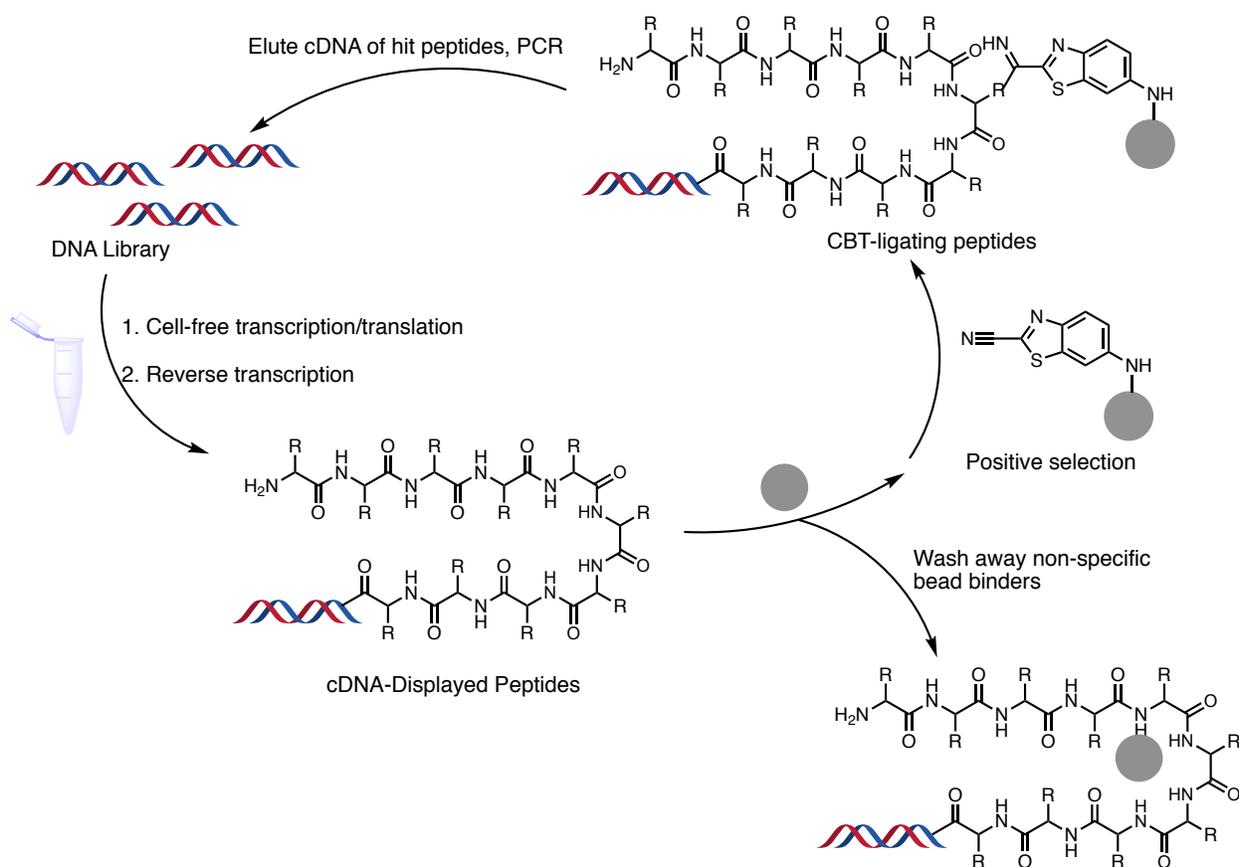


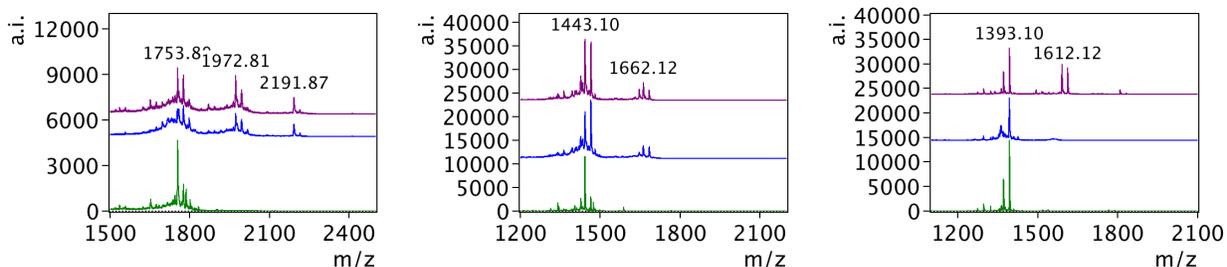
Figure 3.9. DIVERSE workflow. A cell-free transcription/translation system and a reverse transcriptase are used to convert a DNA library to a library of cDNA-displayed peptides. A negative selection is performed to remove bead-binding peptides, then a positive selection with CBT-functionalized beads is performed. After washing away the unreactive peptides, cDNA of the candidate peptides is eluted from the beads directly into a PCR reaction mixture and amplified to yield a new DNA library.

Of these, approximately 24 different sequences with the proper reading frame and inserts were found, and 21 of those contained a Cys-Lys, Lys-Cys, Cys-X-Lys, or Lys-X-Cys motif. Many of the other “hits” did not contain an insert at all; due to a frame-shift mutation, they translated to variations on Val-Glu-Glu-Glu-Val-X-X-Arg-X-X, which we suspected to be a bead-binding rather than a CBT-ligating motif based on the absence of cysteine and lysine residues. In the 24 sequences that did have inserts, the quality of the sequencing data varied greatly, resulting in peptides that had multiple potential amino acid sequences. Additionally, we unexpectedly obtained sequences that contained stop codons. We suspect that the absence of release factors in the cell-free translation system may have allowed increased stop codon readthrough, resulting in incorporation of glutamine, tyrosine, or lysine at UAA and UAG codons and cysteine, tryptophan, or arginine at UGA codons.¹⁸ To determine the most likely amino acid sequences of the peptide hits, we decided to use the PEP-FOLD website¹⁹⁻²⁰ and the Maestro MacroModel software

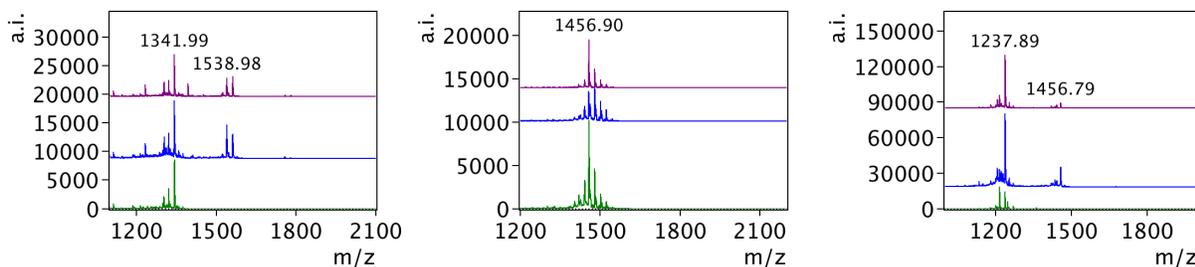
described in Chapter 2 to model each potential peptide. Based on these results, we chose 14 of the most likely candidates for further study.

Thirteen of these peptides were subjected to the CBTNAc binding assay and cysteine competition experiments described in Chapter 2 (Figure 3.10). (The 14th was lost during purification; as it had no cysteine residues but rather a suspected bead-binding sequence, we proceeded without it.) Somewhat unexpectedly, while the majority of the sequences did rapidly react with CBTNAc, the loss of a significant portion of these modifications on treatment with cysteine showed that none of the newer peptides were superior to CBTag 1.0 in terms of forming a stable bond with CBTNAc.

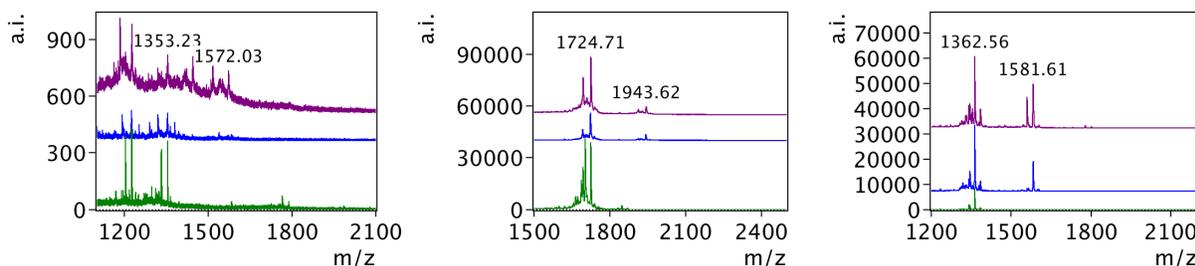
A. Ac-RLTPWKSCMCDWW-NH₂ **B.** Ac-PPWRTKCRAASQ-NH₂ **C.** Ac-VVPPKSCVHRHA-NH₂
 1753 g/mol 1442 g/mol 1371 g/mol



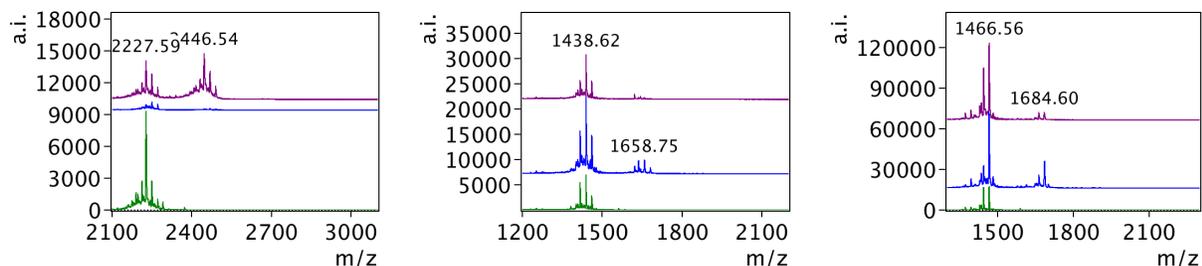
D. Ac-PTRAAKCASTWS-NH₂ **E.** Ac-VEEEVARTGGRR-NH₂ **F.** Ac-PWRARCKGSGG-NH₂
 1319 g/mol 1457 g/mol 1215 g/mol



G. Ac-GSGFPCRRRPGK-NH₂ **H.** Ac-WCPWFKACRHVAR-NH₂ **I.** Ac-ALGTHKCGQQEQ-NH₂
 1331 g/mol 1701 g/mol 1340 g/mol



J. Ac-SYGPRK**L**CGHVEWRRR-NH₂ **K.** Ac-SWGRV**C**KRGEVV-NH₂ **L.** Ac-RVWPG**K**CAAMQR-NH₂
 2228 g/mol 1417 g/mol 1444 g/mol



M. Ac-RRWRF**K**CAPLVQ-NH₂
 2228 g/mol

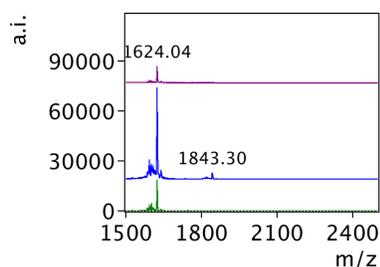


Figure 3.10. MALDI-TOF MS analysis shows that none of the DIVERSE hit peptides formed more of a stable adduct than CBTag 1.0. Cys and Lys residues are bolded. Green (bottom) trace: peptide alone; blue (middle) trace: peptide treated with 1.1 equiv. of CBTNAC for 48 h, then with 10 equiv. DTT for 48 h; purple (top) trace: peptide treated with 1.1 equiv. of CBTNAC for 48 h, then with 10 equiv. cysteine and 10 equiv. DTT for 48 h. Molecular weights are listed under each peptide sequence. Masses that are labeled correspond to [peptide + H/Na/2Na]⁺, [peptide-CBTNAC + H/Na/2Na]⁺, or [peptide-2(CBTNAC) + H/Na/2Na]⁺. Arbitrary intensity (a.i.)

CONCLUSIONS

In this chapter, we have further characterized the substrate scope of CBTag 1.0 and outlined initial efforts to optimize CBTag 1.0 for use in protein labeling. We have shown that although CBTag 1.0 ligates CBTNAC best, some degree of reaction with the azide-functionalized CBT derivatives still occurs. With an optimized peptide sequence, this opens the possibility of introducing a fluorophore or other labeling reagent by way of a bioorthogonal reaction rather than by incubation with a large streptavidin conjugate.

Our initial attempt to use the DIVERSE system did not result in an improved version of CBTag 1.0, leading us to reconsider our selection conditions. In the future, after the positive selection step, the beads will be treated with free cysteine. Following the logic described in Chapter 2, this should leave only stable, irreversible peptide-CBT adducts in the candidate pool. Although we did not obtain an optimized sequence for protein labeling, our selection experiment results highlight the importance of carefully choosing selection conditions in order to remove peptides that undergo rapid but reversible reactions with CBT and clearly confirm our Chapter 2 conclusion that the presence of cysteine and lysine residues alone is insufficient for formation of a stable amidine bond via internal cysteine-lysine transfer of the CBT residue.

MATERIALS AND METHODS

General. Reactions were performed in flame- or oven-dried glassware under an inert nitrogen atmosphere unless otherwise noted. Anhydrous solvents were either purchased or obtained by passing solvent through an activated alumina column via a Pure Process Technology Glass Contour Solvent Purification System. All reagents and solvents were used as received unless otherwise noted. Water was passed through a Milli-Q filtration system prior to use. Where noted, samples were concentrated *in vacuo* at 40 °C using a BÜCHI Rotavapor R-114 equipped with a BÜCHI B-480 heating bath and a Welch Self-Cleaning Dry Vacuum System (Model 2025) or an IKA RV 10 basic rotary evaporator equipped with an IKA HB 10 basic heating bath and a Welch Self-Cleaning Dry Vacuum System (Model 2025). If necessary, compounds were then further dried under high vacuum using an Edwards RV8 Two Stage Rotary Vane Pump or by lyophilization in a LABCONCO FreeZone 4.5Plus.

Thin layer chromatography was performed using SiliCycle SiliaPlate glass-backed silica gel plates containing a fluorescent indicator (Fisher Scientific 50964470). Plates were visualized using a UVGL-25 Compact UV Lamp, 254/365 nm, 4 W (P/N 95-0021-12). For flash column chromatography, the stationary phase was SiliCycle SiliaFlash P60 or Fisher Silica Gel Sorbent (230-400 Mesh, Grade 60) silica gel. For purifications involving preparative or semi-preparative Reversed-Phase High-Performance Liquid Chromatography (RP-HPLC), the following conditions were used: the instrument consisted of an Agilent Technologies ProStar 325 UV-Vis detector, two PrepStar Solvent Delivery Modules, and a 440-LC Fraction Collector; the column was a Varian Microsorb 100Å C18, 8 µm, 21.4 x 250 mm Dynamax preparative column (R0080220C8) equipped with a Microsorb 100Å C18, 8 µm guard column (R0080220G8). Solvent A was 0.1% TFA in Milli-Q water; solvent B was 0.1% TFA in acetonitrile (MeCN). The UV-Vis detector was used to monitor wavelengths at 210 and 254 nm. All pure compounds and stock solutions were stored at -20 °C.

Liquid Chromatography-Mass Spectrometry (LC-MS) experiments were performed on an Agilent Technologies 1260 Infinity attached to a 6120 Quadrupole MS and a Peak Scientific NM32LA nitrogen generator. An Agilent InfinityLab Poroshell 120 EC-C18, 2.7 µm, 4.6 x 50 mm analytical LC column (part number 699975-902) was used. As with the HPLC, solvent A was 0.1% TFA in Milli-Q water; solvent B was 0.1% TFA in MeCN. Wavelengths of 210 nm, 254 nm, and 280 nm were monitored using the diode array detector. For ESI-MS, the 105-2000 *m/z* range was monitored in both positive and negative ion mode. Data was processed and analyzed using LC/MSD ChemStation (Agilent Technologies, Rev. B.04.03[16]). High resolution mass spectrometry (HRMS) data were acquired by ESI-LC/MS on a Waters Acquity UPLC and Thermo Exactive Orbitrap mass spectrometer by Dr. Theresa McLaughlin at the Stanford University Mass Spectrometry facility.

MALDI-TOF MS spectra were obtained on a Bruker microFlex MALDI-TOF (S/N 256969.00028) using a microScout Target MSP 96 target polished steel BC plate (P/N

8280800) and the Bruker Daltonics FlexControl software (Vincent Coates Foundation Mass Spectrometry Laboratory, Stanford University). For peptides, the matrix was α -cyano-4-hydroxycinnamic acid (CHCA) dissolved in 50% MeCN in 0.1% TFA. Plates were spotted with a mixture of 0.5 μ L sample and 0.5 μ L matrix. For intact proteins, 0.5 μ L of a saturated ethanolic sinapinic acid (SA) solution was spotted on the plate and allowed to dry, then a mixture of 0.5 μ L sample and 0.5 μ L of sinapinic acid dissolved in 30/70 MeCN/water with 0.1% TFA was spotted on the plate. Peptide spectra were taken on the Bruker microFlex using the following parameters: reflectron mode, positive ion detection; pulsed nitrogen laser (337 nm, 3 ns pulse width, pulse energy 150 μ J, 60 Hz); detector gain (reflector) 4.0x, 1810 V; sample rate 2.00 GS/s; laser power percentage was adjusted on a sample-by-sample basis to obtain an arbitrary intensity of 10^3 - 10^4 and a flat baseline; mass range 0-4000 Da; spectrum acquisition – random walk partial sample mode, results were the sum of at least 3 x 100 laser shots; calibration of the system was performed using the Bruker Peptide Calibration Standard mixture (part number 8206195). Intact protein spectra were taken using the following parameters: linear mode, positive ion detection; pulsed nitrogen laser (337 nm, 3 ns pulse width, pulse energy 150 μ J, 60 Hz); detector gain (linear) 4.0x, 2810 V; sample rate 1.00 GS/s; laser power percentage was adjusted on a sample-by-sample basis to obtain an arbitrary intensity of 10^3 - 10^4 and a flat baseline; mass range 14520-69960 Da; spectrum acquisition – random walk partial sample mode, results were the sum of at least 3 x 100 laser shots; calibration of the system was performed using the Bruker Protein Calibration Standard II mixture (part number 8207234). Spectra were analyzed using the open-source software mMass (version 5.5.0)³¹⁻³³ after data had been exported as an ASCII file from FlexAnalysis (Bruker Daltonics).

Nuclear magnetic resonance (NMR) spectra were obtained on a Varian Inova 500 in the Stanford University Department of Chemistry NMR Facility. Spectra were processed using MestReNova v10.0.2-15465 (Mestrelab Research S.L., 2015).

Binding assay (analogous to Chapter 2). N-terminal acetylated, C-terminal amidated peptides (0.9 mM) or sfGFP constructs (0.9 mM) were incubated with a CBT derivative (1 mM) in a solution of 10% DMF in PBS for 48 h at room temperature. For the negative control, the peptides or proteins were incubated in a solution of 10% DMF in PBS. An aliquot (0.5 μ L) of each reaction mixture was analyzed by MALDI-TOF MS without prior purification.

Cysteine competition assay. After MALDI-TOF MS analysis, half of the binding assay solution was incubated for 48 h at room temperature after adding an equal volume of cysteine (10 mM or 10 equiv. relative to CBTNAc, unless cysteine would interfere with a later assay, in which case 1 mM or 1.1 equiv. was used; pH adjusted to approximately 7) with or without DTT or TCEP reducing agent (10-15 mM or 10-15 equiv., unless the reducing agent would interfere with a later assay, in which case 1 mM or 1.1 equiv. was used; pH adjusted to approximately 7). The remainder of the binding assay solution was incubated with an equal volume of aqueous DTT or TCEP reducing agent (10-15 mM or

10-15 equiv., unless the reducing agent would interfere with a later assay, in which case 1 mM or 1.1 equiv. was used; pH adjusted to approximately 7). The reaction mixtures were again analyzed by MALDI-TOF MS without prior purification.

Chloroacetate labeling of cysteine residues. Peptide solution (0.9 mM; 0.45 mM if after treatment with a slight excess of cysteine) was incubated with an equal volume of 30 or 500 mM chloroacetate (pH adjusted to approximately 7) overnight at room temperature. The reaction mixtures were analyzed by MALDI-TOF MS without prior purification.

LC-MS assay: Reaction of CBTag 1.0 with the NCL peptide. Samples were prepared as follows: In an LC-MS vial equipped with a 300- μ L insert, a solution of 1 mM CBTag 1.0, CBTag 1.0 (K9R), or CBTag 1.0 (C7S), 2 mM TCEP (adjusted to pH 7 in water), and 1 mM NCL peptide was prepared in 5-6% DMF in PBS (total volume of 20 μ L) and the starting time noted. An LC-MS sequence was prepared such that each sample was injected five times over approximately 45 hours, with an injection volume of 10 μ L each time. The LC-MS gradient was: 5% solvent B for 2 minutes, 5-75% solvent B over 15 minutes, 75-100% solvent B over 1 minute, 100% solvent B for 6 minutes, 100-5% solvent B over 1 minute, and 5% solvent B for 3 minutes, where solvent A was 0.1% TFA in water and solvent B was 0.1% TFA in acetonitrile. Identification of peaks was performed by comparison to the spectra of pure compounds, determining whether predicted masses were present, and/or checking whether addition of an equal volume of 200 mM cysteine (adjusted to pH 7 in water) caused the disappearance of the peak over time. The relevant peaks in the 210 nm chromatographs were integrated manually.

LC-MS time-course assay: Reaction of CBTag 1.0 with S-(4-nitrophenyl) thioacetate. Samples were prepared as follows: In an LC-MS vial equipped with a 300- μ L insert, a solution of 22.5 mM peptide (CBTag 1.0, CBTag 1.0 (K9R), or CBTag 1.0 (C7S); adjusted to pH 7 in PBS), 2 mM TCEP (adjusted to pH 7 in water), and 22.5 mM S-(4-nitrophenyl) thioacetate was prepared in 5-6% DMF in PBS (total volume of 10 μ L) and the starting time noted. An LC-MS sequence was prepared such that each sample was injected five times over approximately 45 hours, with an injection volume of 1 μ L each time. The LC-MS gradient was: 5% solvent B for 2 minutes, 5-75% solvent B over 15 minutes, 75-100% solvent B over 1 minute, 100% solvent B for 6 minutes, 100-5% solvent B over 1 minute, and 5% solvent B for 3 minutes, where solvent A was 0.1% TFA in water and solvent B was 0.1% TFA in acetonitrile. Identification of peaks was performed by comparison to the spectra of pure compounds and determining whether predicted masses were present. The relevant peaks in the 210 nm chromatographs were integrated manually.

sfGFP constructs. *Design.* A pET vector containing a kanamycin resistance cassette and a sequence corresponding to sfGFP with a Tobacco Etch Virus (TEV) protease recognition sequence, HaloTag, and His₆ tag on the C-terminus (for sequence, see Figure 3.S6) as well as primers to introduce an N-terminal His₆ tag were provided by then-lab members Dr. Justin Kim and Dr. Mason Appel. The DNA sequence encoding the CBTag 1.0 peptide (Gly-Gly-His-Pro-Asp-Pro-Cys-Pro-Lys-Gly-Gly) as well as three extra

glycines on each side was optimized for expression in *E. coli* based on Hénaut and Danchin's work²¹ as well as the European Molecular Biology Laboratory (EMBL) website²² describing codons that are known to cause translation issues. MacPyMOL Molecular Graphics System v1.4 (Schrödinger, LLC, New York, NY, 2010) was used to visually assess potential sites of insertion on sfGFP (PDB 2B3P). Primers to remove the TEV protease recognition sequence, the HaloTag, and the C-terminal His₆ tag were designed using the New England Biotechnologies (NEB) primer melting temperature (T_m) calculator then purchased from Elim Biopharmaceuticals, Inc. Primer selection involved determining how much of the sequence on either side of the desired modification was necessary to obtain an annealing temperature and a T_m close to 72 °C. Primers for the simultaneous deletion of those sequences and insertion of the Gly-Gly-Gly-(CBTag 1.0)-Gly-Gly-Gly peptide at the C-terminus were also designed and purchased, as were primers for insertion of the Gly-Gly-Gly-(CBTag 1.0)-Gly-Gly-Gly peptide at the N-terminus or one of four internal sites (Table 3.2).

Table 3.2. Primer sequences used to insert CBTag 1.0 at different sfGFP locations. See Supplementary Table 3.S1 for full sequences of sfGFP constructs.

| Name | Sequence |
|------------------------------------|---|
| GFP_ctermPF3_F | CCGGATCCGTGCCCGAAAGGCGGC TGAGATCCGGC TGCTAACAAAGCCC |
| GFP_ctermPF3_R | GTGGCCGCCGCCGCCGCTTTATACAGTTCATCCAT ACCGTGGTAATGCCCG |
| GFP_halo_del_F | TGAGATCCGGCTGCTAACAAAGCCC |
| GFP_halo_del_R | ATGGTGATGATGATGGTGGCTGCCTTTATACAG |
| GFP_loop10-11PF3_halo_del_F | CCGAAAGGCGGCGGCGGCGGC AATGAAAAACGTGA TCATATGGTGCTGCTGG |
| GFP_loop10-11PF3_halo_del_R | GCACGGATCCGGGTGGCCGCCGCCGCCCGGAT CTTTGCTCAGAACGCTCTGG |
| GFP_loop9-10PF3_halo_del_F | GATCCGTGCCCGAAAGGCGGCGGCGGC CCGGT GCTGCTGCCGATAATC |
| GFP_loop9-10PF3_halo_del_R | CGGGTGCCGCCGCCGCCGCCGCCATCACCATCG GGGTATTCTGCTG |
| GFP_loop6-7PF3_halo_del_F | CCGTGCCCGAAAGGCGGCGGCGGC GGCAACAT TCTGGTCAATAACTGGAATATAATTTCAAC |
| GFP_loop6-7PF3_halo_del_R | ATCCGGGTGGCCGCCGCCGCCGCCATCTTCTTTAAA ATCAATACCTTTTCAGTTCAATGCGGTTCA |
| GFP_loop3-innerhelixPF3_halo_del_F | CCGGATCCGTGCCCGAAAGGCGGCGGCGGC AA ACTGCCGGTTCGTGGCC |
| GFP_loop3-innerhelixPF3_halo_del_R | GTGGCCGCCGCCGCCGCCACCGGTGGTGCAAATAA ATTTCAGGGTCAG |
| GFP_halo_del_ctermHis6_del_R | TTTATACAGTTCATCCATACCGTGGGTAATGCCCG |
| GFP_ntermPF3_halo_del_longprimer_F | CCGGATCCGTGCCCGAAAGGCGGCGGCGGC AT GGTTAGCAAAGGTGAAGAAGTGTACCAGGC |
| GFP_ntermPF3_halo_del_longprimer_R | GTGGCCGCCCATGGTATATCTCCTTCTTAAAGTTAAA CAAAATTATTTCTAGAGGGGAATTGTTATCCGC |
| GFP_round1_F | CATCATCATCACAGCAGCGGC ATGGTTAGCAAA GGTGAAGAAGT |
| GFP_round1_R | ATGCATGGTATATCTCCTTCTTAAAGTTAAACAAAAT TATTCTAGAG |

Site-directed mutagenesis. To create the desired plasmids, the NEB Q5 site-directed mutagenesis kit (E0554) protocol was applied with some modifications. First, the plasmid DNA was amplified with the appropriate primers following the NEB Q5® Hot-Start High-Fidelity DNA Polymerase (M0493) protocol without the Q5 High GC Enhancer, using 1

μL of the plasmid template (1 ng/ μL) per 25 μL reaction. A Bio-Rad C1000 Thermal Cycler or MJ Research PTC-200 Peltier Thermal Cycler-DNA Engine Gradient Cycler was used to perform PCR amplification with an initial denaturation step (30 s at 98 °C), 28 cycles of elongation (10 s at 98 °C, 20 s at 72 °C or the lowest annealing temperature determined by the NEB T_m calculator, and 3 min. at 72 °C), final elongation (2 min. at 72 °C), and cooling (4 °C). These parameters were modified as necessary to optimize the PCR reaction for each set of primers. If the Q5 protocol failed, the Takara PrimeSTAR GXL DNA Polymerase (R050A) protocol was used instead with 5 μL of the plasmid template (1 ng/ μL) per 25 μL reaction and the following PCR protocol: 30 s at 98 °C; 29 cycles of 10 s at 98 °C, 15 s at 60 °C, and 1 min. at 68 °C; 2 min. at 68 °C; 4 °C. Once the PCR amplification was complete, the presence of product was confirmed by gel electrophoresis. Samples were loaded onto a 0.8% w/v agarose gel (TAE buffer, 3 μL Gel Red/50 mL, 4.5 μL PCR reaction and 0.9 μL 6x loading buffer per lane, 5 μL Invitrogen 1 Kb Plus DNA Ladder), the gel was run for 40 min. at 120 V using a Bio-Rad Powerpac 200 or 1000, and a UVP VisiDoc-It Imaging System equipped with a UVP Transilluminator was used to visualize the gel. Crude PCR products were kept at -20 °C when not in use.

Transformation into DH5 α cells, plasmid sequencing, transformation into BL21(DE3) cells. After confirming that the DNA was amplified, the PCR products were subjected to Kinase, Ligation, and Dpn1 (KLD) treatment and transformed into NEB 5-alpha Competent *E. coli* cells (C2987) as described in the NEB Q5 site-directed mutagenesis kit (E0554) protocol. 100 μL of the transformed cells were plated onto LB-kanamycin plates and incubated at 37 °C overnight. Plates were stored at 4 °C when not in use. Single colonies were grown overnight in 5 mL of 50 $\mu\text{g}/\text{mL}$ kanamycin in LB at 37 °C, 250 rpm. Plasmid was purified following the Thermo Scientific GeneJET Plasmid Miniprep Kit (K0503) protocol, with sterile water used in the final elution step. Plasmid concentration was determined using a Thermo Fisher Scientific NanoDrop 2000/c spectrophotometer, then the plasmids were sent to Elim Biopharmaceuticals, Inc for sequencing. A plasmid Editor (ApE) version 2.0.49 was used to align the sequencing data with the target sequence. If the sequences matched, the plasmids were transformed into chemically competent NiCo21 cells (optimized for protein expression). This consisted of thawing a 100 μL aliquot of cells on ice for 5-10 minutes, adding 100 μL of a mixture of 50 ng plasmid, 40 μL 5x KCM buffer, and sterile water, incubating on ice for 30 min., heat-shocking at 42 °C for 90 s, incubating on ice a further 1-2 minutes, adding 800 μL SOC medium (sterile conditions), and shaking at 37 °C for 60 minutes. 100 μL of the cells were plated on pre-warmed LB-kanamycin plates and incubated at 37 °C overnight. Plates were stored at 4 °C until use.

Large-scale expression of proteins. For expression, single colonies were grown overnight in 7 mL of 50 $\mu\text{g}/\text{mL}$ kanamycin in LB at 37 °C, 250 rpm. The next day, these cultures were added to a sterile 1 L Erlenmeyer flask containing approximately 250 mL of LB containing 0.2% D-glucose and 50 $\mu\text{g}/\text{mL}$ kanamycin. Once the OD₆₀₀ (measured in an Eppendorf BioPhotometer) reached 0.4-0.6, a 1-mL aliquot was spun down in a microcentrifuge at 16000 rcf for 3 min. and the pellet saved at -20 °C for later comparison

with a post-induction sample, then IPTG was added to the remaining culture to a final concentration of 1 mM to induce protein expression. The cultures were incubated a further 3 h at 37 °C with shaking (250 rpm). The OD₆₀₀ of the now-green cultures was measured and a 0.2 mL-aliquot spun down and saved in pellet form for gel analysis. The cultures were transferred to 500-mL centrifuge bottles and spun down at 5000 xg at 4 °C for 20 min. in a Thermo Scientific Sorvall RC 6 Plus or 5C Plus Centrifuge. The supernatant was decanted, leaving a green pellet, which was transferred to a 50-mL Falcon tube, weighed, and either flash-frozen and stored at -80 °C or carried on to the next step.

Lysis and homogenization of bacteria to obtain clarified lysate. For each gram of pellet, one protease inhibitor tablet (Roche cOmplete ULTRA tablets, mini) was dissolved in 10 mL of lysis buffer (50 mM sodium phosphate buffer, pH 8.0; 300 mM NaCl; 10 mM imidazole, pH 8.0; final solution was adjusted to pH 8.0 and sterile-filtered). The pellet was resuspended in this solution and vortexed. The resulting mixture was homogenized by using first a Dounce homogenizer (on ice) then an Avestin EmulsiFlex-C3. Then, 10,000x Pierce Universal Nuclease was added to reach a final concentration of 1x, the mixture was gently swirled, and the cell lysate was left on ice for at least 30 min. After that, the sample was centrifuged at 15,000 xg for 30 min. at 4 °C and the clarified lysate decanted into a 50-mL Falcon tube. 1 mL of clarified lysate was set aside for analysis. The pellet was washed with distilled water and also saved for analysis.

Ni-NTA batch purification. The His₆-tagged proteins were initially batch purified using Ni-NTA agarose beads (QIAGEN). The beads were prewashed by transferring 2 mL of the slurry to a clean 50-mL Falcon tube, adding 45 mL wash buffer (50 mM sodium phosphate buffer, pH 8.0; 300 mM NaCl; 20 mM imidazole, pH 8.0; final solution was adjusted to pH 8.0 and sterile-filtered), and gently spinning down the beads by centrifuging at 1000 xg, 4 °C for 1 minute in a Sorvall Legend RT. The supernatant was carefully removed and the beads washed another 2 times, then the clarified lysate was added to the beads. The bead-lysate mixture was gently inverted at 4 °C for 1 h using a Barnstead/Thermolyne Labquake Rotisserie Shaker (Model #C415110). The mixture was then poured into a 25-mL Bio-Rad Econo-Pac Chromotography Column, rinsing the Falcon tube with wash buffer. The flowthrough was collected, then wash buffer was used to rinse the beads until the flowthrough was mostly clear (~50 mL), being cautious not to disturb the beads. Finally, 2x 5 mL of elution buffer (50 mM sodium phosphate buffer, pH 8.0; 300 mM NaCl; 250 mM imidazole, pH 8.0; final solution was adjusted to pH 8.0 and sterile-filtered) was used to remove the His₆-tagged proteins from the column, and 2-mL fractions were collected. The amount of protein in each fraction was determined by using a Thermo Fisher Scientific NanoDrop 2000/c spectrophotometer to measure the protein A₂₈₀, then dividing by the Abs 0.1% value (the measurement at 1 g protein/L) for the sfGFP construct as calculated by ExpASY ProtParam.²³

Gel electrophoresis analysis of protein expression. The contents of the flowthrough, the washes, each fraction, and the earlier pre-induction, post-induction, pellet, pellet, and clarified lysate samples were compared by gel electrophoresis. Each sample was

prepared as follows: the pre- and post-induction pellets were resuspended in $[OD_{600} \times (\text{mL centrifuged}) \times 75]$ μL of 1x SDS loading buffer; a pipette tip was used to pick a small amount of the post-homogenization pellet, which was then resuspended in 1x SDS loading buffer and vortexed thoroughly; the clarified lysate was diluted 8x in 1x SDS loading buffer; and post-Ni-NTA column fractions were diluted to $\sim 0.3 \mu\text{g}$ protein/mL in 1x SDS loading buffer. The samples were placed in a 95 °C Fisher Scientific Dry Bath Incubator for 10 minutes then spun down at 16,100 rcf in a microcentrifuge. 10 μL of sample or 10 μL of 0.5x Thermo Fisher Scientific BenchMark Protein Ladder were loaded into each well of a 4-12% Bis-Tris gel, which was then run at 180 V for 40 min. in 1x XT MES buffer. The gels were imaged using a Typhoon 9410 Variable Mode Imager (488 nm laser, 520 bp 40 filter or no filter, PMT 450-600 V) to detect any fluorescence prior to staining. ImageQuant was used to adjust the brightness and contrast of the images. Although incubating the samples in 1x SDS at 95 °C will denature most proteins, the sfGFP constructs were indeed still detectable by fluorescence, though the signal was weak. Next, the gel was fixed and stained using the Invitrogen Colloidal Blue Staining Kit (LC6025) and the protocol entitled “Staining NuPAGE Novex Bis-Tris Gels.” The gel was stained for 3 h on The Belly Dancer (Stovall Life Science Incorporated) then incubated with distilled water overnight on The Belly Dancer to remove background. Stained gels were scanned into Photoshop using a Canon CanoScan LiDE 30.

MALDI-TOF MS analysis of protein purity. In all cases, sfGFP monomer and dimer were the major peaks (ratio varied), so initial CBT derivative reaction assays were performed without further FPLC purification.

Buffer exchange and concentration of sfGFP constructs. Ni-NTA elution fractions for each construct were combined and loaded into an Amicon Ultra-15 10K Centrifugal Filter Device. The samples were centrifuged in a Sorvall Legend RT for 15 minutes at 3700 xg and 4 °C. Then, four washes were performed, adding 10-15 mL 1x PBS each time and gently inverting before centrifugation. Protein was recovered by transferring the final concentrate into a 2-mL Eppendorf tube, rinsing the filter with $\sim 400 \mu\text{L}$ 1x PBS, and the concentration was determined as described above.

Calculation of percent sfGFP in mature (fluorescent) form. For each protein, the percent of sfGFP in the mature form was calculated by measuring the absorbance of the sample at 485 and 280 nm using a Thermo Fisher Scientific NanoDrop 2000/c spectrophotometer. If percent mature sfGFP = $(c_{485}/c_{280}) \times 100$, and by the Beer-Lambert Law, $c_{\lambda} = A_{\lambda}/(\epsilon_{\lambda}l)$, where A_{λ} is the absorbance at 485 or 280 nm, ϵ_{λ} is the extinction coefficient for wildtype sfGFP (485 nm, $83,300 \text{ M}^{-1}\text{cm}^{-1}$)²⁴ or as calculated by ExPASy ProtParam (280 nm, $18,972.5 \text{ M}^{-1}\text{cm}^{-1}$ on average),²² and l is the pathlength (Nanodrop), then percent mature sfGFP = $(A_{485}/\epsilon_{485})/(A_{280}/\epsilon_{280}) \times 100$.

DIVERSE. In this section, we adapted procedures from references 17 and 25. Ultrapure DNase/RNase-free Distilled Water (Invitrogen) and Eppendorf LoBind microcentrifuge

extraction was transferred to a new tube, ethanol precipitation of the DNA was performed using procedures reported by Walker and Lorsch.²⁶ All aqueous solutions were prepared in Ultrapure DNase-, RNase-free water. Ammonium acetate (7.5 M, 0.1 vol.) was added to the aqueous layer of the extraction. After mixing, 2 vol. of 100% EtOH was added and the solution incubated at -80 °C for 1 h. DNA was pelleted in a microcentrifuge at 12,000 xg, 4 °C for 15 min. The supernatant was removed, then the pellet was allowed to soak for 2 min. in 2 vol. of 70% EtOH. The tube was centrifuged for 2 min. at 12,000 xg, 4 °C, then the supernatant was carefully removed and the pellet allowed to air dry in an open tube covered with a Kimwipe for 1 h at r.t. The pellet was resuspended in 50 µL nuclease-free water, and its concentration was determined using the Thermo Fisher Scientific NanoDrop 2000/c spectrophotometer. The DNA was carried on directly to transcription.

Run-off in vitro transcription with T7 RNA polymerase (NEB). This step was performed following NEB's "Protocol for Standard RNA Synthesis" with a 1 U/µL final concentration of RNaseOUT Recombinant Ribonuclease Inhibitor (Thermo Fisher) and without DTT. The reaction was incubated at 37 °C overnight.

RNA purification. These conditions were the same as those for DNA purification, except 2.5 vol. of 100% EtOH and 2.5 vol. of 70% EtOH were used instead of 2 vol.

Cell-free, release factor-free translation. This step was done using the NEB PURExpress ΔRF123 kit and protocol. One full kit was used, giving a reaction volume of 250 µL. RNaseOUT was used at a final concentration of 1 U/µL, mRNA was used at a final concentration of approximately 2.5 µM (here, ~18.5 µg), and the puromycin linker sequence D11 was used at a final concentration of 2.5 µM. The reaction was incubated at 37 °C for 2.25 h, transferred to ice, and quenched by adding 50 µL of 100 mM EDTA (pH 8.0) and incubating at 37 °C for 30 min. (in later steps, this was reduced to 5 min.).

Reverse transcription. SuperScript IV Reverse Transcriptase (Thermo Fisher) was used and its protocol followed with the following notes: the whole translation mix from the previous step was used; the reverse primer was D9; Ultrapure DNase/RNase-Free Distilled Water was used instead of DEPC-treated water; 65 °C and 55 °C incubations were done in an Eppendorf Thermomixer R. Instead of inactivating with an 80 °C incubation, 50 µL of 100 mM EDTA (pH 8.0) was added, then the solution was incubated at 37 °C for 30 min. (in later steps, this was reduced to 5 min.). 100 µL of HEPES solution (1 M; Thermo Fisher) was used to neutralize the EDTA.

Positive selection, round 1. The solution from the previous step was split equally. In one half, TCEP (0.5 M, pH 6~7) was added to a final concentration of 250 µM (approximately 100x the concentration of the peptides) and the solution was incubated at r.t. for 5-10 min. The other half was left alone. These two conditions (+/-TCEP) were kept separate throughout the selection process. To each solution was added 25 µL of well-mixed alk-CBT beads (see Synthesis section; 10 mg/mL in Ultrapure DNase/RNase-free water).

The tubes were then incubated for 60 min. in an Eppendorf Thermomixer R set to 37 °C and 300 rpm. Using a DynaMag-2 magnet to trap the beads, the supernatant was removed, then the beads were washed with 3x ~400 µL of HBS-T (50 mM HEPES-KOH, pH 7.5; 300 mM NaCl; 0.05% TWEEN 20; final pH 7.5).

Elution of cDNA from beads. Beads were diluted with a premade PCR reaction mixture (all but polymerase) and incubated for 5 min. at 95 °C in a Fisher Scientific Dry Bath Incubator. Q5 Hot-Start High-Fidelity DNA Polymerase was added, and PCR was performed using the protocol described above (*PCR amplification of dsDNA template*). For the first selection round and each set of beads (+/-TCEP), 396 µL of the PCR reaction mixture was used to elute the cDNA, 4 µL polymerase was added, and the solution was separated into 50-µL aliquots for PCR.

Round 2. 10 µL of crude PCR product from the previous step was used directly in the cell-free, release factor-free system. All portions of the translation and the reverse transcription steps were performed on a 1/10th scale relative to the first round of selection.

Negative selection, round 2. After the +TCEP sample was treated with TCEP, to each solution (+/-TCEP) was added 5 µL (it was determined later that this was too low a volume to be effective) of thoroughly mixed alkyne and dibenzocyclooctyne (DBCO) beads (10 mg/mL; bought separately from Click Chemistry Tools, rinsed with and resuspended in nuclease-free water). The mixture was incubated for 60 min. in an Eppendorf Thermomixer R set to 37 °C and 300 rpm. Using a DynaMag-2 magnet to trap the beads, the supernatant was transferred to a clean tube, and the 1x 100 µL of 1 M HEPES used to wash the beads were transferred to the same tube.

Positive selection, round 2. To each of the two tubes of supernatant (+/-TCEP) from the previous step was added 5 µL of DBCO-CBT beads (see Synthesis section; 10 mg/mL). The tubes were then incubated for 60 min. in an Eppendorf Thermomixer R set to 37 °C and 300 rpm. Using a DynaMag-2 magnet to trap the beads, the supernatant was removed, then the beads were washed with 3x ~122 µL of HBS-T (50 mM HEPES-KOH, pH 7.5; 300 mM NaCl; 0.05% TWEEN 20; final pH 7.5). Elution of cDNA from beads was performed on a 50 µL scale.

Rounds 3-13. These selections were performed similarly to those in round 2, with the following conditions (each was applied to the TCEP-treated and untreated solutions separately):

Round 3: 25 µL of beads were used in the negative selection, 5 µL of alk-CBT beads in the positive. Incubation was for 60 min. at 37 °C, 300 rpm.

Round 4: 10 µL of beads were used in the positive selection, 2.5 µL of DBCO-CBT beads in the positive. Incubation was for 30 min. at 37 °C, 300 rpm.

Round 5: 3x 10 µL of beads were used in the negative selection, 2.5 µL of alk-CBT beads in the positive. Incubation was for 15 min. at 37 °C, 300 rpm.

Round 6: 3x 10 μ L of beads were used in the negative selection, 2.5 μ L of DBCO-CBT beads in the positive. Incubation was for 15 min. at 37 °C, 300 rpm. Elution of cDNA from beads was performed on a 25 μ L scale starting in this round.

Round 7: As cDNA elution was done on a 25 μ L scale instead of a 50 μ L scale, everything was done at $\frac{1}{2}$ scale relative to round 6. In this round and subsequent rounds, samples were quenched with EDTA for 5 min. instead of 30 min. 9x 5 μ L of beads were used in the negative selection, 1 μ L of alk-CBT beads in the positive. Incubation was for 5 min. at 37 °C, 300 rpm.

Round 8: 9x 5 μ L of beads were used in the negative selection, 1 μ L of alk-CBT beads in the positive. Incubation was for 3 min. (negative selection) or 5 min. (positive selection) at 37 °C, 300 rpm.

Round 9: 3x 50 μ L of beads were used in the negative selection, 1 μ L of alk-CBT beads in the positive. Incubation was for 5 min. at 37 °C, 300 rpm.

Round 10: Conditions in round 9 were repeated.

Round 11: Only 5 μ L of each quenched reverse transcriptase reaction were incubated with 3x 25 μ L of beads for 3 min. at 37 °C and 300 rpm in the negative selection. The positive selection was performed as in round 10.

Round 12: The negative and positive selections were performed as in round 11. After washing with 1x 50 μ L HBS-T, the beads were incubated with 3x 25-50 μ L of a CBT-PEG₄-azide solution (~20 mM, 10% DMSO in PBS) for 5 min. at 37 °C and 300 rpm. The beads were washed 3x with HBS-T and carried forward as previously.

Round 13: 2x 25 μ L of beads were used in the negative selection, 1 μ L of alk-CBT beads in the positive. Incubation was for 5 min. at 37 °C, 300 rpm. After washing with 1x 50 μ L HBS-T, the beads were incubated with 2x 50 μ L and 1x 100 μ L of a CBT-PEG₄-azide solution (~20 mM, 10% DMSO in PBS) for 5 min. at 37 °C and 300 rpm. The beads were washed 3x with HBS-T and PCR-amplified as previously.

Sequencing of products from round 13. The crude PCR products (+/-TCEP) of the final round of selection were subjected to ZeroBlunt TOPO PCR Cloning (Thermo Fisher) and transformation into Thermo One Shot TOP10 Chemically Competent *E. Coli* following the associated protocols. The cells were grown at 37 °C overnight on LB-kanamycin plates. For each condition (+/-TCEP), 100 colonies were picked and grown in 5 mL of 50 μ g/mL kanamycin in LB overnight at 37 °C, 180 or 210 rpm. Plasmids were purified following the Thermo Scientific GeneJET Plasmid Miniprep Kit (K0503) protocol, with sterile water used in the final elution step. Plasmid concentration was determined using a Thermo Fisher Scientific NanoDrop 2000/c spectrophotometer, then the plasmids were sent to Elim Biopharmaceuticals, Inc for sequencing. A plasmid Editor (ApE) version 2.0.49 was used to determine the hit peptides based on the DNA sequences.

Table 3.4. Full list of hit peptides determined from sequencing data. [/] indicates two or more potential amino acid residues. The DNA sequencing data for the last six entries was not of a high enough quality to enable full determination of the peptide sequences.

| Hit Peptide | Notes |
|-------------------------------|------------------------------|
| RLTPWKSCMCDWW | no TCEP |
| VHDESKRASSHHRVEEEVARTGGGRR | no TCEP; no cysteine residue |
| AHYRCKANR *I | no TCEP; TAA |
| SC[Q/W]RGAARDVTLN | no TCEP; no lysine residue |
| AD[T/P]F[G/R]KGCLGVRL | |
| *VVPPKSCVHRHA | TAG |
| PPWRTKCRAASQ | |
| [A/T]D[T/P]F[G/R]KGC[G/L]GVRL | |
| RVRLPCKIEAVN | |
| RAAPRCKGILRH | |
| PTRAAKCASTWS | |
| P *RARCKGSGG | TGA |
| VRRK *CKRLSQA | TAG |
| RRWRFKCAPLVQ | |
| CDVEGKSCRHAS | |
| GSGFPCKRRPGK | appeared twice |
| WCPWFKACRHHVAR | appeared twice |
| TGQGIKCCAGVA | |
| ALGTHKCG *QEQ | TAA |
| SYGPRKLCGHVEWWRRR | appeared twice |
| LCGDRCKASRRV | |
| ADTWACKTVCGA | |
| AGHARCKVHCYE | |
| LFVRCCGKGAGTL | |
| SWGRVCKRGEVV | |
| RVWPGKCAAMQR | |
| CIRADDGWRRR | |
| | |
| L V X P A N S C VHRHA | very similar to plate 1 G05 |
| * X V P X K S X VHRHA | very similar to plate 1 G05 |
| GGXVSCARDLA*KGEEVAANS | |
| GLRRRYLKDHASRVEEEVARTGGGR | |
| R X VRT C X R X D X C | |
| EVIIGFRVRK C K R H D Q C | |

Computational modeling of DIVERSE peptide hits. PEP-FOLD²⁶⁻²⁷ was used for initial *in silico* modeling of peptide hit structures. For conformational searching and molecular dynamics calculations, Maestro (Version 10.7.015, MMShare Version 3.5.015, Release 2016-3, Platform Darwin-x86_64, Schrödinger, LLC, New York, NY, 2016) and the integrated program Macromodel (Schrödinger, LLC, New York, NY, 2016) were used. In general, standard parameters were applied. An OPLS3 force field with water as the solvent was used. Structures were minimized using the Polak-Ribiere Conjugate Gradient (PRCG) method until a gradient below the convergence threshold of 0.05 kJ/Å-mol was reached, with a maximum of 2500 iterations. For conformational searching, a mixed torsional/low-mode sampling method (multi-ligand) was used, with 1000 steps maximum, 100 steps per rotatable bond, and the energy window for saving structures set to 5.02 kcal/mol. As this was meant to be a fast screen of potential sequences, only one round of minimization and conformational searching was performed for each structure.

Synthesis and characterization (small molecules, alk-CBT and DBCO-CBT beads, peptides)

General procedures for solution-phase amide coupling. In a flame-dried 10-mL pear-shaped flask equipped with a stir bar, a solution of 6-amino-2-cyanobenzothiazole (50 mg, 0.285 mmol) in anhydrous DMF (0.5-1 mL; enough to fully dissolve) was treated with a carboxylic acid (0.314 mmol, 1.05 equiv.). [Oils were measured by mass in a pre-weighed syringe, which was rinsed with ~1 mL anhydrous DCM as necessary.] The resulting solution was cooled to 0 °C, then a solution of EDC•HCl (87.5 mg, 0.457 mmol, 1.6 equiv.; recrystallized from DCM/ether) in anhydrous DCM (1.5 mL) was added. The reaction was allowed to stir at r.t. overnight, then it was diluted with DCM and transferred to a separatory funnel. The organic layer was washed in order with 3x ~30 mL of 1 M LiCl (to remove DMF), 2x ~20 mL of 3% HCl, 2x ~30 mL of 0.5 M sodium carbonate (Na₂CO₃), and 2x ~20 mL brine. The organic layer was then dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*.

CBT-PEG₄-azide. Crude brown oil was purified by silica gel chromatography on an Isolera Prime (Biotage) equipped with a SNAP KP-Sil 100 g column equilibrated with 5 CV of the starting solvent mixture. The flow rate was 100 mL/min., the 254/280 nm wavelengths were monitored, and peaks that absorbed above 23 mAU were collected. The gradient was as follows: 100% EtOAc for 5 CV, 5% MeOH/EtOAc for 3 CV, 10% MeOH/EtOAc for 3 CV. Pure fractions were confirmed by TLC (5% MeOH/EtOAc, visualized by UV) and LC-MS, combined, and concentrated *in vacuo*. This yielded a clear yellow oil that became a yellow solid on trituration with water. The yellow solid was resuspended in water and lyophilized overnight to yield 37.2 mg (29%) of a dull yellow oil. ¹H NMR (500 MHz, Chloroform-*d*) δ 9.38 (s, 1H), 8.70 (t, J = 2.2 Hz, 1H), 8.06 (dd, J = 9.0, 2.3 Hz, 1H), 7.55 (dd, J = 9.0, 2.1 Hz, 1H), 3.83 (t, J = 5.5 Hz, 2H), 3.73 – 3.64 (m, 8H), 3.62 – 3.58 (m, 6H), 3.37 – 3.32 (m, 2H), 2.70 (t, J = 5.5 Hz, 2H). ¹³C{¹H} NMR (126

MHz, Chloroform-*d*) δ 170.82, 148.36, 139.23, 136.77, 134.80, 125.03, 121.09, 113.22, 111.17, 70.63, 70.54, 70.48, 70.46, 70.27, 70.22, 69.99, 67.02, 50.60, 38.03. HRMS (ESI) *m/z*: [M + Na]⁺ Calcd for C₁₉H₂₄N₆O₅SNa 471.1421; Found 471.1413.

azidoCBTNac. Crude product was purified by silica gel chromatography on an Isolera Prime (Biotage) equipped with a SNAP KP-Sil 100 g column equilibrated with 5 CV of the starting solvent mixture. The flow rate was 100 mL/min., the 254/280 nm wavelengths were monitored, and peaks that absorbed above 40 mAU were collected. The gradient was as follows: 25% EtOAc/hexanes for 2 CV, 25-50% EtOAc/hexanes over 7 CV, 50% EtOAc for 2 CV. This yielded 50 mg (67.8%) of a(n off-)white solid. ¹H NMR (500 MHz, Chloroform-*d*) δ 8.64 (d, *J* = 1.9 Hz, 1H), 8.07 (dd, *J* = 8.9, 1.4 Hz, 1H), 7.49 (dd, *J* = 9.0, 1.9 Hz, 1H), 4.07 (s, 2H). ¹³C{¹H} NMR (126 MHz, Chloroform-*d*) δ 166.14, 148.85, 137.98, 136.81, 135.62, 125.34, 120.81, 112.99, 111.70, 52.62. HRMS (ESI) *m/z*: [M + H]⁺ Calcd for C₁₀H₇N₆OS 259.0397; Found 259.0392.

CBT-Fmoc-miniPEG. This reaction was done on a 0.5 mmol scale. By TLC (3:1 EtOAc/hexanes), the reaction was not complete after running overnight, so another 0.8 equiv. of EDC•HCl was added and the reaction allowed to stir another 24 h before extracting as described above. Product was purified by silica gel chromatography on an Isolera Prime (Biotage) equipped with a Zip KP-Sil 120 g column equilibrated with 5 CV of the starting solvent mixture. The flow rate was 50 mL/min., the 254/280 nm wavelengths were monitored, and peaks that absorbed above 23 mAU were collected. The gradient was as follows: 18% EtOAc/hexanes for 1 CV, 18-100% EtOAc/hexanes over 10 CV, 100% EtOAc for 6 CV. This yielded 143.4 mg (55.8%) of a chalky (off-)white solid. ¹H NMR (500 MHz, Chloroform-*d*) δ 8.92 (s, 1H), 8.66 (d, *J* = 2.1 Hz, 1H), 8.08 (d, *J* = 8.9 Hz, 1H), 7.74 (d, *J* = 7.6 Hz, 2H), 7.53 – 7.44 (m, 3H), 7.37 (t, *J* = 7.5 Hz, 2H), 7.29 – 7.25 (m, 2H), 5.12 – 5.08 (m, 1H), 4.35 (d, *J* = 6.9 Hz, 2H), 4.16 (s, 3H), 3.80 (t, *J* = 4.2 Hz, 2H), 3.72 (t, *J* = 4.2 Hz, 2H), 3.66 (t, *J* = 5.3 Hz, 2H), 3.46 (q, *J* = 5.8 Hz, 2H). ¹³C{¹H} NMR (126 MHz, Chloroform-*d*) δ 168.78, 156.74, 149.04, 143.99, 141.50, 138.02, 137.10, 135.67, 127.98, 127.25, 125.61, 125.13, 120.87, 120.27, 113.32, 111.70, 77.59, 77.33, 77.08, 71.50, 70.80, 70.46, 70.12, 66.96, 47.38, 40.87. HRMS (ESI) *m/z*: [M + Na]⁺ Calcd for C₂₉H₂₆N₄O₅SNa 565.1516; Found 565.1509.

CBT-miniPEG. To a 25 mL flame-dried recovery flask equipped with a stir bar was added CBT-Fmoc-miniPEG (94.2 mg, 0.174 mmol) dissolved in anhydrous DMF (3.2 mL). Dry piperidine (0.8 mL, 8.099 mmol, 46.5 equiv.; dried over 4 Å molecular sieves) was added, and the resulting pale/clear dark yellow-orange solution was allowed to stir ~1.5 h under nitrogen at room temperature. The reaction mixture was then diluted slightly with water and DMF and directly purified by reversed-phase C18 chromatography on an Isolera Prime (Biotage) equipped with a SNAP KP-C18 60 g column equilibrated with 5 CV of the starting solvent mixture. The flow rate was 50 mL/min., the 210/254 nm wavelengths were

monitored, and peaks that absorbed above 20 mAU were collected. The gradient was as follows: 5% MeCN/water + 0.1% TFA for 3 CV, 5-40% MeCN/water + 0.1% TFA over 14 CV, 100% MeCN + 0.1% TFA for 3 CV. Pure fractions as determined by LC-MS were combined and concentrated *in vacuo* to an orange-tinted solid which analysis showed still contained piperidine. The product was triturated with hexanes, redissolved in minimal acetone, and precipitated with hexanes to yield 15.8 mg (28.4%) of an off-white/slightly tan solid. ^1H NMR (500 MHz, Methanol- d_4) δ 8.70 (d, $J = 2.1$ Hz, 1H), 8.15 (d, $J = 9.0$ Hz, 1H), 7.76 (dd, $J = 9.0, 2.1$ Hz, 1H), 4.26 (s, 2H), 3.84 – 3.74 (m, 6H), 3.16 (t, $J = 5.0$ Hz, 2H). $^{13}\text{C}\{^1\text{H}\}$ NMR (126 MHz, Methanol- d_4) δ 171.26, 150.07, 140.01, 138.07, 137.04, 126.07, 122.38, 114.04, 113.12, 72.05, 71.41, 71.30, 67.97, 40.55. HRMS (ESI) m/z : $[\text{M} + \text{H}]^+$ Calcd for $\text{C}_{14}\text{H}_{17}\text{N}_4\text{O}_3\text{S}$ 321.1016; Found 321.1012.

CBT-miniPEG-TAMRA (mix of 5- and 6- isomers). In a foil-wrapped flame-dried 10-mL pear-shaped flask equipped with a stir bar, CBT-miniPEG (7.7 mg, 0.024 mmol) and TAMRA-NHS (12.54 mg, 0.024 mmol) were dissolved in anhydrous DMF (1 mL). The solution turned fuchsia. DIPEA (not anhydrous; 8.37 μL , 0.048 mmol, 2 equiv.) was added, turning the solution a slightly paler fuchsia. The reaction was allowed to stir overnight at r.t., then, as starting material remained (LC-MS), another aliquot of DIPEA (anhydrous, 8.37 μL , 0.048 mmol, 2 equiv.) was added. Additional TAMRA-NHS was added 6 h and 23.5 h after the DIPEA (total TAMRA-NHS was 44.393 mg, 0.084 mmol, 3.5 equiv.). One hour after the final TAMRA-NHS was added, LC-MS indicated disappearance of the CBT-miniPEG starting material, suggesting that 3 equiv. of the NHS ester is ideal. The reaction was then directly purified by preparative C18 RP-HPLC with the following gradient: 15% MeCN/water + 0.1% TFA for 7 min., 15-30% MeCN/water + 0.1% TFA over 1 min., 30-50% MeCN/water + 0.1% TFA over 30 min., 50-100% MeCN/water + 0.1% TFA over 1 min., 100% MeCN + 0.1% TFA for 6 minutes. The two product isomers were separable. Pure fractions (determined by LC-MS) were combined and concentrated *in vacuo*, taken up in water, and lyophilized to yield 9 mg (51.1%) of the major product isomer as a fuchsia solid. ^1H NMR (500 MHz, Acetone- d_6) δ 8.71 (d, $J = 1.9$ Hz, 2H), 8.55 (s, 1H), 8.33 (dd, $J = 7.9, 1.8$ Hz, 1H), 8.04 (d, $J = 8.9$ Hz, 1H), 7.86 (dd, $J = 9.0, 2.1$ Hz, 1H), 7.39 (d, $J = 7.9$ Hz, 1H), 7.03 (d, $J = 9.4$ Hz, 2H), 6.95 (dd, $J = 9.4, 2.4$ Hz, 2H), 6.88 (d, $J = 2.4$ Hz, 2H), 4.23 (s, 2H), 3.88 – 3.77 (m, 6H), 3.75 (q, $J = 5.4$ Hz, 2H), 3.34 (s, 12H), 2.67 (s, 1H), 2.09 (s, 1H). $^{13}\text{C}\{^1\text{H}\}$ NMR (126 MHz, Acetone- d_6) δ 157.96, 149.22, 140.12, 137.58, 137.30, 135.83, 132.13, 131.61, 125.71, 121.96, 114.87, 114.10, 113.75, 112.22, 97.16, 71.81, 71.19, 70.58, 70.06, 40.99, 40.51. HRMS (ESI) m/z : $[\text{M} + \text{H}]^+$ Calcd for $\text{C}_{39}\text{H}_{37}\text{N}_6\text{O}_7\text{S}$ 733.2439; Found 733.2421.

CBT-miniPEG-Alexa Fluor 555. The structure of Alexa Fluor 555 is not reported, so no percent yields were calculated. The molecular weight of the Alexa Fluor 555-NHS (tris(triethylammonium) salt) is approximately 1250 g/mol. In a 1-mL conical vial equipped with a conical stir bar and covered with a rubber septum, Alexa Fluor 555-NHS (1 mg, ~ 0.0008 mmol) dissolved in anhydrous DMF (0.075 mL) was added to CBT-miniPEG

(0.438 mg, 0.00137 mmol, 1.7 equiv.). Anhydrous DMF (~0.025 mL, to 0.1 mL) was used to rinse the remaining dye out of the glass pipette used to transfer the DMF solution from the original AF555 vial. DIPEA (0.0006 mL, 0.00352 mmol, 4.4 eq.; dried over 4 Å molecular sieves) was added, and the solution was lightly covered with foil and allowed to stir under nitrogen at room temperature. As the reaction did not proceed over two days, anhydrous triethylamine (~0.01 mL) was added and the reaction stirred overnight in a metal heating block set to 27-29 °C. The reaction mixture was then diluted slightly with water and directly purified by reversed-phase C18 chromatography on an Isolera Prime (Biotage) equipped with a SNAP KP-C18 60 g column equilibrated with 5 CV of the starting solvent mixture. The flow rate was 100 mL/min., the 210/254 nm wavelengths were monitored, and all fractions were collected. The gradient was as follows: 5% MeCN/water + 0.1% TFA for 3 CV, 5-18% MeCN/water + 0.1% TFA over 5.2 CV, 18% MeCN/water + 0.1% TFA for 2.3 CV, 18-30% MeCN/water + 0.1% TFA over 4.7 CV, 100% MeCN + 0.1% TFA for 3 CV. Pure fractions were determined by LC-MS, combined, and concentrated *in vacuo*, then resuspended in Milli-Q water and lyophilized overnight to yield 1.407 mg of a purple-pink solid. ¹H NMR (500 MHz, Methanol-*d*₄) δ 8.64 (s, 1H), 8.53 (s, 1H), 8.15 (d, J = 9.0 Hz, 1H), 7.94 (dd, J = 19.9, 8.7 Hz, 5H), 7.81 (d, J = 9.0 Hz, 1H), 7.51 (s, 2H), 6.65 (d, J = 15.5 Hz, 2H), 4.39 (s, 5H), 4.19 (s, 2H), 3.79 (s, 1H), 3.74 (d, J = 4.7 Hz, 2H), 3.70 – 3.63 (m, 2H), 3.55 (t, J = 5.5 Hz, 2H), 3.45 (d, J = 2.2 Hz, 1H), 3.23 – 3.11 (m, 11H), 3.01 (d, J = 6.0 Hz, 7H), 2.87 (s, 1H), 2.70 (s, 3H), 2.44 – 2.16 (m, 9H), 1.97 (d, J = 11.2 Hz, 4H), 1.85 – 1.65 (m, 19H), 1.44 – 1.24 (m, 10H), 1.13 (s, 3H), 0.90 (s, 2H), 0.61 (s, 2H).

Alk-CBT beads. Alkyne magnetic beads, 1 μm (30-50 nmol alkyne groups/mg, 10 mg/mL, 1 mL) were purchased from Click Chemistry Tools. The beads were washed several times with Ultrapure DNase/RNase-free Distilled Water to remove preservative, using a DynaMag-2 magnet to trap the beads while the supernatant was removed. The beads were resuspended in 1 mL Ultrapure water. Then, the following copper-click reagents were added to a nuclease-free Eppendorf tube in order: 1 μL of 50 mM CuSO₄ (1 equiv.), 6 μL of 50 mM BTAA (6 equiv.), ~0.5 μL of 1 M CBT-PEG₄-azide (10 equiv.), 25 μL of a freshly prepared 100 mM sodium ascorbate solution (50 equiv.), and 100 μL of the alkyne beads (assumed to be 0.5 mM; 1 equiv.). The tube was incubated for 30 min. at r.t., using an Eppendorf Thermomixer R programmed to do cycles of 1400 rpm for 1 min., then 0 rpm for 10 s. The beads were washed as above and resuspended in 100 μL sterile water, transferred to an LC-MS vial, and stored at 4 °C. Infrared (IR) spectroscopy showed a new peak at approximately 1680 cm⁻¹ (medium strength), which matches those reported for azides.

DBCO-CBT beads. DBCO magnetic beads, 1 μm (30-50 nmol alkyne groups/mg, 10 mg/mL, 1 mL) were purchased from Click Chemistry Tools. A 100-μL aliquot of the beads was washed several times with 1x PBS to remove preservative, using a DynaMag-2 magnet to trap the beads while the supernatant was removed. The beads were resuspended in 90 μL 1x PBS and 9.5 μL of DMSO (HybriMax), ~0.5 μL of 1 M CBT-PEG₄-azide (10 equiv.) was added, and the reaction was gently stirred at r.t. overnight.

The beads were washed as above and resuspended in 100 μ L Ultrapure nuclease-free water, transferred to an LC-MS vial, and stored at 4 $^{\circ}$ C. Infrared (IR) spectroscopy was inconclusive, so these beads were used less often than the alk-CBT beads.

NCL peptide (Ac-Ala-Val-Pro-Phe-Gly-benzyl thioester). In a 50-mL peptide synthesis vessel equipped with house nitrogen and vacuum lines, 4-sulfamylbutyryl AM resin (1 g, 1.1 mmol) was allowed to swell in 4:1 DCM/NMP (5 mL) for 1 h. In a 20-mL vial, DIPEA (1.87 mL, 11 mmol, 10 equiv.) was added to a solution of Fmoc-Gly-OH (1.63 g, 5.5 mmol, 5 equiv.) and COMU (2.365 g, 5.5 mmol, 5 equiv.) in 4:1 DCM/NMP (5 mL). The vial was capped and gently inverted with occasional venting until all solid dissolved into the resulting dark red solution. This solution was added to the peptide synthesis vessel after draining the solution used to swell the resin, and the mixture was gently agitated with nitrogen for 24 h at room temperature, rinsing the sides of the vessel with DCM as needed. After 24 h, the solution was drained and the resin thoroughly rinsed, then the Fmoc-Gly-OH coupling was repeated. After another 24 h, the solution was drained, the resin was thoroughly rinsed, and any free functional groups were acetylated using a solution of 10% acetic anhydride and 5% DIPEA in DMF or NMP, agitating with nitrogen for 40 minutes. The solution was drained, then the resin was rinsed and dried under high vacuum for 15 minutes. The resin was shrunk in MeOH for 15 minutes and left under high vacuum overnight to yield 1.4385 g of a pale tan resin. Loading was calculated to be 0.66 mmol/g following the method described below; resin was stored at 4 $^{\circ}$ C until use. Deprotection of Fmoc groups, coupling of amino acids, and acetylation of the N-terminus were performed according to the general procedures for manual peptide synthesis, starting with 0.0618 g (0.0408 mmol) of the Fmoc-Gly-functionalized resin. To begin cleavage of the peptide thioester from the resin, after thoroughly washing the resin with NMP, a mixture of activated alumina (200 mg), NMP (3.5 mL), DIPEA (180 μ L, 1.03 mmol, 25.75 equiv.), and iodoacetonitrile (280 μ L, 3.8 mmol, 95 equiv.) was filtered through a 10-mL syringe equipped with a 0.2 μ m filter and into the peptide synthesis vessel.²⁷ The vessel was wrapped in foil and the resin agitated with nitrogen for 24 h at room temperature. After 24 h, the solution was drained and the resin thoroughly rinsed with DCM. A clean roundbottom flask was attached, then DMF (5 mL), benzyl mercaptan (0.049 mL, 0.4 mmol, 10 equiv.), and catalytic sodium thiophenolate (1.3 mg, 0.01 mmol, 0.25 equiv.) were added and the resin agitated with nitrogen overnight, still with the reaction vessel covered in foil. [Sodium thiophenolate was made by stirring thiophenol (5 mL, 49 mmol) and 4 M NaOH (12.5 mL) together at room temperature for 10 minutes, removing the resulting water *in vacuo*, and drying under vacuum to yield 4.4077 g (68%) of a white crystalline solid.] The next day, the solution was drained into the clean roundbottom flask and the reaction was repeated, but 0.5 equiv. of sodium thiophenolate (2.6 mg, 0.02 mmol) was used instead of 0.25 equiv. The next day, the solution was again drained, rinsing with DCM, and the solvent was removed *in vacuo*. After resuspending in minimal DMF, the peptide was purified by preparative RP-HPLC with the following gradient: 0% solvent B for 2 minutes, 0-60% solvent B over 18 minutes, 60-100% solvent B over 2 minutes, 100% solvent B for 3 minutes. Pure fractions were combined and concentrated *in vacuo*, taken

up in water, and lyophilized to yield 6.838 mg (25.7%) of a tan fluffy solid. HRMS (ESI) m/z: [M + H]⁺ Calcd for C₃₄H₄₆N₅O₆S 652.3163; Found 652.3150.

Resin loading procedures were adapted from reference 28: Fmoc-Gly-functionalized 4-sulfamylbutyryl AM resin (3.078 mg) was stirred with 20% piperidine in NMP (9.234 mL) for 30 minutes. A Thermo Fisher Scientific NanoDrop 2000/c spectrophotometer was used to measure the solution's absorbance at 301 nm three times (blinking with 20% piperidine in NMP each time), giving an average absorbance of 0.171 AU. At 301 nm, the extinction coefficient of the dibenzofulvene-piperidine adduct formed after deprotection is 7800 L mol⁻¹cm⁻¹. By the Beer-Lambert Law, the concentration (M) of the sample is $Abs_{301} \times (\epsilon_{301})^{-1} \times (\text{path length})^{-1}$, where the path length of the instrument is defined as 0.1 cm. Substituting in the appropriate values, the concentration was determined to be 0.22 mM. As the reaction was done at a concentration of 1 g of resin per 3 L of solvent, the final resin loading was 0.66 mmol/g.

DIVERSE peptide hits. Crude peptides were ordered from Elim Biopharmaceuticals, Inc., then purified by preparative RP-HPLC. The gradient was as follows: X% solvent B for 2 minutes, X-Y% solvent B over 18 minutes, Y-100% solvent B over 2 minutes, 100% solvent B for 3 minutes. The X and Y percentages were as follows: 30 and 60 for hit peptides 1 and 9; 15 and 30 for hit peptides 2 and 7; 20 and 40 for hit peptides 3 and 11; 10 and 30 for hit peptides 4, 5, 6, 8, 10, 12, and 13; and 15 and 40 for hit peptide 14. Peptide 2 was lost during purification due to an unexpectedly low UV absorbance. Pure fractions (based on LC-MS) were combined and concentrated *in vacuo*, taken up in water, and lyophilized. The lyophilized products were resuspended to a concentration of 1 mM in PBS and used in the CBTNAc binding assay and cysteine competition assay without further characterization.

SUPPLEMENTARY FIGURES

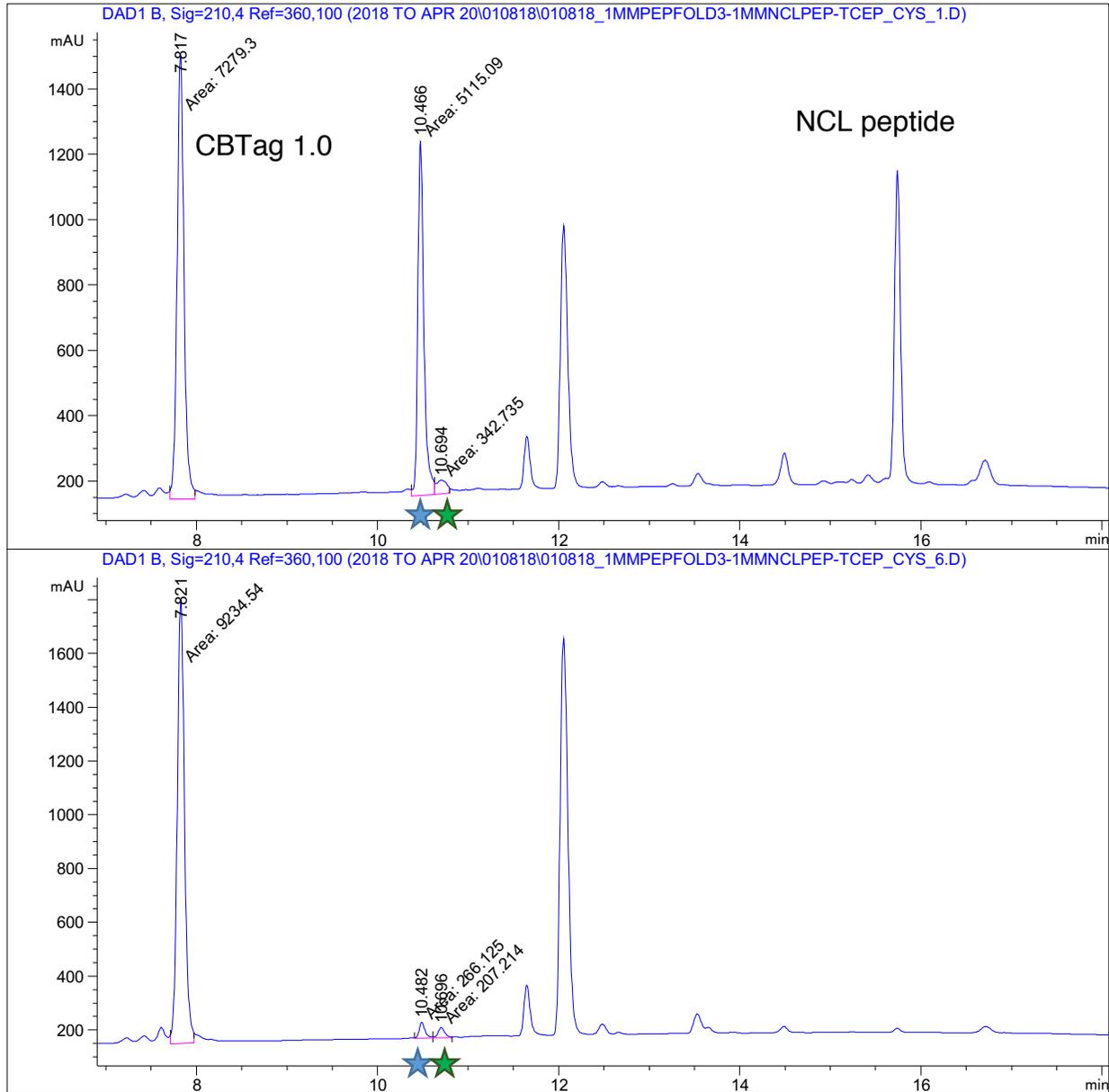
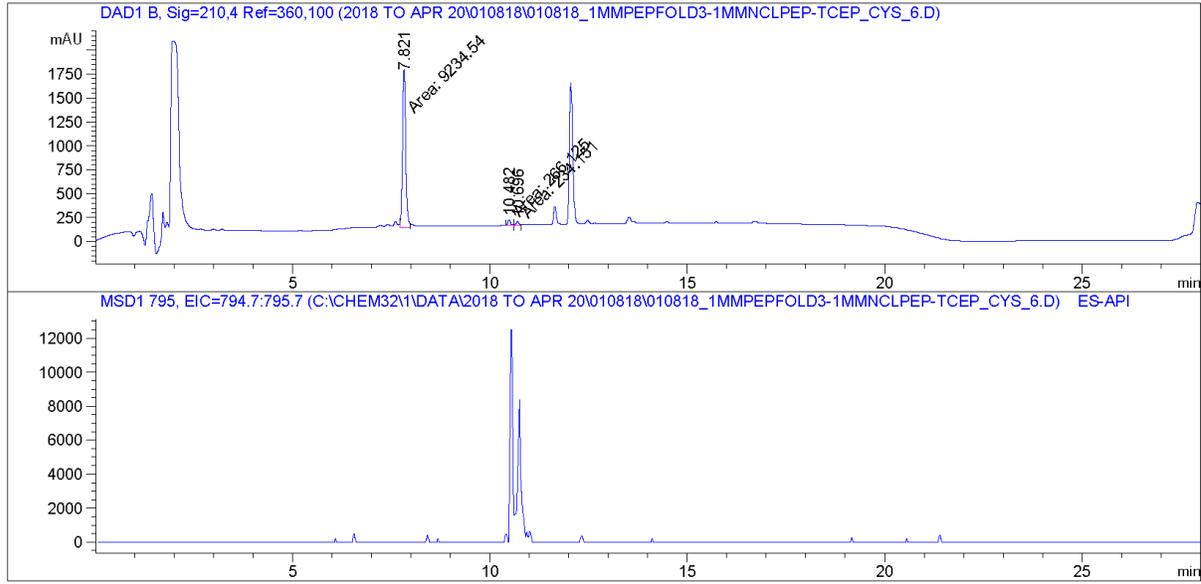


Figure 3.S1. Cysteine treatment identifies peaks that contain cysteine adducts. The top spectrum collected at an early timepoint, while the bottom spectrum was collected at the end of the time-course assay. Based on these results and the mass spectrum in Figure 3.S2, the blue star (minute 10.5) indicates the cysteine adduct, and the green star (minute 10.7) is the lysine adduct.

Injection Date : 1/10/18 3:41:10 PM Inj : 1
 Inj Volume : 15.000 µl
 Different Inj Volume from Sequence ! Actual Inj Volume : 10.000 µl
 Acq. Method : D:\CHEM32\1\METHODS\AP_C18_INSERT.M
 Last changed : 11/10/17 3:07:57 PM by Mason
 Analysis Method : Z:\DESKTOP\BERTOZZI LAB\LCMS INFO\LCMS METHODS TO 112917\SGLK_BERTOZZIBP_ WITHINSERT.M
 Last changed : 5/8/18 2:55:52 AM
 (modified after loading)
 Additional Info : Peak(s) manually integrated



=====
 Area Percent Report
 =====

Sorted By : Signal
 Multiplier: : 1.0000
 Dilution: : 1.0000
 Do not use Multiplier & Dilution Factor with ISTDs

Signal 1: DAD1 B, Sig=210,4 Ref=360,100

| Peak # | RetTime [min] | Type | Width [min] | Area [mAU*s] | Height [mAU] | Area % |
|--------|---------------|------|-------------|--------------|--------------|---------|
| 1 | 7.821 | MF | 0.0931 | 9234.53516 | 1652.48157 | 94.8610 |
| 2 | 10.482 | MF | 0.0756 | 266.12546 | 58.68875 | 2.7338 |
| 3 | 10.696 | MM | 0.0922 | 234.15126 | 42.34366 | 2.4053 |

Figure 3.S2. The LC-MS spectrum and the corresponding extracted mass ions at the final time point of cysteine treatment show that very little CBTag 1.0 forms a stable amide bond with the NCL peptide. 795 m/z, minute 10.5 and 10.7 – [CBTag 1.0-NCL peptide + 2H]²⁺. The minute 10.5 peak disappeared during cysteine treatment, suggesting that that peak is the cysteine-NCL peptide adduct and the minute 10.7 peak is the lysine-NCL adduct. If so, only 2.5% of the peak area that contains CBTag 1.0 corresponds to the lysine adduct.

| | |
|---|---|
| interhelix insert (sfGFP8) | AAATTTATTTGCACCACCGGT GGCGGGGGGGGGCCACCCGGATCCGTGCCGAAAGGCGGGC CGGCGGC AAACTGCCGTTCCGTGGCCGACCCTGGTGACCACCCTGACCTATGGCGTTCAGTGCTTT AGCCGCTATCCGGATCATATGAAACGCCATGATTTCTTTAAAAGCGCGATGCCGGAAGGCTATGTGCA GGAACGTACCATTAGCTTCAAAGATGATGGCACCTATAAAACCCGTGCCGGAAGTTAAATTTGAAGGCG ATACCCTGGTGAACCGCATTGAACTGAAAGGTATTGATTTTAAAGAAGATGGCAACATTCTGGGTCATA AACTGGAATATAATTTCAACAGCCATGCGGTGTATATTACCGCCGATAAACAGAAAAATGGCATCAAAG CGAACTTTAAATCCGTCAACAACGTGGAAGATGGTAGCGTGCAGCTGGCGGATCATTATCAGCAGAAT ACCCCGATTGGTGTATGGCCCGGTGCTGCTGCCGGATAATCATTATCTGAGCACCCAGAGCGTTCTGA GCAAAGATCCGAATGAAAAACGTGATCATATGGTGCTGCTGGAATTTGTTACCGCCGCGGGCATTACC CACGGTATGGATGAACTGTATAAAGGCAGCCACCATCATCATCACCATTGA |
| Sequence of C-term insert (sfGFP9) | ATGGTTAGCAAAGGTGAAGAACTGTTTACCGCGTGTGCGGATTCTGGTGGAACTGGATGGTGTATGT GAATGGCCATAAATTTAGCGTTCGTGGCGAAGGCCAAGGTGATGCGACCAACGGTAAACTGACCCTG AAATTTATTTGCACCACCGGTAAACTGCCGTTCCGTGGCCGACCCTGGTGACCACCCTGACCTATGG CGTTCAGTGCTTTAGCCGCTATCCGGATCATATGAAACGCCATGATTTCTTTAAAAGCGCGATGCCGG AAGGCTATGTGCAGGAACGTACCATTAGCTTCAAAGATGATGGCACCTATAAAACCCGTGCCGGAAGTT AAATTTGAAGGCGATACCCTGGTGAACCGCATTGAACTGAAAGGTATTGATTTTAAAGAAGATGGCAA CATTCTGGGTATAAAGTGAATATAATTTCAACAGCCATGCGGTGTATATTACCGCCGATAAACAGAA AAATGGCATCAAAGCGAACTTTAAATCCGTCAACAACGTGGAAGATGGTAGCGTGCAGCTGGCGGAT CATTATCAGCAGAATACCCCGATTGGTGTATGGCCCGGTGCTGCTGCCGGATAATCATTATCTGAGCAC CCAGAGCGTTCTGAGCAAAGATCCGAATGAAAAACGTGATCATATGGTGCTGCTGGAATTTGTTACCG CCGCGGGCATTACCCACGGTATGGATGAACTGTATAAA GGCGGGGGGGGGCCACCCGGATCCGT GCCGAAAGGCGGC TGA |

REFERENCES

1. Chalfie, M.; Tu, Y.; Euskirchen, G.; Ward, W. W.; Prasher, D. C. Green fluorescent protein as a marker for gene expression. *Science* **1994**, *263*, 802-805.
2. Wang, L.; Brock, A.; Herberich, B.; Schultz, P. G. Expanding the Genetic Code of *Escherichia coli*. *Science* **2001**, *292*, 498-500.
3. Zhang, C.; Welborn, M.; Zhu, T.; Yang, N. J.; Santos, M. S.; Van Voorhis, T.; Pentelute, B. L. π -Clamp-mediated cysteine conjugation. *Nat. Chem.* **2015**, *8*, 120-128.
4. Ojida, A.; Honda, K.; Shinmi, D.; Kiyonaka, S.; Mori, Y.; Hamachi, I. Oligo-Asp Tag/Zn(II) Complex Probe as a New Pair for Labeling and Fluorescence Imaging of Proteins. *J. Am. Chem. Soc.* **2006**, *128*, 10452-10459.
5. Jing, C.; Cornish, V. W. Chemical Tags for Labeling Proteins Inside Living Cells. *Acc. Chem. Res.* **2011**, *44*, 784-792 and references cited therein.
6. Spicer, C. D.; Davis, B. G. Selective chemical protein modification. *Nat. Commun.* **2014**, *5*:4740 and references cited therein.
7. Cal, P. M. S. D.; Bernardes, G. J. L.; Gois, P. M. P. Cysteine-Selective Reactions for Antibody Conjugation. *Angew. Chem. Int. Ed.* **2014**, *53*, 10585-10587 and references cited therein.
8. Stephanopoulos, N.; Francis, M. B. Choosing an effective protein bioconjugation strategy. *Nat. Chem. Biol.* **2011**, *7*, 876-884 and references cited therein.
9. Lotze, J.; Reinhardt, U.; Seitz, O.; Beck-Sickinger, A. G. Peptide-tags for site-specific protein labelling *in vitro* and *in vivo*. *Mol. BioSyst.* **2016**, *12*, 1731-1745 and references cited therein.
10. Krall, N.; da Cruz, F. P.; Boutureira, O.; Bernardes, G. J. L. Site-selective protein-modification chemistry for basic biology and drug development. *Nat. Chem.* **2016**, *8*, 103-113 and references cited therein.
11. Ren, H.; Xiao, F.; Zhan, K.; Kim, Y.-P.; Xie, H.; Xia, Z.; Rao, J. Biocompatible Condensation Reaction for the Labeling of Terminal Cysteine Residues on Proteins. *Angew. Chem. Int. Ed.* **2009**, *48*, 9658-9662.
12. Nguyen, D. P.; Elliott, T.; Holt, M.; Muir, T. M.; Chin, J. W. Genetically Encoded 1,2-Aminothiols Facilitate Rapid and Site-Selective Protein Labeling via a Bio-orthogonal Cyanobenzothiazole Condensation. *J. Am. Chem. Soc.* **2011**, *133*, 11418-11421.
13. Ramil, C. P.; An, P.; Yu, Z.; Lin, Q. Sequence-specific 2-cyanobenzothiazole ligation. *J. Am. Chem. Soc.* **2016**, *138*, 5499-5502.
14. James, A. M.; Hoogewijs, K.; Logan, A.; Hall, A. R.; Ding, S.; Fearnley, I. M.; Murphy, M. P. Non-enzymatic *N*-acetylation of Lysine Residues by AcetylCoA Often Occurs via a Proximal *S*-acetylated Thiol Intermediate Sensitive to Glyoxalase II. *Cell Rep.* **2017**, *18*, 2105-2112.

15. Takamura, Y.; Nomura, G. Changes in the Intracellular Concentration of Acetyl-CoA and Malonyl-CoA in Relation to the Carbon and Energy Metabolism of *Escherichia coli* K12. *J. Gen. Microbiol.* **1988**, *134*, 2249-2253.
16. Alexa Fluor® 555 Carboxylic Acid, tris(triethylammonium) salt. <https://www.thermofisher.com/order/catalog/product/A33080> (accessed Mar. 28, 2017).
17. Kawakami, T.; Ogawa, K.; Goshima, N.; Natsume, T. DIVERSE System: De Novo Creation of Peptide Tags for Non-enzymatic Covalent Labeling by In Vitro Evolution for Protein Imaging Inside Living Cells. *Chem. Biol.* **2015**, *22*, 1671-1679.
18. Blanchet, S.; Cornu, D.; Argentini, M.; Namy, O. New insights into the incorporation of natural suppressor tRNAs at stop codons in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **2014**, *42*, 10061-10072.
19. Thévenet, P.; Shen, Y.; Maupetit, J.; Guyon, F.; Derreumaux, P.; Tufféry, P. PEP-FOLD: an updated de novo structure prediction server for both linear and disulfide bonded cyclic peptides. *Nucleic Acids Res.* **2012**, *40*, W288-293.
20. Shen, Y.; Maupetit, J.; Derreumaux, P.; Tufféry, P. Improved PEP-FOLD approach for peptide and miniprotein structure prediction. *J. Chem. Theor. Comput.* **2014**, *10*, 4745-4758.
21. Hénaut, A.; Danchin, A. Analysis and Predictions from *Escherichia coli* Sequences, or *E. coli* In Silico.; In *Escherichia coli and Salmonella*, Vol. 2; Neidhardt FC ed.; ASM press: Washington, D.C., 1996; Ch. 114:2047-2066.
22. European Molecular Biology Laboratory. "Protein Expression: *E. coli* Optimisation of Expression Levels." https://www.embl.de/pepcore/pepcore_services/protein_expression/ecoli/optimisation_expression_levels/ (accessed May 9, 2018).
23. Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A.; Protein Identification and Analysis Tools on the ExPASy Server; In *The Proteomics Protocols Handbook*; Walker, J. M.: Humana Press, 2005; pp. 571-607.
24. Pedelacq, J. D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **2006**, *24*, 79–88.
25. Yamagishi, Y.; Shoji, I.; Miyagama, S.; Kawakami, T.; Goto, Y.; Suga, H. Natural Product-like Macrocyclic N-Methyl-Peptide Inhibitors against a Ubiquitin Ligase Uncovered from a Ribosome-Expressed De Novo Library. *Chem. Biol.* **2011**, *18*, 1562-1570.
26. Walker, S. E.; Lorsch, J. RNA Purification – Precipitation Methods. *Methods Enzymol.* **2013**, *530*, 337-343.

27. Marques, M. A.; Citron, D. M.; Wang, C. C. Development of Tyrocidine A Analogues with Improved Antibacterial Activity. *Bioorg. Med. Chem.* **2007**, *15*, 6667-6677.
28. *Resin Loading Measurement by Fmoc Cleavage*; Technical Support Information Bulletin 1198; AAPPTec. <https://www.peptide.com/custdocs/1198.pdf> (accessed February 21, 2018).

Chapter 4

Synthesis of solvatochromic probes to label the mycobacterial cell wall

The biological experiments in this chapter were performed in collaboration with Mireille Kamariza.

Chapter 4. Synthesis of solvatochromic probes to label the mycobacterial cell wall

INTRODUCTION

Tuberculosis (TB), an infectious disease caused by the bacterial pathogen *Mycobacterium tuberculosis* (Mtb), is one of the top ten causes of death worldwide and the leading cause of death by a single infectious agent.¹ The World Health Organization (WHO) reports that one-third of the world's population is infected with TB. The reasons for this ongoing pandemic are manifold. Many of these cases are latent, meaning patients are asymptomatic and noninfectious.²⁻³ Without treatment, 5-10% of latent patients will develop the active disease in their lifetimes, but the underlying mechanisms of TB reactivation remain poorly understood. Moreover, no effective TB vaccine is available for adults, and the standard treatment for drug-susceptible TB consists of daily ingestion of multidrug cocktails over 26 weeks.³⁻⁵ While the cure rate is 90% (70% in HIV-positive populations), the length of this treatment course and its side effects decrease patient compliance, a significant problem when non-compliance is considered one of the major drivers of the emergence of multidrug-resistant (MDR) and extremely drug-resistant (XDR) TB.⁶⁻⁸ MDR TB patients have a 52% cure rate after 18-24 months, 14,000 pills, and 240 injections.⁸ XDR TB patients have a 20% cure rate, and their treatment can cost up to \$513,000.⁸ Patients co-infected with HIV present additional challenges, with a lower cure rate overall and a higher likelihood of progression, when left untreated, from latent to active TB.

Current diagnostics also have limitations, particularly those used at the point of care in high-burden, low-resource settings.⁹⁻¹¹ Mtb is notorious for its slow growth, requiring on average four weeks to obtain a conclusive test result using culture methods. Rapid (same-day) and accurate diagnosis is essential, but test sensitivity and specificity can be problematic in low-resource areas, where sputum smear microscopy coupled with the color-based Ziehl-Neelson test¹²⁻¹⁴ or the fluorescent auramine-rhodamine stain¹⁵⁻¹⁶ is the gold-standard TB diagnostic technique. Both tests rely on a hydrophobic layer of the mycobacterial cell wall known as the mycolic acid layer or mycomembrane to retain hydrophobic molecules such as the staining reagents mentioned above. Extensive processing steps to remove excess dye from cellular debris and other bacteria in the sputum sample makes the successful application of these tests highly dependent on technician skill.¹⁷⁻¹⁹ Additionally, these dyes do not provide a readout on bacterial viability, and so other lengthy techniques are required to assess efficacy of drug treatment.

The complex mycobacterial cell wall is both an obstacle to drug penetration and a target for drug development. As discussed in Chapter 1, mycobacteria and other members of the Corynebacterineae suborder have unique cell wall features such as arabinogalactan and mycolic acid layers in addition to the peptidoglycan layer common to Gram-positive and Gram-negative bacteria. The mycolic acid layer contains trehalose-6-monomycolate (TMM) and trehalose-6,6'-dimycolate (TDM), also known as cord factor; only

mycobacteria and certain other members of the Actinobacteria phylum incorporate lipidated trehalose into their cell walls in this manner. The thick and waxy nature of the mycomembrane protects the cell from environmental stresses and provides a physical barrier to most antibiotics. The standard treatment for TB thus includes drugs with cell wall-disrupting mechanisms of action; isoniazid, for example, inhibits mycolic acid synthesis.²⁰ Also known as isonicotinylhydrazide (INH), isoniazid is a prodrug that is converted to an acyl radical by the mycobacterial catalase-peroxidase enzyme KatG. The acyl radical then forms isonicotinic acyl-NADH, which tightly binds and inhibits a long-chain enoyl-acyl carrier protein reductase (InhA) essential for mycolic acid synthesis (Figure 4.1).

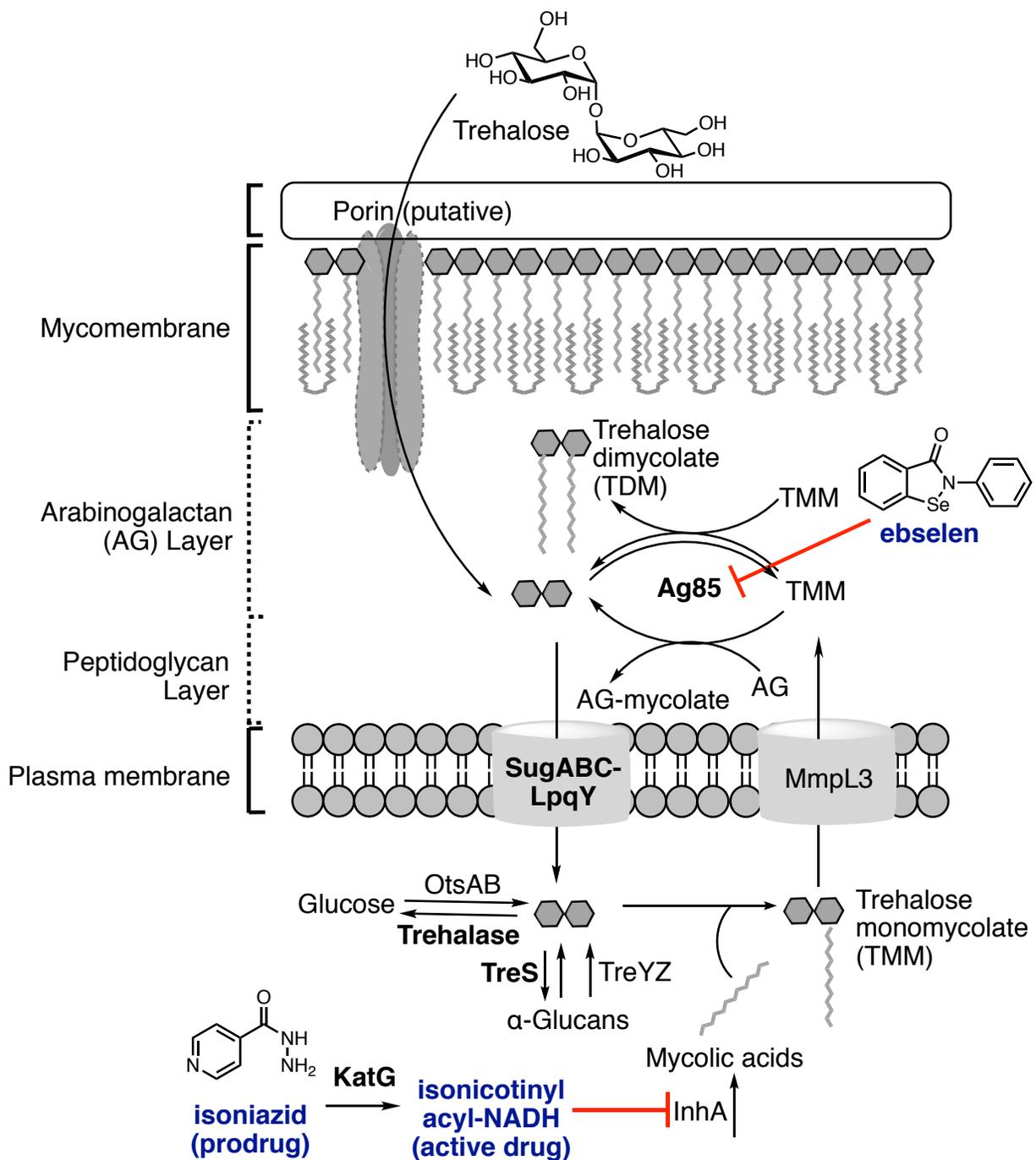


Figure 4.1. An overview of the mycobacterial trehalose metabolic pathways and select drugs that inhibit them. Extracellular trehalose passes into the cell envelope through a putative porin. It is then either transported into the cytosol by the SugABC-LpqY complex for further processing (the “recycling pathway”) or directly modified to trehalose monomycolate (TMM) by the antigen 85 (Ag85) complex. Ebselen can inhibit TMM formation by binding to the Ag85 complex. Isoniazid is activated by KatG before going on to inhibit mycolic acid synthesis. Bolded black words are proteins and bolded blue words are small molecule inhibitors relevant to this chapter.

Many aspects of Mtb's physiology and host-pathogen interactions remain poorly understood, however, hindering development of new drugs. For example, while isoniazid kills Mtb rapidly *in vitro*, it must be taken daily during the 26-week treatment period in order to eradicate infection in humans. This differential potency is thought to be due in part to the effect of the host microenvironment on the epigenetic²¹ and metabolic^{3, 22} state of the bacterium, but a dearth of appropriate chemical probes for noninvasive monitoring of metabolic changes during infection in a living model organism has made this hypothesis difficult to investigate until recently.

In the last decade, metabolic labeling of trehalose glycolipids has been established as an attractive method for visualizing the mycomembrane. While mycobacteria do synthesize trehalose from cytosolic glucose via the OtsAB pathway and from α -glucans via the TreS and TreYZ pathways, they also scavenge extracellular trehalose through what is known as a "recycling pathway" (Figure 4.1).²³ In this process, exogenous trehalose first passes through a putative porin into the cell envelope. There, it can be transported to the cytosol by the transmembrane transporter protein complex SugABC-LpqY,²⁴ where it can be mycolylated to form trehalose monomycolate (TMM), converted to glucose by trehalase, or converted to α -glucans by TreS. Other species-dependent processing can also occur. Cytosolic TMM is then exported by the MmpL3 transporter²⁵⁻²⁷ into the periplasmic space, where the antigen 85 complex (Ag85A [Rv3804c], Ag85B [Rv1885c], Ag85C [Rv0129c])²⁸⁻²⁹ can convert two molecules of TMM into one molecule of trehalose and one molecule of trehalose dimycolate (TDM). Alternatively, rather than being transported into the cytosol, the exogenous trehalose can be converted directly to TMM by the Ag85 complex in the reverse of the reaction in the previous sentence.

Researchers have shown that the enzymes responsible for recycling extracellular trehalose are promiscuous enough to tolerate modifications ranging from small perturbations such as azide groups, alkyne groups, and fluorine atoms to larger fluorophores such as fluorescein (see Figure 1.10 for examples).³⁰⁻³⁸ In particular, the Ag85 complex remains able to catalyze addition of mycolic acids to the 6 position on the sugar, allowing these trehalose derivatives to be incorporated into the mycomembrane as long as the cell is metabolically active. Some probes can also be processed through the cytosolic pathways, though this is less commonly observed.³¹

Previously reported probes have significant disadvantages depending on the application. Fluorophore-bearing probes require extensive washing before imaging in order to reduce background signal. This becomes a problem in situations where washing is not feasible, e.g., in mouse and macrophage infection studies. Azide- and alkyne-based probes require two-step labeling, namely metabolic incorporation of the probe followed by biorthogonal ligation of the fluorophore and/or affinity tag to the probe. The modification of the trehalose is minimal and host membrane penetrance is favorable, but the secondary labeling reagent (e.g., a cyclooctyne-fluorophore) may not be able to reach the mycomembrane easily. Fluorinated probes can be used for positron emission tomography (PET)-based imaging of Mtb in the lungs, but the fluorine-18 isotope that is observed has

a half-life of less than two hours and access to the dedicated, expensive, specialty equipment is required.

Recently, Kamariza and Shieh et al. reported the synthesis and application of DMN-Tre, a fluorogenic trehalose probe that rapidly labels mycobacteria in one step without the necessity of a wash step.³⁸ The probe consists of a trehalose moiety that has been modified at the 6-position with 4-(dimethylamino)-1,8-naphthalimide (DMN), a dye that fluoresces only in very hydrophobic environments and is quenched in the presence of even small amounts of water. DMN is one example of a class of compounds known as solvatochromic dyes. Solvatochromatic reporters change in color or intensity based on the electronic properties of the surrounding environment, making them useful for monitoring changes in hydrophobicity.³⁹⁻⁴² For example, DMN and Nile red derivatives have been previously used to study local hydrophobicity, folding dynamics, and conformational changes in proteins.⁴³⁻⁴⁵ Here, the researchers were able to use this dye to specifically label metabolically active mycobacteria, including Mtb. Mtb was detected in patient sputum samples incubated with 1 mM of DMN-Tre for just 30 minutes at 37 °C without a wash step prior to microscopy. Additionally, the authors showed that pretreating Mtb with a drug cocktail abrogated labeling, opening the possibility of using the probe to monitor treatment efficacy rapidly and on a per cell basis.

However, DMN-Tre has several flaws. For one, its excitation and emission wavelengths (405 and 525 nm, respectively) are not supported on many simpler analytical devices (plate readers, flow cytometers, fluorescence microscopes), which may limit DMN-Tre's potential reach as a point-of-care reagent. Furthermore, the wavelengths overlap with those of many natural molecules which means autofluorescence can be a problem.⁴⁶ These factors, coupled with DMN-Tre's relatively low intensity leave room for significant improvement of the fluorophore. This chapter concerns the synthesis and characterization of four red-shifted solvatochromic trehalose probes, three with high emission intensity relative to DMN-Tre (Figure 4.2). We have shown that the probes reported herein can be used similarly to DMN-Tre. Several of our dyes can be used at concentrations 10x less than were used for DMN-Tre experiments; the one dye used at the same concentration, 3-HC-3-Tre, was much brighter than DMN-Tre and had the least background of all the probes thus far. Differences in labeling behavior between different classes of dye-trehalose conjugates necessitated more in-depth studies of the mechanism of incorporation, some of which are ongoing (e.g., structure-activity relationship studies of the different dyes). Regardless, these probes have potential applications in not only TB diagnostics but also in studies of the effect of host-pathogen interactions and drug treatments on mycobacterial cell wall dynamics.

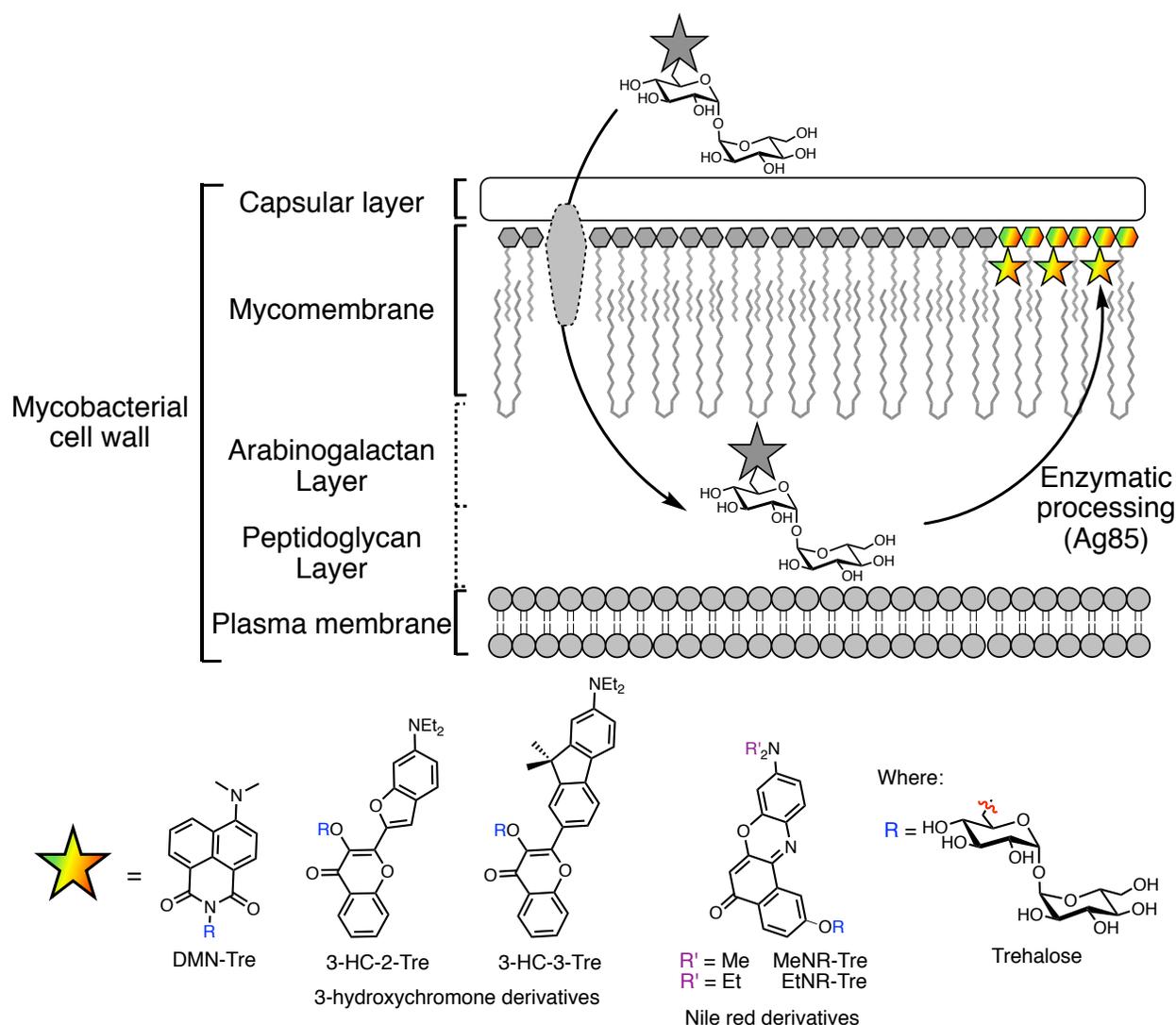


Figure 4.2. Solvatochromic trehalose probes label the mycobacterial mycomembrane. Top panel: Schematic depicting metabolic incorporation of trehalose probes into the mycobacterial cell wall. Bottom panel: Chemical structures of dye-trehalose conjugates described herein.

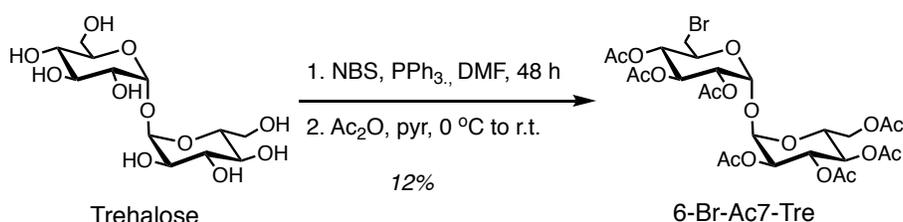
RESULTS AND DISCUSSION

We began by scouring the literature for solvatochromatic structures with strong turn-on in hydrophobic environments to replace DMN. Several such compounds exist that are well characterized, synthetically tractable and currently in use in living systems with minimal perturbation or toxicity (See Fig. 4.3). One well-known example is Nile red, a stain for intracellular neutral lipid droplets.⁴⁷⁻⁵⁰ As might be expected from its name, Nile red has excitation and emission wavelengths similar to those of Red Fluorescent Protein (RFP; maximum excitation/emission at 555/585 nm).⁴⁹ Moreover, Nile red has a very high extinction coefficient ($45,000 \text{ M}^{-1}\text{cm}^{-1}$ in methanol),⁴⁹ causing it to fluoresce brightly at

much lower concentrations than dyes such as DMN ($8800 \text{ M}^{-1}\text{cm}^{-1}$ in TBS buffer).⁴³ Another promising class of solvatochromic dyes is based on a 3-hydroxychromone (3-HC) scaffold.³⁹ These dyes, while not as red-shifted as Nile red, have the advantage of greater tunability thanks to their structure and synthetic modularity. This is discussed in more detail below.

Having identified potential solvatochromes, we considered the method of attachment to trehalose. Unlike DMN, none of these dyes are easily made into phthalic anhydrides that can be reacted with 6-amino-6-deoxy- α,α' -trehalose (6-aminoTre) to form a phthalimide. This left conjugation of a modified dye with 6-aminoTre via functional groups such as carbamates or a route involving another trehalose derivative; we were careful to avoid an amine-based linkage because alkylamines are positively charged at physiological pH, which could undermine incorporation into the mycomembrane. As our synthesis of 6-aminoTre begins with conversion of α,α' -trehalose to 6-bromo-2,2',3,3',4,4',6'-hepta-*O*-acetyl-6-deoxy- α,α' -trehalose (6-Br-Ac7-Tre), we decided to first attempt direct $\text{S}_{\text{N}}2$ displacement of the bromine atom by the aromatic alcohol on a dye.⁵¹ In addition to decreasing the total number of synthetic steps, this approach results in formation of an ether bond, which is significantly more stable to hydrolysis than a naphthalimide functional group.

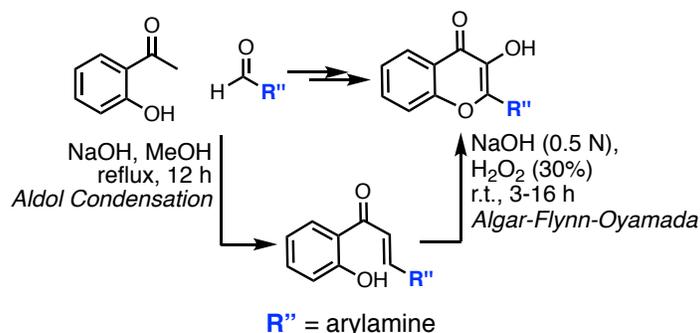
With this synthetic approach decided, we began by synthesizing the 6-Br-Ac7-Tre (Scheme 4.1). An Appel reaction using *N*-bromosuccinimide and triphenylphosphine resulted in a mixture of mono-brominated target compound (6-Br-Tre), a dibrominated side product (6,6'-dibromo-6,6'-dideoxy- α,α' -trehalose) and unreacted starting material. The crude material was then acetylated and purified to give 6-Br-Ac7-Tre.



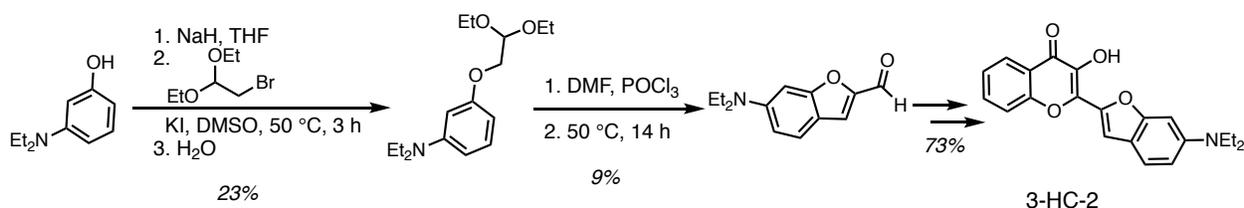
Scheme 4.1. Synthesis of 6-Br-Ac7-Tre from trehalose.

The synthesis of 3-HC dyes ends with an Aldol condensation between a 2'-hydroxyacetophenone derivative and an aminoarylcarbaldehyde, followed by an Algar-Flynn-Oyamada reaction in which the intermediate is treated by hydrogen peroxide and base (Scheme 4.2).⁵² Varying the electronic properties of the dye is as easy or as difficult as choosing an arylamine and functionalizing it with an aldehyde. For initial studies, we chose to synthesize previously reported 3-HC molecules, here referred to as 3-HC-2 and 3-HC-3.⁵³⁻⁵⁶ Following Klymchenko and coworkers' synthesis of the aldehyde precursor to 3-HC-2,⁵³ we first reacted 3-diethylaminophenol with bromoacetaldehyde diethyl acetal (Scheme 4.3). The intermediate was purified then subsequently treated with phosphorus(V) oxychloride and *N,N*-dimethylformamide (DMF) to form the benzofuran as

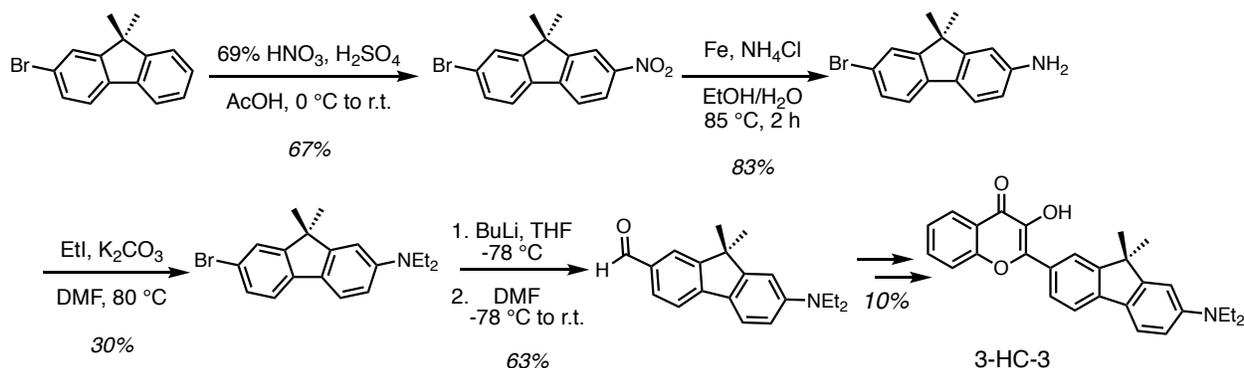
well as to install an aldehyde at the 2-position via a Vilsmeier-Haack reaction. To synthesize the aldehyde precursor to 3-HC-3, we followed the procedures reported by Kucherak et al. (Scheme 4.4).⁵⁵ Briefly, we nitrated 2-bromo-9,9-dimethylfluorene at the 7-position, reduced the nitrate to an amine, alkylated the amine using ethyl iodide, and converted the bromine to an aldehyde through a Bouveault reaction. To form 3-HC-2 and 3-HC-3, each aldehyde was reacted with 2'-hydroxyacetophenone, then treated with hydrogen peroxide.⁵² At first we had unexpected difficulty in reproducibly obtaining the desired products after the hydrogen peroxide reaction, but after much troubleshooting, we optimized the conditions enough to obtain the amount of dye needed for initial experiments.



Scheme 4.2. Synthesis of 3-hydroxychromones.

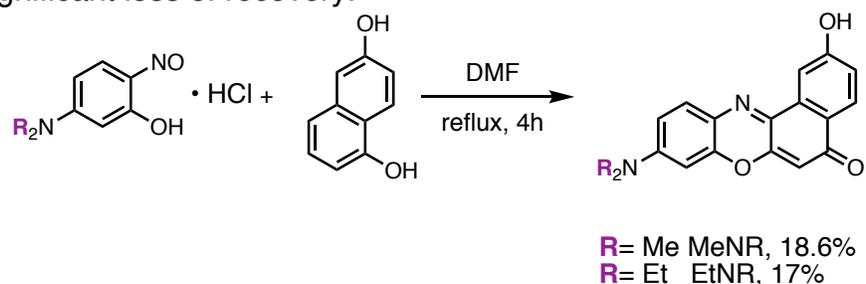


Scheme 4.3. Synthesis of 3-HC-2.



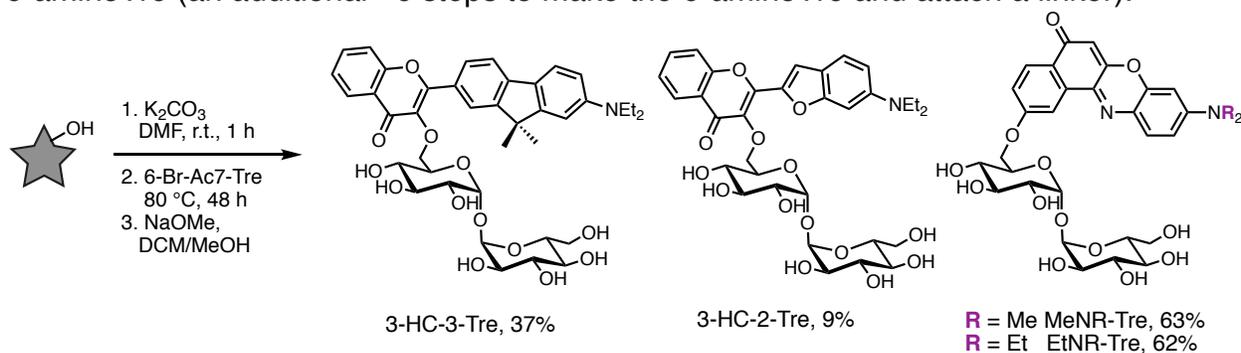
Scheme 4.4. Synthesis of 3-HC-3.

For the Nile red dyes, synthesis was relatively straightforward, but purification was non-trivial. While Nile red itself is commercially available, it does not have a convenient handle for functionalization with trehalose; while the ketone could be reacted with an amine, it would result in an iminium cation at physiological pH, which again could negatively impact mycomembrane incorporation. Thus, we initially focused our efforts on synthesizing 2-hydroxy Nile red and its dimethylamino counterpart, hereafter referred to as EtNR and MeNR, respectively (Scheme 4.5). Refluxing 1,6-dihydroxynaphthalene and a 5-dialkylamino-2-nitrosophenol hydrochloride salt in DMF for 4 hours generated the desired product as well as a dimerized 1,6-dihydroxynaphthalene side product. We empirically found that washing the crude product extensively with water, dissolving the remaining solid in methanol, and purifying by reversed-phase high-performance liquid chromatography (RP-HPLC) using a phenyl-hexyl column with a water/methanol gradient resulted in the highest recovery of pure product. Additionally, because this method avoided introducing additives such as trifluoroacetic acid (TFA), we did not have to neutralize and desalt the product before carrying it forward, a process that had previously resulted in significant loss of recovery.



Scheme 4.5. Synthesis of 2-hydroxy Nile red analogs.

Finally, to create the dye-Tre probes, the aromatic hydroxyl group on the dye was deprotonated with potassium carbonate and used to displace the bromine atom on 6-Br-Ac7-Tre (Scheme 4.6).⁵⁷ The crude dye-Ac7-Tre was then deacetylated with catalytic sodium methoxide and purified to produce the final dye-Tre products in yields ranging from 5-63% over three steps. While the first part of this conjugation is long (48 h), this approach significantly reduced the number of steps necessary to reach the final probes, which otherwise would have been made by forming a carbamate or carbamide bond with 6-aminoTre (an additional ~6 steps to make the 6-aminoTre and attach a linker).



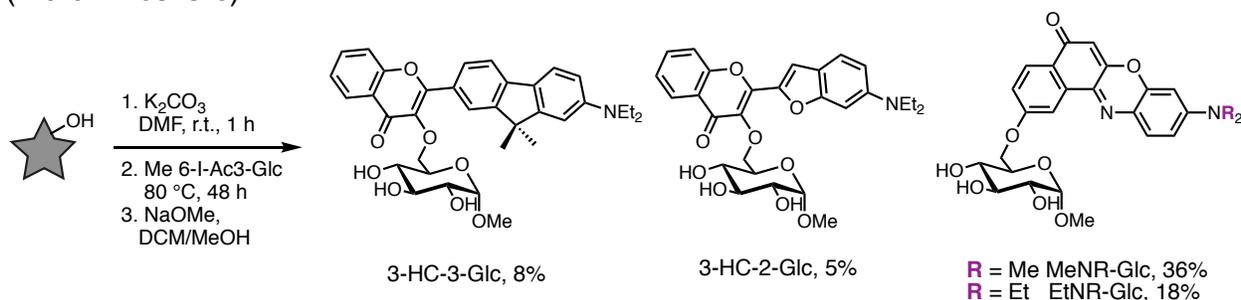
Scheme 4.6. Synthesis of the 6-dye-trehalose (dye-Tre) probes.

For control experiments, we also made 6-modified methyl α -glucoside analogs (Scheme 4.7, 4.8). The addition of the methyl modification to the anomeric hydroxyl group locks the sugar in a configuration that corresponds to one half of trehalose. Although it has been shown that the Ag85 complex can catalyze the formation of glucose monomycolate in addition to TMM,⁵⁸ we assumed that this would not be a problem as long as our dyes were attached to the 6 position, blocking other modifications. For these compounds, we reacted the dyes with methyl 6-iodo-2,3,4-tri-*O*-acetyl-6-deoxy-glucoside (Me 6-I-Ac3-Glc) that we had synthesized by subjecting methyl α -glucoside to triphenylphosphine and molecular iodine followed by acetylation conditions.⁵⁹ Unlike trehalose, glucose has only one 6 position, allowing us to obtain this sugar in higher yield than 6-Br-Ac7-Tre.



60%

Scheme 4.7. Synthesis of methyl 6-iodo-2,3,4-tri-*O*-acetyl-6-deoxy-glucoside (Me 6-I-Ac3-Glc).



Scheme 4.8. Synthesis of the methyl 6-dye-glucoside (dye-Glc) control compounds.

Before testing our new probes in mycobacteria, we characterized their fluorescence intensity and emission spectra in different mixtures of 1,4-dioxane and water (Figure 4.3). We measured the same properties for DMN-Tre in parallel to obtain an accurate comparison. For excitation wavelengths, we used wavelengths that correspond to those commonly used in flow cytometry or fluorescence microscopy (e.g., 405 nm, 488 nm, 561 nm). All probes displayed increased fluorescence as the percentage of 1,4-dioxane increased, i.e., as the hydrophobicity of the surrounding environment increased. DMN-Tre and 3-HC-3-Tre both absorbed most strongly at 405 nm and had comparable fluorescence intensities in each solvent mixture tested, though the spectra for 3-HC-3-Tre were slightly red-shifted. 3-HC-2-Tre responded well to excitation at 405 nm and 488 nm, but its fluorescence intensity was less sensitive to changes in hydrophobicity overall. However, because the emission spectra undergo a bathochromic shift as solvent polarity increases, significant differences in intensity still occur between 500 and 550 nm, the approximate range of wavelengths allowed through the GFP emission filter. Finally, EtNR-Tre and MeNR-Tre both fluoresce more strongly in response to excitation at 561 nm than

at 405 or 488 nm, but their fluorescence intensities in less hydrophobic environments are the same as or greater than that in 99.9% 1,4-dioxane. We determined that this is because the wavelength of maximum absorbance shifts in different solvents – our measurements indicate that in 99.9% 1,4-dioxane, the wavelength of maximum absorbance for EtNR-Tre is 512 nm; EtNR-Tre in 50% 1,4-dioxane has a wavelength of maximum absorption of 567 nm, significantly closer to the 561 nm laser line (Table 4.1). Exciting the dye at 532 nm instead of 561 nm results in the expected decrease in fluorescence as the water content increases (Figure 4.3C). Unfortunately, 532 nm is a less commonly used laser than 561 nm, so for consistency, we used an excitation wavelength of 561 nm in the experiments reported in this chapter.

We also measured the fluorescence properties of two other fluorescent trehalose probes, 6-fluorescein-trehalose (6-FITre) and 6-tetramethylrhodamine-trehalose (6-TMR-Tre). We observed the opposite trend in these probes, with less fluorescence detected as solvent hydrophobicity increased. The excitation and emission spectra of xanthene dyes are known to be dependent on the degree of hydrogen bond formation with the surrounding solvent, with less hydrogen bonding leading to red-shifted excitation and emission maxima as well as decreased fluorescence intensity,⁶⁰ i.e., the opposite effect of that seen in our probes and a less-than-ideal feature for labeling the hydrophobic mycomembrane.

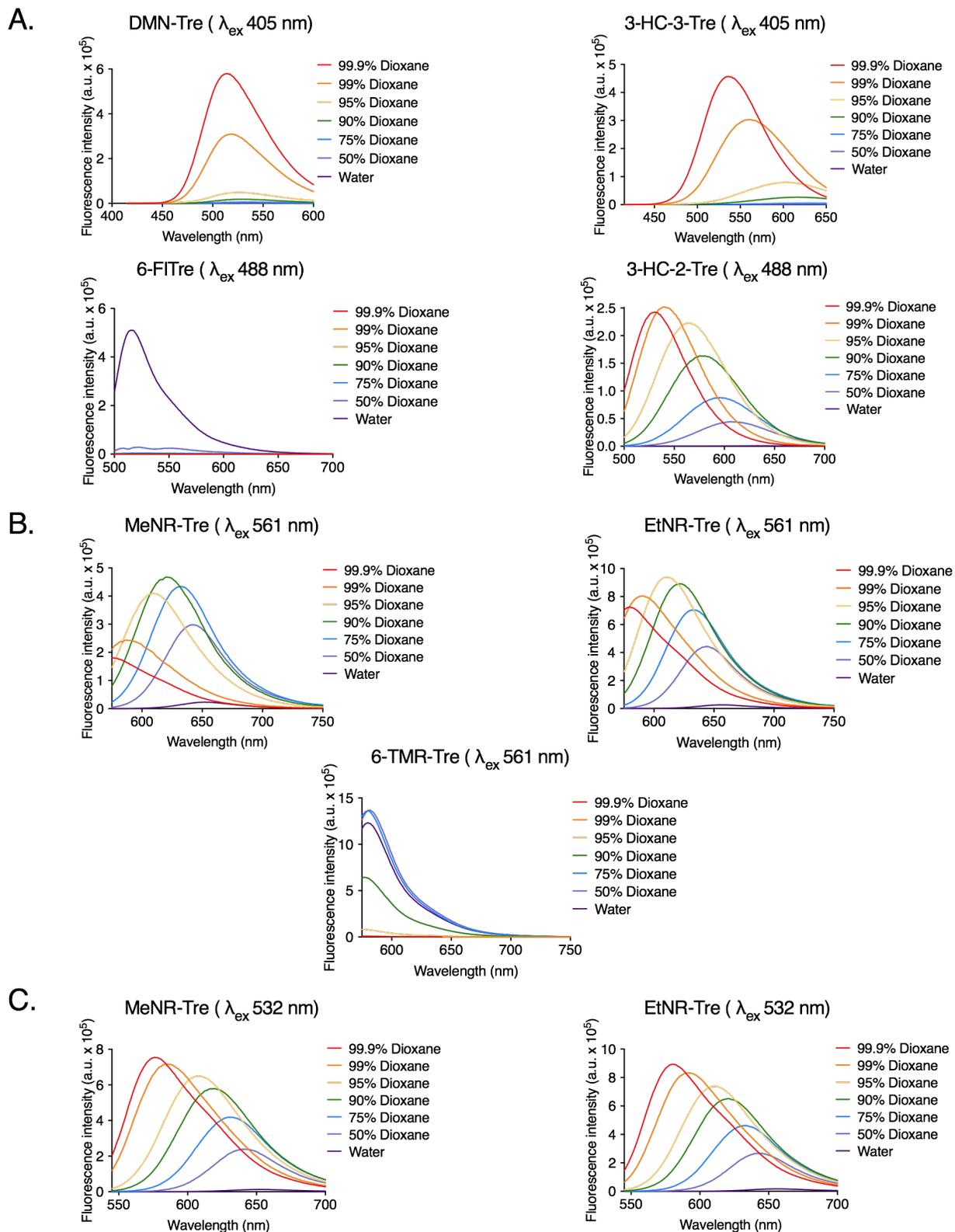


Figure 4.3. Excitation and emission spectra of 3-hydroxychromone and Nile-red trehalose dyes. Fluorescence spectra of (A) DMN-Tre (405 nm), 3-HC-3-Tre (405 nm), 3-HC-2-Tre

(488 nm), 6-FITre (488 nm), (B) MeNR-Tre (561 nm), EtNR-Tre (561 nm) and 6-TMR-Tre (561 nm), and (C) MeNR-Tre (532 nm), EtNR-Tre (532 nm) in solvent systems with various ratios of dioxane/water.

Table 4.1. The EtNR-Tre wavelength of maximum absorbance shifts significantly as the ratio of dioxane to water changes. For each ratio is recorded the wavelength of maximum absorbance, the wavelength of maximum emission, and the maximum emission intensity at the wavelength of maximum absorbance as a percentage of the maximum emission intensity of 99.9% dioxane at 512 nm.

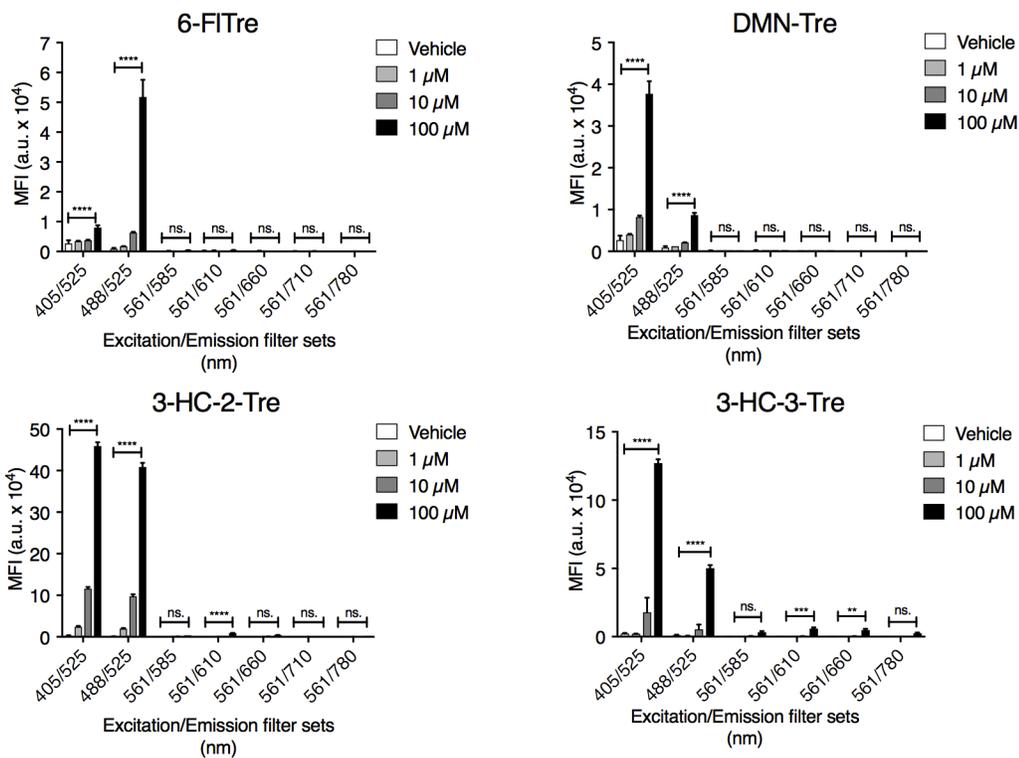
| Percent 1,4-dioxane in water | Wavelength of maximum absorbance (nm) | Wavelength of maximum emission (nm) | Maximum emission intensity (% of emission at 99.9% dioxane) |
|------------------------------|---------------------------------------|-------------------------------------|---|
| 99.9 | 512 | 576 | 100 |
| 90 | 550 | 584 | 63 |
| 75 | 547 | 631 | 19 |
| 50 | 567 | 644 | 5.7 |
| 0 | 585 | 655 | 0.65 |

Having characterized our probes in solution, we moved on to labeling mycobacteria. We first assessed incorporation into *Mycobacterium smegmatis* (Msmeg or Ms), a non-pathogenic and fast-growing member of the *Mycobacterium* genus commonly used as a model for Mtb. The bacteria were incubated with 1, 10, or 100 μM of each probe for 1 hour at 37 $^{\circ}\text{C}$, washed three times, and analyzed by flow cytometry using a variety of excitation and emission filter sets (Figure 4.4). 6-FITre and 6-TMR-Tre were again included for comparison purposes. As expected, DMN-Tre labeled the bacteria most strongly at 100 μM and could be detected using either the 405/525 nm or the 488/525 nm excitation/emission filter sets, though the signal was significantly higher when excited at 405 nm. 3-HC-3-Tre showed the same characteristics but with three times as much signal in both cases; labeling with 10 μM also demonstrated good signal over background. 3-HC-2-Tre worked equally well with the 405/525 and 488/525 nm excitation/emission filter sets and could be used at 1 μM due to its even higher emission intensity. Finally, 6-FITre signal was brightest at 100 μM and using the 488/525 nm excitation/emission filter sets, as expected, though signal was also observed using the 405/525 nm excitation/emission filter sets and when using a labeling concentration of 10 μM with the 488/525 nm excitation/emission filter sets. Its signal intensity was on par with that of DMN-Tre.

For the red dyes, bacteria labeled with 100 μM EtNR-Tre and MeNR-Tre showed significant signal over background at all wavelengths tested, though MeNR-Tre resulted in the lowest fluorescence intensity of all the probes. Concentrations of 1 and 10 μM also resulted in significant labeling at some wavelengths. While in both cases the strongest signal occurred when the 561/610 or 561/660 nm excitation/emission filter sets were

applied, these wavelengths were predicted to be less sensitive to the hydrophobicity of the environment based on the results obtained using the fluorometer in the previous experiment. Consequently, the 561/585 nm excitation/emission filter set was chosen for future experiments. While the intensity was much lower, bacteria incubated with concentrations of 1 and 10 μM EtNR-Tre and MeNR-Tre did still result in more signal than unlabeled bacteria. As expected, incubating with 1 or 10 μM 6-TMR-Tre resulted in brightly labeled bacteria that could be detected with the 561/585, 561/610, or 561/660 nm excitation/emission filter sets. Altogether, the differences in preferred excitation/emission filter sets demonstrate that certain combinations of our solvatochromic probes can potentially be used for such experiments as pulse-chase analysis of changes in cell wall dynamics over time or after drug treatment (Table 4.2).

A.



B.

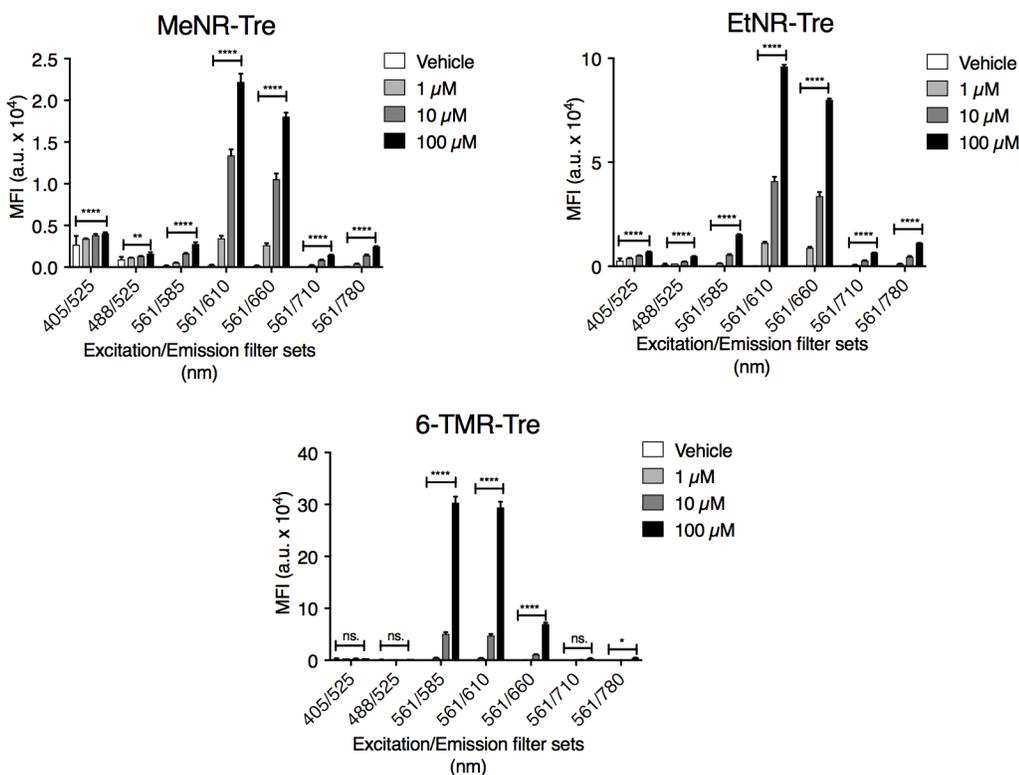
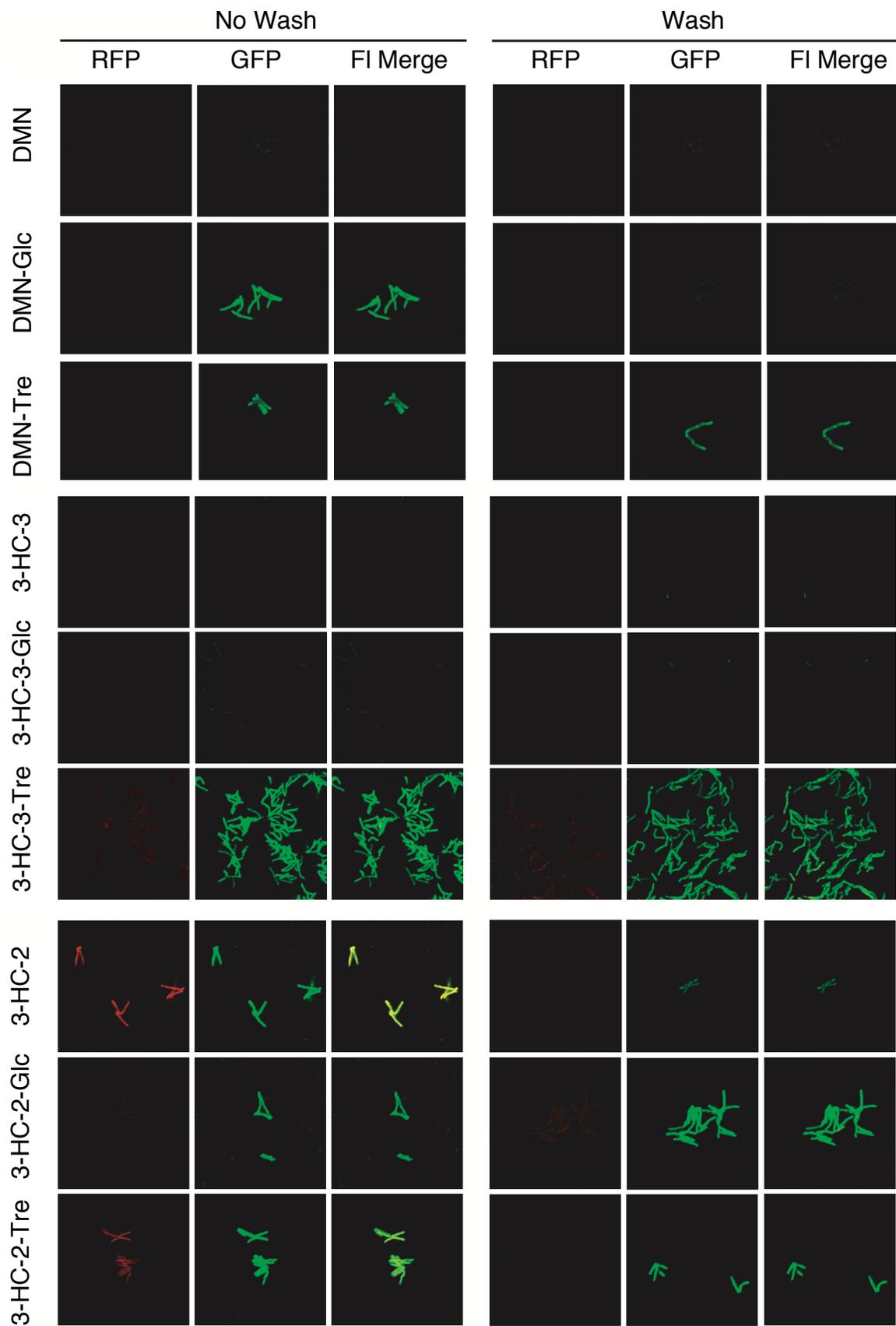


Figure 4.4. Fluorescence analysis of Msmeg cells labeled with 3-hydroxychromone or Nile-red trehalose dyes using various excitation and emission filter sets. Flow cytometry analysis of Msmeg labeled with (A) DMN-Tre, 3-HC-3-Tre, 3-HC-2-Tre, 6-FITre and (B) MeNR-Tre, EtNR-Tre and 6-TMR-Tre. Cells at OD₆₀₀ = 0.5 were incubated with the indicated dye-trehalose probe concentrations for 1 h at 37 °C. MFI = Mean fluorescence intensity. Data are means ± SEM from at least two independent experiments. Data were analyzed by two-way analysis of variance (ANOVA) test (*P < 0.05, **P < 0.01, ***P < 0.001 and ****P < 0.0001; ns, not significant).

Table 4.2. For each probe, the ideal excitation/emission filter set (commercially available, not necessarily common) and the acceptable excitation/emission filter set (common) is outlined.

| Probe | Ideal excitation/emission filter set(s) (nm) | Acceptable excitation/emission filter set(s) (nm) |
|------------|--|---|
| DMN-Tre | 405/525 | 405/525, 488/525 |
| 3-HC-3-Tre | 405/525 | 405/525, 488/525 |
| 3-HC-2-Tre | 440/525 | 488/525 |
| MeNR-Tre | 515/575, 532/575 | 561/585 |
| EtNR-Tre | 515/575, 532/575 | 561/585 |

We next used microscopy to assess the necessity of a wash step and the degree of background signal when Msmeg were labeled with each probe, the glucose-dye control, or free dye (Figure 4.5). Somewhat unexpectedly, all of the glucose-dye controls except 3-HC-3-Glc showed background turn-on localized to the bacteria before a wash step. Surprisingly, 3-HC-2-Glc and MeNR-Glc signal did not disappear on washing and appeared to be just as strong as that generated by labeling with 3-HC-2-Tre and MeNR-Tre. For the remaining glucose-dye controls, we attribute the background to nonspecific hydrophobic interactions between the dye and the cell wall, as signal was localized to the bacteria themselves. Msmeg incubated with 3-HC-2 or MeNR alone also required a wash step to remove background, which we again attribute to nonspecific hydrophobic interactions between the dye and the cell wall. Gratifyingly, 3-HC-2-Tre, 3-HC-3-Tre, MeNR-Tre, and EtNR-Tre all labeled more strongly than DMN-Tre at the same or lower concentrations. Due to the high fluorescence intensity in bacteria labeled with 3-HC-3-Tre and the absence of background signal in bacteria incubated with 3-HC-3-Glc or 3-HC-3 alone, 3-HC-3-Tre was determined to be the most promising of the solvatochromic trehalose probes to date.



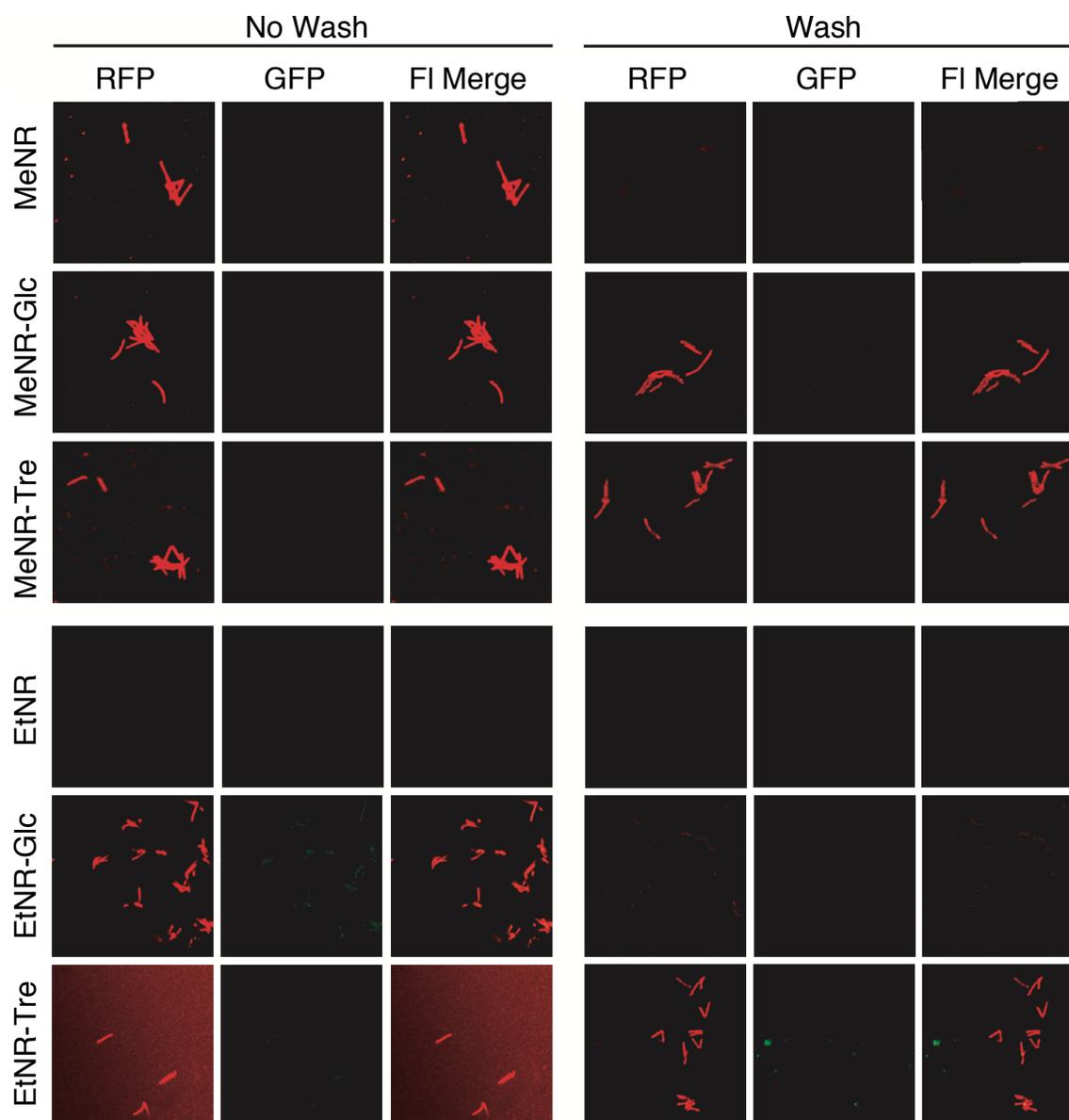


Figure 4.5. Wash steps may be necessary to remove excess unincorporated dyes. Microscopy analysis of Msmeg cells treated with (A) 100 μ M of DMN-Glc or DMN-Tre; (B) 100 μ M of 3-HC-3, 3-HC-3-Glc or 3-HC-3-Tre; (C) 100 μ M of 3-HC-2, 3-HC-2-Glc or 3-HC-2-Tre; (D) 100 μ M MeNR, MeNR-Glc or MeNR-Tre; (E) 10 μ M EtNR, EtNR-Glc or EtNR-Tre. Cells were incubated with the indicated dyes for 1 h at 37 °C. Cells were smeared directly (No wash) or washed 3x with PBS (Wash) then smeared onto a microscope slide. Images were acquired with a Nikon confocal microscope.

To quantitate the extent of Msmeg labeling by the glucose conjugates, we performed a flow cytometry analysis (Figure 4.6). For 3-HC-3-Glc, a small amount of statistically significant labeling occurred, but it was far below the amount of labeling by 3-HC-3-Tre (Figure 4.6A). For 3-HC-2-Glc (B), MeNR-Glc (C), and, contrary to the microscopy data, EtNR-Glc (D), highly significant labeling was observed. In fact, not only was MeNR-Glc labeling highly significant, the fluorescence intensity was approximately 1.5 times that of MeNR-Tre. MeNR alone also showed highly significant labeling compared to untreated bacteria, but its intensity was much lower.

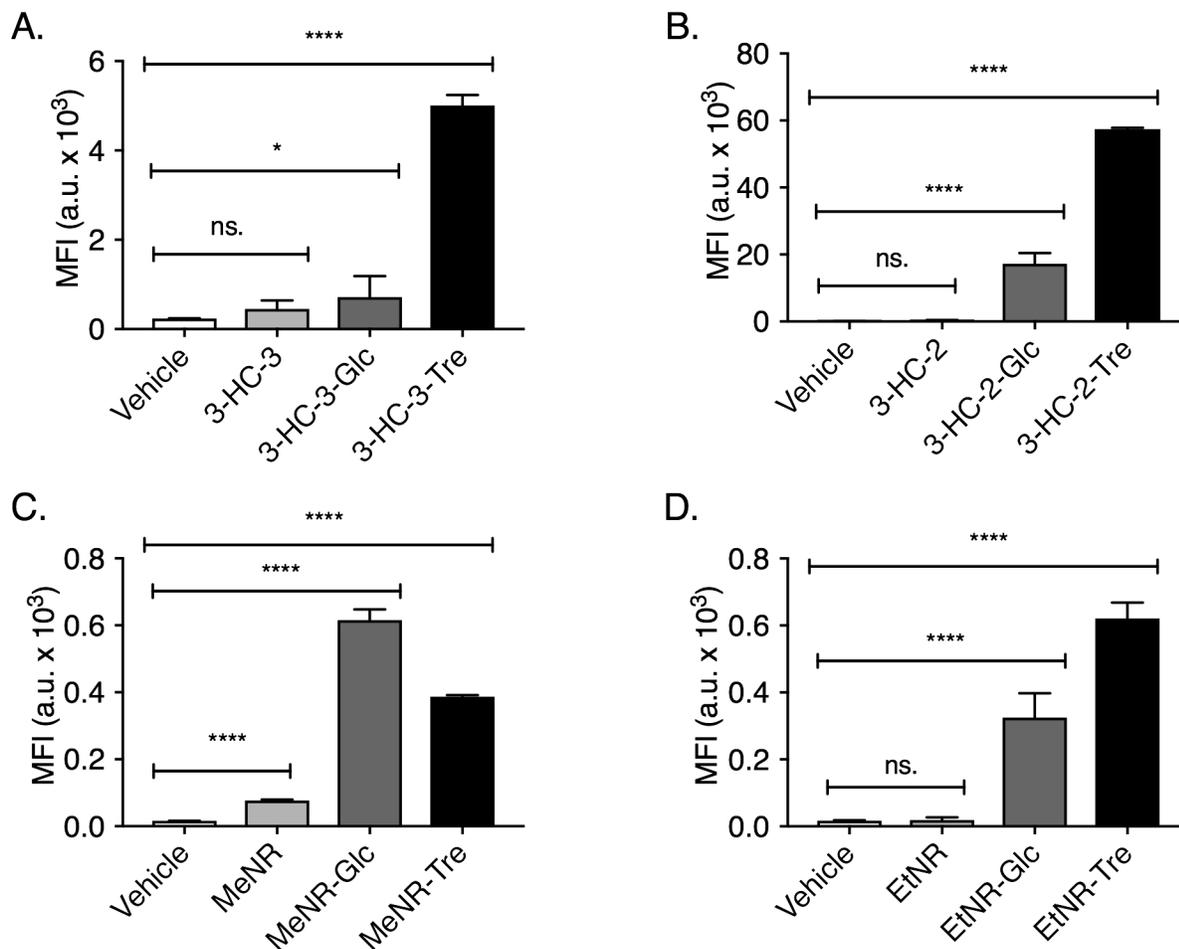


Figure 4.6. Solvatochromic trehalose dyes label Msmeg more strongly than solvatochromic glucose dyes, with the exception of MeNR-Tre. Flow cytometry analysis of Msmeg cells labeled with (A) 100 μ M 3-HC-3, 3-HC-3-Glc, or 3-HC-3-Tre, (B) 10 μ M 3-HC-2, 3-HC-2-Glc, or 3-HC-2-Tre, (C) 10 μ M MeNR, MeNR-Glc, or MeNR-Tre or (D) 10 μ M EtNR, EtNR-Glc, or EtNR-Tre dyes. Samples were incubated at 37 °C for 1 h then analyzed. Data are means \pm SEM from at least two independent experiments. Data were analyzed by ordinary one-way analysis of variance (ANOVA) test (*P < 0.05, **P < 0.01, ***P < 0.001 and ****P < 0.0001; ns, not significant).

Concerned by these results, we decided to confirm that our dye-Tre probes are metabolically incorporated via the same cellular machinery as trehalose. First, we did a competition experiment in which *Msmeg* was incubated with a dye-Tre probe in combination with equal or greater amounts of free trehalose (Figure 4.7). As the concentration of free trehalose increases, the probe should be increasingly outcompeted, and signal should decrease. This phenomenon was observed for all the dye-Tre probes, though some were more sensitive to trehalose competition than others. Additionally, we tested whether MeNR-Glc labeling could be competed away with the addition of trehalose (Figure 4.7E). We did observe a significant decrease in labeling at higher ratios of trehalose to probe, but the response was less pronounced than for MeNR-Tre. Previous studies^{31,34} have shown that trehalose competition sensitivity can be affected by changing the location of the modification on trehalose, raising the possibility that the MeNR-Glc probe is incorporated through the same pathway as MeNR-Tre but is simply more resistant to competition from free trehalose. Thus, this experiment was deemed inconclusive.

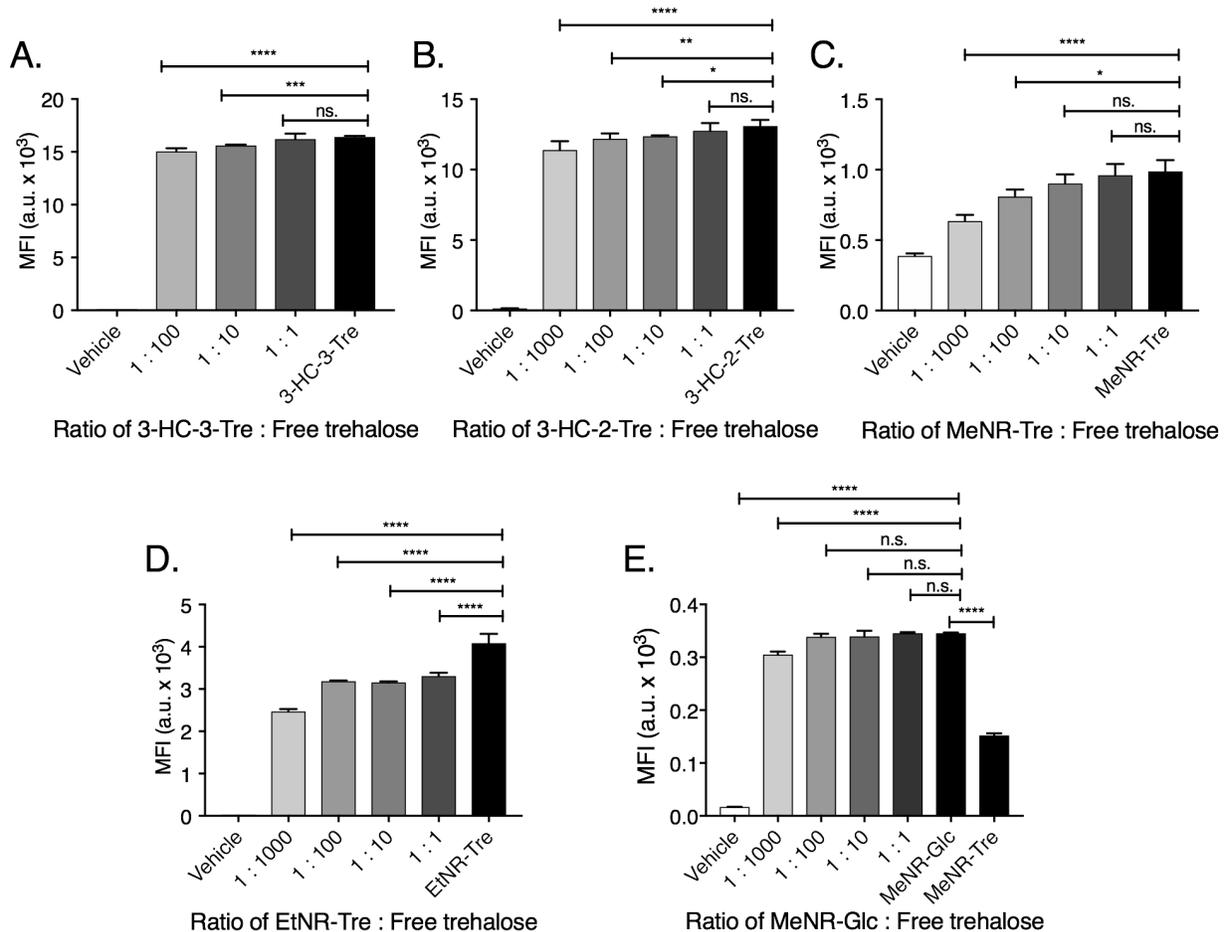


Figure 4.7. Solvatochromic dyes use the same pathway as free trehalose. Flow cytometry MFI analysis of *Msmeg* cells competition of DMN-Tre labeling with free trehalose. Cells were incubated with 100 μ M 3-HC-3-Tre (A) or 10 μ M of 3-HC-2-Tre (B), MeNR-Tre (C), EtNR-Tre (D), or MeNR-Glc (E) probes alone or in combination with equal

or greater concentrations of free trehalose. Samples were incubated at 37 °C for 1 h then analyzed. Data are means \pm SEM from at least two independent experiments. Data were analyzed by one-way analysis of variance (ANOVA) test (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ and **** $P < 0.0001$; ns, not significant).

We also tested 3-HC-2-Tre and EtNR-Tre labeling in Msmeg that had been pretreated with ebselen, an Ag85 inhibitor (Figure 4.8A, B).⁶¹ Inhibition of Ag85 should decrease mycolylation of the dye-Tre probes, in turn decreasing incorporation into the mycomembrane. As the hydrophobic environment of the mycomembrane is required in order for the probes to fluoresce, flow cytometry should show a decrease in signal intensity. As expected, we observed a dose-dependent decrease in 3-HC-2-Tre and EtNR-Tre labeling of pretreated Msmeg cells, though only the EtNR-Tre results were significant due to the large error bars in the 3-HC-2-Tre results. Regardless, these results suggest that the probes are metabolized through the Ag85 pathway, in accordance with previous studies on unnatural trehalose probes. To further confirm that labeling depends on metabolic activity, we treated heat-killed Msmeg with 3-HC-2-Tre, MeNR-Tre, or EtNR-Tre (Figure 4.8C, D). While the 3-HC-2-Tre signal decreased in the heat-killed samples, it did not entirely disappear, and the MeNR-Tre and EtNR-Tre signals actually increased. This could be because the fluorescence intensities of 3-HC-2-Tre, MeNR-Tre, and EtNR-Tre are less sensitive to the presence of water than to that of DMN-Tre, and heat-killing the bacteria may compromise the cell wall enough to allow the probes to nonspecifically interact with lipids without the necessity for metabolic incorporation.

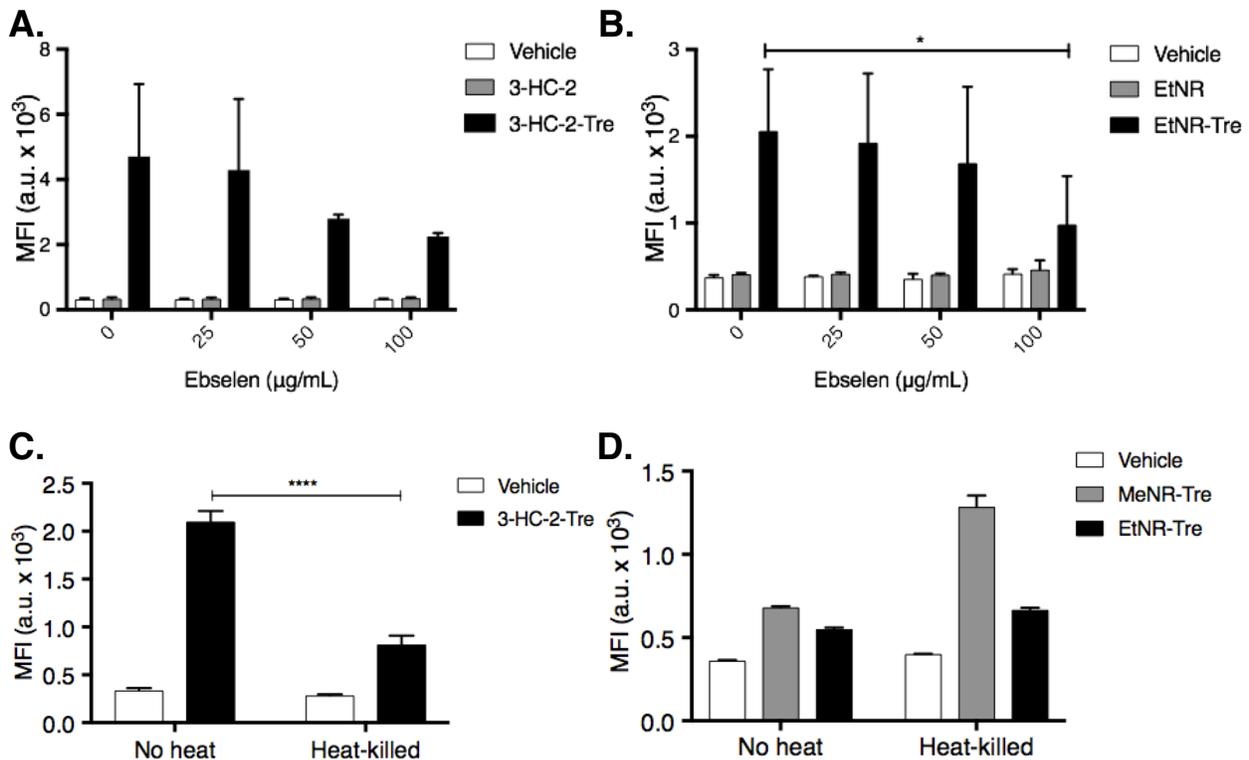


Figure 4.8. Dye-Tre incorporation is dependent on metabolic activity. Top, Msmeg were pretreated with the Ag85 inhibitor ebselen for 3 h then labeled with 10 μ M 3-HC-2-Tre (A) or EtNR-Tre (B) for 1 h at 37 °C prior to flow cytometry analysis. 3-HC-2-Tre results were not statistically significant due to the large error bars at 0 and 25 μ g/mL ebselen. Bottom, Msmeg were heat-killed by incubating at 95 °C for 30 min. then treated with 10 μ M 3-HC-2-Tre (C), MeNR-Tre (D), or EtNR-Tre (D) for 1 h before flow cytometry analysis. Data are means \pm SEM from at least two independent experiments. Data were analyzed by one-way analysis of variance (ANOVA) test (*P < 0.05, **P < 0.01, ***P < 0.001 and ****P < 0.0001; ns, not significant).

To indirectly assess the mechanism of incorporation, we also did a dual-labeling experiment in which Msmeg cells were simultaneously treated with different concentrations of EtNR-Tre and DMN-Tre, which has been shown to undergo Ag85-mediated incorporation (Figure 4.9). Though the fluorescence intensities of EtNR-Tre and DMN-Tre are quite different from each other, the microscopy data show that labeling with 100 μ M EtNR-Tre sharply decreases as the concentration of DMN-Tre is increased from 10 to 100 μ M, suggesting that they are in competition. This further supports the hypothesis that incorporation of EtNR-Tre is processed at least in part by Ag85.

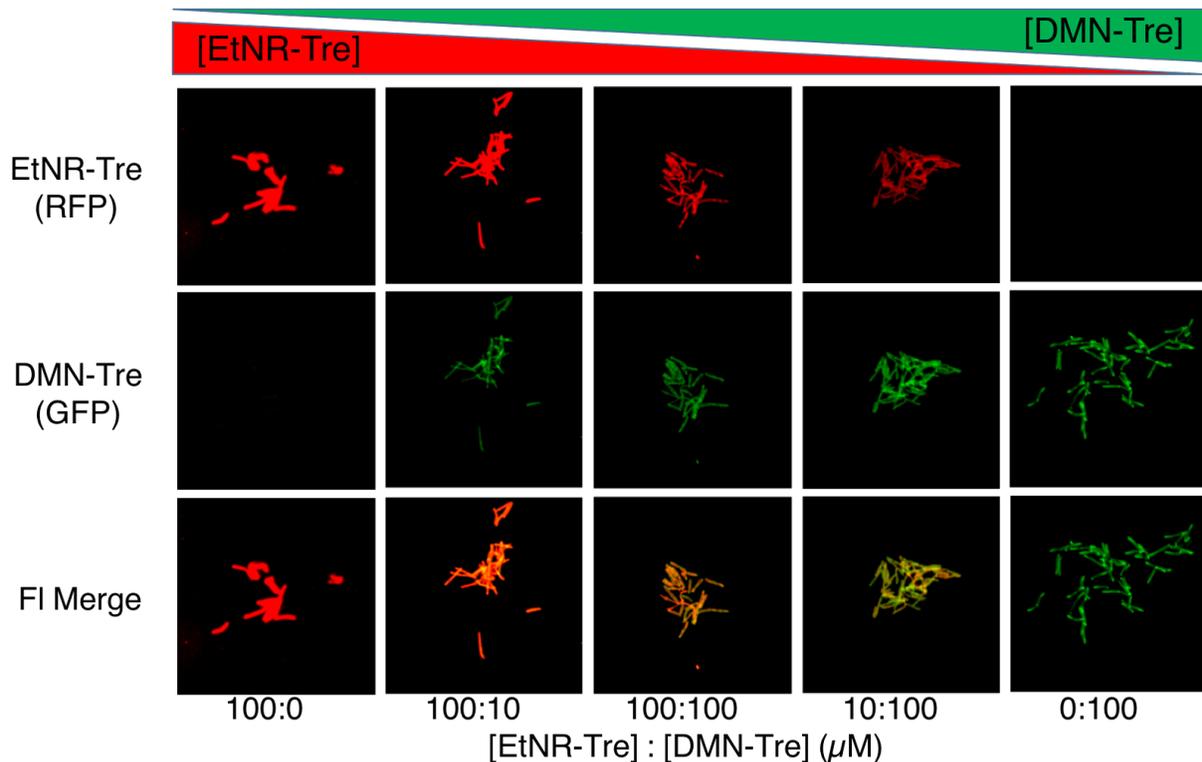


Figure 4.9. DMN-Tre and EtNR-Tre compete for incorporation into Msmeg. Msmeg were treated with different ratios of EtNR-Tre:DMN-Tre for 1 h at 37 °C, washed, and imaged with a Nikon confocal microscope.

Finally, we assessed 3-HC-2-Tre and MeNR-Tre labeling in different Msmeg trehalose transporter mutants in comparison to DMN-Tre labeling to confirm that the probes are not primarily processed via a cytosolic pathway (Figure 4.10). These mutant strains lack SugC, LpqY, TreS (trehalose synthase), or trehalase (MSMEG_4535). SugC and LpqY are components of the trehalose-recycling transmembrane transporter protein complex SugABC-LpqY, while TreS and trehalase are involved in cytoplasmic metabolism of trehalose. For all three probes, labeling decreased to a significant extent in each mutant, but that decrease accounted for less than 50% of probe incorporation in wildtype. The small effect size of the trehalose transporter mutants suggests access to cytosol is not necessary for labeling and that the probes are processed primarily extracellularly.

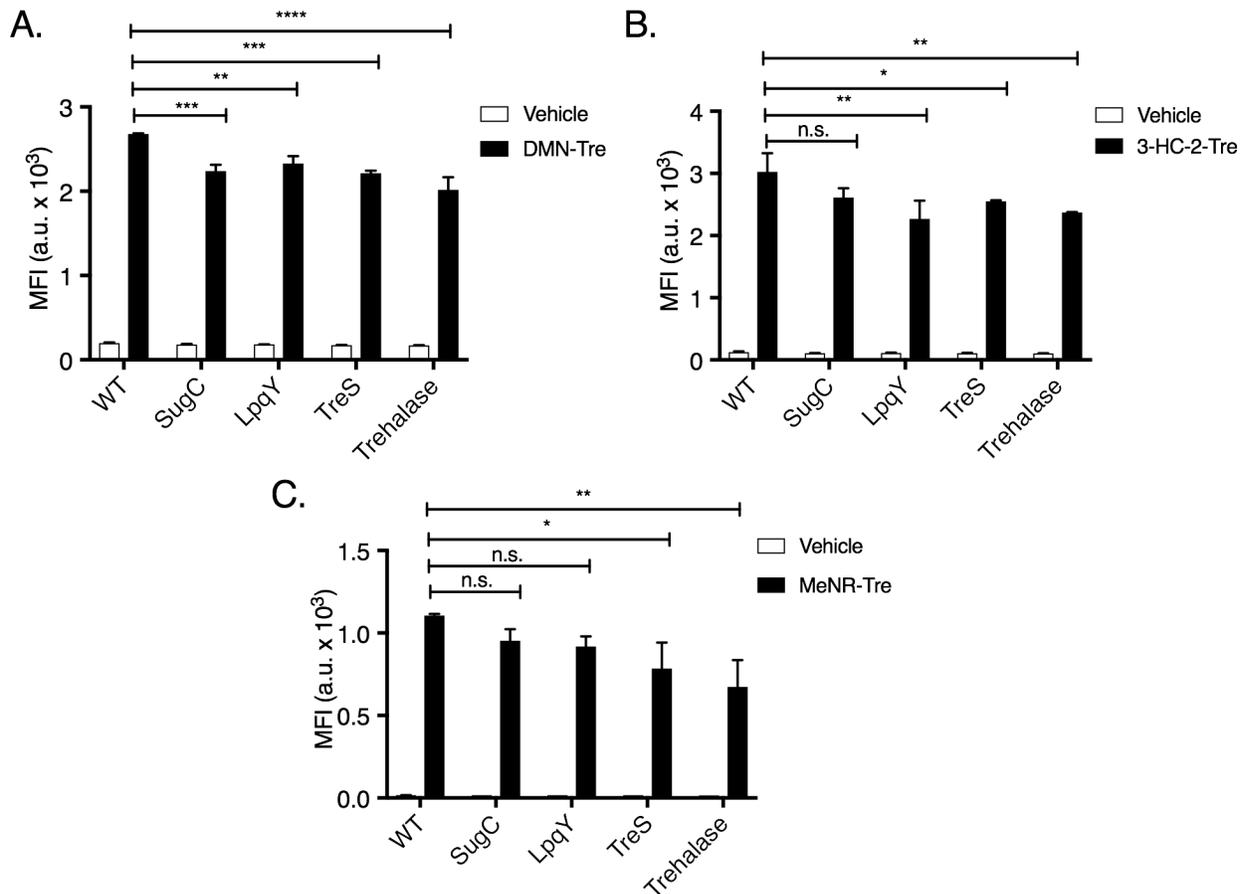


Figure 4.10. Dye-Tre incorporation does not primarily occur through cytosolic trehalose metabolic pathways. Msmeg or Msmeg mutant strains lacking SugC, LpqY, TreS, or trehalase were treated with 100 μ M DMN-Tre (**A**) or 10 μ M 3-HC-2-Tre (**B**) or MeNR-Tre (**C**) for 1 h at 37 $^{\circ}$ C then analyzed by flow cytometry. Data are means \pm SEM from at least two experiments. Data were analyzed by one-way analysis of variance (ANOVA) test (* P < 0.05, ** P < 0.01, *** P < 0.001 and **** P < 0.0001; ns, not significant).

Satisfied that dye-Tre labeling of Msmeg appears to be dependent on the essential features of mycobacterial trehalose metabolism but still unsure whether 3-HC-2-Glc and MeNR-Glc labeling is as well, we next examined the species selectivity of dye-Tre probe

labeling in other mycobacteria, corynebacteria, gram-positive, and gram-negative bacteria (Figure 4.11, 4.12). All mycobacteria and many species from the Corynebacterineae suborder synthesize trehalose glycolipids that incorporate into hydrophobic environments. Most gram-positive and gram-negative bacteria do not produce trehalose glycolipids. We hypothesized that dye-Tre would label *Corynebacterium spp.* but not other gram-negative and gram-positive bacteria. We selected *Corynebacterium glutamicum* (Cg), *Bacillus subtilis* (Bs), *Staphylococcus aureus* (Sa), and *Escherichia coli* (Ec) for further examination. Cg is in the Corynebacterineae suborder and is known to have a mycomembrane including trehalose glycolipids and orthologs of Ag85C³⁸ and MmpL3;⁶² Bs and Sa are gram-positive bacteria; Ec is gram negative.

By flow cytometry, the dye-Tre probes strongly labeled only the bacteria with a mycomembrane, as expected (Figure 4.11A-D). To be certain that these results were not due to differences in bacterial replication rates, we also confirmed that OD₆₀₀ values indicated that Cg, Bs, Ec, and Sa grew by approximately the same amount over a 90-minute period (Figure 4.11E). Less expected was the observation that probes did not label Msmeg and Cg equally well. 3-HC-3-Tre labeled Msmeg approximately 1.5 times as well as it did Cg; 3-HC-2-Tre labeled Msmeg and Cg equally well; EtNR-Tre and MeNR-Tre labeled Cg much more strongly than they did Msmeg. Some labeling was also observed in other bacteria; Bs labeled with 3-HC-3-Tre (100 μ M) to some extent, but much less so than Msmeg and Cg. We are uncertain about the reasons for this observation but are confident that further optimization of 3-HC-3-Tre concentration will more completely remove background signal in gram-positive and gram-negative bacteria. By flow cytometry, we also observed significant labeling in Sa treated with EtNR-Tre, though this was not reflected in our microscopy images (Figure 4.12), prompting us to continue optimization of labeling and analysis in the future.

Additionally, while several of the dye-Glc probes did strongly label at least one of the mycomembrane-containing bacteria (3-HC-3-Glc in Cg, 3-HC-2-Glc in Msmeg, MeNR-Glc in both Msmeg and Cg), no significant labeling was observed in gram-positive or gram-negative bacteria when incubating with 3-HC-3-Glc, 3-HC-2-Glc, or MeNR-Glc. EtNR-Glc gave surprising results, however, with statistically significant labeling across all bacteria. As the fluorescence intensity was unusually low for the EtNR-Tre labeling in Msmeg and Cg, we believe that here again further optimization of labeling and analysis is necessary. Thus, although we are still elucidating the mechanism by which the glucose analogs are incorporated into mycomembrane-containing bacteria, we have shown that the fluorescence signal is mostly dependent on biological machinery that is not present in gram-positive and -negative bacteria, easing concerns that the incorporation of our dye-Tre probes is due to nonspecific interaction of the dye with hydrophobic membranes rather than, predominantly, the trehalose glycolipid with its unique environment.

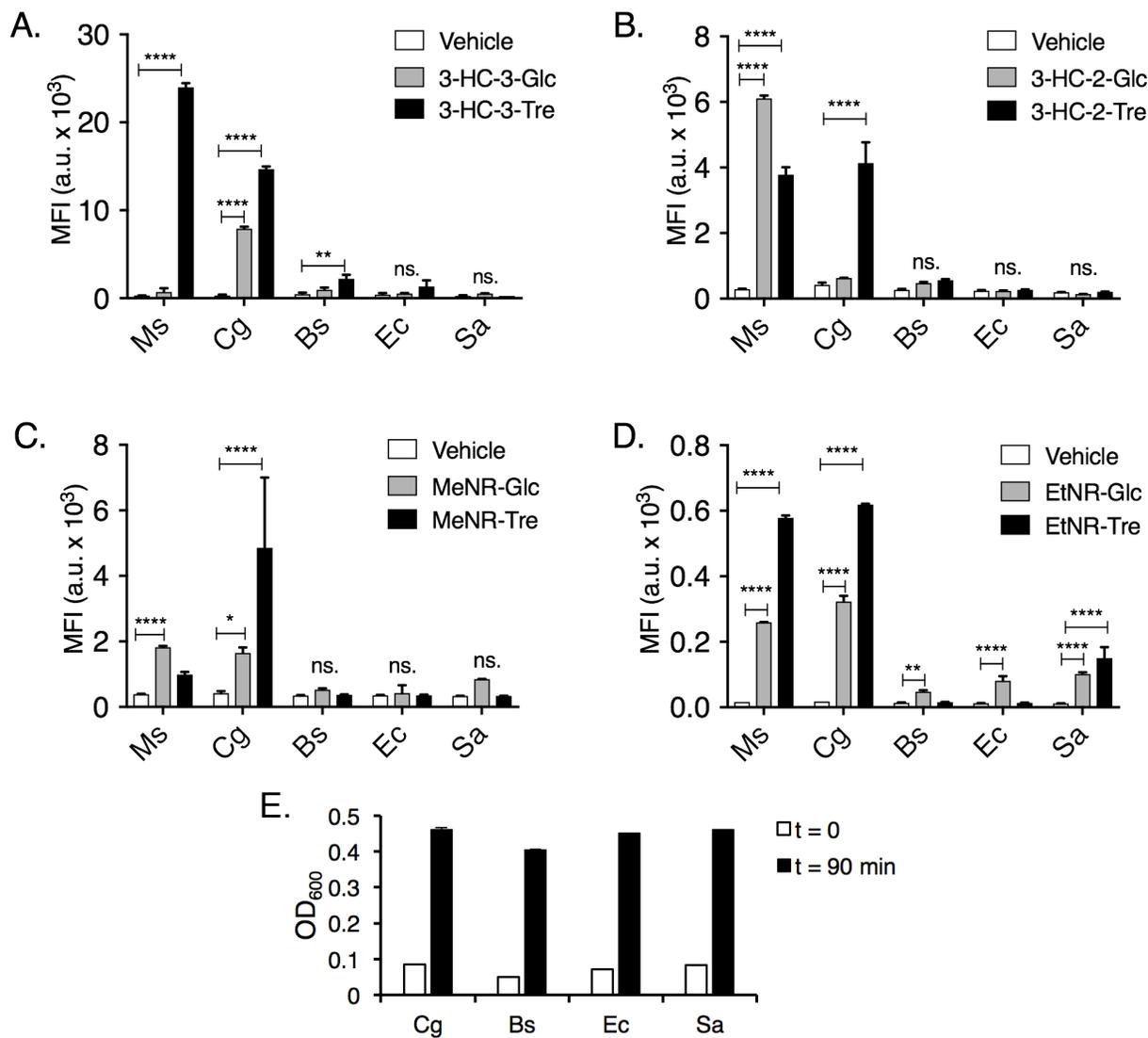


Figure 4.11. Specific labeling of organisms in the Corynebacterineae suborder with solvatochromic trehalose conjugates. Flow cytometry analysis of Msmeg (Ms), *C. glutamicum* (Cg), *B. subtilis* (Bs), *E. coli* (Ec), *S. aureus* (Sa) cells incubated for 1 h at 37 $^{\circ}$ C with (A) 100 μ M of 3-HC-3-Glc or 3-HC-3-Tre; (B) 10 μ M of 3-HC-2-Glc or 3-HC-2-Tre; (C) 10 μ M of MeNR-Glc or MeNR-Tre; (D) 10 μ M EtNR-Glc or EtNR-Tre. (E) Differences in labeling intensity are not an artifact of differing rates of bacterial growth causing variation in fluorescence intensity/cell. OD₆₀₀ at the time points indicated were similar among the studied bacteria. Data are means \pm SEM from at least two independent experiments. Data were analyzed by two-way analysis of variance (ANOVA) test (*P < 0.05, **P < 0.01, ***P < 0.001 and ****P < 0.0001; ns, not significant).

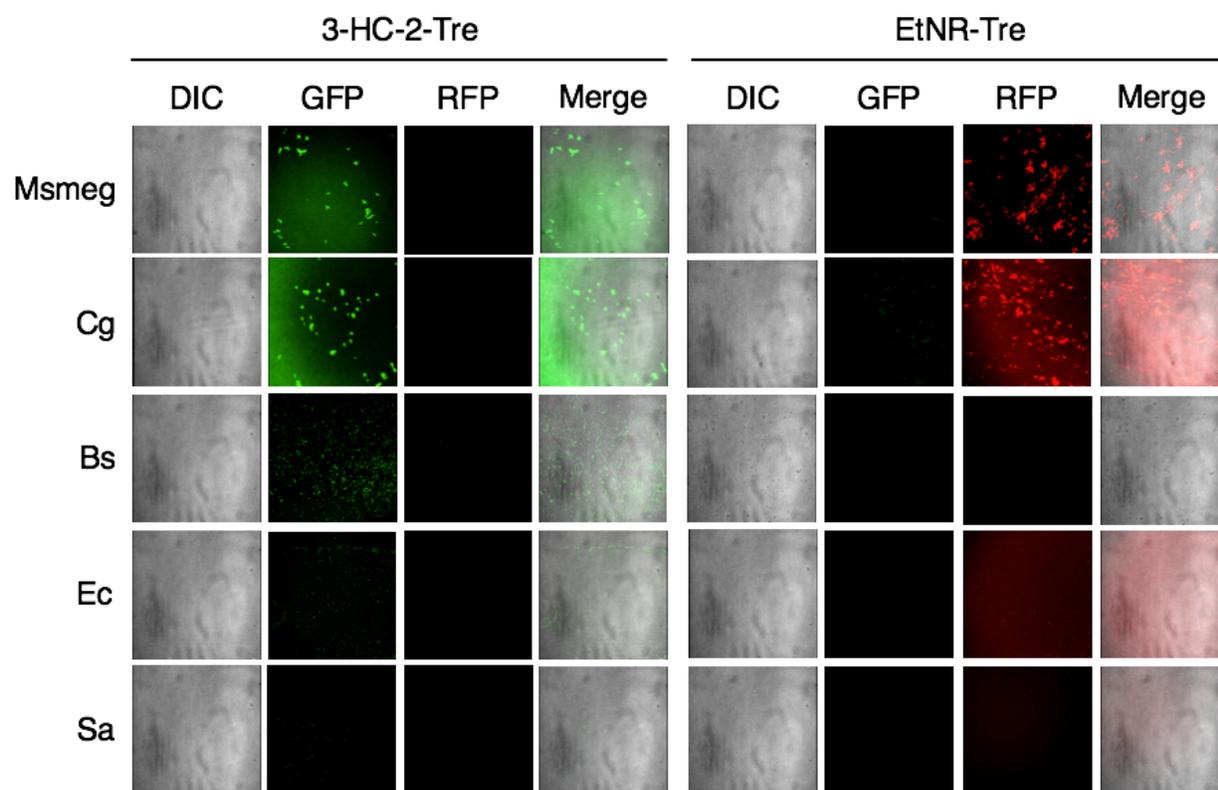
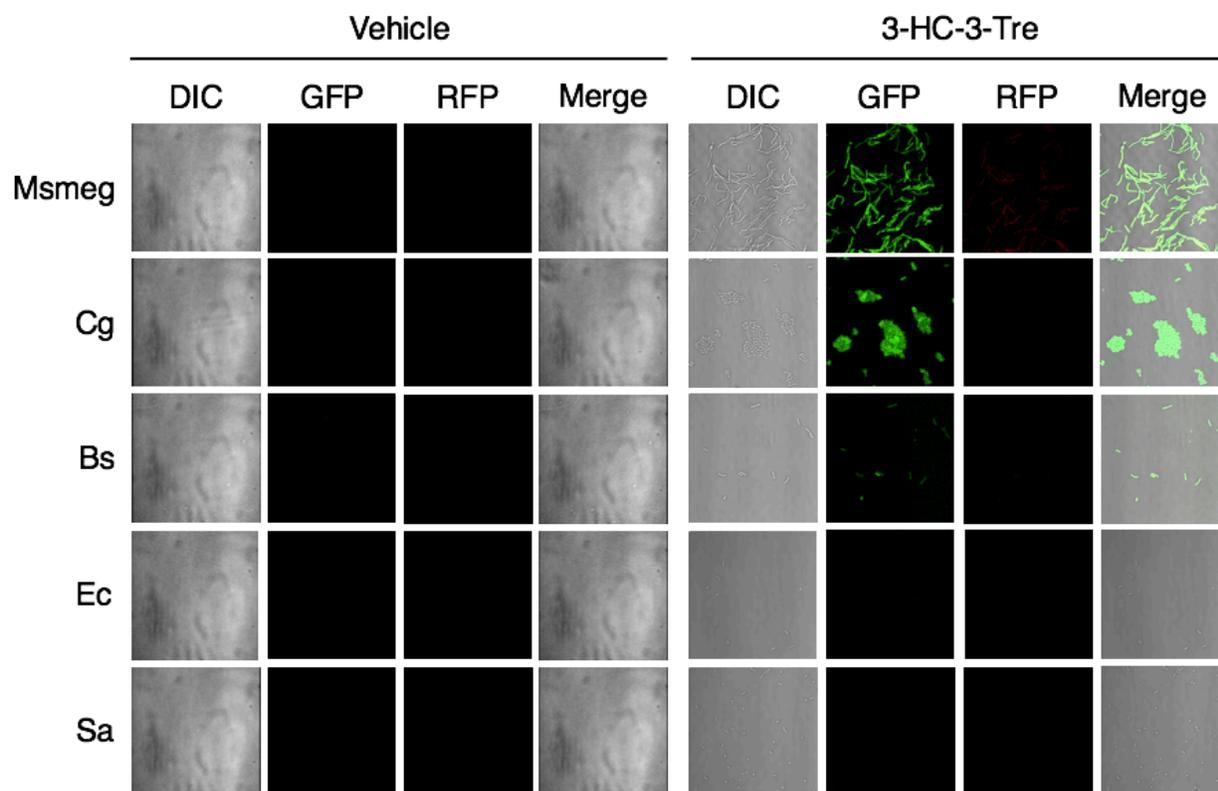


Figure 4.12. Selective imaging of organisms in the Corynebacterineae suborder with solvatochromic trehalose conjugates. Confocal microscopy analysis of Msmeg (Ms), *C. glutamicum* (Cg), *B. subtilis* (Bs), *E. coli* (Ec), *S. aureus* (Sa) cells incubated for 1 h at 37 °C with media alone (top left); 100 μ M of 3-HC-3-Tre (top right); 10 μ M of 3-HC-2-Tre (bottom left); 10 μ M EtNR-Tre (bottom right).

Having completed our initial characterization studies, we wished to determine whether our probes could report on drug susceptibility. We chose to use isoniazid (INH) as our initial proof-of-concept (Figure 4.13). To briefly recapitulate the description in the introduction, INH is a prodrug that becomes activated by the mycobacterial catalase-peroxidase enzyme KatG, then goes on to inhibit synthesis of mycolic acids. Thus, in Msmeg pretreated with INH, labeling by our probes should decrease because their incorporation into the mycomembrane is inhibited. In a $\Delta katG$ mutant, INH cannot be converted to its active form, and thus pretreatment with INH should have no effect on probe incorporation and fluorescence. Our results with 3-HC-2-Tre and MeNR-Tre matched these expectations; the large error bars in the EtNR-Tre results (Figure 4.13C) caused the difference in labeling between INH-treated and untreated Msmeg to be insignificant, complicating analysis. Regardless, this experiment demonstrates that 3-HC-2-Tre and MeNR-Tre may be used to report on bacterial susceptibility to drugs that impact the integrity of the mycomembrane.

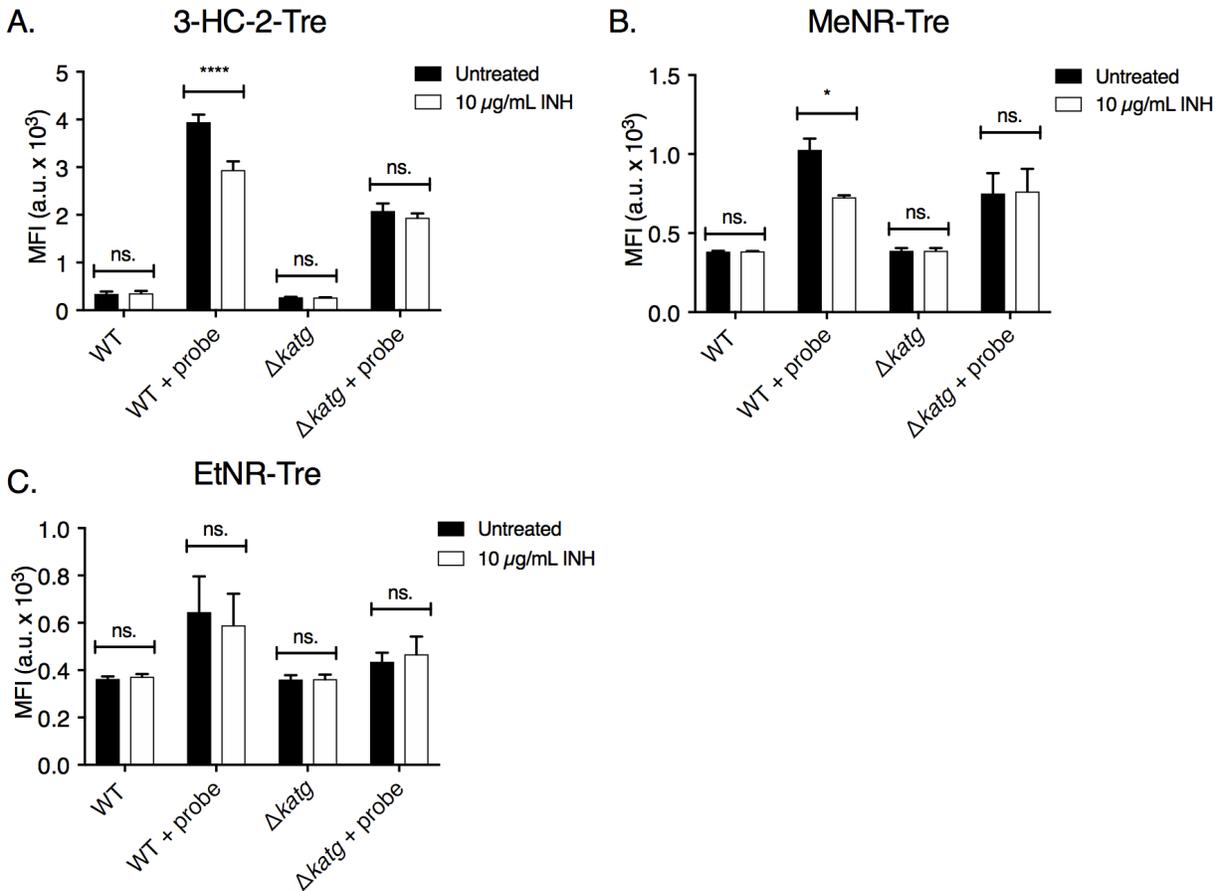


Figure 4.13. Solvatochromic trehalose probes can report on drug susceptibility. Flow cytometry analysis of wild-type (WT) *Msmeg* or *KatG* mutant cells treated with 10 µg/mL isoniazid (INH) for 3 hours, followed by incubation with **A**) 10 µM 3-HC-2-Tre; **B**) 10 µM MeNR-Tre; **C**) 10 µM EtNR-Tre for 1 h. Data are means ± SEM from at least two independent experiments. Data were analyzed by two-way analysis of variance (ANOVA) test (*P < 0.05, **P < 0.01, ***P < 0.001 and ****P < 0.0001; ns, not significant).

Finally, we tested our probes in *Mycobacterium marinum*, a close genetic relative of *Mtb* that naturally infects ectotherms. One of its hosts is zebrafish (*Danio rerio*), a vertebrate model organism popular due to its prolific reproduction, rapid development (only three days separate the zygote stage and the beginning of the larval stage), and genetic tractability.⁶³ Moreover, embryonic zebrafish are transparent, facilitating *in vivo* fluorescence imaging. *M. marinum* infection of zebrafish results in systemic granulomatous disease, analogous to the lung granulomas that are a hallmark of human *Mtb* infection.⁶⁴ Because zebrafish adaptive immune systems do not activate until later developmental stages, embryonic zebrafish can help elucidate mechanisms by which the innate immune system initially responds to *M. marinum* infection.⁶⁴⁻⁶⁵

We first used microscopy to confirm that 3-HC-2-Tre, MeNR-Tre, and EtNR-Tre are incorporated into *M. marinum* (Figure 4.14, top panel). At 10 μM , all three dyes strongly label *M. marinum* after a 3-hour incubation period, with the lowest fluorescence intensity in bacteria labeled by MeNR-Tre. We then infected anaesthetized embryonic zebrafish (2 days post-fertilization) with these bacteria, bacteria pre-labeled with 100 μM DMN-Tre, and unlabeled bacteria via hindbrain injection and immediately imaged (Figure 4.14, bottom panel). Gratifyingly, bacteria incubated with any of the three newer probes were clearly visible, and while some autofluorescence background was observed as a green haze in zebrafish injected with DMN-Tre-treated bacteria, no such background was observed in zebrafish injected with 3-HC-2-Tre-treated bacteria due to the higher fluorescence intensity of 3-HC-2-Tre.

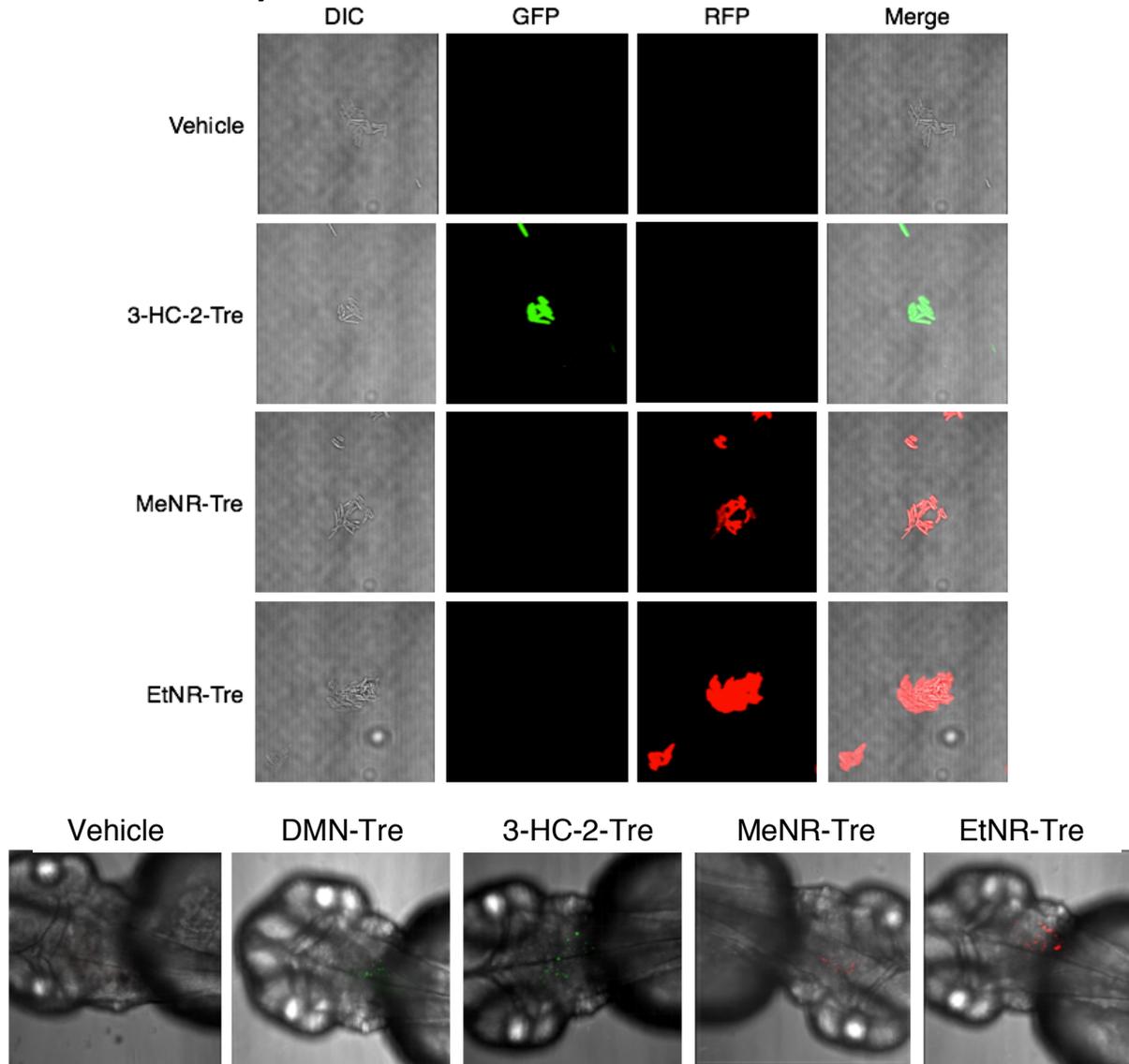


Figure 4.14. 3-HC-2-Tre, MeNR-Tre, and EtNR-Tre label *M. marinum* more brightly than DMN-Tre at 10x lower concentrations and can be used to visualize pre-labeled *M. marinum* in zebrafish. Top, *M. marinum* were incubated with 3-HC-2-Tre (10 μM),

MeNR-Tre (10 μ M), or EtNR-Tre (10 μ M) for 3 h at 33 °C then imaged. Bottom, embryonic zebrafish were injected with *M. marinum* pre-labeled with media only, DMN-Tre (100 μ M), 3-HC-2-Tre (10 μ M), MeNR-Tre (10 μ M), or EtNR-Tre (10 μ M) and immediately imaged in Z-stacks using confocal microscopy. Zebrafish imaging was performed by Shivam Verma (Bertozzi Lab).

CONCLUSIONS

In this chapter, we have introduced several new solvatochromic trehalose probes and have described initial characterization experiments. Some results were unexpected, such as the observation of significant labeling by certain glucose analogs, suggesting a need for more thorough follow-up study. Moreover, as a major problem with the Nile red-based dyes appears to be that they are less water-sensitive at the 561 nm excitation wavelength commonly available on microscopes and flow cytometers. We plan to synthesize compounds for a follow-up structure-activity relationship assay to optimize this feature. It should be possible to find a set of substituents that will shift the wavelength of maximum excitation in hydrophobic environments closer to 561 nm, which would in turn result in more reliable labeling data with the Nile red-based probes. We are also exploring other arylamines to use with our 3-HC scaffold that could potentially result in more red-shifted probes, as the 3-HC dyes have proven to label more strongly and cleanly than the Nile red-based dyes thus far. For additional characterization, we hope to use single-molecule super-resolution microscopy⁶⁶ and fluorescence recovery after photobleaching (FRAP)³⁵ experiments to more conclusively determine whether our probes only become fluorescent on localization to the mycomembrane or whether other subcellular compartments are involved.

Once we have identified the most useful candidates from a set of fully characterized solvatochromic trehalose probes and have established their activity in Mtb, we can develop new assays based on them. One example would be pulse-chase labeling experiments with probes of different colors to determine the effect of varying specific conditions on mycobacterial cell wall dynamics. Mycobacteria could be incubated with the first probe, treated with a compound, then incubated with the second probe; changes in the ratio of labeling could indicate either a bacteriostatic effect or the compound's ability to compromise the integrity of the cell wall. This type of experiment could even be done in combination with other probes, such as the recently reported Ag85-dependent fluorogenic QTF probe from Kiessling and coworkers.³⁷ While QTF reports specifically on Ag85 activity, our probes' sensitivity to water means that they would be affected more generally by disruption of the highly structured cell wall, resulting in decreased fluorescence from either reduced Ag85 activity or decreased hydrophobicity in the surrounding environment.

Beyond *in vitro* screens of compound activity, our probes have the potential to be used in *in vivo* screens as well as studies of host-pathogen interactions more generally. The disconnect between *in vivo* and *in vitro* TB drug activity and rising drug resistance makes development of new assays for phenotypic screening essential; also, our limited understanding of the underlying mechanisms of latency and reactivation of disease would benefit from assessing cell wall dynamics in a minimally perturbed infection setting. Ideally, we would be able to use our probes to label zebrafish or macrophages that had already been infected with mycobacteria in order to perform *in vivo* and *ex vivo* pulse-chase experiments.

Finally, as with DMN-Tre, our probes have the potential to complement existing sputum smear microscopy-based TB diagnostics and accelerate culture-based methods such as the BACTEC MGIT system.⁶⁷ The new probes' higher fluorescence intensity at lower concentrations than DMN-Tre is particularly promising. Outside of TB diagnostics, these probes may also prove useful as a rapid diagnostic technique for such pathogens as *Mycobacterium ulcerans*. Closely related to *M. marinum*, *M. ulcerans* is the causative agent of Buruli ulcer, a disease that can lead to permanent disfigurement and disability.⁶⁸⁻⁷⁰ Diagnosis of *M. ulcerans* infection is arguably more difficult than diagnosis of TB: initial diagnosis is generally based on appearance of the ulcer and geographic location, PCR-based tests are not specific to *M. ulcerans*, the Ziehl-Neelson stain is only 40-80% sensitive, and culture of these slow-growing bacteria is only 20-60% sensitive.⁷¹ Development of better diagnostics is critical, however, as an eight-week antibiotic course cures 80% of cases that are caught early. *M. ulcerans* infections have been reported in 33 countries, largely in Africa, but incidences are also rapidly increasing in Australia. Even the mechanism of transmission is not fully understood. Much about this neglected tropical disease remains to be discovered, and our probes could readily be applied to both basic research and clinical practice, including rapid diagnosis and assessment of treatment response.

MATERIALS AND METHODS.

Biological experiments

Stock solutions. All dye, dye-Glc, and dye-Tre stock solutions were prepared in water at concentrations of 10 mM and 1 mM and stored at -20 °C, protected from light.

Fluorescence spectra. 1 μ L of each dye-Tre conjugate (10 mM in water) was added to 1 mL of water or 99.9%, 99%, 95%, 90%, 75%, or 50% 1,4-dioxane. Fluorescence data were obtained on a Photon Technology International Quanta Master 4 L-format scanning spectrofluorometer equipped with an LPS-220B 75-W xenon lamp and power supply, an A-1010B lamp housing with an integrated igniter, a switchable 814 photon-counting/analog photomultiplier detection unit, and an MD5020 motor driver. Measurements were made in 1 cm x 0.4 cm quartz cuvettes (Starna Cells, Inc. 9F-G-10).

In the associated FelixGX software (Version 4.3.4.2010.6904), spectra were acquired using standard emission scan settings with the exception of the lamp slit widths, which were all set to 1 nm. Compounds were excited at 405 nm, 488 nm, 532 nm, or 561 nm and emission intensity was monitored over 415-600 nm, 500-700 nm, 545-700 nm, or 575-750 nm, respectively. Data was exported as a text file and processed in Excel (e.g., solvent background was subtracted). Prism 7 (GraphPad) was used to create the figure from the final data.

Bacterial cultures. Cultures of Msmeg, Msmeg transporter mutants, Mmar, Cg, Bs, Ec, and Sa were generated as described in reference 38. Briefly, Msmeg and Msmeg mutant overnight cultures were grown at 37 °C from single-colony inoculations of BD Difco Middlebrook 7H9 media supplemented with 10% (v/v) oleate-albumin-dextrose-catalase (OADC) enrichment, 0.5% (v/v) glycerol, and 0.5% (w/v) Tween 80. Cg, Bs, Ec, and Sa overnight cultures were grown from single-colony inoculations of LB medium, Cg at 30 °C and Bs, Ec, and Sa at 37 °C. A 1-mL aliquot of Mmar that had been stored as an OD₆₀₀=0.5 stock frozen in 50% glycerol/50% 7H9 + OADC was thawed, then the bacteria were pelleted at 3300 xg for 3 min., resuspended in 7H9 media supplemented with 10% (v/v) OADC enrichment, 0.5% (v/v) glycerol, and 0.5% (w/v) Tween 80, and incubated at 33 °C overnight.

Flow cytometry. Cells were pelleted at 3300 xg for 3 min., washed twice with 1x PBS, and resuspended in 300 µL 1x PBS. Experiments were performed on either a BD Accuri C6 instrument or a BD LSR II.UV instrument in the shared Fluorescence Activated Cell Sorting (FACS) Facility at Stanford University. The instrument, excitation wavelengths, and filter sets used are noted in each figure or figure caption. Data were obtained for 100,000 cells per sample, processed using FlowJo, and finally imported into Prism 7 (Graphpad) for statistical analysis.

Confocal microscopy. For no-wash experiments, samples (6 µL) were taken directly from the labeled culture; otherwise, samples (6 µL) were washed as described in the flow cytometry section. The samples were spotted onto 1% agarose pads on microscope slides, allowed to dry, and covered with a cover slip which was then sealed with nail polish. Microscopy was performed on a Nikon A1R confocal microscope equipped with a Plan Fluor 60x oil immersion 1.30-numerical aperture objective. Samples were excited with a 405-nm violet laser, 488-nm blue laser, or 561-nm green laser and imaged in the Aqua Amine (425-475 nm), FITC/GFP (500-550 nm), or RFP (570-620 nm) channels, respectively. NIS-Elements AR software (Nikon Inc.) and Fiji (ImageJ)⁷² were used to process images. All image acquisition and processing were executed under identical conditions for control and test samples.

Metabolic labeling experiments. Experiments were performed as described in reference 38 with some modifications. Briefly, overnight bacterial cultures were grown or diluted to an OD₆₀₀ of 0.4-0.5 and aliquoted into the number of culture tubes required for the experiment. To the aliquots were added enough of a 1 or 10 mM stock of dye, dye-

Glc, and/or dye-Tre to reach the final concentration needed for the experiment. Vehicle controls were treated identically, but without the addition of any probes. The bacteria were incubated at 37 °C for 1 h (Msmeg, Msmeg mutants), 37 °C for 2 h (Bs, Ec, Sa), 30 °C for 2 h (Cg), or 33 °C for 3 h (Mmar). At the end of the experiment, samples were analyzed by microscopy and/or flow cytometry as described above.

Trehalose competition. An overnight culture of Msmeg grown to an OD₆₀₀ of 0.4 was divided into 100 µL aliquots. Each aliquot aside from the vehicle control was treated with 3-HC-3-Tre (to 100 µM), 3-HC-2-Tre (to 10 µM), MeNR-Tre (to 10 µM), EtNR-Tre (to 10 µM), or MeNR-Glc (to 10 µM) in addition to trehalose (to 0, 1x, 10x, 100x, or 1000x the concentration of the probe). The bacteria were allowed to grow 1 h at 37 °C then analyzed by flow cytometry as described above.

Ebselen inhibition. An overnight culture of Msmeg grown to an OD₆₀₀ of 0.4 was divided into aliquots and incubated with 0, 25, 50, or 100 µg/mL ebselen for 3 h. For each condition, the ebselen-treated bacteria were incubated with 3-HC-2, 3-HC-2-Tre, MeNR, or MeNR-Tre at a final concentration of 10 µM. After 1 h at 37 °C, samples were analyzed by flow cytometry as described above.

Heat-killed Msmeg. Msmeg were heat-killed by incubating at 95 °C for 30 min.

Dual labeling of Msmeg with DMN-Tre and EtNR-Tre. An overnight culture of Msmeg grown to an OD₆₀₀ of 0.4-0.5 was divided into aliquots. Each aliquot was treated with different amounts of DMN-Tre and EtNR-Tre (100 µM/0 µM, 100 µM/10 µM, 100 µM/100 µM, 10 µM/100 µM, and 0 µM/100 µM) then washed and analyzed by fluorescence microscopy.

Isoniazid treatment of wildtype Msmeg and a $\Delta katg$ mutant. Overnight cultures of wildtype Msmeg and $\Delta katg$ Msmeg were aliquoted and treated with 0 or 10 µg/mL isoniazid for 3 h at 37 °C. The pretreated bacteria were then incubated at 37 °C for 1 h with 3-HC-2-Tre, MeNR-Tre, or EtNR-Tre at a final concentration of 10 µM. After washing, the bacteria were analyzed by flow cytometry.

Imaging of Mmar in zebrafish. Overnight cultures of Mmar were incubated with 100 µM DMN-Tre, 10 µM 3-HC-2-Tre, 10 µM MeNR-Tre, or 10 µM EtNR-Tre for 3 h and imaged. The zebrafish experiments were performed by Shivam Verma using procedures adapted from reference 65. Wildtype embryonic zebrafish (2 days post-fertilization) were anaesthetized by immersing the fish in 1x embryo media containing 0.4 g/L tricaine-s, infected with labeled or unlabeled Mmar via hindbrain ventricle injection (50-100 bacteria/fish using an Eppendorf FemtoJet Microinjector), and immediately imaged in Z stacks with confocal microscopy.

Statistics. Statistical analyses were done in Prism 7 (Graphpad). The method that was applied is noted in each figure caption.

Synthesis and characterization (small molecules)

General. Reactions were performed in flame- or oven-dried glassware under an inert nitrogen atmosphere unless otherwise noted. Anhydrous solvents were either purchased or obtained by passing solvent through an activated alumina column via a Pure Process Technology Glass Contour Solvent Purification System. All reagents and solvents were used as received unless otherwise noted. Water was passed through a Milli-Q filtration system prior to use. Where noted, samples were concentrated *in vacuo* at 40 °C using a BÜCHI Rotavapor R-114 equipped with a BÜCHI B-480 heating bath and a Welch Self-Cleaning Dry Vacuum System (Model 2025) or an IKA RV 10 basic rotary evaporator equipped with an IKA HB 10 basic heating bath and a Welch Self-Cleaning Dry Vacuum System (Model 2025). If necessary, compounds were then further dried under high vacuum using an Edwards RV8 Two Stage Rotary Vane Pump or by lyophilization in a LABCONCO FreeZone 4.5Plus.

Thin layer chromatography was performed using SiliCycle SiliaPlate glass-backed silica gel plates containing a fluorescent indicator (Fisher Scientific 50964470). Plates were visualized using a UVGL-25 Compact UV Lamp, 254/365 nm, 4 W (P/N 95-0021-12). For flash column chromatography, the stationary phase was SiliCycle SiliaFlash P60 or Fisher Silica Gel Sorbent (230-400 Mesh, Grade 60) silica gel. For purifications involving preparative Reversed-Phase High-Performance Liquid Chromatography (RP-HPLC), the following conditions were used: the instrument consisted of an Agilent Technologies ProStar 325 UV-Vis detector, two PrepStar Solvent Delivery Modules, and a 440-LC Fraction Collector; the column was either a Varian Microsorb 100Å C18, 8 µm, 21.4 x 250 mm Dynamax preparative column (R0080220C8) equipped with a Microsorb 100Å C18, 8 µm guard column (R0080220G8) or a Phenomenex Luna® 10 µm Phenyl-Hexyl 100 Å, 250 x 21.2 mm preparative column (00G-4325-P0-AX). For the C18 column, solvent A was 0.1% TFA in Milli-Q water; solvent B was 0.1% TFA in acetonitrile (MeCN). For the phenyl-hexyl column, solvent A was pure Milli-Q water; solvent B was methanol. The UV-Vis detector was used to monitor wavelengths at 210 and 254 or 550 nm. All pure compounds and stock solutions were stored at -20 °C. For Isolera™ Prime purifications, the following general conditions were applied: the instrument was a Biotage Isolera™ Prime with one channel, a single collection bed, and a 200-400 nm detector (model ISO-PSV). Specific solvents, monitored wavelengths, and column types are noted in individual synthesis sections.

High resolution mass spectrometry (HRMS) data were acquired by ESI-LC/MS on a Waters Acquity UPLC and Thermo Exactive Orbitrap mass spectrometer by Dr. Theresa McLaughlin at the Stanford University Mass Spectrometry facility.

Sugars

6-bromo-6-deoxy- α,α -trehalose heptaacetate (6-Br-Ac7-Tre). In a 200 mL flame-dried roundbottom flask equipped with a stir bar, anhydrous DMF (100 mL) was heated to 85 °C, then anhydrous trehalose (4 g, 11.69 mmol) and recrystallized NBS (2.29 g, 12.9 mmol, 1.1 equiv.) were dissolved therein. On addition of the NBS, the solution changed to pale yellow, then orange-yellow, then orange. After 5 minutes, PPh₃ (6.13 g, 23.37 mmol, 2 equiv.) was added, and the solution immediately turned a clear/champagne color. The mixture was allowed to stir 48 h at r.t., then the DMF was removed *in vacuo* to yield a clear off-white oil. Anhydrous pyridine (100 mL) was added, then the resulting solution was cooled to 0 °C and acetic anhydride (~15.5 mL, 164 mmol, 2 equiv. relative to the 7 hydroxyl groups on 6-Br-Tre) was added fast dropwise. The yellow solution was allowed to stir 48 h at r.t., over which time it became a clear orange-red solution. After removing the solvent *in vacuo*, the resulting orange oil was redissolved in ethyl acetate, washed twice with concentrated sodium bicarbonate (orange aqueous layer) and twice with brine (clear aqueous layer). The clear orange organic layer was dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. Product was purified by silica gel chromatography on an Isolera Prime (Biotage) equipped with a Zip KP-Sil 120 g column equilibrated with 5 column volumes (CV) of the starting solvent mixture. The flow rate was 100 mL/min., and all fractions were collected. The gradient was as follows: 5% EtOAc in DCM for 1 CV, 5-30% EtOAc in DCM over 10 CV, 30% EtOAc in DCM for 2 CV. Pure fractions were determined by TLC (4:1 DCM/EtOAc), visualizing first with UV to check for triphenylphosphine oxide side product then by 5% H₂SO₄ in MeOH and charring. Of the three sugar spots observed, the top one was dibrominated side product, the middle was product, and the bottom was peracetylated trehalose side product. Pure fractions were concentrated *in vacuo* to a white crystalline foam and lyophilized to yield 0.9727 g (11.9%) of hard white crystals. ¹H NMR (500 MHz, Chloroform-*d*) δ 5.47 (q, J = 10.0 Hz, 2H), 5.32 (t, J = 4.0 Hz, 2H), 5.13 (dd, J = 10.3, 3.9 Hz, 1H), 5.07 – 4.91 (m, 3H), 4.21 (dd, J = 12.4, 6.0 Hz, 1H), 4.10 (q, J = 7.1 Hz, 2H), 4.06 – 3.97 (m, 2H), 3.41 – 3.27 (m, 2H), 2.14 – 2.05 (m, 11H), 2.05 – 1.98 (m, 9H), 1.24 (t, J = 7.1 Hz, 1H). ¹³C{¹H} NMR (126 MHz, Chloroform-*d*) δ 170.67, 170.02, 169.96, 169.66, 169.63, 169.57, 92.27, 91.82, 71.18, 70.14, 69.89, 69.73, 69.63, 69.30, 68.56, 68.26, 61.79, 60.45, 30.48, 21.02, 20.74, 20.67, 20.63, 14.25. HRMS (ESI) *m/z*: [M + Na]⁺ Calcd for C₂₆H₃₅BrO₁₇Na 721.0950; Found 721.0932.

Methyl 6-iodo-6-deoxy- α -D-glucopyranoside triacetate (Me 6-I-Ac3-Glc). In a 500-mL flame-dried two-necked roundbottom flask equipped with a stir bar, methyl α -D-glucoside (2.5 g, 12.9 mmol), PPh₃ (5.065 g, 19.3 mmol, 1.5 equiv.), and imidazole (1.753 g, 25.7 mmol, 2 equiv.) were combined and refluxed in anhydrous THF (100 mL). A solution of iodine (I₂; 4.902 g, 19.3 mmol, 1.5 equiv.) in anhydrous THF (25 mL) was added fast dropwise and the reaction was allowed to reflux a further 2 h. The reaction was allowed to cool to r.t., then the solution was vacuum-filtered to remove the white precipitate (imidazole salt), transferred to a 500-mL roundbottom flask, and concentrated *in vacuo* to a slightly cloudy viscous oil. The oil was dissolved in anhydrous pyridine (100 mL) and cooled to 0 °C, then acetic anhydride (7.3 mL, 77.2 mmol, 2 equiv. relative to the 3 free hydroxyl groups) was added fast dropwise. The reaction was allowed to stir overnight at

r.t. under nitrogen gas then concentrated *in vacuo* to a white solid/yellow oil. Product was purified by silica gel chromatography on an Isolera Prime (Biotage) equipped with a Zip KP-Sil 120 g column equilibrated with 3 CV of the starting solvent mixture. The flow rate was 100 mL/min., the 254/280 nm wavelengths were monitored (iodine absorbs at 254 nm), and peaks that absorbed above 25 mAu were collected. The gradient was as follows: 10% EtOAc in hexanes for 2 CV, 10-65% EtOAc in hexanes over 6 CV, 65-100% EtOAc in hexanes over 1 CV, and 100% EtOAc for 2.5 CV. Pure fractions were determined by TLC (1:1 hexanes/EtOAc), visualizing first with UV then by 5% H₂SO₄ in MeOH and charring. Pure fractions were combined and concentrated *in vacuo* then recrystallized from acetone and hexanes. After high vacuum, 3.3066 g (59.7%) of white fluffy crystals were obtained. ¹H NMR (500 MHz, Chloroform-*d*) δ 5.45 (dd, J = 10.2, 9.3 Hz, 1H), 4.94 (d, J = 3.7 Hz, 1H), 4.88 – 4.83 (m, 2H), 3.77 (ddd, J = 10.2, 8.3, 2.5 Hz, 1H), 3.46 (s, 3H), 3.28 (dd, J = 10.9, 2.5 Hz, 1H), 3.12 (dd, J = 10.9, 8.3 Hz, 1H), 2.06 (s, 3H), 2.04 (s, 3H), 1.99 (s, 3H). ¹³C{¹H} NMR (126 MHz, Chloroform-*d*) δ 170.34, 170.26, 169.89, 96.91, 72.68, 71.12, 69.87, 68.84, 55.99, 20.98, 20.92, 3.93. HRMS (ESI) *m/z*: [M + Na]⁺ Calcd for C₁₃H₁₉IO₈Na 453.0017; Found 453.0010.

Dyes

3-HC-3. The aldehyde precursor was synthesized used procedures adapted from reference 55.

2-Bromo-9,9-dimethyl-7-nitrofluorene. 2-bromo-9,9-dimethylfluorene (5 g, 18.3 mmol) was dissolved in glacial acetic acid (75 mL) in a 250-mL roundbottom flask equipped with a stir bar. The solution was cooled to 0 °C, then sulfuric acid was added (25 mL). To this solution was added 69% nitric acid (22.5 mL) fast dropwise at 0 °C with vigorous stirring. After the addition was complete, the reaction was stirred at r.t. for ~30 min., at which point TLC (9:1 hexanes/EtOAc, visualized with shortwave UV) showed significant disappearance of starting material (R_f 0.75) accompanied by appearance of one new major spot (R_f 0.54) and ~3 very faint minor ones (R_f 0.47, 0.41, 0.34). Then, the reaction mixture was poured into water (600 mL), and the green precipitate was collected by vacuum filtration. The crude product was recrystallized from acetonitrile and dried under high vacuum to yield 3.9276 g (67.4%) of yellow needle-like crystals. ¹H NMR (500 MHz, Chloroform-*d*) δ 8.29 – 8.24 (m, 2H), 7.79 (dd, J = 8.2, 0.7 Hz, 1H), 7.66 (d, J = 8.1 Hz, 1H), 7.63 (d, J = 1.7 Hz, 1H), 7.54 (dd, J = 8.1, 1.8 Hz, 1H), 1.53 (s, 6H). ¹³C{¹H} NMR (126 MHz, Chloroform-*d*) δ 157.18, 154.48, 147.66, 144.87, 136.06, 131.13, 126.86, 123.91, 123.82, 123.01, 120.54, 118.87, 118.64, 77.54, 77.29, 77.04, 47.83, 26.91, 26.69.

7-Bromo-9,9-dimethylfluorenyl-2-amine. In a 150-mL roundbottom flask equipped with a stir bar, 2-bromo-9,9-dimethyl-7-nitrofluorene (2 g, 6.3 mmol), iron powder (1 g, 17 mmol, 2.7 equiv.), and ammonium chloride (740 mg, 12.6 mmol, 2 equiv.) were dissolved in a mixture of ethanol (90 mL) and water (25 mL). The solution was refluxed at 85 °C for 2 h under N₂. Then, saturated sodium bicarbonate (50 mL) was added, turning the reaction a mustard yellow/brown. The mixture was vacuum-filtered to remove the iron powder, washing with ethanol. The organic solvent was removed *in vacuo*, and the remaining suspension was vacuum-filtered. The brown-orange solid was then collected and purified

by reversed-phase C18 chromatography on an Isolera Prime (Biotage) equipped with a SNAP Ultra 60 g column equilibrated with 5 CV of the starting solvent mixture. The flow rate was 75 mL/min., the 254/280 nm wavelengths were monitored, and peaks that absorbed above 30 mAu were collected. The gradient was as follows: 40% MeCN/water + 0.1% TFA for 6 CV, 40-70% MeCN/water + 0.1% TFA over 7 CV, 70-100% MeCN/water + 0.1% TFA over 3 CV, and 100% MeCN + 0.1% TFA for 5 CV. Pure fractions (determined by LC-MS) were combined and concentrated *in vacuo* to remove MeCN and TFA, then the product was collected by vacuum filtration and left under high vacuum overnight. This yielded 1.4954 g (83.1%) of a white solid. ¹H NMR (500 MHz, Chloroform-*d*) δ 7.47 (s, 2H), 7.40 (d, J = 12.3 Hz, 2H), 6.73 (s, 1H), 6.66 (s, 1H), 3.80 (s, 2H), 1.43 (s, 6H). ¹³C{¹H} NMR (126 MHz, Chloroform-*d*) δ 207.34, 155.09, 146.81, 130.09, 126.05, 121.26, 120.25, 119.26, 114.35, 109.57, 47.08, 31.21, 27.41.

(7-Bromo-9,9-dimethylfluorenyl)-2-diethylamine. In a flame-dried 200-mL recovery flask equipped with a stir bar, 7-bromo-9,9-dimethylfluorenyl-2-amine (1.5608 g, 5.4 mmol), iodoethane (3.61 mL, 58 mmol, 5.4 equiv.), and potassium carbonate (9.67 g, 70 mmol) were dissolved in anhydrous DMF (40 mL). The reaction was stirred at 80 °C for 5 h, cooled, and poured into water (150 mL). The resulting solution was neutralized using 2 M HCl and extracted 3x with chloroform. The combined organic extracts (yellow) were dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo* to a yellow oil. The product was purified by silica gel chromatography on an Isolera Prime (Biotage) equipped with a SNAP KP-Sil 50 g column in line with a SNAP KP-Sil 100 g column equilibrated with 4 CV of the starting solvent mixture. The flow rate was 100 mL/min., the 254/280 nm wavelengths were monitored, and peaks that absorbed above 40 mAu were collected. The gradient was as follows: 20% DCM/hexanes for 9 CV, 20-30% DCM/hexanes over 15 CV, 30% DCM/hexanes for 1 CV, 100% DCM for 2 CV. Pure fractions were determined by LC-MS and concentrated *in vacuo*. After high vacuum, this yielded 555.8 mg (30%) of a pale pink solid. ¹H NMR (500 MHz, Chloroform-*d*) δ 7.59 (t, J = 1.2 Hz, 1H), 7.58 (d, J = 8.4 Hz, 1H), 7.48 (d, J = 1.1 Hz, 2H), 6.80 (d, J = 2.4 Hz, 1H), 6.74 (dd, J = 8.5, 2.5 Hz, 1H), 3.51 (q, J = 7.1 Hz, 4H), 1.56 (s, 6H), 1.31 (t, J = 7.1 Hz, 6H). ¹³C{¹H} NMR (126 MHz, Chloroform-*d*) δ 155.51, 155.20, 148.42, 139.50, 130.18, 126.34, 126.04, 121.47, 120.05, 118.66, 111.19, 105.87, 47.27, 45.08, 27.78, 13.03.

7-Diethylamino-9,9-dimethylfluorene-2-carbaldehyde. In a 200-mL recovery flask equipped with a stir bar and flame-dried under three cycles of vacuum and nitrogen gas, (7-Bromo-9,9-dimethylfluorenyl)-2-diethylamine (555.8 mg, 1.62 mmol) was dissolved in anhydrous THF (8 mL). The solution was cooled to -78 °C and n-butyllithium (2.5 M in hexanes; 0.72 mL, 1.8 mmol, 1.1 equiv.; **CAUTION – PYROPHORIC**) was added dropwise under nitrogen. The reaction was allowed to stir for 1 h at -78 °C, over which time the solution color changed from yellow to green and back to yellow. Then, anhydrous DMF (0.25 mL, 3.2 mmol, 2 equiv.) was added and the reaction allowed to stir a further 2 h at -78 °C. At this time, the reaction was allowed to warm to r.t. over 1 h, turning a goldenrod yellow. After quenching with 1 M HCl (to a pH of ~5), the solution was extracted 3x with EtOAc. The combined light yellow-green EtOAc layers were dried over anhydrous

sodium sulfate, filtered, and concentrated *in vacuo* yield a brownish green oil that solidified to green crystals on standing at 4 °C. The product was purified by silica gel chromatography on an Isolera Prime (Biotage) equipped with a SNAP KP-Sil 100 g column equilibrated with 4 CV of the starting solvent mixture. The flow rate was 100 mL/min., the 254/280 nm wavelengths were monitored, and peaks that absorbed above 40 mAu were collected. The gradient was as follows: 100% hexanes for 1 CV, 0-15% EtOAc/hexanes over 10 CV, 15% EtOAc/hexanes for 2 CV. Pure fractions were determined by LC-MS, concentrated *in vacuo*, and dried under high vacuum overnight to yield approximately 300 mg (63%) of a yellow solid.

3-HC-3. In a flame-dried three-necked 25-mL roundbottom flask equipped with a stir bar, 7-Diethylamino-9,9-dimethylfluorene-2-carbaldehyde (50 mg, 0.17 mmol) was dissolved in anhydrous MeOH (5 mL). 2'-hydroxyacetophenone (23 μ L, 0.193 mmol, 1.14 equiv.) and crushed NaOH (20.4 mg, 0.51 mmol, 3 equiv.) were added and the bright yellow reaction was refluxed overnight at 75 °C. Due to the small volume, the reaction dried to a red-orange oil overnight; 5 mL of anhydrous MeOH was added to redissolve it and the reaction was allowed to reflux a further 2 h. After checking that no more product had formed by TLC (9:1 hexanes/EtOAc, visualized by UV), the orange solution was removed from heat and allowed to cool completely to r.t. over approximately 1 h. Then, 0.5 M NaOH (1.02 mL) was added slow dropwise while stirring thoroughly, making sure that no solid crashed out of the reaction mixture. To the resulting clear red solution was added 30% hydrogen peroxide (70 μ L, 0.7 mmol, 4.1 equiv.) dropwise. The reaction was covered with foil and allowed to stir under nitrogen at r.t. overnight. The reaction mixture was directly purified by reversed-phase C18 chromatography on an Isolera Prime (Biotage) equipped with a SNAP Ultra 60 g column equilibrated with 5 CV of the starting solvent mixture. The flow rate was 75 mL/min., the 210/254 nm wavelengths were monitored, and peaks that absorbed above 5 mAu were collected. The column was visually monitored for elution of colored bands – the solvent system was held at 72% MeCN/water (no TFA) until a pale orange band (yellow fractions) and a red-orange band (orange and red fractions) had mostly eluted (~26 CV), then the %MeCN was increased to 100% over 2 CV and held at 100% for 5 CV. Product eluted around CV 27~30 (a smaller peak very soon after/overlapping with the tail end of a large peak). Pure fractions (determined by LC-MS) were combined and concentrated *in vacuo*, then the product was dried under high vacuum overnight to yield 6.8 mg (9.7%) of an orange solid. ¹H NMR (500 MHz, Acetone-*d*₆) δ 8.39 (d, *J* = 1.7 Hz, 1H), 8.29 (dd, *J* = 8.2, 1.7 Hz, 1H), 8.18 (dd, *J* = 8.0, 1.6 Hz, 1H), 7.85 – 7.74 (m, 3H), 7.67 (d, *J* = 8.5 Hz, 1H), 7.51 – 7.46 (m, 1H), 6.92 (d, *J* = 2.4 Hz, 1H), 6.74 (dd, *J* = 8.5, 2.4 Hz, 1H), 3.51 (q, *J* = 7.1 Hz, 4H), 1.53 (s, 6H), 1.21 (t, *J* = 7.0 Hz, 6H). HRMS (ESI) *m/z*: [M + H]⁺ Calcd for C₂₈H₂₈NO₃ 426.2064; Found 426.2070.

3-HC-2. The aldehyde precursor was synthesized used procedures adapted from reference 53.

N,N-diethyl-3-(2,2-diethoxyethoxy)aniline. In a flame-dried 250-mL roundbottom flask equipped with a stir bar, 3-diethylaminophenol (10 g, 60.5 mmol) was dissolved in anhydrous THF (53 mL). While stirring, 60% sodium hydride (2.905 g, 121 mmol, 2 equiv.)

was added slowly, releasing hydrogen gas. The solvent was removed *in vacuo* until approximately 10 mL of a dark brown oil remained; this oil was redissolved in anhydrous DMSO (42 mL). To this solution was added bromoacetaldehyde diethylacetal (10.95 mL, 73.5 mmol, 1.2 equiv.) and potassium iodide (15.31 g, 92.2 mmol, 1.5 equiv.). The reaction was heated at 50 °C for 3 h, then water (75 mL) was added. The solution was extracted 4x with benzene, then the organic layers were combined, dried over anhydrous sodium sulfate, filtered, and concentrated *in vacuo*. The product was purified by silica gel chromatography on an Isolera Prime (Biotage) equipped with a SNAP KP-Sil 100 g column equilibrated with 3 CV of the starting solvent mixture. Multiple runs were performed. The flow rate was 100 mL/min., the 254/280 nm wavelengths were monitored, and peaks that absorbed above 40 mAu were collected. The gradient was as follows: 5% EtOAc/hexanes for 1 CV, 5-36% EtOAc/hexanes over 9 CV. Pure fractions were determined by TLC (4:1 hexanes/EtOAc, visualized with UV) and LC-MS, concentrated *in vacuo*, and dried under high vacuum overnight to yield 3.9902 g (23.3%) of a viscous pale yellow oil. ¹H NMR (500 MHz, Chloroform-*d*) δ 7.13 (t, J = 8.2 Hz, 1H), 6.37 – 6.23 (m, 3H), 4.87 (t, J = 5.2 Hz, 1H), 4.04 (d, J = 5.2 Hz, 2H), 3.79 (dq, J = 9.3, 7.1 Hz, 2H), 3.66 (dq, J = 9.4, 7.0 Hz, 2H), 3.34 (q, J = 7.1 Hz, 4H), 1.29 (t, J = 7.0 Hz, 7H), 1.18 (t, J = 7.1 Hz, 6H). ¹³C{¹H} NMR (126 MHz, Chloroform-*d*) δ 160.32, 149.36, 130.15, 105.55, 101.15, 100.86, 99.22, 68.51, 62.50, 44.64, 15.66, 12.92.

6-diethylaminobenzofuran-2-carbaldehyde. In a flame-dried 25-mL three-necked roundbottom flask equipped with a stir bar, POCl₃ (2.186 mL, 67.2 mmol, 9.1 equiv.) was cooled to 0 °C, then anhydrous DMF (0.574 mL, 7.35 mmol, 1.03 equiv.) was added, followed by N,N-diethyl-3-(2,2-diethoxyethoxy)aniline (2 g, 7.12 mmol). The reaction was stirred at 50 °C for 14 h, resulting in a pale brown paste which was then scraped into ice, rinsing with ~150 mL water. The dark red solution was neutralized with 4 M NaOH, becoming cloudy orange. The mixture was extracted with 3x EtOAc, then the combined organic layers were dried over sodium sulfate, filtered, and concentrated *in vacuo* to an orange-brown oil. The product was purified by silica gel chromatography on an Isolera Prime (Biotage) equipped with a Zip KP-Sil 80 g column equilibrated with 5 CV of the starting solvent mixture. The flow rate was 75 mL/min., the 254/280 nm wavelengths were monitored, and all fractions were collected. The gradient was as follows: 5% EtOAc/hexanes for 1 CV, 5-40% EtOAc/hexanes over 10 CV, 40% EtOAc/hexanes for 1 CV. Pure fractions were determined by TLC (4:1 hexanes/EtOAc, visualized with UV) and LC-MS, concentrated *in vacuo*, and dried under high vacuum overnight to yield 136.2 mg (9%) of a red-brown oil. HRMS (ESI) *m/z*: [M + H]⁺ Calcd for C₁₃H₁₆NO₂ 218.1176; Found 218.1174.

3-HC-2. A flame-dried 15-mL two-necked roundbottom flask equipped with a stir bar was charged with a solution of 6-diethylaminobenzofuran-2-carbaldehyde (41.8 mg, 0.19 mmol) in anhydrous MeOH (1.16 mL). 2'-hydroxyacetophenone (23 μL, 0.19 mmol) and crushed NaOH (23.1 mg, 0.58 mmol, 3 equiv.) were added and the reaction was refluxed overnight at 75 °C. Due to the small volume, the reaction dried to a red-orange solid overnight; 5 mL of MeOH was added to redissolve it and the reaction was allowed to

reflux a further 2 h. After checking that no more product had formed by TLC (2:1 hexanes/EtOAc, visualized by UV), the orange solution was removed from heat and allowed to cool completely to r.t. over approximately 1 h. Then, 0.5 M NaOH (1.16 mL) was added slow dropwise while stirring thoroughly, making sure that no solid crashed out of the reaction mixture. To the resulting clear red solution was added 35% hydrogen peroxide (79 μ L, 0.8 mmol, 4.1 equiv.) dropwise. The reaction was covered with foil and allowed to stir for 3 h at r.t. under nitrogen. TLC and LC-MS confirmed completion of the reaction, which was then poured into water. The MeOH was removed *in vacuo* and the remaining aqueous solution was extracted several times with EtOAc. The combined EtOAc layers (clear orange solution tinted green) were dried over anhydrous sodium sulfate, filtered, concentrated *in vacuo*, and dried under high vacuum overnight to yield 49.2 mg (73.2%) of an orange-red solid. [If necessary, the product was purified by silica gel chromatography on an Isolera Prime (Biotage) equipped with a SNAP KP-Sil 50 g column equilibrated with 5 CV of the starting solvent mixture. The flow rate was 100 mL/min., the 254/280 nm wavelengths were monitored, and peaks that absorbed above 40 mAu were collected. The gradient was as follows: 8% EtOAc/hexanes for 1 CV, 8-66% EtOAc/hexanes over 10 CV, 66% EtOAc/hexanes for 2 CV. Pure fractions were determined by TLC (4:1 hexanes/EtOAc, visualized with UV) and LC-MS, concentrated *in vacuo*, and dried under high vacuum overnight.] HRMS (ESI) *m/z*: [M + H]⁺ Calcd for C₂₁H₁₉NO₄ 350.1387; Found 350.1387.

9-(diethylamino)-2-hydroxy-5H-benzo[a]phenoxazin-5-one (EtNR). Synthetic procedures were adapted from reference 49. In a flame-dried 100-mL roundbottom flask equipped with a stir bar, 5-(diethylamino)-2-nitrosophenol hydrochloride (703.3 mg, 3.1 mmol, 1 eq.) and 1,6-dihydroxynaphthalene (490 mg, 3.1 mmol, 1 eq.) were dissolved in anhydrous DMF (60 mL). The wine-red solution was refluxed in a 170 °C oil bath under nitrogen for 5 h, then removed from heat and concentrated to a solid. Product was purified by reversed-phase C18 chromatography on an Isolera Prime (Biotage) equipped with a SNAP KP-C18 60 g column equilibrated with 4 CV of the starting solvent mixture. The following method was used to purify EtNR before a better method based on a phenylhexyl column was developed using MeNR: The flow rate was 100 mL/min., the 254/280 nm wavelengths were monitored, and peaks that absorbed above 40 mAu were collected. The gradient was as follows: 5% MeCN/water + 0.1% TFA for 2 CV, 5-45% MeCN/water + 0.1% TFA over 10 CV, 45% MeCN/water + 0.1% TFA for 5.5 CV, 45-65% MeCN/water + 0.1% TFA over 5 CV, 65% MeCN/water + 0.1% TFA for 2 CV, and 100% MeCN + 0.1% TFA for 8 CV. LC-MS was used to determine that two peaks contained the same compound, with the desired mass. These peaks were assumed to be different salt forms of EtNR based on this and the NMR data (see MeNR synthesis for a purification that does not introduce salts). Pure fractions for each peak (determined by LC-MS) were combined and concentrated *in vacuo*, then left under high vacuum overnight. Peak 1 yielded 88 mg of a dark black/blue solid (likely the salt), while peak 2 yielded 87.1 mg of a dark maroon/purple solid. Together, the yield was approximately 17.1%. The peak 2 compound was used in future reactions. Peak 1: ¹H NMR (500 MHz, DMF-*d*₇) δ 8.05 (d, J = 8.7 Hz, 1H), 8.00 (d, J = 2.4 Hz, 1H), 7.60 (d, J = 8.9 Hz, 1H), 7.20 (dd, J = 8.5, 2.5

Hz, 1H), 6.85 (dd, J = 9.1, 2.6 Hz, 1H), 6.64 (d, J = 2.6 Hz, 1H), 6.16 (s, 1H), 3.57 (q, J = 7.0 Hz, 4H), 1.23 (t, J = 6.9 Hz, 6H). $^{13}\text{C}\{^1\text{H}\}$ NMR (126 MHz, DMF- d_7) δ 182.00, 161.45, 152.26, 151.37, 147.09, 139.50, 134.54, 131.28, 127.76, 124.67, 124.61, 118.63, 110.32, 108.71, 104.53, 96.46, 44.98, 12.43. **Peak 2:** ^1H NMR (500 MHz, DMF- d_7) δ 8.12 – 7.96 (m, 3H), 7.58 (d, J = 8.7 Hz, 1H), 7.23 – 7.12 (m, 1H), 6.86 (d, J = 8.6 Hz, 1H), 6.63 (s, 1H), 6.17 (s, 1H), 3.56 (q, J = 7.3 Hz, 4H), 1.23 (t, J = 6.9 Hz, 6H). $^{13}\text{C}\{^1\text{H}\}$ NMR (126 MHz, DMF- d_7) δ 181.59, 161.45, 152.20, 151.53, 147.15, 139.16, 134.54, 131.37, 127.72, 125.05, 124.37, 118.62, 110.64, 108.66, 104.21, 96.44, 45.06, 12.44. HRMS (ESI) m/z : $[\text{M} + \text{H}]^+$ Calcd for $\text{C}_{20}\text{H}_{18}\text{N}_2\text{O}_3$ 335.1390; Found 335.1391.

9-(dimethylamino)-2-hydroxy-5H-benzo[a]phenoxazin-5-one (MeNR). Synthetic procedures were adapted from reference 49. In a flame-dried 100-mL roundbottom flask equipped with a stir bar, 5-(dimethylamino)-2-nitrosophenol hydrochloride (608.1 mg, 3.1 mmol, 1 eq.) and 1,6-dihydroxynaphthalene (490 mg, 3.1 mmol, 1 eq.) were dissolved in anhydrous DMF (60 mL). The burgundy solution was refluxed in a 170 °C oil bath under nitrogen for 4 h, then removed from heat and concentrated to a solid. Product was purified by first washing the crude product extensively with water to remove a dark blue impurity that appeared to be a dimer of the naphthalene starting material. The remaining solid was dissolved in methanol and purified by RP-HPLC using a preparative phenyl-hexyl column over 13 5-mL injections (no more than 100 mg/injection). The flow rate was 20 mL/min., the 210/550 nm wavelengths were monitored, and peaks that absorbed at 550 nm were collected using a fraction collector. The gradient was as follows: 68% MeOH/water (no TFA) for 18 min., 68-100% MeOH/water over 1 min., 100% MeOH over 11 min. Product eluted between approximately min. 13 and 16 as fuchsia-colored fractions. Fractions that were deemed pure by LC-MS were combined and concentrated *in vacuo* to a red-purple solid. The compound was resuspended in water (with sonication) and lyophilized overnight to yield 177 mg (18.6%) of a dark maroon powder. HRMS (ESI) m/z : $[\text{M} + \text{H}]^+$ Calcd for $\text{C}_{18}\text{H}_{15}\text{N}_2\text{O}_3$ 307.1077; Found 307.1076.

General procedures for conjugation of dyes with trehalose or glucose: For an approximately 0.02 mmol scale, the following steps were performed (adapted from reference 57). In a flame-dried 10-mL pear-shaped flask, dye (0.02 mmol), anhydrous potassium carbonate (0.0354 mmol, 1.77 equiv.), and anhydrous DMF (~1 mL or enough to fully dissolve the dye, whichever is greater) were combined and stirred for 1 h at r.t. Then, 6-Br-Ac7-Tre or Me 6-I-Ac3-Glc (0.06 mmol, 3 equiv.) was added, and the reaction was allowed to stir 48 h at 80 °C, protected from light. The reaction was allowed to completely cool to r.t., then anhydrous 0.5 M NaOMe in MeOH (~1 mL) was added directly to the mixture. After stirring 1.5 h at r.t., the reaction was directly purified by reversed-phase C18 chromatography on an Isolera Prime (Biotage) equipped with a SNAP Ultra C18 60 g column (75 mL/min.) or a SNAP Ultra C18 30 g column (50 mL/min.) equilibrated with 5-7 CV of the starting solvent mixture. No TFA was added. The gradient was modified on-the-fly; when a 210/254 nm peak appeared or colored bands on the column began to move, the gradient was held at that solvent ratio until the compound eluted. The method

was set to collect peaks absorbing above 3 mAu (210/254 nm), but in practice the user told the machine when to collect based on whether a colored band was eluting. The basic gradient was 100% water for 3 CV, 0-100% MeCN/water over 15 CV, 100% MeCN. Pure fractions were determined by LC-MS, combined, and concentrated *in vacuo*, then resuspended in Milli-Q water and lyophilized overnight.

3-HC-3-Tre. 3-HC-3 (5 mg, 0.012 mmol) yielded 3.291 mg (37.4% over three steps) of an orange powder. NMR signal was too weak to report accurate proton peaks; see Appendix 2 for spectrum. HRMS (ESI) m/z : $[M + H]^+$ Calcd for $C_{40}H_{48}NO_{13}$ 750.3120; Found 750.3115.

3-HC-3-Glc. 3-HC-3 (3 mg, 0.007 mmol) yielded 0.338 mg (8% over three steps) of a yellow-orange powder. 1H NMR (500 MHz, Acetone- d_6) δ 8.56 (s, 1H), 8.21 (t, $J = 9.3$ Hz, 1H), 7.90 – 7.75 (m, 3H), 7.70 (d, $J = 8.5$ Hz, 1H), 7.55 – 7.50 (m, 1H), 7.21 (s, 1H), 6.93 (s, 1H), 6.75 (d, $J = 8.5$ Hz, 1H), 5.26 (s, 1H), 5.11 (s, 1H), 4.76 (s, 1H), 4.35 (s, 1H), 4.26 (d, $J = 10.7$ Hz, 1H), 4.00 (s, 1H), 3.82 (d, $J = 9.0$ Hz, 1H), 3.77 (s, 1H), 3.71 (s, 1H), 3.59 (d, $J = 7.4$ Hz, 1H), 3.52 (q, $J = 7.7, 7.3$ Hz, 4H), 3.33 (d, $J = 24.9$ Hz, 3H), 1.56 (d, $J = 19.1$ Hz, 6H), 1.22 (t, $J = 7.2$ Hz, 6H). HRMS (ESI) m/z : $[M + H]^+$ Calcd for $C_{35}H_{40}NO_8$ 602.2748; Found 602.2760.

3-HC-2-Tre. 3-HC-2 (10 mg, 0.031 mmol) yielded 1.672 mg (8.7% over three steps) of an orange fluffy solid. 1H NMR (500 MHz, Methanol- d_4) δ 8.19 (d, $J = 8.0$ Hz, 1H), 7.99 (s, 1H), 7.84 – 7.78 (m, 1H), 7.72 (d, $J = 8.3$ Hz, 1H), 7.55 (d, $J = 9.2$ Hz, 1H), 7.49 (s, 1H), 6.85 (s, 2H), 5.27 (s, 1H), 5.17 (s, 1H), 4.46 (d, $J = 14.9$ Hz, 2H), 4.19 (d, $J = 9.0$ Hz, 1H), 3.90 (d, $J = 10.4$ Hz, 2H), 3.81 (d, $J = 12.1$ Hz, 2H), 3.73 (t, $J = 8.5$ Hz, 2H), 3.61 (d, $J = 9.5$ Hz, 1H), 3.54 – 3.44 (m, 6H), 1.27 – 1.21 (m, 6H). HRMS (ESI) m/z : $[M + H]^+$ Calcd for $C_{33}H_{40}NO_{14}$ 674.2443; Found 674.2440.

3-HC-2-Glc. 3-HC-2 (10 mg, 0.031 mmol) yielded 0.845 mg (5.2% over three steps) of an orange powder. 1H NMR (500 MHz, Methanol- d_4) δ 8.19 (d, $J = 8.0$ Hz, 1H), 8.01 (d, $J = 2.0$ Hz, 1H), 7.81 (dd, $J = 9.5, 7.5$ Hz, 1H), 7.71 (d, $J = 8.6$ Hz, 1H), 7.53 – 7.47 (m, 3H), 7.29 (d, $J = 3.6$ Hz, 2H), 6.85 (d, $J = 4.8$ Hz, 2H), 4.80 (d, $J = 3.8$ Hz, 1H), 4.52 (d, $J = 11.1$ Hz, 1H), 4.47 (d, $J = 5.0$ Hz, 1H), 3.89 (s, 1H), 3.72 (t, $J = 9.3$ Hz, 1H), 3.61 (t, $J = 9.5$ Hz, 1H), 3.54 – 3.46 (m, 5H), 3.42 (s, 3H), 1.24 (t, $J = 7.0$ Hz, 6H). HRMS (ESI) m/z : $[M + H]^+$ Calcd for $C_{28}H_{32}NO_9$ 526.2072; Found 526.2070.

EtNR-Tre. EtNR (10.6 mg, 0.032 mmol) yielded 4.3 mg (62.3% over three steps) of a dark purple fluffy solid. NMR signal was too weak to report here. HRMS (ESI) m/z : $[M + H]^+$ Calcd for $C_{32}H_{39}N_2O_{13}$ 659.2447; Found 659.2452.

EtNR-Glc. EtNR (10.2 mg, 0.03) yielded 2.228 mg (18% over three steps) of a dark purple powder. NMR for peracetylated intermediate: 1H NMR (500 MHz, Chloroform- d) δ 8.21 (d, $J = 8.7$ Hz, 1H), 8.03 (d, $J = 2.6$ Hz, 1H), 7.60 (d, $J = 9.0$ Hz, 1H), 7.18 (dd, $J = 8.7, 2.6$ Hz, 1H), 6.65 (dd, $J = 9.1, 2.7$ Hz, 1H), 6.45 (d, $J = 2.7$ Hz, 1H), 6.30 (s, 1H), 5.56 (dd,

J = 10.2, 9.3 Hz, 1H), 5.26 (t, J = 9.6 Hz, 1H), 5.01 (d, J = 3.7 Hz, 1H), 4.97 (dd, J = 10.2, 3.7 Hz, 1H), 4.31 – 4.20 (m, 3H), 3.46 (d, J = 7.9 Hz, 7H), 2.09 (s, 3H), 2.05 (s, 3H), 2.04 (s, 3H), 1.26 (t, J = 7.1 Hz, 7H).

MeNR-Tre. MeNR (5 mg, 0.016 mmol) yielded 5.133 mg (63.1% over three steps; 2.2 mg peracetylated intermediate had been set aside) of a dark maroon/purple powder. ¹H NMR (500 MHz, Methanol-*d*₄) δ 8.12 (dd, J = 5.7, 3.1 Hz, 2H), 7.63 (d, J = 9.1 Hz, 1H), 7.26 (dd, J = 8.8, 2.6 Hz, 1H), 6.84 (dd, J = 9.2, 2.6 Hz, 1H), 6.61 (d, J = 2.7 Hz, 1H), 6.25 (s, 1H), 5.19 (t, J = 3.8 Hz, 2H), 4.44 (d, J = 9.0 Hz, 1H), 4.40 (d, J = 5.0 Hz, 1H), 4.26 (d, J = 10.4 Hz, 1H), 3.91 – 3.84 (m, 2H), 3.83 (s, 1H), 3.80 (d, J = 3.1 Hz, 1H), 3.69 (dd, J = 11.8, 5.4 Hz, 1H), 3.62 – 3.57 (m, 2H), 3.52 (dd, J = 9.7, 3.7 Hz, 1H), 3.35 (d, J = 9.5 Hz, 1H), 3.14 (s, 6H). HRMS (ESI) *m/z*: [M + H]⁺ Calcd for C₃₀H₃₅N₂O₁₃ 631.2134; Found 631.2135.

MeNR-Glc. MeNR (5 mg, 0.016 mmol) yielded 2.834 mg (36.1% over three steps) of a dark maroon/purple powder. ¹H NMR (500 MHz, Methanol-*d*₄) δ 7.99 (d, J = 8.7 Hz, 1H), 7.86 (d, J = 2.6 Hz, 1H), 7.36 (d, J = 9.0 Hz, 1H), 7.14 (dd, J = 8.7, 2.6 Hz, 1H), 6.58 (dd, J = 9.1, 2.7 Hz, 1H), 6.30 (d, J = 2.5 Hz, 1H), 6.07 (s, 1H), 4.75 (d, J = 3.8 Hz, 1H), 4.44 – 4.41 (m, 1H), 4.30 (dd, J = 10.2, 5.7 Hz, 1H), 3.96 – 3.91 (m, 1H), 3.75 – 3.69 (m, 1H), 3.56 – 3.50 (m, 2H), 3.49 (s, 3H), 3.01 (s, 6H).

DMN-Tre, DMN-Glc, 6-FITre, and 6-TMR-Tre were on hand from previous studies.^{35-36,}
38

REFERENCES

1. *Global tuberculosis report 2016*; World Health Organization: Geneva, Switzerland, 2016.
2. Dutta, N. K.; Karakousis, P. C. Latent Tuberculosis Infection: Myths, Models, and Molecular Mechanisms. *Microbiol. Mol. Biol. Rev.* **2014**, *78*, 343-371.
3. Gomez, J. E.; McKinney, J. D. *M. tuberculosis* persistence, latency, and drug tolerance. *Tuberculosis* **2004**, *84*, 29-44.
4. Zhang, Y. The magic bullets and tuberculosis drug targets. *Annu. Rev. Pharmacol. Toxicol.* **2005**, *45*, 529-564.
5. Sulis, G.; Roggi, A.; Matteelli, A.; Raviglione, M. C. Tuberculosis: Epidemiology and Control. *Mediterr. J. Hematol. Infect. Dis.* **2014**, *6*, e2014070.
6. Zignol, M.; Dean, A. S.; Falzon, D.; van Gemert, W.; Wright, A.; van Deun, A.; Portaels, F.; Laszlo, A.; Espinal, M. A.; Pablos-Méndez, A.; Bloom, A.; Aziz, M. A.; Weyer, K.; Jaramillo, E.; Nunn, P.; Floyd, K.; Raviglione, M. C. Twenty Years of Global Surveillance of Antituberculosis-Drug Resistance. *N. Engl. J. Med.* **2016**, *375*, 1081-1089.
7. Multidrug-Resistant Tuberculosis (MDR-TB) 2016 Update. World Health Organization: Geneva, Switzerland, 2016.
http://www.who.int/tb/challenges/mdr/mdr_tb_factsheet.pdf (accessed October 20, 2016).
8. Marks, S. M.; Flood, J.; Seaworth, B.; Hirsch-Moverman, Y.; Armstrong, L.; Mase, S.; Salcedo, K.; Oh, P.; Graviss, E. A.; Colson, P. W.; Armitage, L.; Revuelta, M.; Sheeran, K.; the TB Epidemiologic Studies Consortium. Treatment Practices, Outcomes, and Costs of Multidrug-Resistant and Extensively Drug-Resistant Tuberculosis, United States, 2005–2007. *Emerg. Infect. Dis.* **2014**, *20*, 812-821.
9. L. M. Parsons, Á. Somoskövi, C. Gutierrez, E. Lee, C. N. Paramasivan, A. Abimiku, S. Spector, G. Roscigno, J. Nkengasong, Laboratory diagnosis of tuberculosis in resource-poor countries: Challenges and opportunities. *Clin. Microbiol. Rev.* *24*, 314–350 (2011).
10. World Health Organization (WHO), “High-priority target product profiles for new tuberculosis diagnostics: Report of a consensus meeting” (WHO Press, 2014); http://apps.who.int/iris/bitstream/10665/135617/1/WHO_HTM_TB_2014.18_eng.pdf.
11. World Health Organization (WHO), “Molecular line probe assays for rapid screening of patients at risk of multidrug-resistant tuberculosis (MDR-TB)” (WHO Press, 2008) www.who.int/tb/features_archive/policy_statement.pdf.
12. Ziehl, F. Zur farbung des tuberkelbacillus. *Dtsch. Med. Wschr.* **1882**, *8*, 451.
13. Ziehl, F. Ueber die farbung des tuberkelbacillus. *Dtsch. Med. Wschr.* **1883**, *9*, 247–249.

14. Neelsen, F. Ein casuistischer beitrage zur lehre von der tuberkulose. *Centralbl. Med. Wissensch.* **1883**, *28*, 497–501 (1883).
15. Hagemann, P. K. H. Fluoreszenzfärbung von tuberkelbakterien mit auramin. *Münch. Med. Wschr.* **1938**, *85*, 1066–1068.
16. Truant, J. P.; Brett, W. A.; Thomas, W. Jr. Fluorescence microscopy of tubercle bacilli stained with auramine and rhodamine. *Henry Ford Hosp. Med. Bull.* **1962**, *10*, 287–296.
17. Boyd, J. C.; Marr, J. J. Decreasing reliability of acid-fast smear techniques for detection of tuberculosis. *Ann. Intern. Med.* **1975**, *82*, 489–492.
18. Filho, C. F. F. C.; Costa, M. G. F. Sputum smear microscopy for tuberculosis: Evaluation of autofocus functions and automatic identification of tuberculosis mycobacterium; In *Understanding Tuberculosis—Global Experiences and Innovative Approaches to the Diagnosis*; Cardona, P.J., Ed.; InTech, 2012; Ch. 13.
19. Singhal, R.; Myneedu, V. P. Microscopy as a diagnostic tool in pulmonary tuberculosis. *Int. J. Mycobacteriol.* **2015**, *4*, 1–6.
20. Timmins, G. S.; Deretic, V. Mechanisms of action of isoniazid. *Mol. Microbiol.* **2006**, *62*, 1220–1227.
21. Adams, K. N.; Takaki, K.; Connolly, L. E.; Wiedenhoft, H.; Winglee, K.; Humbert, O.; Edelstein, P. H.; Cosma, C. L.; Ramakrishnan, L. Drug tolerance in replicating mycobacteria mediated by a macrophage-induced efflux mechanism. *Cell* **2011**, *145*, 39.
22. Baek, S. H.; Li, A. H.; Sasseti, C. M. Metabolic regulation of mycobacterial growth and antibiotic sensitivity. *PLOS Biol.* **2011**, *9*, e1001065.
23. Kalscheuer, R.; Koliwer-Brandl, H. Genetics of Mycobacterial Trehalose Metabolism. *Microbiol. Spectrum* **2014**, *2*:MGM2-0002-2013.
24. Kalscheuer, R.; Weinrick, B.; Veeraraghavan, U.; Besra, G. S.; Jacobs, W. R., Jr. Trehalose-recycling ABC transporter LpqY-SugA-SugB-SugC is essential for virulence of *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci.* **2010**, *107*, 21761–21766.
25. Grzegorzewicz, A. E.; Pham, H.; Gundi, V. A.; Scherman, M. S.; North, E. J.; Hess, T.; Jones, V.; Gruppo, V.; Born, S. E.; Kordulakova, J.; Chavadi, S. S.; Morisseau, C.; Lenaerts, A. J.; Lee, R. E.; McNeil, M. R.; Jackson, M. Inhibition of mycolic acid transport across the *Mycobacterium tuberculosis* plasma membrane. *Nat. Chem. Biol.* **2012**, *8*, 334.
26. Xu, A.; Meshcheryakov, V. A.; Poce, G.; Chng, S.-S. MmpL3 is the flippase for mycolic acids in mycobacteria. *Proc. Natl. Acad. Sci.* **2017**, *114*, 7993–7998.
27. Tahlan, K.; Wilson, R.; Kastrinsky, D. B.; Arora, K.; Nair, V.; Fischer, E.; Barnes, S. W.; Walker, J. R.; Alland, D.; Barry, C. E., III; Boshoff, H. I. SQ109 targets MmpL3, a

- membrane transporter of trehalose monomycolate involved in mycolic acid donation to the cell wall core of *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemother.* **2012**, *56*, 1797.
28. Sathyamoorthy, N.; Takayama, K. Purification and characterization of a novel mycolic acid exchange enzyme from *Mycobacterium smegmatis*. *J. Biol. Chem.* **1987**, *262*, 13417.
 29. Belisle, J. T.; Vissa, V. D.; Sievert, T.; Takayama, K.; Brennan, P. J.; Besra, G. S. Role of the major antigen of *Mycobacterium tuberculosis* in cell wall biogenesis. *Science* **1997**, *276*, 1420–1422.
 30. Nobre, A.; Alarico, S.; Maranha, A.; Mendes, V.; Empadinhas, N. The molecular biology of mycobacterial trehalose in the quest for advanced tuberculosis therapies. *Microbiology* **2014**, *160*, 1547-1570.
 31. Swarts, B. M.; Holsclaw, C. M.; Jewett, J. C.; Alber, M.; Fox, D. M.; Siegrist, M. S.; Leary, J. A.; Kalscheuer, R.; Bertozzi, C. R. Probing the Mycobacterial Trehalome with Bioorthogonal Chemistry. *J. Am. Chem. Soc.* **2012**, *134*, 16123-16126.
 32. Foley, H. N.; Stewart, J. A.; Kavunja, H. W.; Rundell, S. R.; Swarts, B. M. Bioorthogonal Chemical Reporters for Selective In Situ Probing of Mycomembrane Components in Mycobacteria. *Angew. Chem. Int. Ed.* **2016**, *55*, 2053–2057.
 33. Rundell, S. R.; Wagar, Z. L.; Meints, L. M.; Olson, C. D.; O'Neill, M. K.; Piligian, B. F.; Poston, A. W.; Hood, R. J.; Woodruff, P. J.; Swarts, B. M. Deoxyfluoro-D-trehalose (FDTre) analogues as potential PET probes for imaging mycobacterial infection. *Org. Biomol. Chem.* **2016**, *14*, 8598-8609.
 34. Backus, K. M.; Boshoff, H. I.; Barry, C. S.; Boutureira, O.; Patel, M. K.; D'Hooge, F.; Lee, S. S.; Via, L. E.; Tahlan, K.; Barry, C. E. III.; Davis, B. G. Uptake of unnatural trehalose analogs as a reporter for *Mycobacterium tuberculosis*. *Nat. Chem. Biol.* **2011**, *7*, 228-235.
 35. Rodríguez-Rivera, F. P.; Zhou, X.; Theriot, J. A.; Bertozzi, C. R. Visualization of mycobacterial membrane dynamics in live cells. *J. Am. Chem. Soc.* **2017**, *139*, 3488-3495.
 36. Rodríguez-Rivera, F. P.; Zhou, X.; Theriot, J. A.; Bertozzi, C. R. Acute modulation of mycobacterial cell envelope biogenesis by front-line TB drugs. *Angew. Chem. Int. Ed.* **2018**, *57*, 5267-5272.
 37. Hodges, H. L.; Brown, R. A.; Crooks, J. A.; Weibel, D. B.; Kiessling, L. L. Imaging mycobacterial growth and division with a fluorogenic probe. *Proc. Natl. Acad. Sci.* **2018**. 201720996; published ahead of print April 27, 2018. <https://doi.org/10.1073/pnas.1720996115>.
 38. Kamariza, M.*; Shieh, P.*; Ealand, C. S.; Peters, J. S.; Chu, B.; Rodríguez-Rivera, F. P.; Babu Sait, M. R.; Treuren, W. V.; Martinson, N.; Kalscheuer, R.; Kana, B. D.;

- Bertozzi, C. R. Rapid detection of *Mycobacterium tuberculosis* in sputum with a solvatochromic trehalose probe. *Sci. Transl. Med.* **2018**, *10*, eaam6310.
39. Klymchenko, A. S.; Mély, Y. Fluorescent Environment-Sensitive Dyes as Reporters of Biomolecular Interactions. *Prog. Mol. Biol. Transl. Sci.* **2013**, *113*, 35-58
 40. Loving, G. S.; Sainlos, M.; Imperiali, B. Monitoring Protein Interactions and Dynamics with Solvatochromic Fluorophores. *Trends Biotechnol.* **2010**, *28*, 73-83.
 41. Reichardt, C. Solvatochromic Dyes as Solvent Polarity Indicators. *Chem. Rev.* **1994**, *94*, 2319-2358.
 42. Klymchenko, A. S. Solvatochromic and Fluorogenic Dyes as Environment-Sensitive Probes: Design and Biological Applications. *Acc. Chem. Res.* **2017**, *50*, 366-375.
 43. Loving, G.; Imperiali, B. A Versatile Amino Acid Analogue of the Solvatochromic Fluorophore 4-*N,N*-Dimethylamino-1,8-naphthalimide: A Powerful Tool for the Study of Dynamic Protein Interactions. *J. Am. Chem. Soc.* **2008**, *130*, 13630-13638.
 44. Maity, J.; Honcharenko, D.; Strömberg, R. Synthesis of fluorescent D-amino acids with 4-acetamidobiphenyl and 4-*N,N*-dimethylamino-1,8-naphthalimido containing side chains. *Tetrahedron Lett.* **2015**, *56*, 4780-4783.
 45. Nakanishi, J.; Nakajima, T.; Sato, M.; Ozawa, T.; Tohda, K.; Umezawa, Y. Imaging of Conformational Changes of Proteins with a New Environment-Sensitive Fluorescent Probe Designed for Site-Specific Labeling of Recombinant Proteins in Live Cells. *Anal. Chem.* **2001**, *73*, 2920-2928.
 46. Bachmann, L.; Zezell, D. M.; da Costa Ribeiro, A.; Gomez, L.; Ito, A. S. Fluorescence Spectroscopy of Biological Tissues—A Review. *Appl. Spectrosc. Rev.* **2006**, *41*, 575-590.
 47. Greenspan, P.; Mayer, E. P.; Fowler, S. D. Nile Red: A Selective Fluorescent Stain for Intracellular Lipid Droplets. *J. Cell Biol.* **1985**, *100*, 965-973.
 48. Martinez, V.; Henary, M. Nile Red and Nile Blue: Applications and Syntheses of Structural Analogues. *Chem. Eur. J.* **2016**, *22*, 13764-13782.
 49. Briggs, M. S. J.; Bruce, I.; Miller, J. N.; Moody, C. J.; Simmonds, A. C.; Swann, E. Synthesis of functionalised fluorescent dyes and their coupling to amines and amino acids. *J. Chem. Soc. Perkin Trans. 1* **1997**, 1051-1058.
 50. Kucherak, O. A.; Oncul, S.; Darwich, Z.; Yushchenko, D. A.; Arntz, Y.; Didier, P.; Mély, Y.; Klymchenko, A. S. Switchable Nile Red-Based Probe for Cholesterol and Lipid Order at the Outer Leaflet of Biomembranes. *J. Am. Chem. Soc.* **2010**, *132*, 4907-4916.
 51. Hanessian, S.; Lavaller, P. Synthesis of 6-amino-6-deoxy- α,α -trehalose: A positional isomer of trehalosamine. *J. Antibiot.* **1972**, *25*, 683-684.
 52. Gunduz, S.; Goren, A. C.; Ozturk, T. Facile Syntheses of 3-Hydroxyflavones. *Org. Lett.* **2012**, *14*, 1576-1579.

53. Klymchenko, A. S.; Pivovarenko, V. G.; Ozturk, T.; Demchenko, A. P. Modulation of the solvent-dependent dual emission in 3-hydroxychromones by substituents. *New J. Chem.* **2003**, *27*, 1336-1343.
54. Kucherak, O. A.; Richert, L.; Mély, Y.; Klymchenko, A. S. Dipolar 3-methoxychromones as bright and highly solvatochromic fluorescent dyes. *Phys. Chem. Chem. Phys.* **2012**, *14*, 2292-2300.
55. Kucherak, O. A.; Didier, P.; Mély, Y.; Klymchenko, A. S. Fluorene Analogues of Prodan with Superior Fluorescence Brightness and Solvatochromism. *J. Phys. Chem. Lett.* **2010**, *1*, 616-620.
56. Klymchenko, A. S.; Ozturk, T.; Pivovarenko, V. G.; Demchenko, A. P. A 3-hydroxychromone with dramatically improved fluorescence properties. *Tetrahedron Lett.* **2001**, *42*, 7967-7970.
57. Junior, C. O. R.; Castro, S. B. R.; Pereira, A. A.; Alves, C. C. S.; Oliveira, E. E.; Rêgo, R. T.; Ferreira, A. P.; de Almeida, M. V. Synthesis of genistein coupled with sugar derivatives and their inhibitory effect on nitric oxide production in macrophages. *Eur. J. Med. Chem.* **2014**, *85*, 615-620.
58. Matsunaga, I.; Naka, T.; Talekar, R. S.; McConnell, M. J.; Katoh, K.; Nakao, H.; Otsuka, A.; Behar, S. M.; Yano, I.; Moody, D. B.; Sugita, M. Mycolyltransferase-mediated Glycolipid Exchange in Mycobacteria. *J. Biol. Chem.* **2008**, *283*, 28835-28841.
59. Zief, M.; Hockett, R. C. Methyl 6-Iodo-6-desoxy- α -D-glucopyranoside. *J. Am. Chem. Soc.* **1945**, *67*, 1267-1268.
60. Martin, M. M. Hydrogen bond effects on radiationless electron transitions in xanthene dyes. *Chem. Phys. Lett.* **1975**, *35*, 105-111.
61. Favrot, L.; Grzegorzewicz, A. E.; Lajiness, D. H.; Marvin, R. K.; Boucau, J.; Isailovic, D.; Jackson, M.; Ronning, D. R. Mechanism of inhibition of *Mycobacterium tuberculosis* antigen 85 by ebselen. *Nat. Commun.* **2013**, *4*, 2748.
62. Belardinelli, J. M.; Yazidi, A.; Yang, L.; Fabre, L.; Li, W.; Jacques, B.; Angala, S. K.; Rouiller, I.; Zgurskaya, H. I.; Sygusch, J.; Jackson, M. Structure–Function Profile of MmpL3, the Essential Mycolic Acid Transporter from *Mycobacterium tuberculosis*. *ACS Infect. Dis.* **2016**, *2*, 702-713.
63. van Leeuwen, L. M.; van der Sar, A. M.; Bitter, W. Animal Models of Tuberculosis: Zebrafish. *Cold Spring Harb. Perspect. Med.* **2015**, *5*: a018580.
64. Swaim, L. E.; Connolly, L. E.; Volkman, H. E.; Humbert, O.; Born, D. E.; Ramakrishnan, L. *Mycobacterium marinum* Infection of Adult Zebrafish Causes Caseating Granulomatous Tuberculosis and Is Moderated by Adaptive Immunity. *Infect. Immun.* **2006**, *74*, 6108-6117.
65. Cambier, C. J.; Takaki, K. K.; Larson, R. P.; Hernandez, R. E.; Tobin, D. M.; Urdahl, K. B.; Cosma, C. L.; Ramakrishnan, L. Mycobacteria manipulate macrophage

- recruitment through coordinated use of membrane lipids. *Nature* **2014**, *505*, 218-222.
66. Gahlmann, A.; Moerner, W. E. Exploring bacterial cell biology with single-molecule tracking and super-resolution imaging. *Nat. Rev. Microbiol.* **2014**, *12*, 9–22.
67. Tortoli, E.; Cichero, P.; Piersimoni, C.; Simonetti, M. T.; Gesu, G.; Nista, D. Use of BACTEC MGIT 960 for Recovery of Mycobacteria from Clinical Specimens: Multicenter Study. *J. Clin. Microbiol.* **1999**, *37*, 3578-3582.
68. Demangel, C.; Stinear, T. P.; Cole, S. T. Buruli ulcer: reductive evolution enhances pathogenicity of *Mycobacterium ulcerans*. *Nat. Rev. Microbiol.* **2009**, *7*, 50-60.
69. Duker, A. A.; Portaels, F.; Hale, M. Pathways of *Mycobacterium ulcerans* infection: A review. *Env. Int.* **2006**, *32*, 567-573.
70. “Buruli ulcer (*Mycobacterium ulcerans* infection)” [Fact Sheet]. *World Health Organization*. April 2018. [http://www.who.int/news-room/fact-sheets/detail/buruli-ulcer-\(mycobacterium-ulcerans-infection\)](http://www.who.int/news-room/fact-sheets/detail/buruli-ulcer-(mycobacterium-ulcerans-infection)) (accessed May 9, 2018).
71. Herbinger, K.-H.; Adjei, O.; Awua-Boateng, N.-Y.; Nienhuis, W. A.; Kunaa, L.; Siegmund, V.; Nitschke, J.; Thompson, W.; Klutse, E.; Agbenorku, P.; Schipf, A.; Reu, S.; Racz, P.; Fleischer, B.; Beissner, M.; Fleischmann, E.; Helfrich, K.; van der Werf, T. S.; Lüscher, T.; Bretzel, G. Comparative study of the sensitivity of different diagnostic methods for the laboratory diagnosis of Buruli ulcer disease. *Clin. Infect. Dis.* **2009**, *48*, 1055–1064.
72. Schindelin, J.; Arganda-Carreras, I.; Frise, E.; Kaynig, V.; Longair, M.; Pietzsch, T.; Preibisch, S.; Rueden, C.; Saalfeld, S.; Schmid, B.; Tinevez, J.-Y.; White, D. J.; Hartenstein, V.; Eliceiri, K.; Tomancak, P.; Cardona, A. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **2012**, *9*, 676-682.

Appendix 1.

LC-MS time-course spectra and integration results (Chapter 2)

Peak identification based on Supplementary Figures 2.S2-2.S3:

CBTag 1.0 – minute 7.9

CBTag 1.0-13CBTNAc, lysine adduct – minute 9

CBTag 1.0-13CBTNAc, cysteine adduct – minute 9.2

CBTag 1.0-2(13CBTNAc) – minute 9.7

13CBTNAc – minute 12.9

Integration results were consolidated previously in Supplementary Table 2.S2-2.S4.

Data File C:\CHEM32\1\DATA\2018 TO MAR 07\012618\012418_PF3-PH7-13CBTNAC_1.D
Sample Name: 012218_PF3-ph7-13cbtnac_1

=====
Integration Results
=====

Signal 1: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\012618\012418_PF3-PH7-13CBTNAC_1.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.854 | FM | 1.90725e4 | 2084.91626 | 0.1525 | 7.697 | 8.004 |
| 2 | 9.061 | MF | 875.47845 | 209.35365 | 0.0697 | 8.970 | 9.065 |
| 3 | 9.195 | MF | 1.81111e4 | 2078.43286 | 0.1452 | 9.065 | 9.298 |
| 4 | 9.711 | FM | 2091.56250 | 275.67270 | 0.1265 | 9.644 | 9.877 |
| 5 | 12.882 | MF | 1.77146e4 | 2086.62402 | 0.1415 | 12.720 | 12.976 |

Signal 2: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\012618\012418_PF3-PH7-13CBTNAC_2.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.894 | FM | 2.17422e4 | 2129.17847 | 0.1702 | 7.699 | 8.060 |
| 2 | 9.058 | FM | 2265.41406 | 433.72296 | 0.0871 | 8.952 | 9.098 |
| 3 | 9.232 | MF | 1.81665e4 | 2077.62915 | 0.1457 | 9.098 | 9.349 |
| 4 | 9.728 | FM | 6531.01660 | 759.14935 | 0.1434 | 9.640 | 9.934 |
| 5 | 12.956 | MF | 1.59388e4 | 2136.41089 | 0.1243 | 12.824 | 13.046 |

Signal 3: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\012618\012418_PF3-PH7-13CBTNAC_3.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.416 | FM | 3.00385e4 | 1894.61292 | 0.2642 | 7.077 | 7.685 |
| 2 | 8.906 | MF | 2944.28687 | 611.75421 | 0.0802 | 8.846 | 8.999 |
| 3 | 9.129 | MF | 1.37661e4 | 1695.87073 | 0.1353 | 8.999 | 9.250 |
| 4 | 9.662 | MF | 8471.92676 | 1076.19922 | 0.1312 | 9.538 | 9.835 |
| 5 | 12.802 | MF | 1.34734e4 | 1632.25354 | 0.1376 | 12.676 | 12.937 |

Signal 4: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\012618\012418_PF3-PH7-13CBTNAC_4.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.884 | MF | 1.49431e4 | 1688.32275 | 0.1475 | 7.747 | 8.001 |
| 2 | 9.010 | FM | 3431.95166 | 649.64691 | 0.0880 | 8.979 | 9.113 |
| 3 | 9.231 | FM | 1.33222e4 | 1571.53259 | 0.1413 | 9.113 | 9.360 |
| 4 | 9.729 | FM | 1.02633e4 | 1182.65515 | 0.1446 | 9.627 | 9.932 |
| 5 | 12.922 | MF | 1.41186e4 | 1700.74683 | 0.1384 | 12.764 | 13.027 |

Signal 5: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\012618\012418_PF3-PH7-13CBTNAC_5.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.844 | MF | 1.13900e4 | 1649.19177 | 0.1151 | 7.726 | 7.969 |
| 2 | 9.010 | MF | 2775.56274 | 557.69238 | 0.0829 | 8.951 | 9.081 |
| 3 | 9.193 | MF | 8668.32715 | 1219.24573 | 0.1185 | 9.081 | 9.297 |
| 4 | 9.696 | MF | 8322.69238 | 1001.90900 | 0.1384 | 9.578 | 9.886 |
| 5 | 12.800 | MF | 1.29913e4 | 1721.24487 | 0.1258 | 12.678 | 12.910 |

Data File C:\CHEM32\1\DATA\2018 TO MAR 07\012618\012418_PF3-PH7-13CBTNAC_1.D
Sample Name: 012218_PF3-ph7-13cbtnac_1

=====
Integration Results
=====

Signal 1: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\012618\012418_PF3-PH7-13CBTNAC_1.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 9.081 | MF | 455.95853 | 76.43199 | 0.0994 | 8.948 | 9.085 |
| 2 | 9.195 | MF | 4273.00830 | 647.51642 | 0.1100 | 9.085 | 9.285 |
| 3 | 9.713 | FM | 864.22162 | 103.05989 | 0.1398 | 9.631 | 9.836 |
| 4 | 12.880 | MF | 1.54454e4 | 2768.33911 | 0.0930 | 12.757 | 12.973 |
| 5 | 12.973 | FM | 383.86099 | 201.20613 | 0.0318 | 12.973 | 13.073 |

Signal 2: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\012618\012418_PF3-PH7-13CBTNAC_2.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 9.057 | MF | 416.94113 | 78.41770 | 0.0886 | 8.983 | 9.115 |
| 2 | 9.232 | MF | 3347.27686 | 497.54984 | 0.1121 | 9.115 | 9.320 |
| 3 | 9.728 | MF | 1268.26978 | 193.05396 | 0.1095 | 9.657 | 9.807 |
| 4 | 12.956 | MM | 1.15061e4 | 2109.27319 | 0.0909 | 12.842 | 13.088 |

Signal 3: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\012618\012418_PF3-PH7-13CBTNAC_3.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 8.906 | MF | 851.79407 | 162.02739 | 0.0876 | 8.852 | 8.984 |
| 2 | 9.129 | FM | 3439.48486 | 476.80533 | 0.1202 | 8.984 | 9.230 |
| 3 | 9.662 | FM | 2666.17871 | 344.44800 | 0.1290 | 9.541 | 9.762 |
| 4 | 12.803 | MM | 1.12936e4 | 2014.62061 | 0.0934 | 12.677 | 12.918 |

Signal 4: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\012618\012418_PF3-PH7-13CBTNAC_4.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 9.046 | MF | 906.43396 | 164.82642 | 0.0917 | 8.992 | 9.120 |
| 2 | 9.231 | MF | 2879.00098 | 392.18085 | 0.1224 | 9.120 | 9.325 |
| 3 | 9.730 | MF | 2489.67114 | 360.50821 | 0.1151 | 9.647 | 9.817 |
| 4 | 12.924 | MM | 1.27247e4 | 2235.55273 | 0.0949 | 12.818 | 13.074 |

Signal 5: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\012618\012418_PF3-PH7-13CBTNAC_5.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 9.011 | MF | 638.57806 | 124.45227 | 0.0855 | 8.962 | 9.095 |
| 2 | 9.193 | FM | 1658.48438 | 245.42241 | 0.1126 | 9.095 | 9.280 |
| 3 | 9.696 | FM | 2168.59033 | 286.15845 | 0.1263 | 9.591 | 9.827 |
| 4 | 12.800 | MM | 9266.77344 | 1678.98987 | 0.0920 | 12.687 | 12.978 |

Print of window 38: Current Chromatogram(s)

Data File : C:\CHEM32\1\DATA\2018 TO MAR 07\020218\013118_PF3-PH7-13CBTNAC_1.D

Sample Name : 013118_PF3-ph7-13cbtnac_1

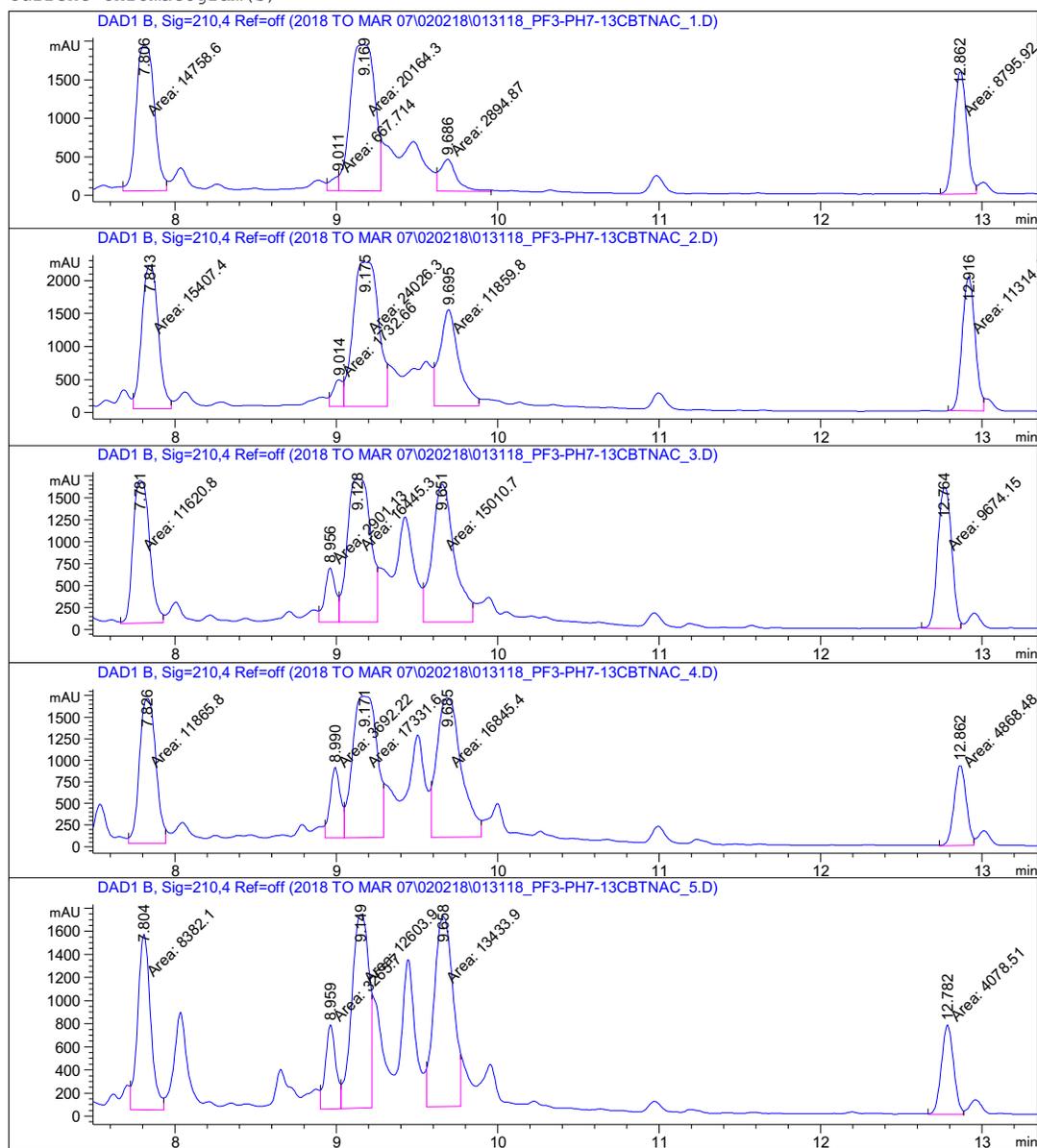
=====
Acq. Operator : SGLK Seq. Line : 5
Acq. Instrument : Instrument 1 Location : Vial 5
Injection Date : 2/2/18 6:47:32 PM Inj : 1
Inj Volume : 1.000 µl

Acq. Method : D:\CHEM32\1\METHODS\AP_C18_INSERT.M
Last changed : 1/22/18 3:04:38 PM by SGLK
Analysis Method : Z:\DESKTOP\BERTOZZI LAB\LCMS INFO\LCMS METHODS TO 112917\SGLK_BERTOZZI_BP_WITHINSERT.M

Last changed : 3/8/18 8:51:08 AM
(modified after loading)

Additional Info : Peak(s) manually integrated

Current Chromatogram(s)



Data File C:\CHEM32\1\DATA\2018 TO MAR 07\020218\013118_PF3-PH7-13CBTNAC_1.D
Sample Name: 013118_PF3-ph7-13cbtnac_1

=====
Integration Results
=====

Signal 1: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\020218\013118_PF3-PH7-13CBTNAC_1.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.806 | MF | 1.47586e4 | 1877.88623 | 0.1310 | 7.677 | 7.947 |
| 2 | 9.011 | MF | 667.71442 | 192.94965 | 0.0577 | 8.939 | 9.012 |
| 3 | 9.169 | MF | 2.01643e4 | 1890.88965 | 0.1777 | 9.012 | 9.273 |
| 4 | 9.686 | FM | 2894.87231 | 414.46405 | 0.1164 | 9.622 | 9.956 |
| 5 | 12.862 | MF | 8795.91797 | 1592.83997 | 0.0920 | 12.740 | 12.963 |

Signal 2: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\020218\013118_PF3-PH7-13CBTNAC_2.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.843 | FM | 1.54074e4 | 2173.26514 | 0.1182 | 7.740 | 7.976 |
| 2 | 9.014 | MF | 1732.66248 | 402.80444 | 0.0717 | 8.955 | 9.044 |
| 3 | 9.175 | MF | 2.40263e4 | 2183.97681 | 0.1834 | 9.044 | 9.314 |
| 4 | 9.695 | MF | 1.18598e4 | 1464.29858 | 0.1350 | 9.603 | 9.883 |
| 5 | 12.916 | MF | 1.13144e4 | 2036.28650 | 0.0926 | 12.788 | 13.007 |

Signal 3: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\020218\013118_PF3-PH7-13CBTNAC_3.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.781 | MF | 1.16208e4 | 1624.52429 | 0.1192 | 7.661 | 7.925 |
| 2 | 8.956 | MF | 2901.13184 | 622.10889 | 0.0777 | 8.891 | 9.015 |
| 3 | 9.128 | MF | 1.64453e4 | 1625.91223 | 0.1686 | 9.015 | 9.254 |
| 4 | 9.651 | MF | 1.50107e4 | 1571.71680 | 0.1592 | 9.536 | 9.845 |
| 5 | 12.764 | MF | 9674.14648 | 1609.95398 | 0.1001 | 12.623 | 12.868 |

Signal 4: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\020218\013118_PF3-PH7-13CBTNAC_4.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.826 | FM | 1.18658e4 | 1681.22424 | 0.1176 | 7.712 | 7.941 |
| 2 | 8.990 | MF | 3692.21680 | 820.23694 | 0.0750 | 8.929 | 9.047 |
| 3 | 9.171 | MF | 1.73316e4 | 1636.02942 | 0.1766 | 9.047 | 9.292 |
| 4 | 9.685 | MF | 1.68454e4 | 1617.40186 | 0.1736 | 9.587 | 9.896 |
| 5 | 12.862 | MF | 4868.47998 | 934.37787 | 0.0868 | 12.734 | 12.947 |

Signal 5: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\020218\013118_PF3-PH7-13CBTNAC_5.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.804 | FM | 8382.09668 | 1519.23242 | 0.0920 | 7.725 | 7.928 |
| 2 | 8.959 | MF | 3265.69629 | 730.16943 | 0.0745 | 8.901 | 9.028 |
| 3 | 9.149 | MF | 1.26039e4 | 1678.56458 | 0.1251 | 9.028 | 9.219 |
| 4 | 9.658 | MF | 1.34339e4 | 1646.90576 | 0.1360 | 9.559 | 9.769 |
| 5 | 12.782 | MF | 4078.51001 | 775.32263 | 0.0877 | 12.664 | 12.883 |

Data File C:\CHEM32\1\DATA\2018 TO MAR 07\020218\013118_PF3-PH7-13CBTNAC_1.D
Sample Name: 013118_PF3-ph7-13cbtnac_1

=====
Integration Results
=====

Signal 1: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\020218\013118_PF3-PH7-13CBTNAC_1.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 8.980 | MF | 167.63275 | 40.41625 | 0.0691 | 8.935 | 9.025 |
| 2 | 9.158 | MF | 6653.14258 | 1013.87598 | 0.1094 | 9.025 | 9.279 |
| 3 | 9.688 | FM | 1015.97882 | 131.04523 | 0.1292 | 9.622 | 10.154 |
| 4 | 12.862 | MF | 4049.37402 | 772.72168 | 0.0873 | 12.714 | 12.963 |

Signal 2: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\020218\013118_PF3-PH7-13CBTNAC_2.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 9.012 | FM | 449.12659 | 88.06937 | 0.0850 | 8.959 | 9.058 |
| 2 | 9.181 | FM | 6981.57227 | 1022.56451 | 0.1138 | 9.058 | 9.295 |
| 3 | 9.695 | FM | 3533.27783 | 434.13306 | 0.1356 | 9.614 | 9.872 |
| 4 | 12.916 | FM | 5390.01611 | 1040.12830 | 0.0864 | 12.820 | 13.013 |

Signal 3: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\020218\013118_PF3-PH7-13CBTNAC_3.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 8.956 | FM | 668.34357 | 142.29138 | 0.0783 | 8.902 | 9.017 |
| 2 | 9.130 | FM | 5847.83105 | 930.69971 | 0.1047 | 9.017 | 9.242 |
| 3 | 9.651 | FM | 5163.11035 | 692.02686 | 0.1243 | 9.536 | 9.782 |
| 4 | 12.764 | MF | 5270.82471 | 997.93738 | 0.0880 | 12.628 | 12.878 |

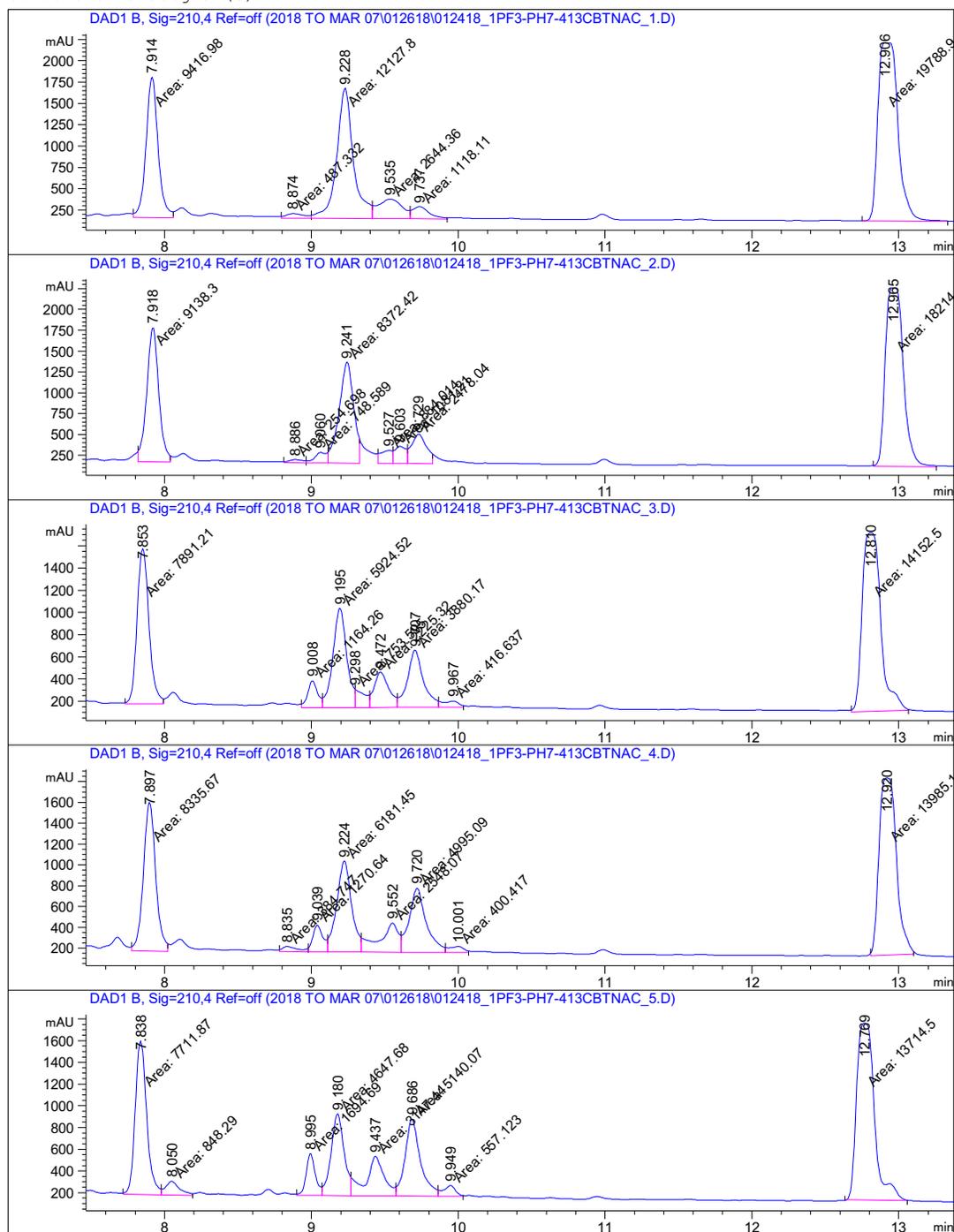
Signal 4: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\020218\013118_PF3-PH7-13CBTNAC_4.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 8.989 | FM | 900.47656 | 190.13754 | 0.0789 | 8.927 | 9.049 |
| 2 | 9.172 | FM | 5806.35840 | 870.80627 | 0.1111 | 9.049 | 9.279 |
| 3 | 9.685 | FM | 6531.10059 | 858.56616 | 0.1268 | 9.577 | 9.835 |
| 4 | 12.862 | FM | 2182.54004 | 418.45169 | 0.0869 | 12.759 | 12.955 |

Signal 5: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\020218\013118_PF3-PH7-13CBTNAC_5.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 8.959 | FM | 765.35132 | 166.98950 | 0.0764 | 8.906 | 9.029 |
| 2 | 9.149 | FM | 3172.03857 | 503.40402 | 0.1050 | 9.029 | 9.217 |
| 3 | 9.658 | FM | 5281.03174 | 725.90582 | 0.1213 | 9.536 | 9.794 |
| 4 | 12.782 | MF | 1728.99561 | 334.71692 | 0.0861 | 12.661 | 12.886 |

Current Chromatogram(s)



=====
Integration Results
=====

Signal 1: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\012618\012418_1PF3-PH7-413CBTNAC_1.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.914 | MF | 9416.98242 | 1647.93689 | 0.0952 | 7.787 | 8.060 |
| 2 | 8.874 | FM | 487.33218 | 54.21212 | 0.1498 | 8.795 | 8.999 |
| 3 | 9.228 | FM | 1.21278e4 | 1529.12659 | 0.1322 | 8.999 | 9.416 |
| 4 | 9.535 | FM | 2644.35767 | 228.19458 | 0.1931 | 9.416 | 9.674 |
| 5 | 9.737 | FM | 1118.11084 | 142.95132 | 0.1304 | 9.674 | 9.924 |
| 6 | 12.906 | FM | 1.97889e4 | 2092.50806 | 0.1576 | 12.750 | 13.333 |
| 7 | 17.096 | MM | 341.25854 | 47.20979 | 0.1205 | 16.977 | 17.244 |

Signal 2: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\012618\012418_1PF3-PH7-413CBTNAC_2.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.918 | FM | 9138.29590 | 1610.67407 | 0.0946 | 7.820 | 8.038 |
| 2 | 8.886 | MF | 254.69795 | 37.63269 | 0.1128 | 8.811 | 8.965 |
| 3 | 9.060 | FM | 748.58875 | 126.74625 | 0.0984 | 8.965 | 9.112 |
| 4 | 9.241 | FM | 8372.42285 | 1212.87939 | 0.1150 | 9.112 | 9.326 |
| 5 | 9.527 | FM | 884.01447 | 154.15724 | 0.0956 | 9.452 | 9.556 |
| 6 | 9.603 | FM | 1081.20557 | 205.80222 | 0.0876 | 9.556 | 9.654 |
| 7 | 9.729 | FM | 2478.03564 | 353.92752 | 0.1167 | 9.654 | 9.825 |
| 8 | 12.965 | FM | 1.82142e4 | 2143.19775 | 0.1416 | 12.826 | 13.256 |
| 9 | 17.135 | FM | 137.25032 | 18.79482 | 0.1217 | 17.041 | 17.286 |

Signal 3: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\012618\012418_1PF3-PH7-413CBTNAC_3.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.853 | MF | 7891.20898 | 1398.38062 | 0.0941 | 7.731 | 7.992 |
| 2 | 9.008 | MF | 1164.25757 | 244.43816 | 0.0794 | 8.933 | 9.075 |
| 3 | 9.195 | MF | 5924.52051 | 898.11829 | 0.1099 | 9.075 | 9.298 |
| 4 | 9.298 | MF | 753.59515 | 167.25101 | 0.0595 | 9.298 | 9.396 |
| 5 | 9.472 | MF | 2225.31738 | 319.45935 | 0.1161 | 9.396 | 9.586 |
| 6 | 9.707 | MF | 3880.17480 | 515.12500 | 0.1255 | 9.586 | 9.867 |
| 7 | 9.967 | FM | 416.63672 | 54.89103 | 0.1050 | 9.867 | 10.036 |
| 8 | 12.810 | MM | 1.41525e4 | 1605.68469 | 0.1469 | 12.676 | 13.068 |
| 9 | 17.018 | FM | 103.15417 | 13.08915 | 0.1313 | 16.907 | 17.166 |

Data File C:\CHEM32\1\DATA\2018 TO MAR 07\012618\012418_1PF3-PH7-413CBTNAC_1.D
Sample Name: 012418_1PF3-ph7-413CBTNac_1

Signal 4: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\012618\012418_1PF3-PH7-413CBTNAC_4.D)

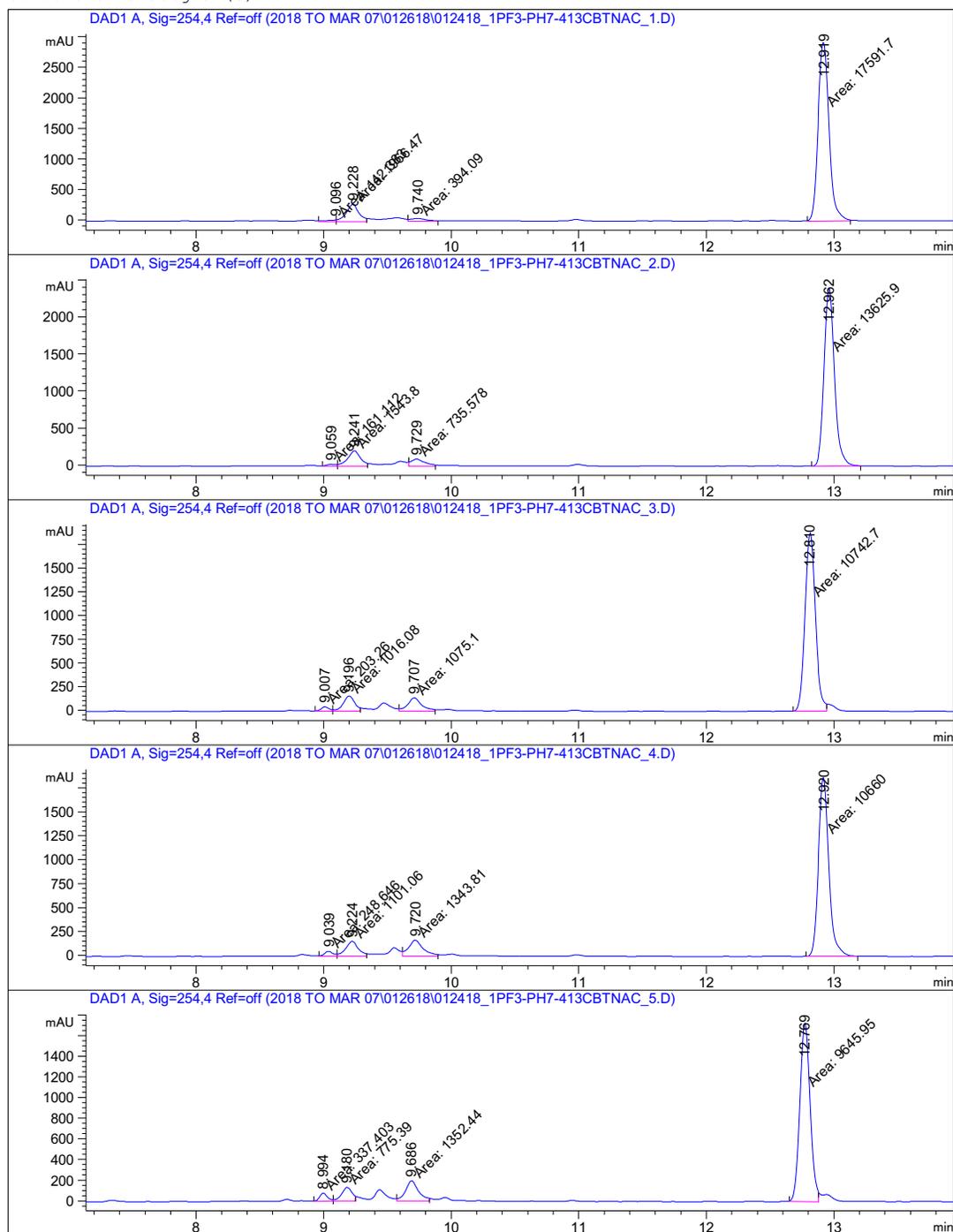
| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.897 | MF | 8335.66602 | 1429.31128 | 0.0972 | 7.776 | 8.022 |
| 2 | 8.835 | MF | 384.74734 | 50.63771 | 0.1266 | 8.782 | 8.979 |
| 3 | 9.039 | FM | 1270.64465 | 258.62827 | 0.0819 | 8.979 | 9.111 |
| 4 | 9.224 | FM | 6181.45459 | 876.76721 | 0.1175 | 9.111 | 9.339 |
| 5 | 9.552 | FM | 2548.07251 | 281.66565 | 0.1508 | 9.339 | 9.610 |
| 6 | 9.720 | FM | 4995.09375 | 616.77344 | 0.1350 | 9.610 | 9.914 |
| 7 | 10.001 | FM | 400.41724 | 57.92731 | 0.1152 | 9.914 | 10.071 |
| 8 | 12.920 | MM | 1.39851e4 | 1702.83252 | 0.1369 | 12.809 | 13.102 |
| 9 | 17.103 | FM | 84.47837 | 10.66243 | 0.1320 | 16.993 | 17.267 |

Signal 5: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\012618\012418_1PF3-PH7-413CBTNAC_5.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.838 | MF | 7711.86670 | 1410.79956 | 0.0911 | 7.715 | 7.976 |
| 2 | 8.050 | FM | 848.29047 | 127.37299 | 0.1110 | 7.976 | 8.190 |
| 3 | 8.995 | MF | 1694.69202 | 389.25287 | 0.0726 | 8.900 | 9.071 |
| 4 | 9.180 | MF | 4647.68359 | 753.63672 | 0.1028 | 9.071 | 9.268 |
| 5 | 9.437 | MF | 3147.43799 | 364.28168 | 0.1440 | 9.268 | 9.575 |
| 6 | 9.686 | MF | 5140.07275 | 710.37439 | 0.1206 | 9.575 | 9.864 |
| 7 | 9.949 | FM | 557.12329 | 97.50262 | 0.0952 | 9.864 | 10.032 |
| 8 | 12.769 | MM | 1.37145e4 | 1626.81152 | 0.1405 | 12.634 | 13.059 |
| 9 | 16.978 | FM | 75.99796 | 9.27865 | 0.1365 | 16.875 | 17.135 |

=====

Current Chromatogram(s)



Data File C:\CHEM32\1\DATA\2018 TO MAR 07\012618\012418_1PF3-PH7-413CBTNAC_1.D
Sample Name: 012418_1PF3-ph7-413CBTNac_1

=====
Integration Results
=====

Signal 1: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\012618\012418_1PF3-PH7-413CBTNAC_1.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 9.096 | MF | 142.38287 | 26.04919 | 0.0911 | 8.961 | 9.097 |
| 2 | 9.228 | FM | 1966.46570 | 281.31848 | 0.1165 | 9.097 | 9.336 |
| 3 | 9.740 | FM | 394.09006 | 45.81873 | 0.1434 | 9.662 | 9.896 |
| 4 | 12.919 | MM | 1.75917e4 | 2931.18726 | 0.1000 | 12.792 | 13.128 |

Signal 2: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\012618\012418_1PF3-PH7-413CBTNAC_2.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 9.059 | FM | 161.11209 | 28.80318 | 0.0932 | 8.991 | 9.109 |
| 2 | 9.241 | FM | 1543.80066 | 213.12289 | 0.1207 | 9.109 | 9.346 |
| 3 | 9.729 | FM | 735.57843 | 97.84830 | 0.1253 | 9.668 | 9.876 |
| 4 | 12.962 | MM | 1.36259e4 | 2410.98779 | 0.0942 | 12.827 | 13.210 |

Signal 3: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\012618\012418_1PF3-PH7-413CBTNAC_3.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 9.007 | FM | 203.26022 | 46.58055 | 0.0727 | 8.934 | 9.072 |
| 2 | 9.196 | FM | 1016.08386 | 161.26431 | 0.1050 | 9.072 | 9.288 |
| 3 | 9.707 | FM | 1075.09631 | 142.35010 | 0.1259 | 9.590 | 9.875 |
| 4 | 12.810 | MF | 1.07427e4 | 1886.81433 | 0.0949 | 12.681 | 12.946 |

Signal 4: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\012618\012418_1PF3-PH7-413CBTNAC_4.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 9.039 | FM | 248.64555 | 52.22098 | 0.0794 | 8.962 | 9.105 |
| 2 | 9.224 | FM | 1101.06201 | 155.99211 | 0.1176 | 9.105 | 9.337 |
| 3 | 9.720 | FM | 1343.80579 | 168.67032 | 0.1328 | 9.618 | 9.895 |
| 4 | 12.920 | MM | 1.06600e4 | 1872.36560 | 0.0949 | 12.783 | 13.186 |

Signal 5: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\012618\012418_1PF3-PH7-413CBTNAC_5.D)

| | | | | | | | |
|---|--------|----|------------|------------|--------|--------|--------|
| 1 | 8.994 | MF | 337.40302 | 79.62730 | 0.0706 | 8.922 | 9.075 |
| 2 | 9.180 | MF | 775.38977 | 133.16174 | 0.0970 | 9.075 | 9.250 |
| 3 | 9.686 | MF | 1352.44153 | 195.07692 | 0.1155 | 9.573 | 9.831 |
| 4 | 12.769 | MF | 9645.94922 | 1733.51257 | 0.0927 | 12.652 | 12.880 |

Print of window 38: Current Chromatogram(s)

Data File : C:\CHEM32\1\DATA\2018 TO MAR 07\020218\013118_1PF3-PH7-413CBTNAC_1.D

Sample Name : 013118_1PF3-ph7-413CBTNac_1

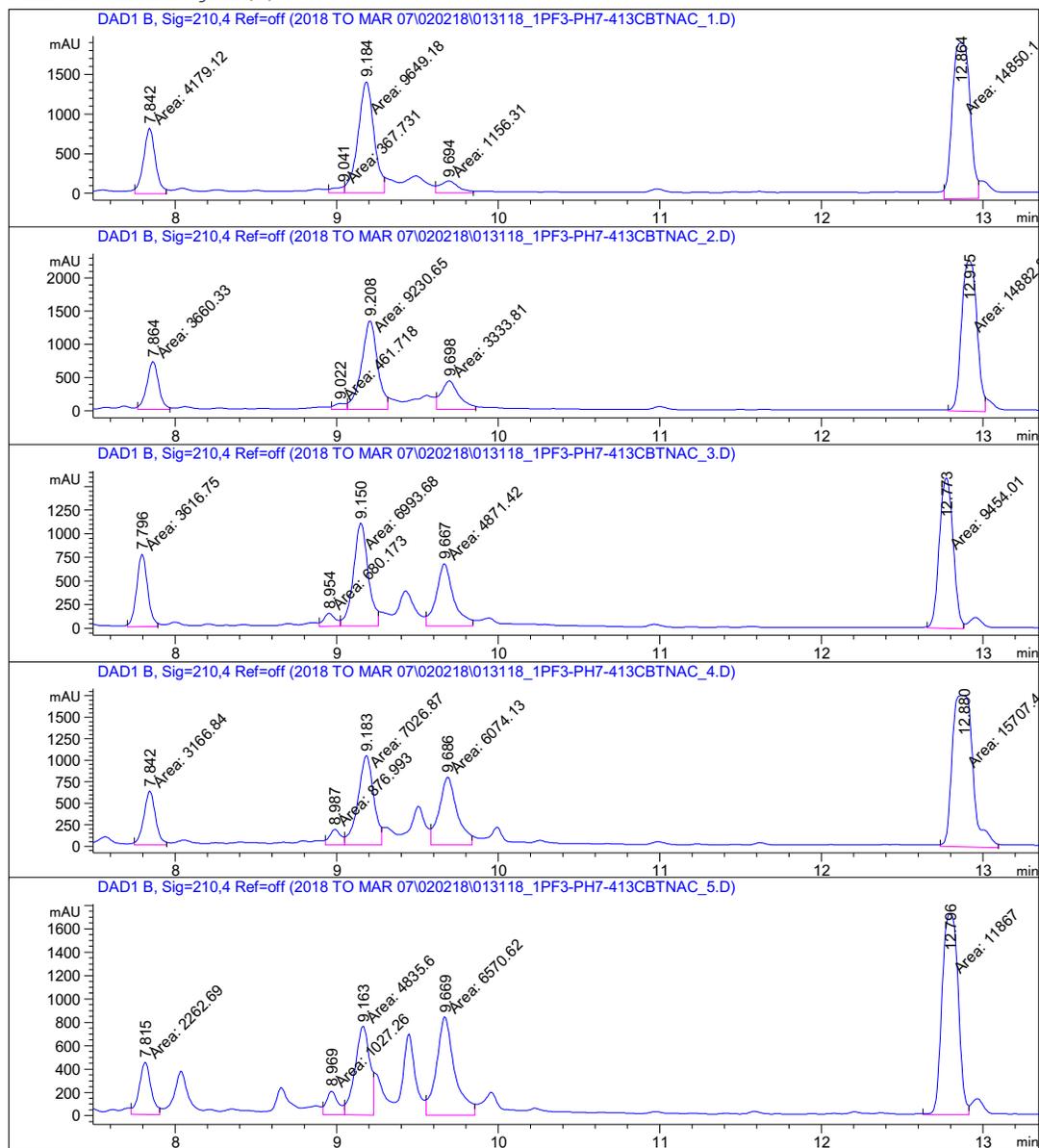
=====
Acq. Operator : SGLK Seq. Line : 7
Acq. Instrument : Instrument 1 Location : Vial 7
Injection Date : 2/2/18 7:45:39 PM Inj : 1
Inj Volume : 1.000 µl

Acq. Method : D:\CHEM32\1\METHODS\AP_C18_INSERT.M
Last changed : 1/22/18 3:04:38 PM by SGLK
Analysis Method : Z:\DESKTOP\BERTOZZI LAB\LCMS INFO\LCMS METHODS TO 112917\SGLK_BERTOZZI_BP_WITHINSERT.M

Last changed : 3/8/18 8:51:08 AM
(modified after loading)

Additional Info : Peak(s) manually integrated

Current Chromatogram(s)



=====
Integration Results
=====

Signal 1: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\020218\013118_1PF3-PH7-413CBTNAC_1.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.842 | FM | 4179.11914 | 821.41675 | 0.0848 | 7.749 | 7.942 |
| 2 | 9.041 | FM | 367.73087 | 78.21339 | 0.0784 | 8.950 | 9.047 |
| 3 | 9.184 | FM | 9649.18066 | 1400.78040 | 0.1148 | 9.047 | 9.293 |
| 4 | 9.694 | FM | 1156.30530 | 145.55722 | 0.1324 | 9.609 | 9.843 |
| 5 | 12.864 | FM | 1.48501e4 | 1956.26233 | 0.1265 | 12.758 | 12.970 |

Signal 2: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\020218\013118_1PF3-PH7-413CBTNAC_2.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.864 | FM | 3660.32935 | 723.93250 | 0.0843 | 7.767 | 7.966 |
| 2 | 9.022 | FM | 461.71805 | 91.62712 | 0.0840 | 8.968 | 9.064 |
| 3 | 9.208 | FM | 9230.65137 | 1339.48608 | 0.1149 | 9.064 | 9.313 |
| 4 | 9.698 | FM | 3333.80762 | 432.63248 | 0.1284 | 9.615 | 9.858 |
| 5 | 12.915 | MF | 1.48828e4 | 2264.37549 | 0.1095 | 12.782 | 13.012 |

Signal 3: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\020218\013118_1PF3-PH7-413CBTNAC_3.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.796 | FM | 3616.75317 | 768.20435 | 0.0785 | 7.702 | 7.893 |
| 2 | 8.954 | FM | 680.17310 | 138.71642 | 0.0817 | 8.891 | 9.020 |
| 3 | 9.150 | FM | 6993.68018 | 1092.80615 | 0.1067 | 9.020 | 9.257 |
| 4 | 9.667 | FM | 4871.42480 | 657.31110 | 0.1235 | 9.550 | 9.840 |
| 5 | 12.773 | FM | 9454.00977 | 1592.20349 | 0.0990 | 12.652 | 12.879 |

Signal 4: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\020218\013118_1PF3-PH7-413CBTNAC_4.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.842 | FM | 3166.83960 | 631.48499 | 0.0836 | 7.746 | 7.945 |
| 2 | 8.987 | FM | 876.99335 | 181.88477 | 0.0804 | 8.929 | 9.047 |
| 3 | 9.183 | FM | 7026.86963 | 1040.78650 | 0.1125 | 9.047 | 9.275 |
| 4 | 9.686 | FM | 6074.12842 | 789.74097 | 0.1282 | 9.580 | 9.835 |
| 5 | 12.880 | MM | 1.57074e4 | 1758.29065 | 0.1489 | 12.735 | 13.092 |

Signal 5: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\020218\013118_1PF3-PH7-413CBTNAC_5.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.815 | FM | 2262.69141 | 448.04358 | 0.0842 | 7.726 | 7.901 |
| 2 | 8.969 | FM | 1027.25977 | 204.35941 | 0.0838 | 8.915 | 9.047 |
| 3 | 9.163 | FM | 4835.60400 | 760.72070 | 0.1059 | 9.047 | 9.228 |
| 4 | 9.669 | FM | 6570.62402 | 840.52429 | 0.1303 | 9.550 | 9.852 |
| 5 | 12.796 | MF | 1.18670e4 | 1718.16589 | 0.1151 | 12.625 | 12.910 |

Data File C:\CHEM32\1\DATA\2018 TO MAR 07\020218\013118_1PF3-PH7-413CBTNAC_1.D
Sample Name: 013118_1PF3-ph7-413CBTNac_1

=====
Integration Results
=====

Signal 1: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\020218\013118_1PF3-PH7-413CBTNAC_1.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 9.041 | FM | 55.88944 | 11.06242 | 0.0842 | 8.942 | 9.044 |
| 2 | 9.184 | FM | 1641.74548 | 253.71167 | 0.1078 | 9.044 | 9.283 |
| 3 | 9.697 | FM | 286.40692 | 40.31698 | 0.1184 | 9.622 | 9.811 |
| 4 | 12.864 | FM | 1.08668e4 | 2075.80518 | 0.0872 | 12.765 | 12.958 |

Signal 2: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\020218\013118_1PF3-PH7-413CBTNAC_2.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 9.023 | FM | 105.11598 | 19.44548 | 0.0901 | 8.956 | 9.064 |
| 2 | 9.208 | FM | 1616.72717 | 230.14301 | 0.1171 | 9.064 | 9.324 |
| 3 | 9.699 | FM | 901.97675 | 117.54403 | 0.1279 | 9.631 | 9.870 |
| 4 | 12.915 | FM | 8779.70313 | 1712.29663 | 0.0855 | 12.808 | 13.006 |

Signal 3: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\020218\013118_1PF3-PH7-413CBTNAC_3.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 8.953 | FM | 152.07585 | 30.38887 | 0.0834 | 8.899 | 9.024 |
| 2 | 9.151 | FM | 1291.37805 | 203.47014 | 0.1058 | 9.024 | 9.252 |
| 3 | 9.667 | FM | 1378.61804 | 184.58427 | 0.1245 | 9.552 | 9.823 |
| 4 | 12.773 | FM | 4896.59473 | 931.85321 | 0.0876 | 12.659 | 12.874 |

Signal 4: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\020218\013118_1PF3-PH7-413CBTNAC_4.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 8.987 | FM | 195.99939 | 38.38641 | 0.0851 | 8.929 | 9.051 |
| 2 | 9.183 | FM | 1271.32251 | 188.63309 | 0.1123 | 9.051 | 9.274 |
| 3 | 9.686 | FM | 1673.51196 | 218.45514 | 0.1277 | 9.586 | 9.832 |
| 4 | 12.869 | MF | 1.48578e4 | 2598.30322 | 0.0953 | 12.759 | 12.964 |

Signal 5: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\020218\013118_1PF3-PH7-413CBTNAC_5.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 0.980 | BB | 31.18398 | 7.66578e-1 | 0.4986 | 1.45e-3 | 1.063 |
| 2 | 8.969 | FM | 189.97842 | 40.02207 | 0.0791 | 8.906 | 9.035 |
| 3 | 9.164 | FM | 791.44257 | 130.94086 | 0.1007 | 9.035 | 9.213 |
| 4 | 9.670 | FM | 1595.67419 | 225.92827 | 0.1177 | 9.545 | 9.782 |
| 5 | 12.796 | FM | 7938.74023 | 1519.81384 | 0.0871 | 12.690 | 12.891 |

Data File C:\CHEM32\1\DATA\2018 TO MAR 07\012618\012418_4PF3-PH7-13CBTNAC_1.D
Sample Name: 012218_PF3-ph8-azidoCBTNac_1

=====
Integration Results
=====

Signal 1: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\012618\012418_4PF3-PH7-13CBTNAC_1.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.860 | MF | 1.84950e4 | 2050.75513 | 0.1503 | 7.720 | 8.024 |
| 2 | 8.089 | MF | 2392.83325 | 382.49109 | 0.1043 | 8.024 | 8.210 |
| 3 | 8.870 | MF | 765.58057 | 85.67575 | 0.1489 | 8.776 | 8.971 |
| 4 | 9.060 | MF | 642.70343 | 166.08194 | 0.0645 | 8.971 | 9.062 |
| 5 | 9.208 | MF | 1.85275e4 | 2048.39893 | 0.1507 | 9.062 | 9.343 |
| 6 | 9.715 | MF | 2025.30261 | 257.27994 | 0.1312 | 9.638 | 9.885 |
| 7 | 12.896 | MF | 1810.43713 | 333.72260 | 0.0904 | 12.793 | 12.981 |
| 8 | 17.091 | MM | 3022.47485 | 399.32254 | 0.1262 | 16.939 | 17.285 |

Signal 2: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\012618\012418_4PF3-PH7-13CBTNAC_2.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.894 | FM | 2.06262e4 | 2095.16553 | 0.1641 | 7.694 | 8.040 |
| 2 | 8.117 | FM | 2581.28320 | 382.76730 | 0.1124 | 8.040 | 8.218 |
| 3 | 9.056 | FM | 2080.65186 | 425.99606 | 0.0814 | 8.969 | 9.100 |
| 4 | 9.232 | FM | 1.89702e4 | 2035.62708 | 0.1553 | 9.100 | 9.433 |
| 5 | 9.726 | FM | 6051.60547 | 719.79962 | 0.1401 | 9.636 | 9.945 |
| 6 | 12.955 | MM | 1533.16333 | 254.23740 | 0.1005 | 12.864 | 13.138 |
| 7 | 17.131 | MM | 1801.86560 | 237.32285 | 0.1265 | 16.935 | 17.304 |
| 8 | 27.940 | BBA | 1001.94745 | 111.75077 | 0.1494 | 27.807 | 28.000 |

Signal 3: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\012618\012418_4PF3-PH7-13CBTNAC_3.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.386 | MF | 7936.72314 | 765.12000 | 0.1729 | 7.089 | 7.572 |
| 2 | 7.842 | MF | 1.41918e4 | 1574.01428 | 0.1503 | 7.712 | 7.979 |
| 3 | 8.066 | MF | 2933.36401 | 388.90378 | 0.1257 | 7.979 | 8.220 |
| 4 | 9.008 | MF | 3157.13452 | 698.73083 | 0.0753 | 8.918 | 9.071 |
| 5 | 9.190 | MF | 1.21071e4 | 1499.39441 | 0.1346 | 9.071 | 9.325 |
| 6 | 9.704 | MF | 7429.56299 | 972.61151 | 0.1273 | 9.604 | 9.896 |
| 7 | 12.807 | MF | 1474.88635 | 268.96881 | 0.0914 | 12.678 | 12.932 |
| 8 | 17.020 | MM | 1165.90271 | 149.83099 | 0.1297 | 16.844 | 17.177 |
| 9 | 27.049 | HH | 38.29931 | 3.93957 | 0.1554 | 26.843 | 27.199 |
| 10 | 27.931 | BBA | 964.60730 | 97.13586 | 0.1655 | 27.769 | 27.996 |

Data File C:\CHEM32\1\DATA\2018 TO MAR 07\012618\012418_4PF3-PH7-13CBTNAC_1.D
Sample Name: 012218_Pf3-ph8-azidoCBTNac_1

Signal 4: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\012618\012418_4PF3-PH7-13CBTNAC_4.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.681 | FM | 3636.06616 | 680.77594 | 0.0890 | 7.600 | 7.738 |
| 2 | 7.881 | FM | 1.56507e4 | 1654.62610 | 0.1576 | 7.738 | 8.014 |
| 3 | 8.118 | FM | 9029.92383 | 1248.92078 | 0.1205 | 8.014 | 8.327 |
| 4 | 9.038 | FM | 3078.17871 | 652.84564 | 0.0786 | 8.975 | 9.099 |
| 5 | 9.219 | FM | 1.22704e4 | 1517.02673 | 0.1348 | 9.099 | 9.339 |
| 6 | 9.714 | FM | 9018.74121 | 1109.87085 | 0.1354 | 9.616 | 9.914 |
| 7 | 12.914 | MF | 1438.77722 | 274.10233 | 0.0875 | 12.795 | 12.995 |
| 8 | 17.100 | MM | 878.28180 | 117.26001 | 0.1248 | 16.955 | 17.257 |
| 9 | 27.939 | HBA | 5135.85693 | 416.95944 | 0.2053 | 27.132 | 27.999 |

=====

=====
Integration Results
=====

Signal 1: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\012618\012418_4PF3-PH7-13CBTNAC_1.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 9.073 | FM | 109.53465 | 26.67127 | 0.0684 | 8.975 | 9.077 |
| 2 | 9.208 | FM | 3892.34985 | 579.80493 | 0.1119 | 9.077 | 9.332 |
| 3 | 9.718 | FM | 553.98206 | 74.16487 | 0.1245 | 9.638 | 9.879 |
| 4 | 12.896 | MF | 767.04785 | 139.52420 | 0.0916 | 12.774 | 12.987 |

Signal 2: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\012618\012418_4PF3-PH7-13CBTNAC_2.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 9.055 | FM | 369.18109 | 77.76151 | 0.0791 | 8.970 | 9.104 |
| 2 | 9.232 | FM | 3423.59888 | 478.64450 | 0.1192 | 9.104 | 9.360 |
| 3 | 9.726 | FM | 1544.29590 | 191.12616 | 0.1347 | 9.642 | 9.902 |
| 4 | 12.955 | MF | 589.09991 | 108.67957 | 0.0903 | 12.857 | 13.029 |

Signal 3: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\012618\012418_4PF3-PH7-13CBTNAC_3.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 9.008 | MF | 556.58862 | 136.95667 | 0.0677 | 8.942 | 9.077 |
| 2 | 9.191 | FM | 2360.06812 | 378.14111 | 0.1040 | 9.077 | 9.295 |
| 3 | 9.703 | FM | 2024.40527 | 277.56799 | 0.1216 | 9.582 | 9.879 |
| 4 | 12.807 | MF | 596.59241 | 112.93456 | 0.0880 | 12.713 | 12.913 |

Signal 4: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\012618\012418_4PF3-PH7-13CBTNAC_4.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 9.038 | FM | 612.48633 | 130.93390 | 0.0780 | 8.951 | 9.104 |
| 2 | 9.219 | FM | 2387.05029 | 345.88968 | 0.1150 | 9.104 | 9.322 |
| 3 | 9.715 | FM | 2560.03564 | 324.88141 | 0.1313 | 9.615 | 9.893 |
| 4 | 12.914 | MF | 623.71161 | 117.50446 | 0.0885 | 12.825 | 13.001 |

Print of window 38: Current Chromatogram(s)

Data File : C:\CHEM32\1\DATA\2018 TO MAR 07\020218\013118_4PF3-PH7-13CBTNAC_1.D

Sample Name : 013118_4PF3-ph7-13CBTNAC_1

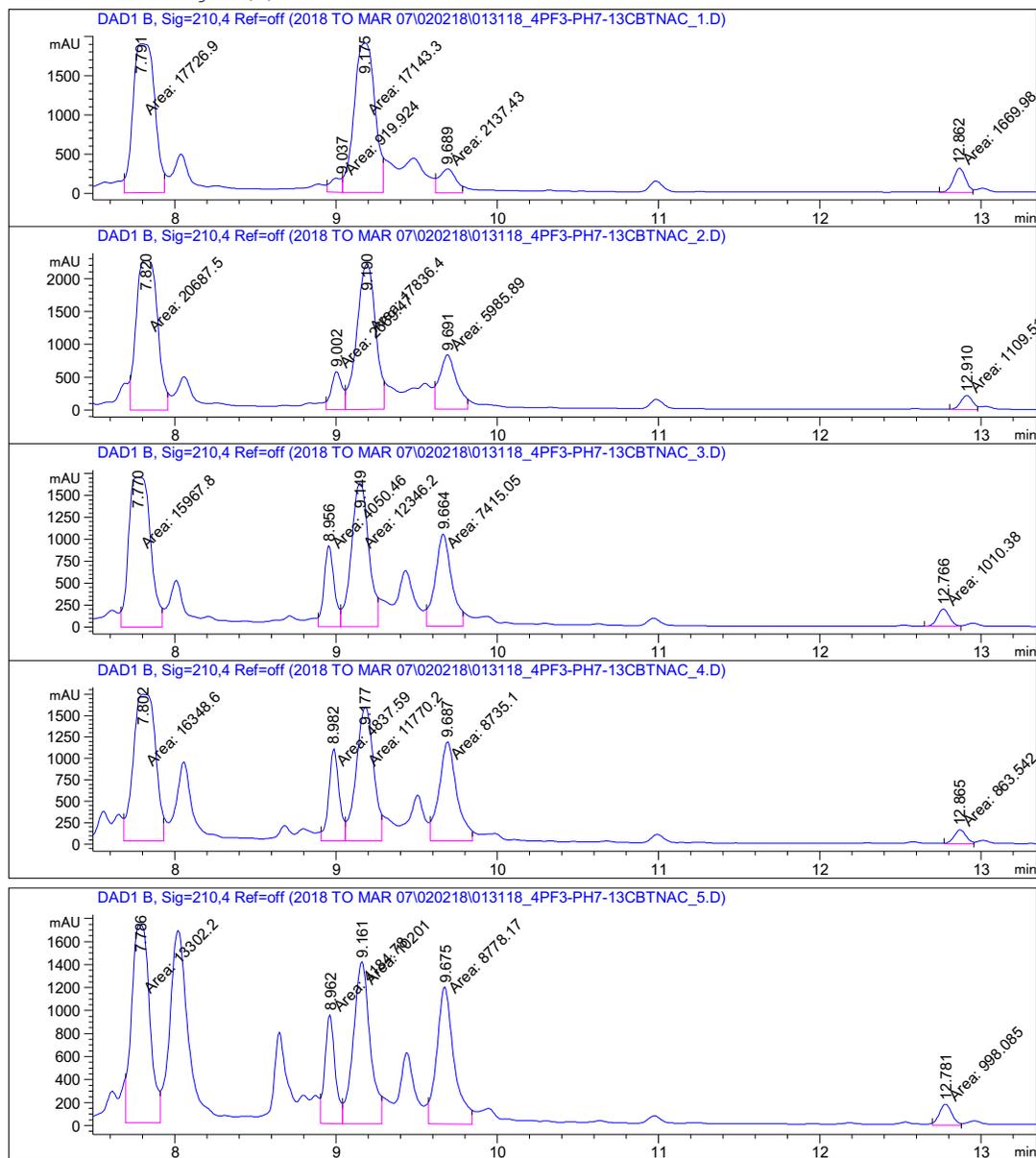
=====
Acq. Operator : SGLK Seq. Line : 6
Acq. Instrument : Instrument 1 Location : Vial 6
Injection Date : 2/2/18 7:16:34 PM Inj : 1
Inj Volume : 1.000 µl

Acq. Method : D:\CHEM32\1\METHODS\AP_C18_INSERT.M
Last changed : 1/22/18 3:04:38 PM by SGLK
Analysis Method : Z:\DESKTOP\BERTOZZI LAB\LCMS INFO\LCMS METHODS TO 112917\SGLK_BERTOZZI_BP_WITHINSERT.M

Last changed : 3/8/18 8:51:08 AM
(modified after loading)

Additional Info : Peak(s) manually integrated

Current Chromatogram(s)



Data File C:\CHEM32\1\DATA\2018 TO MAR 07\020218\013118_4PF3-PH7-13CBTNAC_1.D
Sample Name: 013118_4PF3-ph7-13CBTNAC_1

=====
Integration Results
=====

Signal 1: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\020218\013118_4PF3-PH7-13CBTNAC_1.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.791 | FM | 1.77269e4 | 1896.11926 | 0.1558 | 7.684 | 7.935 |
| 2 | 9.037 | FM | 919.92407 | 184.64258 | 0.0830 | 8.943 | 9.039 |
| 3 | 9.175 | FM | 1.71433e4 | 1902.55383 | 0.1502 | 9.039 | 9.290 |
| 4 | 9.689 | FM | 2137.42554 | 306.13031 | 0.1164 | 9.615 | 9.783 |
| 5 | 12.862 | MF | 1669.97876 | 311.27527 | 0.0894 | 12.740 | 12.947 |

Signal 2: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\020218\013118_4PF3-PH7-13CBTNAC_2.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.820 | MF | 2.06875e4 | 2273.26367 | 0.1517 | 7.722 | 7.955 |
| 2 | 9.002 | MF | 2669.47339 | 576.34473 | 0.0772 | 8.938 | 9.056 |
| 3 | 9.190 | MF | 1.78364e4 | 2205.12256 | 0.1348 | 9.056 | 9.297 |
| 4 | 9.691 | MF | 5985.88770 | 826.99255 | 0.1206 | 9.614 | 9.815 |
| 5 | 12.910 | MF | 1109.50574 | 216.23482 | 0.0855 | 12.802 | 12.978 |

Signal 3: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\020218\013118_4PF3-PH7-13CBTNAC_3.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.770 | MF | 1.59678e4 | 1708.92773 | 0.1557 | 7.666 | 7.921 |
| 2 | 8.956 | MF | 4050.46216 | 925.19824 | 0.0730 | 8.888 | 9.028 |
| 3 | 9.149 | MF | 1.23462e4 | 1635.00671 | 0.1259 | 9.028 | 9.260 |
| 4 | 9.664 | MF | 7415.04932 | 1055.82605 | 0.1170 | 9.561 | 9.787 |
| 5 | 12.766 | MF | 1010.38171 | 196.10500 | 0.0859 | 12.645 | 12.871 |

Signal 4: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\020218\013118_4PF3-PH7-13CBTNAC_4.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.802 | FM | 1.63486e4 | 1706.57397 | 0.1597 | 7.682 | 7.930 |
| 2 | 8.982 | FM | 4837.59375 | 1077.47766 | 0.0748 | 8.907 | 9.056 |
| 3 | 9.177 | FM | 1.17702e4 | 1562.65039 | 0.1255 | 9.056 | 9.282 |
| 4 | 9.687 | FM | 8735.09570 | 1156.13318 | 0.1259 | 9.583 | 9.843 |
| 5 | 12.865 | FM | 863.54230 | 162.56932 | 0.0885 | 12.768 | 12.954 |

Signal 5: DAD1 B, Sig=210,4 Ref=off (2018 TO MAR 07\020218\013118_4PF3-PH7-13CBTNAC_5.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 7.786 | MF | 1.33022e4 | 1725.83533 | 0.1285 | 7.694 | 7.908 |
| 2 | 8.962 | MF | 4184.78467 | 947.06653 | 0.0736 | 8.904 | 9.040 |
| 3 | 9.161 | MF | 1.02010e4 | 1408.12524 | 0.1207 | 9.040 | 9.282 |
| 4 | 9.675 | MF | 8778.17090 | 1192.27075 | 0.1227 | 9.573 | 9.840 |
| 5 | 12.781 | MF | 998.08514 | 184.60497 | 0.0901 | 12.695 | 12.875 |

Data File C:\CHEM32\1\DATA\2018 TO MAR 07\020218\013118_4PF3-PH7-13CBTNAC_1.D
Sample Name: 013118_4PF3-ph7-13CBTNAC_1

=====
Integration Results
=====

Signal 1: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\020218\013118_4PF3-PH7-13CBTNAC_1.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 8.997 | MF | 200.26137 | 38.45741 | 0.0868 | 8.950 | 9.045 |
| 2 | 9.173 | MF | 4234.27588 | 634.43121 | 0.1112 | 9.045 | 9.299 |
| 3 | 9.691 | MF | 737.03558 | 96.44783 | 0.1274 | 9.614 | 9.811 |
| 4 | 12.862 | MF | 729.25427 | 132.65248 | 0.0916 | 12.754 | 12.951 |

Signal 2: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\020218\013118_4PF3-PH7-13CBTNAC_2.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 9.002 | FM | 572.63983 | 118.38261 | 0.0806 | 8.938 | 9.062 |
| 2 | 9.191 | FM | 3598.41650 | 516.68695 | 0.1161 | 9.062 | 9.304 |
| 3 | 9.691 | FM | 1752.47876 | 226.24167 | 0.1291 | 9.619 | 9.856 |
| 4 | 12.910 | MF | 474.20395 | 90.58943 | 0.0872 | 12.811 | 12.985 |

Signal 3: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\020218\013118_4PF3-PH7-13CBTNAC_3.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 8.956 | FM | 871.46686 | 199.60472 | 0.0728 | 8.888 | 9.034 |
| 2 | 9.149 | FM | 2728.26025 | 419.56812 | 0.1084 | 9.034 | 9.276 |
| 3 | 9.664 | FM | 2236.65991 | 306.16095 | 0.1218 | 9.546 | 9.811 |
| 4 | 12.766 | FM | 453.96454 | 85.18710 | 0.0888 | 12.670 | 12.873 |

Signal 4: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\020218\013118_4PF3-PH7-13CBTNAC_4.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 8.981 | FM | 1078.54163 | 244.56328 | 0.0735 | 8.910 | 9.057 |
| 2 | 9.178 | FM | 2427.26831 | 358.23096 | 0.1129 | 9.057 | 9.276 |
| 3 | 9.687 | FM | 2653.67725 | 342.92633 | 0.1290 | 9.580 | 9.861 |
| 4 | 12.865 | FM | 379.64224 | 70.02226 | 0.0904 | 12.766 | 12.951 |

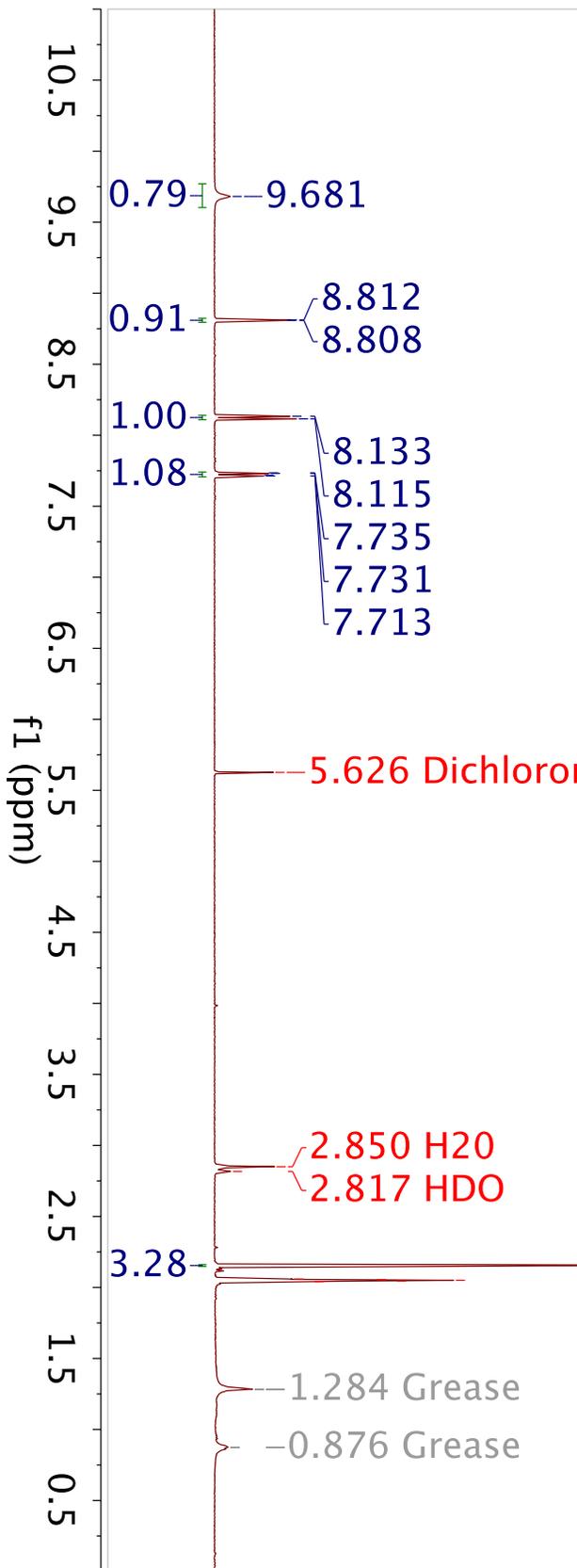
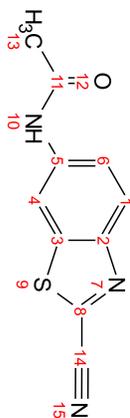
Signal 5: DAD1 A, Sig=254,4 Ref=off (2018 TO MAR 07\020218\013118_4PF3-PH7-13CBTNAC_5.D)

| Peak # | Time [min] | Type | Area [mAU*s] | Height [mAU] | Width [min] | Start [min] | End [min] |
|--------|------------|------|--------------|--------------|-------------|-------------|-----------|
| 1 | 8.962 | MF | 886.41290 | 200.69423 | 0.0736 | 8.893 | 9.051 |
| 2 | 9.162 | MF | 1947.91333 | 281.73981 | 0.1152 | 9.051 | 9.282 |
| 3 | 9.675 | MF | 2469.42773 | 342.71152 | 0.1201 | 9.546 | 9.816 |
| 4 | 12.781 | MF | 429.56952 | 78.42221 | 0.0913 | 12.687 | 12.873 |

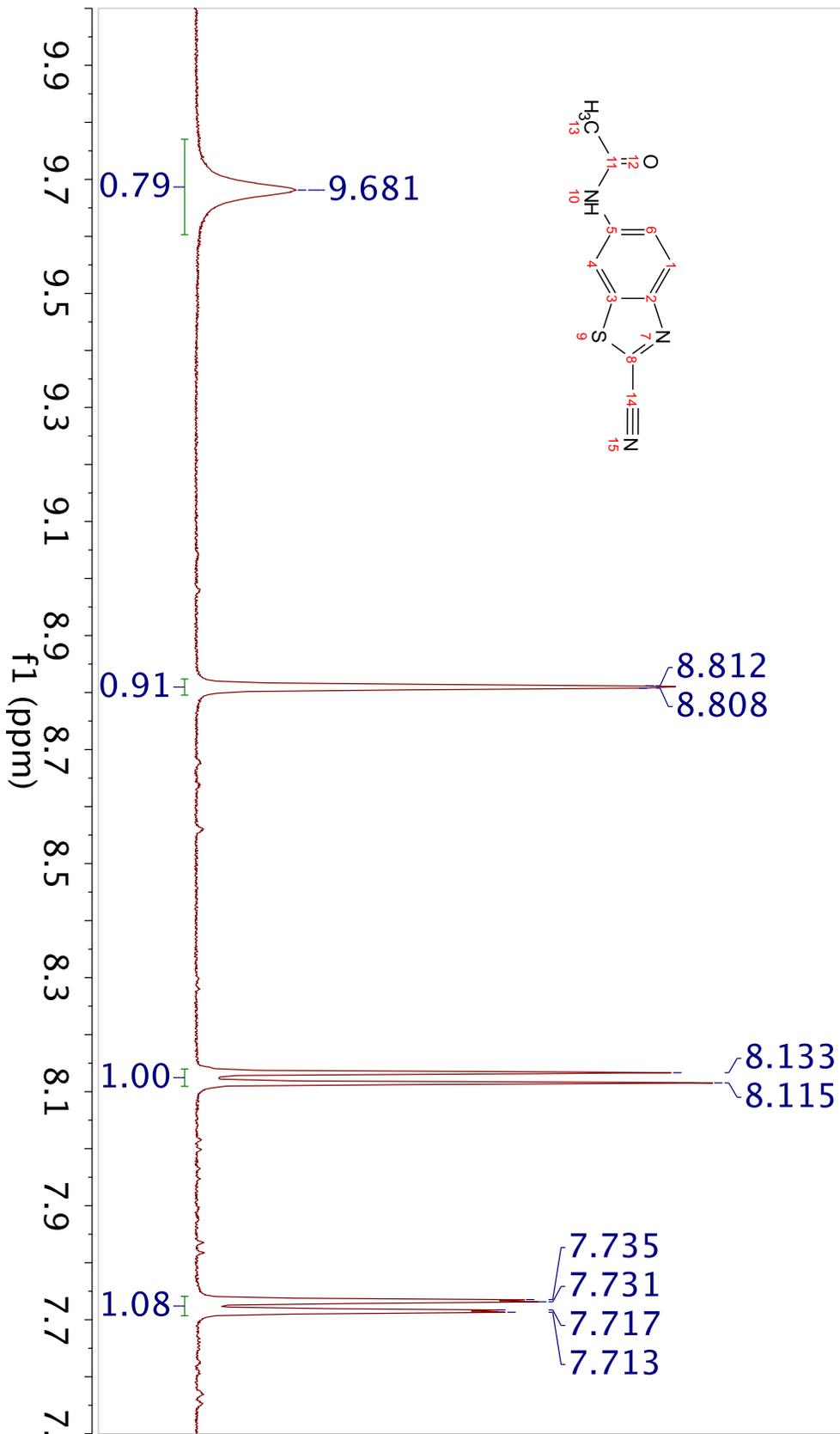
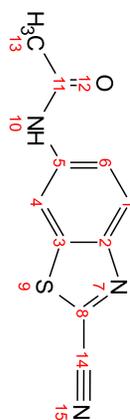
Appendix 2.

^1H and ^{13}C NMR Spectra

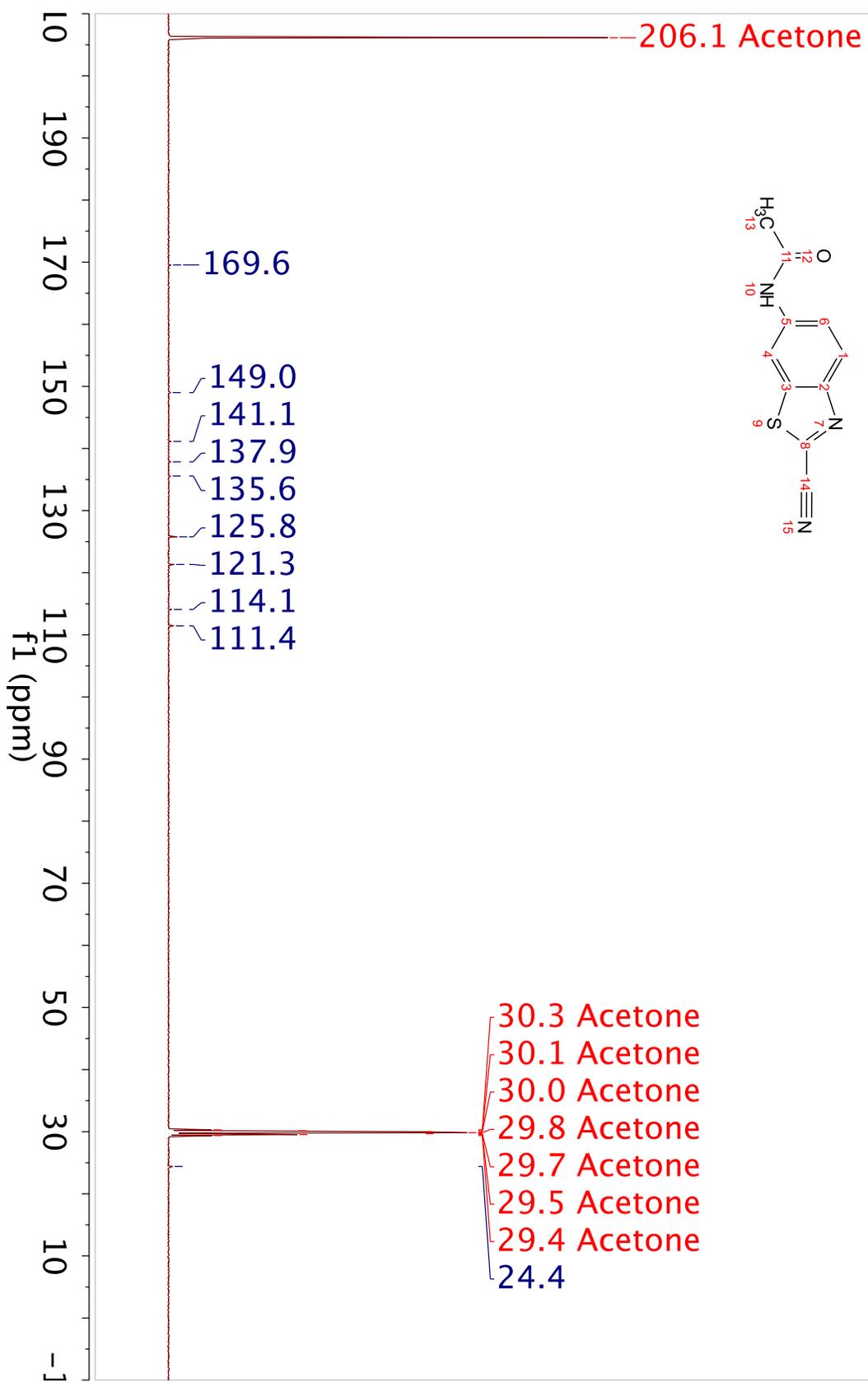
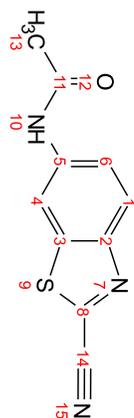
021218_CBTNAC_acetone_proton
STANDARD PROTON PARAMETERS



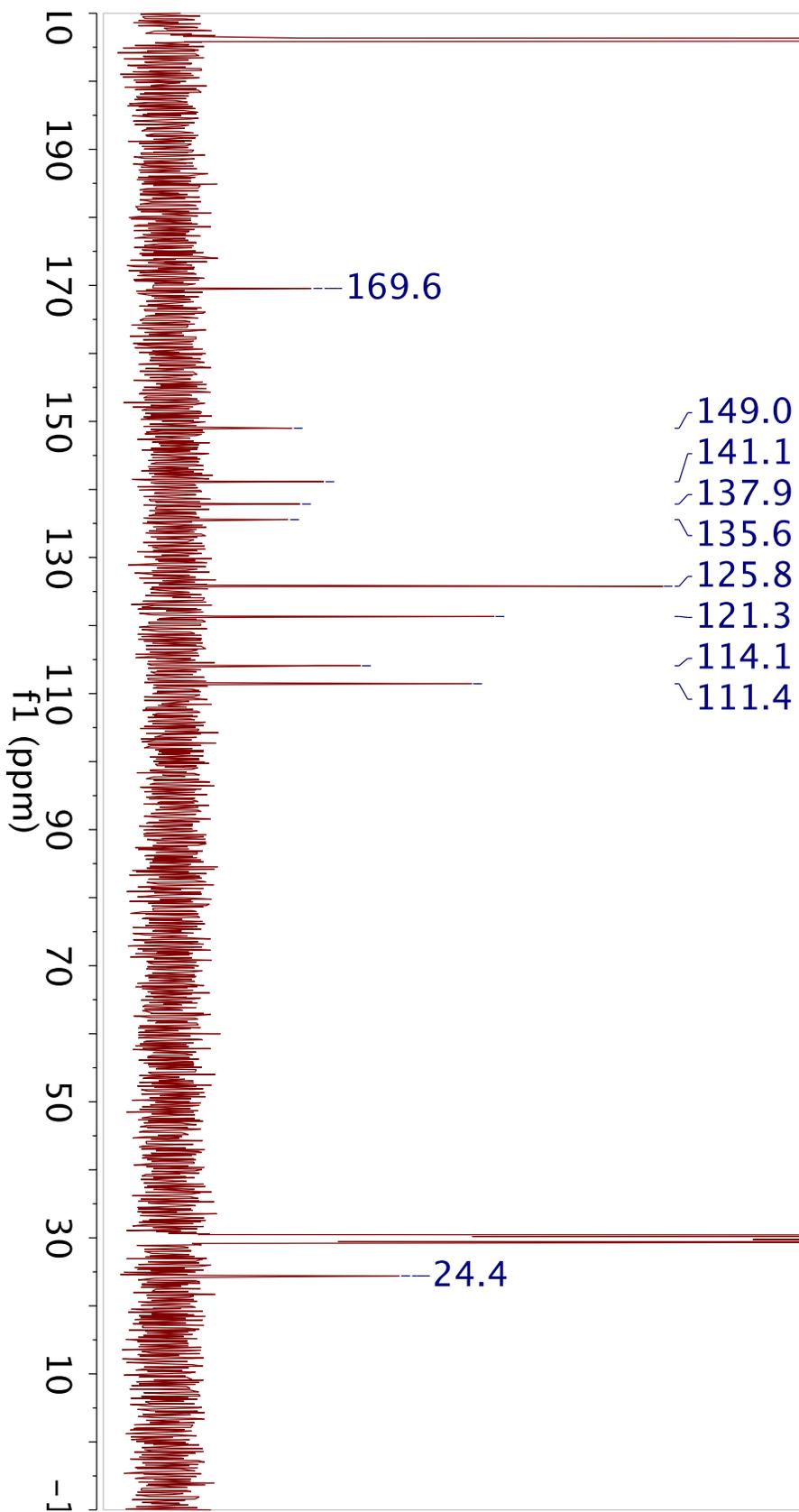
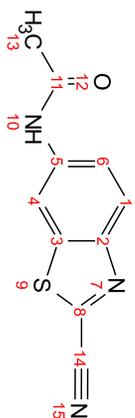
021218_CBTNAC_acetone_proton
STANDARD PROTON PARAMETERS



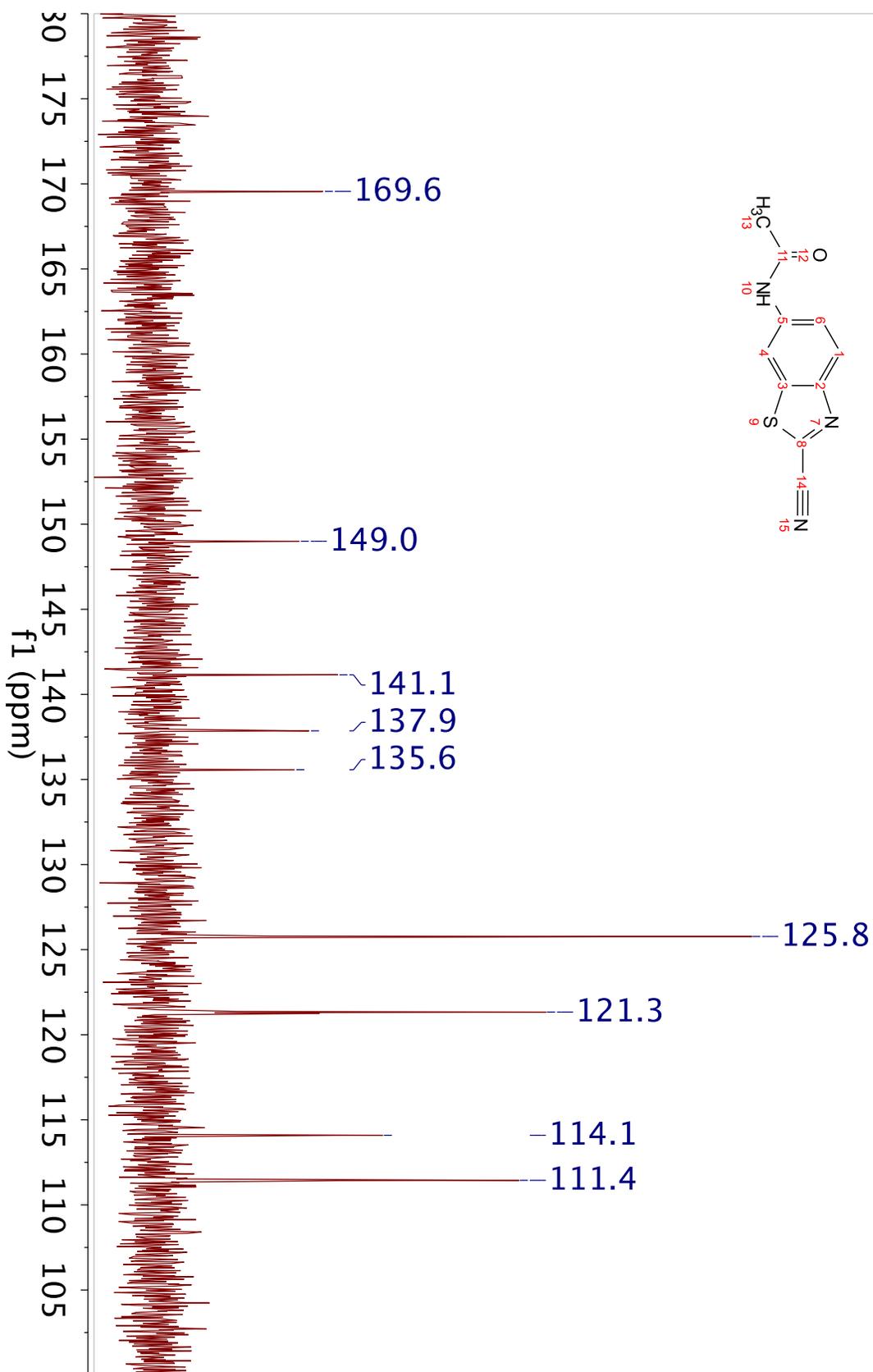
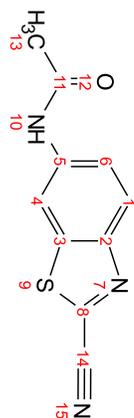
021218_cbtnac_acetone_13c
C13par



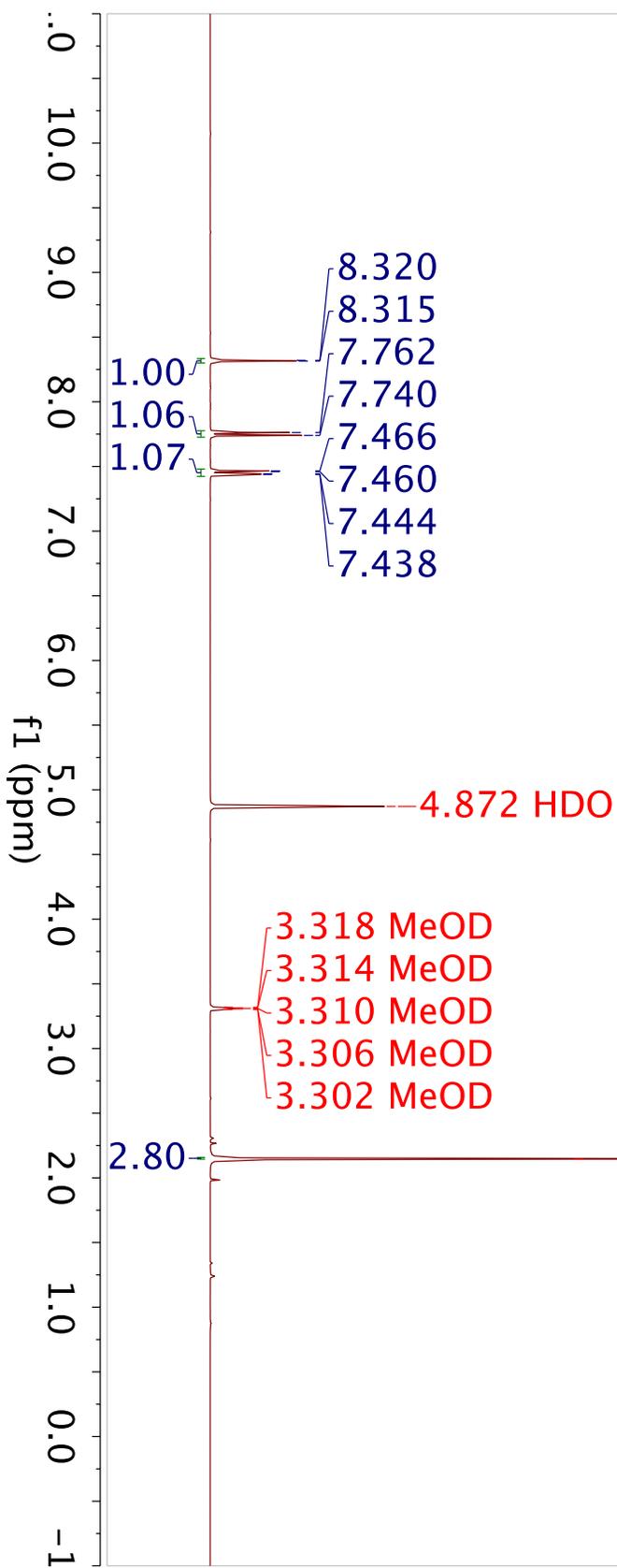
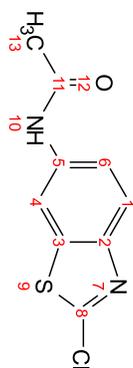
021218_cbtnac_acetone_13c
C13par



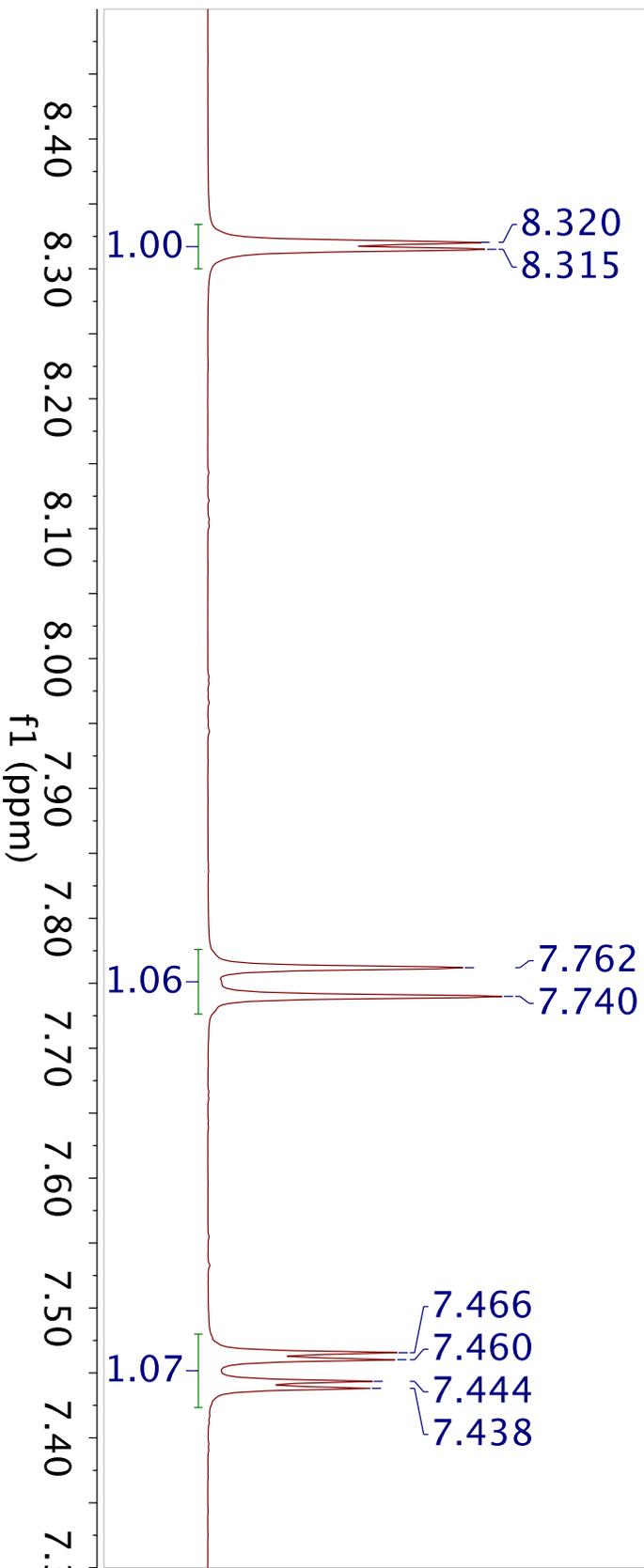
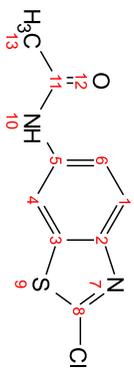
021218_cbtnac_acetone_13c
C13par



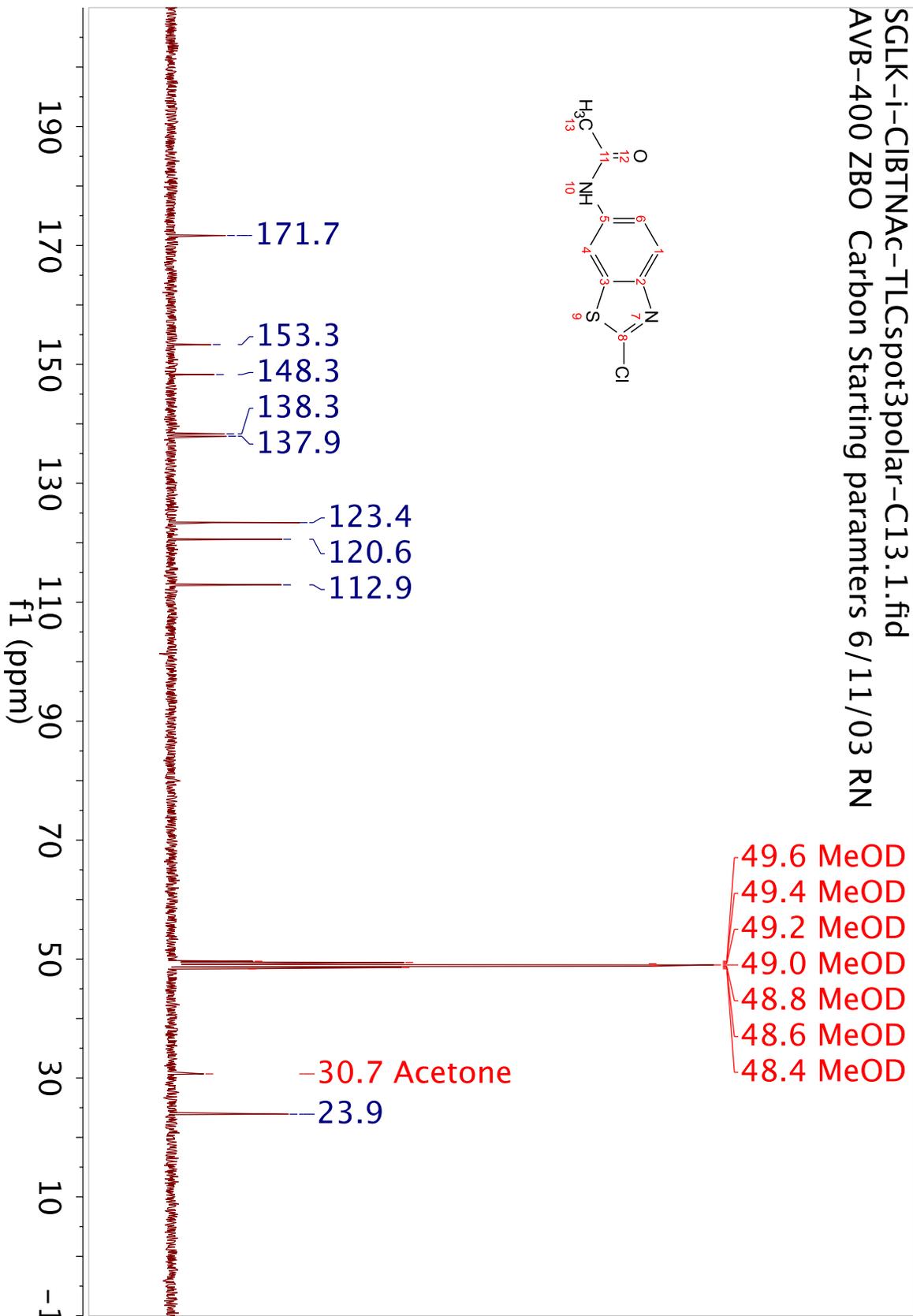
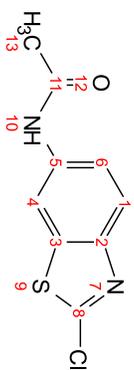
SGLK-i-CIBTNAC-TLCspot3polar.1.fid
AVB-400 ZBO Proton starting parameters. 6/11/03 RN



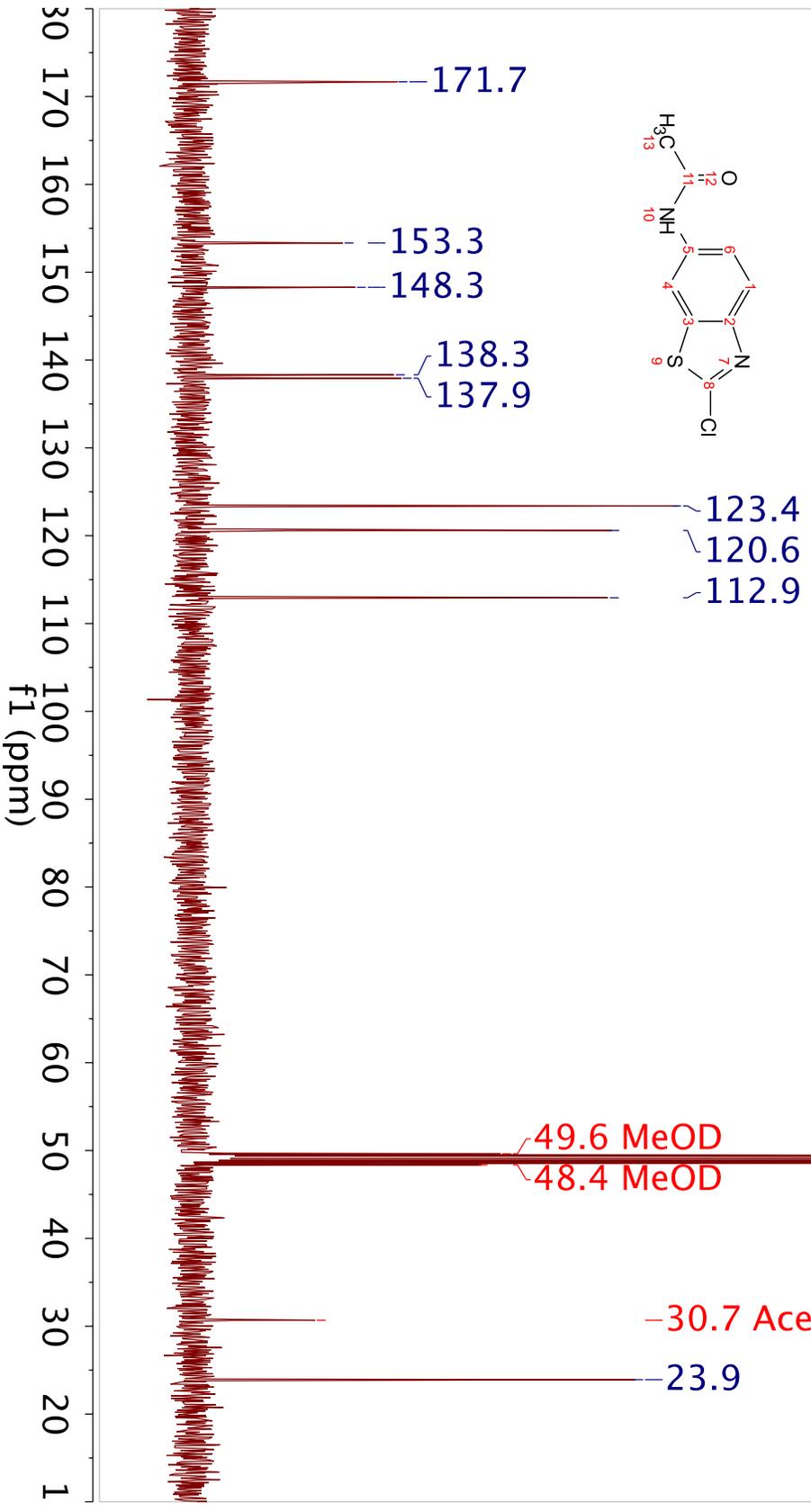
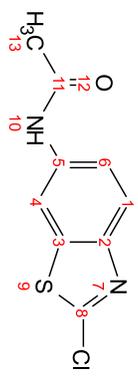
SGLK-i-CIBTNAC-TLCspot3polar.1.fid
AVB-400 ZBO Proton starting parameters. 6/11/03 RN



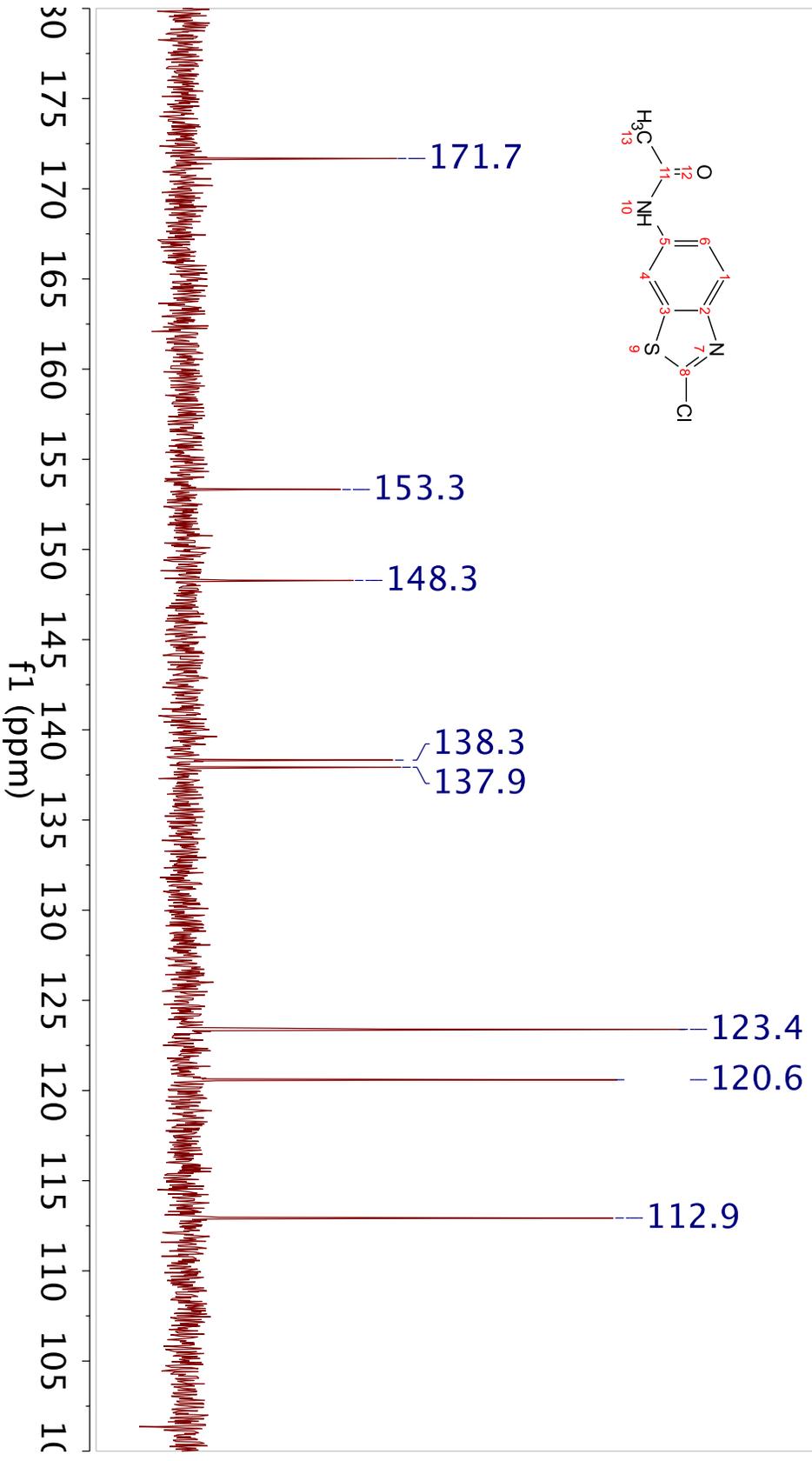
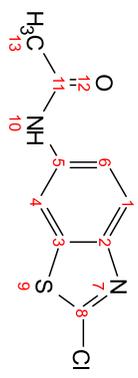
SGLK-i-CIBTNAC-TLCspot3polar-C13.1.fid
AVB-400 ZBO Carbon Starting paramters 6/11/03 RN



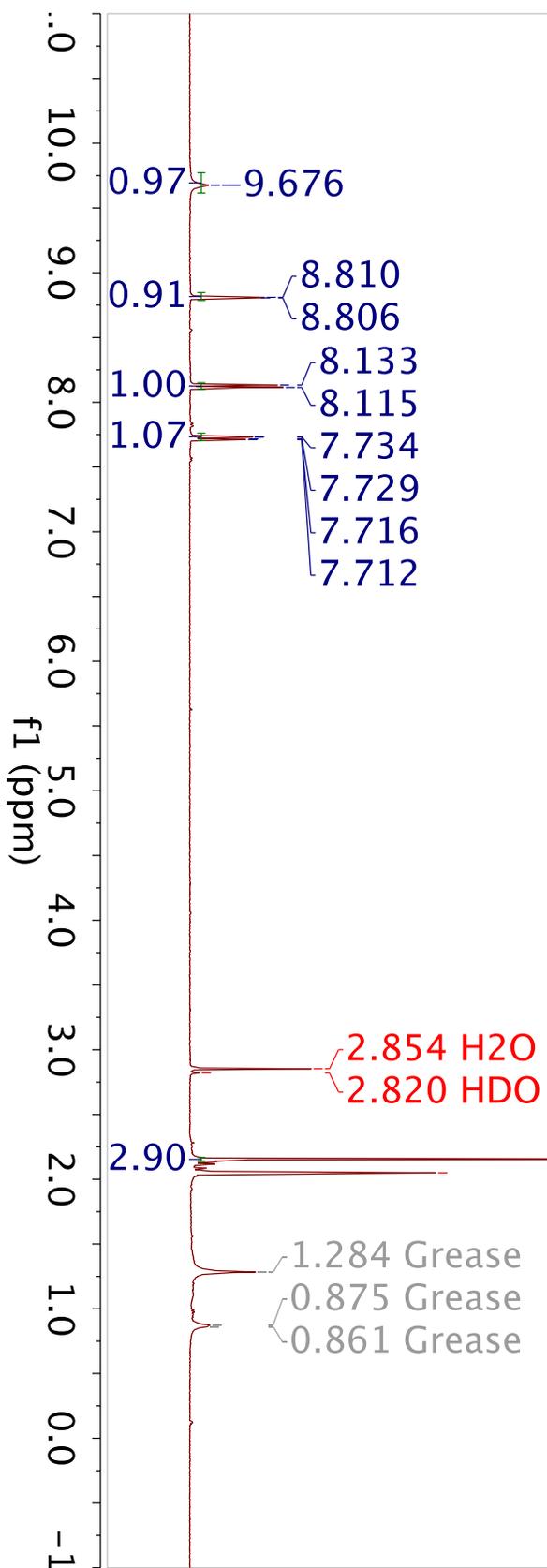
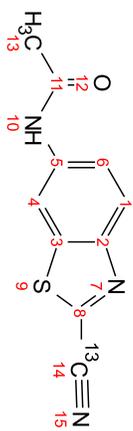
SGLK-i-CIBTNAC-TLCspot3polar-C13.1.fid
AVB-400 ZBO Carbon Starting paramters 6/11/03 RN



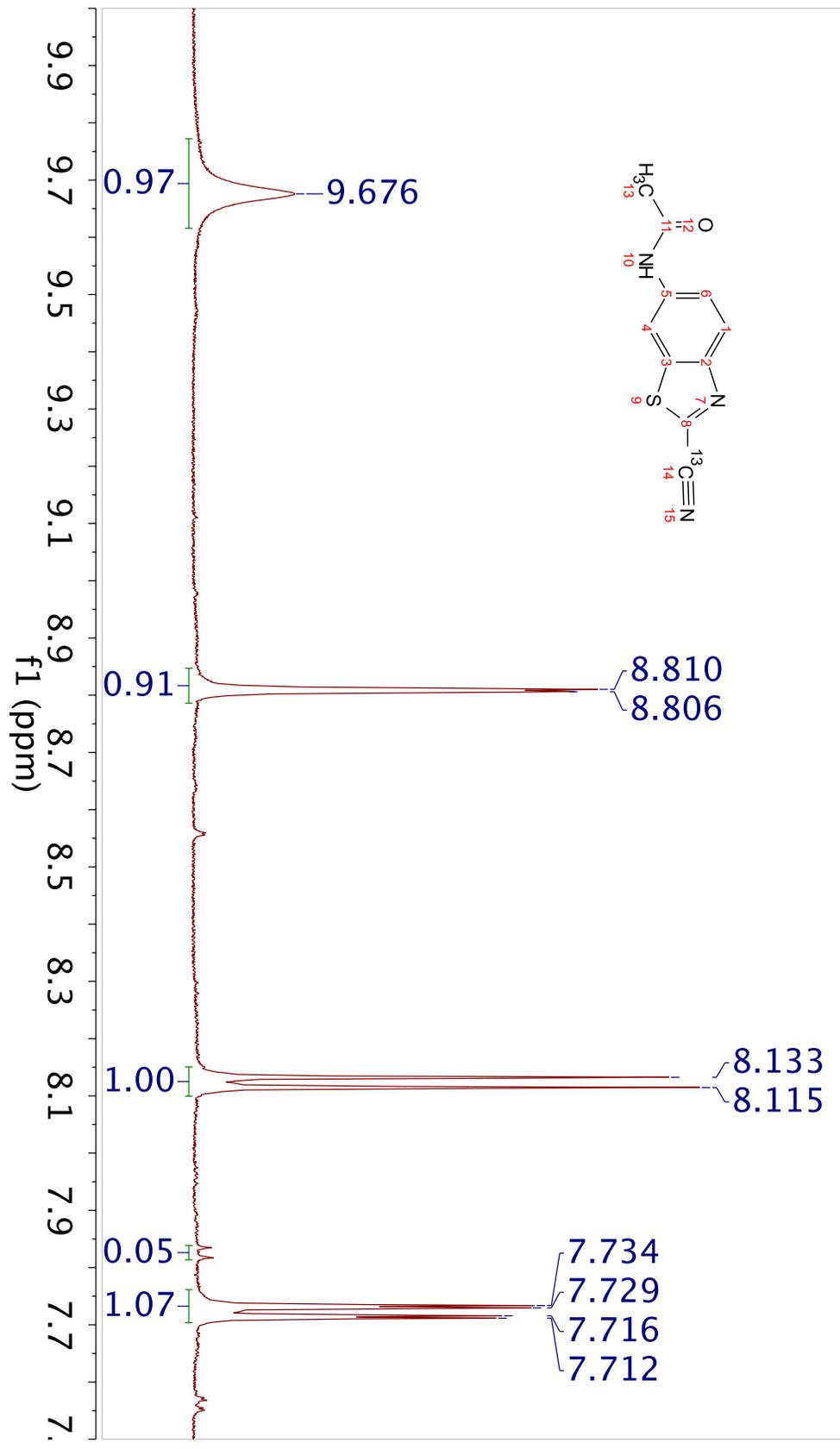
SGLK-i-CIBTNAC-TLCspot3polar-C13.1.fid
AVB-400 ZBO Carbon Starting paramters 6/11/03 RN



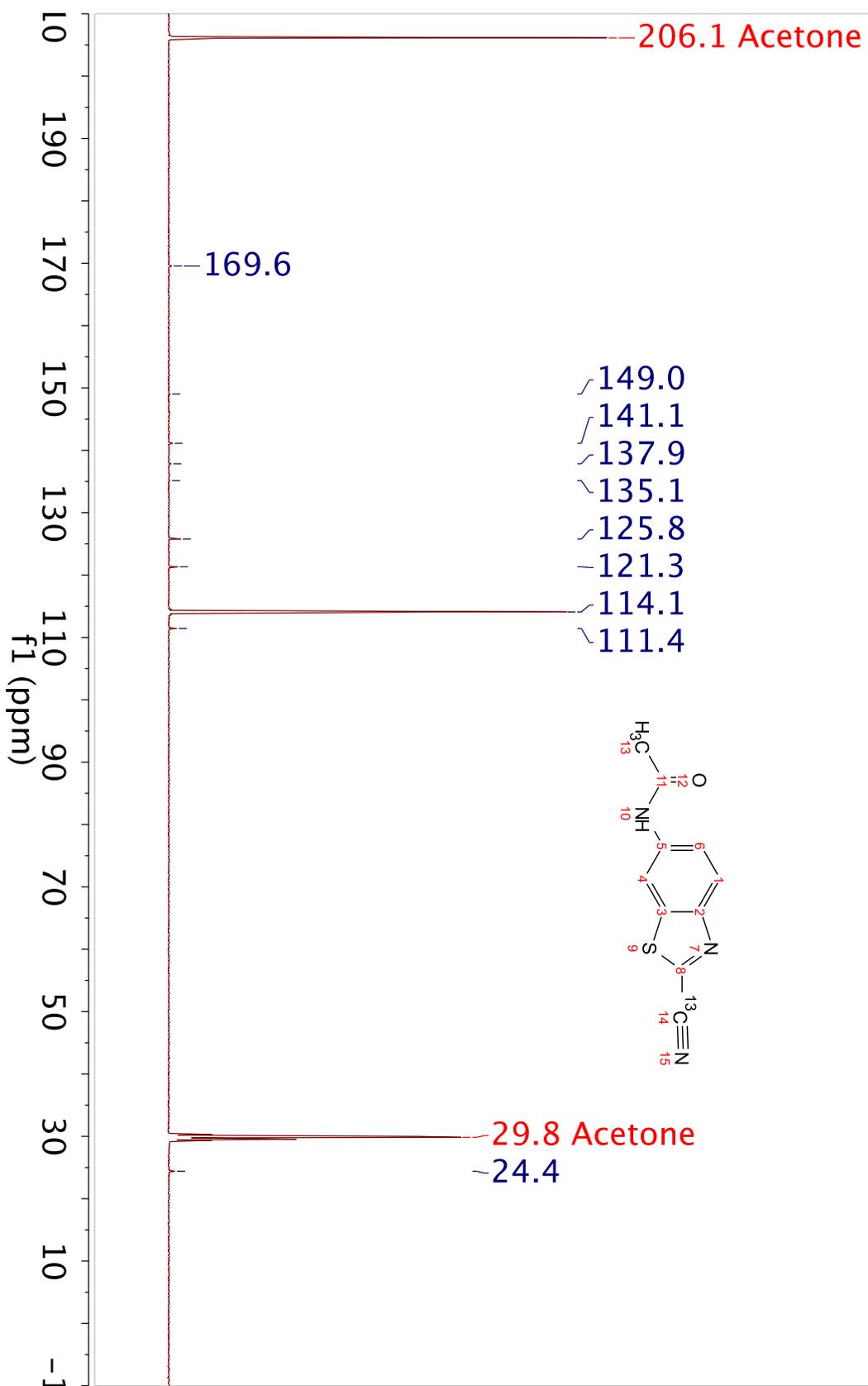
022318_13CBTNAc_meoh-shigemii_acetone_32s_proton
STANDARD PROTON PARAMETERS



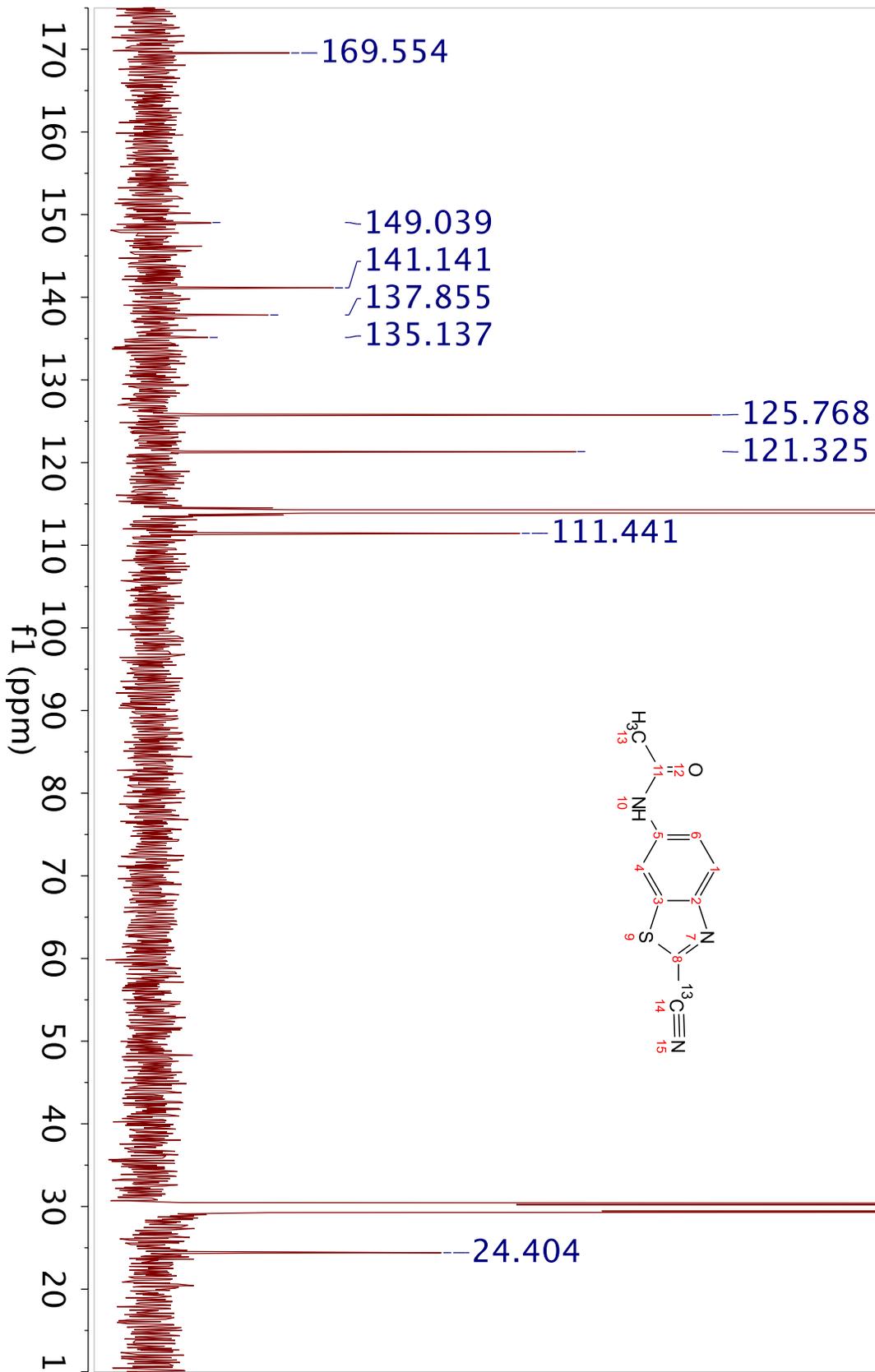
022318_13CBTNAc_meoh-shigemi_acetone_32s_proton
STANDARD PROTON PARAMETERS



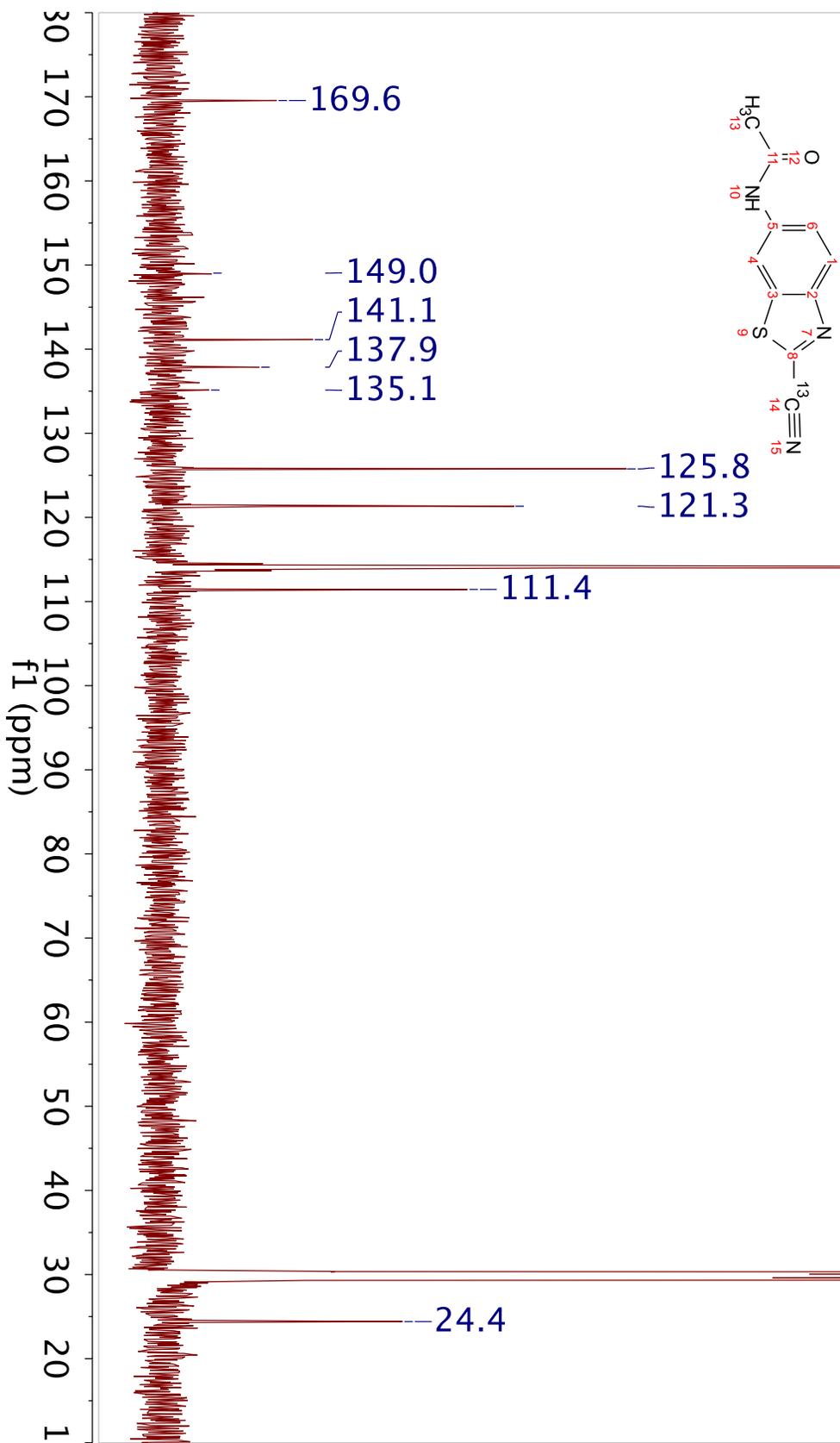
022318_13cbtncac_mcoh-shigem_i_acetone_d1-5s_13C
C13par



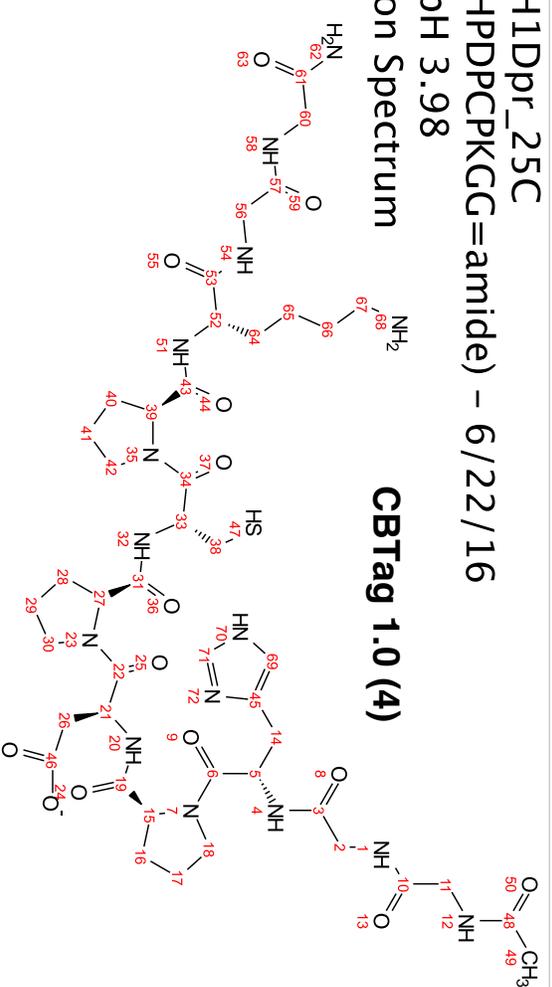
022318_13cbtncac_meoh-shigemi_acetone_d1-5s_13C
C13par



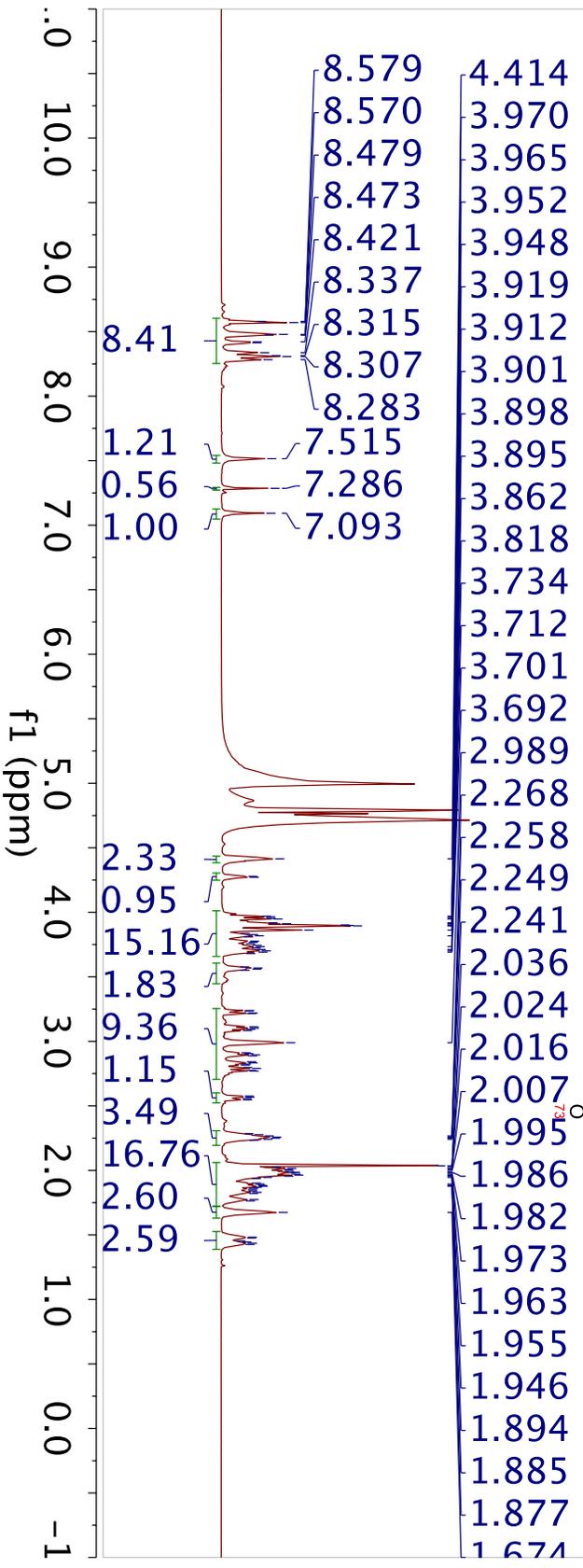
022318_13cbtncac_meoh-shigemii_acetone_d1-5s_13C
C13par



160622_PEPFOLD-3_1H1Dpr_25C
 PEPFOLD-3 (acetyl-GHPPDPCPKG=amide) - 6/22/16
 ~7.4 mM in 10% D2O, pH 3.98
 Presat Water Suppression Spectrum



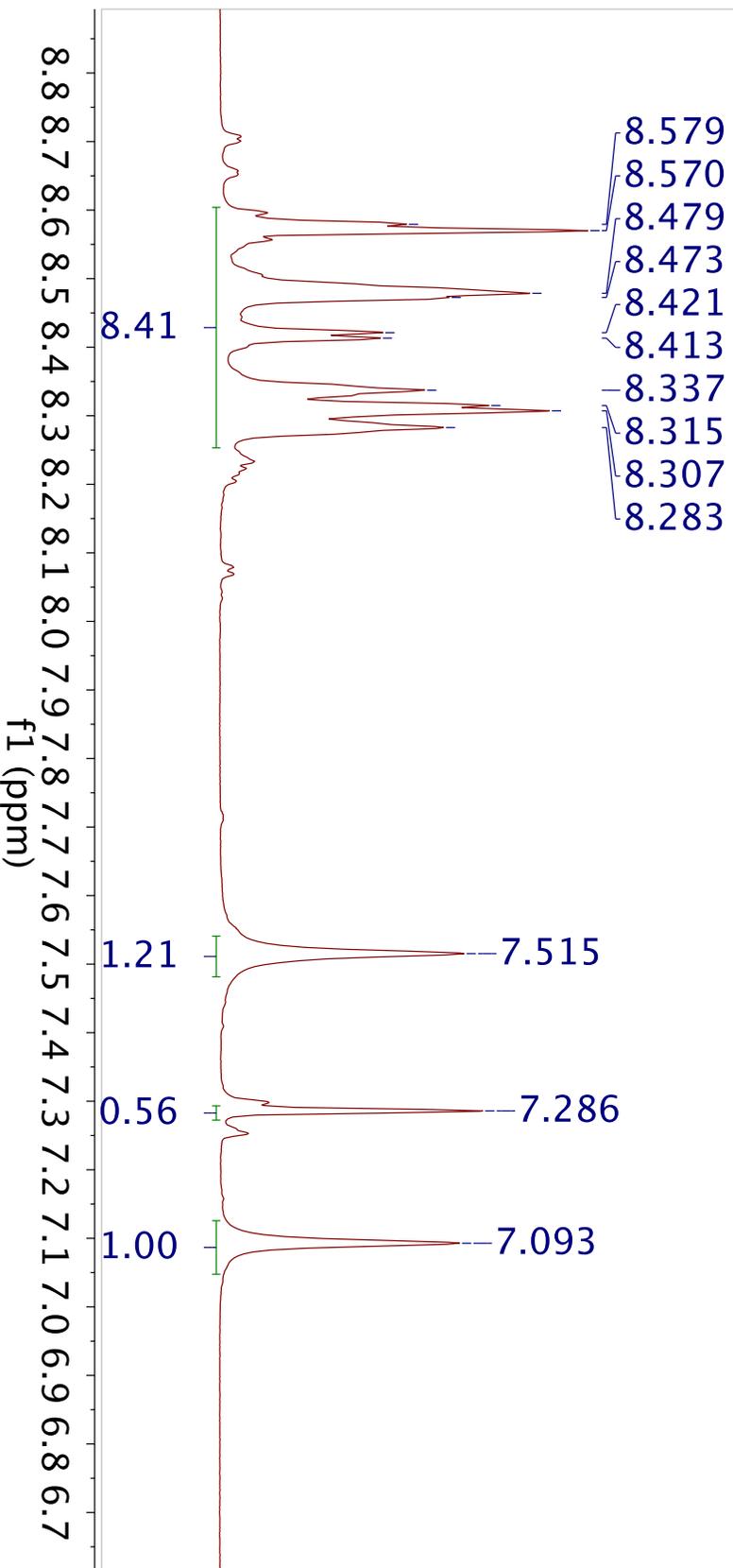
CBTag 1.0 (4)



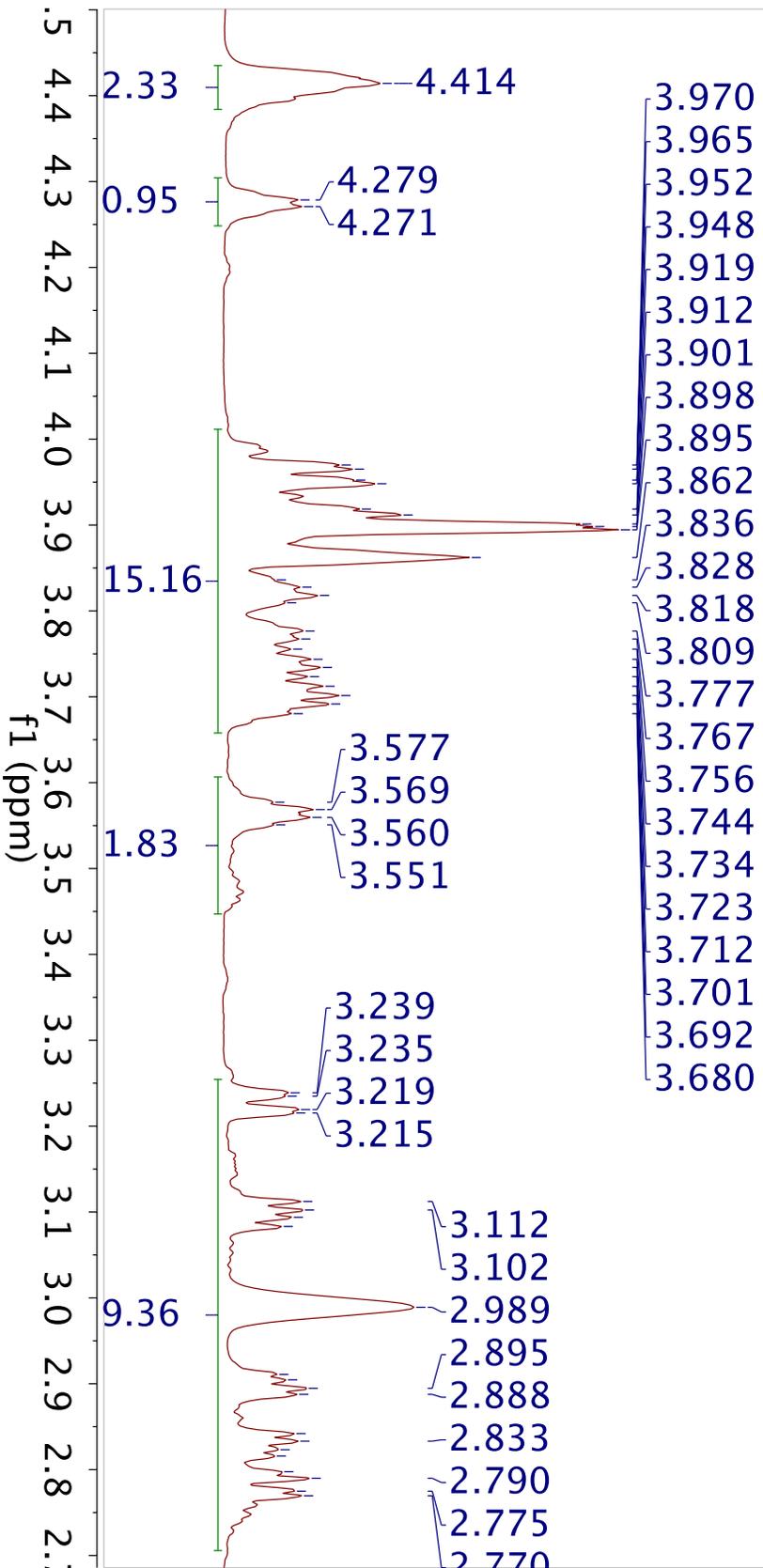
160622_PEPFOLD-3_1H1Dpr_25C
PEPFOLD-3 (acetyl-GGHPDPCPKGG=amide) - 6/22/16
~7.4 mM in 10% D2O, pH 3.98
Presat Water Suppression Spectrum



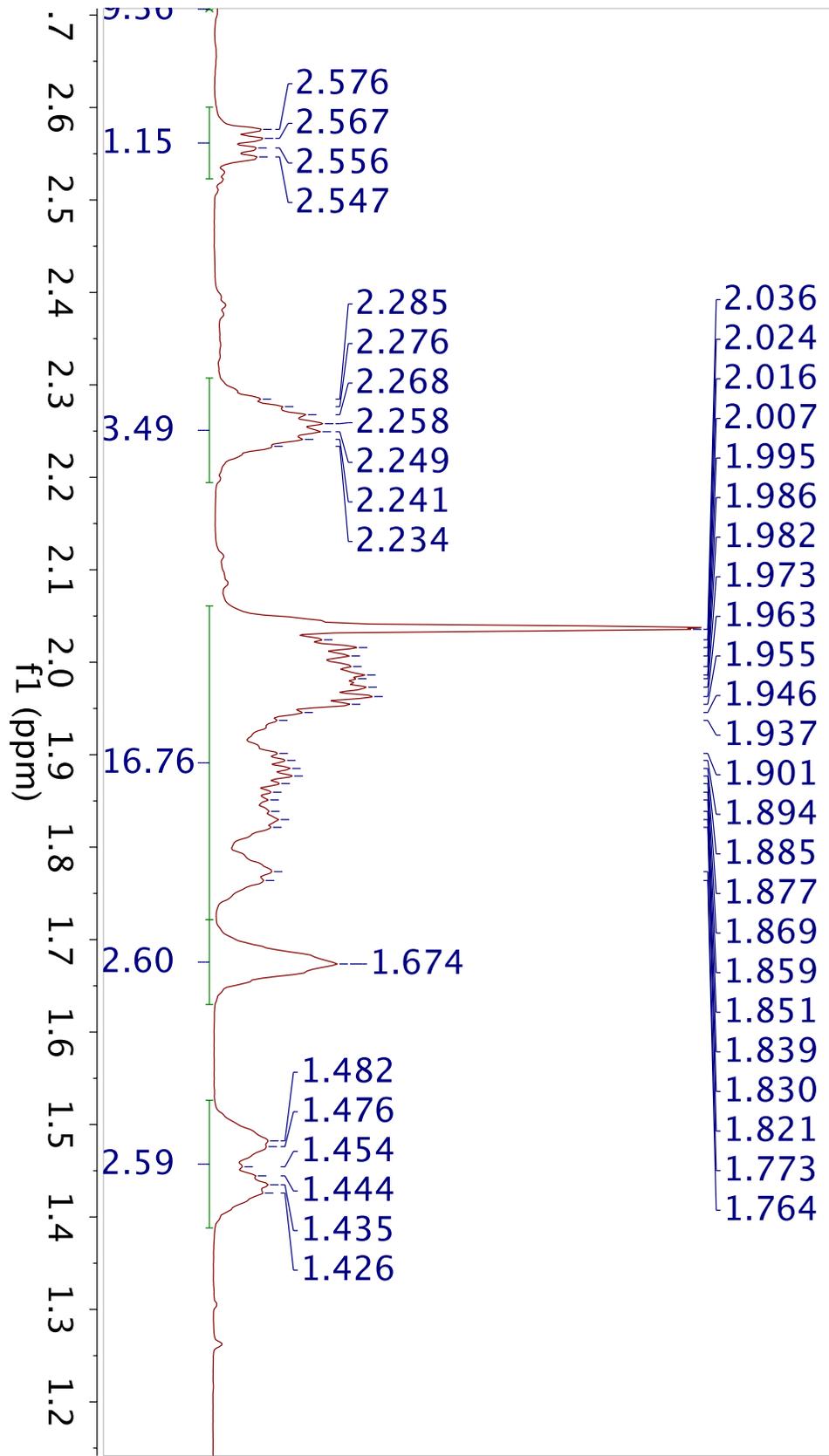
160622_PEPFOLD-3_1H1Dpr_25C
PEPFOLD-3 (acetyl-GGHPDPCPKGG=amide) - 6/22/16
~7.4 mM in 10% D2O, pH 3.98
Presat Water Suppression Spectrum



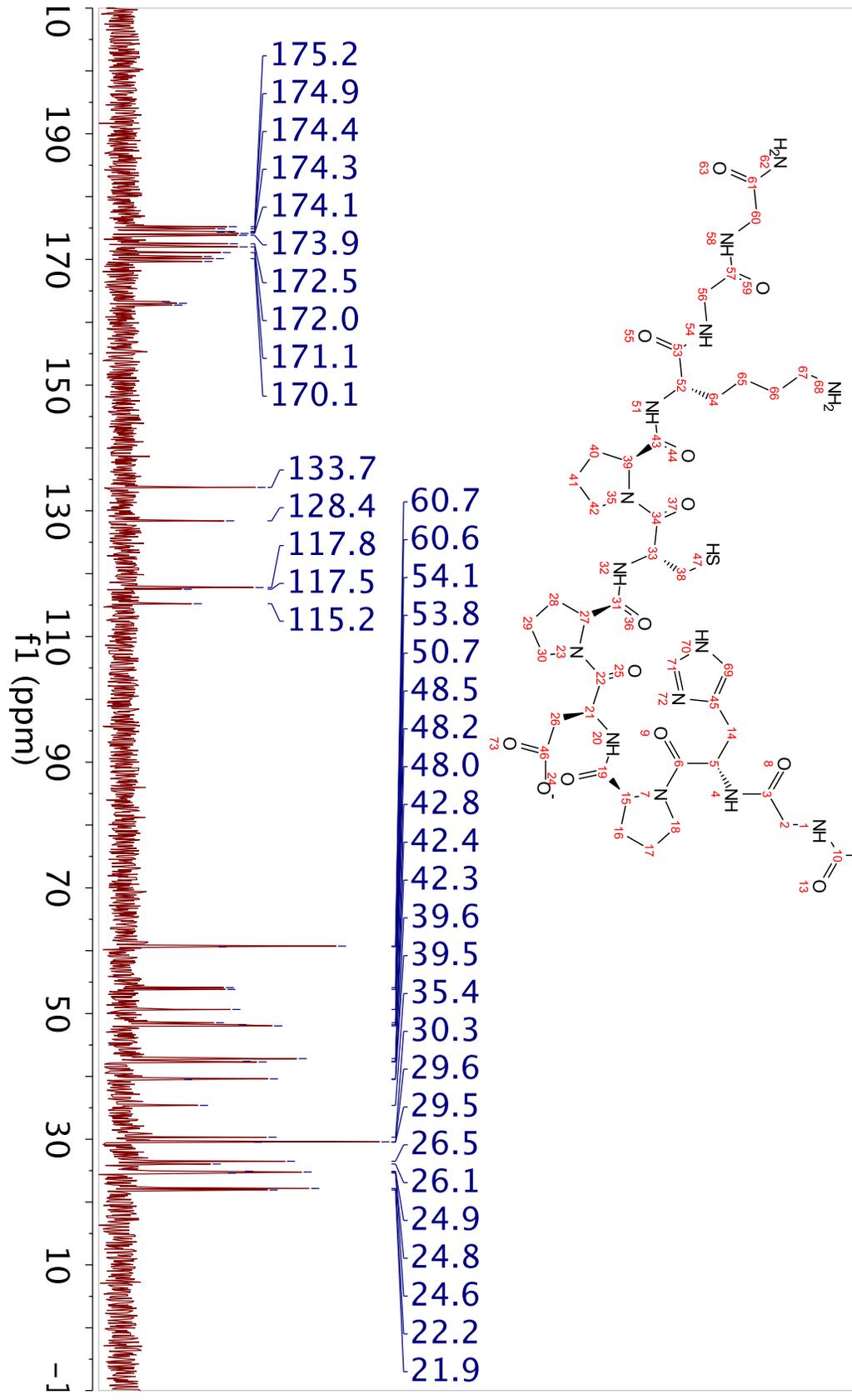
160622_PEPFOLD-3_1H1Dpr_25C
 PEPFOLD-3 (acetyl-GHPPDPCPKG=amide) - 6/22/16
 ~7.4 mM in 10% D2O, pH 3.98
 Presat Water Suppression Spectrum



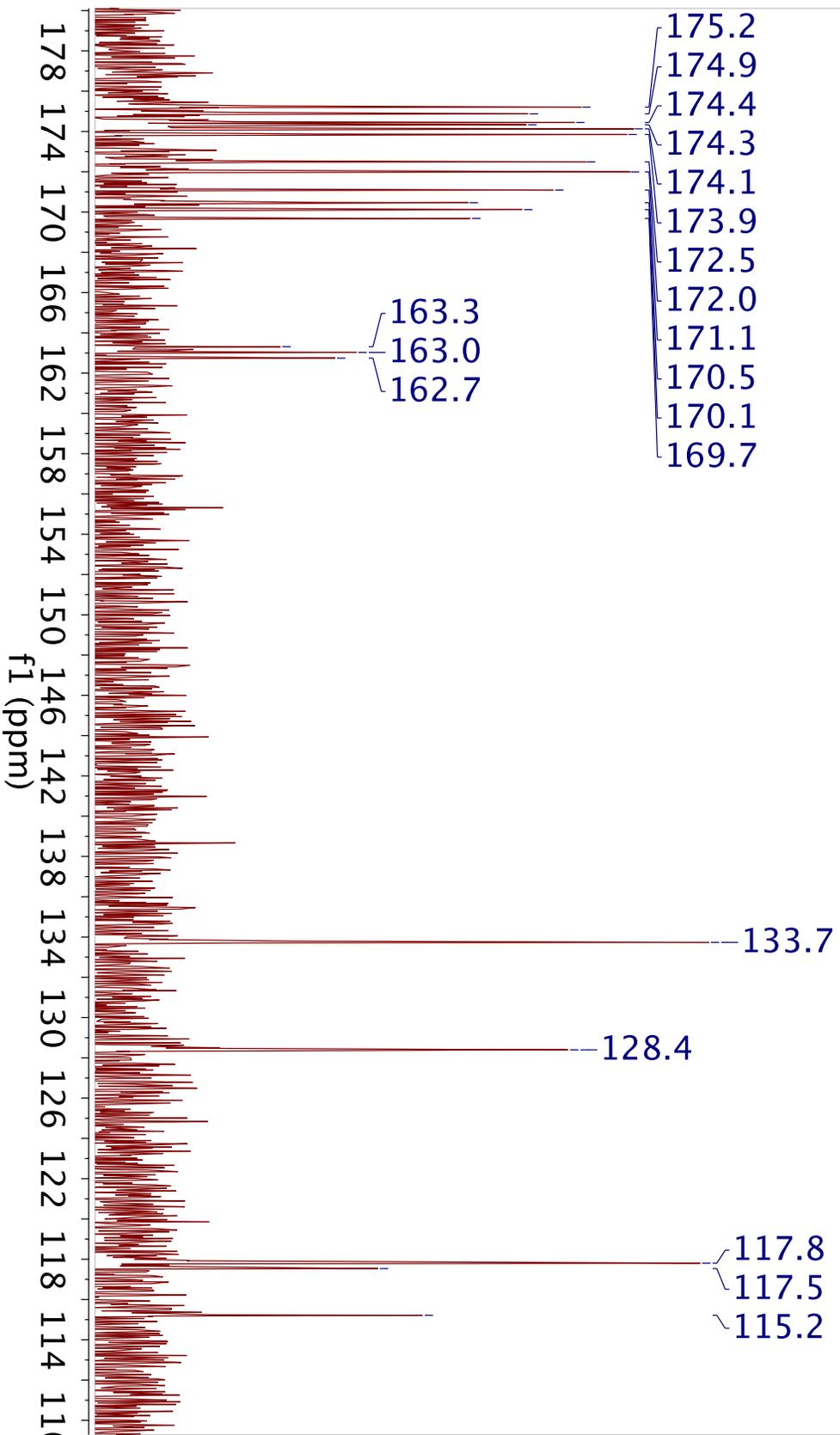
160622_PEPFOLD-3_1H1Dpr_25C
PEPFOLD-3 (acetyl-GHPPDPCPKG=amide) - 6/22/16
~7.4 mM in 10% D2O, pH 3.98
Presat Water Suppression Spectrum



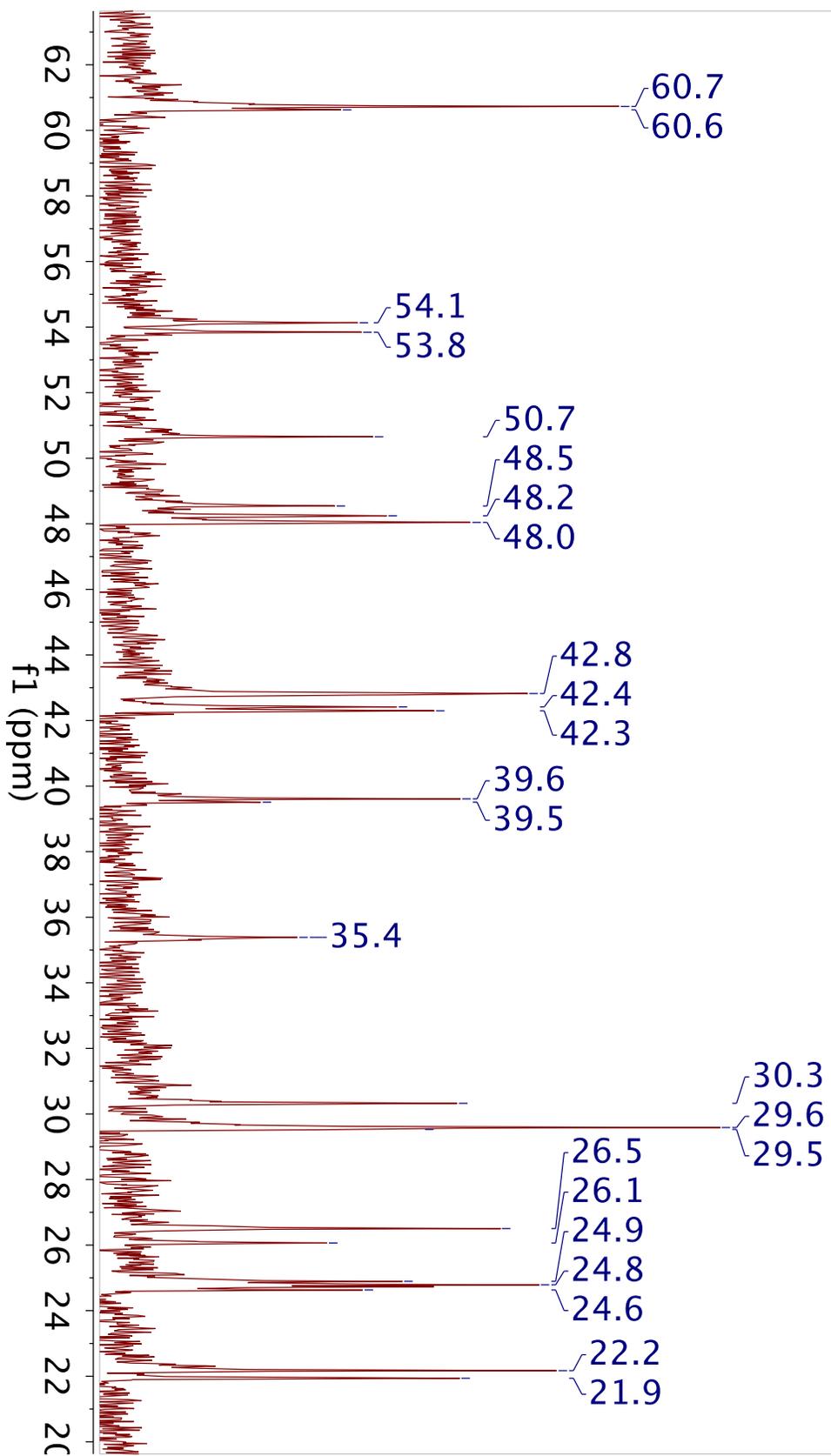
060816_PEPFOLD-3_100mM_water_13C_240scans
C13par



060816_PEPFOLD-3_100mM_water_13C_240scans
C13par

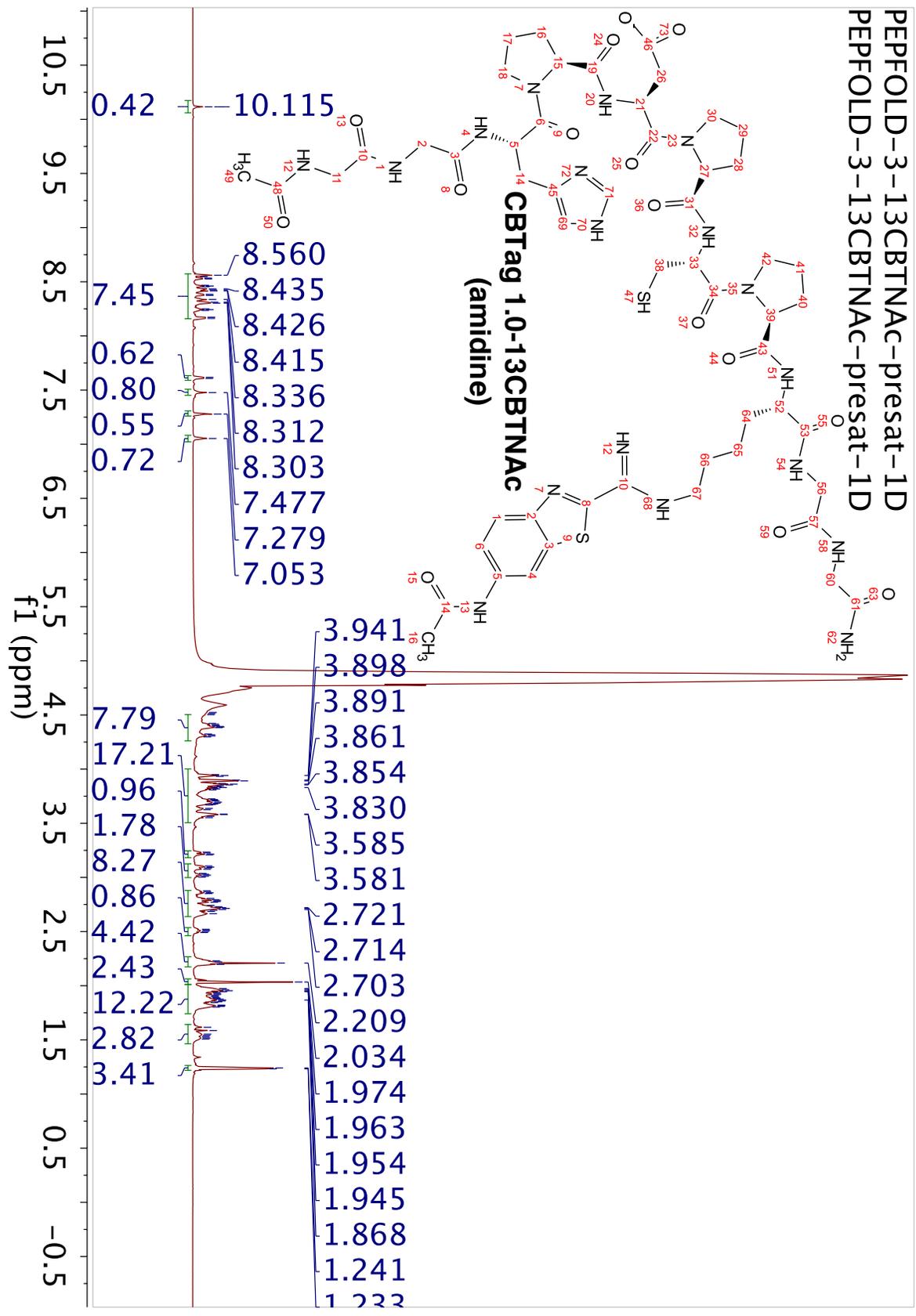


060816_PEPFOLD-3_100mM_water_13C_240scans
C13par

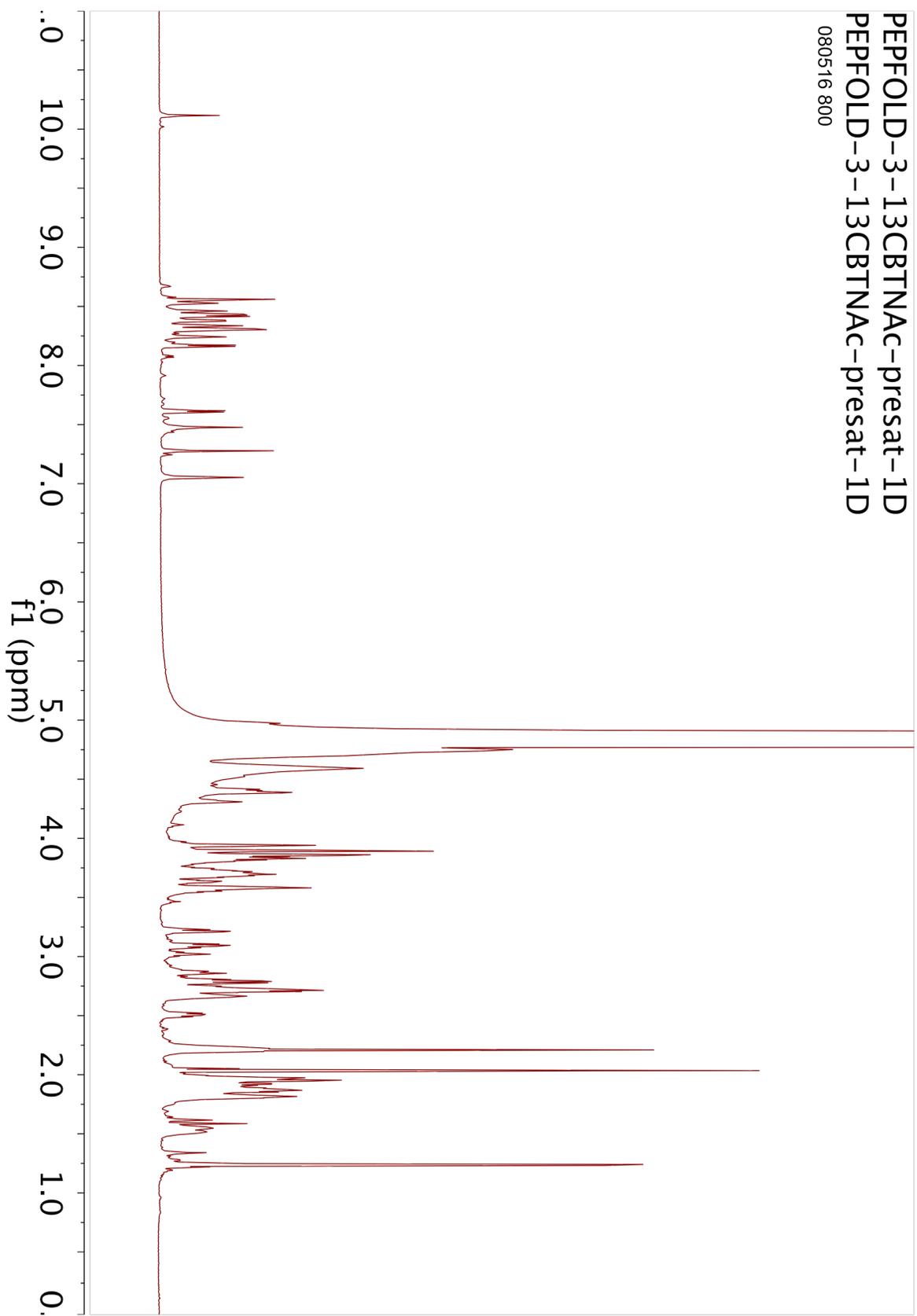


PEPFOLD-3-13CBTNAC-presat-1D
 PEPFOLD-3-13CBTNAC-presat-1D

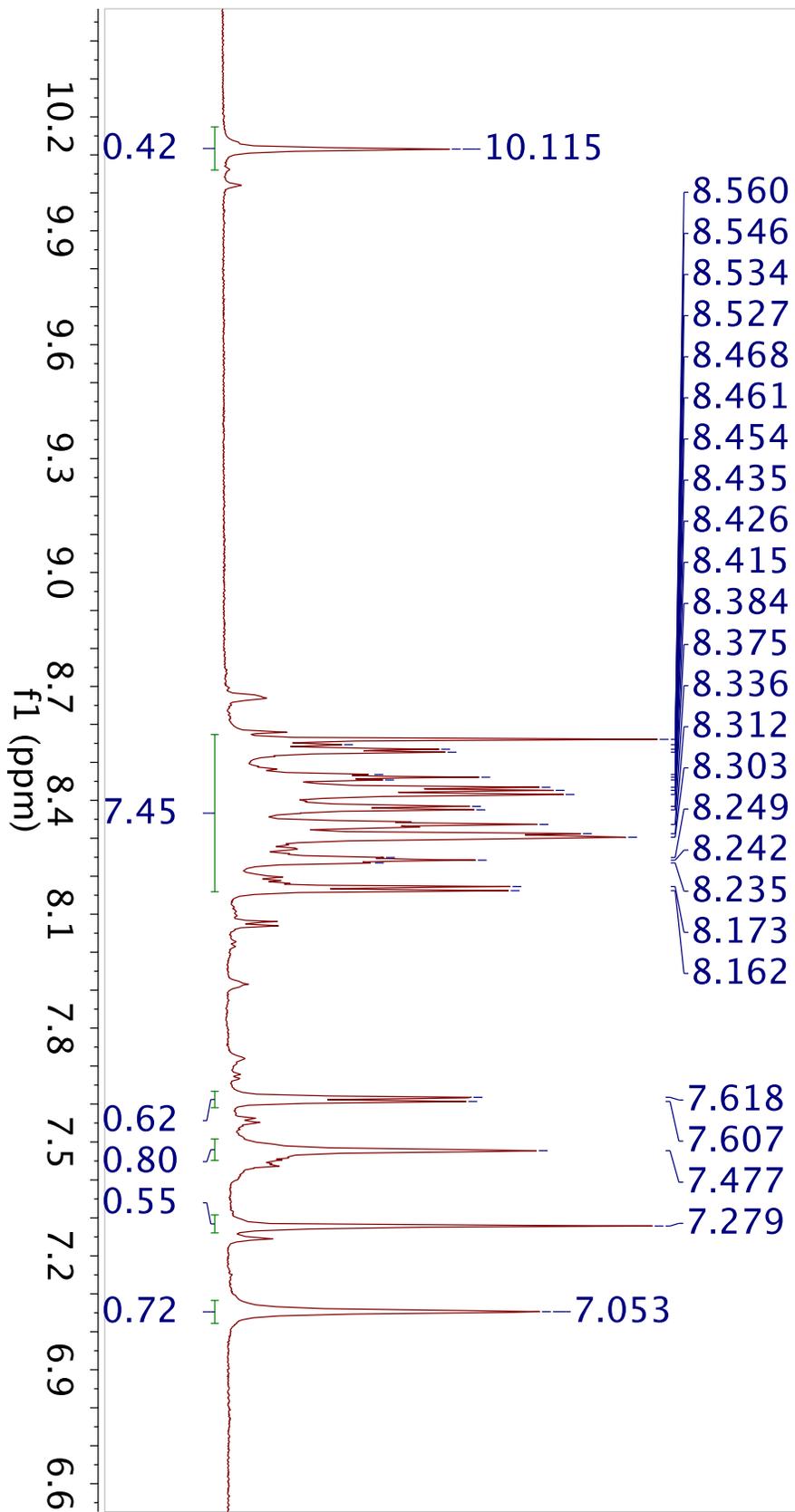
CBTag 1.0-13CBTNAC
 (amidine)



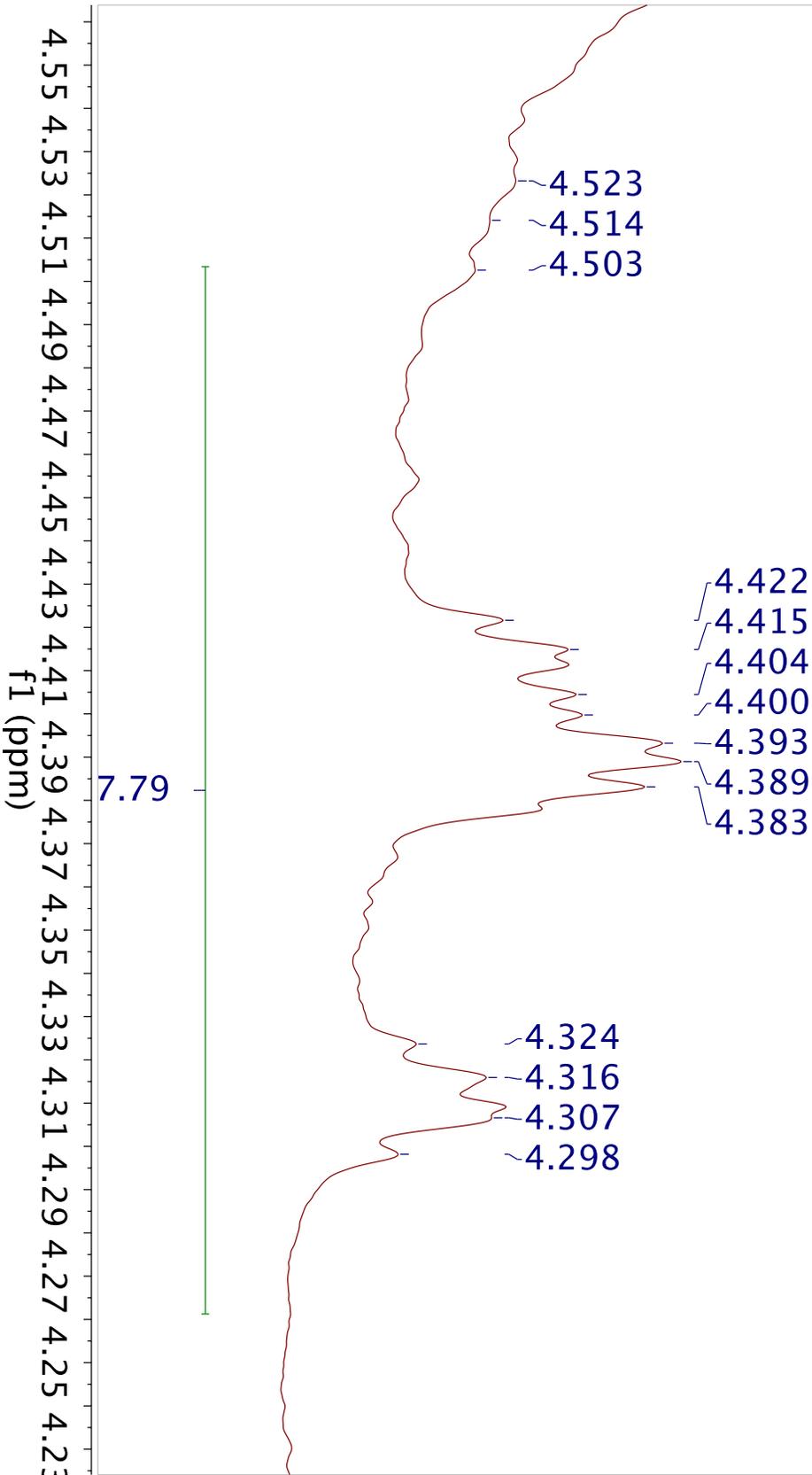
PEPFOLD-3-13CBTNAC-presat-1D
PEPFOLD-3-13CBTNAC-presat-1D
080516 800

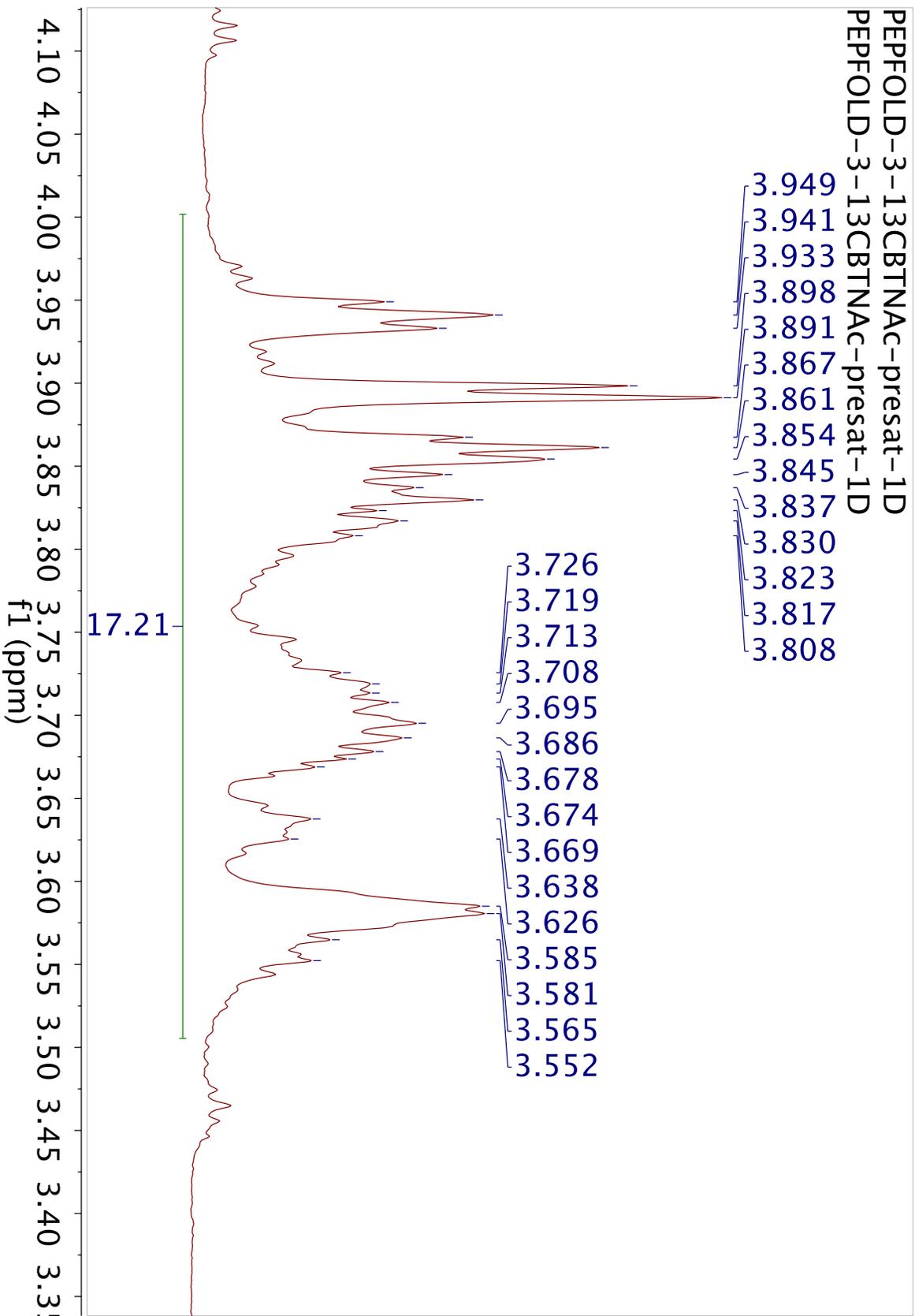


PEPFOLD-3-13CBTNAc-presat-1D
PEPFOLD-3-13CBTNAc-presat-1D

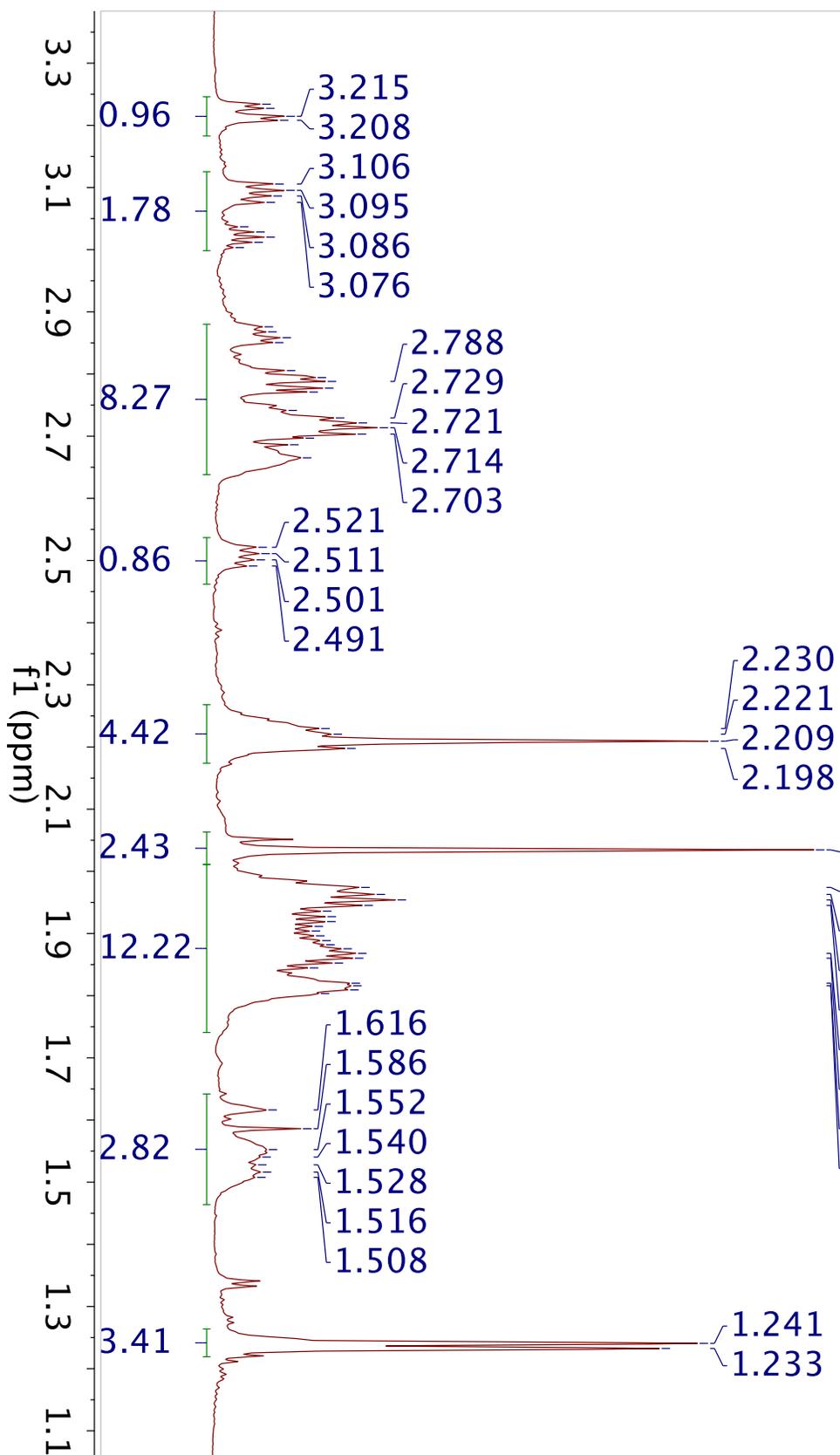


PEPFOLD-3-13CBTNAC-presat-1D
PEPFOLD-3-13CBTNAC-presat-1D

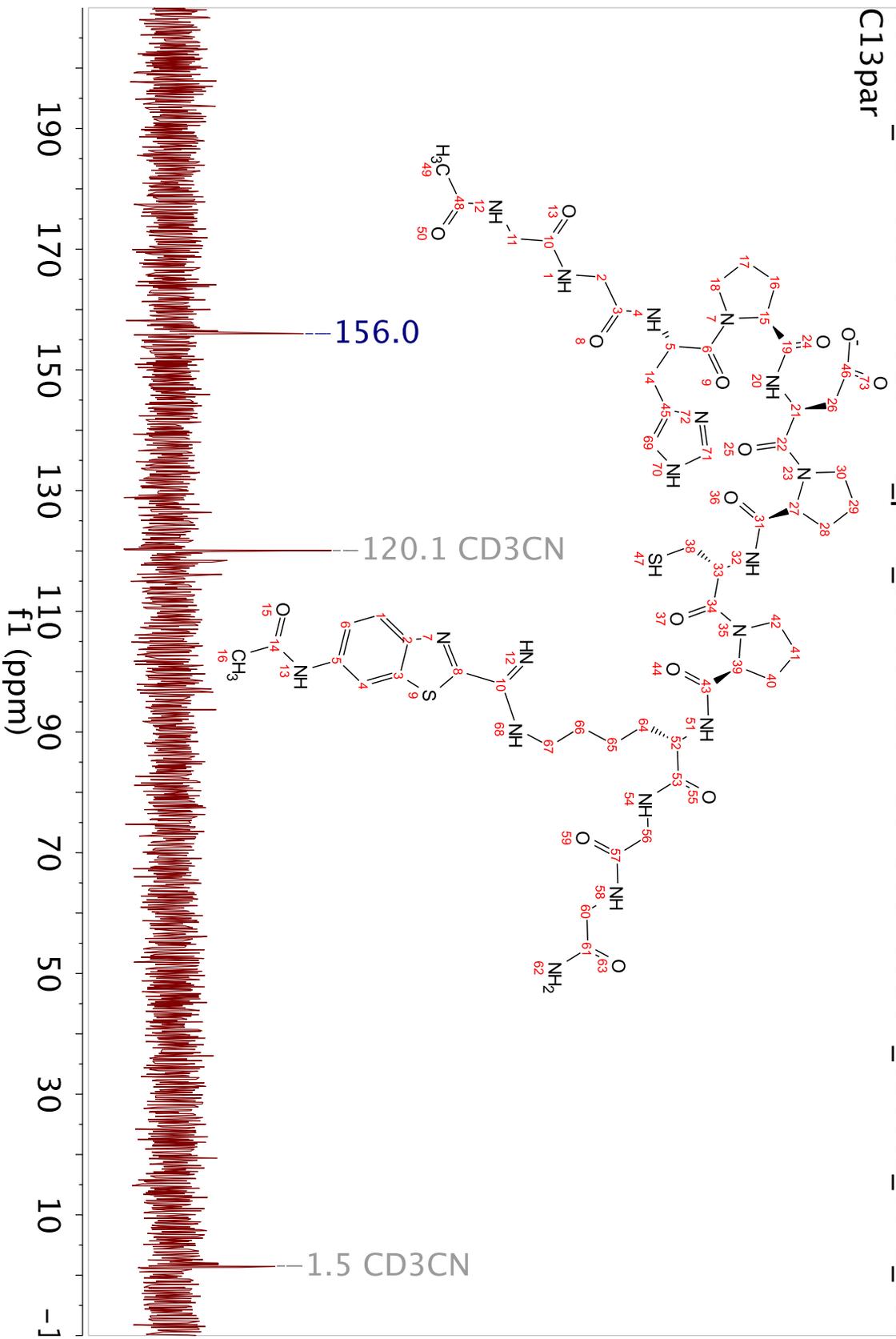




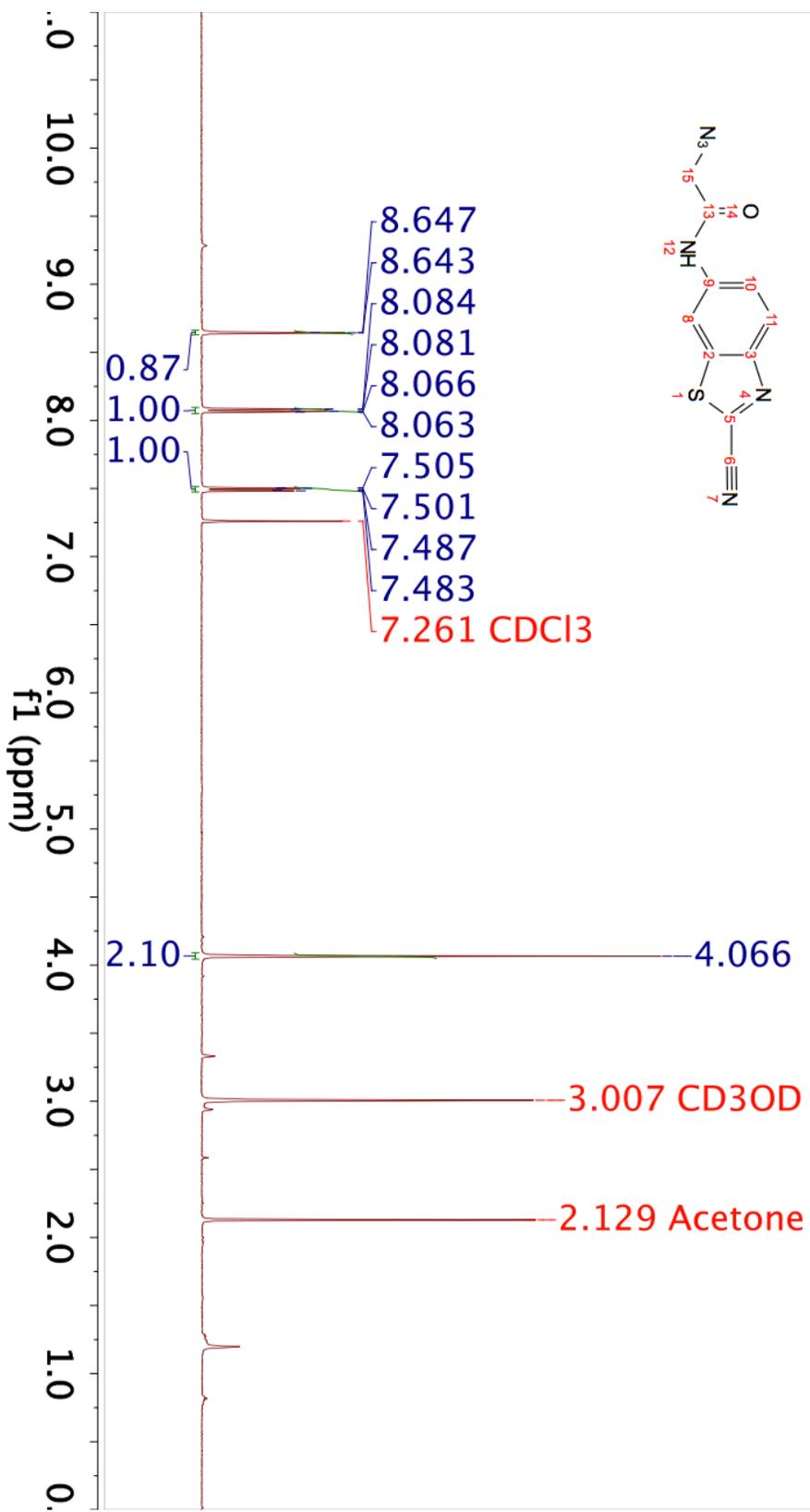
PEFOLD-3-13CBTNAC-presat-1D
PEFOLD-3-13CBTNAC-presat-1D



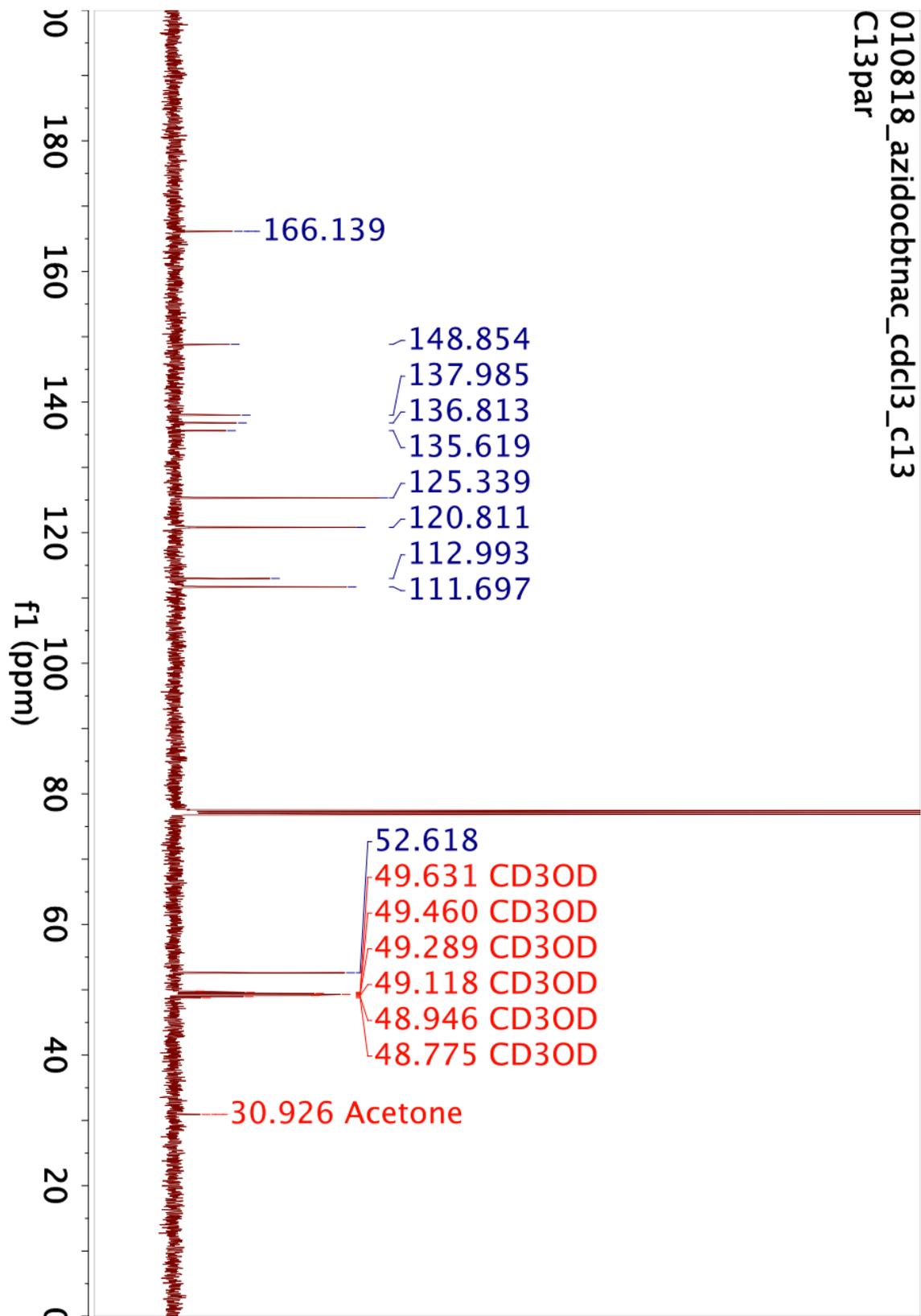
030118_PEPFOLD-3-13CBTNAC_pure_90-10-2uL-H2O-D2O-CD3CN_12000s_d1-2_13C
C13par



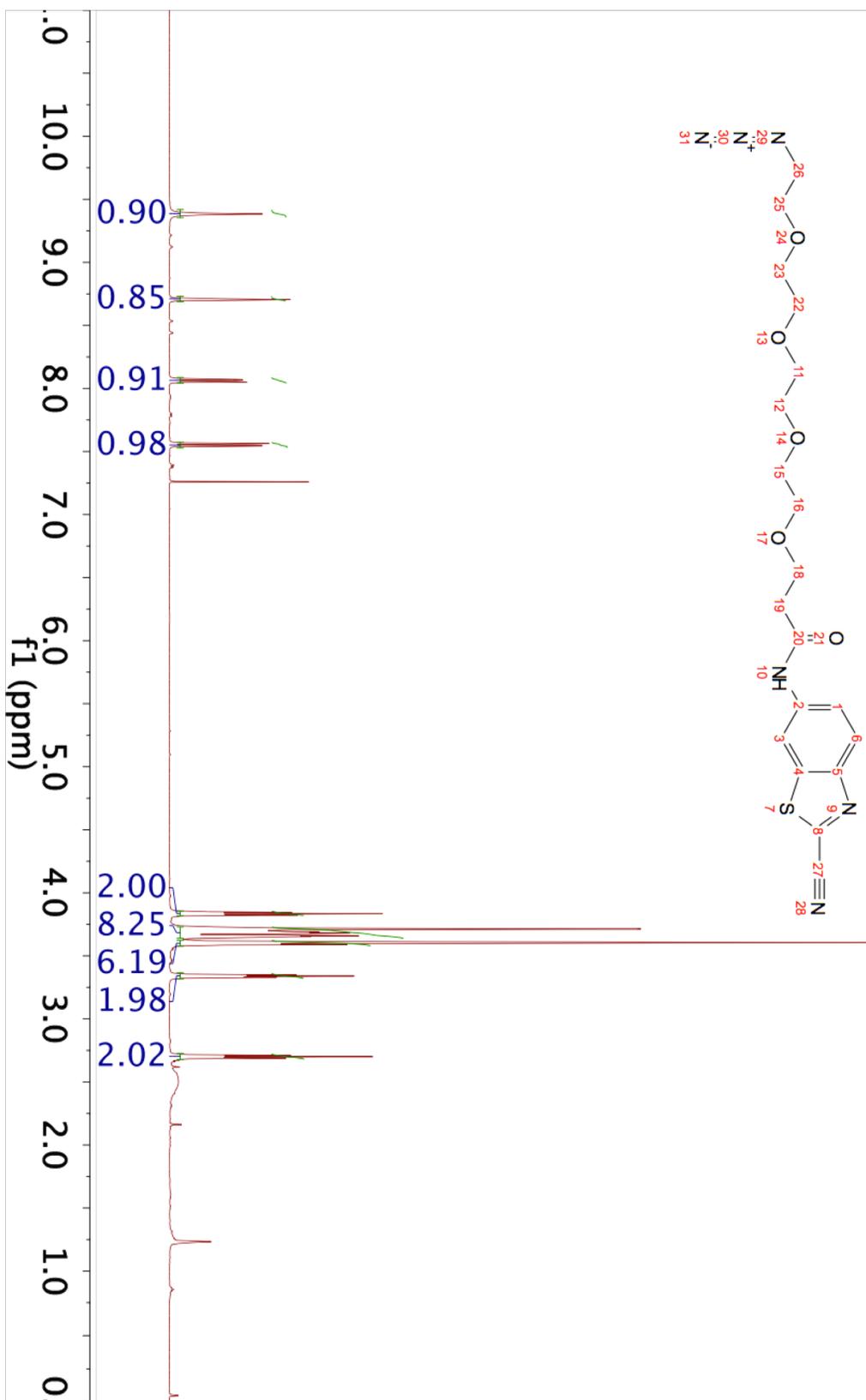
010818_azidoctnac_cdcl3_proton
STANDARD PROTON PARAMETERS



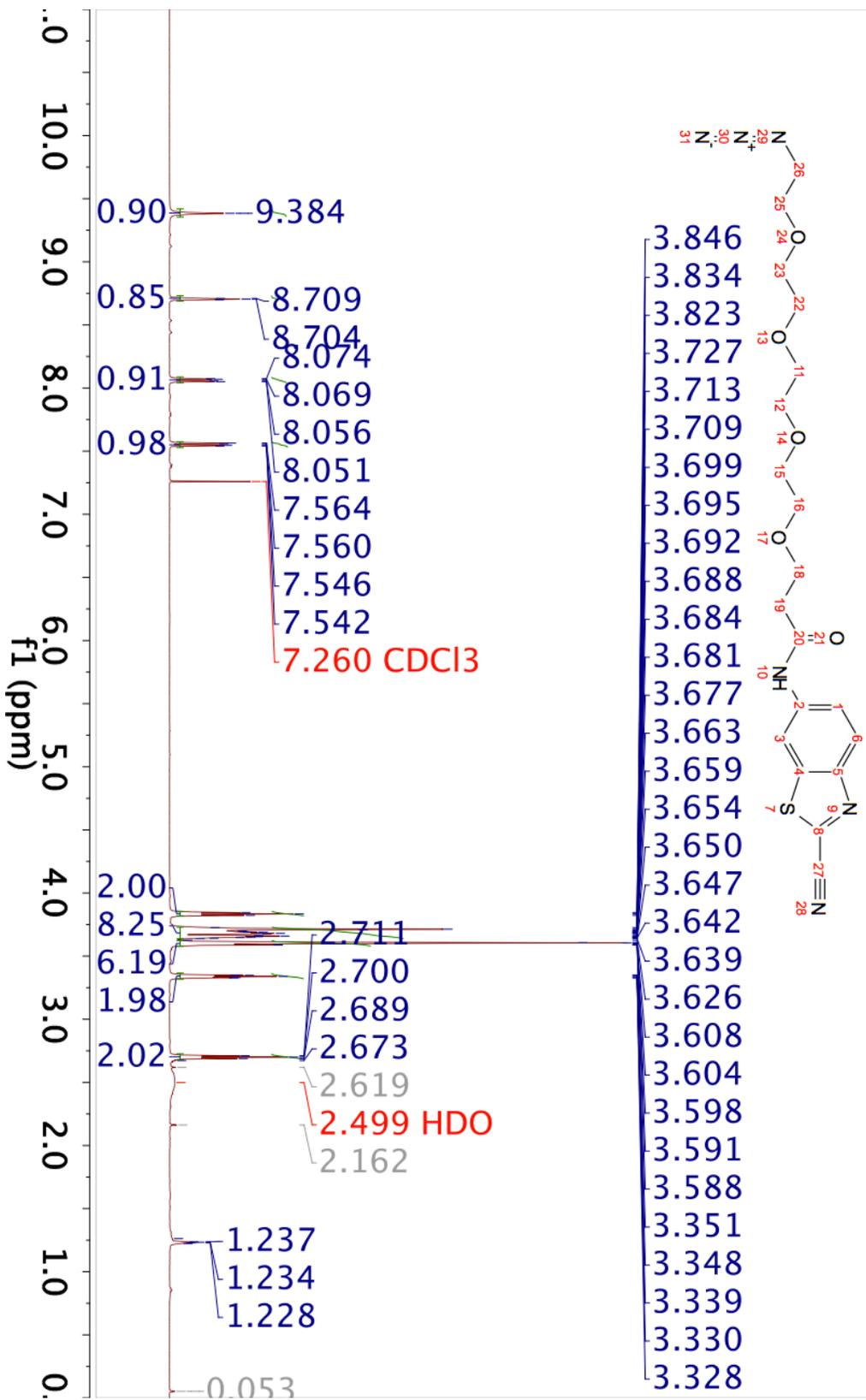
010818_azidobctnac_cdcl3_c13
C13par



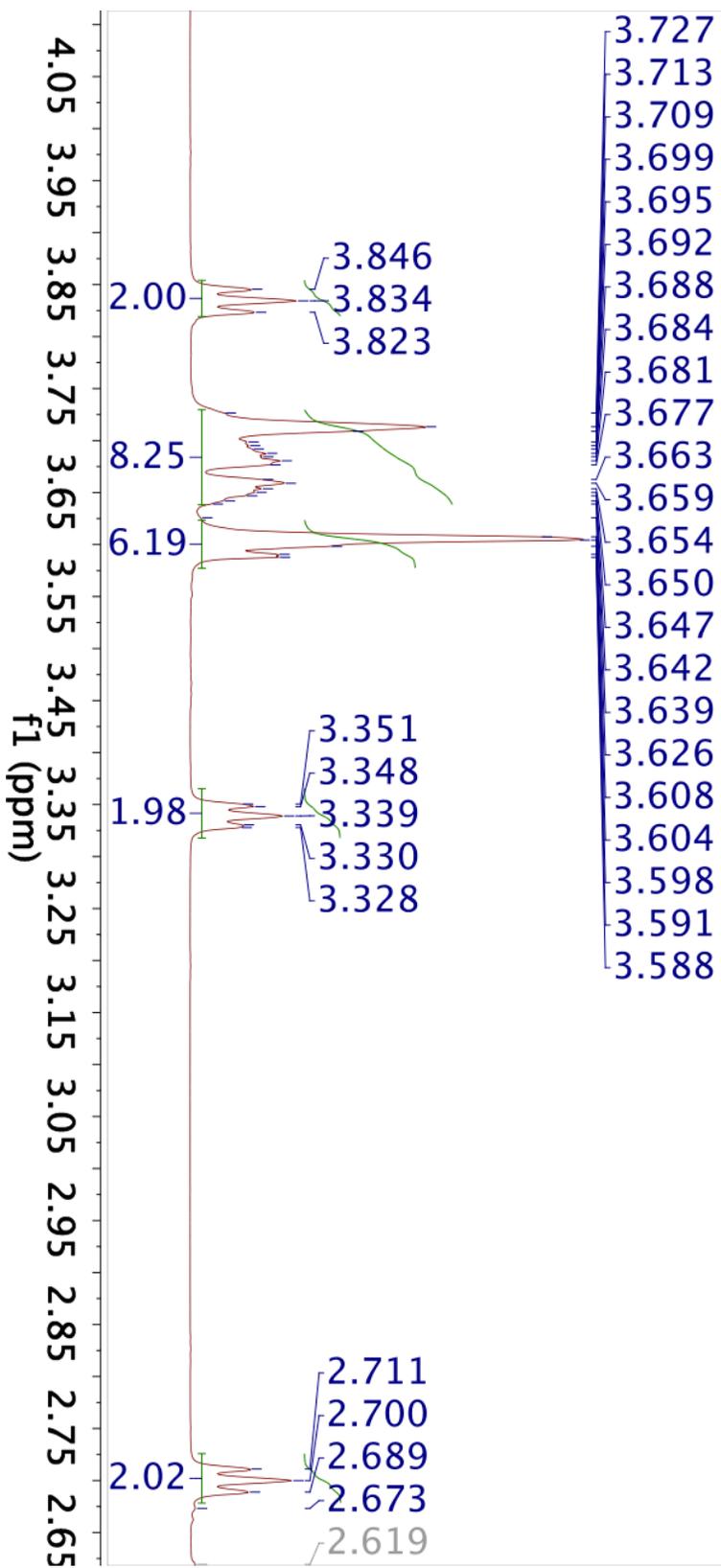
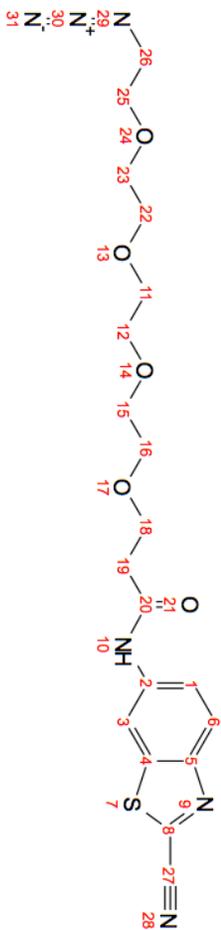
110317_CBT-PEG4-azide_CDCl3_proton
STANDARD PROTON PARAMETERS



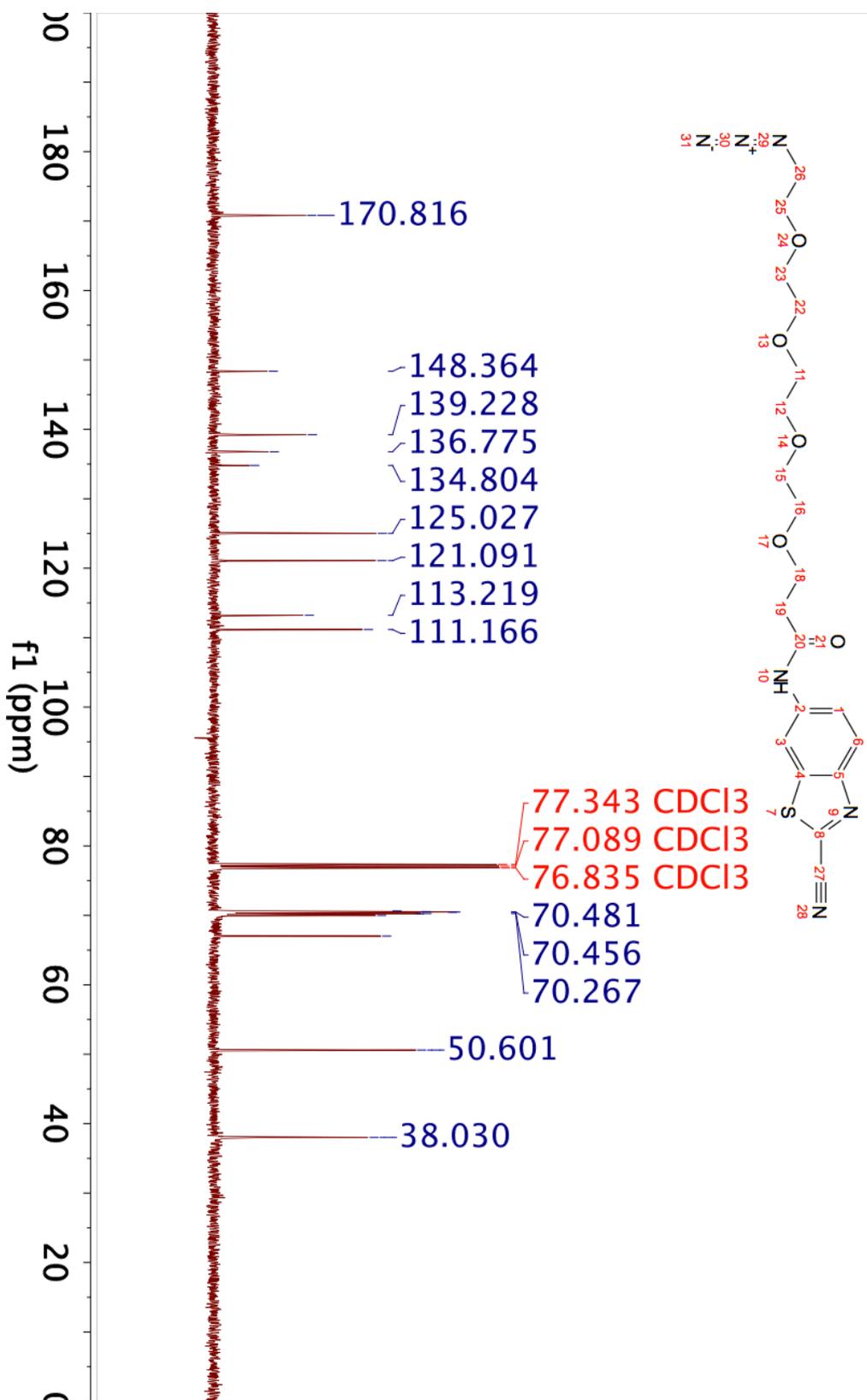
110317_CBT-PEG4-azide_CDCI3_proton
 STANDARD PROTON PARAMETERS



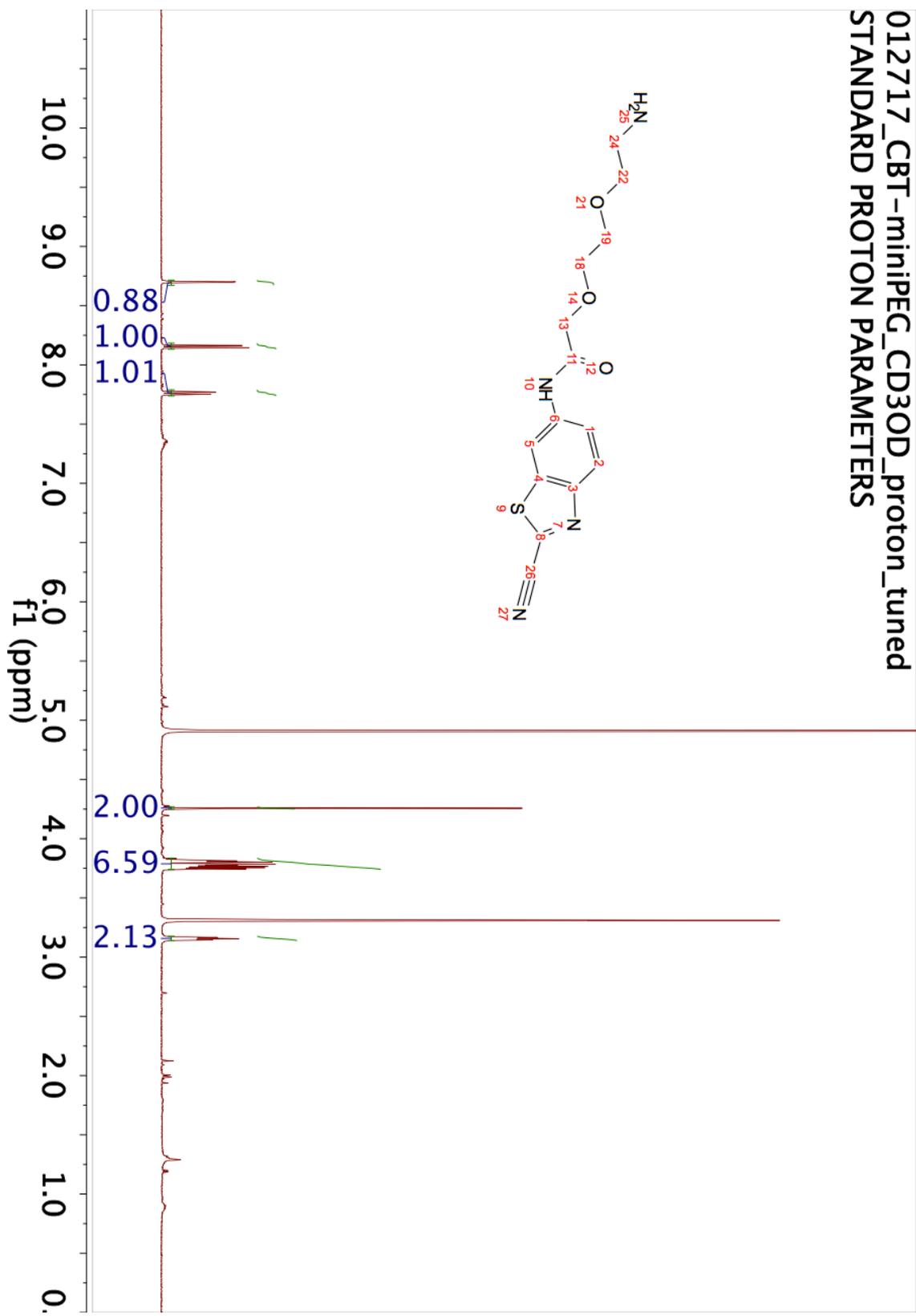
110317_CBT-PEG4-azide_CDCl3_proton
 STANDARD PROTON PARAMETERS



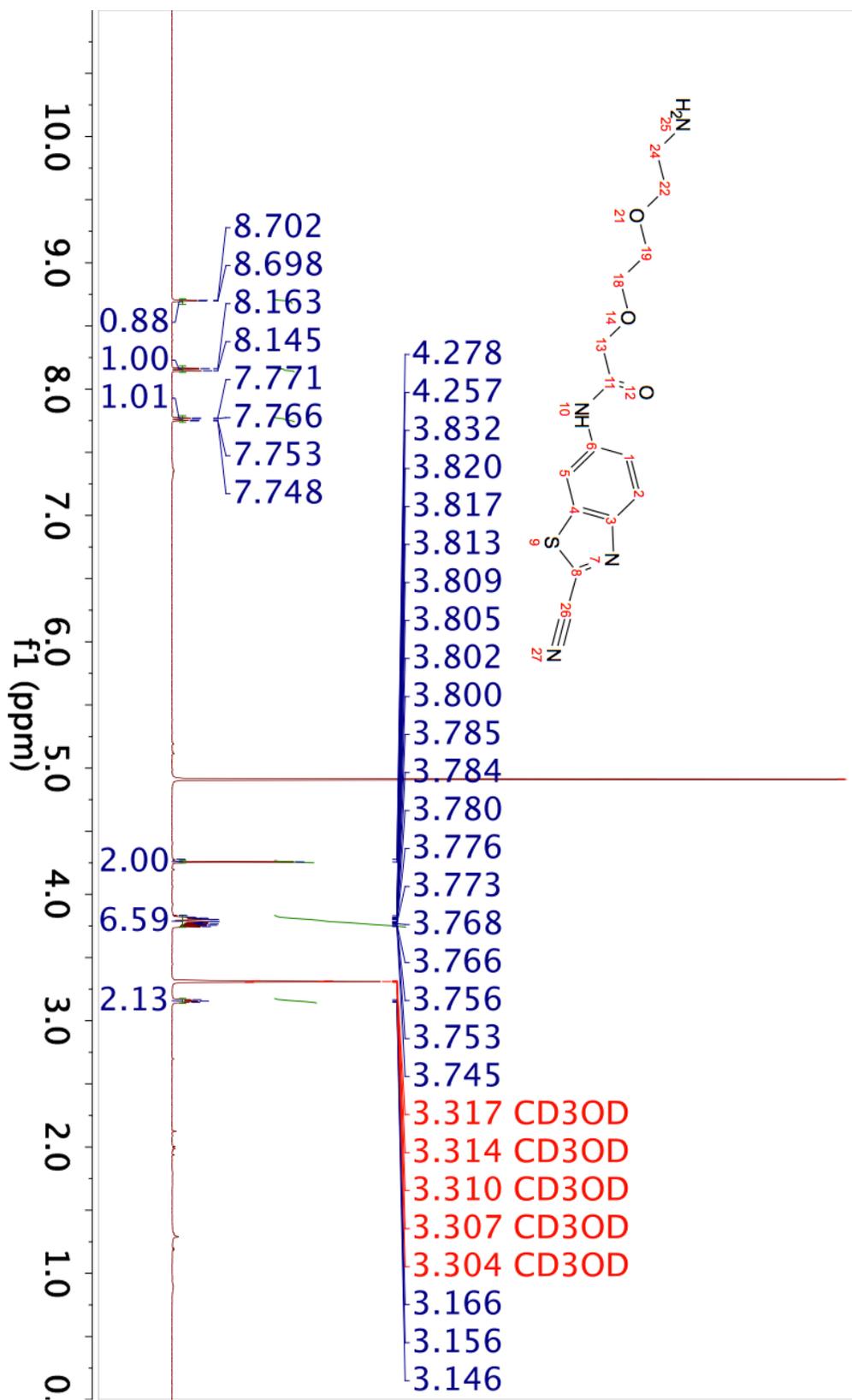
110317_CBT-PEG4-azide_CDCl3_13C
C13par



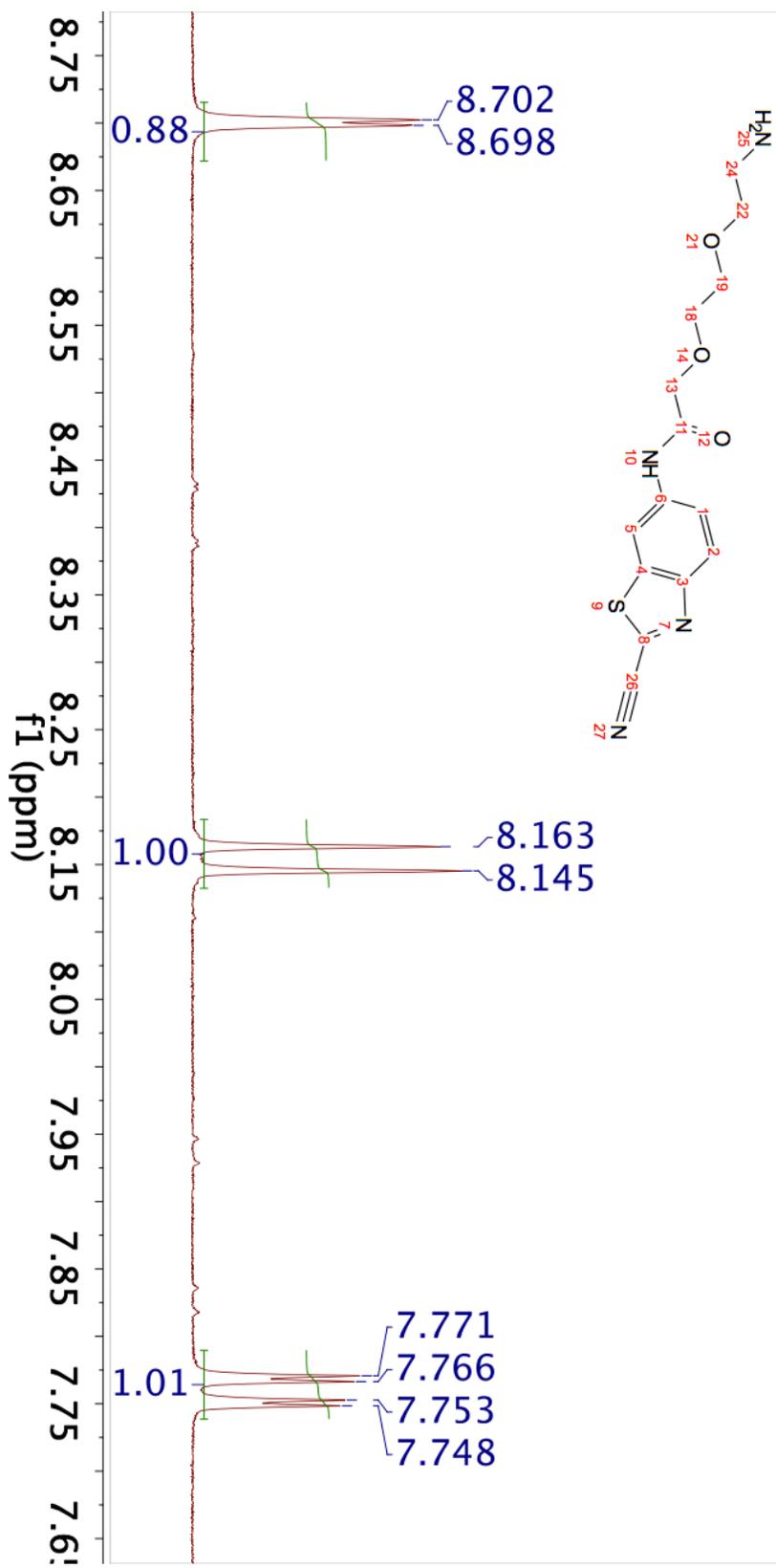
012717_CBT-miniPEG_CD3OD_proton_tuned
STANDARD PROTON PARAMETERS



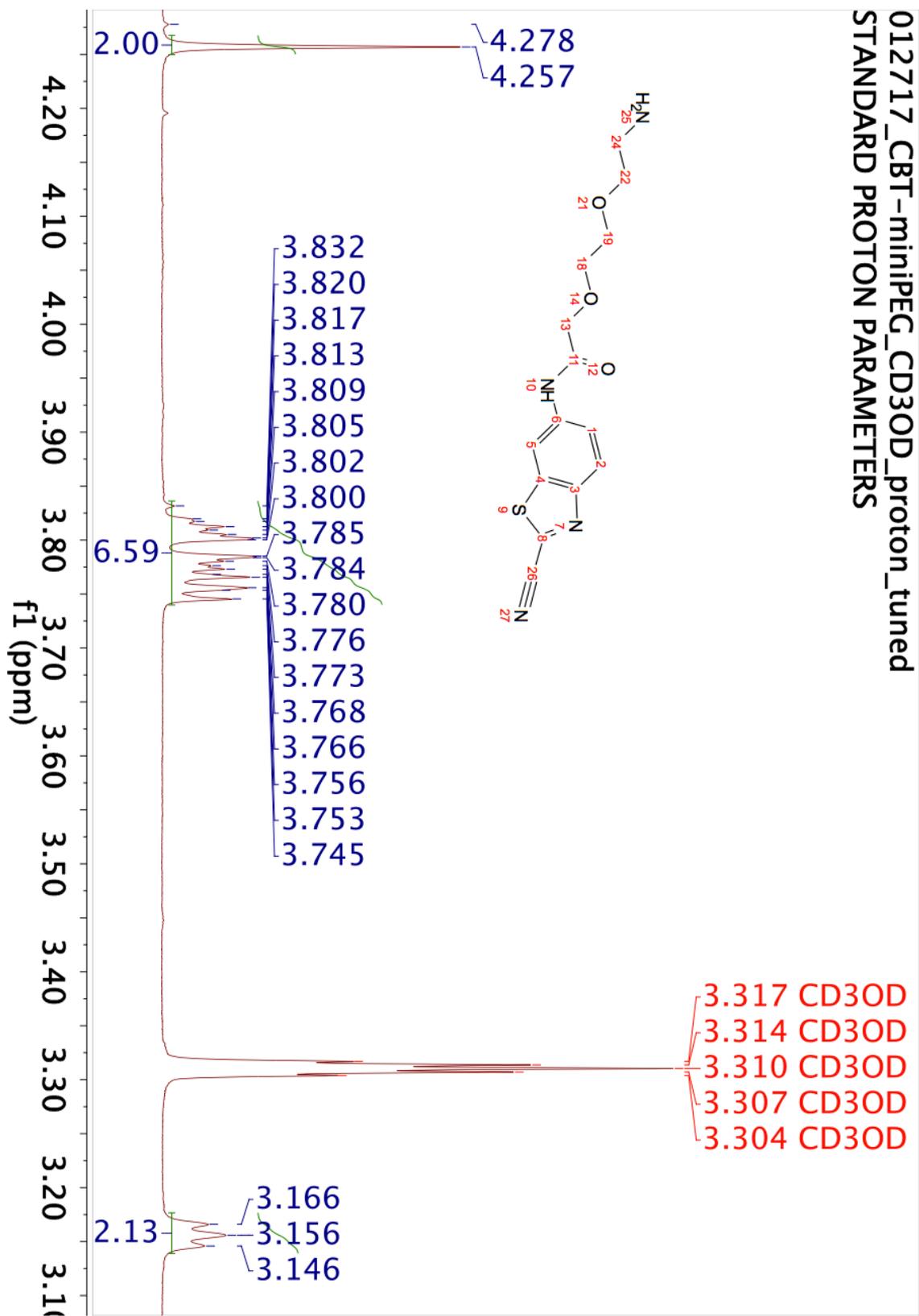
012717_CBT-miniPEG_CD3OD_proton_tuned
 STANDARD PROTON PARAMETERS



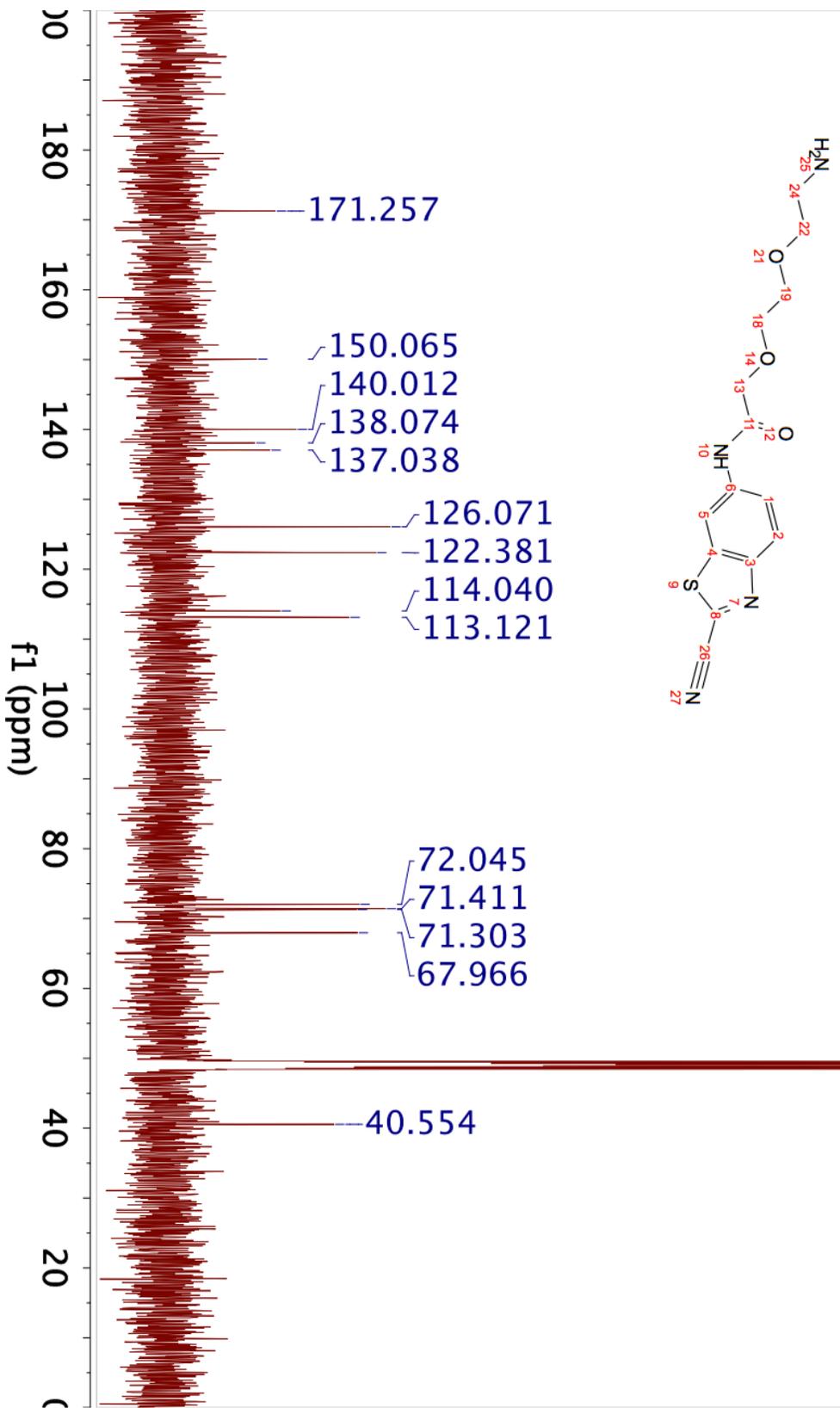
012717_CBT-miniPEG_CD3OD_proton_tuned
STANDARD PROTON PARAMETERS



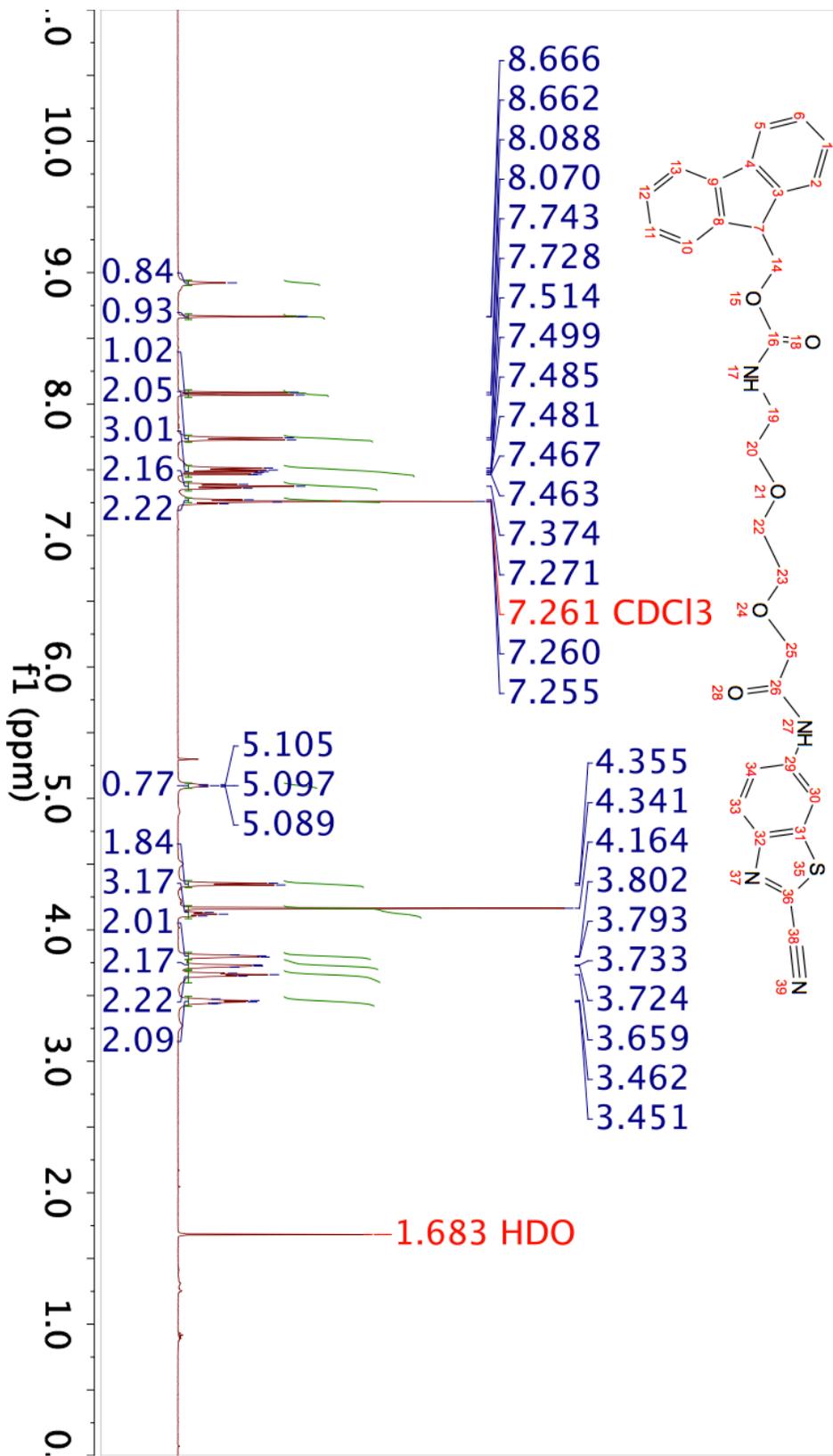
012717_CBT-miniPEG_CD3OD_proton_tuned
 STANDARD PROTON PARAMETERS



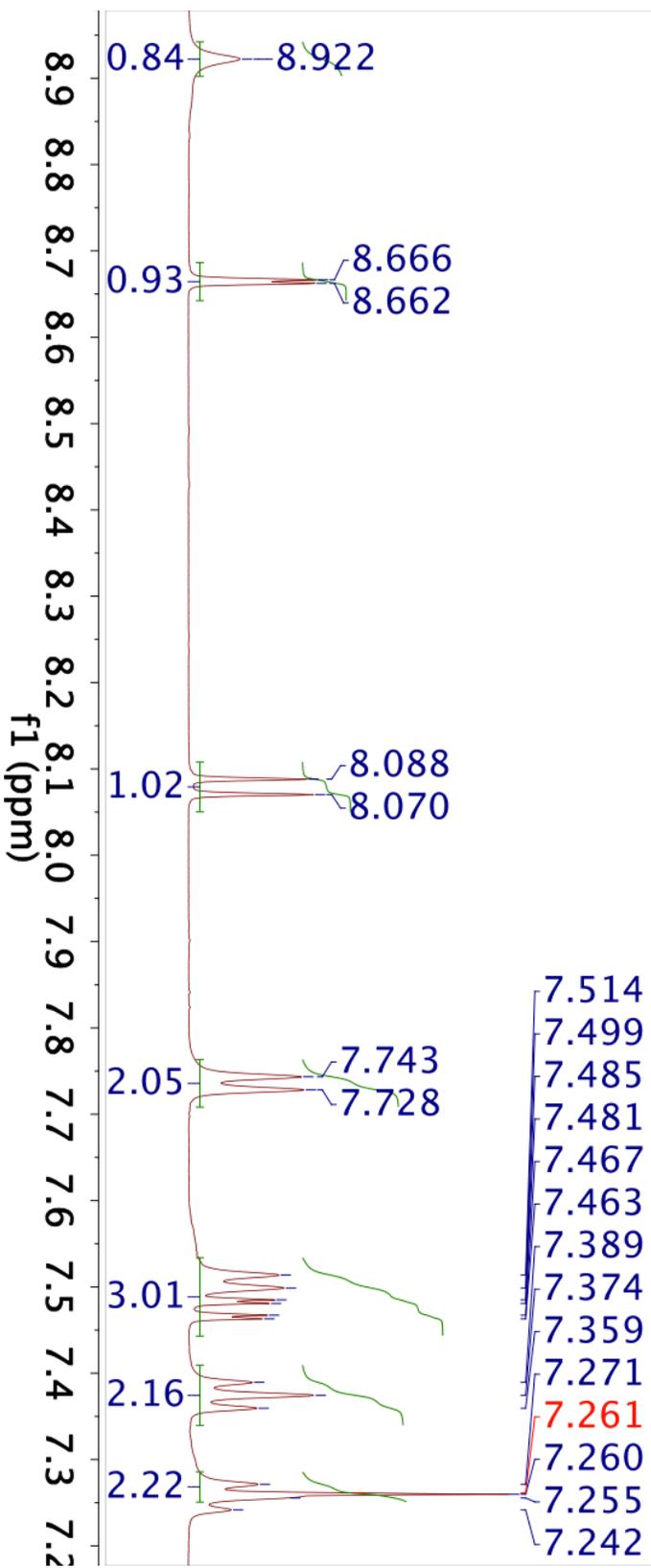
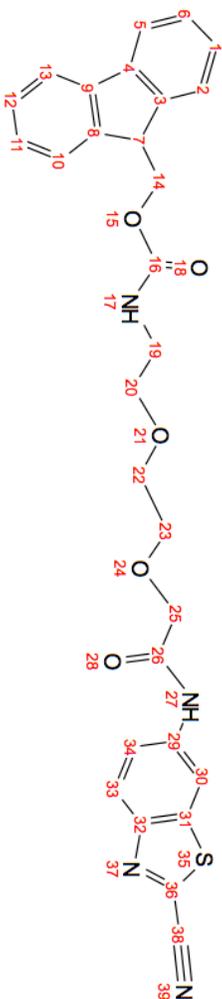
012717_CBT-miniPEG_CD3OD_13C_tuned
C13par



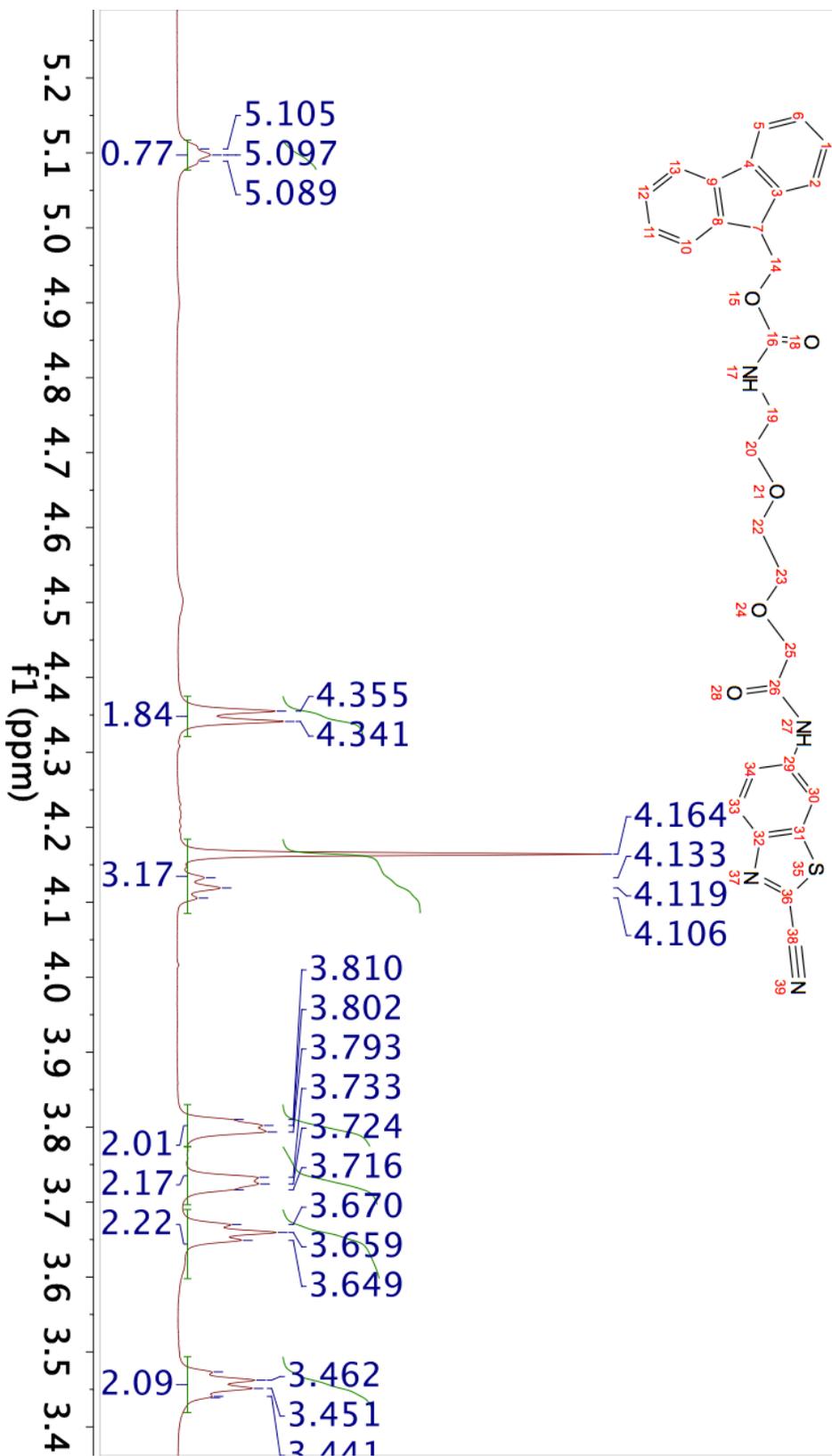
011617_CBT-Fmoc-miniPEG_fr13-21_CDCl3_proton
 STANDARD PROTON PARAMETERS



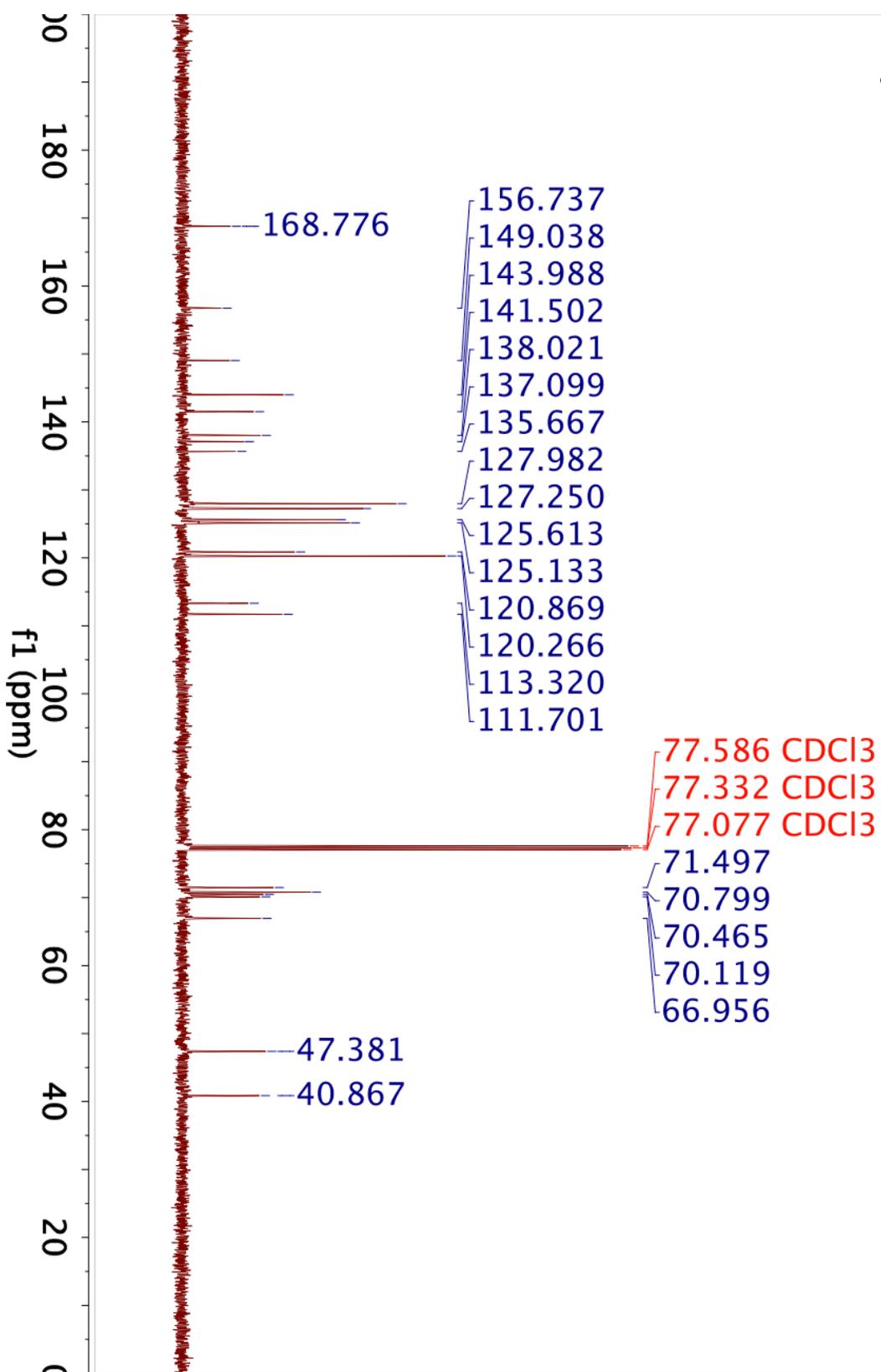
011617_CBT-Fmoc-miniPEG_fr13-21_CDCl3_proton
 STANDARD PROTON PARAMETERS



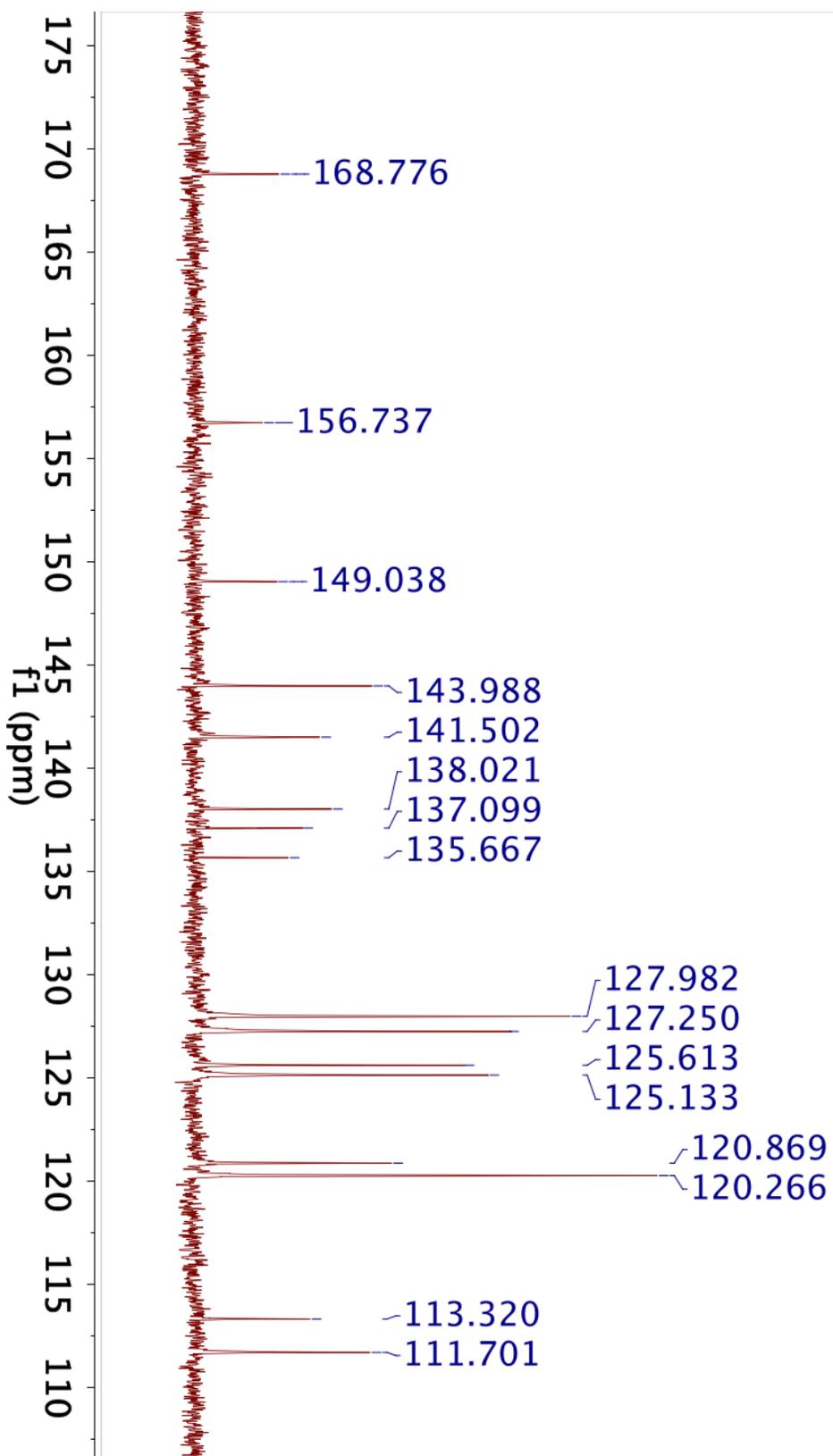
011617_CBT-Fmoc-miniPEG_fr13-21_CDCl3_proton
 STANDARD PROTON PARAMETERS



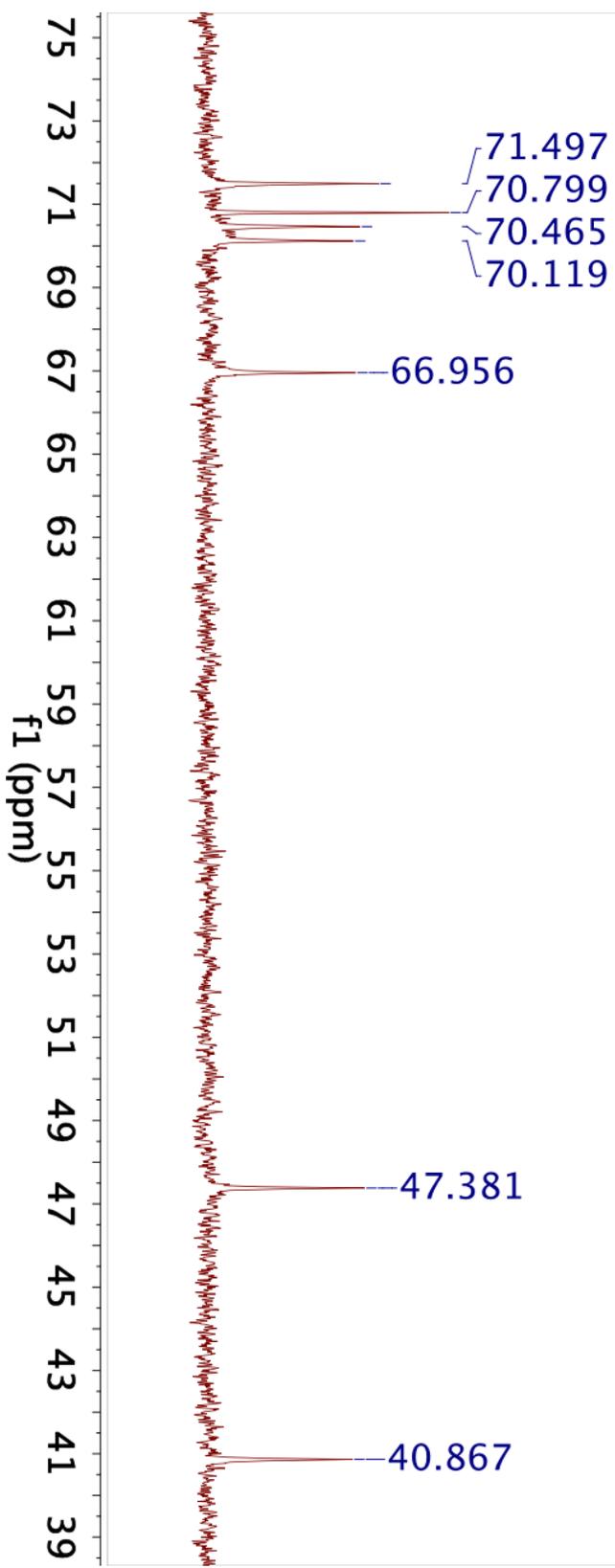
011617_CBT-Fmoc-miniPEG_fr22-26_CDCl3_13C
C13par



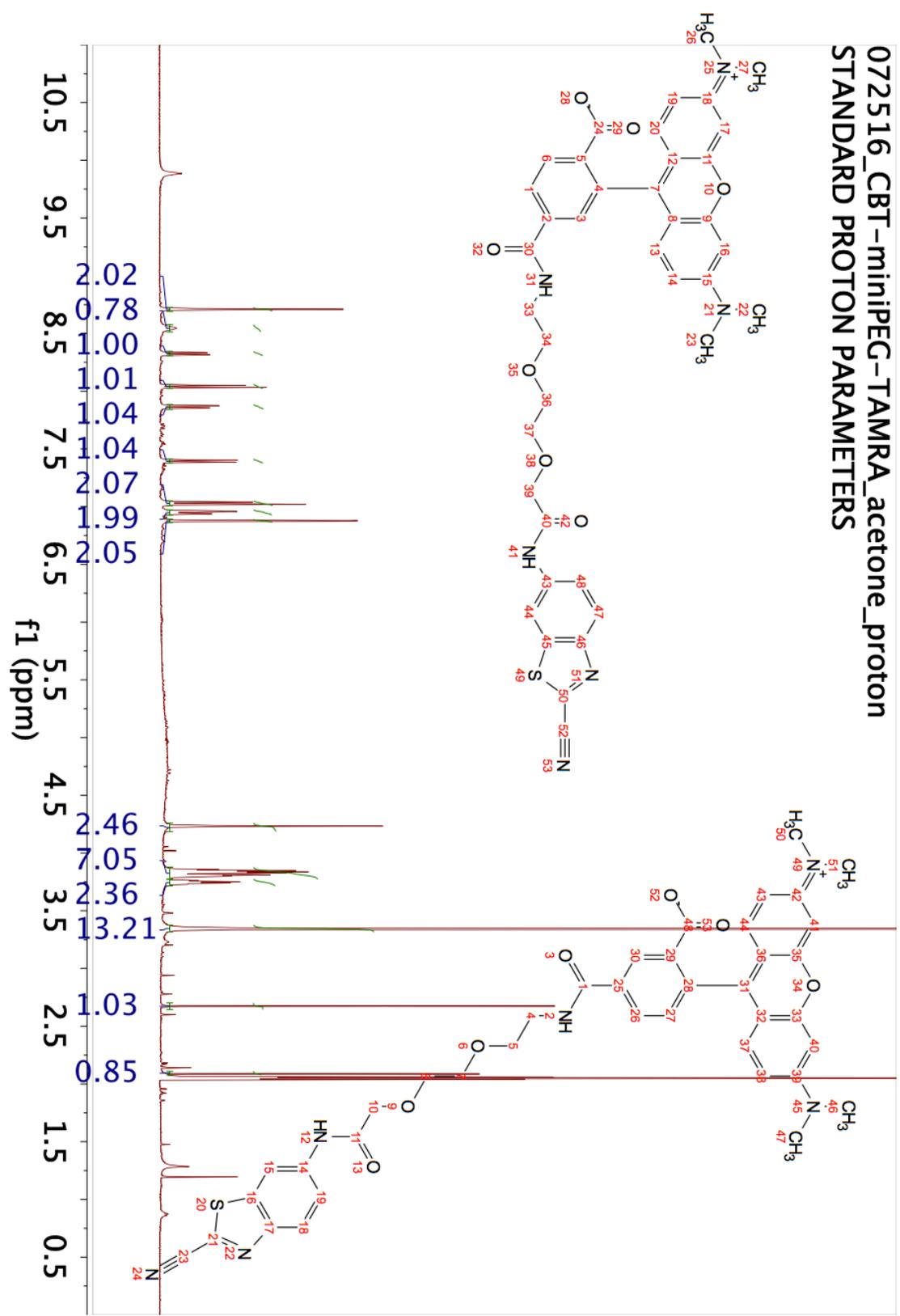
011617_CBT-Fmoc-miniPEG_fr22-26_CDCl3_13C
C13par



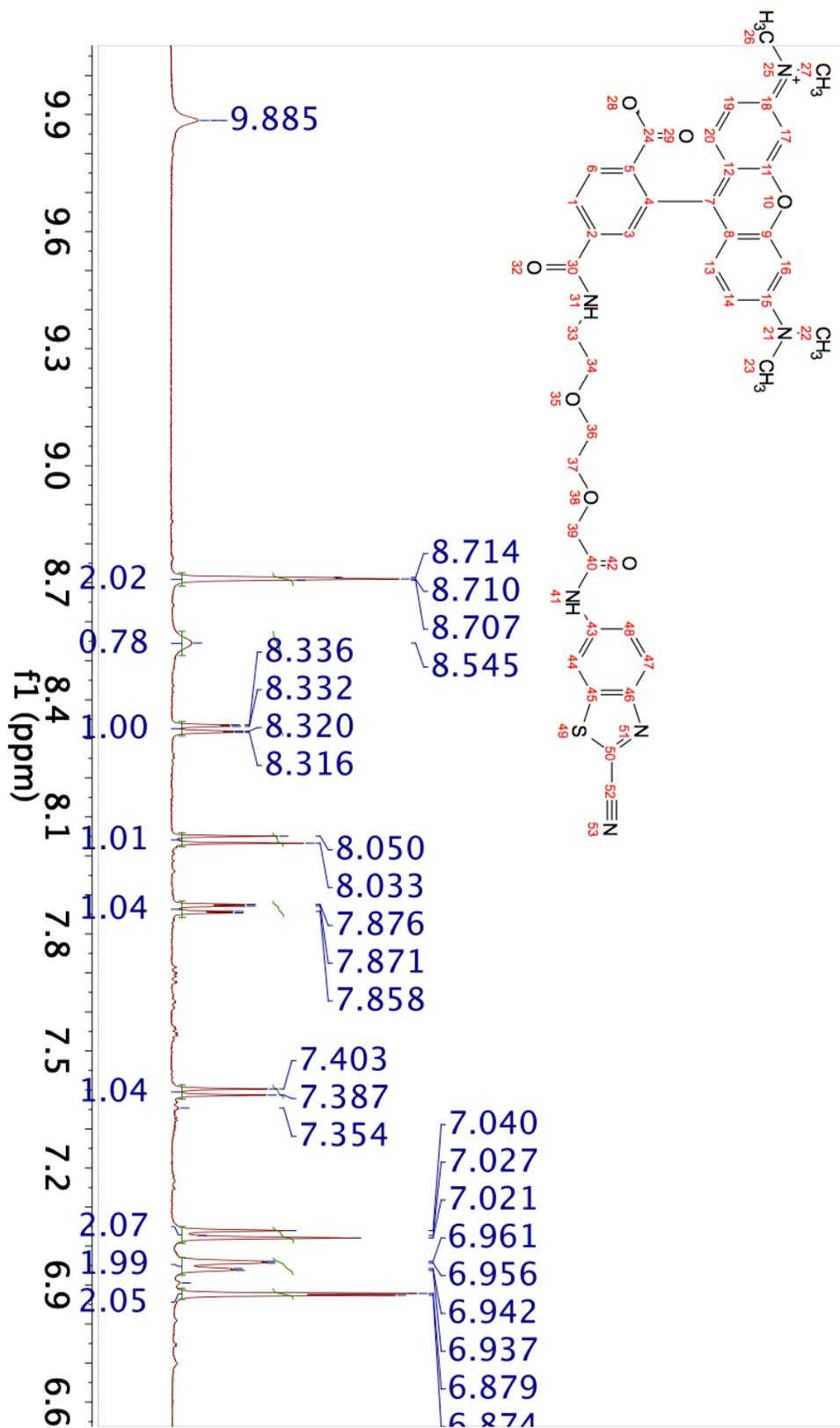
011617_CBT-Fmoc-miniPEG_fr22-26_CDCl3_13C
C13par



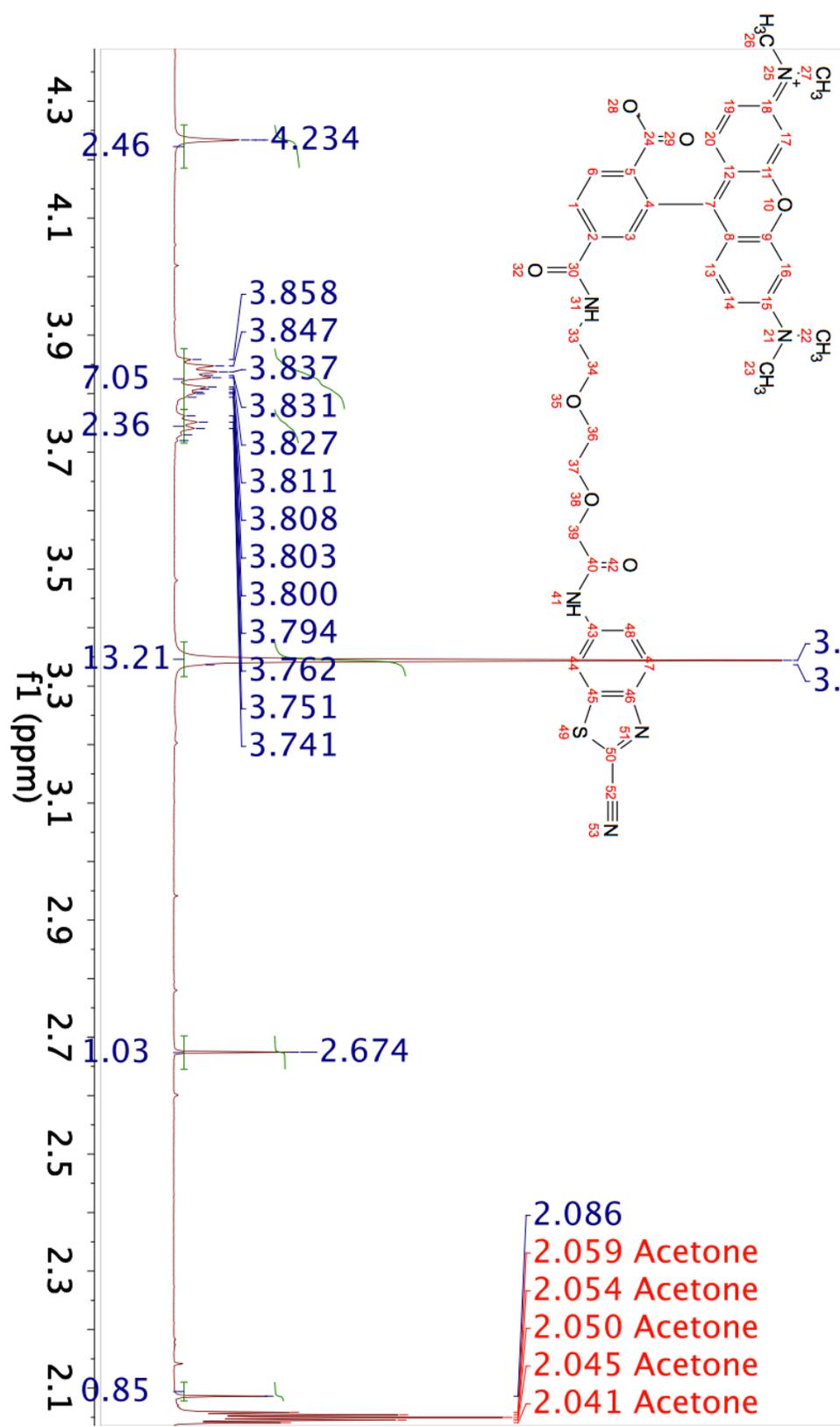
072516_CBT-miniPEG-TAMRA_acetone_proton
 STANDARD PROTON PARAMETERS

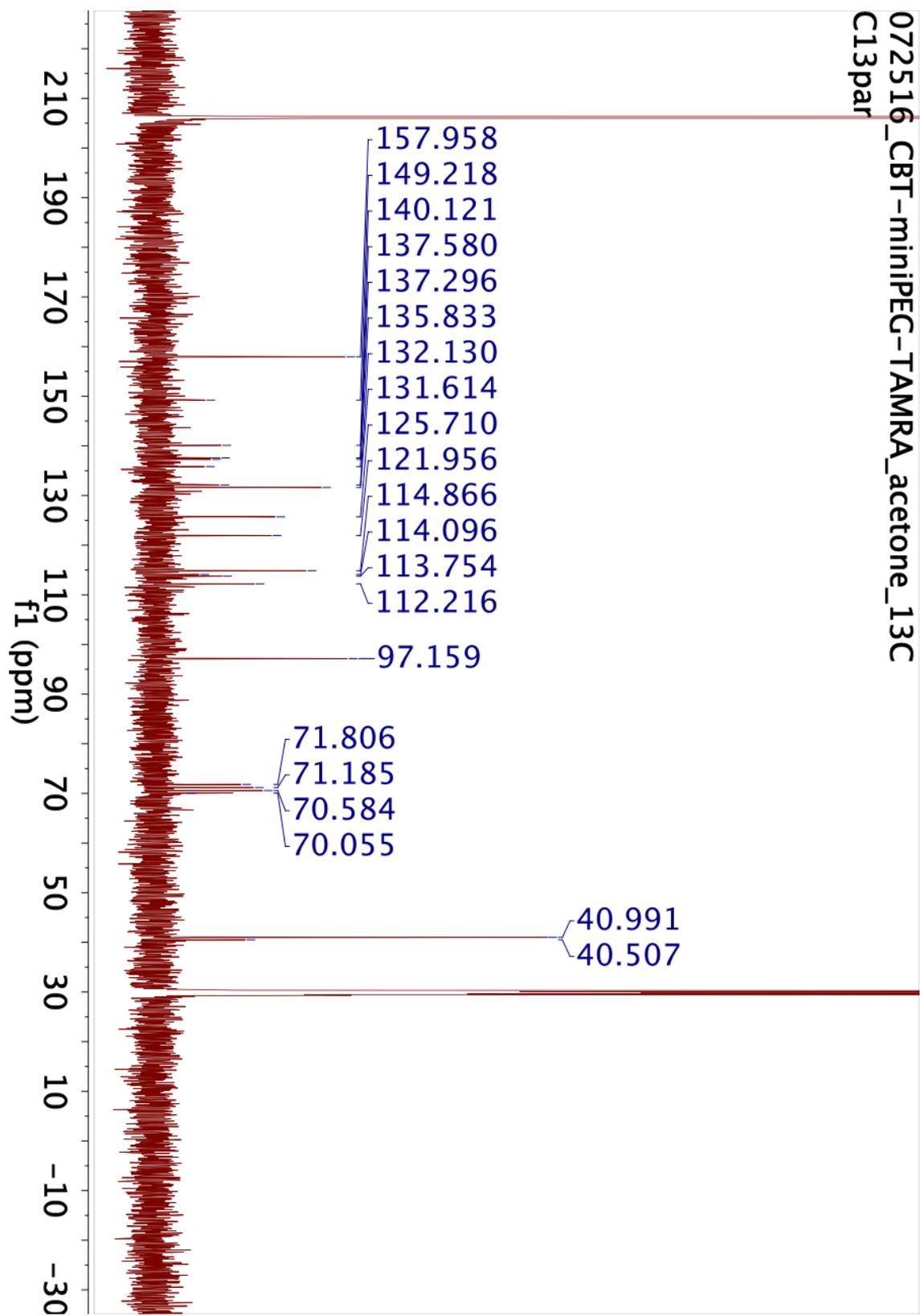


072516_CBT-miniPEG-TAMRA_acetone_proton
 STANDARD PROTON PARAMETERS

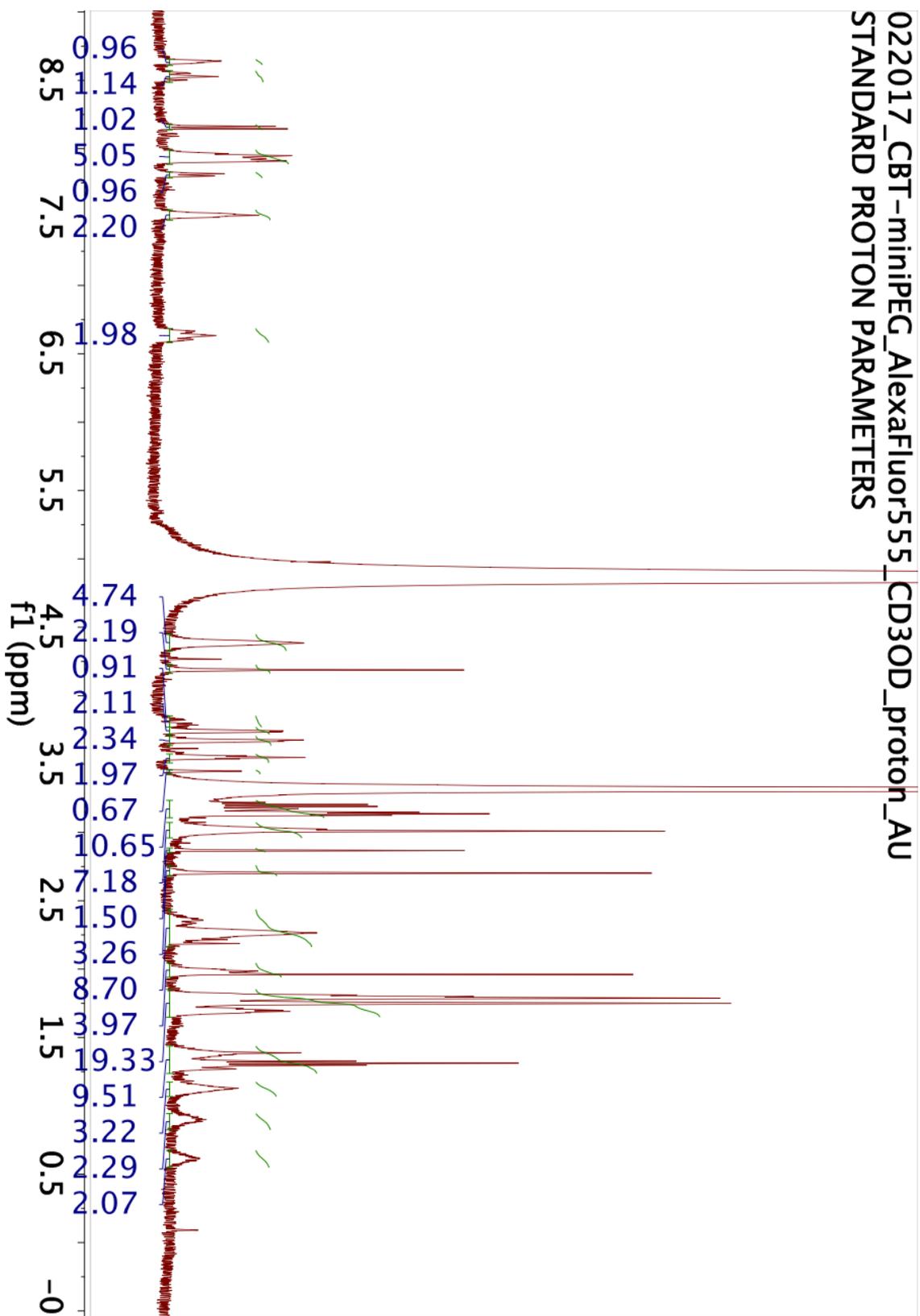


072516_CBT-miniPEG-TAMRA_acetone_proton
 STANDARD PROTON PARAMETERS

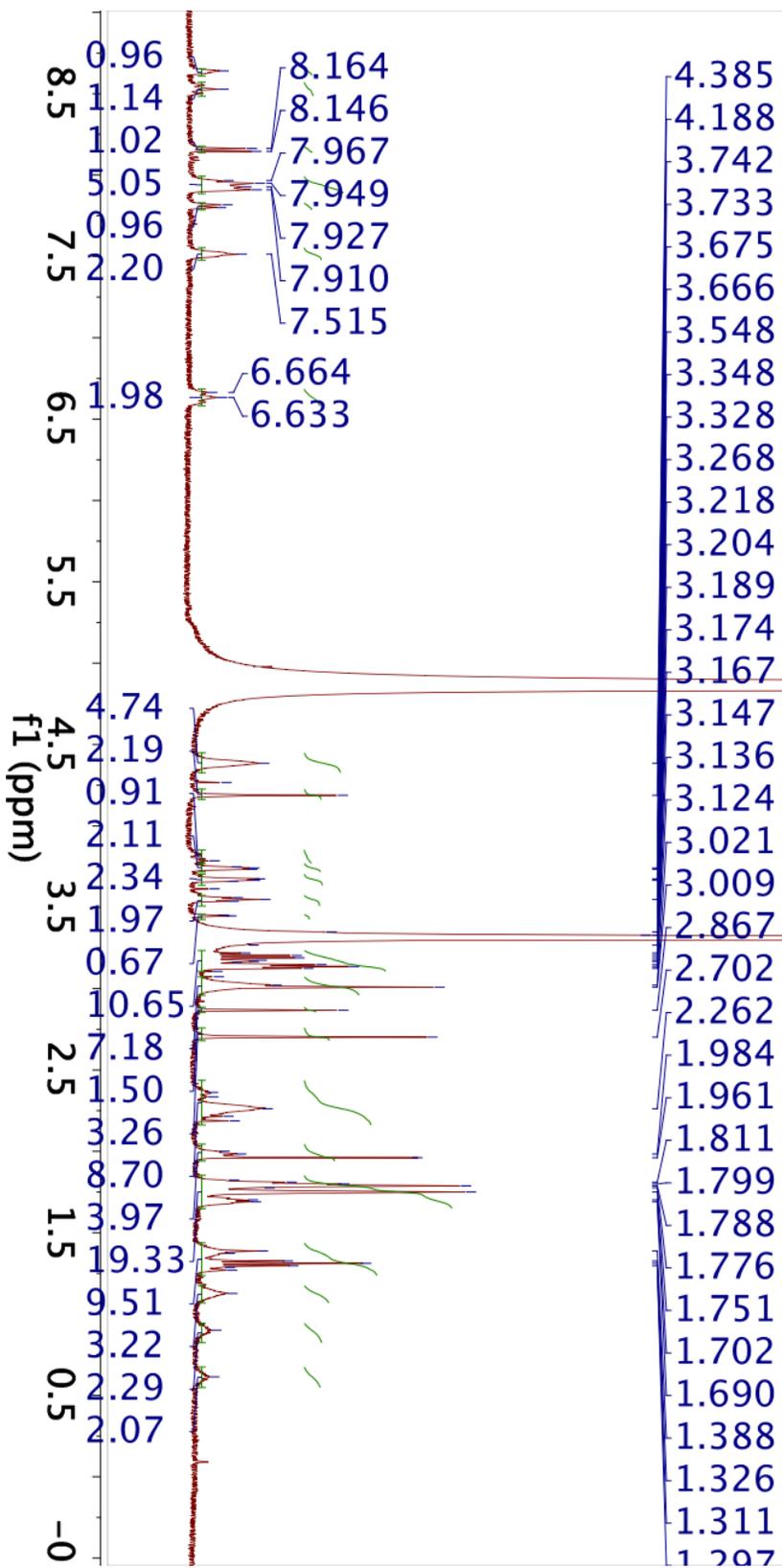




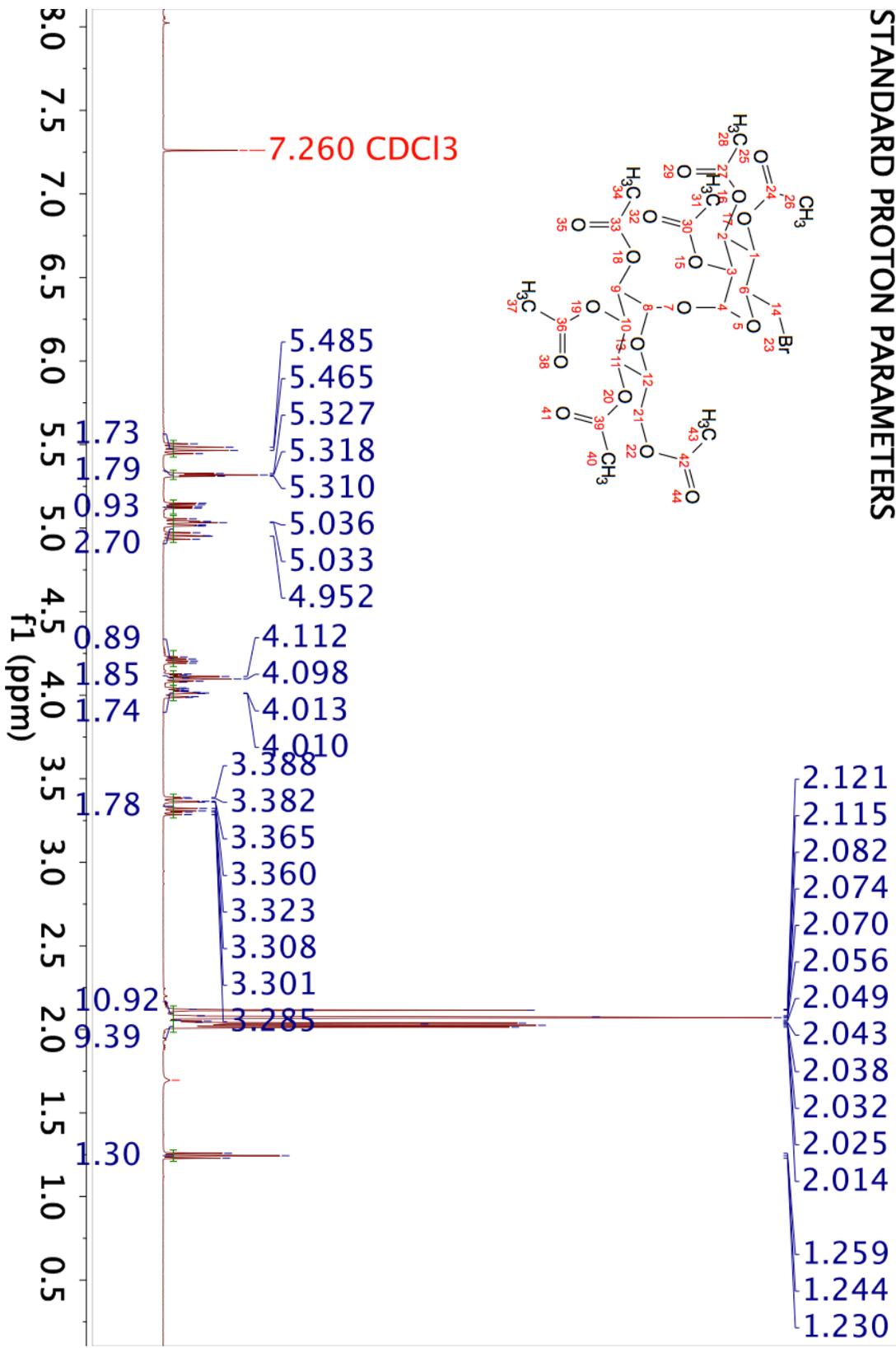
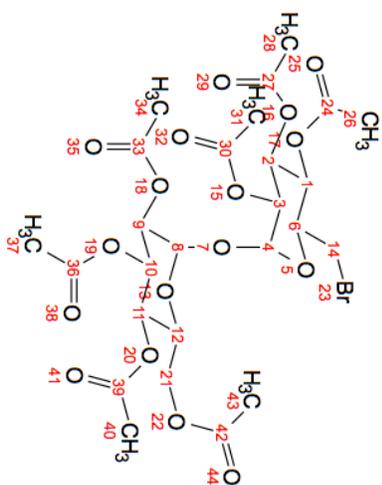
022017_CBT-miniPEG_AlexaFluor555_CD3OD_proton_AU
STANDARD PROTON PARAMETERS



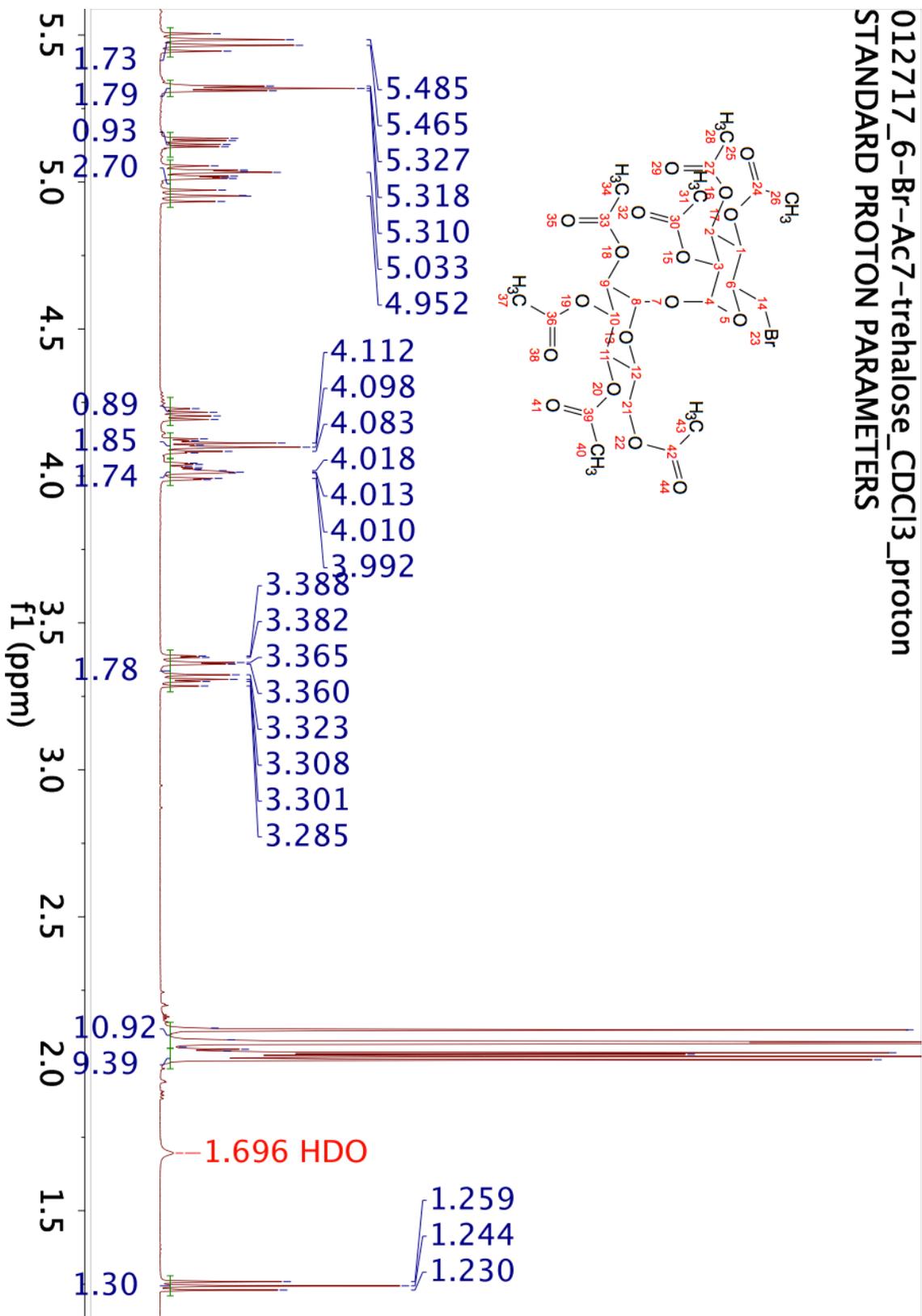
022017_CBT-miniPEG_AlexaFluor555_CD3OD_proton_AU
 STANDARD PROTON PARAMETERS



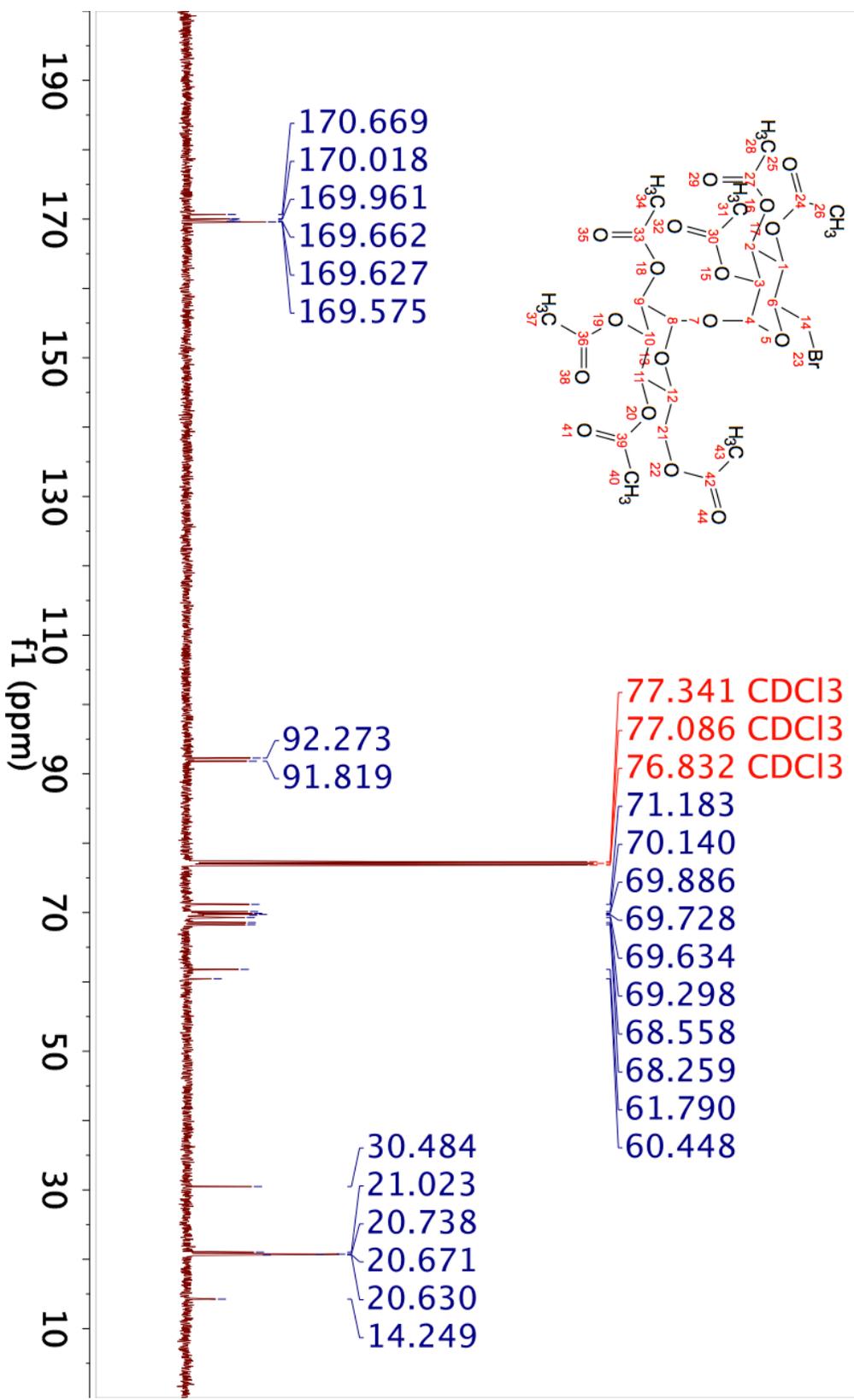
012717_6-Br-Ac7-trehalose_CDCl3_proton
 STANDARD PROTON PARAMETERS



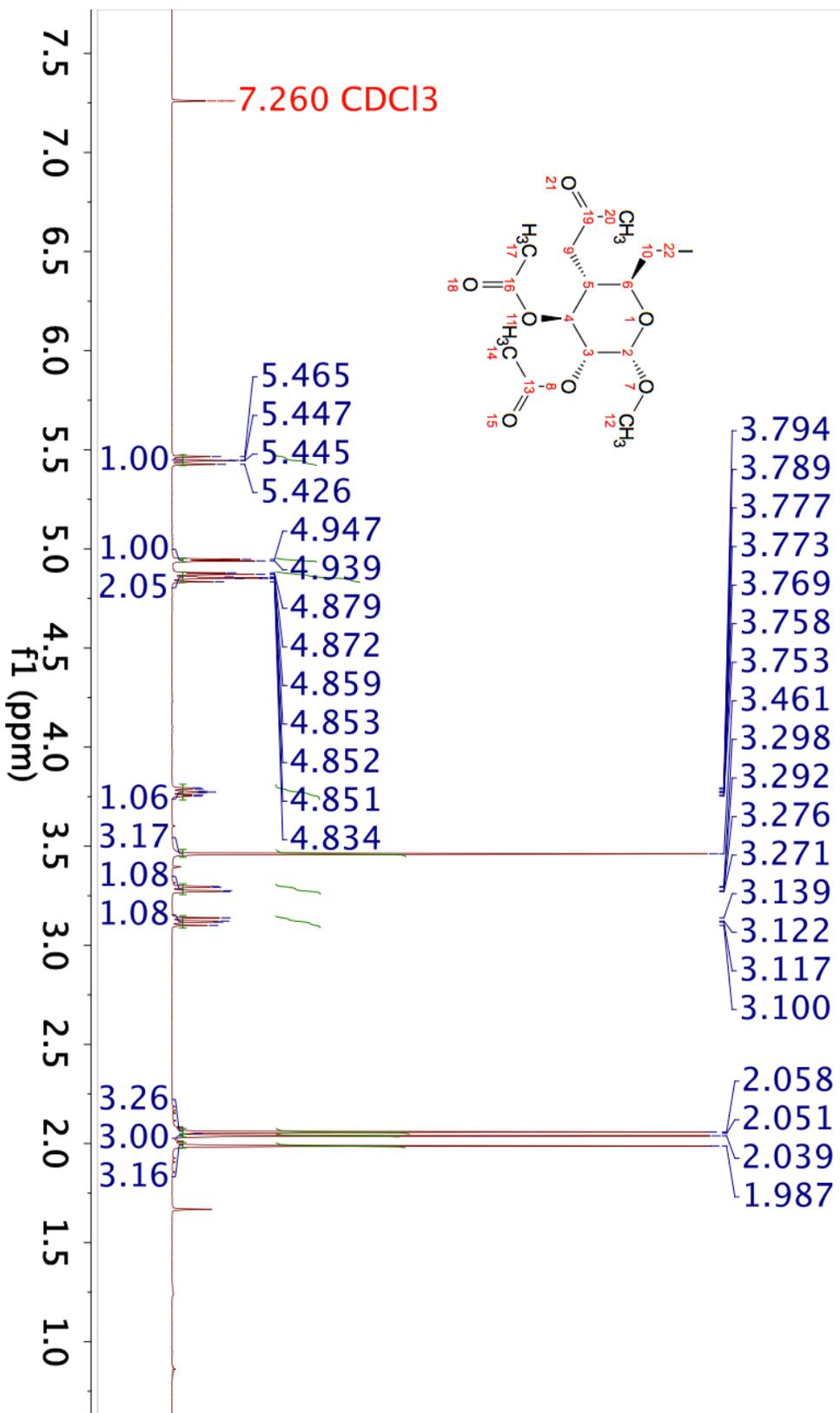
012717_6-Br-Ac7-trehalose_CDCl3_proton
 STANDARD PROTON PARAMETERS



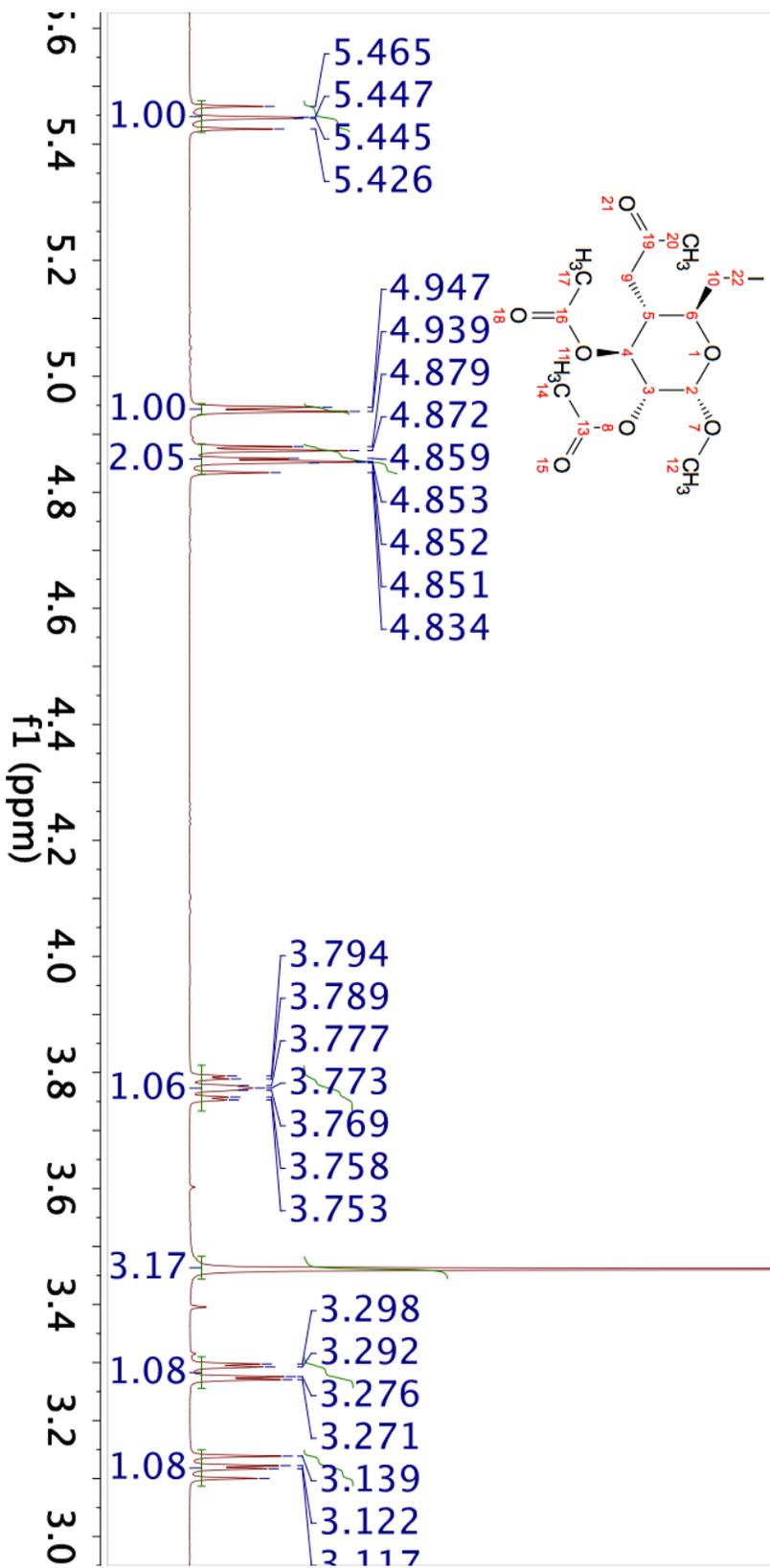
012717_6-Br-Ac7-Tre_CDCI3_13C
 C13par



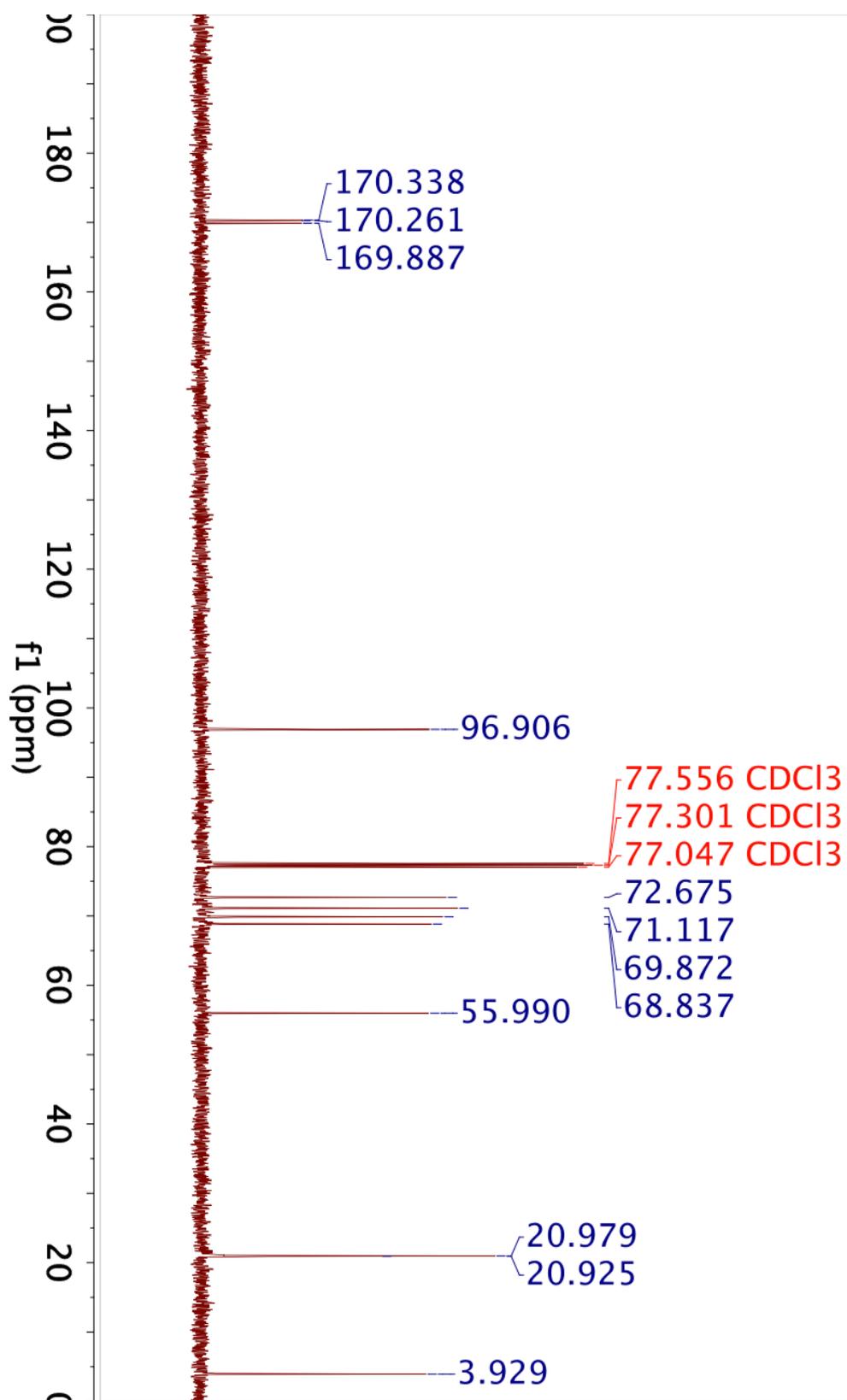
062617_methyl-6-iodo-ac7-glucopyranoside_pure_cdcl3_proton
 STANDARD PROTON PARAMETERS



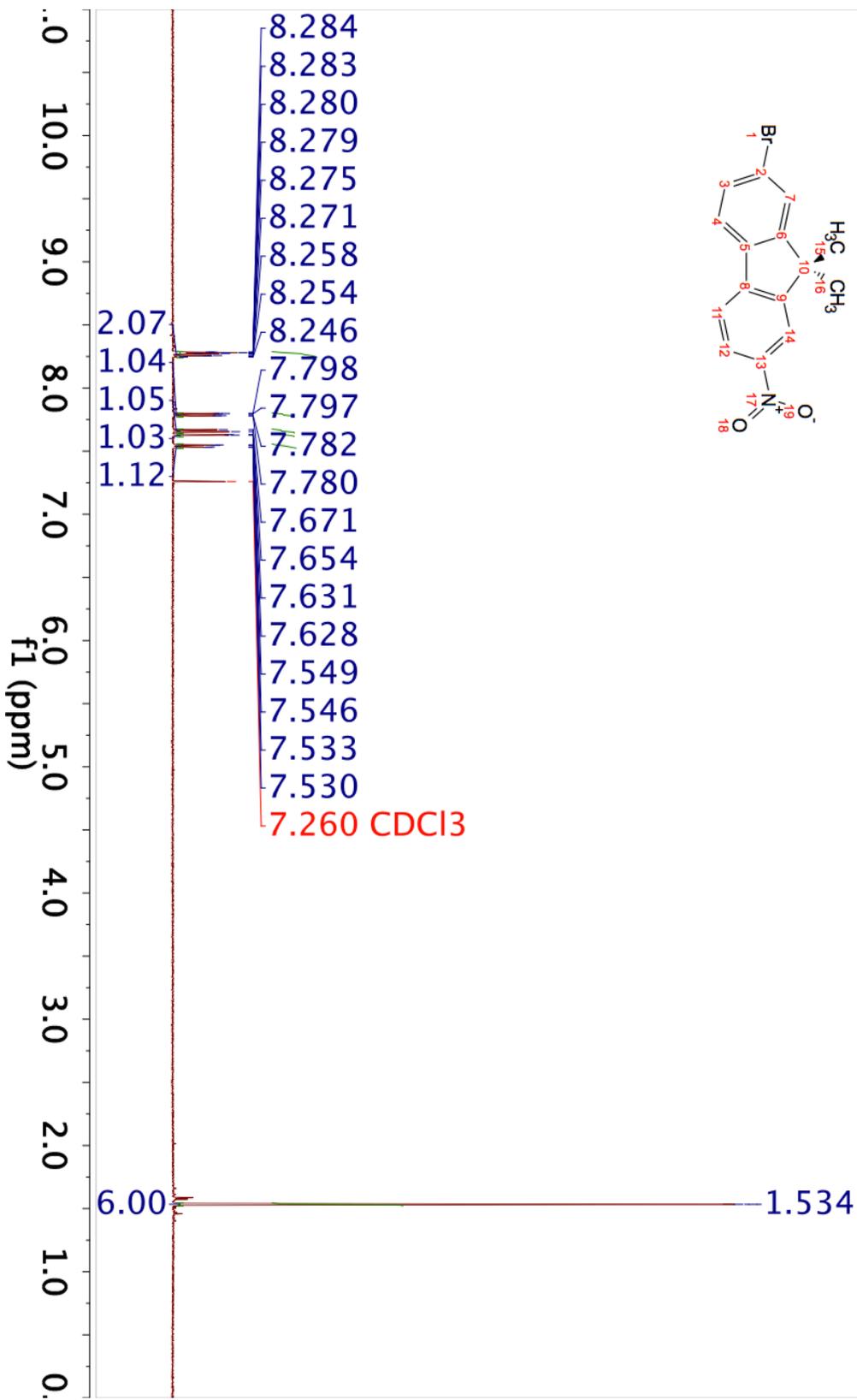
062617_methyl-6-iodo-ac7-glucopyranoside_pure_cdcl3_proton
 STANDARD PROTON PARAMETERS



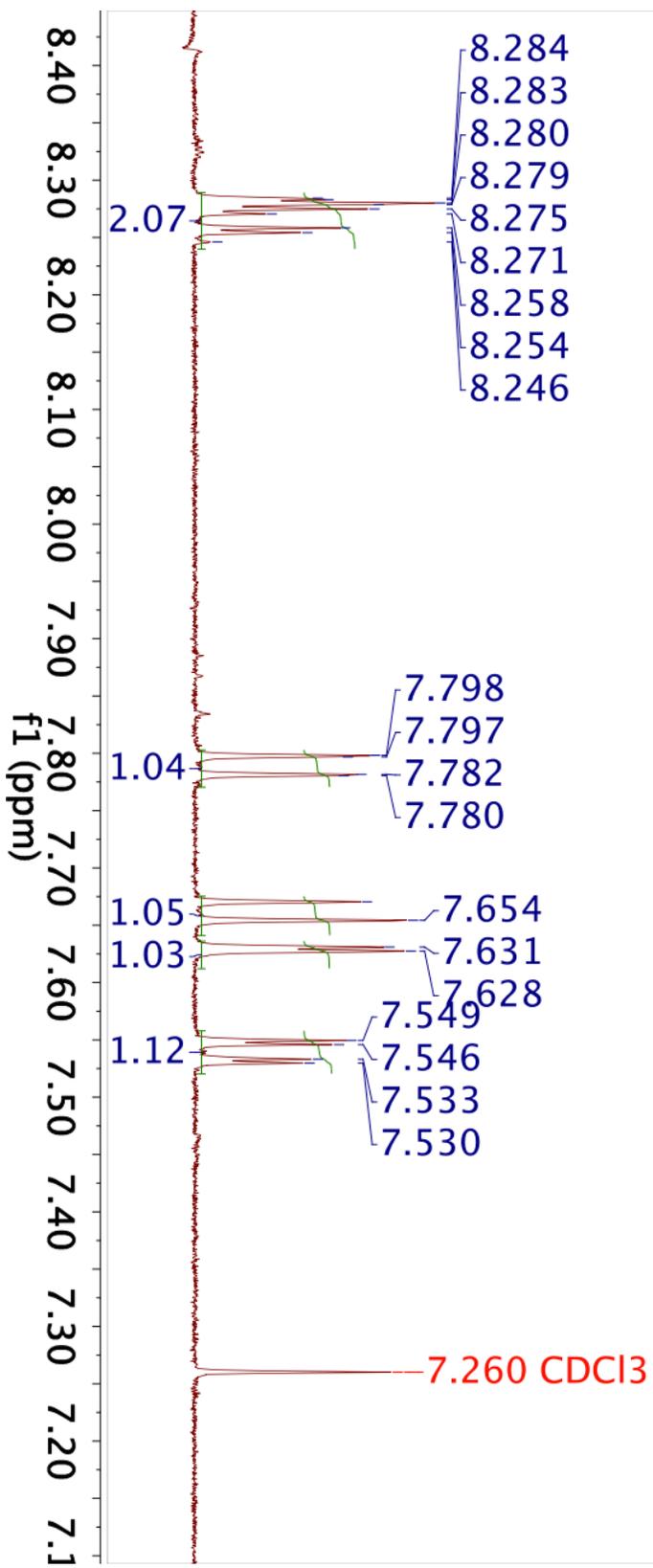
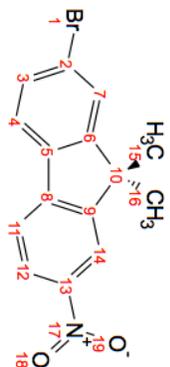
062617_methyl-6-iodo-ac7-glucopyranoside_pure_cdcl3_13C
C13par



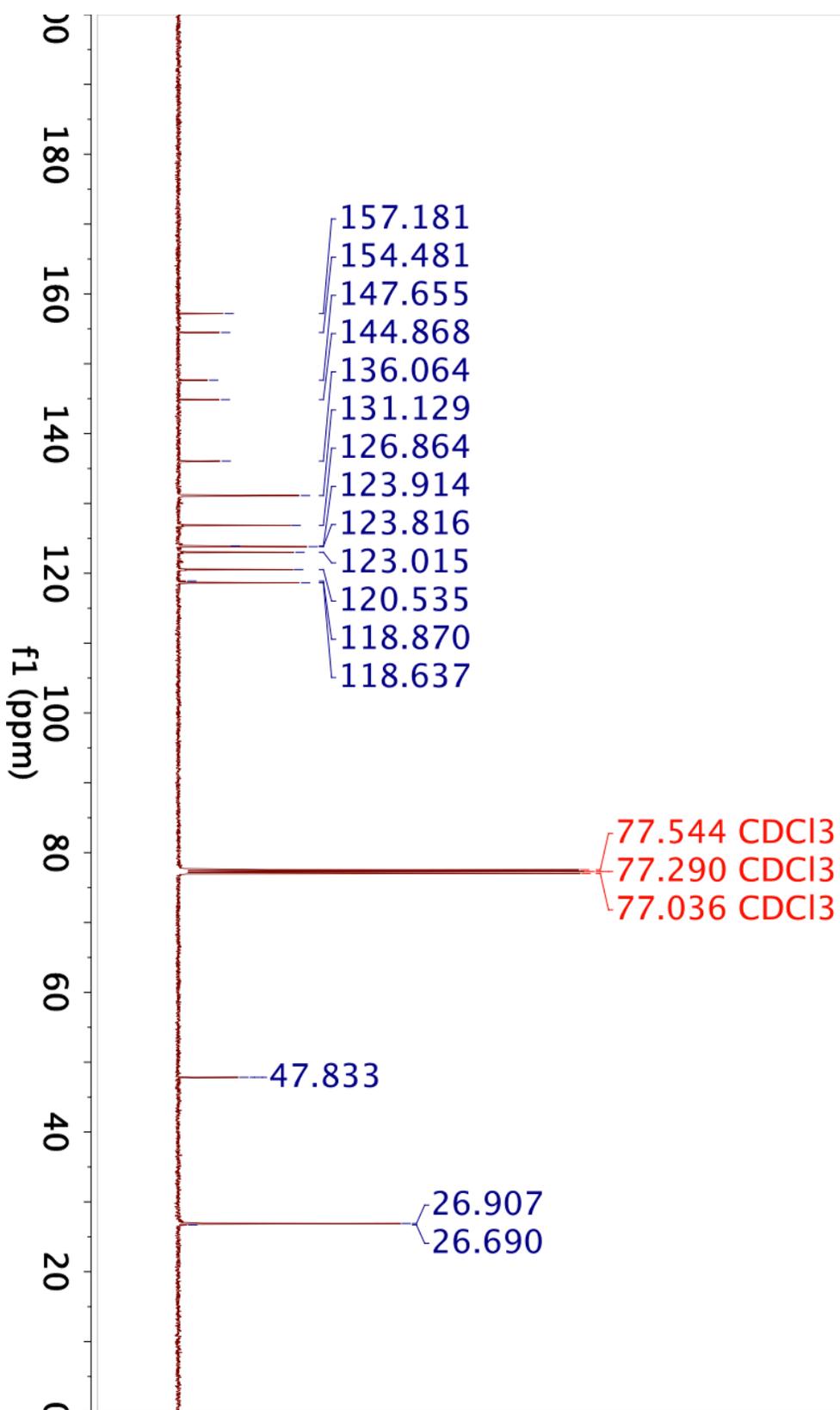
011717_2-Br-9-9-me2-7-no2-fluorene_CDCl3_proton
STANDARD PROTON PARAMETERS



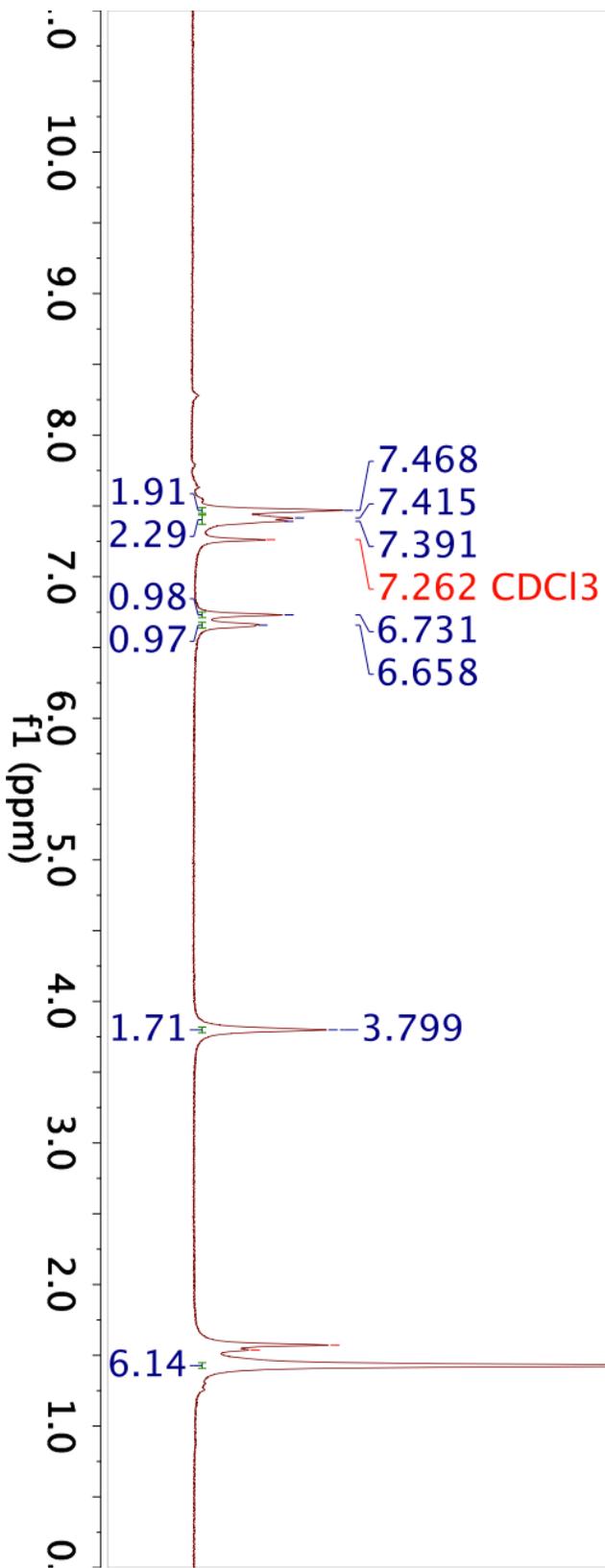
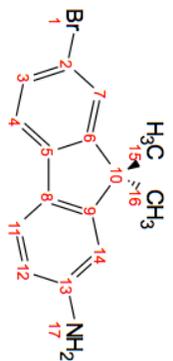
011717_2-Br-9-me2-7-no2-fluorene_CDCl3_proton
STANDARD PROTON PARAMETERS



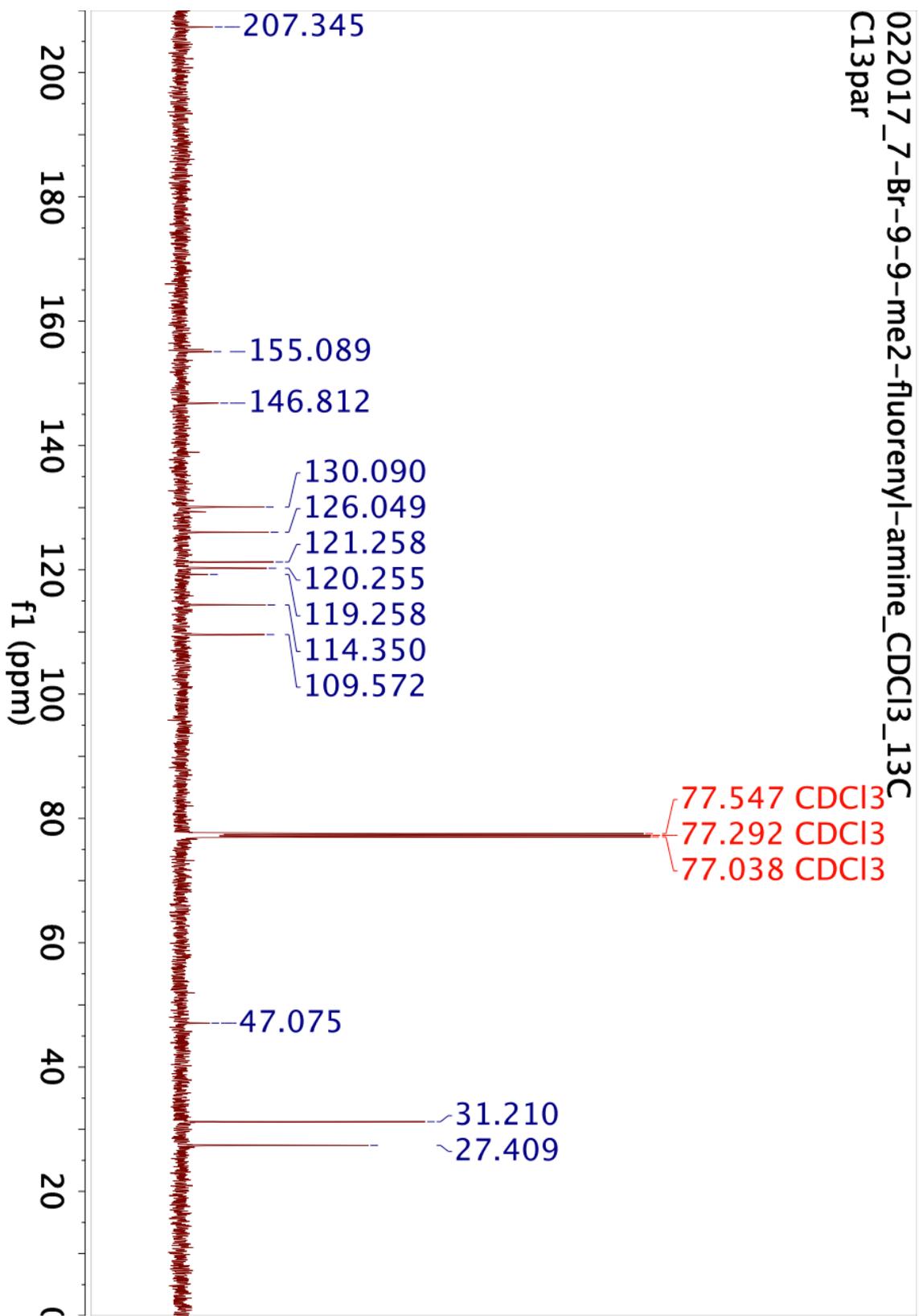
011717_2-Br-9-9-me2-7-no2-fluorene_cdcl3_C13
C13par



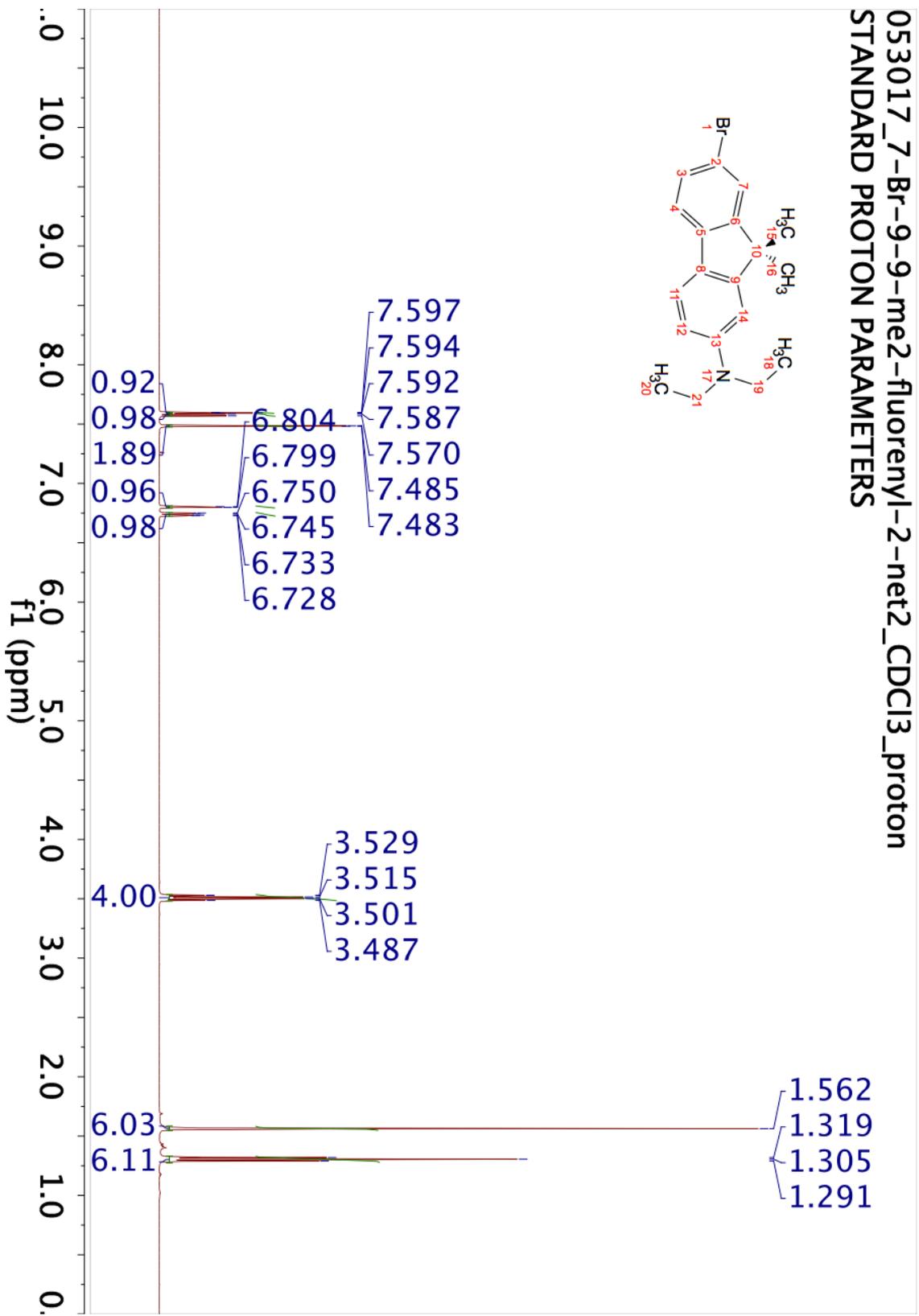
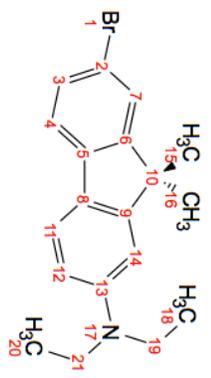
021717_7-Br-9-9-me2-fluorenyl-2-amine_CDCI3_proton
STANDARD PROTON PARAMETERS



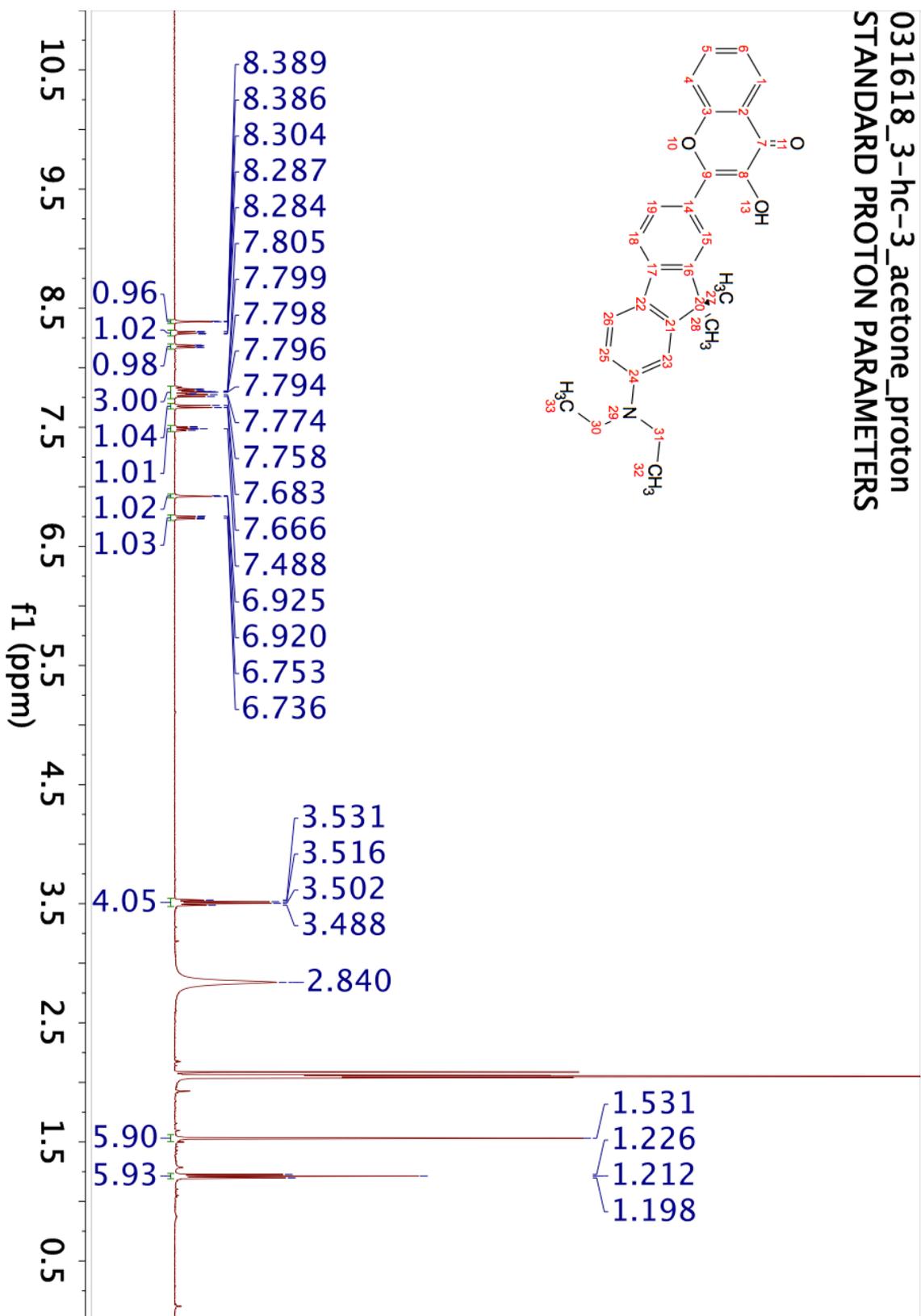
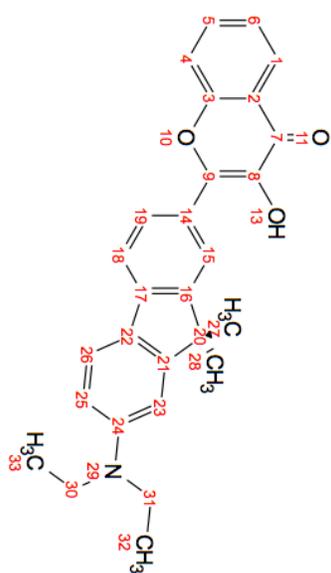
022017_7-Br-9-9-me2-fluorenyl-amine_CDCl3_13C
C13par



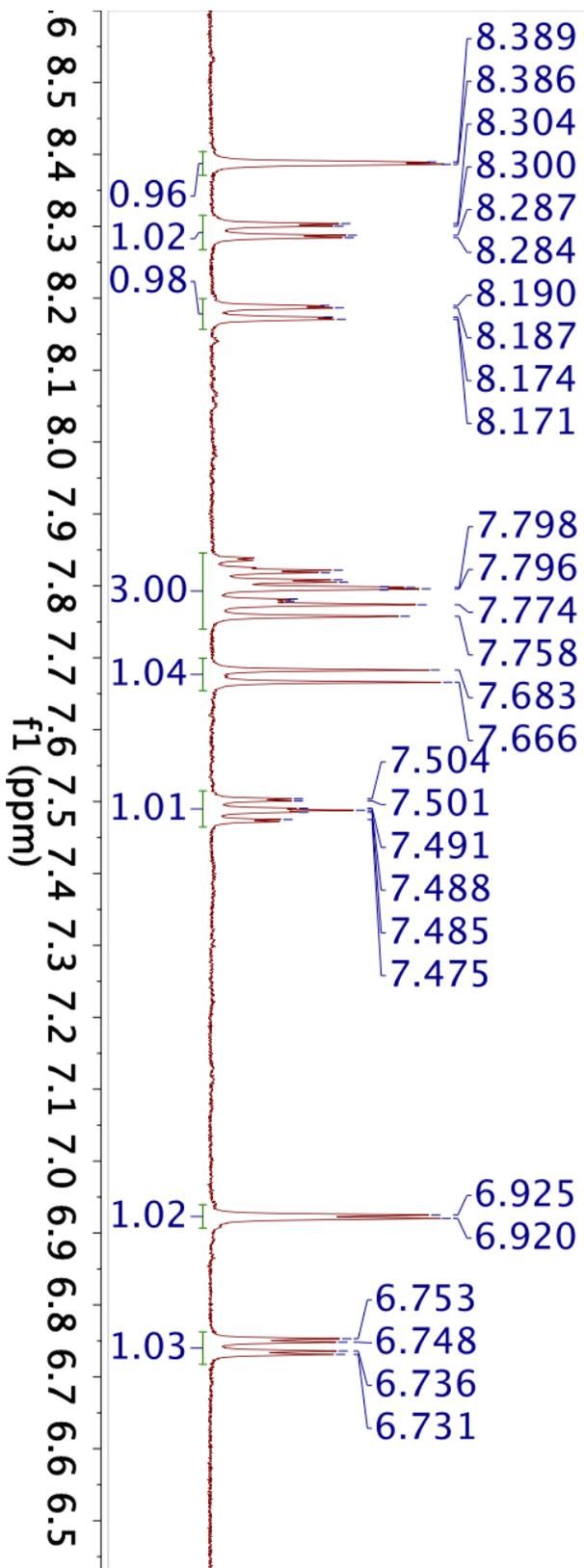
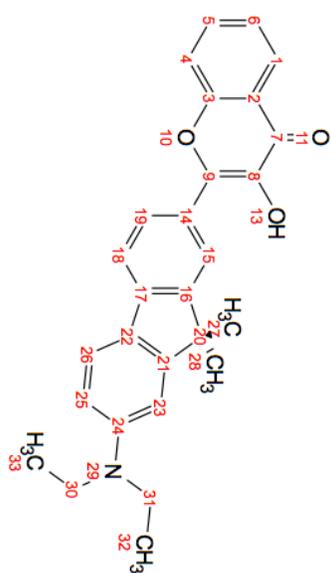
053017_7-Br-9-9-me2-fluorenyl-2-net2_CDCI3_proton
 STANDARD PROTON PARAMETERS



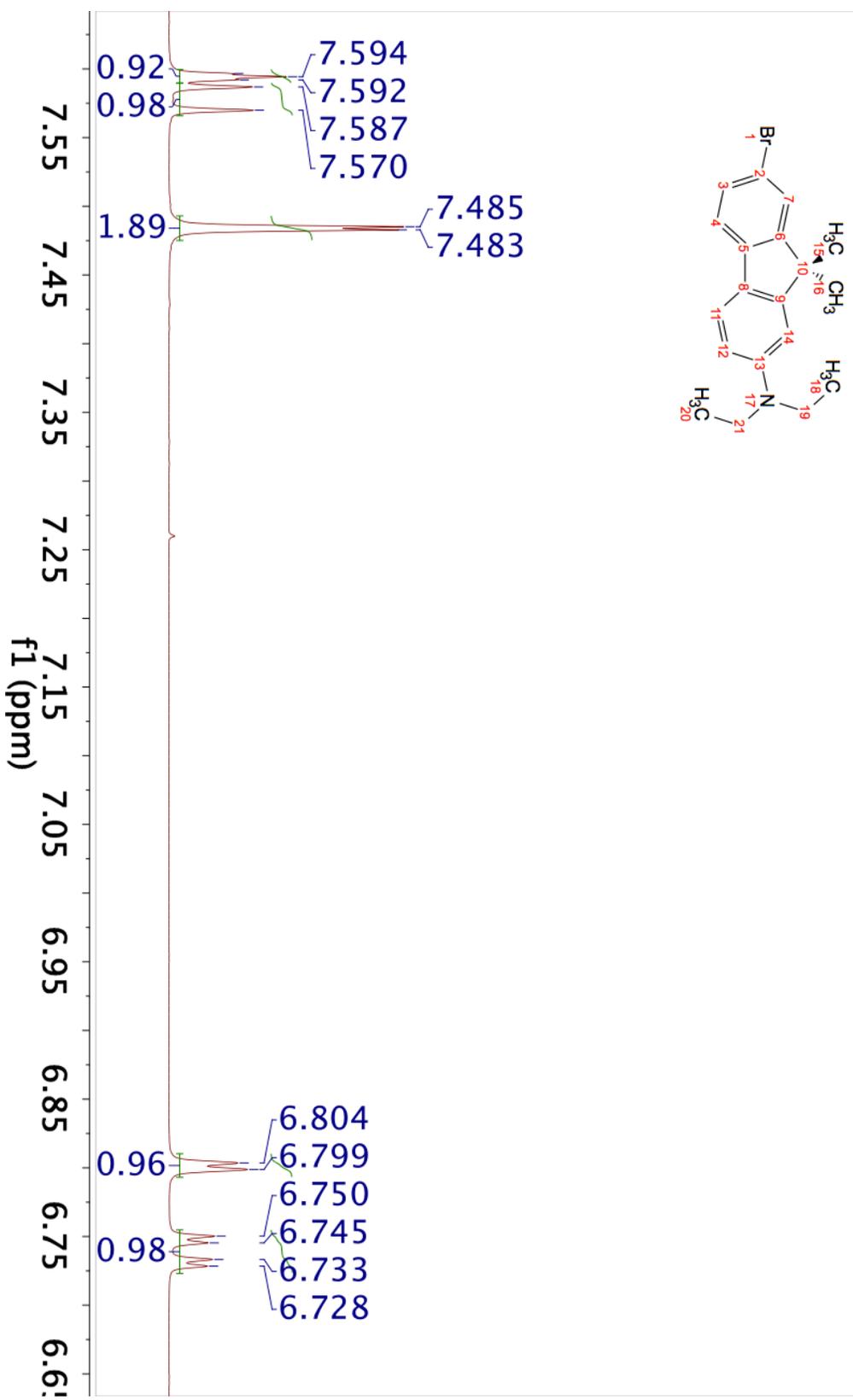
031618_3-hc-3_acetone_proton
STANDARD PROTON PARAMETERS



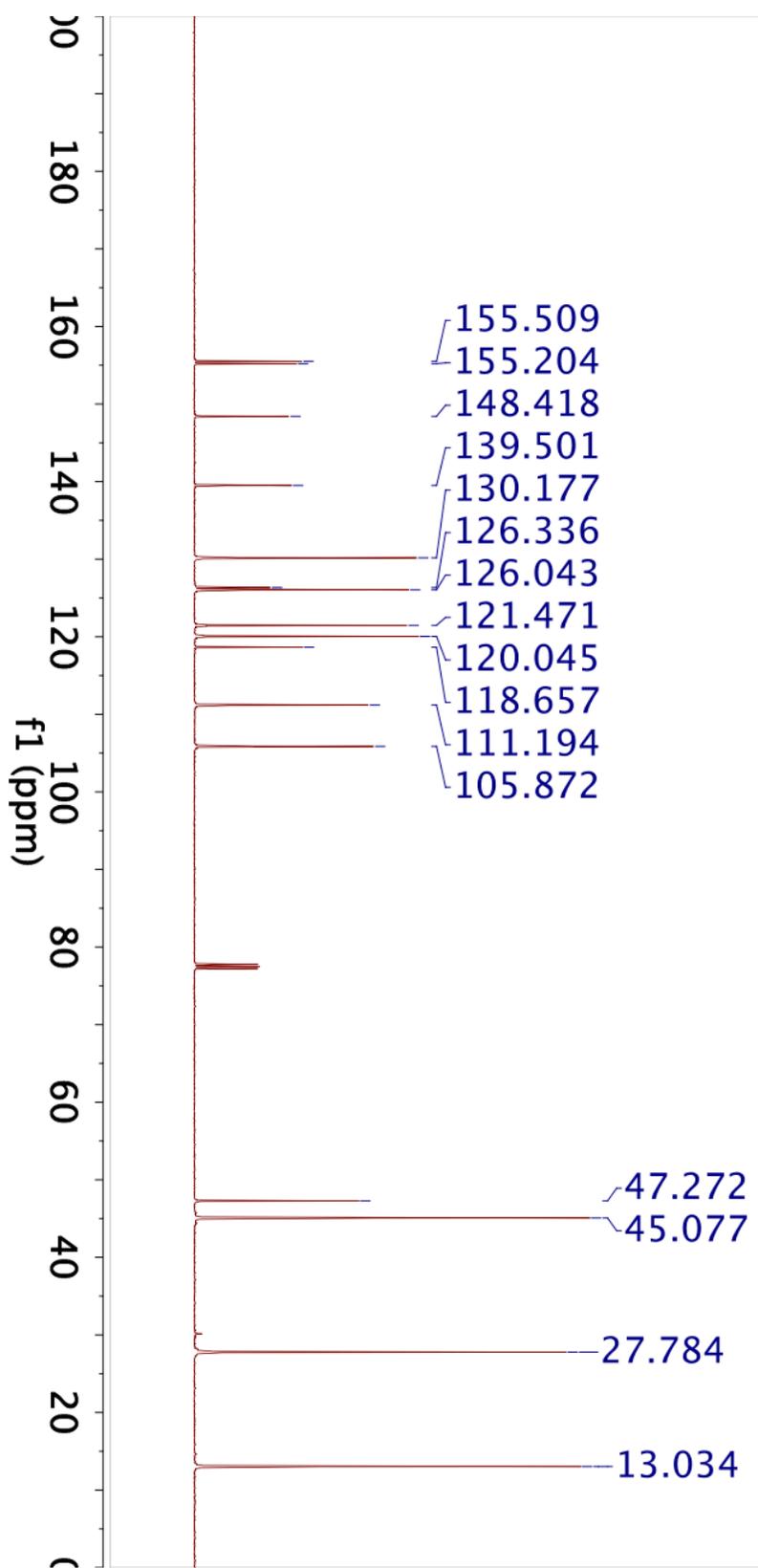
031618_3-hc-3_acetone_proton
 STANDARD PROTON PARAMETERS



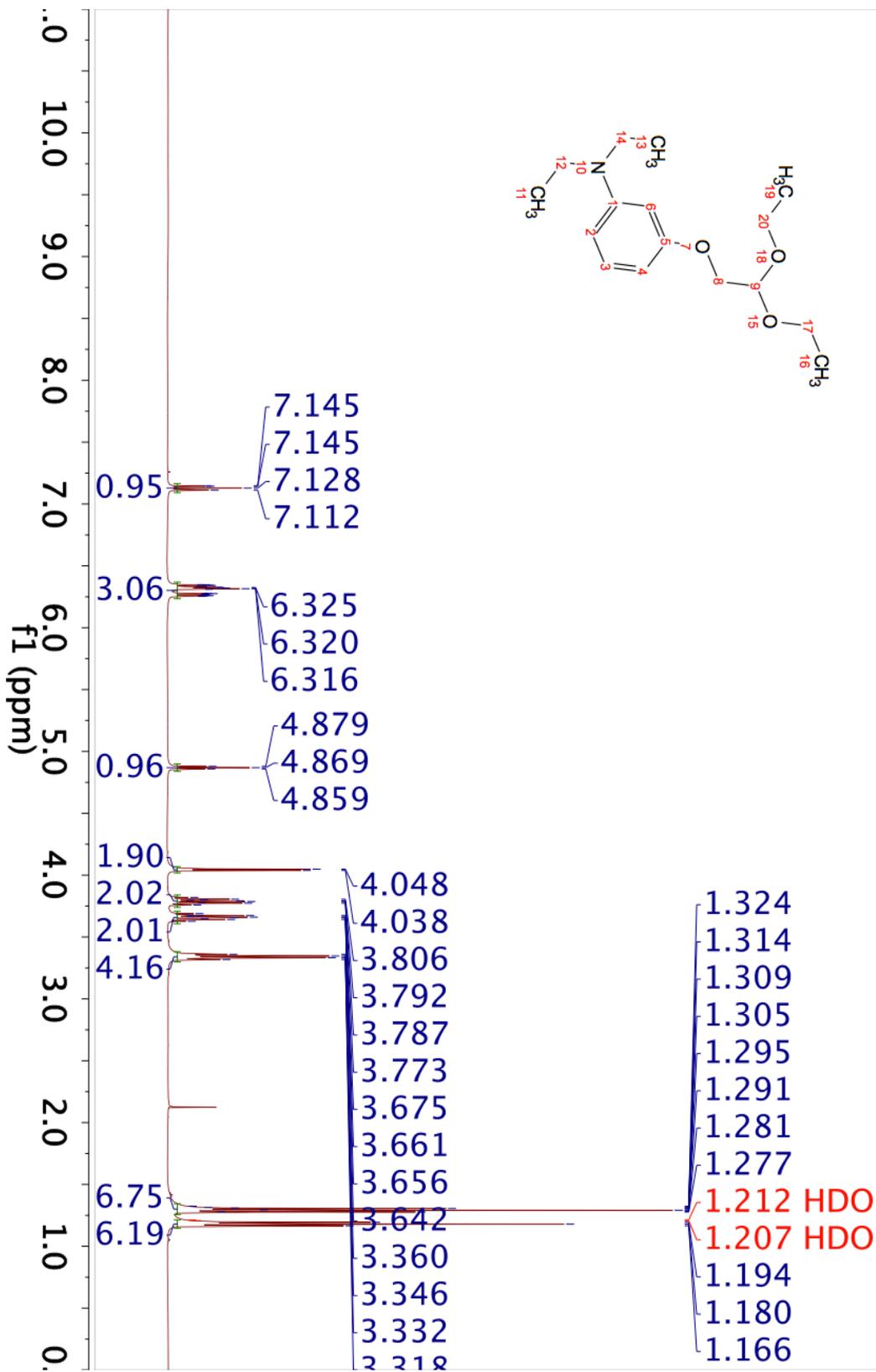
053017_7-Br-9-9-me2-fluorenyl-2-net2_CDCI3_proton
STANDARD PROTON PARAMETERS



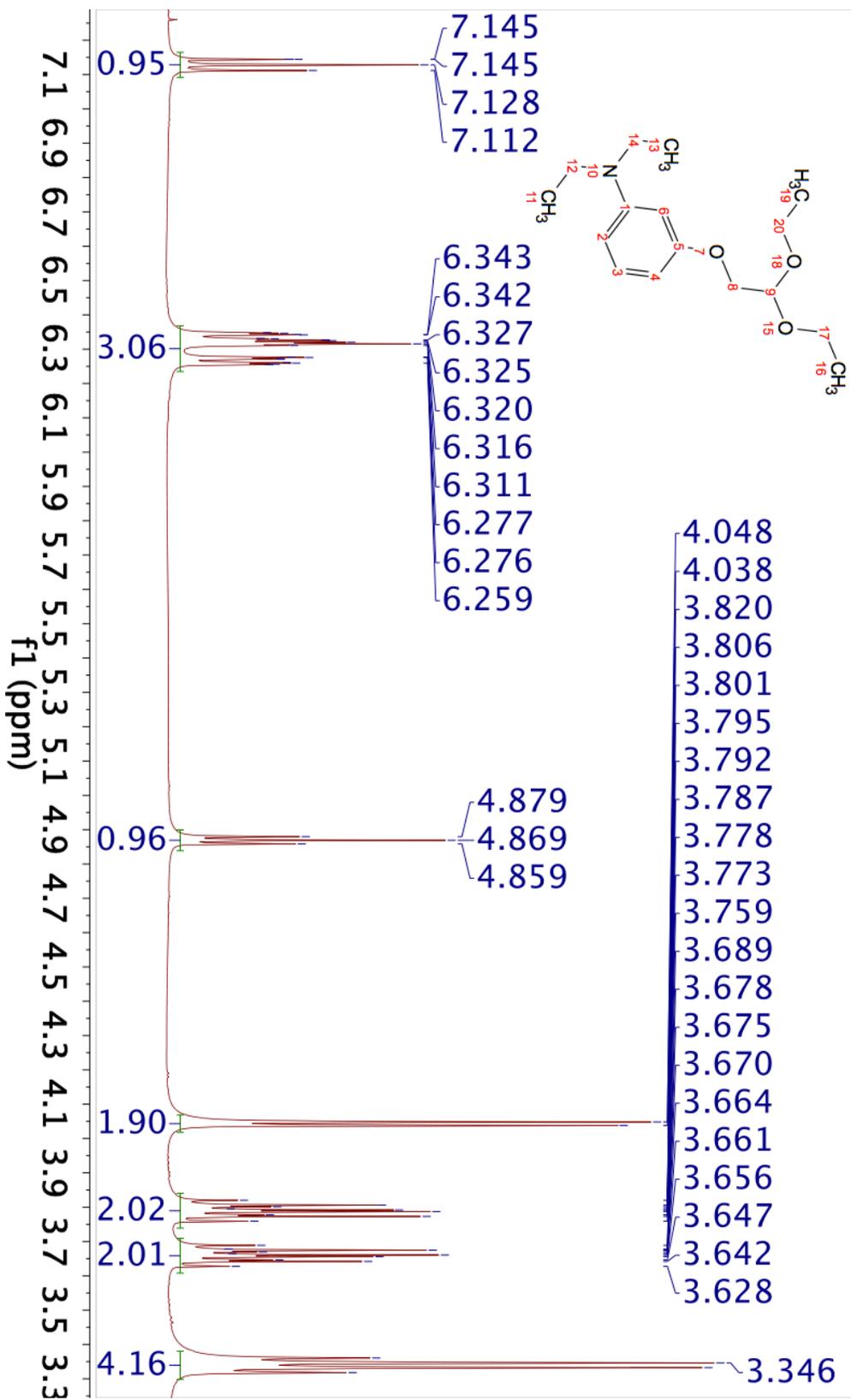
053017_7-br-9-9-me2-fluorenyl-2-net2_CDCl3_13C
C13par



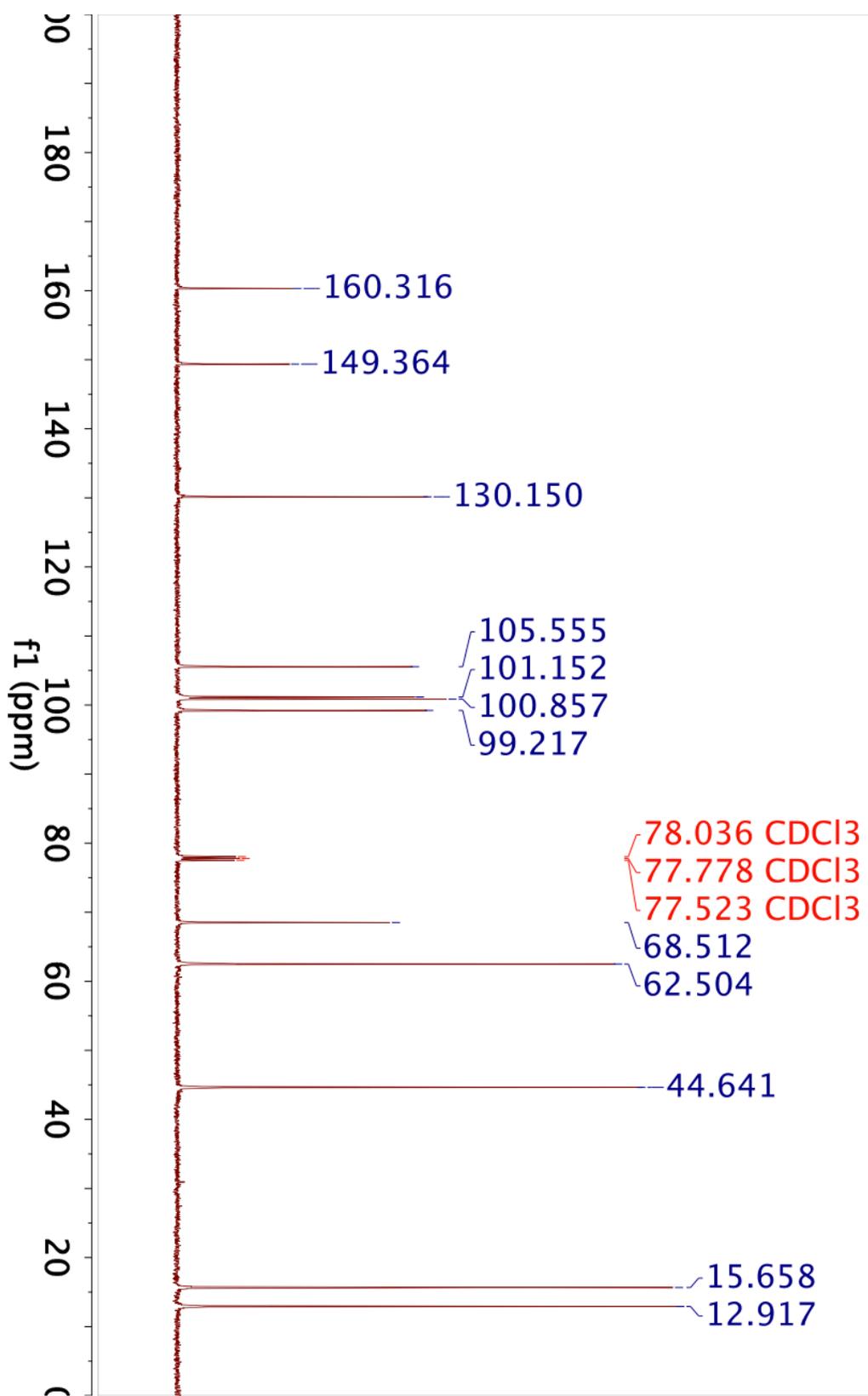
060817_diethoxy-etc_CDCl3_proton
 STANDARD PROTON PARAMETERS



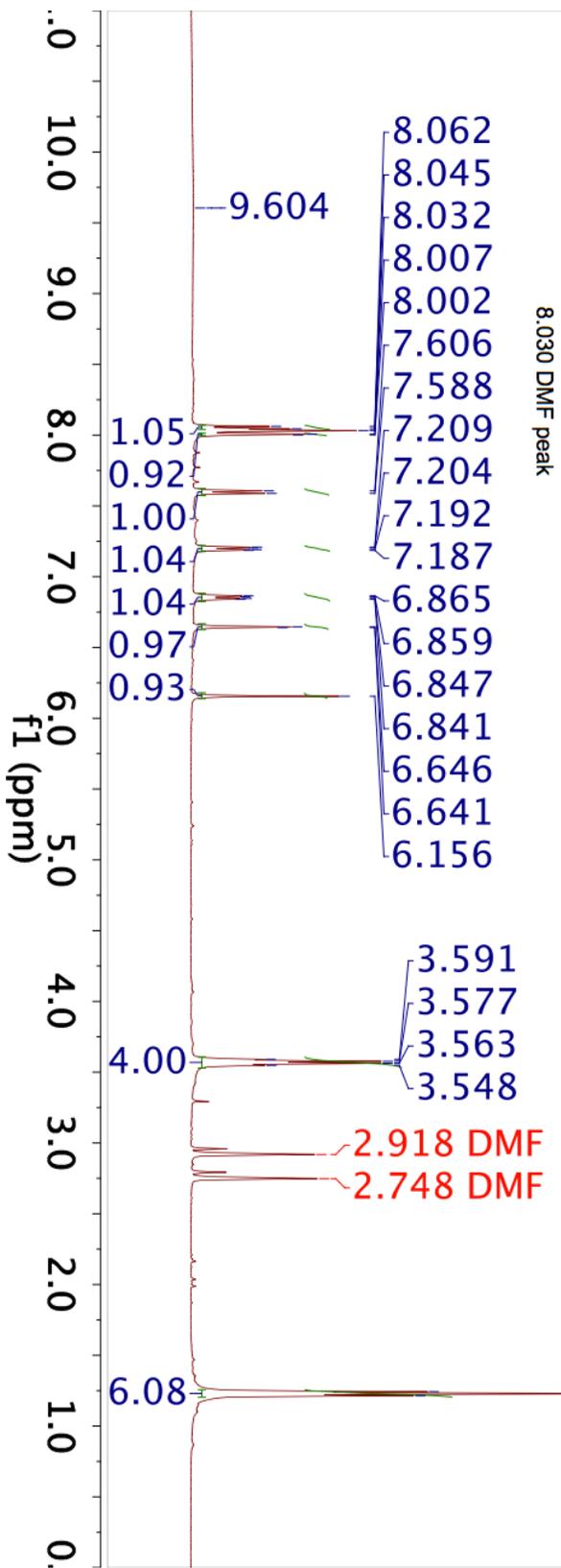
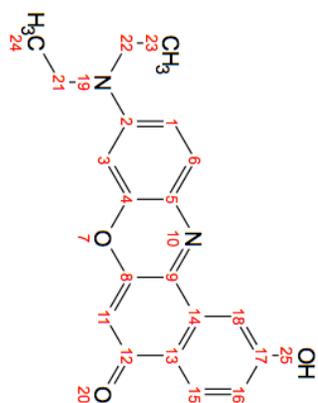
060817_diethoxy-etc_CDCl3_proton
 STANDARD PROTON PARAMETERS



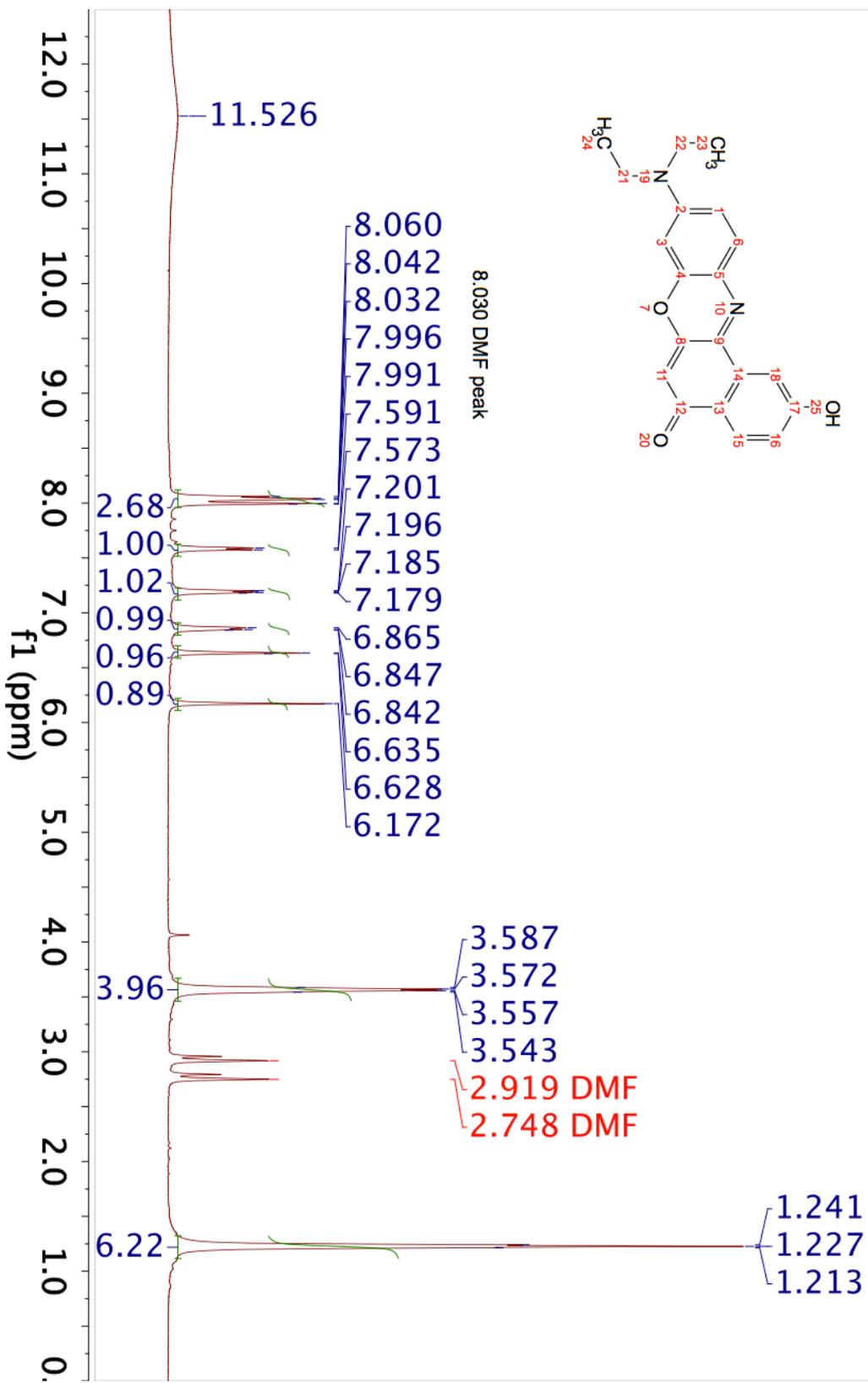
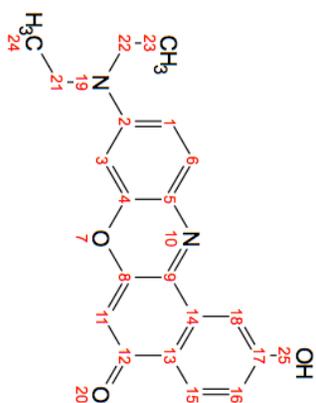
060817_diethoxy-ethoxy-aniline_CDCl3_13C
C13par



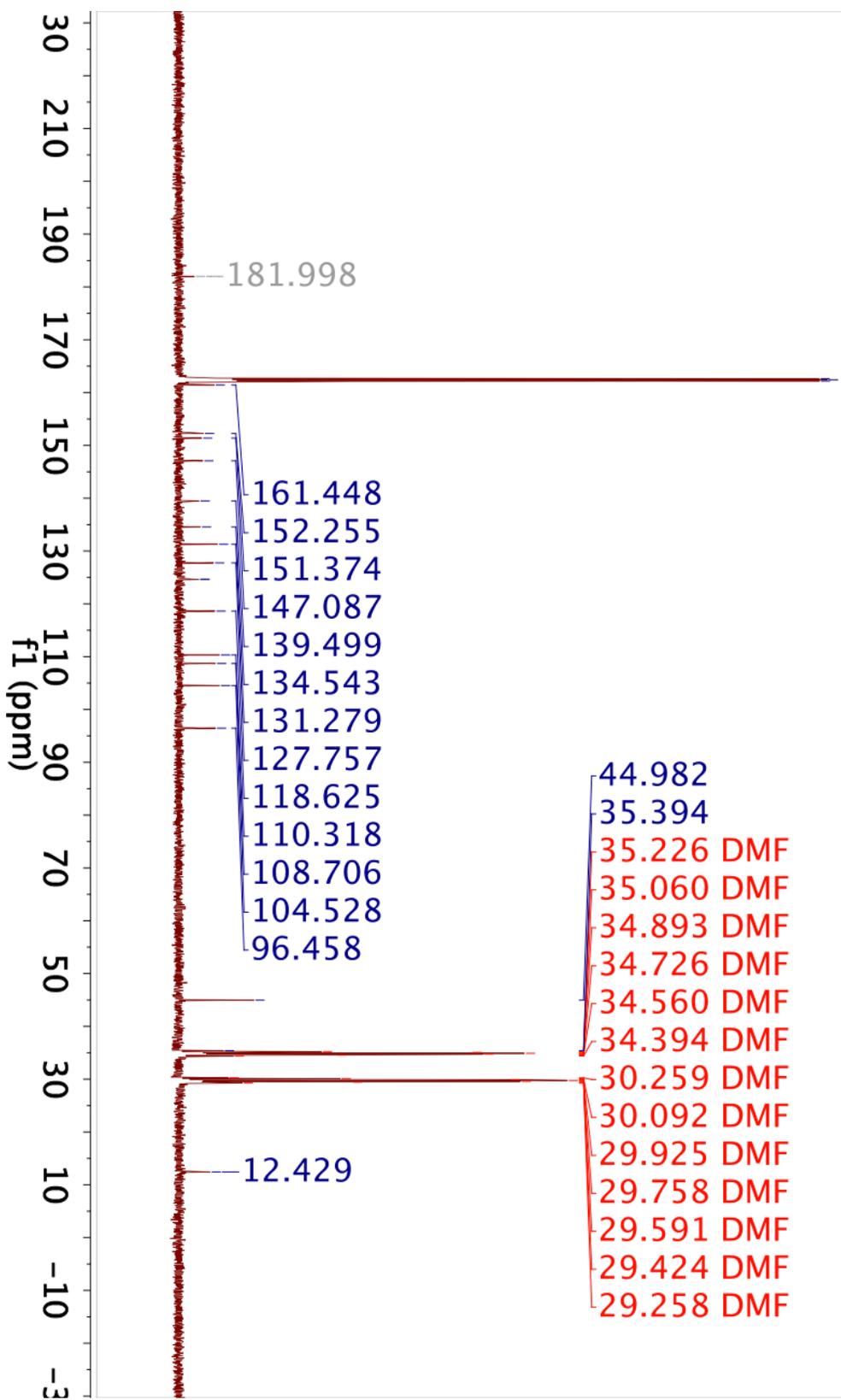
020717_2-hydroxy-Nile-Red-peak1_d7-DMF_proton
 STANDARD PROTON PARAMETERS



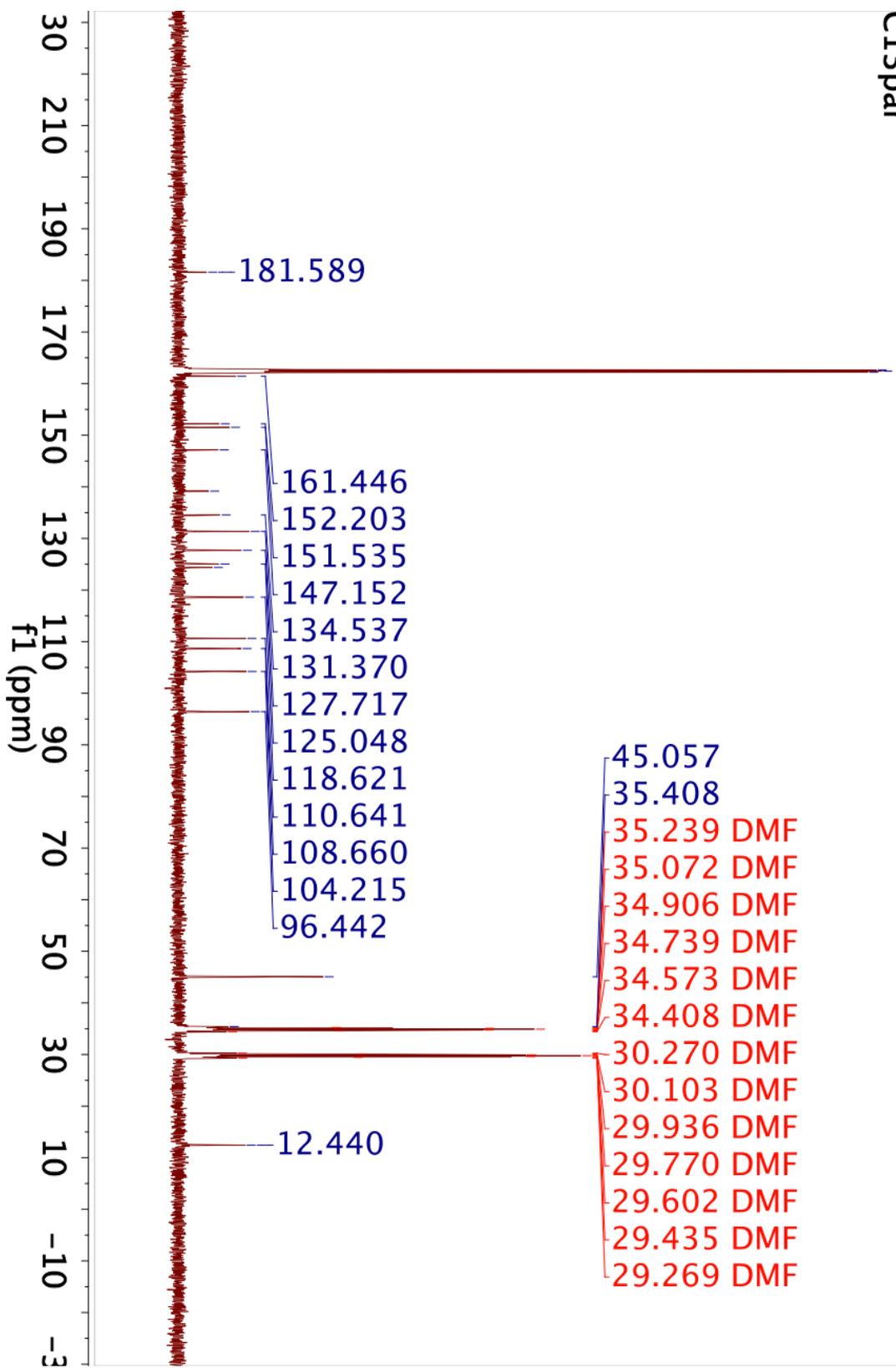
020717_2-hydroxy-Nile-Red-peak2_d7-DMF_proton
STANDARD PROTON PARAMETERS



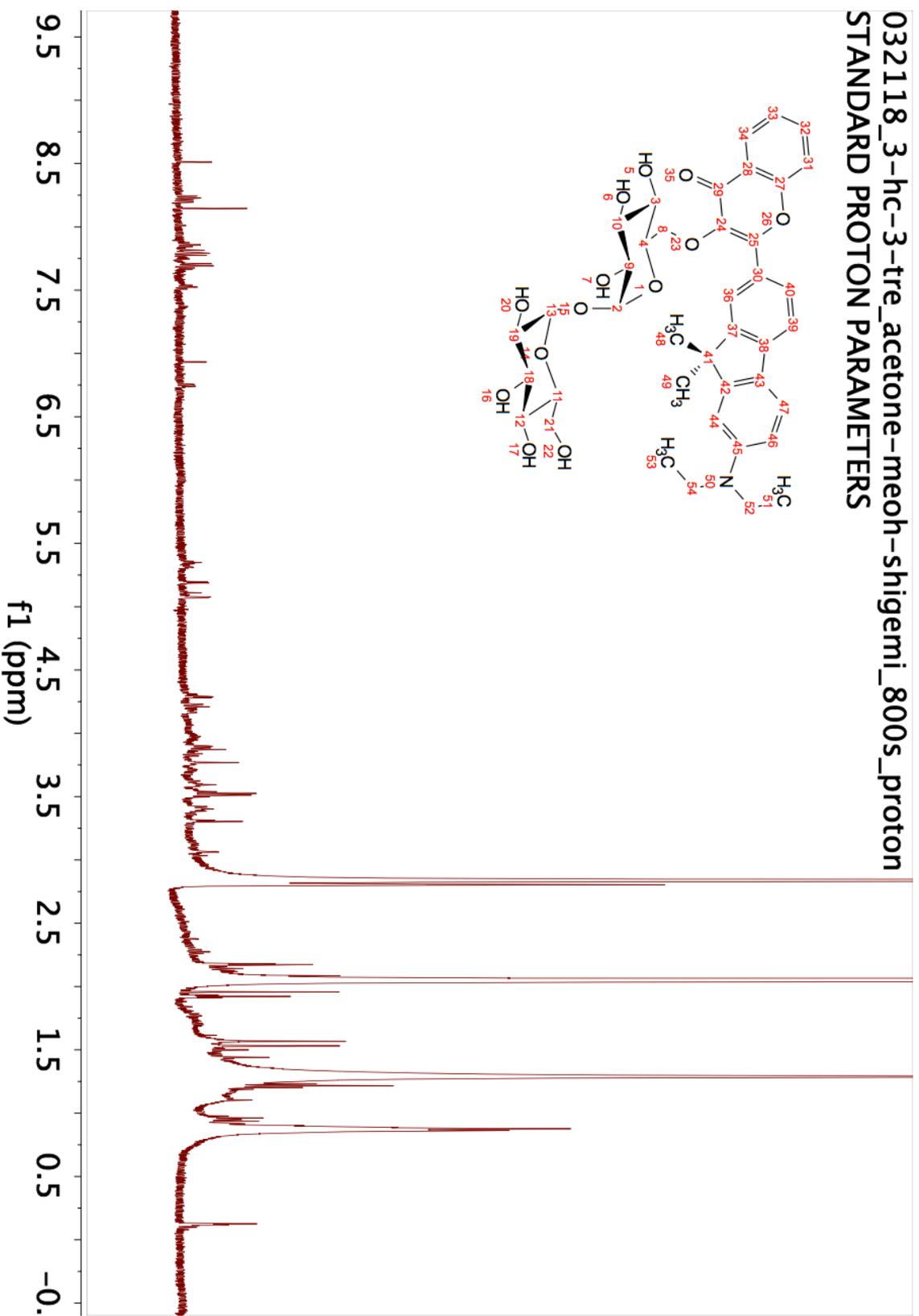
020717_2-hydroxy-Nile-Red-peak1_d7-DMF_13C
C13par



020717_2-hydroxy-Nile-Red-peak2_d7-DMF_13C
C13par

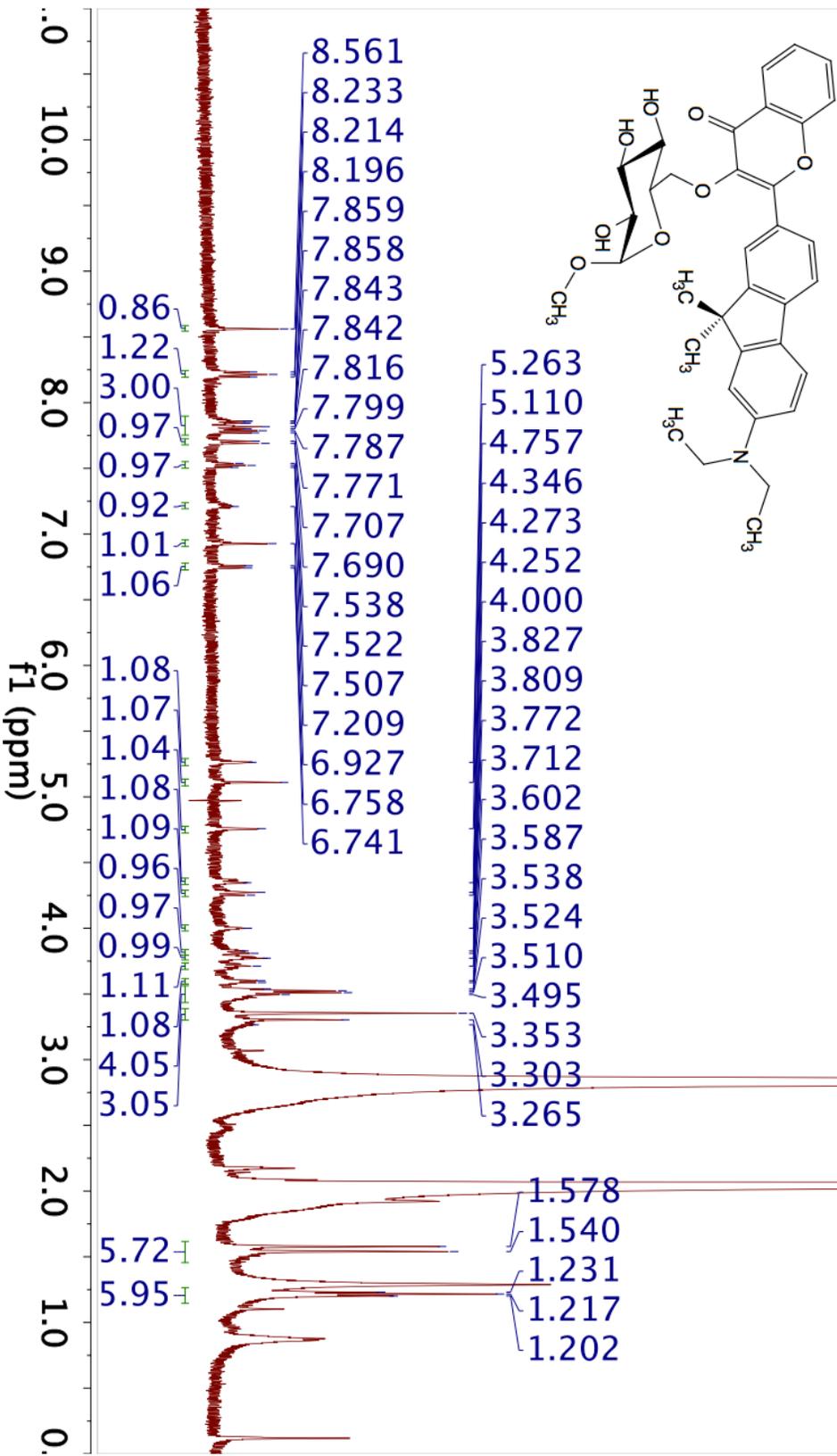


032118_3-hc-3-tre_acetone-meoh-shigemi_800s_proton
STANDARD PROTON PARAMETERS

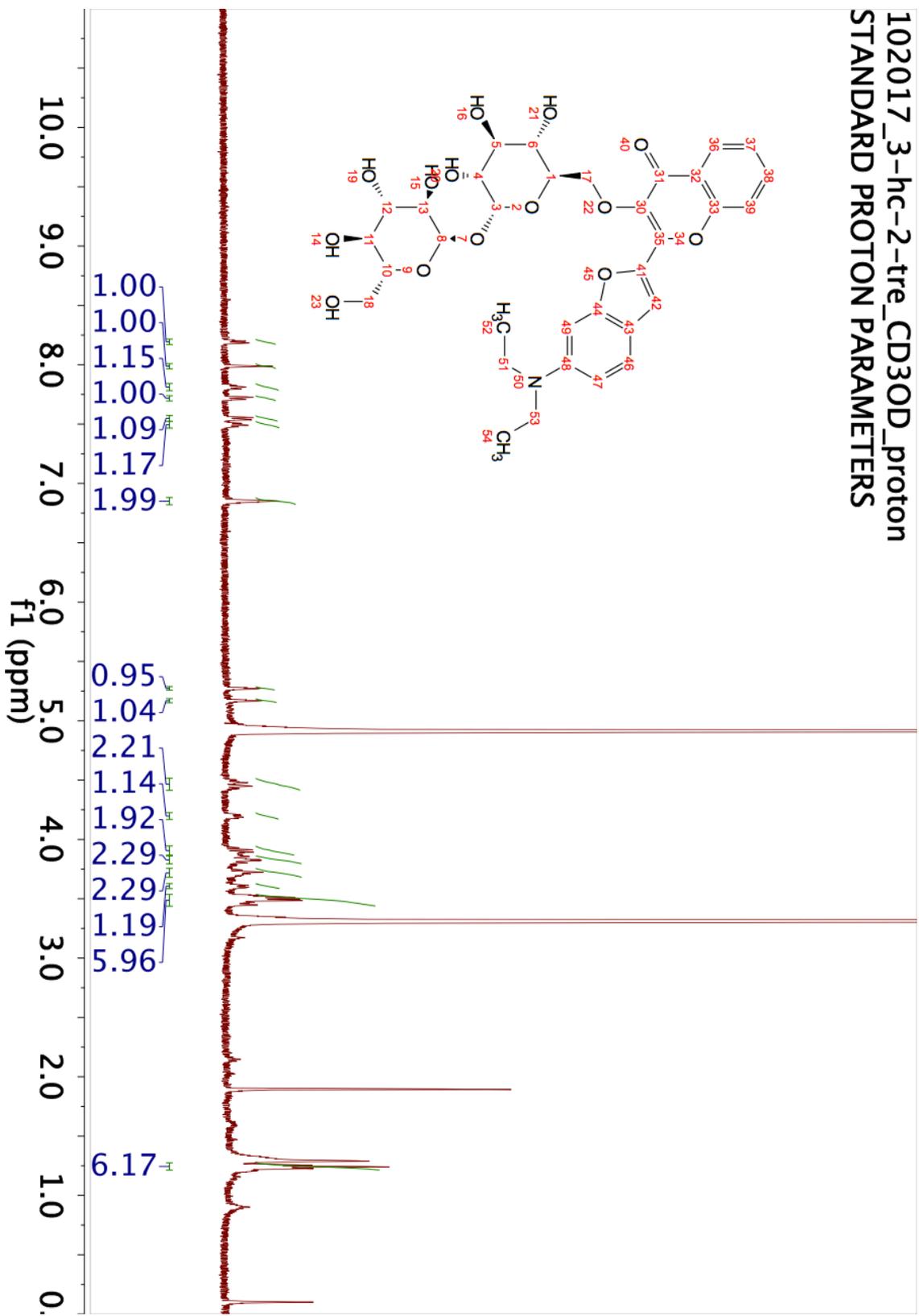


032118_3-hc-3-glc_acetone-mcoh-shigemi_1000s_proton
 STANDARD PROTON PARAMETERS

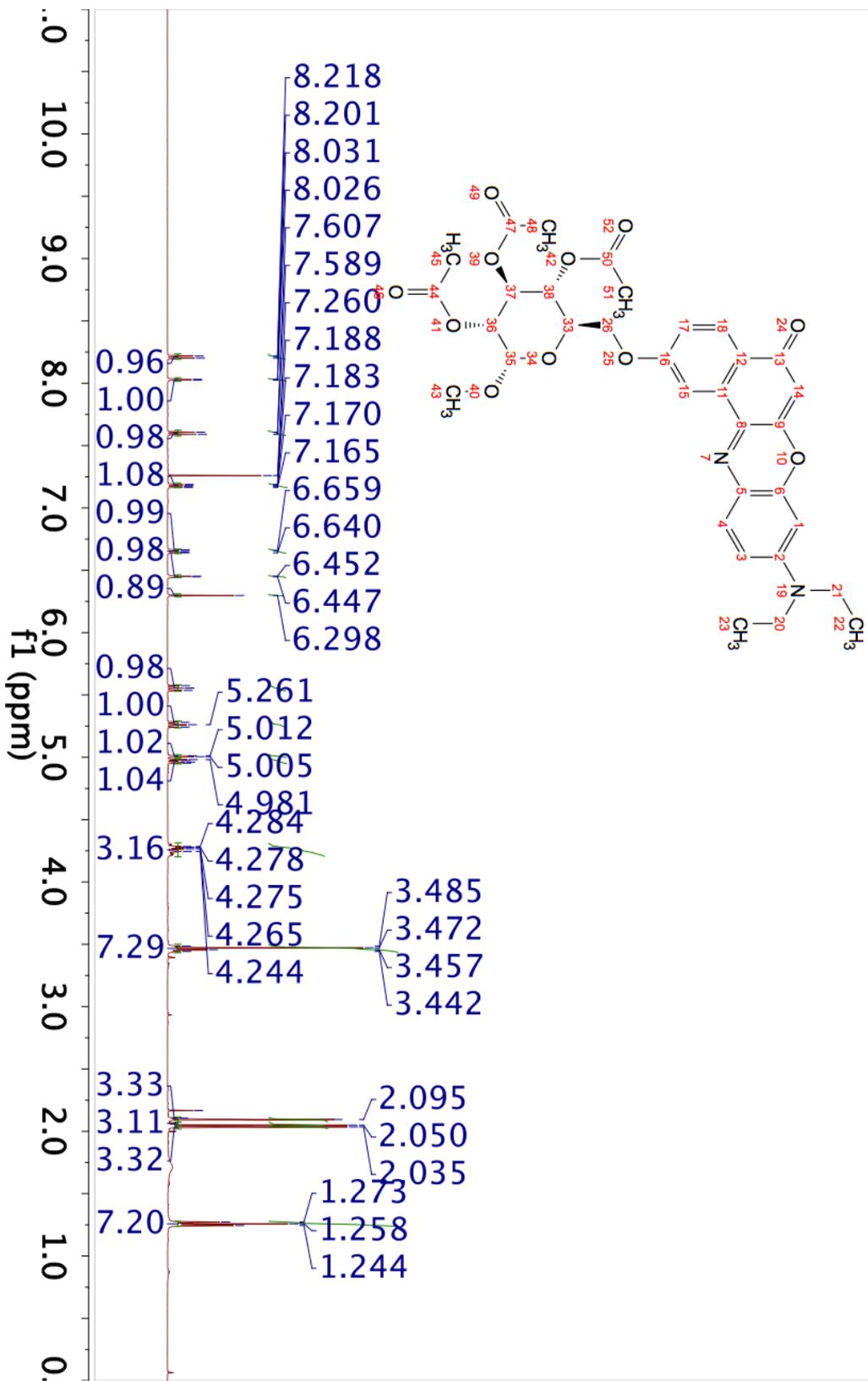
grease (0.88, 1.29); residual methanol from wash (3.31); spinning sidebands for water and acetone

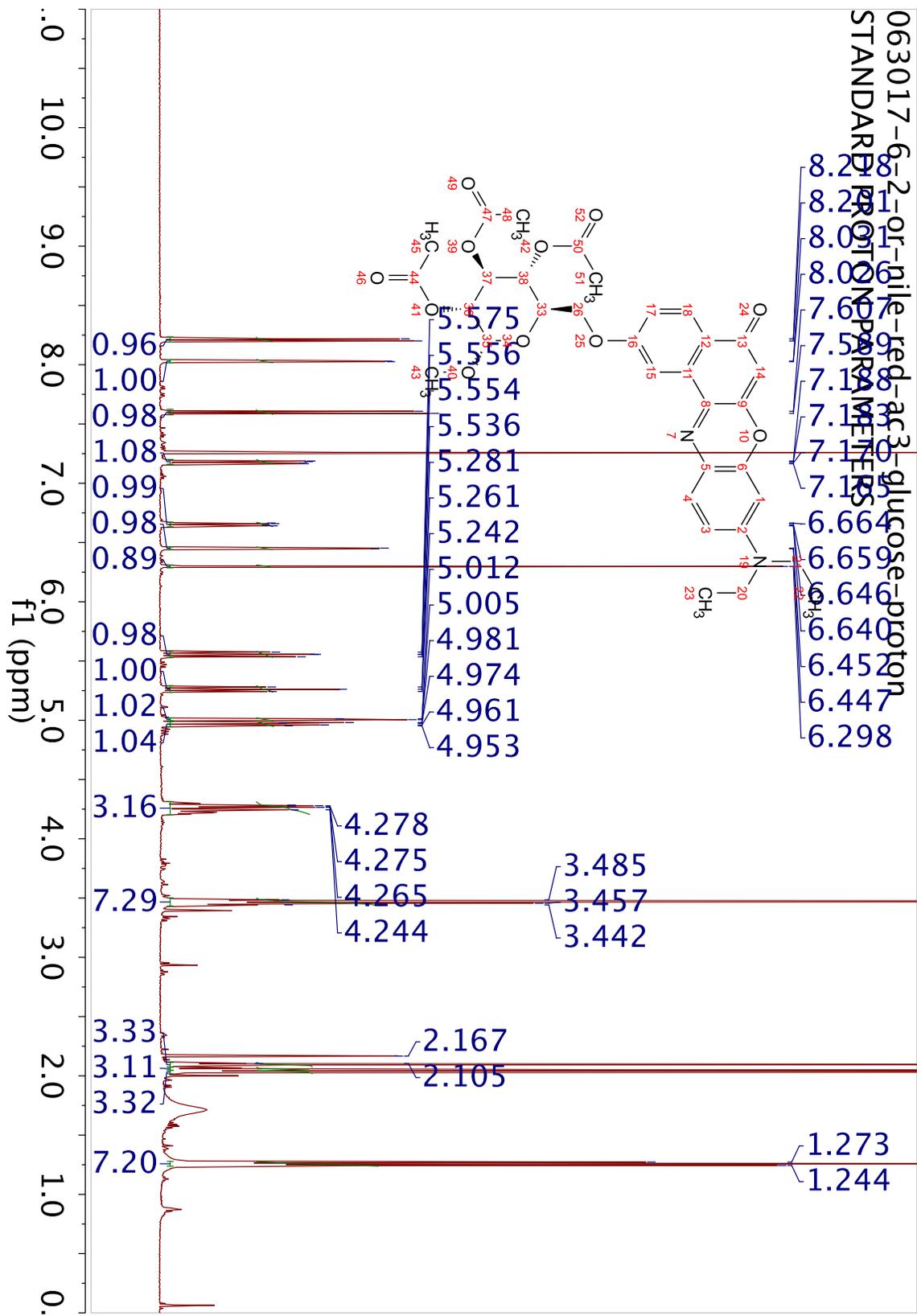


102017_3-hc-2-tre_CD3OD_proton
 STANDARD PROTON PARAMETERS

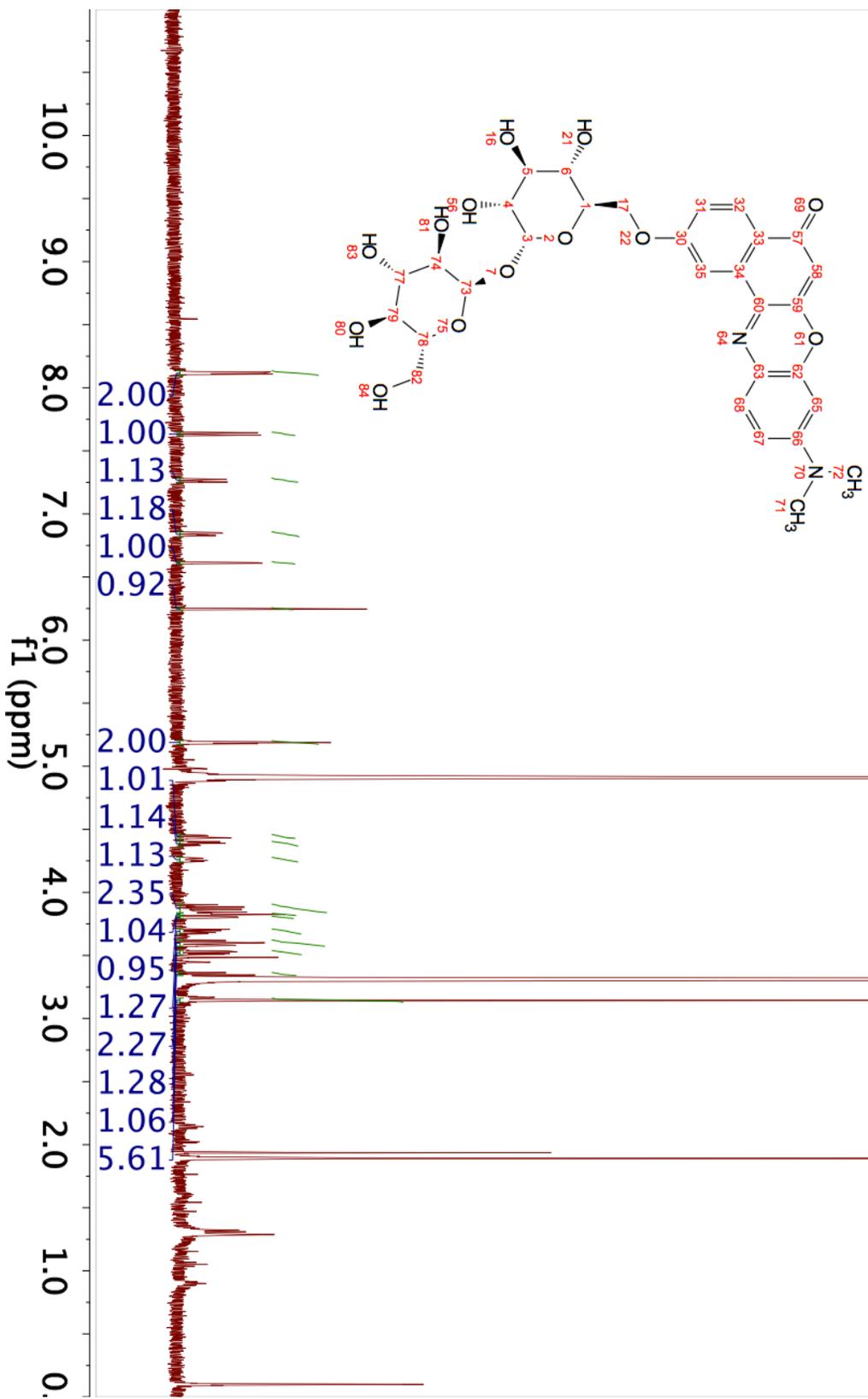


063017-6-2-or-nile-red-ac3-glucose-proton
 STANDARD PROTON PARAMETERS

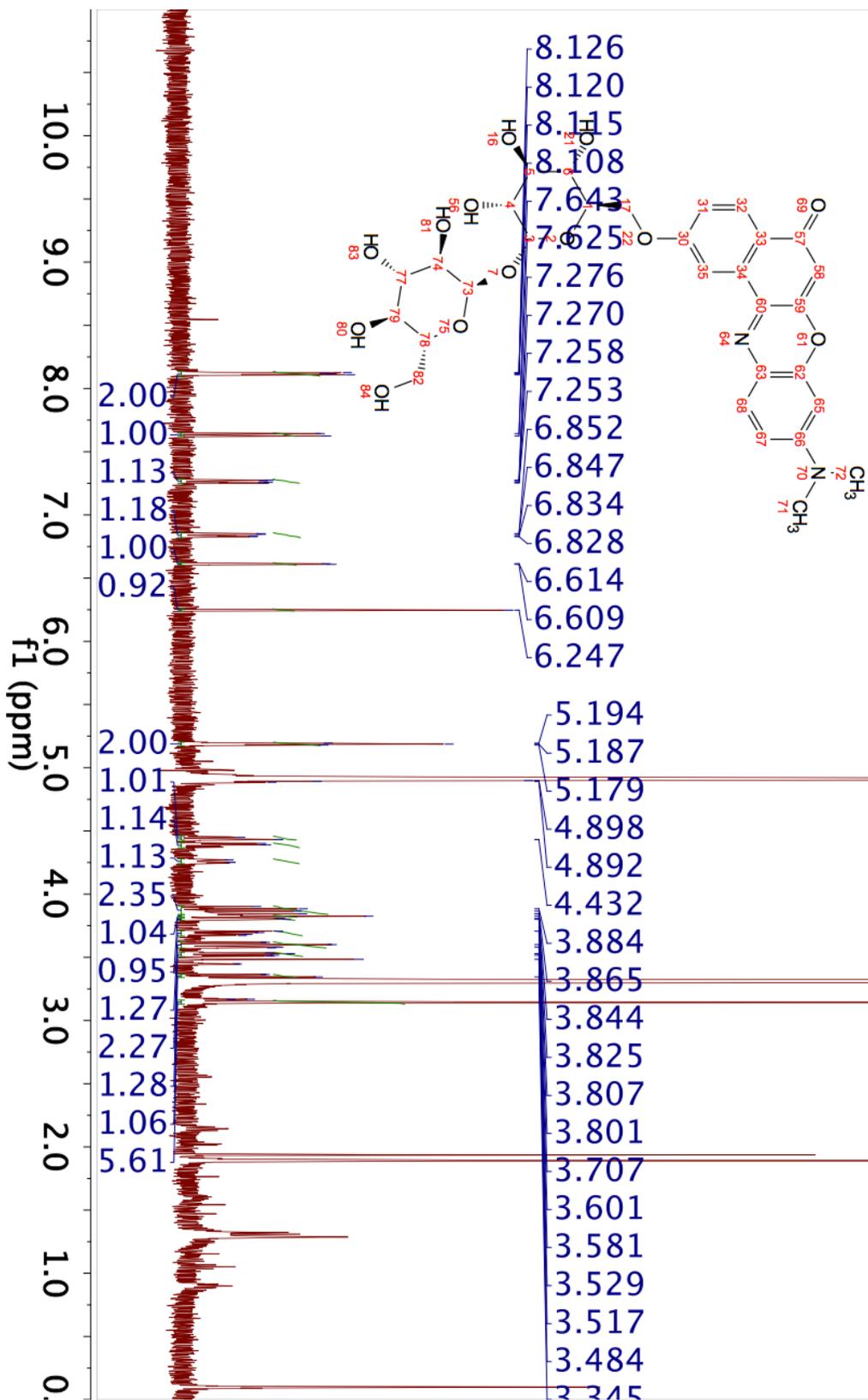




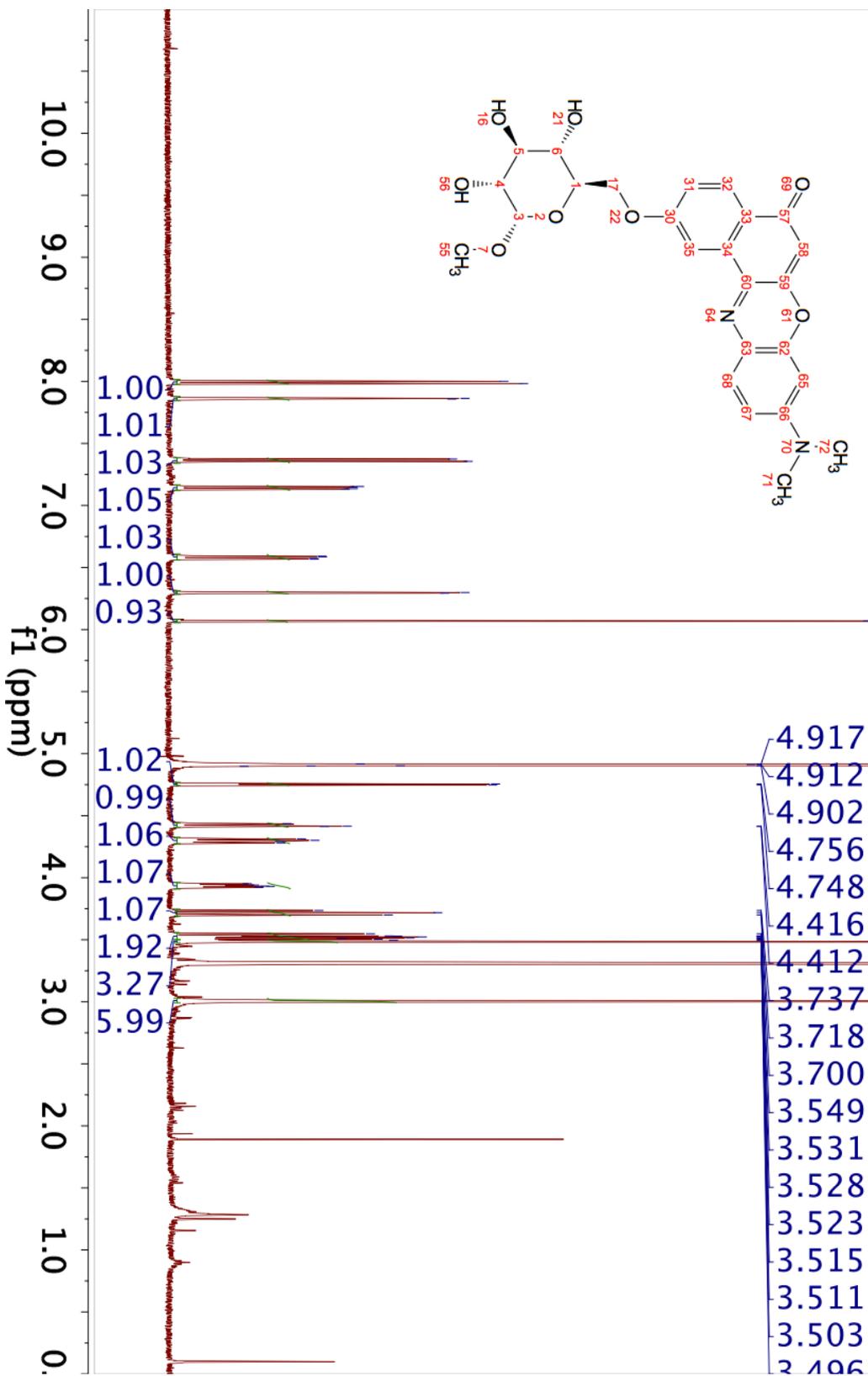
102017_menr-tre_CD3OD_proton
STANDARD PROTON PARAMETERS



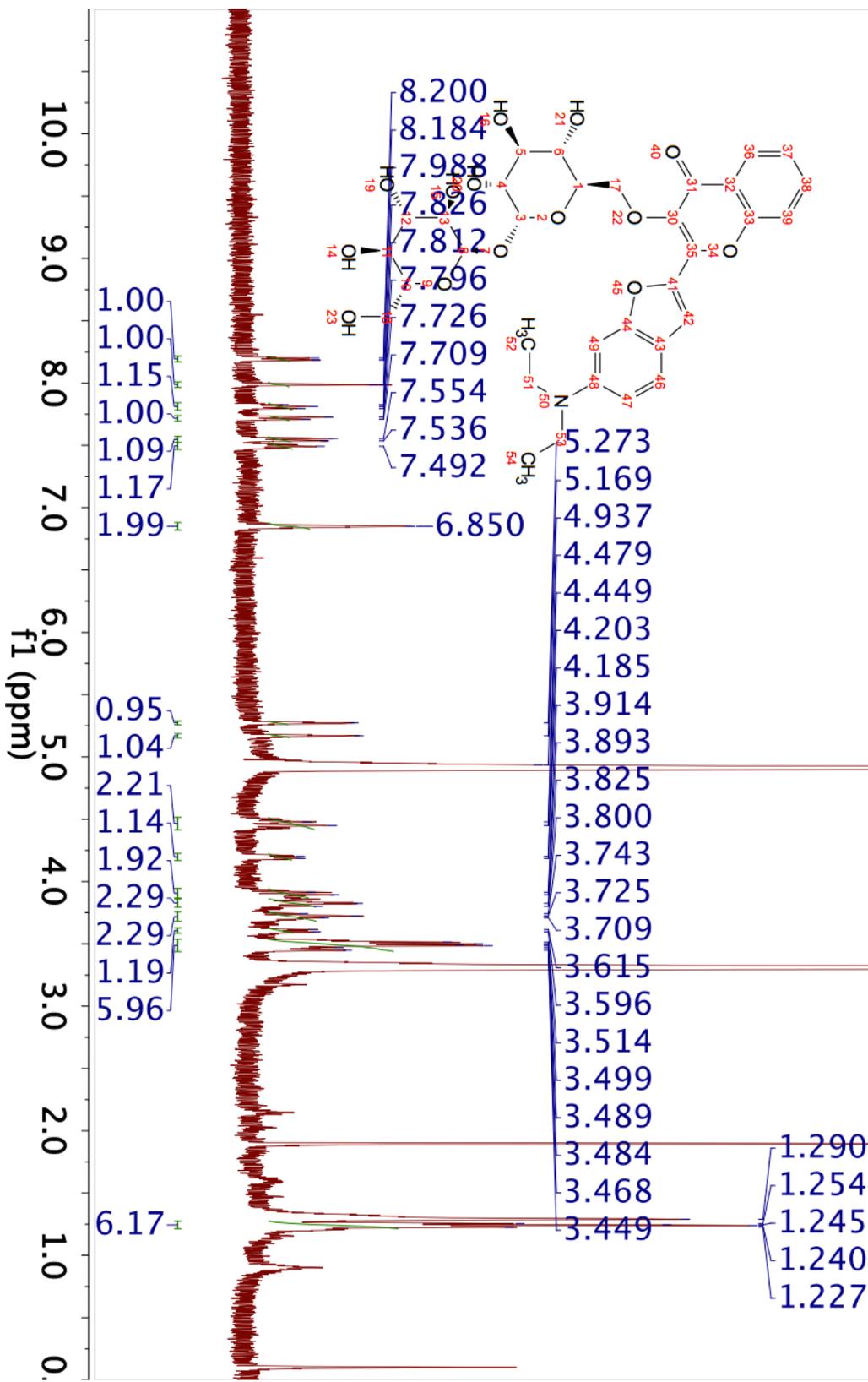
102017_menr-tre_CD3OD_proton
 STANDARD PROTON PARAMETERS



102017_MeNR-glc_CD3OD_proton
 STANDARD PROTON PARAMETERS



102017_3-hc-2-tre_CD3OD_proton
 STANDARD PROTON PARAMETERS



102017_3-hc-2-glc_CD3OD_Proton
 STANDARD PROTON PARAMETERS

