

Chemical Methods for Protein Modification and Cellular Delivery

By

Kalie A. Mix

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Biochemistry)

at the

UNIVERSITY OF WISCONSIN–MADISON

2017

Date of final oral examination: 12/12/17

This dissertation is approved by the following members of the Final Oral Committee:

Ronald T. Raines, *Henry Lardy Professor of Biochemistry*, Biochemistry and Chemistry

David J. Pagliarini, Associate Professor, Biochemistry

M. Thomas Record, *Steenbock Professor in Chemical Sciences*, Biochemistry and Chemistry

Eric S. Shusta, *Howard Curler Distinguished Professor*, Chemical and Biological Engineering

Douglas B. Weibel, Professor, Biochemistry, Biomedical Engineering, and Chemistry

Chemical Methods for Protein Modification and Cellular Delivery

Kalie A. Mix

Under the supervision of Ronald T. Raines

at the University of Wisconsin–Madison

Protein therapeutics comprise a rapidly growing class of drug that has been highly impactful in the clinic. The specificity of protein binding and activity enables effective modulation of biological interactions with lower risk of adverse effects than with small-molecule drugs. Unlike small-molecule drugs, protein therapeutics are limited, to those that exert their activity in an extracellular environment. Major classes of protein therapeutics include hormones such as insulin, monoclonal antibodies that bind to a cell-surface receptor, or enzymes such as lactase and blood clotting factors that act in the digestive tract or bloodstream, respectively. The ability to expand this class of therapeutics to include proteins and enzymes that effect their activity in the intracellular milieu would be transformational because approximately $2/3$ of the proteome is localized within the cell.¹ The ability to deliver proteins to the cytosol could enable replacement of enzymes harboring loss-of-function mutations, and modulation of cellular signaling events. This thesis describes chemical methods for modification of proteins to enable cytosolic delivery.

In Chapter One, I introduce the history of diazo chemistry and its application to the modification of biomolecules. In Chapter Two, I describe the optimization of reactivity and selectivity of a diazo amide to engender chemoselective esterification of protein carboxyl groups. In Chapter Three, I describe derivatization of this molecule with hydrophobic tags. One of these reagents, an α -diazo dimethyl amide, efficiently labels green fluorescent protein and enables its

passage through the plasma membrane of mammalian cells. In Chapter Four, I employ this dimethyl amide diazo reagent for the cellular delivery of an antibody fragment. Finally, in Chapter Five, I describe a method to address a second challenge in the development of antibody therapeutics, namely, site-specific modification. Together, the methods described here provide valuable tools for the development of new protein therapeutics.

Acknowledgments

I am grateful for the many people who have provided advice and guidance throughout my graduate career. First, I would like to thank my advisor, Ronald Raines, for creating a lab environment that fosters creativity, independence, and accountability. I have had the opportunity to learn a wide variety of skills ranging from organic synthesis to cell biology, and to take intellectual ownership of the directions of my project. The “hands-off” management style of the Raines lab takes a lot of trust, and I’ve really appreciated this freedom because it has encouraged me to be accountable for all the details of my research.

I would like to thank my colleagues in the Raines lab for their thought-provoking discussions, feedback on my project, and companionship. John Lukesh, Brett Vanveller, Trish Hoang, Leland Hyman, Kristen Andersen, Nick McGrath and Caglar Tanrikulu all patiently helped me learn new technical skills that I needed for my project to succeed. Sam Orke, Val Ressler, Emily Garnett, and Aubrey Ellison, who joined the lab around the same time as I did, have provided a great deal of support over the years by helping troubleshoot experiments in lab or just taking a break to go get cupcakes at Hilldale. I am grateful to Matt Aronoff for sharing his enthusiasm for diazo chemistry and organic reaction mechanisms, for co-authoring our review on diazo compounds, and for graciously hosting me when I’ve visited Zurich since his graduation. I am also especially grateful to Robert Newberry for patiently sharing an office with me and changing the way I planned experiments by making sure I was thinking critically about the big-picture purpose.

I would like to thank my committee members, both past and present, for providing feedback on my research from a breadth of backgrounds: Professors Laura Kiessling, Patricia Keely,

David Pagliarini, Thomas Record, Eric Shusta, and Douglas Weibel. Their insight has challenged me to think creatively about the future directions of my project and perform experiments I would not have devised on my own. Additionally, Professors Christina Hull and James Keck have provided career advice and guidance as my Molecular Biosciences Training Grant mentors. I would also like to thank Andrew Gardner, Rick Bunt, and Jim Larrabee, who mentored me at Middlebury as I was beginning my scientific career.

I have also been fortunate to work with a number of collaborators during my graduate career. At UW–Madison: Carrie Marshall, Vanessa Grosskopf, Ben Umlauf, Eric Shusta, Danielle Lohman, Matt Stefely, David Pagliarini, Sandy Tseng and Aseem Ansari. Outside UW–Madison: Maria Glanz and Christian Hackenberger (FMP-Berlin), Henry Herce (Dana Farber Cancer Institute), Emily Derbyshire (Duke U.), Amy Weeks and James Wells (UCSF). At Massachusetts Institute of Technology and the Broad Institute, I would like to thank Glen Paraddis, Wendy Salmon, and Nadine Elowe for help with instrumentation. Amit Choudhary, Mike Palte, and the students in Stuart Schreiber’s group were especially kind and welcoming as I was getting settled at the Broad Institute.

I would also like to thank scientific and program staff from UW–Madison for providing exceptional resources and making the program run so smoothly. Martha Vestling, Mark Anderson, Milo Westler, Darrell McCaslin, Dan Stevens, Rachael Sheridan, Elle Grevstad, and Greg Barrett-Wilt provided high quality core facilities as well as technical and experimental advice over the years that has been invaluable. Robin Davies, Laura Vanderploeg, Kaine Korzekwa, Colleen Clary, Matthew Jones, Crystal Peterson, Brenda Renaud, Paul Daniels, and Julie Kennedy keep all aspects of the department, such as facilities, purchasing, AV/tech, and travel running seamlessly. Cara Jenkins (UW–Madison) and Christiana Kalfas (MIT) have

helped manage the move to MIT and other tasks within the Raines group. I am especially grateful to Kate Ryan for her help managing the many logistical challenges of graduate school.

I have been very lucky to have such a close group of friends in Madison, and my experience here would not have been the same without them. All of my IPiB classmates, especially Danielle Lohman, Jake Chung, Sandy Tseng, Brandon Feinen, and Shane Bernard have provided a strong network of encouragement.

Last but not least, I cannot thank my family members enough for their support over the years. I am fortunate to have an extended family that has taken continued interest in my research and traveled all around the U.S. for my undergraduate graduation, my thesis defense, and watching various sporting events. My sister, Anna Mix, inspires me to pursue my goals with the same kind of passion and drive as she has always had. Finally, I cannot express enough gratitude to my parents, Joe and Lorna Mix. They have been role models in their own careers, and they have always pushed me to do my best in mine.

Table of Contents

Abstract.....	i
Acknowledgments	iii
Table of Contents.....	vi
List of Tables	xiii
List of Schemes	xiv
List of Figures.....	xv
List of Abbreviations	xviii
Chapter One	
Introduction: Diazo Compounds: Versatile Tools for Chemical Biology.....	1
1.1 Introduction	2
1.2 Natural Products	3
1.3 Amino Acids.....	6
1.4 Preparation.....	7
1.5 Cycloadditions.....	10
1.6 Probes	12
1.7 Protein Alkylation.....	13
1.8 Bioreversible Protein Modification	17
1.9 Peptide and Protein Modification with Carbenoids.....	18
1.10 Nucleic Acid Alkylation.....	19
1.11 Outlook.....	21

1.12 Acknowledgments	22
Chapter Two	
Optimized Diazo Scaffold for Protein Esterification	23
2.1 Introduction	24
2.2 Results and Discussion	25
2.3 Acknowledgments	32
2.4 Materials and Methods	32
2.4.1 General.....	32
2.4.2 Chemical Synthesis.....	33
2.4.3 Measurement of Reaction Rate Constants.....	46
2.4.4 Esterification of BocGlyOH	48
2.4.5 Esterification of Other Small Molecules	54
2.4.6 Protein Labeling	57
2.4.7 Ultraviolet Spectra of Diazo Compound 2.2	58
2.4.8 NMR Spectra	59
Chapter Three	
Cytosolic Delivery of Proteins by Bioreversible Esterification	96
3.1 Introduction	97
3.2 Results and Discussion	98
3.3 Acknowledgments	108
3.4 Materials and Methods	108

3.4.1 General.....	108
3.4.2 Chemical Synthesis.....	109
3.4.3 Protein Preparation	117
3.4.4 Protein Esterification	122
3.4.5 Mammalian Cell Culture	127
3.4.6 Flow Cytometry.....	128
3.4.7 Confocal Microscopy	129
3.4.8 Esterification Reversibility	130
3.4.9 Cytotoxicity Assay	131
3.4.10 NMR Spectra	132

Chapter Four

Cellular Delivery of anti-GFP Antigen-binding Fragment (Fab).....	143
4.1 Introduction	144
4.2 Results and Discussion.....	146
4.3 Acknowledgments	152
4.4 Materials and Methods	152
4.4.1 General.....	152
4.4.2 Chemical Synthesis.....	154
4.4.3 Production of anti-GFP Fab.....	154
4.4.4 Modification of Fab with Oxaziridine–Azide 4.1	154
4.4.5 Modification of Fab–oxa–N ₃ with DIBAC–Cy3.....	155
4.4.6 Modification of Fab–Cy3 with Diazo Compound 3.1	155

4.4.7 Cell Culture.....	155
4.4.8 Flow Cytometry.....	156
4.4.9 Confocal Microscopy	156

Chapter Five

Site-Specific Antibody Functionalization Using Tetrazine–Styrene Cycloaddition.....	157
5.1 Introduction	158
5.2 Results	161
5.3 Discussion.....	170
5.4 Acknowledgments	172
5.5. Materials and Methods	173
5.5.1 General.....	173
5.5.2 Chemical Synthesis.....	173
5.5.3 Styrene and <i>trans</i> -Cyclooctene Stability	176
5.5.4 Tetrazine–Styrene NMR Kinetics	177
5.5.5 Yeast Surface Display	177
5.5.6 EPL + IEDDA Cycloaddition of Yeast Surface-Displayed Proteins	177
5.5.7 SDS–PAGE and Immunoblotting of Reacted Proteins	178
5.5.8 FITC Titration.....	178
5.5.9 Flow Cytometry.....	179
5.5.10 Fluorescence Microscopy	179
5.5.11 NMR Spectra	180
5.5.12 LC–MS Chromatograms.....	182

Chapter Six

Future Directions	184
6.1. Further Optimization of Esterification Chemistry	184
6.2. Improvement of Cytosolic Delivery Efficiency	186
6.3. Investigation of Ester Stability in Cells and Serum.....	187
6.4. Cellular Delivery of Functional Proteins and Enzymes	189

Appendix One

Cellular Delivery of Green Fluorescent Protein by Cell-Penetrating Peptides Using Diazo Compound-Mediated Esterification	191
A1.1 Introduction.....	192
A1.2 Results and Discussion	194
A1.3 Future Directions	198
A1.4 Acknowledgments	199
A1.5 Materials and Methods	199
A1.5.1 General.....	199
A1.5.2 Chemical Synthesis.....	200
A1.5.3 Peptide Synthesis	201
A1.5.4 GFP Labeling	204
A1.5.5 Cell Culture and Confocal Microscopy	205

Appendix Two

Cellular Delivery of Cas9 for Genome Editing	206
A2.1 Introduction.....	207

A2.2 Results and Discussion	208
A2.3 Future Directions	217
A2.4 Acknowledgments	218
A2.5 Materials and Methods	219
A2.5.1 General.....	219
A2.5.2 Chemical Synthesis.....	219
A2.5.3 Design of crRNA	220
A2.5.4 In Vitro DNA Cleavage.....	220
A2.5.5 RNA Esterification	220
A2.5.6 Cas9 Protein Labeling.....	221
A2.5.7 Delivery of Protein in Cell Culture.....	221
A2.5.8 Detection of Genomic Modifications Using T7E1 Assay	222
A2.5.9 Confocal Microscopy of Cas9–sgRNA/ATTO Internalization	223
Appendix Three	
Synthesis of a New Collagen Mimetic Peptide	225
A3.1 Introduction.....	226
A3.2 Results and Discussion	230
A3.3 Future Directions	232
A3.4 Materials and Methods	232
A3.4.1 General.....	232
A3.4.2 Chemical Synthesis.....	233
A3.4.3 Peptide Synthesis	236

Appendix Four

Synthesis of Demethoxy Q Derivatives for Biochemical Investigation of COQ9 Structure and Function	237
A4.1 Introduction.....	238
A4.2 Results and Discussion	239
A4.3 Future Directions	242
A4.4 Acknowledgments	242
A4.5 Materials and Methods	242
A4.5.1 General.....	242
A4.5.2 Chemical Synthesis.....	243
A4.5.3 Liposome Flootation Assay	244
A4.5.4 NMR Spectra	246
References	247

List of Tables

Table 1.1 Diazo compounds that esterify proteins	15
Table 3.1 Notional effect of esterification on the electrostatic surface of GFP	102

List of Schemes

Scheme 2.1 Mechanism of the reaction between a diazo compound and carboxylic acid in an aqueous solution	24
Scheme 3.1 Synthetic route to diazo compound 3.1	109
Scheme 5.1 Route for the two-step site-specific functionalization of yeast surface-displayed scFv	160
Scheme A3.1 Synthetic route to Ac-(flpHypGly) ₇	231
Scheme A4.1 Synthetic route to DMQ ₂ and DMQ ₉	240

List of Figures

Figure 1.1 Structure and reactivity of some natural products that contain diazo groups	4
Figure 1.2 Preparation of diazo compounds.....	8
Figure 1.3 Diazo compounds in dipolar cycloadditions with strained alkynes	11
Figure 1.4 Diazo compounds for covalent modification of proteins.....	17
Figure 1.5 Covalent modification of nucleic acids using diazo compounds	20
Figure 2.1 Scaffold for testing reactivity and selectivity of diazo compounds	26
Figure 2.2 Rate constants for esterification of BocGlyOH and Hammett analysis.....	27
Figure 2.3 Effect of σ_p value on chemoselectivity of diazo compounds.....	28
Figure 2.4 Chemoselectivity of esterification reactions in aqueous solution.....	29
Figure 2.5 Chemoselectivity of esterification in the presence of an amino group	30
Figure 2.6 MALDI–TOF mass spectrometry data for esterification of RNase A	31
Figure 2.7 ^1H NMR kinetic data for the reaction between compounds 2.1–2.6 and BocGlyOH	47
Figure 2.8 Ultraviolet spectra of diazo compound 2.2	58
Figure 3.1 Esterification and cellular uptake of esterified superfolder GFP	99
Figure 3.2 Optimization of solvent conditions for esterification of superfolder GFP.....	100
Figure 3.3 Identification of residues esterified by diazo compound 3.1	103
Figure 3.4 Images of cellular internalization of GFP and its super-charged and esterified variants.....	104
Figure 3.5 Images of the nuclear internalization of a protein that contains a nuclear localization signal and its esterified variant.....	106
Figure 3.6 MALDI–TOF spectra to assess the reversibility of protein esterification with diazo compound 3.1.....	107

Figure 3.7 Graph of the viability of CHO-K1 cells treated with α -hydroxy dimethylamide 3.7	107
Figure 3.8 MALDI–TOF mass spectra of purified super-charged GFP and nlsGFP	121
Figure 3.9 Representative MALDI–TOF spectra of GFP esterified with diazo compounds 3.1–3.6	124
Figure 3.10 MALDI–TOF spectrum of nlsGFP esterified with diazo compound 3.1	127
Figure 4.1 Modification of anti-GFP Fab with Cy3 dye	147
Figure 4.2 Modification of Fab–Cy3 with diazo compound 3.1	149
Figure 4.3 Quantification of cell uptake using flow cytometry	150
Figure 4.4 Confocal microscopy images of CHO-K1 cells treated with Fab–Cy3 or Fab–Cy3– 3.1	151
Figure 5.1 Stability of candidate reagents	163
Figure 5.2 Kinetics of the tetrazine–styrene reaction	164
Figure 5.3 Functionalization of 4-4-20 scFv by EPL followed by IEDDA	166
Figure 5.4 4-4-20 scFv maintains function after modification with a styrene and labeling with a tetrazine–Cy5	167
Figure 5.5 Internalization of Cy5-labeled scFv's into rat brain endothelial (RBE4) cells ..	169
Figure A1.1 Ligation reagents used in this study	194
Figure A1.2 Conjugation of cyclic cell-penetrating peptides to GFP	195
Figure A1.3 Characterization of GFP conjugates	196
Figure A1.4 Confocal microscopy and DIC images of HeLa cells treated with GFP–cR10	197
Figure A1.5 Confocal microscopy and DIC images of HeLa cells treated with GFP–cR10–Cy3	198
Figure A1.6 MALDI–TOF mass spectrum of azido–cR10–Cy3	202
Figure A1.7 HPLC chromatogram and ESI mass spectrum of diazo–cR10	203

Figure A2.1 GFP-targeting crRNA sequences and in vitro cleavage of GFP gene.....	209
Figure A2.2 Esterification of Cas9 using diazo compound 3.1	210
Figure A2.3 Reaction of AUGC with diazo compound 3.1	212
Figure A2.4 Reaction of pAUGC with diazo compound 3.1	213
Figure A2.5 GFP fluorescence after 48 h incubation with CRISPR components.....	214
Figure A2.6 Characterization of genomic DNA cleavage.....	216
Figure A2.7 Confocal microscopy images of HEK293T–GFP cells treated with Cas9–sgRNA/ATTO components	217
Figure A3.1 CMP invasion of collagen helix at sites of proteolytic or mechanical degradation.....	226
Figure A3.2 Tumor-associated collagen signatures imaged by second harmonic generation	228
Figure A3.3 Structure of CMPs.....	230
Figure A3.4 MALDI–TOF mass spectrum of Ac-(flpHypGly) ₇	231
Figure A4.1 Putative biosynthetic route to coenzyme Q.....	238
Figure A4.2 MALDI–TOF mass spectrum of DMQ ₉	240
Figure A4.3 Association of COQ9 enzyme with liposomes	241

List of Abbreviations

A	adenine
AIBN	azobisisobutyronitrile
Ala or A	alanine
AMP	ampicillin
Arg or R	arginine
ASAP	atmospheric solids analysis probe
Asp or D	aspartic acid
ATCC	American Type Culture Collection
atm	atmosphere
ATP	adenosine triphosphate
BCA	bicinchoninic acid
bFGF	basic fibroblast growth factor
Boc	<i>tert</i> -butoxycarbonyl
BRCA2	breast cancer type 2 susceptibility protein
BSA	bovine serum albumin
C	cytosine
calcd	calculated
Cas9	CRISPR-associated protein 9
CD ₃ CN	acetonitrile
CHO-K1	Chinese hamster ovary -K1
CMP	collagen mimetic peptide
CO ₂	carbon dioxide
CoQ	coenzyme Q
COQ(<i>n</i>)	genes encoding coenzyme Q biosynthetic enzymes
Coq(<i>n</i>)	coenzyme Q biosynthetic enzymes (<i>M. musculus</i>)
COQ7	coenzyme Q biosynthetic enzyme 7 (<i>H. sapiens</i>)
COQ9	coenzyme Q biosynthetic enzyme 9 (<i>H. sapiens</i>)

CPP	cell-penetrating peptide
cR10	cyclic arginine-10
CRISPR	clustered regularly interspaced palindromic repeats
crRNA	CRISPR ribonucleic acid
cTAT	cyclic <i>trans</i> activator of transcription (cell-penetrating peptide)
CuAAC	copper-catalyzed azide–alkyne cycloaddition
Cy3	cyanine 3 (dye)
Cy5	cyanine 5 (dye)
Cys	cysteine
d	doublet
Da	Dalton
DBCO	dibenzocyclooctyne
DBU	1,8-diazabicyclo[5.4.0]undec-7-ene
DCC	N,N'-dicyclohexylcarbodiimide
DCM	dichloromethane
dd	doublet of doublets
Dde	N-(1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl) protecting group
Dha	dehydroalanine
DIEA	diisopropylethylamine
DMEM	Dulbecco's modified Eagle's medium
DMF	dimethylformamide
DMQ	demethoxy Q
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
DON	6-diazo-5-oxo-norleucine
DONV	5-diazo-4-oxo-norvaline
DPBS	Dulbecco's phosphate-buffered saline
DTT	dithiothreitol
ECM	extracellular matrix
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
EDTA	ethylenediaminetetraacetic acid

em	emission
EPL	expressed protein ligation
equiv	equivalents
ESI	electrospray ionization
EtOAc	ethyl acetate
ex	excitation
Fab	Fragment antigen-binding
FBS	fetal bovine serum
FITC	fluorescein isothiocyanate
Flp	(2 <i>S</i> ,4 <i>R</i>)-4-fluoroproline
flp	(2 <i>S</i> ,4 <i>S</i>)-4-fluoroproline
Fmoc	fluorenylmethoxycarbonyl
FRET	Förster resonance energy transfer
g	grams
G (DNA)	guanine
GFP	green fluorescent protein
Glu or E	glutamic acid
Gly or G	glycine
HATU	1-[bis(dimethylamino)methylene]-1 <i>H</i> -1,2,3-triazolo[4,5- <i>b</i>]pyridinium 3-oxid hexafluorophosphate
HBS	HEPES buffered saline
HCl	hydrochloric acid
HEK293T	human embryonic kidney-293 cells
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His or H	histidine
HIV	human immunodeficiency virus
HOAt	1-hydroxy-7-azabenzotriazole
HOMO	highest occupied molecular orbital
HPLC	high-performance liquid chromatography
HRMS	high resolution mass spectrometry
Hyp	(2 <i>S</i> ,4 <i>R</i>)-4-hydroxyproline

hyp	(2 <i>S</i> ,4 <i>S</i>)-4-hydroxyproline
IEDDA	inverse-electron demand Diels–Alder
IPTG	isopropyl β-D-1-thiogalactopyranoside
kDa	kilodalton
kPsi	kilo pounds per square inch
LB	Luria–Bertani
LC–MS	liquid chromatography–mass spectrometry
LL2	<i>Mus musculus</i> lung cells
log <i>P</i>	octanol–water partition coefficient
m	multiplet
<i>m/z</i>	mass-to-charge ratio
MALDI–TOF	matrix-assisted laser desorption ionization–time-of-flight
MeOH	methanol
MES	2-(<i>N</i> -morpholino)ethanesulfonic acid
MESNA	2-mercaptoethanesulfonic acid
mg	milligrams
MHz	megahertz
min	minutes
mL	milliliters
mM	millimolar
mmol	millimoles
MRI	magnetic resonance imaging
mRNA	messenger ribonucleic acid
MS/MS	tandem mass spectrometry
MTS	3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)- 2H-tetrazolium
M_w	molecular mass
MWCO	molecular weight cut-off
NBS	<i>N</i> -bromosuccinimide
NHS	<i>N</i> -hydroxysuccinimide
NLS	nuclear localization signal

nlsGFP	green fluorescent protein construct with nuclear localization signal
nm	nanometer
NMR	nuclear magnetic resonance
OD ₆₀₀	optical density at 600 nm
oxa	oxaziridine
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PC	phosphatidylcholine
PCR	polymerase chain reaction
PDB	Protein Data Bank
PE	phosphatidylethanolamine
PEG	polyethylene glycol
PES	polyethersulfone
PET	positron emission tomography
PG lipid	phosphatidylglycerol
pI	isoelectric point
p <i>K</i> _a	logarithm of the acid dissociation constant
Pro	proline
PTEN	phosphatase and tensin homolog protein
pVEC	murine vascular endothelial-cadherin protein-derived protein
q	quartet
RBE4	rat brain endothelial cell line
ReACT	redox-activated chemical tagging
RNA	ribonucleic acid
RNase A	bovine ribonuclease A
RNP	ribonucleoprotein
rpm	rotations per minute
s	seconds
s	singlet
scFv	single-chain variable fragment
scFv 4-4-20	anti-FITC scFv

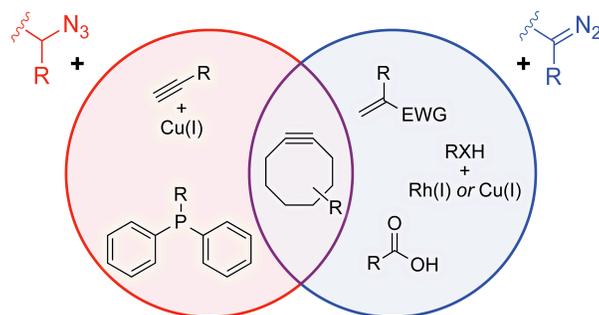
scFVA	endothelial cell surface-binding scFv
SDS	sodium dodecyl sulfate
Ser or S	serine
sfGFP	superfolder green fluorescent protein construct
sgRNA	single guide ribonucleic acid
SHG	second harmonic generation
siRNA	small interfering ribonucleic acid
SPAAC	strain-promoted azide–alkyne cycloaddition
sv40	Simian vacuolating virus 40
T7E1	T7 endonuclease I
TACS	tumor-associated collagen signature
TALEN	transcription activator-like effector nuclease
TAT	<i>trans</i> -activator of transcription
<i>t</i> BuOH	<i>tert</i> -butanol
TCEP	tris(2-carboxyethyl)phosphine
TCO	<i>trans</i> -cyclooctene
TE	tris-ethylenediaminetetraacetic acid
TEA	triethylamine
TEV	tobacco etch virus
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TIS	triisopropylsilane
TLC	thin-layer chromatography
TMS	trimethylsilane
TP10	transportan peptide 10
tracrRNA	trans-activating CRISPR RNA
U	uracil
v/v	volume-to-volume ratio
WGA	wheat germ agglutinin
Z	net molecular charge
ZFN	zinc finger nuclease

ε	extinction coefficient
μL	microliters
μmol	micromols

Chapter One

Diazo Compounds: Versatile Tools for Chemical Biology*

*This chapter has been published, in part, under the same title. Reference: Mix, K.A., Aronoff, M.R. Raines, R.T. (2016) *ACS Chem. Biol.* 11, 3233–3244.



Abstract: Diazo groups have broad and tunable reactivity. That and other attributes endow diazo compounds with the potential to be valuable reagents for chemical biologists. The presence of diazo groups in natural products underscores their metabolic stability and anticipates their utility in a biological context. The chemoselectivity of diazo groups, even in the presence of azido groups, presents many opportunities. Already, diazo compounds have served as chemical probes and elicited novel modifications of proteins and nucleic acids.

Here, we review advances that have facilitated the chemical synthesis of diazo compounds, and we highlight applications of diazo compounds in the detection and modification of biomolecules. In Chapter Two, I describe optimization of the reactivity and selectivity of a series of diazo amides for esterification of proteins in an aqueous solution. The modularity of these diazo amides also enables facile derivatization with any amine-bearing functional group. In Chapter Three, I utilize this modularity to synthesize a second series of diazo compounds that maintains optimized reactivity and selectivity but contain functional groups that alter the polarity of the reagent. One compound, an α -diazo dimethyl amide, enables cytosolic delivery of green fluorescent protein (GFP). This esterification method is easily adapted to any native protein of

interest and could be especially valuable for cellular delivery of protein therapeutics. In Chapter Four, I use this reagent to esterify an antibody fragment (Fab) and enable its cellular uptake. In addition to cellular delivery, a second challenge in the development of antibody-based therapeutics is site-selective modification. In Chapter Five, I describe a new protein chemistry method that combines expressed protein ligation (EPL) with Diels–Alder cycloaddition to site-selectively modify antibody single-chain variable fragments (scFv) with probes.

Author Contributions: Kalie A. Mix, Matthew R. Aronoff, and Ronald T. Raines wrote this chapter.

1.1 Introduction

Azido groups dominate the current landscape of chemoselective reactions in chemical biology. Yet, diazo groups have attributes that are even more desirable than those of azido groups. For example, diazo groups ($R^1R^2C=N_2$) are smaller than analogous azido groups ($R^1R^2HC-N_3$), and diazo groups display a broader range of reactivity.^{2,3}

The simplest diazo compound, diazomethane, is a yellow gas that was discovered by von Pechmann in 1894^{4,5} and is a common reagent in synthetic organic chemistry. Diazomethane and other diazoalkanes are, however, highly toxic⁶⁻⁸ and explosively reactive,^{9,10} and have little utility in the context of chemical biology. The problem arises from their high basicity, as protonation of the α carbon of a diazo group leads to the formation of a diazonium species ($R^1R^2HC-N_2^+$) poised for a rapid S_N2 reaction that releases nitrogen gas.

Recent advances in synthetic methodology provide ready access to “stabilized” diazo compounds that are compatible with living systems. The stability arises from diminished basicity

due to delocalization of the electrons on the α carbon to another functional group. Such stabilized diazo compounds have the potential for widespread application in chemical biology.

Here, we review the use of diazo compounds in chemical biology. We begin with an overview of natural products and amino acids that contain a diazo group. That is followed by a summary of methods for the chemical synthesis of diazo compounds. We then highlight the remarkable versatility of diazo compounds in the context of chemical biology, and we end with a brief prospectus for the future.

1.2 Natural Products

In contrast to azido groups,¹¹ diazo groups are found in many natural products.¹² Isotopic labeling studies and genome mining have provided insight into their biosynthesis.¹³⁻¹⁶ No enzyme is known to catalyze the formation of an N–N bond, though a gene cluster that encodes a nitrous acid-producing enzyme could be a source.¹⁷ Intrinsic antitumor and antibiotic activities endow some natural diazo compounds with potential clinical utility, but mechanisms of action *in vivo* are unclear. As the isolation and synthesis of diazo-containing natural products have been reviewed extensively elsewhere,^{18,19} we summarize only key findings and recent advances. We focus, in particular, on the kinamycins and lomaiviticins, two classes of natural products with unusual structures and intriguing mechanisms of reactivity (Figures 1.1A and 1.1B).

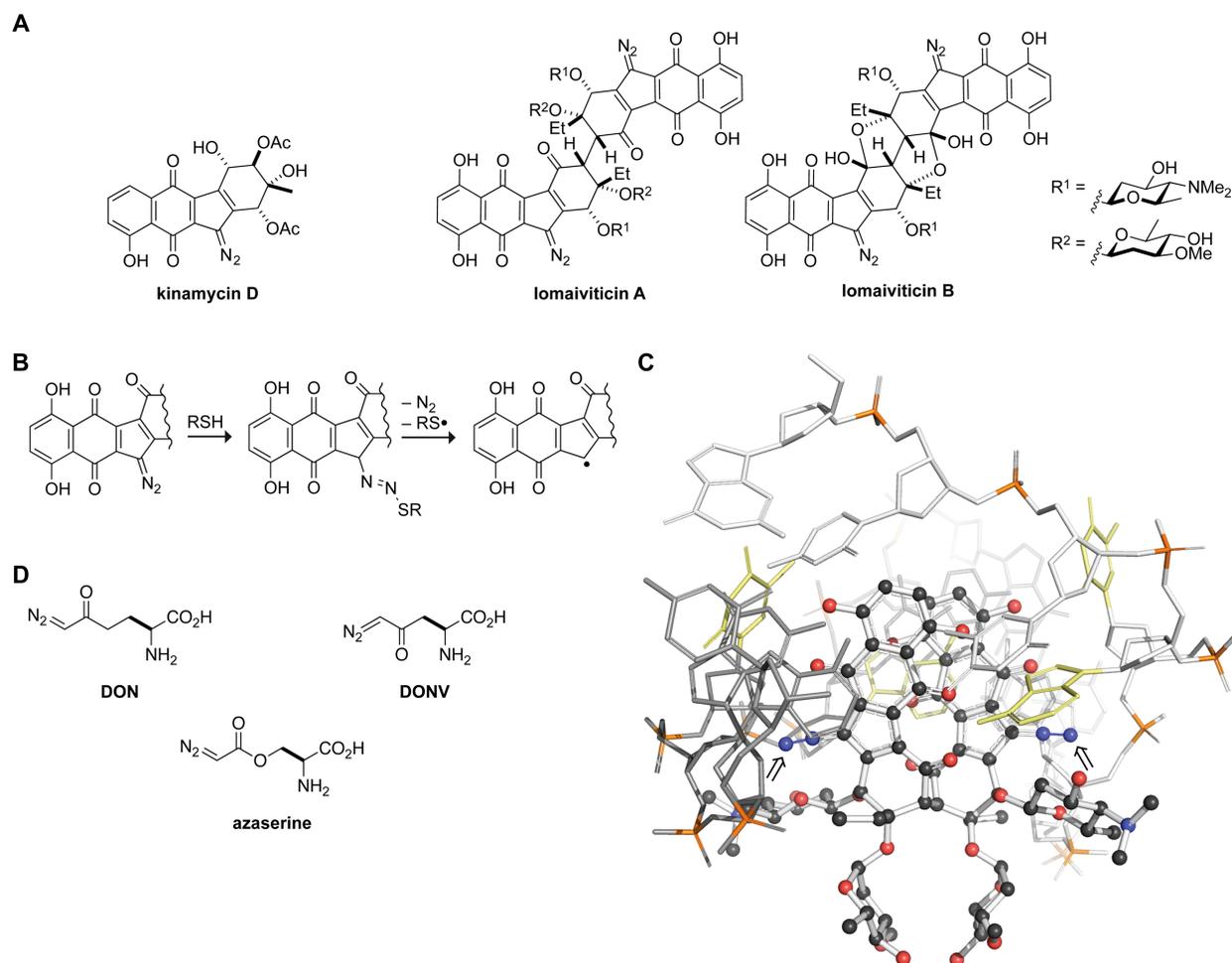


Figure 1.1 Structure and reactivity of some natural products that contain diazo groups. (A) Kinamycin D, lomaiviticin A, and lomaiviticin B. (B) Putative mechanism for the generation of a reactive vinylogous radical from lomaiviticin A.²⁰ (C) Solution structure of the complex of lomaiviticin A with a G-C-T-A-T-A-G-C duplex.²¹ Displaced A-T base pairs are depicted in yellow. Phosphorous atoms are depicted in orange. Hydrogen atoms are not shown. Arrows point to the two diazo groups. Image was created with Protein Data Bank entry 2n96 and the program PyMOL from Schrödinger (New York, NY). (D) Amino acids that contain diazo groups.

The kinamycins were isolated from *Streptomyces murayamaensis* in 1970 and displayed antimicrobial activity against gram-positive bacteria.²² Initially, the compounds were thought to contain a cyanamide group due to their infrared absorption near $\sim 2155\text{ cm}^{-1}$, but were later established to have a diazo moiety.²³ The complex architecture of these molecules, which consist

of a 4-ring carbocyclic skeleton that contains several stereogenic centers, challenged synthetic chemists until routes were developed a decade ago.²⁴⁻²⁶

Like the kinamycins, the lomaiviticins are analogs of 9-diazofluorene (Figure 1.1A). Lomaiviticins A and B were isolated in 2001 from the marine ascidian symbiont *Salinispora pacifica* and displayed antitumor activity at sub-micromolar concentrations.²⁷ Lomaiviticins C–E were isolated in 2012 from *Salinispora pacifica* and demonstrated similar potency.²⁸ Although synthetic routes to the lomaiviticins are unrealized to date, progress has been made towards intermediates and analogues.²⁹⁻³³

Diazofluorene analogues have long been used to investigate possible mechanisms of DNA cleavage *in vitro*. Using 9-diazofluorene, Arya and Jebaratnam were among the first to suggest that a diazo group could mediate DNA cleavage.³⁴ Kinafluorenone, which contains a ketone oxygen in lieu of a diazo group, displayed no antibiotic activity and thus supported the hypothesis that the diazo moiety is the active pharmacophore.³⁵ A variety of reactive intermediates that elicit cytotoxicity have been proposed, including a covalent adduct,^{36,20} *ortho*-quinone methide,^{37,20} acylfulvene,²⁸ or vinyl radical^{36-38,20,39} (Figure 1.1B). Certain lomaiviticins, such as (–)-lomaiviticin A, are nearly a hundred-fold more toxic to cancer cells than are kinamycins,³⁹ despite similar reactive intermediates being accessible from both kinamycins and lomaiviticins. (–)-Lomaiviticin A is especially potent, exhibiting cytotoxic activity at nanomolar–picomolar concentrations.

To reveal the basis for the superior cytotoxicity of (–)-lomaiviticin A, Herzon and coworkers performed a thorough comparison of (–)-lomaiviticin A, (–)-lomaiviticin C, and (–)-kinamycin C.³⁹ They found that the reduction of (–)-lomaiviticin A *in vitro* proceeds more rapidly than does that of (–)-kinamycin C. Moreover, only (–)-lomaiviticin A causes double-stranded breaks in

DNA and activates the double-strand break repair pathway in cells. This combination of attributes likely accounts for the superior potency of (–)-lomaiviticin A. Further, these authors provided evidence that DNA cleavage is instigated by a vinylic carbon radical (Figure 1.1B) and is independent of iron and reactive oxygen species. A solution structure of (–)-lomaiviticin A in complex with DNA revealed that both subunits of lomaiviticin A intercalate into DNA at AT-rich sequences and cause base pairs to be twisted out of the duplex (Figure 1.1C).²¹ The α carbon of the diazo group lies in close proximity to the DNA strand, facilitating hydrogen abstraction by an incipient radical.

One challenge in the investigation and application of lomaiviticins is their limited availability. Smaller analogues that are easier to synthesize provide a partial solution.⁴⁰ One such analogue, a monomeric lomaiviticin aglycon, is capable of inducing DNA damage, albeit at higher concentrations than does (–)-lomaiviticin A. Both (–)-lomaiviticin A and this monomeric lomaiviticin aglycon activate homologous recombination and the non-homologous end-joining repair of DNA in cells.⁴¹ Dysfunctional DNA-repair pathways underlie many human cancers,⁴² rendering lomaiviticins as a potential treatment strategy. In support of this strategy, cell lines with defective DNA-repair pathways (*e.g.*, BRCA2- and PTEN-deficient cells), are more sensitive to (–)-lomaiviticin A and monomeric lomaiviticin aglycon than are isogenic cell lines with intact damage repair pathways.

1.3 Amino Acids

Some natural amino acids contain diazo groups.^{43,44} Notable examples include azaserine and 6-diazo-5-oxo-norleucine (DON), which are nearly isosteric to glutamine (Figure 1.1D).⁴⁵ Both amino acids were isolated initially from *Streptomyces* cultures and exhibit antibiotic and tumor

inhibitory properties.^{46,43} These diazo compounds effectively inhibit amidotransferases involved in the biosynthesis of pyrimidines and purines.⁴⁷⁻⁴⁹ DON entered early-stage clinical trials based on its beneficial activity against various carcinomas, lymphomas, and Hodgkin's disease.⁵⁰ The ability of DON to inhibit amidotransferases revealed the mechanism by which γ -glutamyl transferase acts in tandem with aminopeptidase M to transfer the glutamyl group of glutathione to amino acids and peptides.⁵¹⁻⁵³ DON was also used to determine the catalytic nucleophile and characterize the substrate specificity of glutaminase–asparaginases from various organisms.^{54,55}

Likewise, diazo-containing analogs of asparagine have found utility in medicine as well as enzymology. 5-Diazo-4-oxo-norvaline (DONV; Figure 1.1D) inhibits the growth of asparagine-dependent tumors by interfering with the synthesis and utilization of asparagine.^{44,56} DONV is also a specific inhibitor of L-asparaginase, which is used routinely in the treatment of leukemia.⁵⁷ Clinical assays that aim to determine the blood concentration of asparagine in patients treated with L-asparaginase suffer from degradation of asparagine in the serum sample due to L-asparaginase. The addition of DONV to the assay mixture improves the reliability of asparagine detection.⁵⁷

1.4 Preparation

The synthesis of diazo compounds has become facile. Common methods include (i) diazo transfer,^{58,59} (ii) diazotization,^{60,61} (iii) hydrazone decomposition^{62,63} or hydrazone oxidation,^{64,65} (iv) rearrangement of *N*-alkyl *N*-nitroso compounds,^{9,66} (v) 1,3-disubstituted acyl (or aryl) triazine fragmentation,^{67,68} and (vi) elaboration of other diazo compounds (Figure 1.2).⁶⁹⁻⁷³ Most of these routes have been reviewed extensively for their merits in the context of synthetic chemistry.^{74,75} Nevertheless, the preparation of diazo compounds for applications in chemical

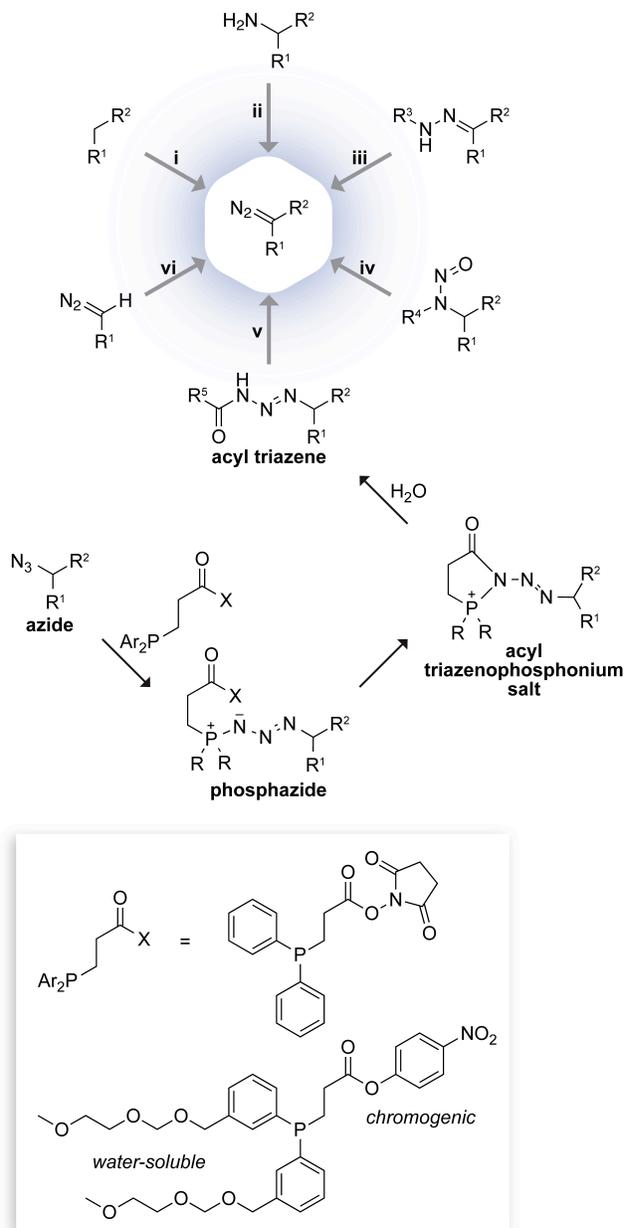


Figure 1.2 Preparation of diazo compounds by (i) diazo transfer,^{58,59} (ii) diazotization,^{60,61} (iii) hydrazone decomposition^{62,63} or hydrazone oxidation,^{64,65} (iv) rearrangement of *N*-alkyl *N*-nitroso compounds,^{9,66} (v) 1,3-disubstituted acyl or aryl triazine fragmentation,^{67,68} and (vi) elaboration of other diazo compounds. Diazo compounds can be accessed from azides via acyl triazenes in a process mediated by a phosphinoester.^{78,79}

biology entails additional challenges because of restrictions on the compatibility of ancillary functional groups and on solubility.

Diazo transfer is a simple and effective way to introduce the diazo group when the pK_a of a proton on the acceptor carbon is low enough to be extracted with a mild base, as is necessary in the stabilized diazo compounds useful in chemical biology. For example, 1,8-diazabicycloundec-7-ene (DBU) can generate α -diazocarbonyl groups after a diazo transfer reaction using sulfonyl azide reagents (e.g., *p*-acetamidobenzenesulfonyl azide and imidazolesulfonyl azide).^{59,76,77} The electronic delocalization that enables diazo transfer also stabilizes the ensuing diazo compound.

Recently, our group reported on a general method to prepare a stabilized diazo group directly from a parent azide.^{78,79} Fragmentation of acyl triazines uses a phosphinoester to convert an azido group into its corresponding diazo group. The reactivity underlying this loss of NH, or

“deimidogenation”, was derived from insight on the mechanism of the Staudinger ligation.⁸⁰⁻⁸⁴ In the Staudinger ligation as well as the Staudinger reaction,^{85,86} the incipient phosphazide quickly extrudes molecular nitrogen to generate an iminophosphorane. A highly reactive acylating group subverts nitrogen extrusion by trapping the phosphazide (Figure 1.2). The ensuing triazenophosphonium intermediate hydrolyzes quickly in water to form an acyl triazene, which is a known precursor to a diazo group.^{67,68}

Azide deimidogenation benefits from the extraordinary chemoselectivity of phosphine for an azide. This approach has a high tolerance for other functional groups, including ketones, esters, aldehydes, thiols, α -chloroesters, epoxides, and disulfide bonds. Chemoselectivity was demonstrated by converting an azido group into a diazo group in aqueous solution containing an enzyme, which was not modified covalently and retained full catalytic activity.⁷⁹ Notably, appropriate azides for deimidogenation (that is, azides with an electron-withdrawing group on the α carbon) are readily accessible via S_N2 reactions with inorganic azide.⁸⁷

Finally, diazo compounds that contain sensitive functional groups can be prepared by the late-stage installation of a prefabricated diazo group. This strategy typically relies on acyl transfer. In 1962, Westheimer and coworkers introduced the concept of photoaffinity labeling by acylating chymotrypsin with *p*-nitrophenyl diazoacetate and then forming an intramolecular crosslink upon photolysis.⁸⁸ Most recent late-stage installations have employed an *N*-hydroxysuccinimide (NHS) ester containing a pendant α -diazocarbonyl group. Badet and coworkers developed a clever synthetic route to the simplest reagent of this class, *N*-hydroxysuccinimidyl diazoacetate.⁸⁹ Such NHS esters have been used to install diazo groups on small molecules^{90,91} as well as biomolecules of varying complexity, including biotin,⁹² mannosamine,⁹³ heparan-sulfate fragments,⁹⁴ lysozyme,⁹³ and bovine serum albumin (BSA).⁹⁵

1.5 Cycloadditions

The archetypal reaction for the diazo group is the 1,3-dipolar cycloaddition. Soon after the synthesis of ethyl diazoacetate by Curtius,⁶⁰ Buchner observed its reaction with an α,β -unsaturated carboxylic ester to form a pyrazole.⁹⁶ Over the last century the reactivity of diazo groups in cycloadditions has engaged theoretical, synthetic, and biological chemists, and these explorations have been reviewed for their use and merits in synthetic chemistry.^{97,98} Here, we focus on recent work that is relevant to biological systems.

Copper-catalyzed azide–alkyne cycloadditions (CuAAC)^{99,100} and strain-promoted azide–alkyne cycloadditions (SPAAC)^{101–103} are two of the most enabling advances in the field of chemical biology.^{104,83,105} The diazo group shares the ability of the azido group to undergo cycloadditions with alkynes, forming a pyrazole rather than a triazole.^{106,107,95} The reactivity of diazo groups is remarkably predictable and tunable¹⁰⁸—the diazo compounds can react with a strained alkyne at much higher or much lower rates than analogous azides (Figure 1.3A).^{106,107,109} Because a diazo group can be generated directly from an azido group^{78,79} and reacts with strained alkynes in common use, the diazo group fits easily into extant methodology.

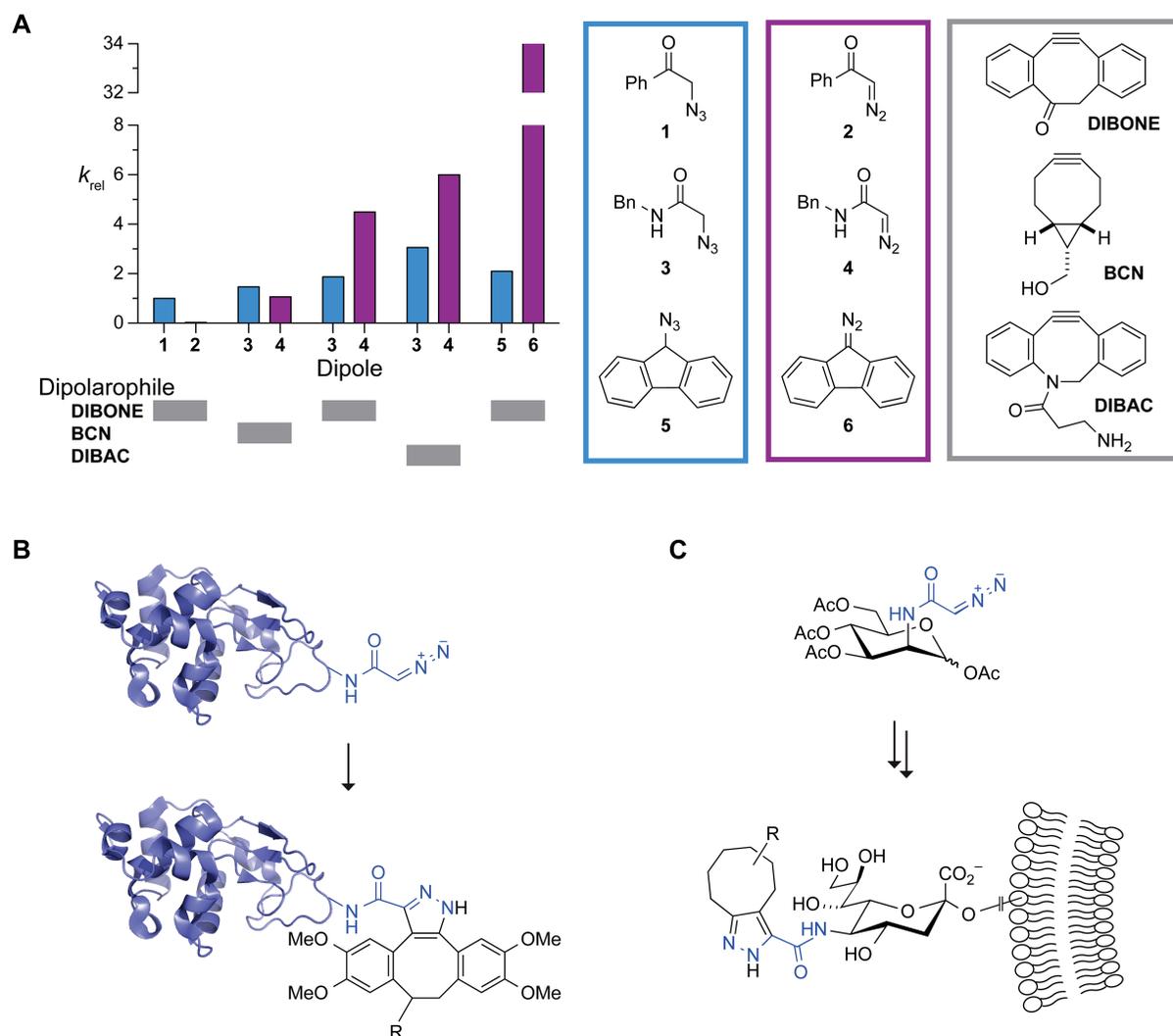


Figure 1.3 Diazo compounds in dipolar cycloadditions with strained alkynes. (A) Relative rate constants of diazo compounds and analogous azides with various cyclooctynes.^{109,92} (B) Labeling of a diazo-modified lysozyme with a cyclooctyne.⁹³ (C) Labeling of a metabolized diazo sugar displayed on the surface of human cells with a cyclooctyne.⁹²

In addition to reacting with strained alkynes, diazo groups undergo uncatalyzed cycloadditions with unstrained dipolarophiles, including terminal alkenes and alkynes. Moreover, diazo compounds can react chemoselectively with certain alkenes and alkynes in the presence of an azide. In essence, a diazo group is more electron-rich, and thus a better nucleophile in normal-

electron-demand cycloadditions with electron-deficient dipolarophiles.¹¹⁰⁻¹¹³ Detailed insight is attainable from computational analyses. Distortion energies account for a majority (80%) of the activation energy for 1,3-dipolar cycloadditions. Due to their increased nucleophilicity and higher HOMO energy, diazo compounds have lower distortion energies than do their azide analogues.^{110,113} The reactions can occur at ambient temperature in aqueous cosolvent with reaction rates similar to or greater than those of SPAACs with azides. Notably, a diazo group can react chemoselectively with the naturally occurring amino acid dehydroalanine (Dha), which contains an electronically activated alkene.¹¹⁰ Selective biotinylation of activated alkenes could enable enrichment and isolation of compounds from a complex lysate, facilitating discovery of new natural products.

1.6 Probes

The diazo group is found in the natural products of microorganisms (*vide supra*). In contrast, its absence in higher organisms enables its utility there as a chemical reporter. The reactivity of the diazo group with many common SPAAC dipolarophiles spawned the use of a diazo group as a chemical reporter for cell-surface glycosylation.

Leeper and coworkers prepared an *N*-diazoacetyl galactosamine and incubated this synthetic sugar with LL2 cells.⁹³ Treatment with a biotin-bearing cyclooctyne and subsequent addition of an avidin fluorophore produced some increase in fluorescence of cells incubated with the diazo-bearing glycan compared to untreated cells. In the same study, an α -diazo NHS ester was reacted with a lysine residue on lysozyme to append the diazo group. Following modification, the appendage was used to attach a fluorophore to the protein via a cycloaddition between the diazo group and a cyclooctyne (Figure 1.3B).

Our group demonstrated the suitability of a diazoacetamide derivative of *N*-acetylmannosamine as a chemical reporter of glycosylation on the surface of CHO K1, Jurkat, HEK293T, and HeLa cells (Figure 1.3C).⁹² The degree of labeling was determined by SPAAC between the diazo group and a biotin-bearing cyclooctyne, followed by treatment with an avidin fluorophore. Metabolic incorporation of the diazo-bearing sugar was evidenced through live-cell microscopy and flow cytometry, and labeling was abolished by treatment with a sialidase. Diazo and alkynyl sugars could be labeled independently on the cell surface. Notably, such dual labeling was not possible on cells displaying azido and alkynyl sugars due to the reactivity of the azide in both CuAAC and SPAAC reactions.

Diazo compounds have long been incorporated into biomolecules as photoaffinity probes.^{114,115} Upon irradiation with ultraviolet light, the diazo group fragments into molecular nitrogen and a carbene, which can undergo either an insertion reaction or a Wolff rearrangement^{116,117} followed by nucleophilic attack on the ensuing ketene, both of which crosslink the diazo compound to proximal functional groups. This strategy has been used to map the architecture of chymotrypsin (*vide supra*),⁸⁸ reveal antibody combining sites,¹¹⁸ examine the structure of lipid membranes,¹¹⁹ and identify isoprenoid-binding sites on proteins.¹²⁰

1.7 Protein Alkylation

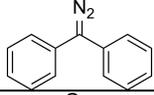
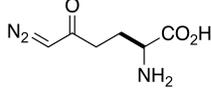
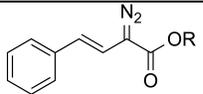
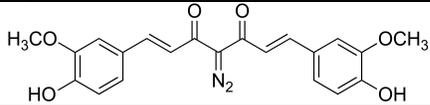
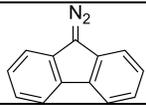
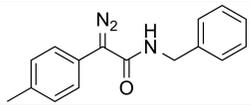
The ability of diazo reagents to alkylate oxygen, nitrogen, sulfur, and even carbon exemplifies their diverse reactivity.^{2,121-124} When applied to protein modification, these reactions are typically catalyzed by acid or transition metals. Despite the apparent promiscuity of this mode of reactivity, even highly reactive compounds such as diazomethane have historically found utility in elucidating structural and functional aspects of proteins.¹²⁵ Stabilized diazo reagents enable

O-alkylation of carboxyl groups and were valuable tools in classical protein chemistry and enzymology.^{126,127} Later, the discovery of diazo-containing amino-acid analogues led medicinal chemists and structural biologist to employ these compounds as covalent inhibitors of metabolic enzymes.⁴⁵ Modern applications of diazo chemistry in chemical biology aim to capitalize on the versatility of diazo compounds to access linkages that cannot be achieved by other methods. Maintaining chemoselectivity in the presence of water and other biological nucleophiles has been a primary challenge in developing diazo compounds as useful tools for protein chemistry.^{128,129}

The earliest uses of diazo reagents for protein labeling sought to characterize structural features of proteins. In 1914, Geake and Nierenstein used diazomethane to alkylate caseinogen so as to characterize the structure of amino-acid side chains (Table 1.1).¹²⁵ By comparing the methylated and unmethylated protein, they identified and quantified side chains that contain amino or hydroxyl groups. Later studies addressed large-scale structural characterization of proteins, such as quantification of the number of peptide chains in a protein and identification of carboxyl groups in the binding region of the anti-hapten antibody.^{130,131}

The last 100 years have seen many attempts to limit the promiscuity of the diazo reagent by using stabilized α -diazo amides (Table 1.1). Doscher and Wilcox used α -diazoacetamide to label chymotrypsin in work that laid the foundation for modern protein-labeling endeavors.¹²⁶ They demonstrated that, although the rate of esterification was much greater than the rate of diazo-compound hydrolysis, the large excess of water molecules limits the efficiency of esterification. The authors suggested that employing a mixed aqueous–organic solvent could increase esterification efficiency by both limiting diazo hydrolysis and increasing the pK_a of enzymic carboxyl groups. This idea was later explored, and did indeed increase the efficiency of protein

esterification.¹²⁸ Although α -diazooacetamide was more selective than diazomethane, it still *S*-alkylated sulfhydryl groups.

Table 1.1 Diazo compounds that esterify proteins.			
Diazo Compound	Protein	Year	Reference
	caseinogen	1914	¹²⁵
	insulin	1958	130
	β -lactoglobulin		
	lysozyme		
	polyclonal antibody	1960	¹³¹
	chymotrypsinogen	1961	126
	ribonuclease A	1965	127
	pepsin	1966–1968	132–134
	acid proteases	1972–1973	135–138
	prorenin	1980	139
	<i>O</i> -sulfotransferase	2015	94
	pepsin	1966	¹⁴⁰
	phosphoribosyl pyrophosphate amidotransferase	1963	47
	glutaminase A	1973	48
	glutamyl transpeptidase	1978	52,54,53
	asparaginase	1977	⁵⁶
	myoglobin	2004	⁵⁴
	subtilisin		
	Yes kinase	2015	141
	β -lactoglobulin	2007	¹⁴²
	ribonuclease A	2015	¹²⁸
	red fluorescent protein		
	ribonuclease A	2015	¹²⁹

In 1917, Staudinger and Gaule became the first to use a diazo compound, diphenyldiazomethane, to form an ester.¹⁴³ The mechanism of this reaction was established in elegant work by Roberts and coworkers in 1951 (Figure 1.4A).¹⁴⁴ The heightened reactivity of carboxyl groups *versus* carboxylates inspired subsequent esterification experiments. Riehm and Scheraga used α -diazooacetamide to esterify the carboxyl groups in ribonuclease A.¹²⁷ They found that one aspartic acid residue was esterified preferentially, and proposed that this

residue resides in a solvent-accessible area of local negative charge, which would raise its pK_a value and lead to its selective esterification. Shortly thereafter, Delpierre and Fruton used an α -diazoketone to label a single residue in the active site of pepsin, causing near-complete inhibition of the enzyme.¹⁴⁰ These workers proposed that this residue was in a privileged environment that enabled its selective labeling, as was posited for the aspartic acid in ribonuclease A,¹²⁷ though neither of these speculations has been explored further. Instead, the inhibition of pepsin using α -diazoketones gave rise to a breadth of studies characterizing the active site of pepsin and comparing pepsin to its zymogen form (*i.e.*, pepsinogen), in which the active-site residue is inaccessible to solvent and thus does not react with the diazo reagent.^{145,132,146,133,134,147-150} The combination of covalent labeling using a diazo reagent with Edman degradation (which was invented concurrently) provided a robust method for determining the identity of a catalytically important residue and its surrounding sequence.¹⁵¹ Using these techniques, novel acid proteases were classified based on their propensity to be inactivated by a diazo compound.^{135-138,152,139,153} Nonetheless, with the advent of site-directed mutagenesis, the use of diazo compounds to characterize proteins became rare.

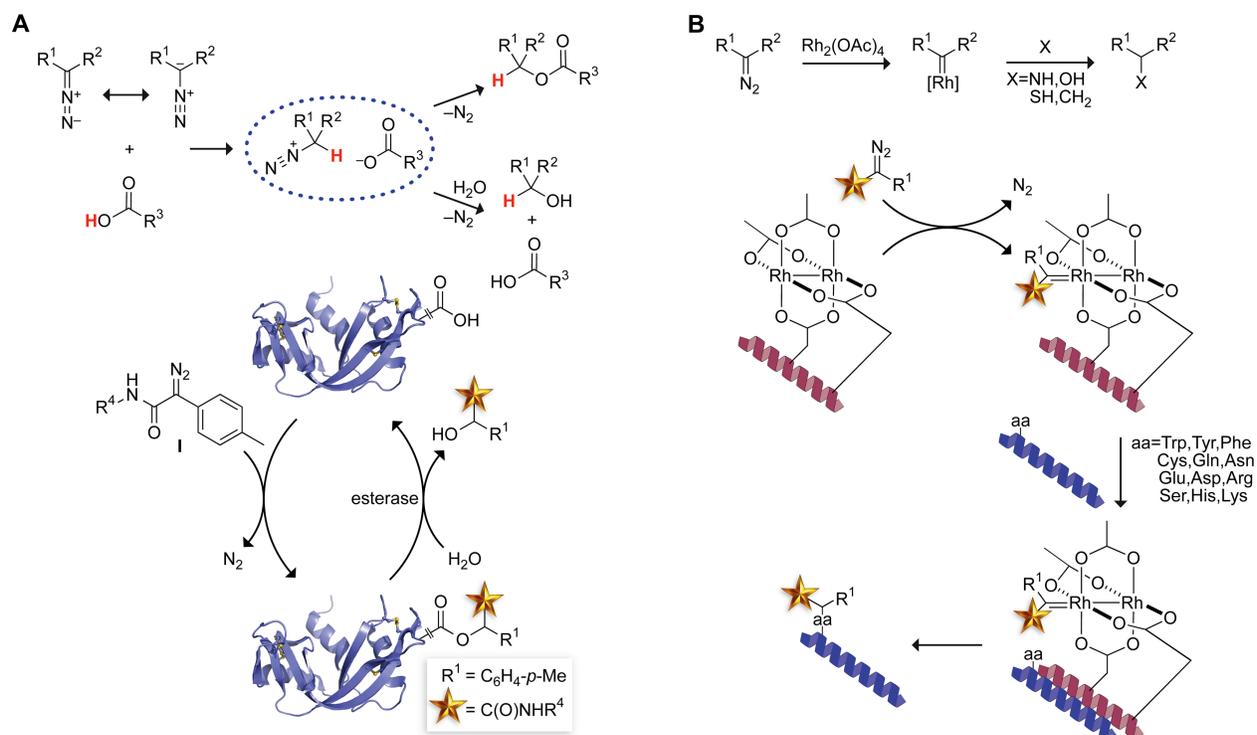


Figure 1.4 Diazo compounds for covalent modification of proteins. (A) Putative mechanism for the esterification of carboxylic acids with a diazo compound,¹⁴⁴ and its application to the bioreversible labeling of a protein.^{128,129} Diazo compound **I** is optimized for protein esterification.¹²⁹ (B) Putative mechanism of a diazo carbenoid insertion reaction, and its application to the site-specific modification of a proximal amino-acid residue.¹⁵⁴

1.8 Bioreversible Protein Modification

The abundance and promiscuity of cellular esterases has been utilized in prodrug strategies in which chemotherapeutic agents are masked as esters and converted to their active forms upon cellular uptake.¹⁵⁵⁻¹⁵⁷ Our group envisioned a similar strategy for proteins in which carboxyl moieties are esterified by a diazo compound to install a molecular tag, such as a pharmacokinetic-enhancing, cell-type-targeting, or cell-penetrating moiety. Upon cellular uptake, the ester-linked tags are removed by endogenous esterases to recreate the native protein (Figure 1.4A). This strategy would be especially valuable for the delivery of proteins whose activities decrease significantly upon irreversible modification.¹⁵⁸

In an initial study, structurally and electronically diverse diazo compounds were screened for their reactivity and selectivity in an aqueous environment.¹²⁸ Of these compounds, only 9-diazofluorene esterified a panel of carboxylic acids efficiently in the presence of water. This diazo compound was used to label two model proteins, ribonuclease A and red fluorescent protein. The nascent esters were hydrolyzed upon treatment with a HeLa-cell extract, regenerating wild-type protein.

Later, a more systematic study investigated the rate and selectivity of a series of structurally similar but electronically diverse α -diazo amides.¹²⁹ A Hammett analysis of these compounds, which were derived from phenylglycine, revealed that electron-donating or electron-withdrawing groups on the aryl ring had a dramatic effect on the rate of esterification. Still, the compounds were similar in their selectivity for ester formation over hydrolysis of the diazo reagent. The comparable selectivity among the compounds in this study supports the proposed mechanism in which the diazonium and carboxylate species, formed as intermediates, are held together in a solvent cage as an intimate ion pair (Figure 1.4A),¹⁴⁴ and the ratio of ester to alcohol product is determined by the diffusion out of this solvent cage rather than the reactivity of the diazo compound.^{144,159} An α -diazo(*p*-methylphenyl)-glycinamide (**I**) demonstrated the fastest rate while maintaining selectivity, and esterifies proteins more efficiently than any known diazo reagent. The amide of compound **I** allows for facile incorporation of an amine of interest.

1.9 Peptide and Protein Modification with Carbenoids

An early example of asymmetric catalysis employed a chiral transition-metal catalyst to generate a carbenoid from a diazo compound.¹⁶⁰ Carbenoids generated similarly can access a broad scope of insertion reactions and are hence powerful reagents for modifying peptides and proteins. In a

seminal study, Francis and coworkers used vinylic α -diazo esters to modify tryptophan residues in horse heart myoglobin.⁷⁶ Then, Ball and coworkers employed metallopeptides to combine proximity-driven and transition metal-driven catalysis.^{161,162} In this system, the rhodium catalyst is displayed on a peptide, which is designed to bind a second peptide or protein of interest by forming a coiled-coil (Figure 1.4B).¹⁵⁴ The catalyst on the metallopeptide is oriented such that the incipient carbenoid is generated proximal to the target residue, focusing its high reactivity and enabling modification of many types of amino acids.¹⁶³ For example, although tryptophan can be modified by the addition of a diazo compound and rhodium acetate catalyst alone, employing a metallopeptide to orient the catalyst enables modification of the phenyl group of phenylalanine, imidazolyl group of histidine, and guanidinium group of arginine.

In a proof-of-concept study, Popp and Ball alkylated the aromatic amino-acid side chains by tethering the dirhodium center to a lysine-rich K3 peptide, which binds to and reacts with a glutamate-rich E3 peptide at a specific tryptophan residue.¹⁵⁴ In a follow-up investigation, the scope of the E3/K3 system was extended to the alkylation of a broad range of functional groups, including a carboxamide.¹⁶³ This system has since been used to modify maltose-binding protein fused to the E3 peptide,¹⁶⁴ as well as for the site-selective modification of the native Fyn protein using a peptide ligand bearing the rhodium catalyst.^{165,141}

1.10 Nucleic Acid Alkylation

Natural nucleobases can be modified *in situ* with diazo compounds. Gillingham and coworkers used rhodium(II) to catalyze the conversion of a diazo ester into a carbenoid that inserted into exocyclic N–H bonds (Figure 1.5A).¹⁶⁶ Because this reactivity does not extend to double-helical

regions, the strategy can target hairpins and single-stranded regions. This selectivity is useful, for example, in studies on the mechanism of RNA interference, which entails 3' overhangs.

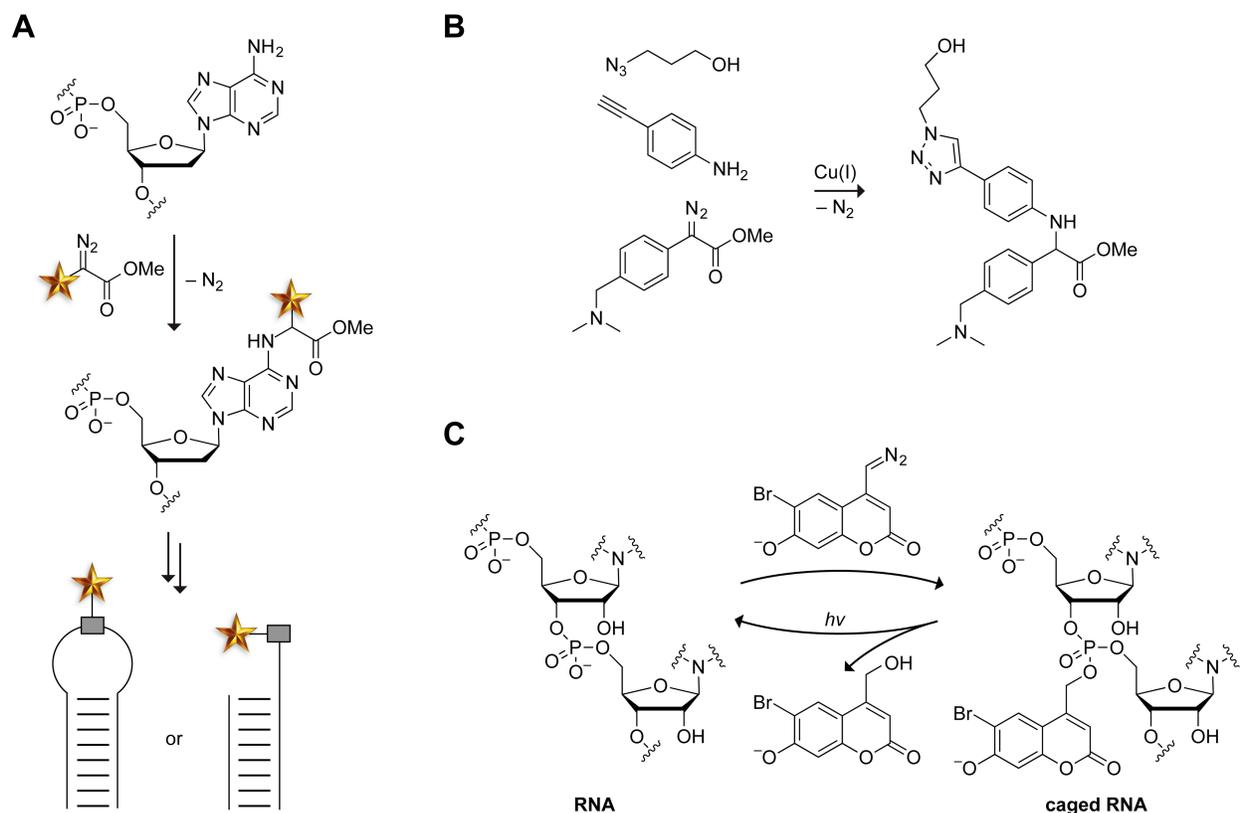


Figure 1.5 Covalent modification of nucleic acids using diazo compounds. (A) Representative alkylation of DNA by a diazo compound. Alkylation occurs on solvent-accessible nucleobases.¹⁶⁶ (B) One-pot N-H insertion and azide-alkyne cycloaddition with a copper(I) catalyst.¹⁶⁷ (C) Photoreversible O-alkylation of a phosphoryl group in RNA by a diazo coumarin.¹⁶⁸

Rhodium(II) has been used most widely as a catalyst for the generation of carbenoids in chemical biology.¹⁶⁹ Gillingham and coworkers showed, however, that copper(I)-carbenoid chemistry for N-H insertion is likewise effective.¹⁶⁷ Their work demonstrated novel synergy of the diazo group with “copper-click” chemistry by combining N-H insertion with CuAAC in a one-pot single-catalyst process (Figure 1.5B).

An alternative strategy for nucleic-acid modification involves *O*-alkylation of the phosphoryl group. Okamoto and coworkers employed this method to modify an mRNA using a photolabile derivative of coumarin bearing a diazo moiety (Figure 1.5C).¹⁶⁸ The ensuing “caged” mRNA, which encoded green fluorescent protein, was delivered to zebrafish embryos, where its translation could be modulated spatially and temporally by uncaging using ultraviolet light. Photolabile diazo groups have also been used to control RNA interference, in which a double-stranded precursor to an siRNA is inactivated upon modification with the diazo reagent and then uncaged with ultraviolet light.¹⁷⁰ Diazo compounds have been employed to label and detect nucleic acids on microarrays without disrupting base pairing.¹⁷¹ Recently, Gillingham and coworkers reported on a diazo compound that modifies the phosphoryl groups of nucleic acids selectively in the presence of carboxylic acids.¹⁷² Their methodology could be useful for the labeling and detection of phosphorylated peptides and proteins as well.

1.11 Outlook

Diazo compounds were discovered over 120 years ago. Recent advances in chemical synthesis have enabled the facile preparation of stabilized diazo compounds that are compatible with living systems. Like azido groups, diazo groups are chemoselective. Unlike azido groups, diazo groups have reactivity with natural and nonnatural functional groups that is tunable. The ability to tune their reactivity by delocalization of the electrons on the α carbon renders diazo compounds as attractive reagents in physiological contexts. Moreover, the versatility of diazo-group reactivity is extraordinary. Their ability to react rapidly, selectively, and autonomously with nonnatural functional groups (*e.g.*, strained alkynes) as well as natural carboxyl groups, phosphoryl groups, and even the alkene in dehydroalanine residues anoints diazo groups as special. Accordingly, we

envision an expansion in the use of diazo compounds to probe biological phenomena and to treat human disease, and even foresee an era of “diazophilia”.¹⁷³

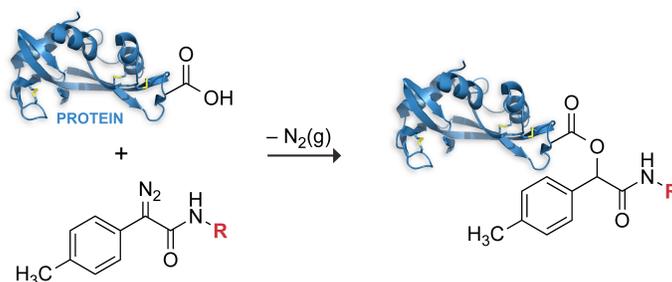
1.12 Acknowledgments

We are grateful to C. L. Jenkins for comments on the manuscript. K.A.M. was supported by Molecular Biosciences Training Grant T32 GM007215 (NIH). Work on diazo compounds in the Raines laboratory is supported by Grant R01 GM044783 (NIH).

Chapter Two

Optimized Diazo Scaffold for Protein Esterification*

*This chapter has been published, in part, under the same title. Reference: Mix, K.A., Raines, R.T. (2015) *Org. Lett.* 17, 2359–2361.



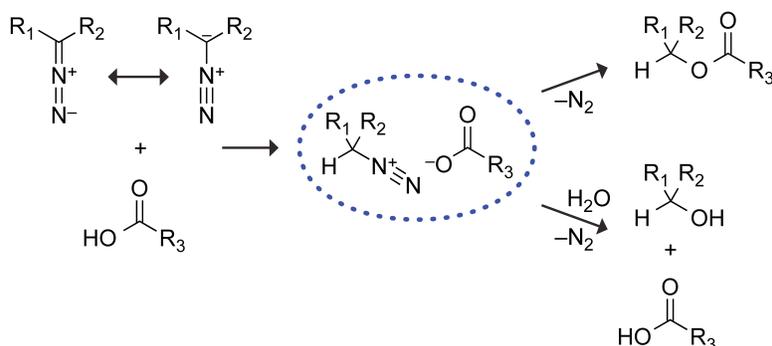
Abstract

The *O*-alkylation of carboxylic acids with diazo compounds provides a means to esterify carboxylic acids in aqueous solution. A Hammett analysis of the reactivity of diazo compounds derived from phenylglycinamide revealed that the *p*-methylphenylglycinamide scaffold has an especially high reaction rate and ester:alcohol product ratio, and esterifies protein carboxyl groups more efficiently than does any known reagent.

2.1 Introduction

Broad reactivity has made diazo compounds one of the most versatile functional groups in synthetic organic chemistry.^{2,121,174,175,124} Recently, this broad utility has been expanded into the field of chemical biology. For example, the diazo group has been shown to undergo 1,3-dipolar cycloadditions with strained alkynes in a tunable manner. The rates can greatly exceed those of the analogous azide,¹⁰⁹ and the reactions are chemoselective in the presence of mammalian cells.⁹² In addition, diazo compounds have been used to label proteins via C–H, N–H, and S–H insertion reactions.^{76,176}

Diazo compounds have another well-known mode of reactivity—esterification of carboxylic acids. We realized that this reactivity could provide unique opportunities in chemical biology. For example, unlike the alkylation of other functional groups, *O*-alkylation of a carboxyl group is bioreversible because mammalian cells contain non-specific esterases.^{177,156,178,179} The esterification of carboxyl groups in proteins and other biomolecules is, however, difficult to effect, as solvent water competes effectively with alcohols for electrophilic acyl groups. In contrast, esterification reactions mediated by diazo groups rely on the carboxyl group serving as a nucleophile (Scheme 2.1).^{144,180}



Scheme 2.1 Mechanism of the reaction between a diazo compound and carboxylic acid in an aqueous solution.

The use of diazo compounds to label proteins was attempted 60 years ago.^{131,181,182,145} These initial results were not compelling. A large molar excess (up to 10^3 -fold) of diazo compound was required to overcome hydrolytic decomposition. Moreover, the reaction was not chemoselective, as amino, sulfhydryl, and phenolic side chains suffered alkylation. Such modifications are potentially deleterious to protein function and not bioreversible.⁸

Previous work in our laboratory suggested that the obstacles in reactivity can be overcome by tuning the reactivity of a diazo group. In particular, we found that the basicity of 9-diazafluorene endows this diazo compound with the ability to label a protein in an aqueous environment.⁹ The fluorenyl scaffold is, however, unduly large and not readily amenable to synthetic modification, and its reaction rate and chemoselectivity are not necessarily maximal.

Accordingly, we sought a scaffold that is optimal for the esterification of carboxyl groups in an aqueous environment. Towards that end, we have examined derivatives of phenylglycinamide (Figure 2.1A). This scaffold delocalizes the electron density on C^α into an amidic carbonyl group as well as a phenyl group that enables a Hammett analysis¹⁸³⁻¹⁸⁶ of the esterification reaction. Moreover, the amide linkage enables facile installation of useful moieties.

2.2 Results and Discussion

Diazo compounds **2.1–2.6** were accessed from derivatives of phenylacetic acid (Figure 2.1B). Briefly, an azide was installed at the benzylic position of the acid either through displacement of a bromide or by diazo transfer to an existing amine. The ensuing α -azido acids were then coupled to benzylamine and converted to the diazo compound by deimidogenation using a phosphinoester.^{78,187}

In initial experiments, we probed the effect of electron distribution on the reactivity of diazo groups by measuring the rate of esterification in acetonitrile. To do so, we reacted diazo compounds **2.1–2.6** with BocGlyOH, and measured the second-order rate constants with ^1H NMR spectroscopy. The effect of electron distribution on the reaction rate was dramatic: rate constants spanned over two orders of magnitude and increased with the electron-donating character of the phenyl substituents (Figure 2.2A). Hammett analysis of these rate constants gave a slope of $\rho = -2.7$ (Figure 2.2B). This value, which is comparable to those for typical $\text{S}_{\text{N}}1$ reactions, indicates that the

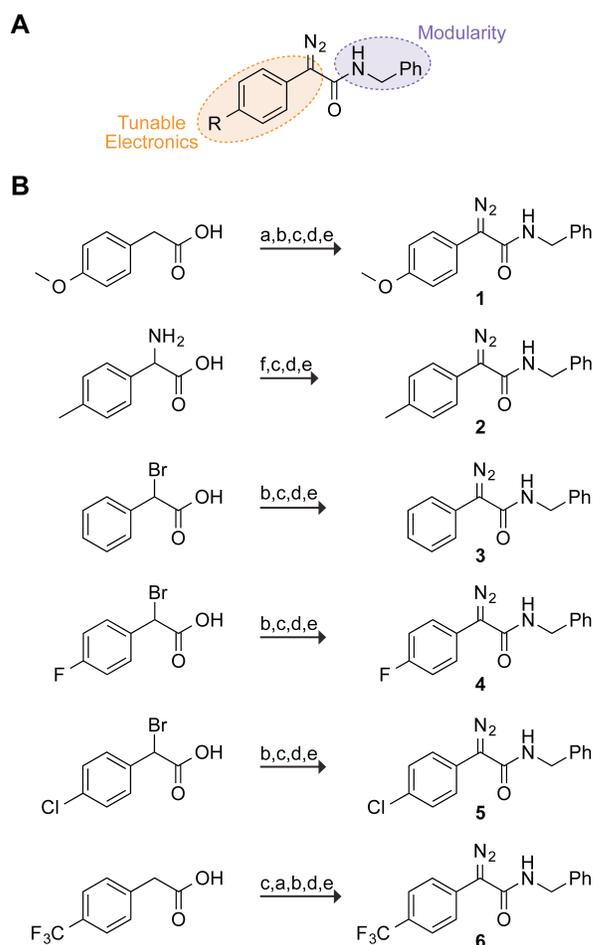


Figure 2.1 (A) Scaffold for testing the reactivity and selectivity of diazo compounds. (B) Synthetic route to diazo compounds **2.1–2.6**. Steps: a) NBS, AIBN; b) NaN_3 , $\text{THF}:\text{H}_2\text{O}$; c) NHS, DCC, THF; d) PhCH_2NH_2 , DCM; e) *N*-succinimidyl 3-(diphenylphosphino)propionate, then NaHCO_3 or DBU;¹⁰ f) imidazole-1-sulfonyl azide hydrochloride, DBU, CuSO_4 , MeOH.¹¹

esterification reaction is highly sensitive to substituents and that substantial positive charge accumulates during its course,¹⁸⁸ as expected from a mechanism involving an intermediate diazonium ion (Scheme 2.1).^{144,180}

Next, we sought to find the one compound that demonstrates the greatest selectivity for esterification over hydrolysis in an aqueous environment. Towards that end, we reacted diazo compounds **2.1–2.6** with equimolar BocGlyOH in a 1:1 mixture of acetonitrile and 2-(*N*-morpholino)ethanesulfonic acid (MES)–HCl buffer at pH 5.5, and we determined the ratio of ester-to-alcohol product with ¹H NMR spectroscopy.

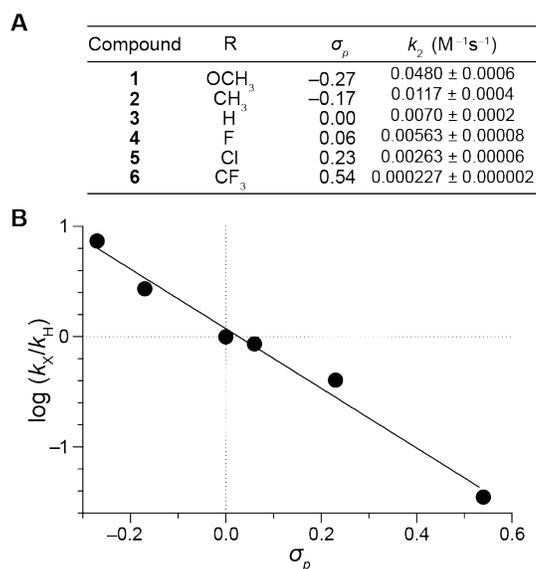


Figure 2.2 (A) Second-order rate constants for the esterification of BocGlyOH by diazo compounds **2.1–2.6** in CD₃CN. (B) Hammett plot of the data in panel A. Values of σ_p are from ref. ¹⁸⁹. $\rho = -2.7$

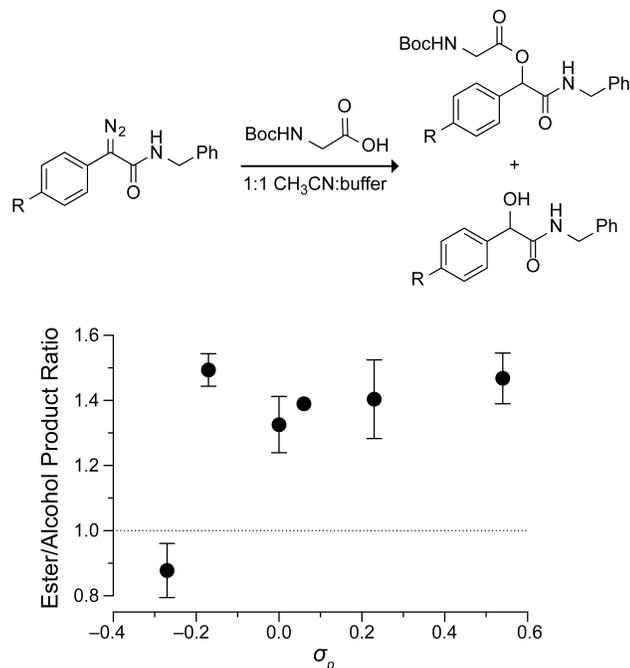


Figure 2.3 Effect of σ_p value on the chemoselectivity of diazo compounds **2.1–2.6** in aqueous solution.

Surprisingly, the ester:alcohol ratio reached a maximum of 1.4:1 and remained unchanged despite increasing electron-withdrawal by the substituents (Figure 2.3). This result is consistent with a sharp cutoff for the formation of a carboxylate-diazonium intimate ion-pair intermediate that is maintained in a solvent cage by a Coulombic interaction (Scheme 2.1).^{144,180,159}

Based on these experiments diazo compound **2.2** was selected for further study, as it demonstrated the fastest rate of those compounds that retain chemoselectivity in an aqueous environment. Because certain diazo compounds undergo O–H and S–H insertion reactions,^{174,76,176} we sought to ensure that diazo compound **2.2** would esterify acids selectively in the presence of the sulfhydryl, hydroxyl, or phenolic moieties found on protein side chains. We were gratified to find that diazo compound **2.2** esterified BocSerOH, *p*-hydroxybenzoic acid,

and 3-mercaptopropionic acid in 1:1 acetonitrile/100 mM MES–HCl buffer at pH 5.5, and that no other coupling products were observable by ^1H NMR spectroscopy. We also attempted to esterify AlaOH to probe for reaction with an amino group. Consistent with previous observations,¹⁹⁰ diazo compound **2.2** did not react with either the amino group or the carboxyl group of AlaOH, which was largely zwitterionic in the reaction mixture (Figure 2.5).

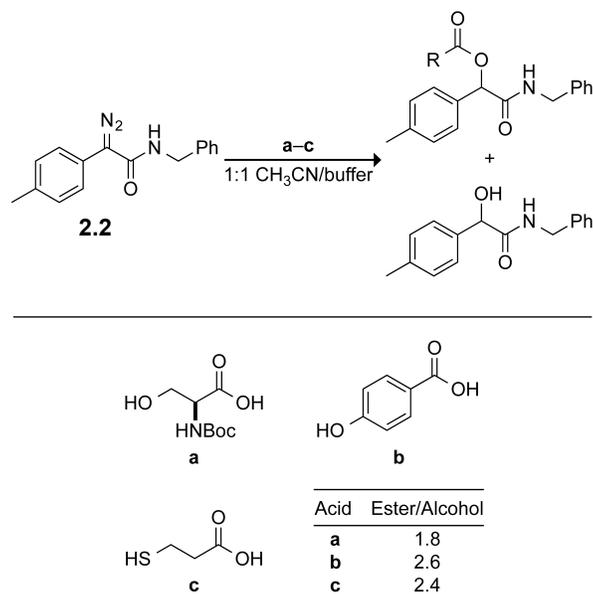


Figure 2.4 Chemoselectivity of esterification reactions in aqueous solution.

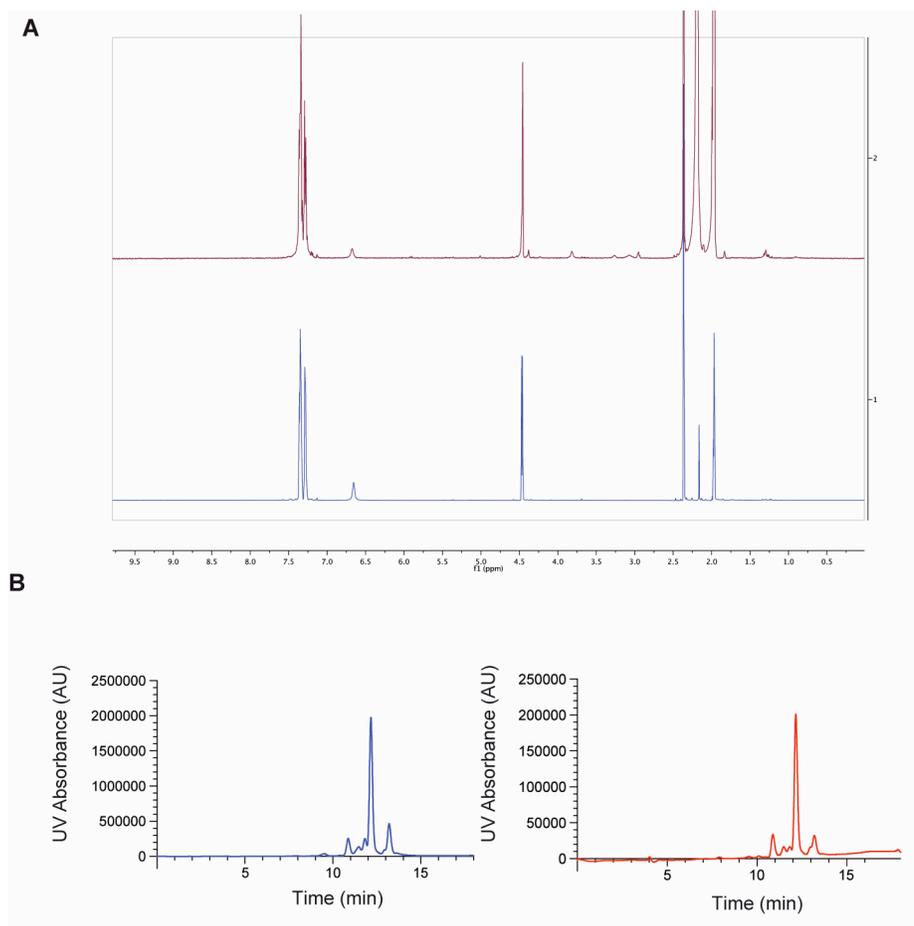


Figure 2.5 (A) ^1H NMR (400 MHz, CD_3CN) overlay of diazo compound **2.2** (bottom, blue) and a crude reaction mixture of diazo compound **2.2** treated with AlaOH in 1:1 acetonitrile/100 mM MES-HCl buffer at pH 5.5 (top, red). (B) LC-MS chromatograms of diazo compound **2.2** (left, blue) and a crude reaction mixture of diazo compound **2.2** treated with AlaOH in 1:1 acetonitrile/100 mM MES-HCl buffer at pH 5.5 (right, red). The trace impurities with retention times of 11 and 13 min are present in both chromatograms and are likely decomposition products of diazo compound **2.2** in the acidic conditions used for chromatography.

Finally, we compared diazo compound **2.2** to 9-diazafluorene for the labeling of a protein. To do so, we treated a well-known model protein, ribonuclease A,¹⁹¹ with 10 equiv of each diazo compound. The reactions were allowed to proceed for 4 h at 37 °C in 1:1 acetonitrile/10 mM MES-HCl buffer at pH 5.5. We then determined the extent of esterification with MALDI-TOF

mass spectrometry. We found that diazo compound **2.2** was approximately twofold more efficient than was 9-diazafluorene in effecting esterification (Figure 2.6).

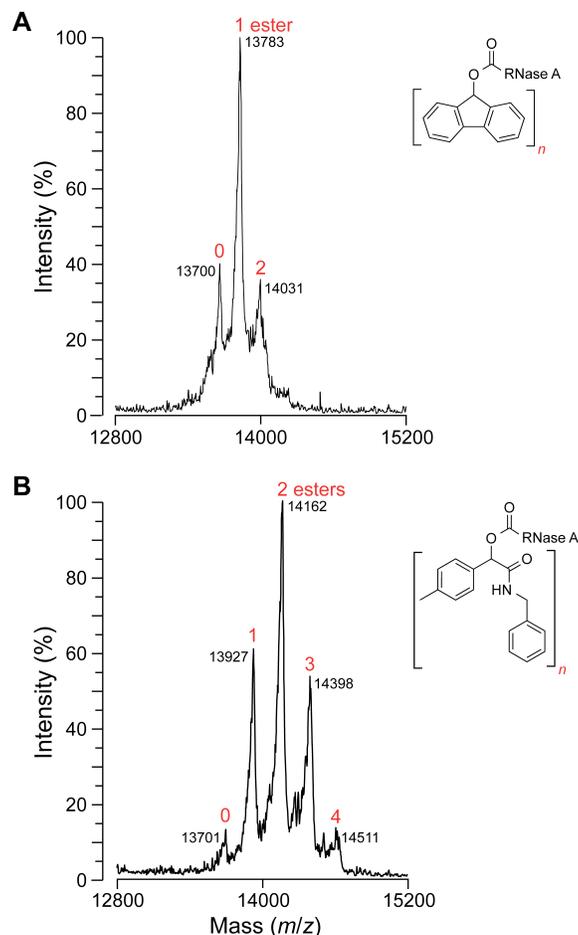


Figure 2.6 MALDI-TOF mass spectrometry data for esterification of RNase A with (A) 9-diazafluorene and (B) diazo compound **2.2**.

We conclude that diazo compound **2.2** can be used to esterify proteins in an aqueous environment more efficiently than any other known reagent. Moreover, its modular design enables facile modification with useful moieties. We are now using this diazo compound to attach cell-type targeting, cell-penetration, and pharmacokinetic enhancing modules to proteins of interest.

2.3 Acknowledgments

This work was supported by grant R01 GM044783 (NIH). K.A.M. was supported by Molecular Biosciences Training Grant T32 GM007215 (NIH). This work made use of the National Magnetic Resonance Facility at Madison, which is supported by grant P41 GM103399 (NIH), and the Biophysics Instrumentation Facility, which was established with grants BIR-9512577 (NSF) and S10 RR013790 (NIH). We thank Dr. N. A. McGrath (University of Wisconsin–La Crosse) for contributive discussions and critical reading of the manuscript, and Dr. B. VanVeller (Iowa State University) for suggesting the phenylglycine scaffold.

2.4. Materials and Methods

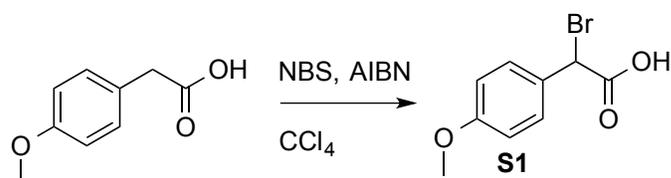
2.4.1 General

Silica gel (40 μm ; 230–400 mesh) was from SiliCycle. Reagents were obtained from commercial sources and used without further purification. Dichloromethane and tetrahydrofuran were dried over a column of alumina. Thin-layer chromatography (TLC) was performed on plates of EMD 250 μm silica 60-F₂₅₄. The phrase “concentrated under reduced pressure” refers to the removal of solvents and other volatile materials using a rotary evaporator at water aspirator pressure (<20 torr) while maintaining a water bath below 40 °C. Residual solvent was removed from samples at high vacuum (<0.1 torr). ¹H and ¹³C NMR spectra for all compounds were acquired with Bruker spectrometers in the National Magnetic Resonance Facility at Madison operating at 400, 500, 600, or 750 MHz. Chemical shift data are reported in units of δ (ppm) relative to an internal standard (residual solvent or TMS). Electrospray ionization (ESI) mass spectrometry for small-molecule characterization was performed with a Micromass LCT at the Mass

Spectrometry Facility in the Department of Chemistry at the University of Wisconsin–Madison. Matrix-assisted laser desorption-ionization–time-of-flight (MALDI–TOF) mass spectrometry for protein characterization was performed with a Voyager DE-Pro instrument at the Biophysics Instrumentation Facility at the University of Wisconsin–Madison.

2.4.2 Chemical Synthesis

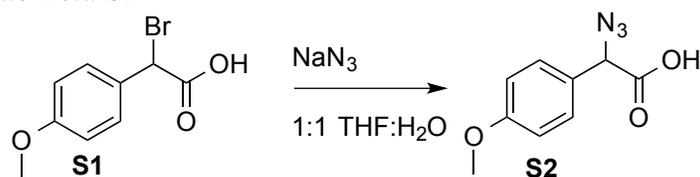
Preparation of α -Bromoacid **S1**



4-Methoxyphenylacetic acid (5.000 g, 30.10 mmol) was dissolved in CCl₄ (50 mL). *N*-Bromosuccinimide (5.625 g, 31.6 mmol) and AIBN (0.985 g, 6.0 mmol) were added. The resulting solution was heated to 80 °C and allowed to reflux overnight. The succinimide by-product was removed by filtration, and the solution was concentrated under reduced pressure. The residue was purified by chromatography on silica gel, eluting with 1:1 EtOAc/hexanes to afford **S1** (5.705 g, 78%) as a white solid.

Data for S1: ¹H NMR (500 MHz, CDCl₃, δ): 7.50 (d, 2H, J = 8.8 Hz), 6.90 (d, 2H, J = 8.8 Hz), 5.36 (s, 1H), 3.82 (s, 1H.) ¹³C NMR (125 MHz, CDCl₃, δ): 173.4, 160.5, 130.2, 126.8, 114.3, 55.4, 45.9. HRMS (ESI⁻) m/z calcd for C₉H₉BrO₃ [M–H]⁻ 242.9662; found, 242.9660.

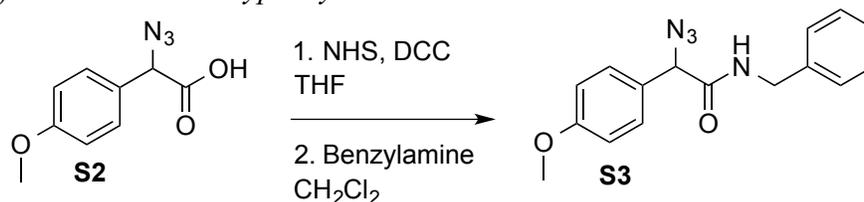
Preparation of α -Azido Acid **S2**



α -Bromo-4-methoxyphenylacetic acid **S1** (0.802 g, 3.3 mmol) was dissolved in 1:1 THF/H₂O (4 mL). Sodium azide (0.429 g, 6.6 mmol) was added, and the resulting solution was stirred overnight. The solution was then concentrated under reduced pressure, and the residue was dissolved in EtOAc (50 mL). The resulting solution was washed with 0.1 M HCl (2 \times 50 mL). The organic layer was dried over anhydrous Na₂SO₄(s) and concentrated under reduced pressure to afford **S2** (0.412 g, 62%) as a white solid.

Data for S2: ¹H NMR (500 MHz, CDCl₃, δ): 7.35 (d, 2H, J = 8.7 Hz), 6.95 (d, 2H, J = 8.7 Hz), 5.00 (s, 1H), 3.83 (s, 3H). ¹³C NMR (125 MHz, CDCl₃, δ): 173.5, 160.5, 129.1, 125.2, 114.6, 64.6, 55.4. HRMS (ESI⁻) m/z calcd for C₉H₉N₃O₃ [M-H]⁻ 206.0571; found, 206.0577.

Preparation of α -azido 4-Methoxyphenylacetic Amide S3

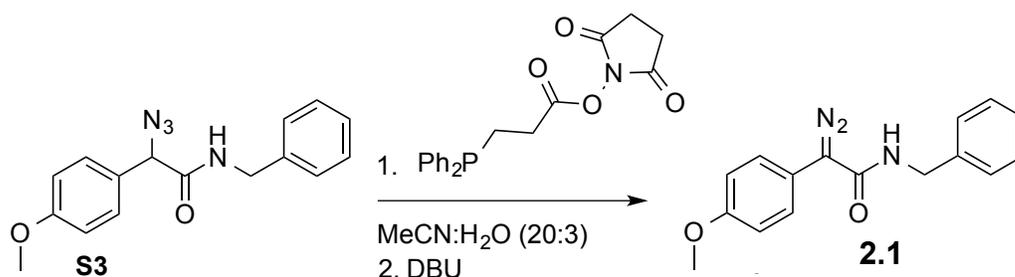


α -Azido-4-methoxyphenylacetic acid **S2** (0.412 g, 2.0 mmol) was dissolved in THF (5 mL), and the resulting solution was cooled in an ice bath. *N*-Hydroxysuccinimide (0.230 g, 2.0 mmol) was added, followed by the portion-wise addition of DCC (0.453 g, 2.2 mmol). The resulting solution was warmed to ambient temperature and stirred overnight. The slurry was removed by filtration, and the solution was concentrated under reduced pressure. The residue was dissolved in EtOAc (10 mL) and washed with saturated aqueous NaHCO₃ (2 \times 10 mL). The organic layer was dried over anhydrous Na₂SO₄(s) and concentrated under reduced pressure. The residue was purified by chromatography on silica gel, eluting with 3:7 EtOAc/hexanes, and used immediately. The NHS ester (0.4 g, 1.2 mmol) was dissolved in CH₂Cl₂ (10 mL). Benzylamine (0.10 mL, 1.3 mmol) was added dropwise, and the resulting solution was stirred overnight. The

solution was then concentrated under reduced pressure. The residue was dissolved in EtOAc (10 mL) and washed with 0.1 M HCl (2×10 mL) and saturated aqueous NaHCO₃ (2×10 mL). The organic layer was dried over anhydrous Na₂SO₄(s) and concentrated under reduced pressure to afford **S3** (0.255 g, 43%) as a white solid.

Data for S3: ¹H NMR (500 MHz, CD₃CN, δ): 7.34–7.30 (m, 4H), 7.27–7.23 (m, 3H), 6.97 (d, 2H, $J = 8.8$ Hz), 4.99 (s, 1H), 4.37 (m, 2H), 3.80 (s, 3H). ¹³C NMR (125 MHz, CD₃CN, δ): 169.4, 161.0, 139.8, 130.2, 129.4, 128.4, 128.2, 128.0, 115.1, 66.6, 55.9, 43.6. HRMS (ESI⁺) m/z calcd for C₁₆H₁₆N₄O₂ [M+H]⁺ 297.1347; found, 297.1346.

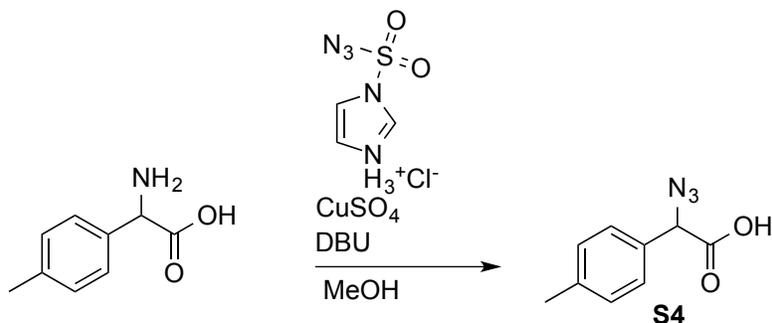
Preparation of α -Diazo Amide **2.1**



α -Azidoamide **S3** (0.356 g, 1.2 mmol) was dissolved in 20:3 MeCN/H₂O (12 mL), and the resulting solution was cooled in an ice bath. *N*-Succinimidyl 3-(diphenylphosphino)propionate (0.440 g, 1.24 mmol) was added slowly. The solution was warmed to ambient temperature and stirred until all azide was consumed (~12 h as monitored by TLC). DBU (0.21 mL, 1.4 mmol) was added, and the solution was stirred for 1 h. The solution was then diluted with brine (10 mL) and extracted with CH₂Cl₂ (2×20 mL). The organic layer was dried over anhydrous Na₂SO₄(s) and concentrated under reduced pressure. The residue was purified by chromatography on silica gel, eluting with 1:1 EtOAc/hexanes to afford **2.1** (0.095 g, 28%) as an orange solid.

Data for 2.1: ^1H NMR (500 MHz, CD_3CN , δ): 7.37 (d, 2H, $J = 8.9$ Hz), 7.34–7.29 (m, 4H), 7.26–7.23 (m, 1H), 4.43 (d, 2H, $J = 6.2$ Hz), 3.80 (s, 3H). ^{13}C NMR (125 MHz, CDCl_3 , δ): 165.4, 159.7, 138.4, 130.3, 128.7, 127.7, 127.5, 117.5, 115.3, 63.1, 55.4, 44.1. HRMS (ESI^+) m/z calcd for $\text{C}_{16}\text{H}_{15}\text{N}_3\text{O}_2$ $[\text{M}+\text{H}]^+$ 282.1238; found, 282.1232.

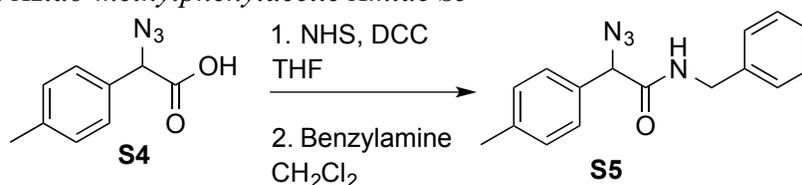
Preparation of α -Azido Acid S4



Imidazole-1-sulfonyl-azide hydrochloride was prepared as reported previously.¹⁹² Spectral data and yields match those reported previously. α -Amino-4-methylphenylacetic acid (2.000 g, 12.1 mmol) was dissolved in MeOH (24 mL). DBU (3.61 mL, 24.2 mmol), CuSO_4 (0.300 g, 1.2 mmol), and azide (3.030 g, 14.5 mmol) were added sequentially. The resulting solution was heated to 40 °C and stirred overnight. The solution was then concentrated under reduced pressure. The residue was dissolved in EtOAc (30 mL) and washed twice with 1 M aqueous HCl (2×30 mL). The organic layers were combined and dried over anhydrous Na_2SO_4 (s). The solution was concentrated under reduced pressure. The residue was dissolved in benzene and recrystallized from benzene and hexanes to afford **S4** (0.390 g, 17%) as a white solid.

Data for S4: ^1H NMR (600 MHz, CDCl_3 , δ): 7.30 (d, 2H, $J = 8.1$ Hz), 7.24 (d, 2H, $J = 7.8$ Hz), 5.01 (s, 1H), 2.37 (s, 3H). ^{13}C NMR (150 MHz, CDCl_3 , δ): 173.4, 139.7, 130.2, 129.9, 127.6, 64.9, 21.2. HRMS (ESI^-) m/z calcd for $\text{C}_9\text{H}_9\text{N}_3\text{O}_2$ $[\text{M}-\text{H}]^-$ 190.0622; found, 190.0625.

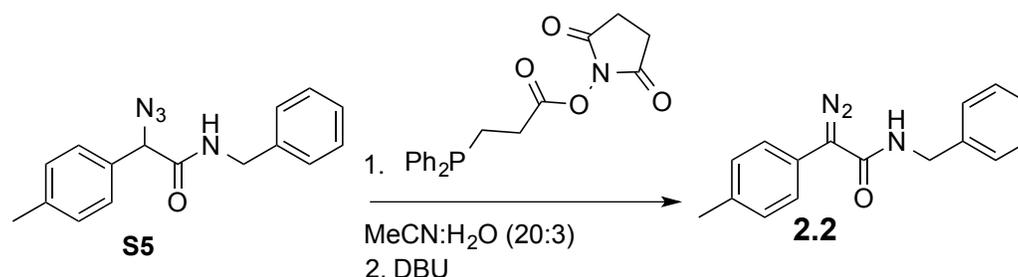
Preparation of α -Azido-methylphenylacetic Amide S5



α -Azido 4-methylphenylacetic acid **S4** (2.204 g, 11.6 mmol) was dissolved in THF (30 mL) and cooled in an ice bath. *N*-Hydroxysuccinimide (1.334 g, 11.6 mmol) was added, followed by portion-wise addition of DCC (2.637 g, 12.8 mmol). The resulting solution was warmed to ambient temperature and stirred overnight. The slurry was removed by filtration, and the solution was concentrated under reduced pressure. The residue was dissolved in EtOAc (30 mL). The resulting solution was washed with saturated aqueous NaHCO_3 (2×30 mL). The organic layer was dried over anhydrous $\text{Na}_2\text{SO}_4(\text{s})$, concentrated under reduced pressure, and used immediately. The NHS ester (2.5 g, 8.7 mmol) was dissolved in CH_2Cl_2 (30 mL). Benzylamine (0.98 mL, 9.6 mmol) was added dropwise, and the resulting solution was stirred overnight. The solution was then concentrated under reduced pressure. The residue was dissolved in EtOAc (30 mL) and washed with 0.1 M HCl (2×30 mL) and saturated aqueous NaHCO_3 (2×30 mL). The organic layer was dried over anhydrous $\text{Na}_2\text{SO}_4(\text{s})$ and concentrated under reduced pressure to afford **S5** (1.988 g, 61%) as a white solid.

Data for S5: ^1H NMR (500 MHz, CD_3CN , δ): 7.33–7.28 (m, 4H), 7.26–7.22 (m, 5H), 5.00 (s, 1H), 4.36 (dd, 2H, $J = 1.8, 6.2$ Hz), 2.35 (s, 3H). ^{13}C NMR (125 MHz, CD_3CN , δ): 169.2, 140.0, 139.8, 133.5, 130.4, 129.4, 128.8, 128.2, 128.0, 66.9, 43.6, 21.1. HRMS (ESI^+) m/z calcd for $\text{C}_{16}\text{H}_{16}\text{N}_4\text{O}$ $[\text{M}+\text{H}]^+$ 281.1397; found, 281.1395.

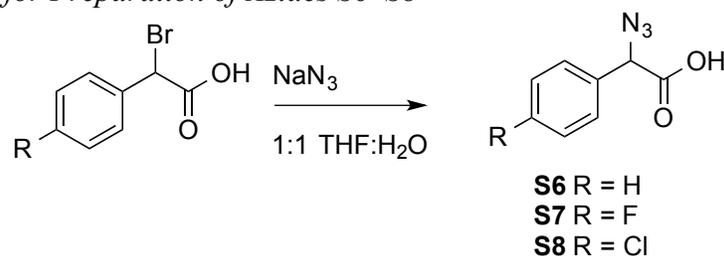
Preparation of α -Diazo-methylphenylacetic Amide **2.2**



α -Azido 4-methylphenylacetic amide **S5** (1.995 g, 7.1 mmol) was dissolved in 20:3 MeCN/H₂O (50 mL), and the resulting solution was cooled in an ice bath. *N*-Succinimidyl 3-(diphenylphosphino)propionate (2.769 g, 7.8 mmol) was added slowly. The solution was warmed to ambient temperature and stirred until all azide was consumed (~24 h as monitored by TLC). DBU (1.27 mL, 8.5 mmol) was added, and the solution stirred for 45 min. The solution was then diluted with brine (10 mL) and extracted with CH₂Cl₂ (2 × 30 mL). The organic layer was dried over anhydrous Na₂SO₄(s) and concentrated under reduced pressure. The residue was purified by chromatography on silica gel, eluting with 4:6 EtOAc/hexanes to afford **2.2** (1.038 g, 55%) as an orange solid.

Data for 2.2: ¹H NMR (600 MHz, CD₃CN, δ): 7.33–7.23 (m, 9H), 6.63 (s, 1H), 4.44 (d, 2H, J = 6.2 Hz), 2.34 (s, 3H). ¹³C NMR (150 MHz, CD₃CN, δ): 165.5, 140.7, 138.1, 130.9, 129.3, 128.2, 128.1, 127.9, 124.1, 63.74, 44.0, 21.1. HRMS (ESI⁺) m/z calcd for C₁₆H₁₅N₃O [M+H]⁺ 266.1288; found, 266.1292.

General Procedure for Preparation of Azides **S6–S8**



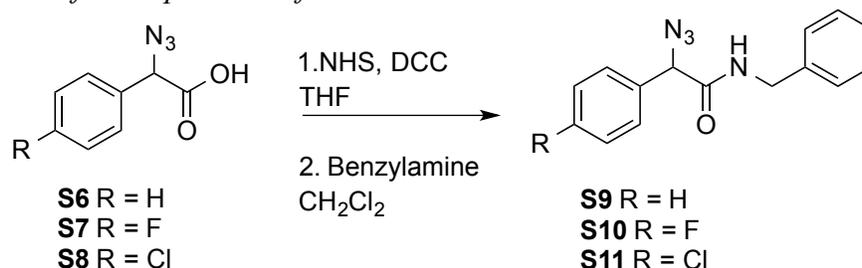
Each α -bromophenylacetic acid (23.3 mmol) was dissolved in a solution of 1:1 THF/H₂O (24 mL). Sodium azide (1.512 g, 46.5 mmol) was added, and the resulting solution was stirred overnight. The solution was then concentrated under reduced pressure. The residue was dissolved in EtOAc (50 mL), and washed with 0.1 M HCl (2 \times 50 mL). The organic layer was dried over anhydrous Na₂SO₄(s) and concentrated under reduced pressure to afford a white solid (**S6**: 4.076 g, 99%; **S7**: 4.016 g, 89%; **S8**: 3.761 g, 77%).

Data for Azide S6: ¹H NMR (400 MHz, CDCl₃, δ): 7.43 (m, 5H), 5.05 (s, 1H). ¹³C NMR (400 MHz, CDCl₃, δ): 174.0, 133.1, 129.6, 129.2, 127.7, 65.1. HRMS (ESI⁺) m/z calcd for C₈H₇N₃O₂ [M+H]⁺ 177.0533; found, 177.0538.

Data for Azide S7: ¹H NMR (400 MHz, CDCl₃, δ): 7.41 (dd, 2H, J = 5.1, 8.5 Hz), 7.12 (t, 2H, J = 8.4 Hz), 5.05 (s, 1H). ¹³C NMR (100 MHz, CDCl₃, δ): 175.0, 163.5 (d, J = 249.6 Hz), 129.8 (d, J = 8.5 Hz) 129.1 (d, J = 2.6 Hz), 116.5 (d, J = 22.1 Hz), 64.5. HRMS (ESI⁻) m/z calcd for C₈H₆FN₃O₂ [M-H]⁻ 194.0371; found, 194.0378.

Data for Azide S8: ¹H NMR (400 MHz, CDCl₃, δ): 7.41 (d, 2H, J = 8.4 Hz), 7.37 (d, 2H, J = 8.3 Hz), 5.06 (s, 1H). ¹³C NMR (125 MHz, CDCl₃, δ): 174.7, 135.8, 131.5, 129.5, 129.0, 64.3. HRMS (ESI⁻) m/z calcd for C₈H₆ClN₃O₂ [M-H]⁻ 210.0075; found, 210.0078.

General Procedure for Preparation of Amides S9–S11



Each α -azidoacetic acid (**S6–S8**) (15.4 mmol) was dissolved in THF (30 mL), and the resulting solution was cooled in an ice bath. *N*-Hydroxysuccinimide (NHS) (1.772 g, 15.4 mmol) was

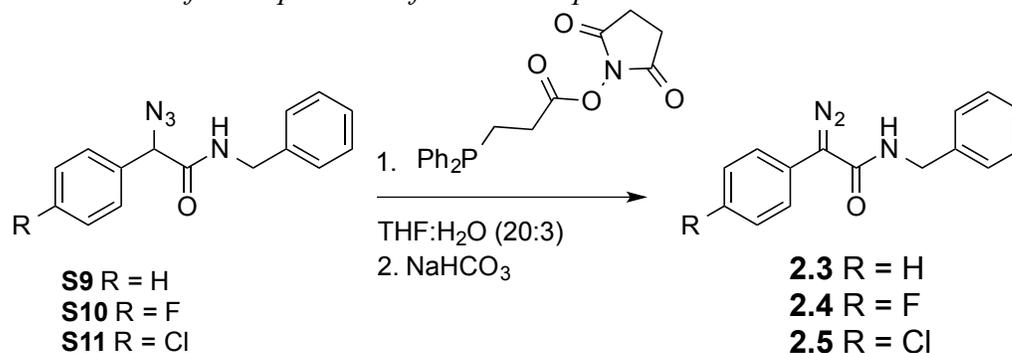
added, followed by portion-wise addition of DCC (3.177 g, 15.4 mmol). The solution was warmed to ambient temperature and stirred overnight. The slurry was removed by filtration, and the solution was concentrated under reduced pressure. The residue was dissolved in EtOAc (50 mL) and washed with saturated aqueous NaHCO₃ (2 × 50 mL). The organic layer was dried over anhydrous Na₂SO₄(s) and concentrated under reduced pressure. The residue was purified by chromatography on silica gel, eluting with 1:1 EtOAc/hexanes. The resulting solution was then concentrated under reduced pressure and used immediately. The NHS ester (10.5 mmol) was dissolved in CH₂Cl₂ (105 mL). Benzylamine (1.16 mL, 10.6 mmol) was added drop-wise, and the resulting solution was stirred overnight. The solution was concentrated under reduced pressure. The residue was dissolved in EtOAc (50 mL) and washed with 0.1 M HCl (2 × 50 mL) and saturated aqueous NaHCO₃ (2 × 50 mL). The organic layer was dried over anhydrous Na₂SO₄(s) and concentrated under reduced pressure. The residue was purified by chromatography on silica gel, eluting with 30% EtOAc/hexanes to afford a white solid (**S9**: 2.384 g, 58% for 2 steps; **S10**: 2.062 g, 47% for 2 steps; **S11**: 2.179 g, 47% for 2 steps).

Data for Amide S9: ¹H NMR (500 MHz, CD₃CN, δ): 7.43–7.42 (m, 5H), 7.31–7.29 (m, 2H), 7.26–7.22 (m, 3H), 5.06 (s, 1H), 4.37 (d, 2H, *J* = 6.2). ¹³C NMR (125 MHz, CDCl₃, δ): 167.8, 137.5, 134.9, 129.2, 129.1, 128.8, 127.8, 127.73, 127.67, 67.4, 43.7. HRMS (ESI⁺) *m/z* calcd for C₁₅H₁₄N₄O [M+H]⁺ 267.1241; found, 267.1241.

Data for Amide S10: ¹H NMR (600 MHz, CD₃CN, δ): 7.45–7.42 (dd, 2H, *J* = 5.4, 8.7 Hz), 7.23–7.30 (m, 2H), 7.26–7.22 (m, 3H), 7.18–7.15 (m, 2H), 5.08 (s, 1H), 4.37 (dd, 2H, *J* = 3.0, 6.2 Hz). ¹³C NMR (100 MHz, CDCl₃, δ): 167.6, 163.1 (d, *J* = 249.2 Hz), 137.5, 130.9 (d, *J* = 2.0 Hz), 129.5 (d, *J* = 8.5 Hz), 128.8, 127.8, 116.2 (d, *J* = 21.8 Hz), 105.0, 66.6, 43.7. 43.7, HRMS (ESI⁺) *m/z* calcd for C₁₅H₁₃FN₄O [M+H]⁺ 285.1147; found, 285.1150.

Data for Amide S11: ^1H NMR (500 MHz, CD_3CN , δ): 7.44–7.39 (m, 4H), 7.33–7.27 (m, 2H), 7.25–7.22 (m, 3H), 5.08 (s, 1H), 4.36 (m, 2H). ^{13}C NMR (125 MHz, CD_3CN , δ): 168.8, 139.7, 135.5, 135.2, 130.4, 129.9, 129.4, 128.2, 128.0, 66.3, 43.6. HRMS (ESI⁺) m/z calcd for $\text{C}_{15}\text{H}_{13}\text{ClN}_4\text{O}$ [$\text{M}+\text{H}$]⁺ 301.0851; found, 301.0850.

General Procedure for Preparation of Diazo Compounds 2.3–2.5



Each α -azidobenzamide (**S9**–**S11**) (7.3 mmol) was dissolved in a solution of 20:3 THF:H₂O (75 mL) and cooled in an ice bath. *N*-Succinimidyl 3-(diphenylphosphino)propionate (2.734 g, 7.7 mmol) was added slowly. The resulting solution was warmed to ambient temperature and stirred until all azide was consumed (6–12 h as monitored by TLC). Saturated aqueous NaHCO_3 (73 mL) was added, and the solution was stirred overnight. The solution was then diluted with brine (50 mL) and extracted with CH_2Cl_2 (2×70 mL). The organic layer was dried over anhydrous $\text{Na}_2\text{SO}_4(\text{s})$ and concentrated under reduced pressure. The residue was purified by chromatography on silica gel, eluting with 1:1 EtOAc/hexanes to afford an orange solid (**2.3**: 1.012 g, 55%; **2.4**: 0.887 g, 45%; **2.5**: 0.877 g, 42%).

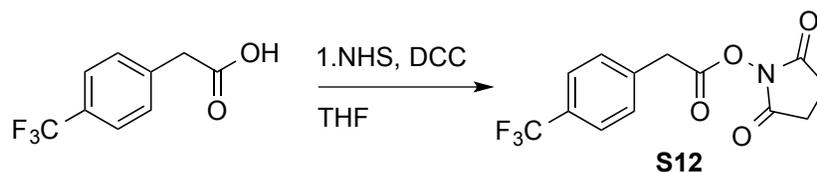
Data for Diazo 2.3: ^1H NMR (600 MHz, CD_3CN , δ): 7.46–7.41 (m, 4H), 7.34–7.28 (m, 4H), 7.28–7.23 (m, 2H), 6.73 (s, 1H), 4.44 (d, 2H, $J = 6.1$ Hz). ^{13}C NMR (125 MHz, CD_3CN , δ):

165.1, 140.6, 130.2, 129.3, 128.2, 127.8, 127.7, 127.6, 127.4, 64.0, 43.9. HRMS (ESI⁺) m/z calcd for C₁₅H₁₃N₃O [M+H]⁺ 252.1132; found, 252.1125.

Data for Diazo 2.4: ¹H NMR (500 MHz, CD₃CN, δ): 7.49–7.46 (dd, 2H, J = 5.4, 8.6 Hz), 7.34–7.29 (m, 4H), 7.26–7.23 (m, 1H), 7.20–7.16 (t, 2H, J = 8.8), 6.70 (s, 1H), 4.43 (d, 2H, J = 6.2). ¹³C NMR (125 MHz, CD₃CN, δ): 165.2, 162.5 (d, J = 244.9 Hz), 140.6, 130.2 (d, J = 8.3 Hz), 129.2, 128.1, 127.8, 123.4 (d, J = 3.1 Hz), 116.9 (d, J = 22.1 Hz), 62.99, 43.8. HRMS (ESI⁺) m/z calcd for C₁₅H₁₂FN₃O [M+H]⁺ 270.1038; found, 270.1032.

Data for Diazo 2.5: ¹H NMR (500 MHz, CD₃CN, δ): 7.45 (d, 2H, J = 8.8 Hz), 7.42 (d, 2H, 8.9 Hz), 7.35–7.30 (m, 4H), 7.28–7.26 (m, 1H), 6.79 (s, 1H), 4.44 (d, 2H, J = 6.1 Hz). ¹³C NMR (125 MHz, CDCl₃, δ): 164.1, 138.1, 133.5, 129.9, 128.8, 128.5, 127.8, 127.7, 124.7, 63.5, 44.2. HRMS (ESI⁺) m/z calcd for C₁₅H₁₂ClN₃O [M+H]⁺ 286.0742; found, 286.0748.

Preparation of Ester S12

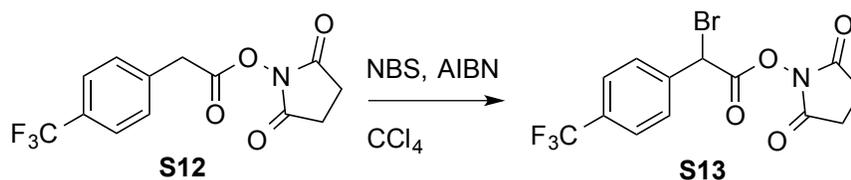


4-(Trifluoromethyl)phenylacetic acid (5.000 g, 24.5 mmol) was dissolved in THF (50 mL), and the resulting solution was cooled in an ice bath. *N*-Hydroxysuccinimide (2.818 g, 24.5 mmol) was added, followed by DCC (5.047 g, 24.5 mmol). The solution was warmed to ambient temperature and stirred overnight. The slurry was removed by filtration, and the solution was concentrated under reduced pressure. The residue was dissolved in EtOAc (50 mL) and washed with saturated aqueous NaHCO₃ (2 × 50 mL). The organic layer was dried over anhydrous Na₂SO₄(s) and concentrated under reduced pressure. The residue was purified by

chromatography on silica gel, eluting with 1:1 EtOAc/hexanes to afford **S12** (7.301 g, 99%) as a white solid.

Data for Ester S12: ^1H NMR (400 MHz, CDCl_3 , δ): 7.63 (d, 2H, $J = 7.99$ Hz), 7.48 (d, 2H, $J = 7.92$ Hz), 4.00 (s, 2H), 2.84 (s, 4H). ^{13}C NMR (125 MHz, CDCl_3 , δ): 168.9, 166.1, 135.27, 130.2 (q, $J = 32.6$ Hz), 129.7, 125.8 (q, $J = 3.7$ Hz), 123.9 (q, $J = 272.1$ Hz), 37.4, 25.6. HRMS (EI^+) m/z calcd for $\text{C}_{13}\text{H}_{10}\text{F}_3\text{NO}_4$ $[\text{M}+\text{H}]^+$ 301.0557; found, 301.0565.

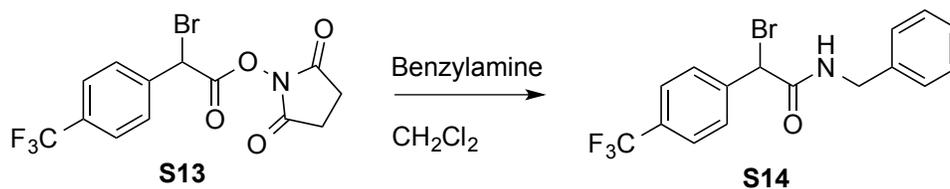
Preparation of α -Bromoester S13



Ester **S12** (3.763 g, 12.5 mmol) was dissolved in CCl_4 (25 mL). *N*-Bromosuccinimide (3.329 g, 18.7 mmol) and AIBN (0.394 g, 2.4 mmol) were added. The resulting solution was heated to 80 °C and allowed to reflux overnight. The succinimide by-product was removed by filtration, and solution was concentrated under reduced pressure. The residue was purified by chromatography on silica gel, eluting with 1:1 EtOAc/hexanes to afford **S13** (2.037 g, 43%) as a white solid.

Data for S13: ^1H NMR (500 MHz, CDCl_3 , δ): 7.72 (d, 2H, $J = 8.3$ Hz), 7.69 (d, 2H, $J = 8.6$ Hz), 5.68 (s, 1H), 2.86 (s, 4H). ^{13}C NMR (125 MHz, CDCl_3 , δ): 168.2, 163.8, 137.7, 131.9 (q, $J = 32.8$ Hz), 129.2, 126.1 (q, $J = 3.7$ Hz), 123.6 (q, $J = 272.5$ Hz), 40.7, 25.6. HRMS (EI^+) m/z calcd for $\text{C}_{13}\text{H}_9\text{BrF}_3\text{NO}_4$ $[\text{M}+\text{H}]^+$ 378.9662; found, 378.9667.

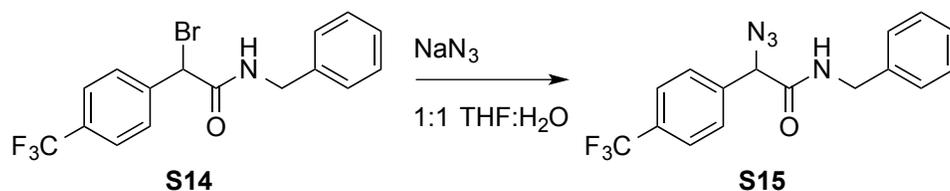
Preparation of α -Bromoamide S14



α -Bromoester **S13** (3.297 g, 8.7 mmol) was dissolved in CH_2Cl_2 (80 mL). Benzylamine (0.91 mL, 8.7 mmol) was added drop-wise, and the resulting solution was stirred overnight. The solution was concentrated under reduced pressure, and the residue was dissolved in EtOAc (50 mL). The solution was washed with 0.1 M HCl (2×50 mL) and saturated aqueous NaHCO_3 (2×50 mL). The organic layers were dried over anhydrous $\text{Na}_2\text{SO}_4(\text{s})$ and concentrated under reduced pressure. The residue was purified with chromatography on silica gel, eluting with 1:1 EtOAc/hexanes to afford **S14** (1.456 g, 45%) as a white solid.

Data for S14: ^1H NMR (500 MHz, CD_3CN , δ): 7.76 (d, 2H, $J = 8.3$ Hz), 7.72 (d, 2H, $J = 2\text{H}$), 7.51 (s, 1H), 7.35 (t, 3H, $J = 7.4$ Hz), 7.29 (t, 3H, $J = 7.7$ Hz), 5.59 (s, 1H), 4.40 (m, 2H). ^{13}C NMR (125 MHz, CDCl_3 , δ): 166.2, 141.2, 137.1, 131.1 (q, $J = 32.8$ Hz), 128.9, 128.8, 128.0, 127.8, 125.9 (q, $J = 3.7$ Hz), 123.7 (q, $J = 272.3$ Hz), 49.8, 44.6. HRMS (ESI $^+$) m/z calcd for $\text{C}_{16}\text{H}_{13}\text{BrF}_3\text{NO}$ $[\text{M}+\text{H}]^+$ 372.0206; found, 372.0210.

Preparation of α -Azidoamide S15



(30 mL) and extracted with CH_2Cl_2 (2×30 mL). The organic layer was dried over anhydrous $\text{Na}_2\text{SO}_4(\text{s})$ and concentrated under reduced pressure. The residue was purified by chromatography on silica gel, eluting with 1:1 EtOAc/hexanes to afford **2.6** (0.382 g, 40%) as an orange solid.

Data for 2.6: ^1H NMR (400 MHz, CDCl_3 , δ): 7.65 (d, 2H, $J = 8.0$ Hz), 7.50 (d, 2H, $J = 8.1$ Hz), 7.38–7.31 (m, 5H), 5.70 (s, 1H), 4.59 (d, 2H, $J = 4.6$ Hz). ^{13}C NMR (125 MHz, CD_3CN , δ): 164.2, 140.4, 132.9, 128.3, 127.9, 127.6 (q, $J = 32.4$ Hz), 126.5 (q, $J = 3.9$ Hz), 126.3, 125.3 (q, $J = 270.8$ Hz), 64.0, 43.9. HRMS (ESI^+) m/z calcd for $\text{C}_{16}\text{H}_{12}\text{F}_3\text{N}_3\text{O}$ $[\text{M}+\text{H}]^+$ 320.1006; found, 320.0993.

2.4.3 Measurement of Reaction Rate Constants

Each diazo compound and BocGlyOH were dissolved separately in CD_3CN at a concentration of 50 mM. The solutions were combined in an NMR tube at an equimolar ratio, mixed, and then inserted immediately into an NMR spectrometer. A 16-scan ^1H NMR spectrum was acquired every 10 min. Percent conversion was monitored by disappearance of starting material and appearance of product as determined by integration of multiple ^1H NMR spectral peaks. No other products were apparent by ^1H NMR spectroscopy. The value of the second-order rate constant was determined by linear regression analysis of a plot of $1/[\text{diazo}]$ versus time. All reactions were performed in triplicate.

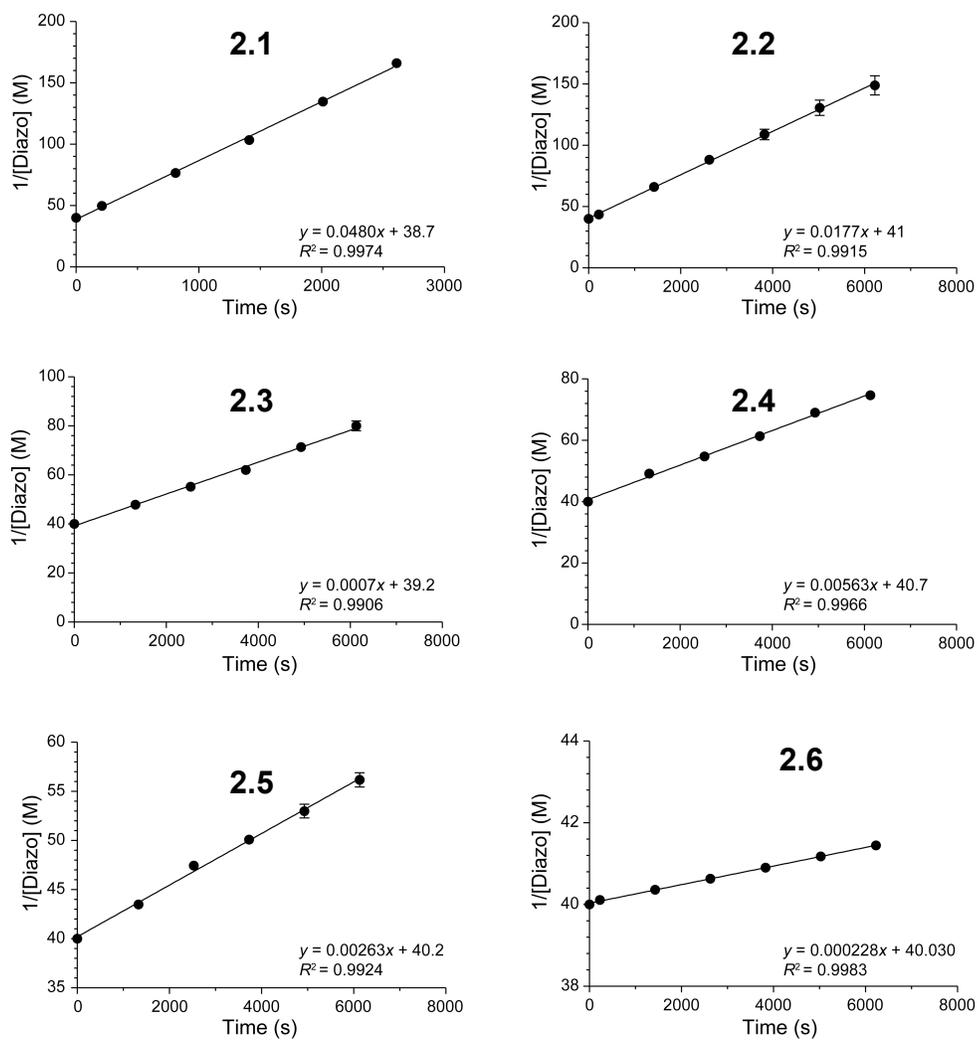
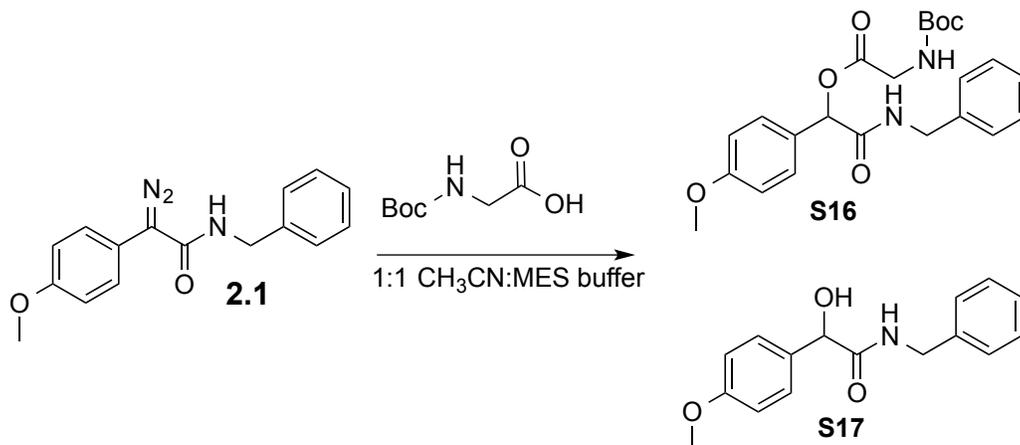


Figure 2.7 ^1H NMR kinetic data for reaction between compounds 2.1–2.6 and BocGlyOH.

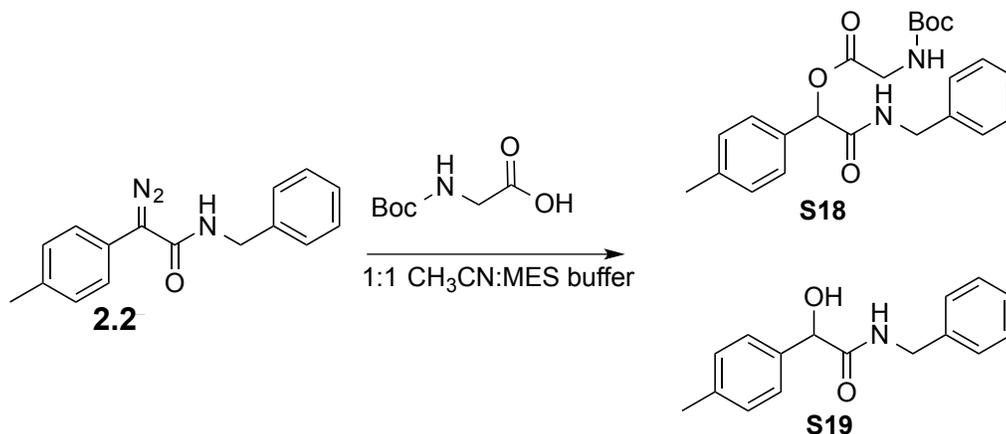
2.4.4 Esterification of BocGlyOH



Diazo compound **2.1** (0.005 g, 0.02 mmol) and BocGlyOH (0.003 g, 0.02 mmol) were added to a 1:1 solution of acetonitrile/100 mM MES–HCl buffer at pH 5.5, and the resulting solution was stirred for 6 h at ambient temperature. The reaction mixture was concentrated under reduced pressure, and the ratio of products was determined by integration of ^1H NMR spectral peaks.

Data for S16: ^1H NMR (400 MHz, CD_3CN , δ): 7.60 (s, 1H), 7.37–7.22 (m, 7H), 6.93 (d, 2H, $J = 8.4$ Hz), 5.91 (s, 1H), 5.74 (s, 1H), 4.43–4.31 (m, 2H), 3.94–3.82 (m, 2H), 3.79 (s, 3H), 1.38 (s, 9H). ^{13}C NMR (100 MHz, CD_3CN , δ): 170.4, 169.3, 161.1, 157.4, 139.9, 129.9, 129.3, 128.6, 128.1, 127.9, 114.8, 80.3, 76.7, 55.9, 43.2, 43.2, 28.4. HRMS (ESI $^+$) m/z calcd for $\text{C}_{23}\text{H}_{28}\text{N}_2\text{O}_6$ $[\text{M}+\text{H}]^+$ 429.2021; found, 429.2021.

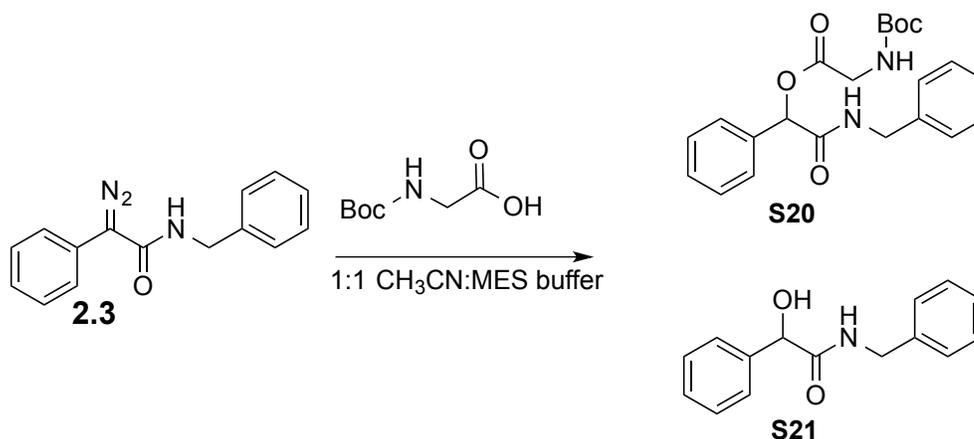
Data for S17: ^1H NMR (500 MHz, CD_3CN , δ): 7.47 (s, 1H), 7.33–7.25 (m, 4H), 7.23–7.21 (m, 3H), 6.90 (d, 2H, $J = 8.8$ Hz), 4.97 (d, 1H, $J = 4.5$ Hz), 4.40–4.32 (m, 2H), 4.16 (d, 2H, $J = 4.5$ Hz), 3.78 (s, 3H). ^{13}C NMR (125 MHz, CD_3CN , δ): 173.3, 160.4, 140.3, 133.8, 129.3, 129.0, 128.1, 127.8, 114.5, 74.3, 55.8, 43.1. HRMS (ESI $^+$) m/z calcd for $\text{C}_{16}\text{H}_{17}\text{NO}_3$ $[\text{M}+\text{H}]^+$ 272.1282; found, 272.1278.



Diazo compound **2.2** (0.005 g, 0.02 mmol) and BocGlyOH (0.003 g, 0.02 mmol) were added to 1:1 acetonitrile/100 mM MES–HCl buffer at pH 5.5, and the resulting solution was stirred for 6 h at ambient temperature. The solution was then concentrated under reduced pressure, and the ratio of products was determined by integration of ^1H NMR spectral peaks.

Data for S18: ^1H NMR (500 MHz, CD_3CN , δ): 7.65 (s, 1H), 7.33–7.28 (m, 4H), 7.25–7.20 (m, 5H), 5.92 (s, 1H), 5.77 (s, 1H), 4.42–4.31 (m, 2H), 3.92–3.82 (m, 2H), 2.34 (s, 3H), 1.38 (s, 9H). ^{13}C NMR (125 MHz, CD_3CN , δ) 170.4, 169.2, 157.4, 140.0, 139.8, 133.7, 130.1, 129.3, 128.3, 128.1, 127.9, 80.3, 76.8, 43.2, 43.2, 28.4, 21.2. HRMS (ESI $^+$) m/z calcd for $\text{C}_{23}\text{H}_{28}\text{N}_2\text{O}_5$ [$\text{M}+\text{NH}_4$] $^+$ 430.2337; found, 430.2336.

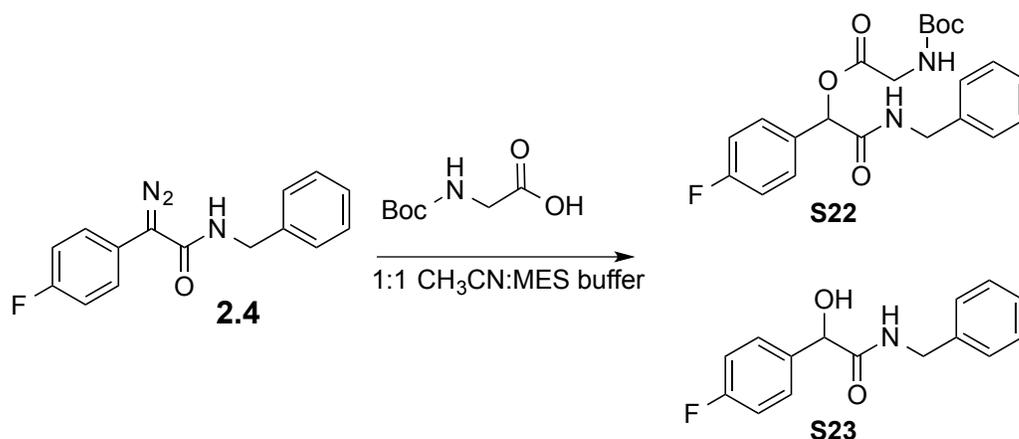
Data for S19: ^1H NMR (500 MHz, CD_3CN , δ): 7.46 (s, 1H), 7.31–7.28 (m, 4H), 7.25–7.21 (m, 3H), 7.17 (d, 2H, $J = 7.9$ Hz), 4.99 (d, 1H, $J = 4.2$ Hz), 4.40–4.32 (m, 2H), 4.18 (d, 1H, $J = 4.5$ Hz), 2.32 (s, 1H). ^{13}C NMR (125 MHz, CD_3CN , δ): 173.3, 140.3, 138.74, 138.71, 129.8, 129.3, 128.1, 127.9, 127.6, 74.6, 43.1, 21.1. HRMS (ESI $^+$) m/z calcd for $\text{C}_{16}\text{H}_{17}\text{NO}_2$ [$\text{M}+\text{H}$] $^+$ 256.1333; found, 256.1330.



Diazo compound **2.3** (0.005 g, 0.02 mmol) and BocGlyOH (0.004 g, 0.02 mmol) were added to 1:1 acetonitrile/100 mM MES–HCl buffer at pH 5.5, and the resulting solution was stirred for 6 h at ambient temperature. The reaction mixture was then concentrated under reduced pressure, and the ratio of products was determined by integration of ^1H NMR spectral peaks.

Data for S20: ^1H NMR (750 MHz, CD_3CN , δ): 7.65 (s, 1H), 7.46 (m, 2H), 7.40 (m, 3H), 7.30 (t, 2H, $J = 7.4$ Hz), 7.23 (m, 3H), 5.99 (s, 1H), 5.78 (s, 1H), 4.41 (dd, 1H, $J = 6.3, 15.2$ Hz), 4.35 (dd, 1H, $J = 6.1, 15.2$ Hz), 3.92 (dd, 1H, $J = 6.2, 17.9$ Hz), 3.88 (dd, 1H, $J = 5.7, 18.0$ Hz), 1.40 (s, 9H). ^{13}C NMR (125 MHz, CDCl_3 , δ): 168.7, 168.0, 156.4, 137.9, 135.0, 129.1, 128.8, 128.6, 127.8, 127.5, 127.4, 80.6, 76.2, 43.4, 43.0, 28.2. HRMS (ESI^+) m/z calcd for $\text{C}_{22}\text{H}_{26}\text{N}_2\text{O}_5$ $[\text{M}+\text{H}]^+$ 399.1915; found, 399.1917.

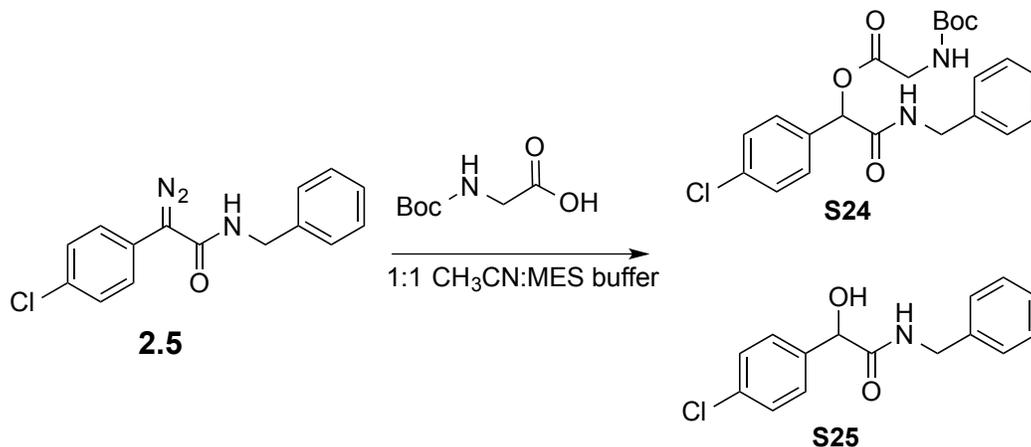
Data for S21: ^1H NMR (750 MHz, CD_3CN , δ): 7.48 (s, 1H), 7.43 (d, 2H, $J = 7.4$ Hz), 7.36 (t, 2H, $J = 7.4$ Hz), 7.31 (m, 3H), 7.24 (m, 3H), 5.04 (d, 1H, $J = 2.8$ Hz), 4.37 (m, 2H), 4.28 (d, 1H, $J = 3.8$ Hz). ^{13}C NMR (125 MHz, CD_3CN , δ): 173.1, 141.6, 140.3, 129.3, 129.2, 128.8, 128.1, 127.9, 127.6, 74.7, 43.1. HRMS (ESI^+) m/z calcd for $\text{C}_{15}\text{H}_{15}\text{NO}_2$ $[\text{M}+\text{H}]^+$ 242.1176; found, 242.1169.



Diazo compound **2.4** (0.005 g, 0.02 mmol) and BocGlyOH (0.003 g, 0.02 mmol) were added to 1:1 acetonitrile/100 mM MES–HCl buffer at pH 5.5, and the resulting solution was stirred for 6 h at ambient temperature. The reaction mixture was then concentrated under reduced pressure, and the ratio of products was determined by integration of ^1H NMR spectral peaks.

Data for S22: ^1H NMR (500 MHz, CD_3CN , δ): 7.66 (s, 1H), 7.48 (dd, 2H, $J = 5.4, 8.6$ Hz), 7.30 (t, 2H, $J = 7.3$ Hz), 7.25–7.20 (m, 3H), 7.14 (t, 2H, $J = 8.9$ Hz), 5.97 (s, 1H), 5.77 (s, 1H), 4.40 (dd, 1H, $J = 6.3, 15.2$ Hz), 4.34 (dd, 1H, $J = 6.1, 15.2$ Hz), 3.94–3.84 (m, 2H), 1.38 (s, 9H). ^{13}C NMR (125 MHz, CDCl_3 , δ): 168.6, 167.9, 163.1 (d, $J = 248.2$ Hz), 156.4, 137.8, 131.0 (d, $J = 3.3$ Hz), 129.4 (d, $J = 8.5$ Hz), 127.8, 127.5, 115.8 (d, $J = 21.8$ Hz), 80.7, 75.5, 43.4, 43.0, 28.2. HRMS (ESI^+) m/z calcd for $\text{C}_{22}\text{H}_{25}\text{FN}_2\text{O}_5$ $[\text{M}+\text{H}]^+$ 417.1821; found, 417.1816.

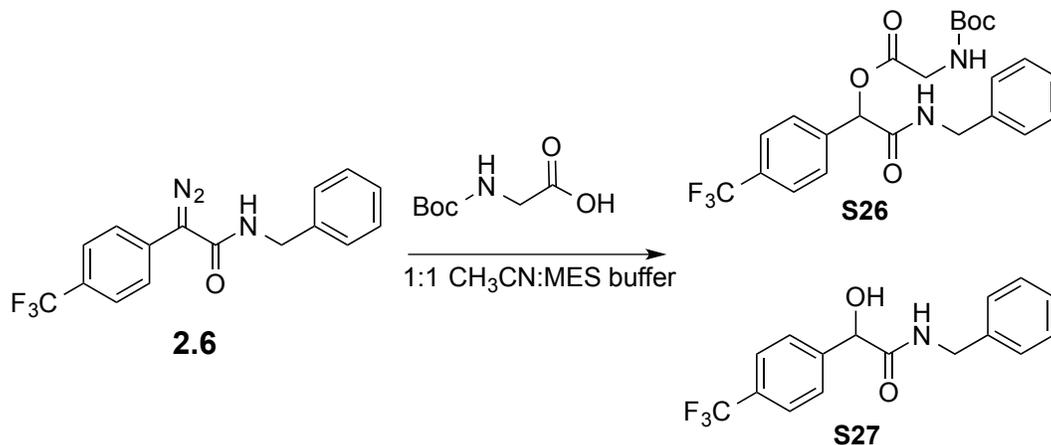
Data for S23: ^1H NMR (400 MHz, CD_3CN , δ): 7.53 (s, 1H), 7.45–7.42 (m, 2H), 7.32–7.28 (m, 2H), 7.24–7.20 (m, 3H), 7.09 (t, 2H, $J = 8.9$ Hz), 5.04 (s, 1H), 4.41–4.31 (m, 2H). ^{13}C NMR (125 MHz, CD_3CN , δ): 174.7, 165.0 (d, $J = 243.7$ Hz), 142.0, 139.6, 131.3 (d, $J = 8.3$ Hz), 131.1., 129.8, 129.6, 117.6 (d, $J = 21.7$ Hz), 75.7, 44.8. HRMS (ESI^+) m/z calcd for $\text{C}_{15}\text{H}_{14}\text{FNO}_2$ $[\text{M}+\text{H}]^+$ 260.1082; found, 260.1080.



Diazo compound **2.5** (0.005 g, 0.02 mmol) and BocGlyOH (0.003 g, 0.02 mmol) were added to 1:1 acetonitrile/100 mM MES–HCl buffer at pH 5.5, and the resulting solution was stirred for 6 h at ambient temperature. The reaction mixture was then concentrated under reduced pressure, and the ratio of products was determined by integration of ¹H NMR spectral peaks.

Data for S24: ¹H NMR (500 MHz, CD₃CN, δ): 7.61 (s, 1H), 7.45–7.40 (m, 4H), 7.31–7.29 (m, 2H), 7.25–7.21 (m, 3H), 5.98 (s, 1H), 5.74 (s, 1H), 4.42–4.32 (m, 2H), 3.90 (m, 2H), 1.39 (s, 9H). ¹³C NMR (125 MHz, CDCl₃, δ): 168.5, 167.6, 156.4, 137.7, 135.1, 135.6, 128.9, 128.8, 128.6, 127.8, 127.5, 80.8, 75.4, 43.4, 43.0, 28.2. HRMS (ESI⁺) *m/z* calcd for C₂₂H₂₅ClN₂O₅ [M+NH₄]⁺ 450.1791; found, 450.1785.

Data for S25: ¹H NMR (500 MHz, CD₃CN, δ): 7.47 (s, 1H), 7.42 (d, 2H, *J* = 8.5 Hz), 7.37 (d, 2H, 8.6 Hz), 7.32–7.29 (m, 2H), 7.25–7.21 (m, 3H), 5.04 (d, 1H, *J* = 1.8 Hz), 4.36 (m, 2H), 4.31 (d, 1H, *J* = 3.4 Hz). ¹³C NMR (125 MHz, CD₃CN, δ): 172.7, 140.5, 140.2, 134.0, 129.3, 129.21, 129.18, 128.1, 127.9, 73.9, 43.1. HRMS (ESI⁺) *m/z* calcd for C₁₅H₁₄ClNO₂ [M+H]⁺ 276.0786; found, 276.0789.

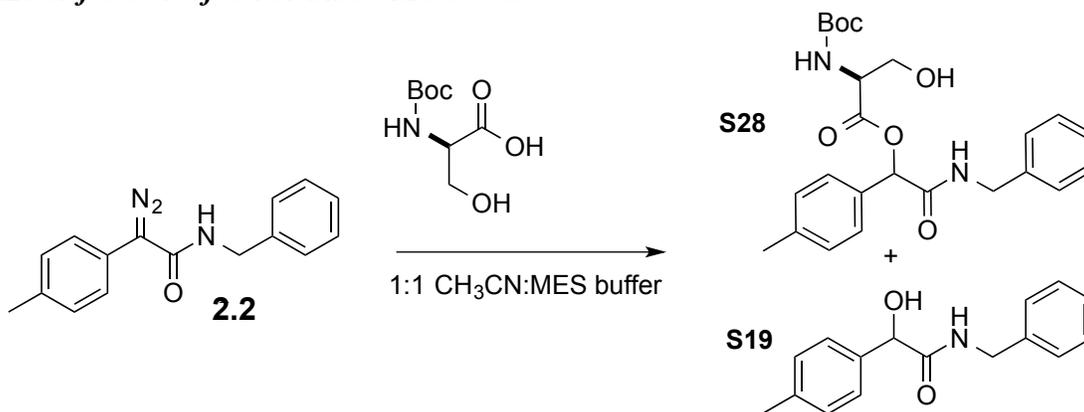


Diazo compound **2.6** (0.005 g, 0.02 mmol) and BocGlyOH (0.003 g, 0.02 mmol) were added to 1:1 acetonitrile/100 mM MES–HCl buffer at pH 5.5, and the resulting solution was stirred for 6 h at ambient temperature. The reaction mixture was then concentrated under reduced pressure, and the ratio of products was determined by integration of ^1H NMR spectral peaks.

Data for S26: ^1H NMR (500 MHz, CD_3CN , δ): 7.73–7.71 (m, 3H), 7.65 (d, 2H, $J = 8.3$ Hz), 7.31–7.28 (m, 2H), 7.25–7.20 (m, 3H), 6.06 (s, 1H), 5.77 (s, 1H), 4.42–4.32 (m, 2H), 3.97–3.87 (m, 2H), 1.38 (s, 1H). ^{13}C NMR (125 MHz, CD_3CN , δ): 170.3, 168.4, 157.4, 141.1, 139.7, 131.1 (q, $J = 32.4$ Hz), 129.4, 128.8, 128.1, 128.0, 126.3 (q, $J = 3.9$ Hz), 125.1 (q, $J = 271.3$ Hz), 80.4, 76.1, 43.4, 43.2, 28.4. HRMS (ESI $^+$) m/z calcd for $\text{C}_{23}\text{H}_{25}\text{F}_3\text{N}_2\text{O}_5$ $[\text{M}+\text{NH}_4]^+$ 484.2037; found, 484.2054.

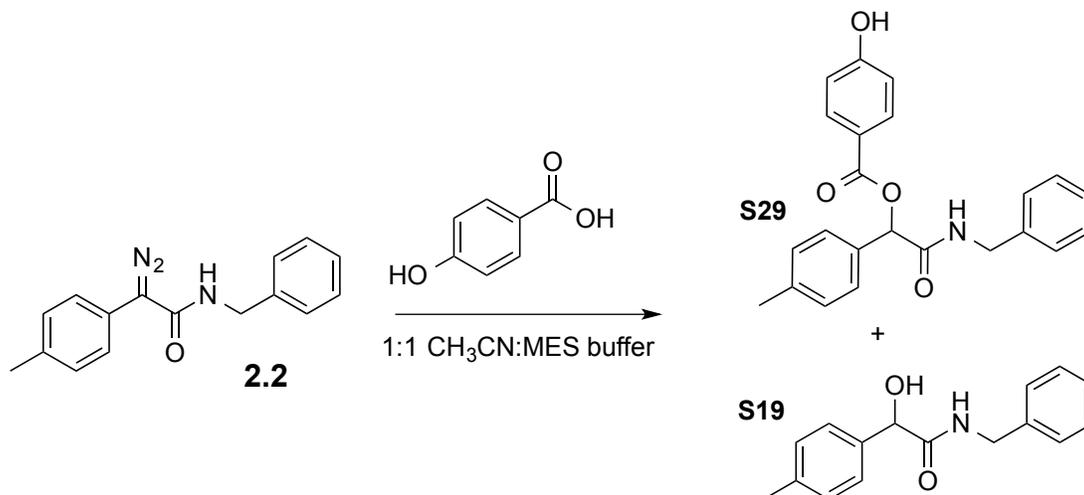
Data for S27: ^1H NMR (400 MHz, CD_3CN , δ): 7.69–7.62 (m, 4H), 7.56 (s, 1H), 7.31–7.20 (m, 5H), 5.54 (s, 1H), 5.14 (d, 1H, $J = 4.6$ Hz), 4.45 (d, 1H, $J = 4.8$ Hz), 4.37–4.35 (m, 2H). ^{13}C NMR (125 MHz, CD_3CN , δ): 172.3, 146.0, 140.1, 130.1 (q, $J = 32.3$ Hz), 129.3, 128.1, 128.9, 126.2 (q, $J = 41.3$ Hz), 125.3 (q, $J = 271.3$ Hz), 74.0, 43.1. HRMS calcd for $(\text{C}_{16}\text{H}_{14}\text{F}_3\text{NO}_2)$ $[\text{M}+\text{H}]^+$ 310.1050; found, 310.1043.

2.4.5 Esterification of Other Small Molecules



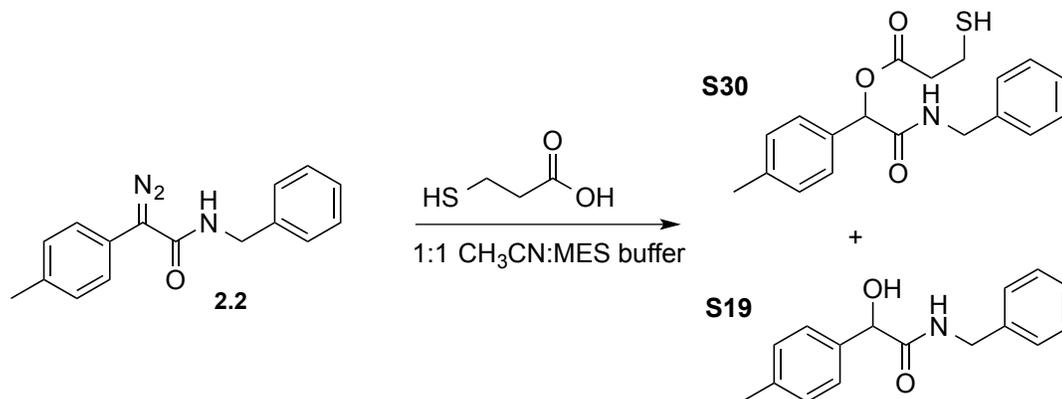
Diazo compound **2.2** (0.005 g, 0.02 mmol) and BocSerOH (0.004 g, 0.02 mmol) were added to 1:1 acetonitrile/100 mM MES–HCl buffer at pH 5.5, and the resulting solution was stirred for 6 h at ambient temperature. The solution was then concentrated under reduced pressure, and the ratio of products was determined by integration of ¹H NMR spectral peaks. Data for **S19** are reported above; data for **S28** are reported below (both diastereomers). No other products were observed by TLC or ¹H NMR spectroscopy.

Data for S28: ¹H NMR (500 MHz, CD₃CN, Diastereomer **A**, δ): 7.72 (s, 1H), 7.35 (d, 2H, $J = 8.0$ Hz), 7.30 (t, 2H, $J = 7.3$ Hz), 7.24 (t, 3H, $J = 7.7$ Hz), 7.18 (d, 2H, $J = 7.2$ Hz), 5.96 (s, 1H), 5.79 (d, 1H, $J = 6.8$ Hz), 4.38–4.33 (m, 2H), 4.32–4.29 (m, 1H), 4.08–4.03 (m, 1H), 3.77–3.69 (m, 2H), 2.34 (s, 3H), 1.40 (s, 9H). ¹H NMR (500 MHz, CD₃CN, Diastereomer **B**, δ): 7.64 (s, 1H), 7.36–7.28 (m, 4H), 7.25–7.17 (m, 5H), 5.95 (s, 1H), 5.84 (d, 1H, $J = 7.8$ Hz), 4.41–4.30 (m, 2H), 4.28–4.25 (m, 1H), 3.86–3.82 (m, 1H), 3.79–3.72 (m, 1H), 3.41 (t, 3H, $J = 5.7$ Hz), 2.34 (s, 3H), 1.36 (s, 9H). ¹³C NMR (125 MHz, CD₃CN, Diastereomer **A**, δ): 171.3, 169.7, 157.0, 140.2, 139.6, 133.2, 130.2, 129.3, 128.5, 128.1, 128.0, 80.3, 77.0, 63.3, 57.1, 43.4, 28.4, 21.2. ¹³C NMR (125 MHz, CD₃CN, Diastereomer **B**, δ): 171.2, 169.3, 156.7, 139.9, 139.8, 133.6, 130.1, 129.3, 128.4, 128.1, 127.9, 80.3, 77.0, 62.8, 57.1, 43.3, 28.4, 21.1. HRMS (ESI⁺) m/z calcd for C₂₄H₃₀N₂O₆ [M+H]⁺ 443.2177; found, 443.2185 (Diastereomer **A**), 443.2183 (Diastereomer **B**).



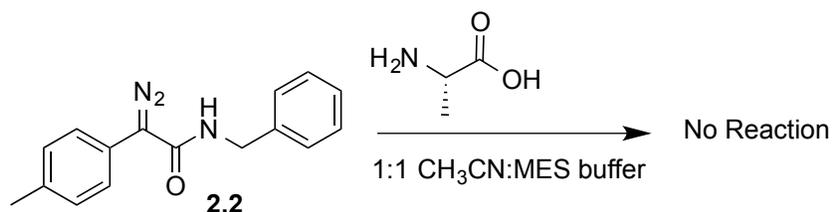
Diazo compound **2.2** (0.005 g, 0.02 mmol) and *p*-hydroxybenzoic acid (0.003 g, 0.02 mmol) were added to 1:1 acetonitrile/100 mM MES–HCl buffer at pH 5.5, and the resulting solution was stirred for 6 h at ambient temperature. The solution was then concentrated under reduced pressure, and the ratio of products was determined by integration of ¹H NMR spectral peaks. Data for **S19** are reported above; data for **S29** are reported below. No other products were observed by TLC or ¹H NMR spectroscopy.

Data for S29: ¹H NMR (500 MHz, CD₃CN, δ): 7.98 (d, 2H, $J = 8.8$ Hz), 7.76 (s, 1H), 7.44 (d, 2H, $J = 8.1$ Hz), 7.39 (s, 1H), 7.29–7.18 (m, 7H), 6.89 (d, 2H, $J = 8.8$ Hz), 6.06 (s, 1H), 4.36 (d, 2H, $J = 6.2$ Hz), 2.35 (s, 3H). ¹³C NMR (125 MHz, CD₃CN, δ): 169.7, 165.8, 162.6, 140.0, 139.8, 134.2, 133.0, 130.1, 129.3, 128.3, 128.0, 127.9, 121.9, 116.1, 76.8, 43.1, 21.2. HRMS (ESI⁺) m/z calcd for C₂₃H₂₁NO₄ [M+H]⁺ 376.1544; found, 376.1539.



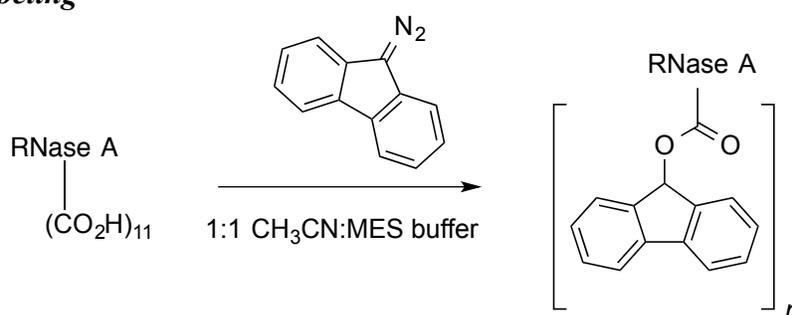
Diazo compound **2.2** (0.005 g, 0.02 mmol) and 3-mercaptopropanoic acid (0.002 g, 0.02 mmol) were added to 1:1 acetonitrile/100 mM MES-HCl buffer at pH 5.5, and the resulting solution was stirred for 6 h at ambient temperature. The solution was then concentrated under reduced pressure, and the ratio of products was determined by integration of ¹H NMR spectral peaks. Data for **S19** are reported above; data for **S30** are reported below. No other products were observed by TLC or ¹H NMR spectroscopy.

Data for S30: ¹H NMR (500 MHz, CD₃CN, δ): 7.38 (s, 1H), 7.34 (d, 2H, *J* = 8.1 Hz), 7.29 (t, 2H, *J* = 7.3 Hz), 7.25–7.19 (m, 5H), 5.91 (s, 1H), 4.35 (d, 2H, *J* = 6.2 Hz), 2.80–2.70 (m, 4H), 2.34 (s, 3H), 1.89 (t, 1H, *J* = 8.2 Hz). ¹³C NMR (125 MHz, CD₃CN, δ): 171.5, 169.4, 139.9, 139.8, 133.9, 130.1, 129.3, 128.3, 128.1, 127.9, 76.6, 43.1, 39.1, 21.1, 20.2. HRMS (ESI⁺) *m/z* calcd for (C₁₉H₂₁NO₃S) [M+H]⁺ 344.1315; found, 344.1315.

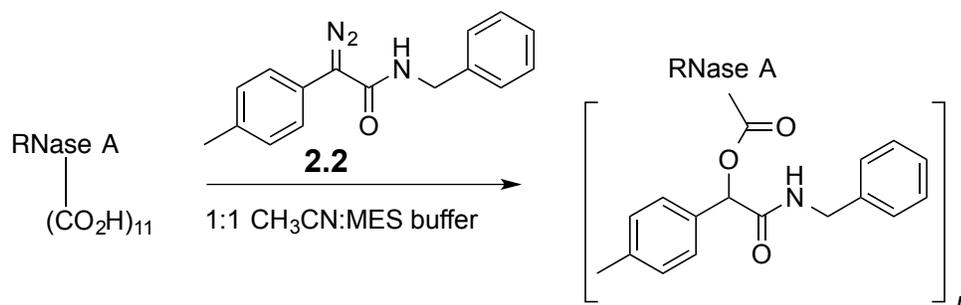


Diazo compound **2.2** (0.005 g, 0.02 mmol) and AlaOH (0.002 g, 0.02 mmol) were added to 1:1 acetonitrile/100 mM MES–HCl buffer at pH 5.5, and the resulting solution was stirred for 6 h at ambient temperature. The solution was then concentrated under reduced pressure, and the crude reaction mixture was analyzed by ^1H NMR spectroscopy (Figure 2.5A) and LC–MS (Figure 2.5B), which revealed no reaction.

2.4.6 Protein Labeling



9-Diazafluorene was prepared as described previously.⁷⁸ Yields and spectra matched the published data. Ribonuclease A (0.010 g, 0.73 μmol) was dissolved in 1 mL of 10 mM MES–HCl buffer at pH 5.5. 9-Diazafluorene (0.007 g, 0.036 mmol) was dissolved in 5 mL of CH_3CN . A 100- μL aliquot of the diazo stock solution was added to a 100- μL aliquot of the RNase A stock solution. The resulting mixture was mixed by nutation for 4 h at 37 $^\circ\text{C}$. Any remaining diazo compound was then quenched by addition of 10 μL of 17.4 M acetic acid. Acetonitrile was removed by concentration under reduced pressure, and the aqueous solution of labeled protein was analyzed by MALDI–TOF mass spectrometry (Figure 2.6).



Ribonuclease A (0.010 g, 0.73 μmol) was dissolved in 1 mL of 10 mM MES–HCl buffer at pH 5.5. Diazo compound **2.2** (0.095 g, 0.036 mmol) was dissolved in 5 mL of CH_3CN . A 100- μL aliquot of the diazo stock solution was added to a 100- μL aliquot of the RNase A stock solution. The resulting mixture was mixed by nutation for 4 h at 37 $^\circ\text{C}$. Any remaining diazo compound was then quenched by addition of 10 μL of 17.4 M acetic acid. Acetonitrile was removed by concentration under reduced pressure, and the aqueous solution of labeled protein was analyzed by MALDI–TOF mass spectrometry (Figures 2.6).

2.4.7 Ultraviolet Spectra of Diazo Compound 2.2

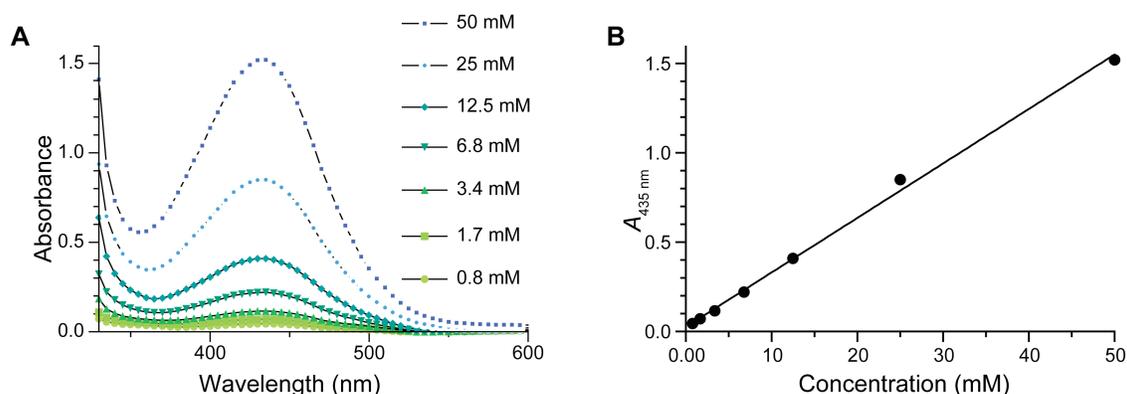
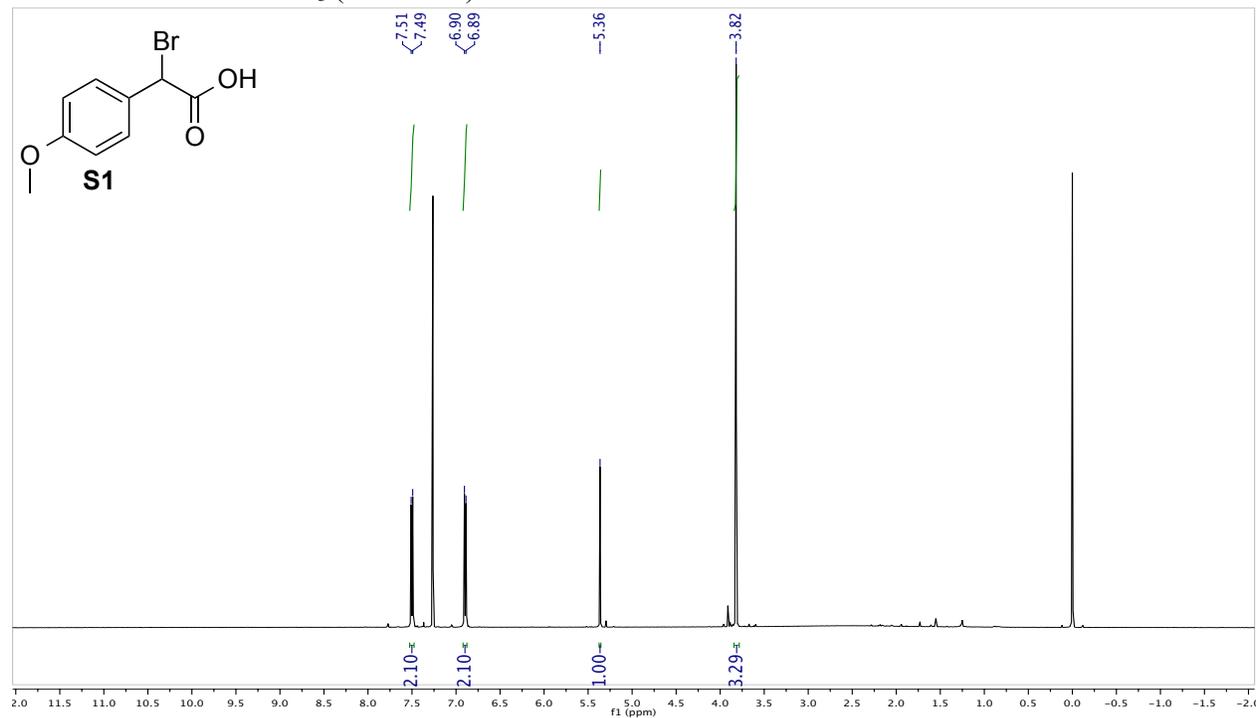


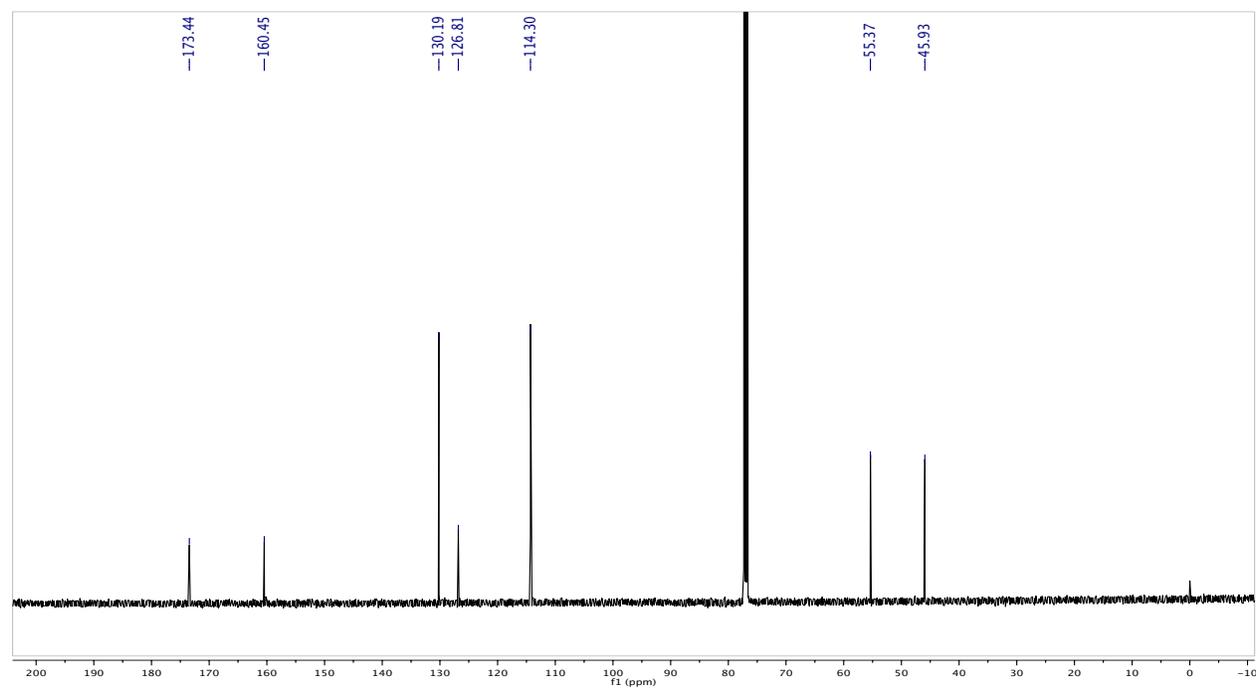
Figure 2.8 (A) Ultraviolet spectra of diazo compound **2** (0.8–50 mM). (B) Plot of the concentration dependence of the absorbance of diazo compound **2** (0.8–50 mM) at $\lambda_{\text{max}} = 435 \text{ nm}$, giving $\epsilon = 30.5 \text{ M}^{-1} \text{ cm}^{-1}$.

2.4.9 NMR Spectra

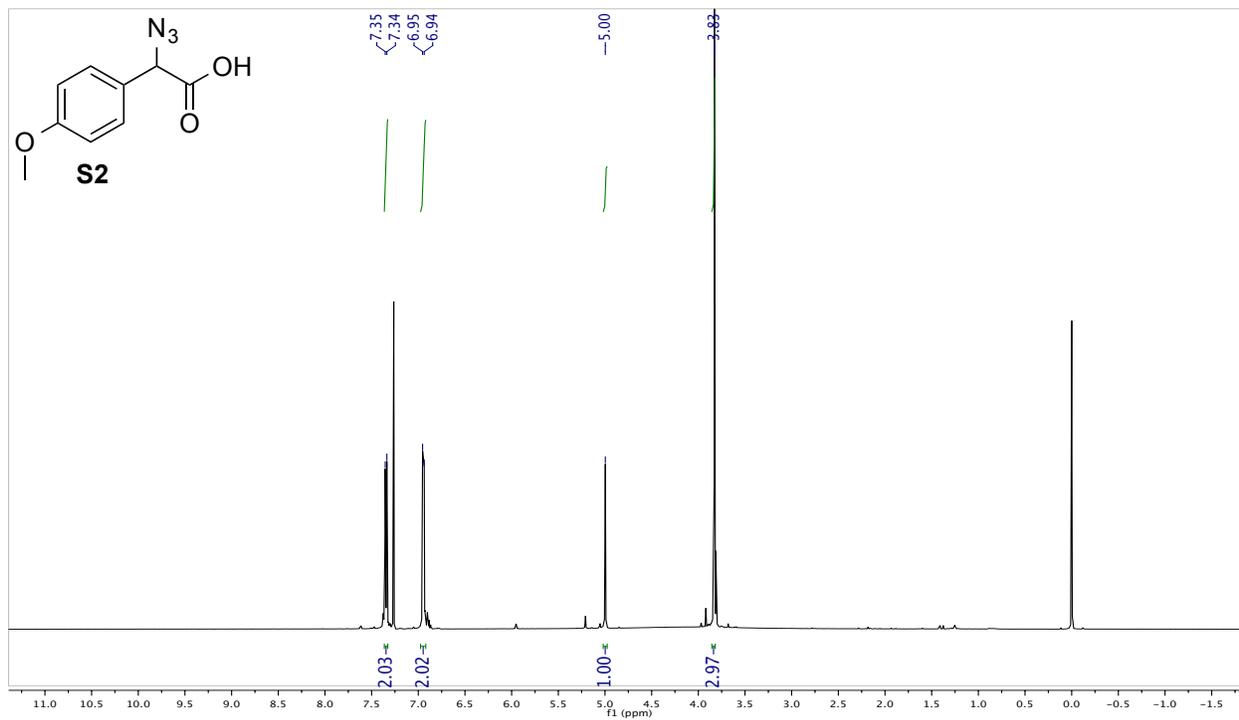
^1H NMR of S1 in CDCl_3 (500 MHz):



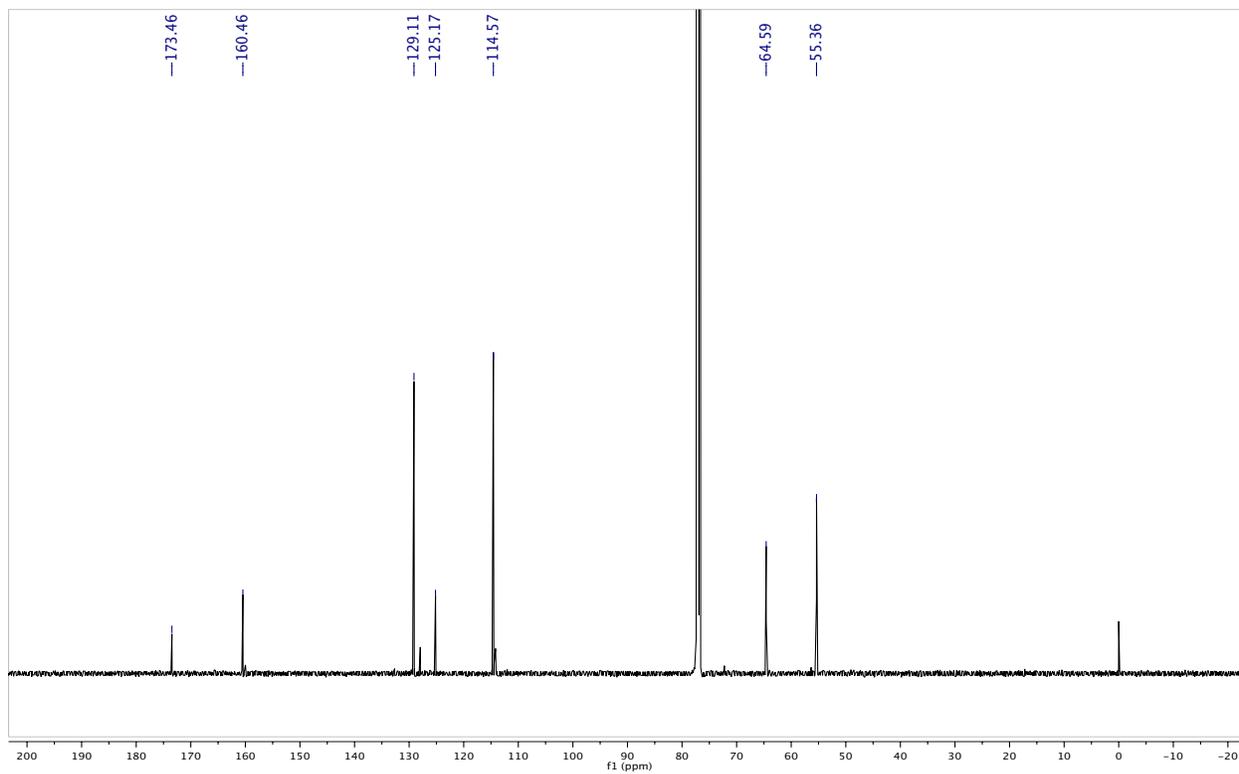
^{13}C NMR of S1 in CDCl_3 (125 MHz):



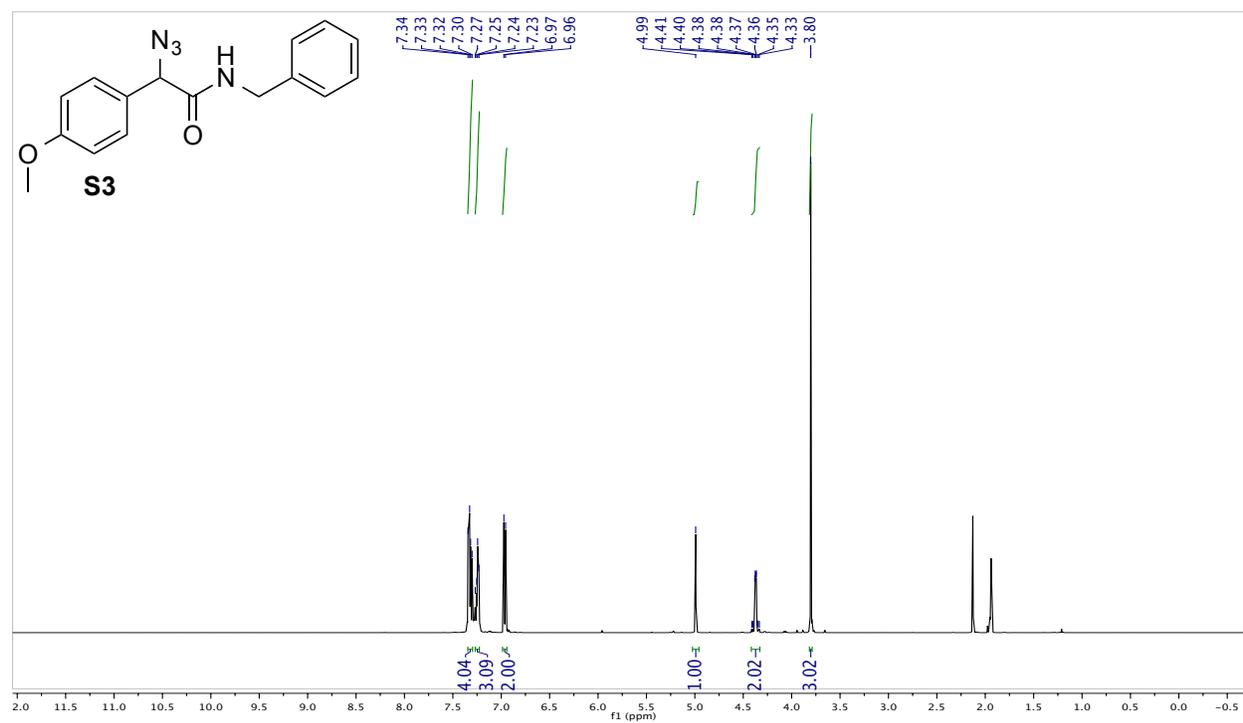
^1H NMR of S2 in CDCl_3 (500 MHz):



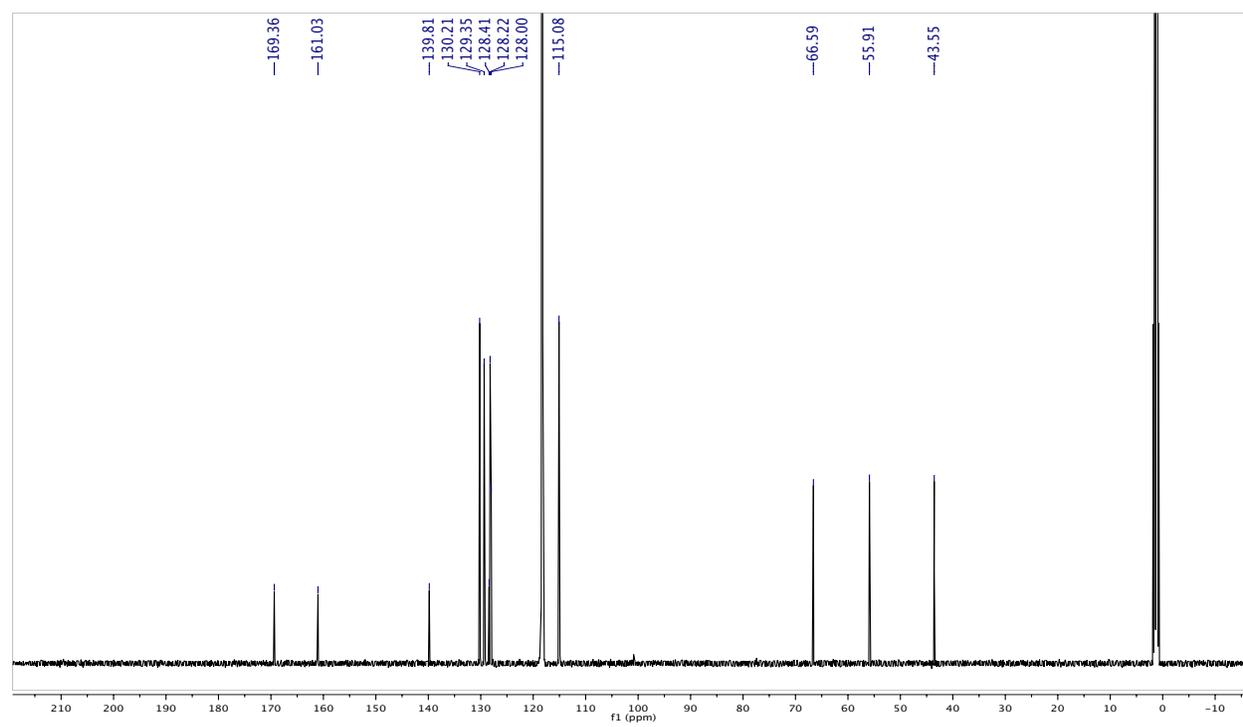
^{13}C NMR of S2 in CDCl_3 (125 MHz):



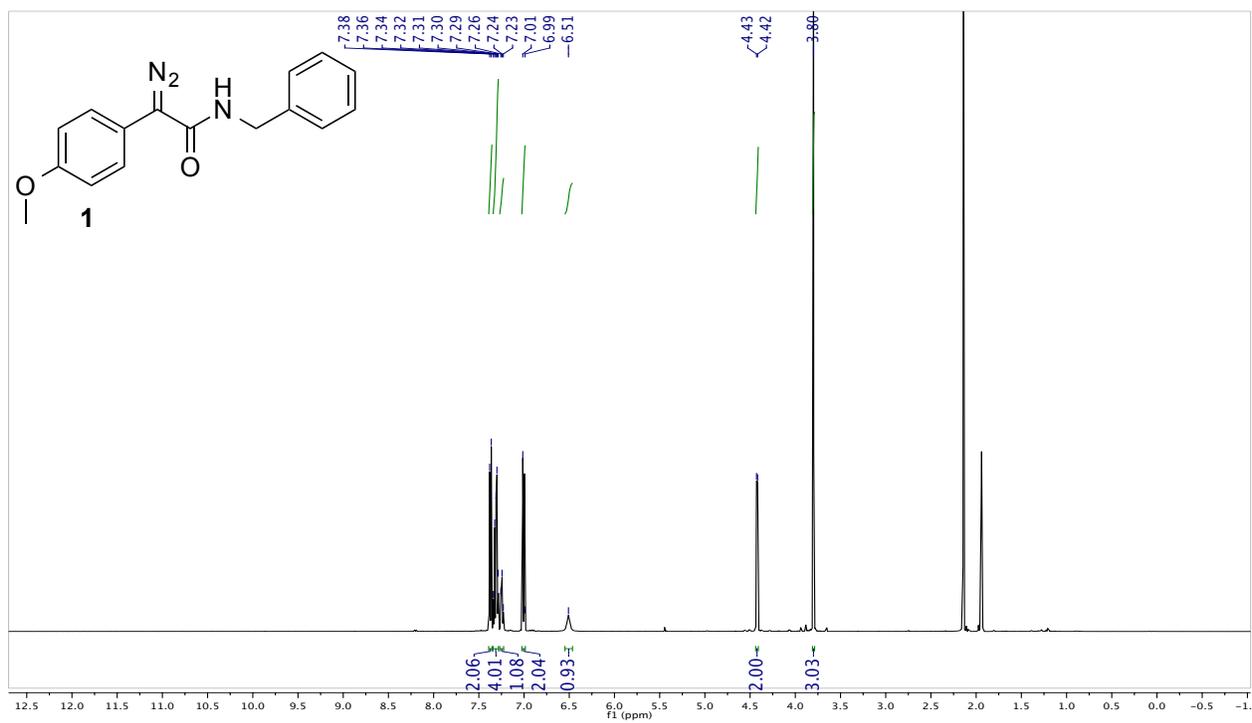
^1H NMR of S3 in CD_3CN (500 MHz):



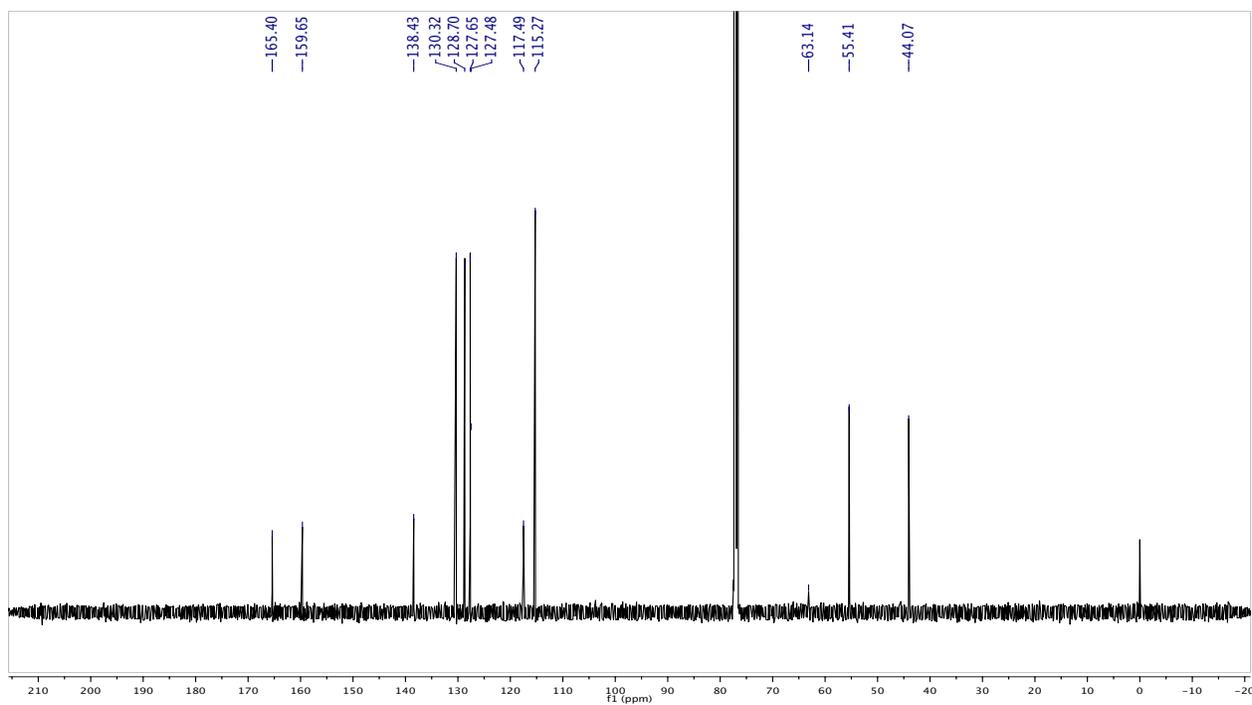
^{13}C NMR of S3 in CD_3CN (125 MHz):



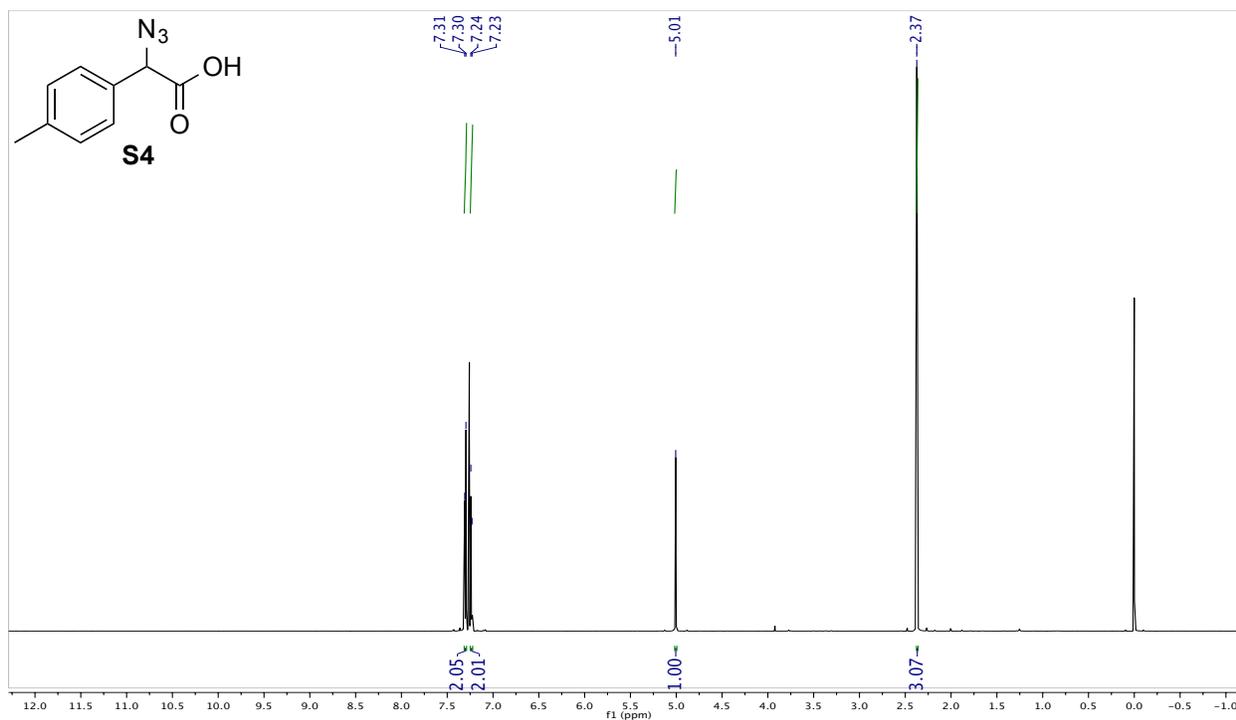
^1H NMR of 1 in CD_3CN (500 MHz):



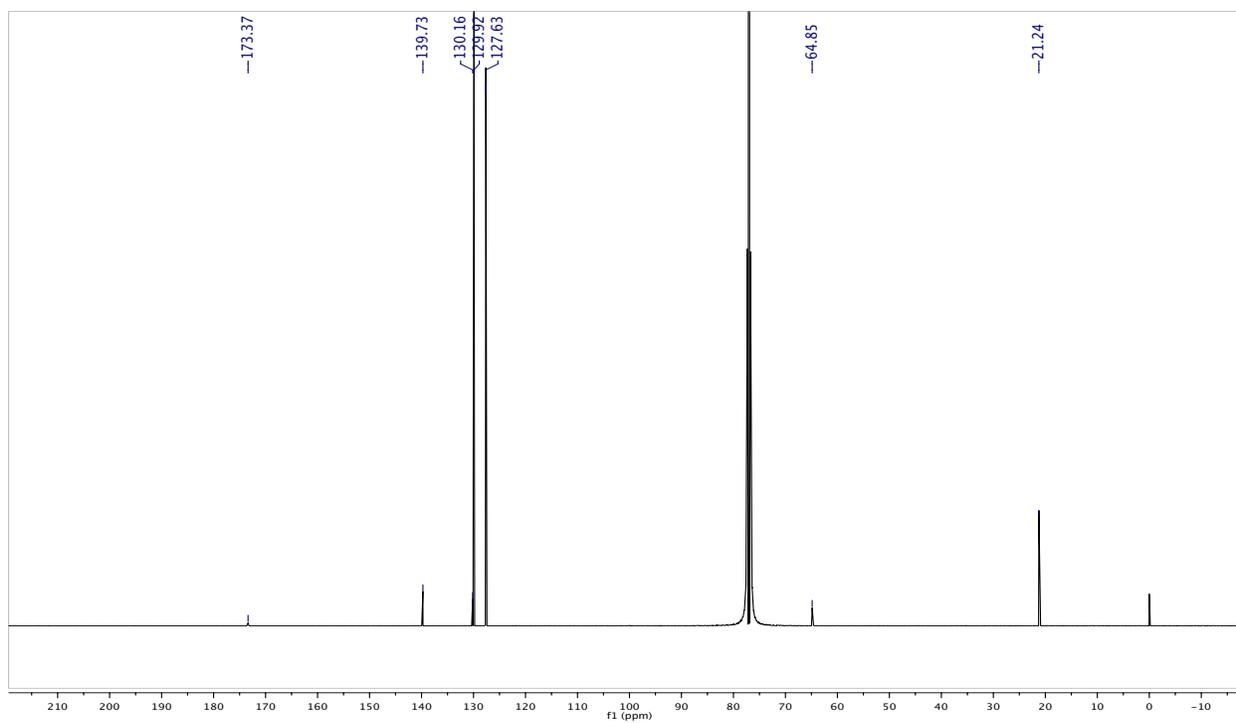
^{13}C NMR of 1 in CDCl_3 (125 MHz):



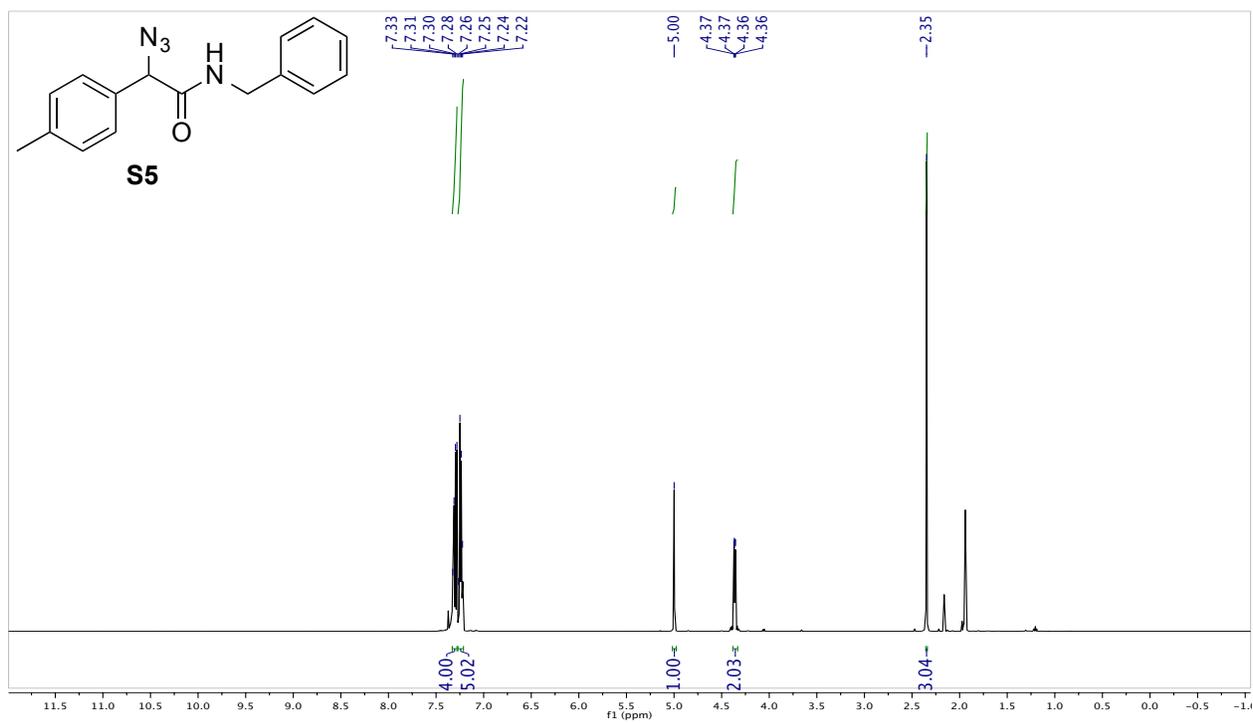
^1H NMR of S4 in CDCl_3 (600 MHz):



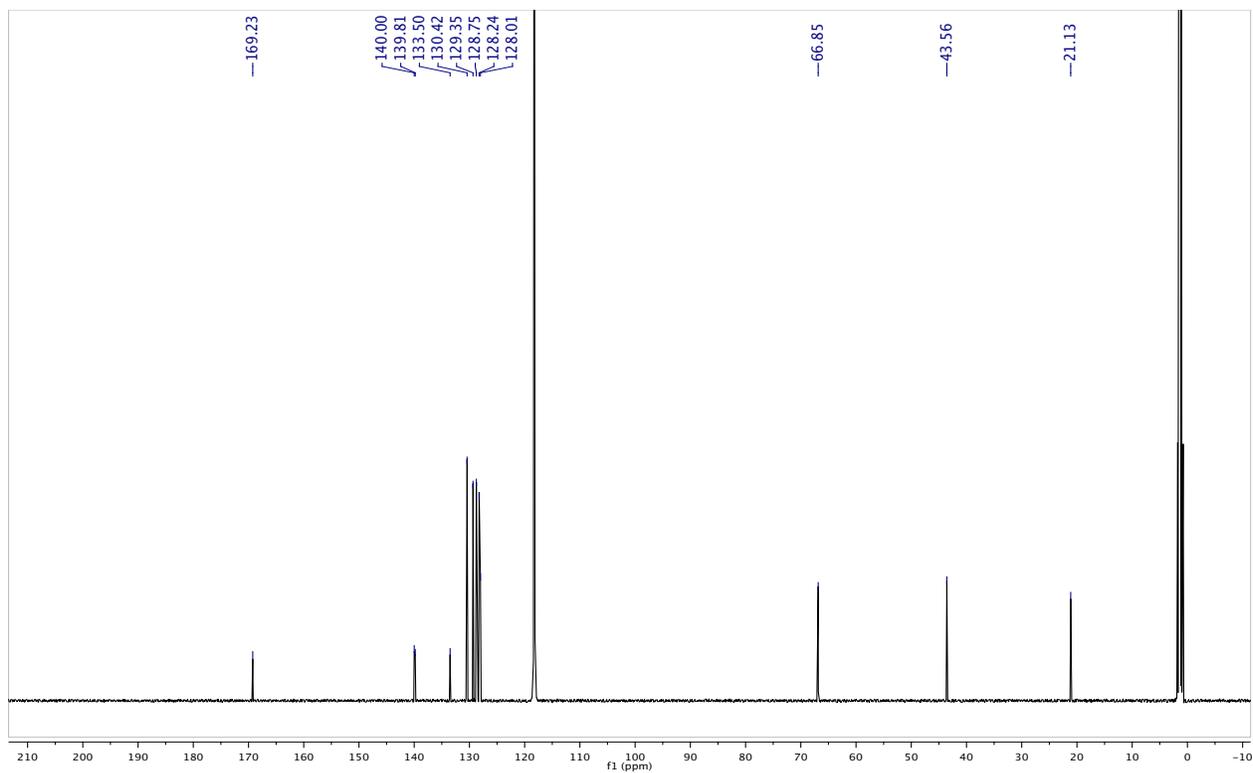
^{13}C NMR of S4 in CDCl_3 (150 MHz):



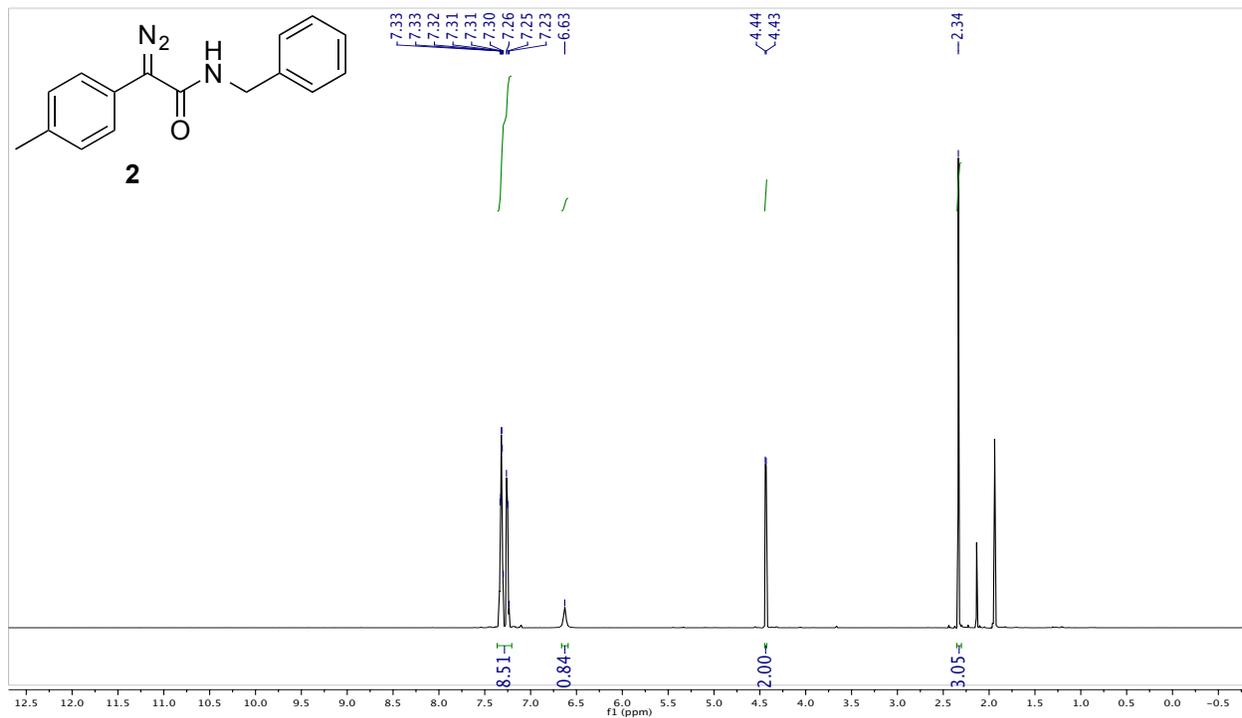
^1H NMR of S5 in CD_3CN (500 MHz):



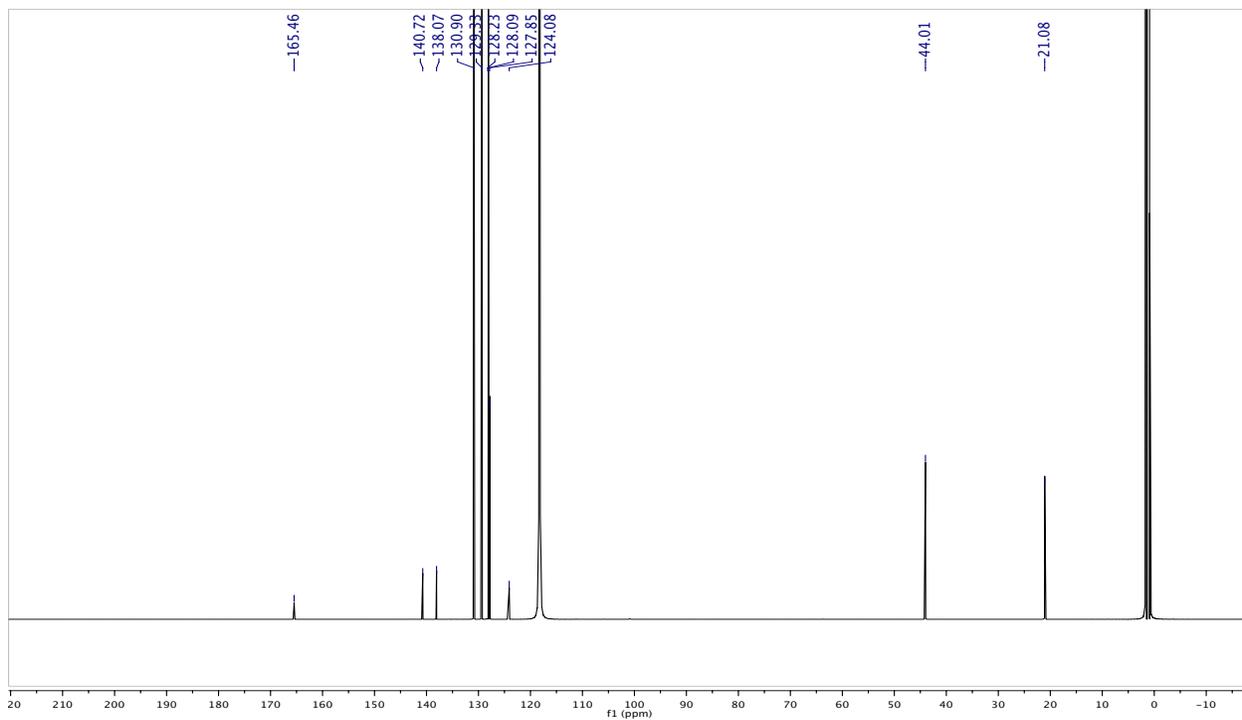
^{13}C NMR of S5 in CD_3CN (125 MHz):



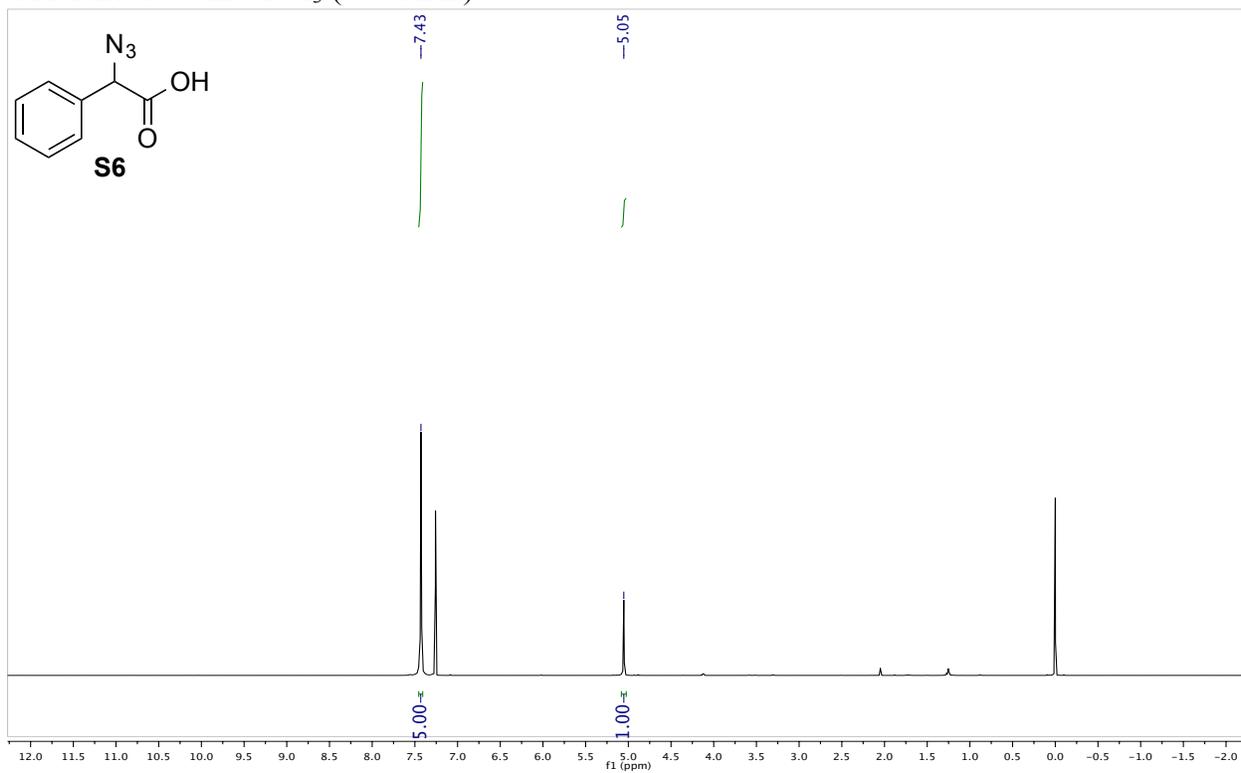
^1H NMR of 2 in CD_3CN (600 MHz):



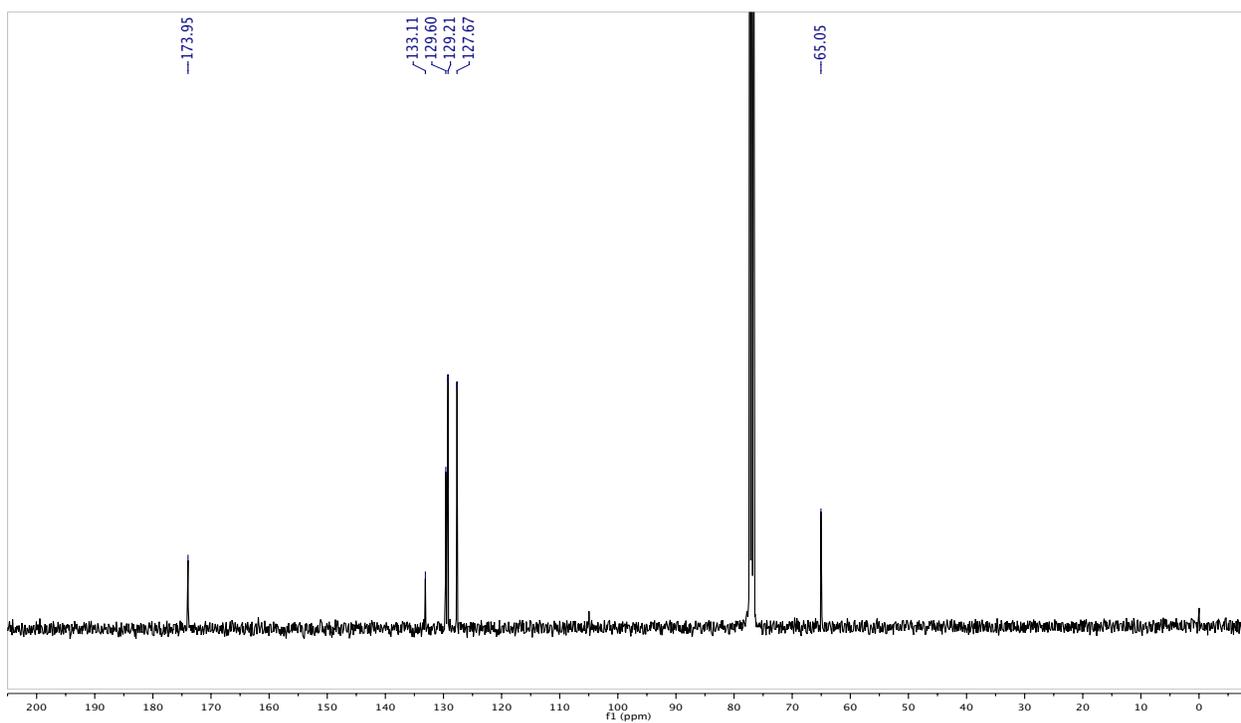
^{13}C NMR of 2 in CD_3CN (150 MHz):



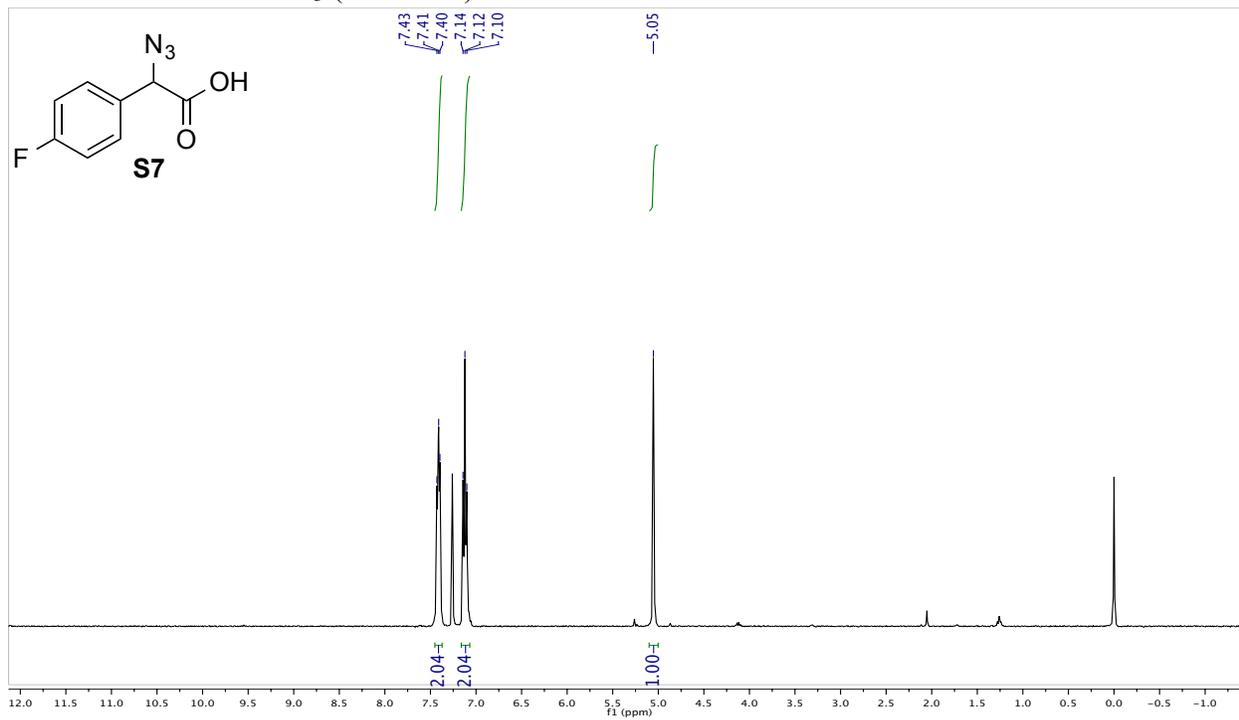
^1H NMR of S6 in CDCl_3 (400 MHz):



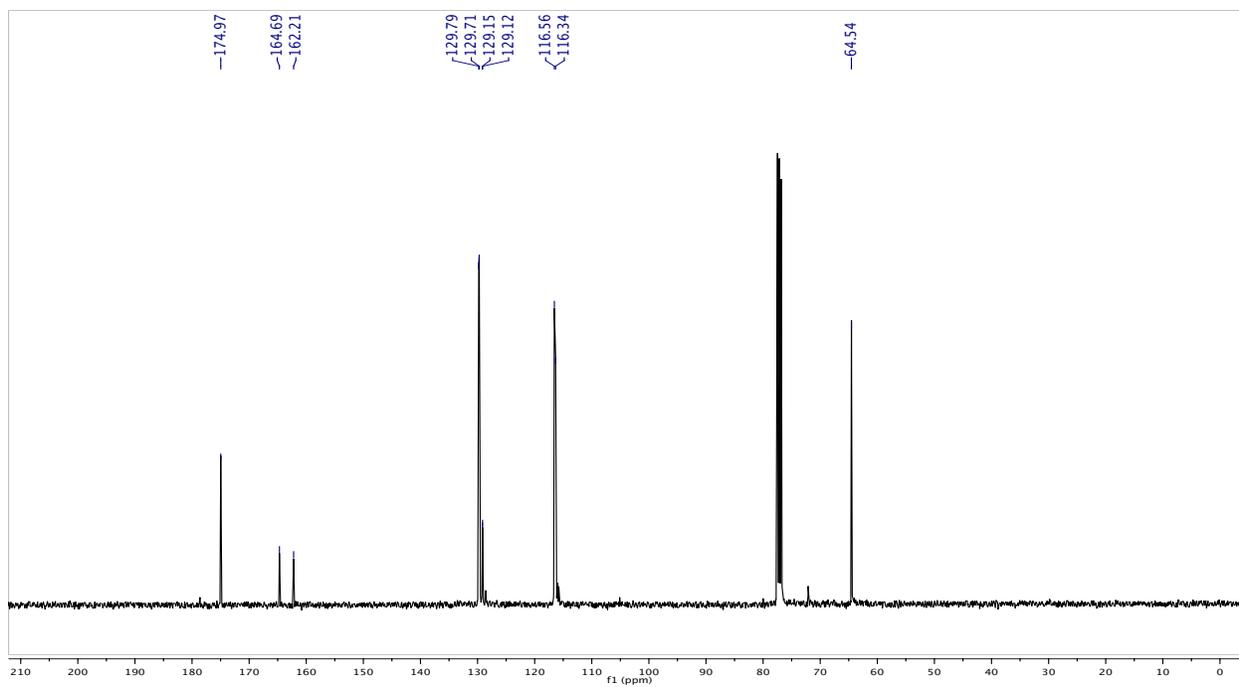
^{13}C NMR of S6 in CDCl_3 (100 MHz):



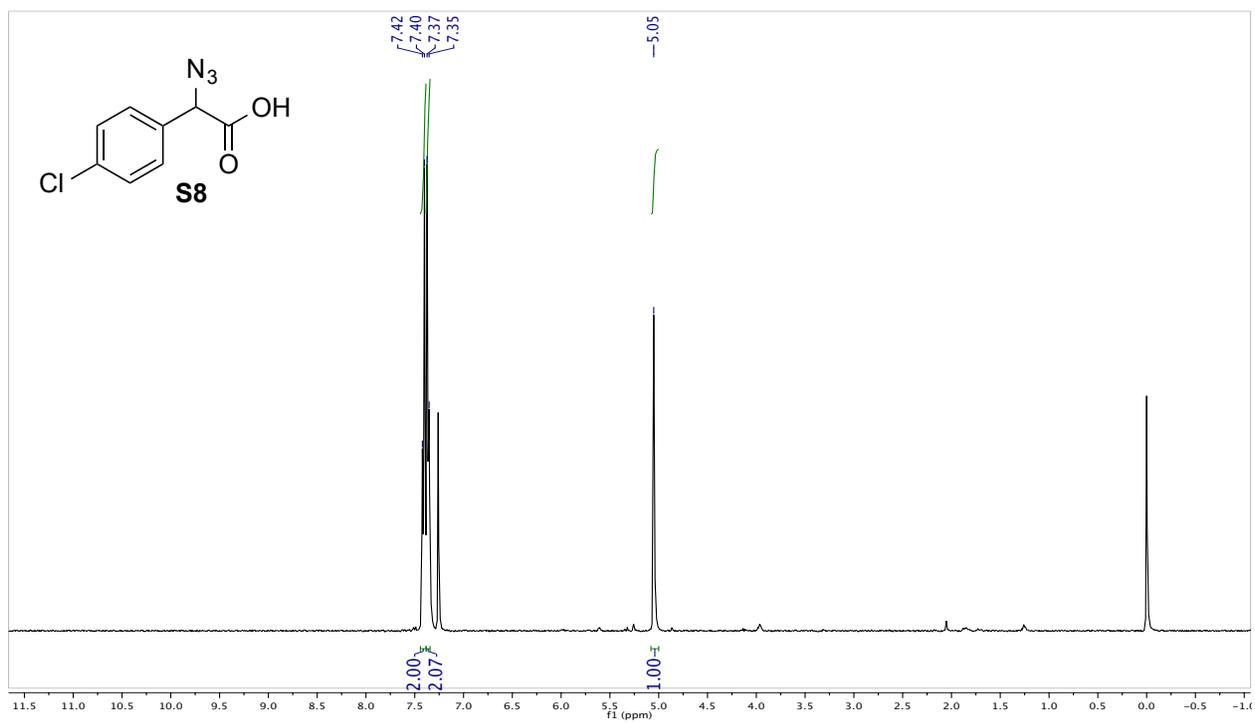
^1H NMR of S7 in CDCl_3 (400 MHz):



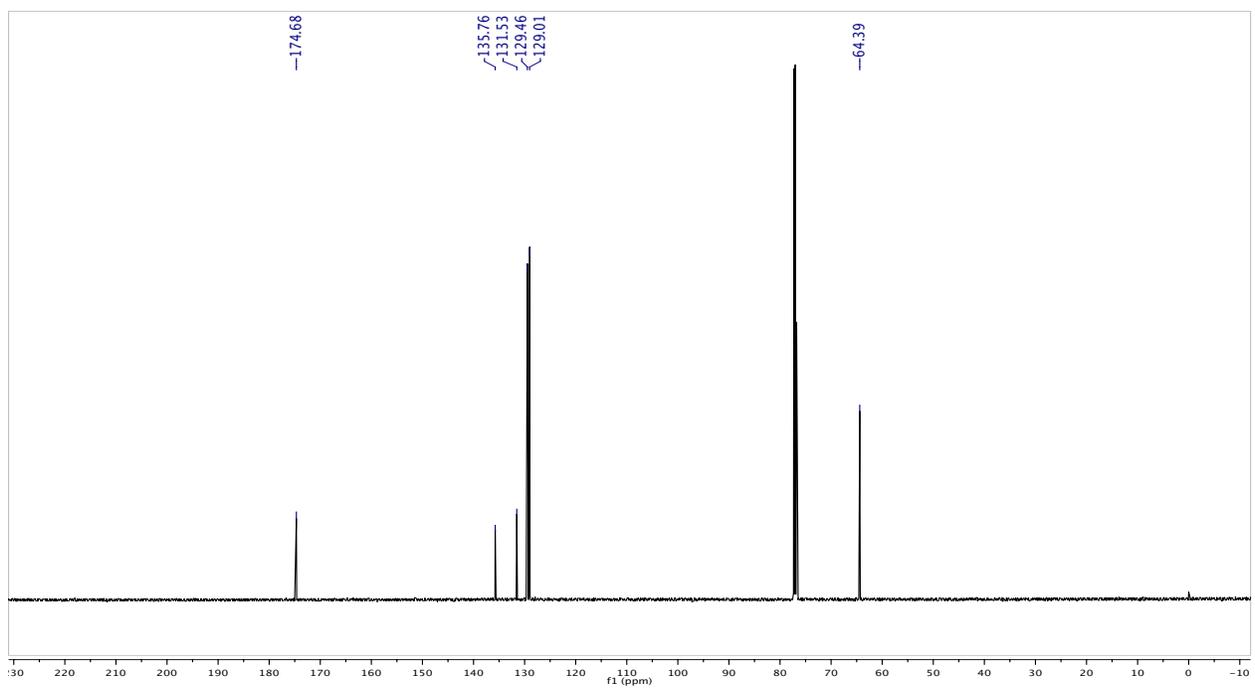
^{13}C NMR of S7 in CDCl_3 (100 MHz):



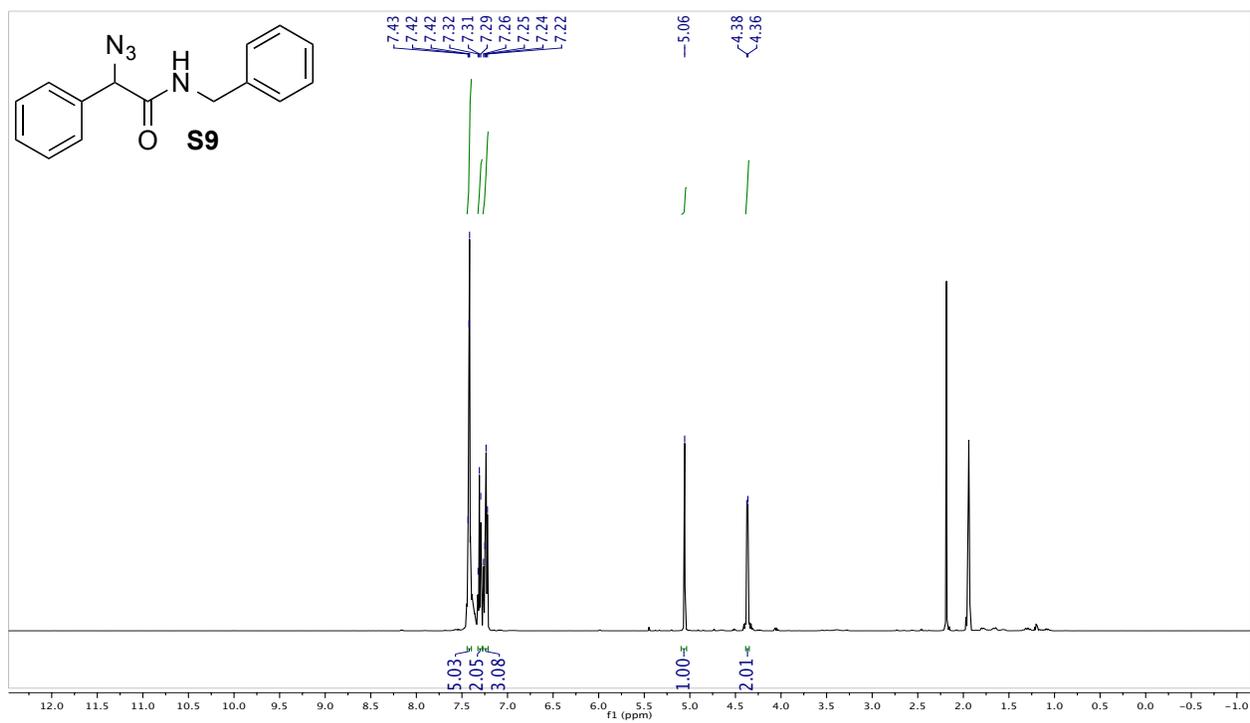
^1H NMR of S8 in CDCl_3 (400 MHz):



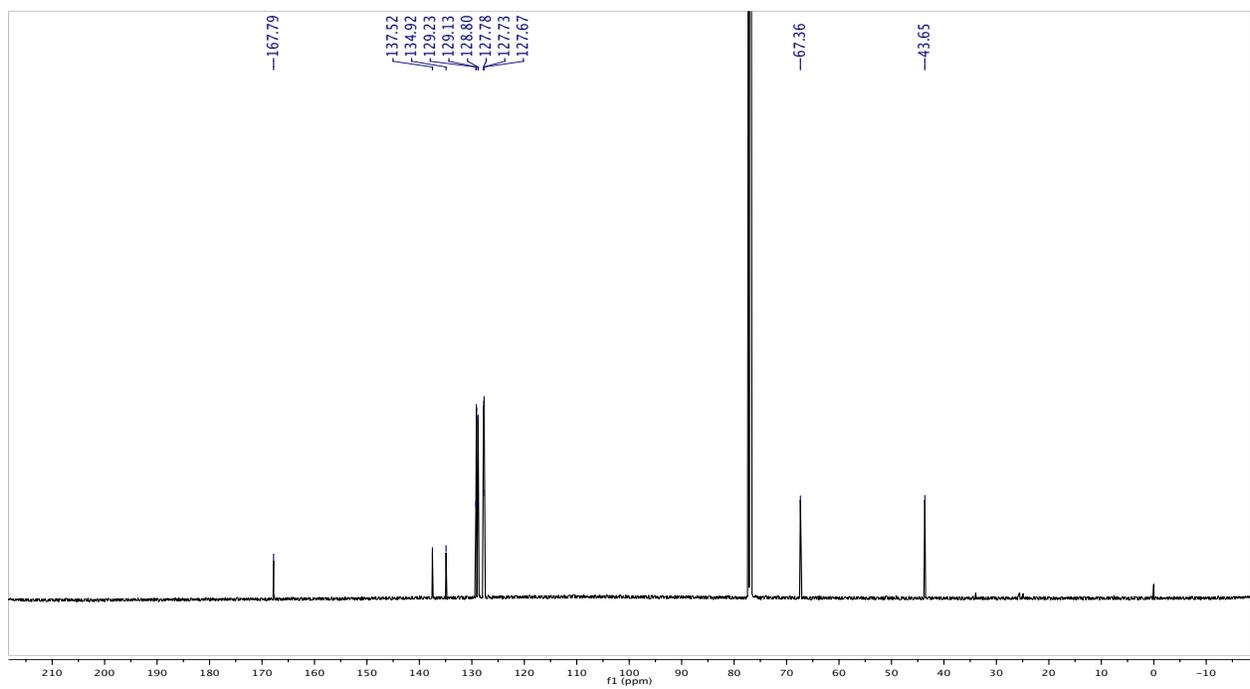
^{13}C NMR of S8 in CDCl_3 (125 MHz):



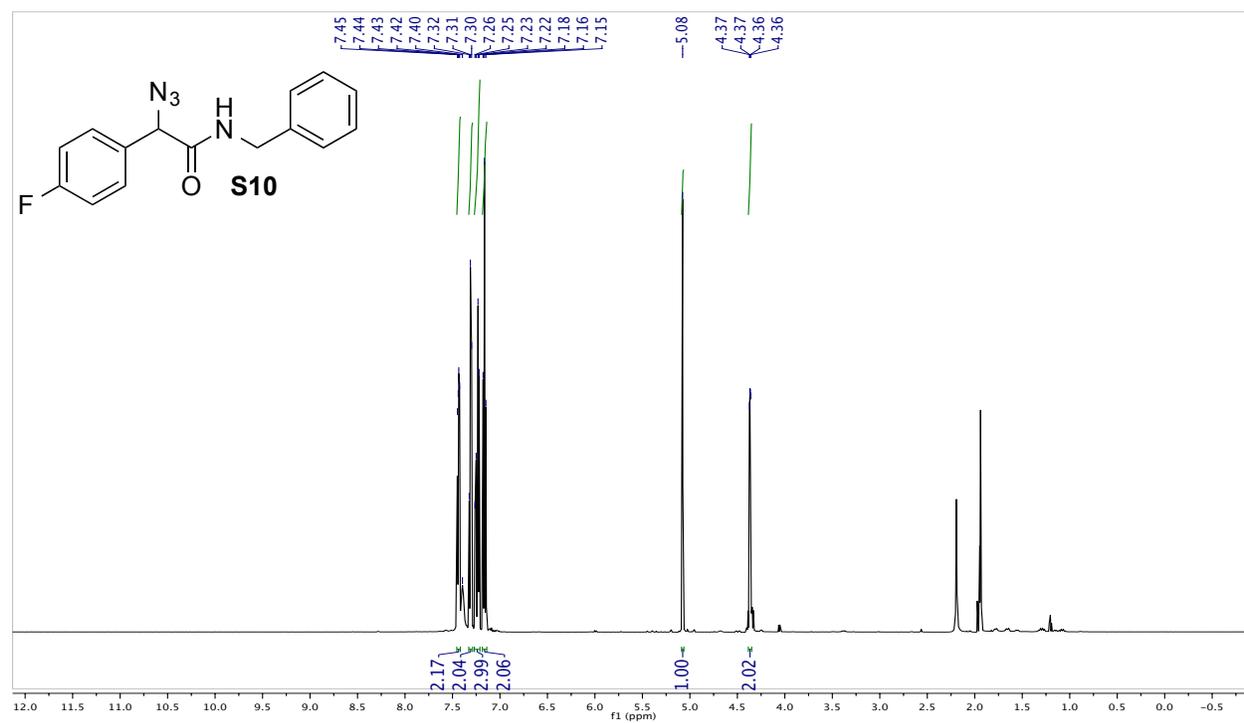
^1H NMR of S9 in CD_3CN (500 MHz):



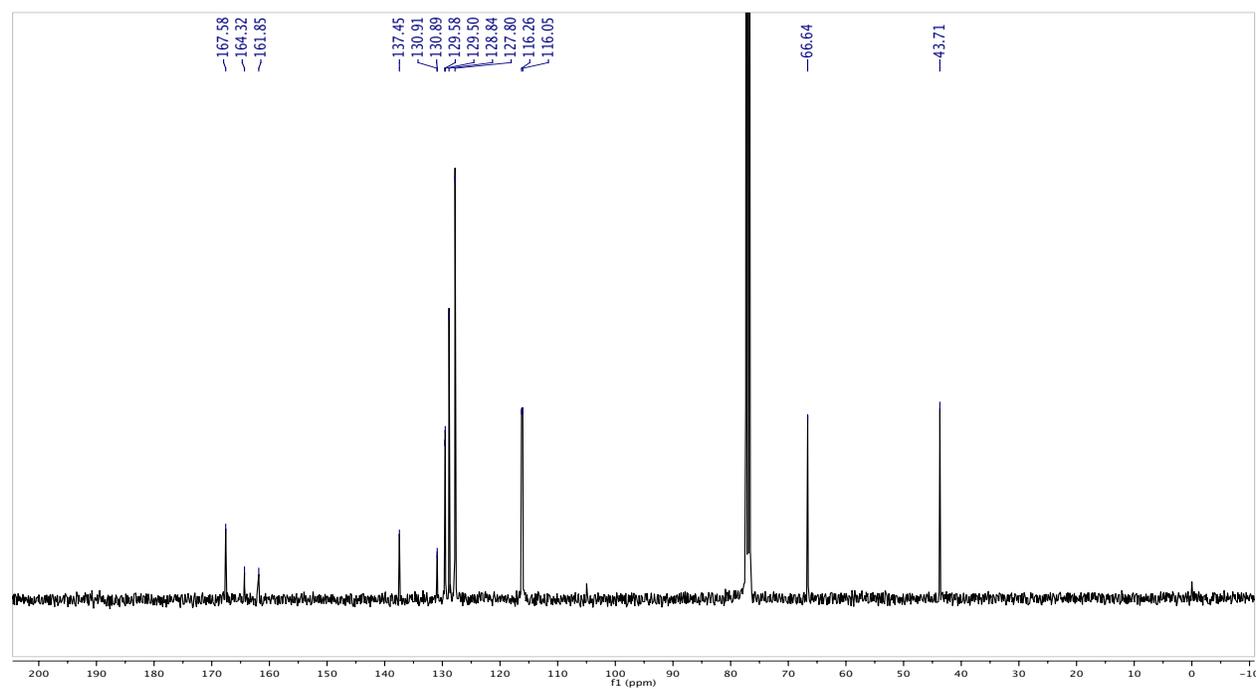
^{13}C NMR of S9 in CDCl_3 (125 MHz):



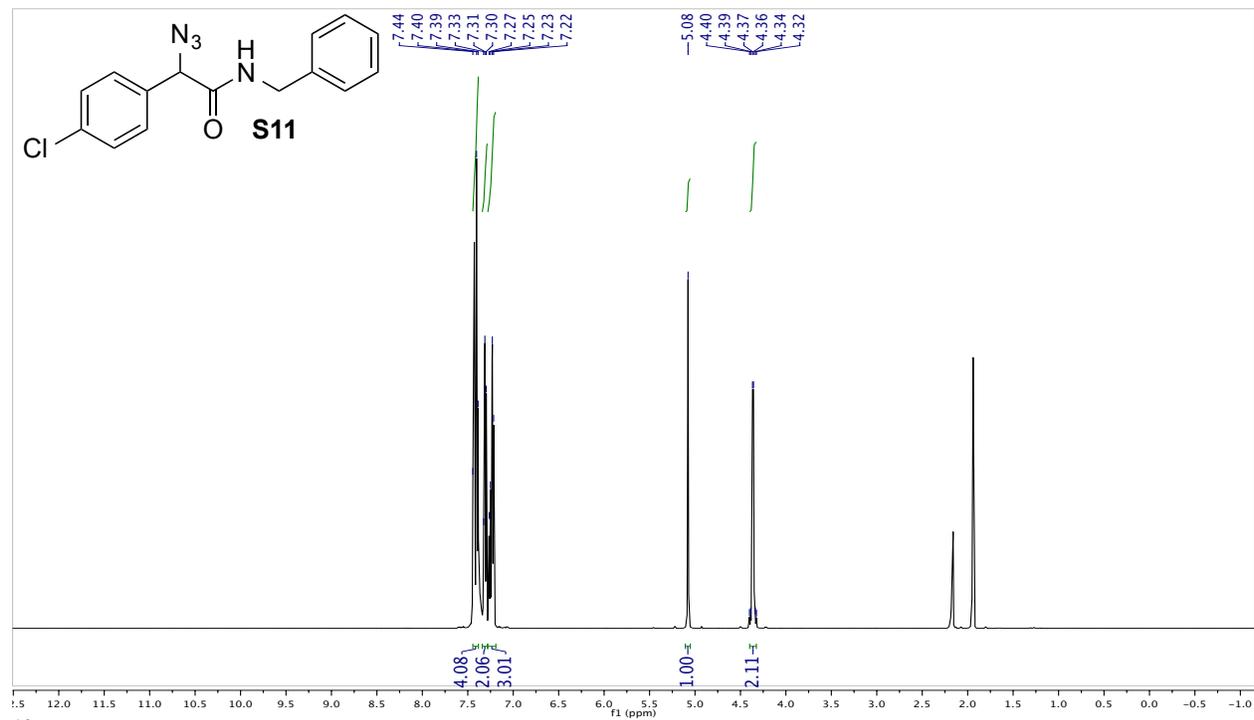
^1H NMR of S10 in CD_3CN (600 MHz):



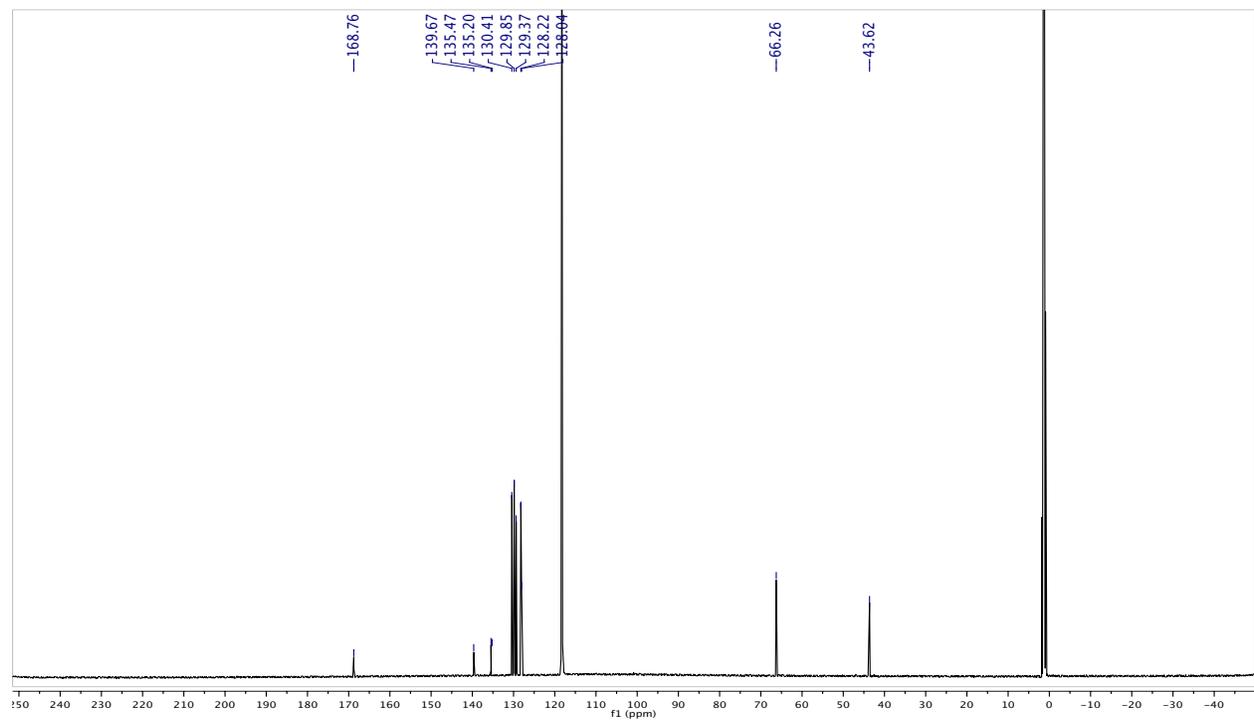
^{13}C NMR of S10 in CDCl_3 (100 MHz):



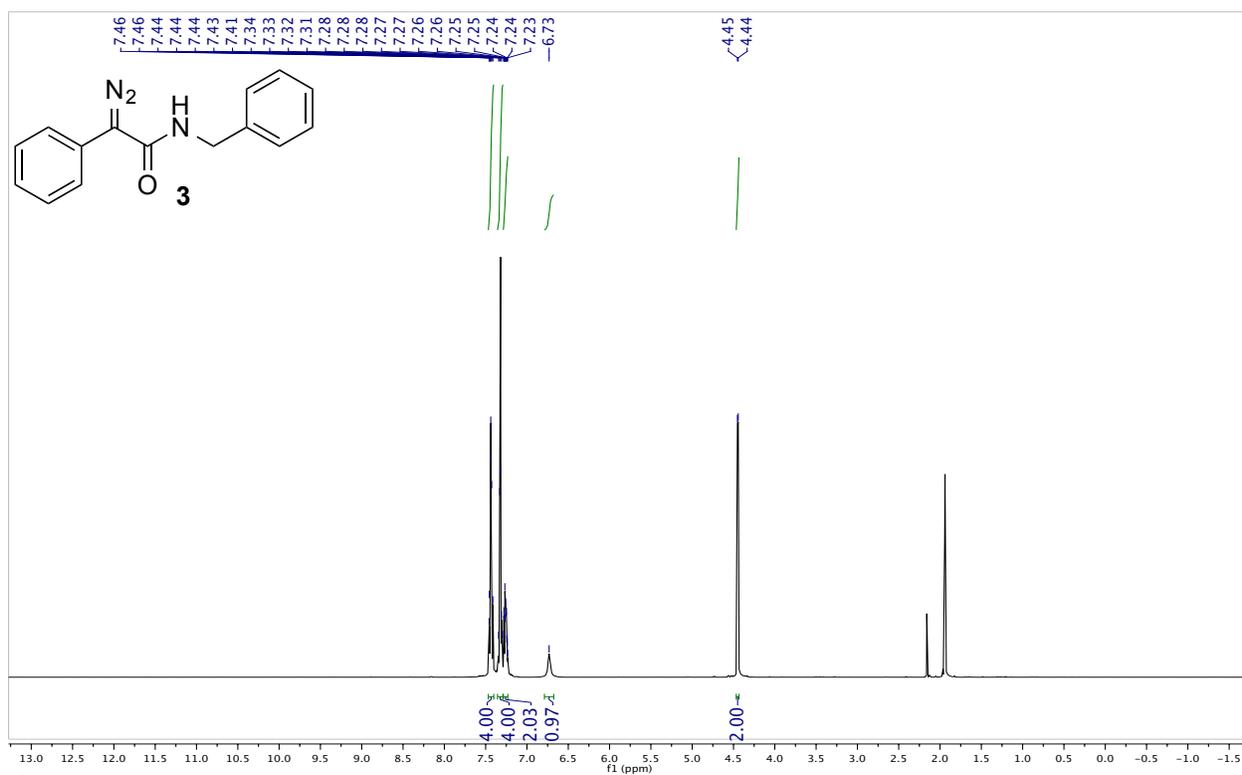
^1H NMR of S11 in CD_3CN (500 MHz):



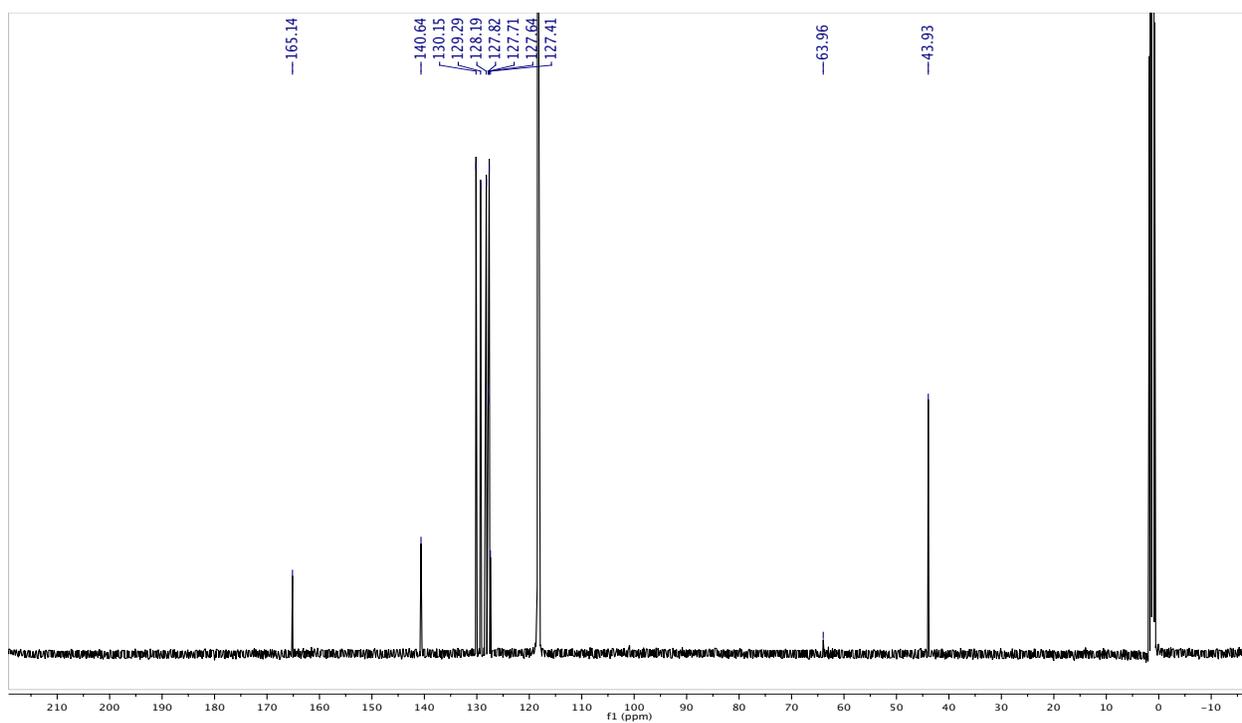
^{13}C NMR of S11 in CD_3CN (125 MHz):



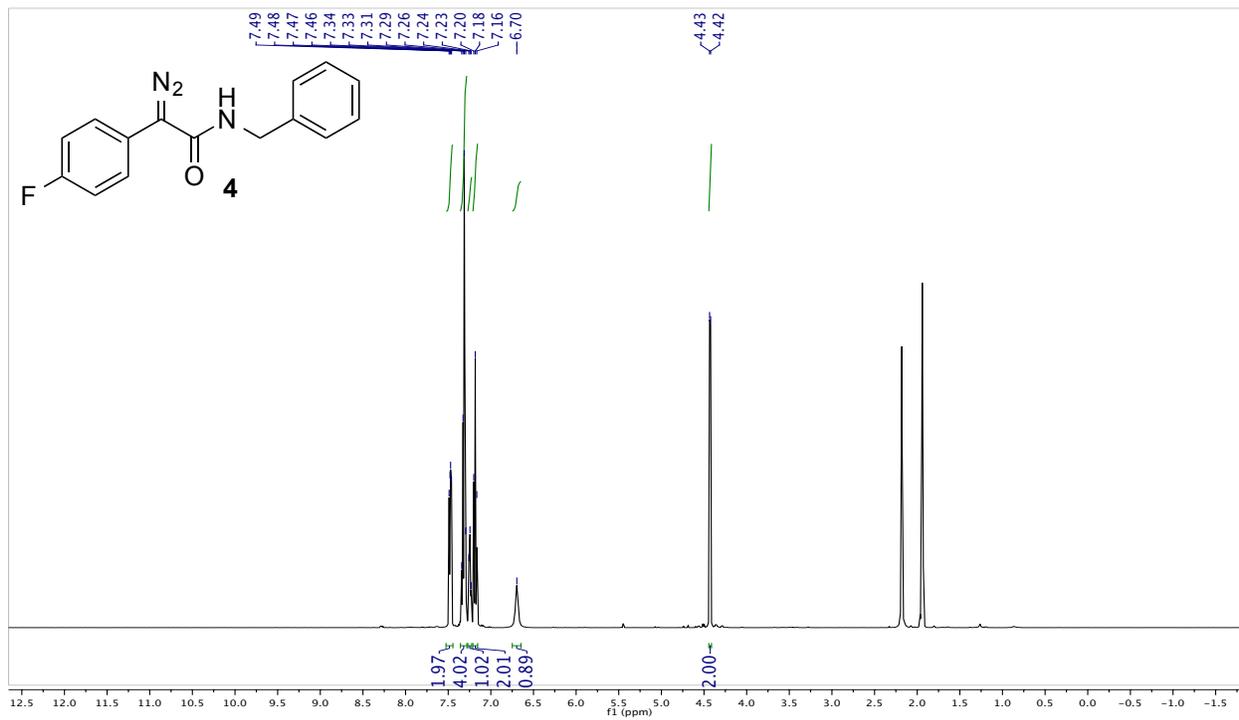
^1H NMR of 3 in CD_3CN (600 MHz):



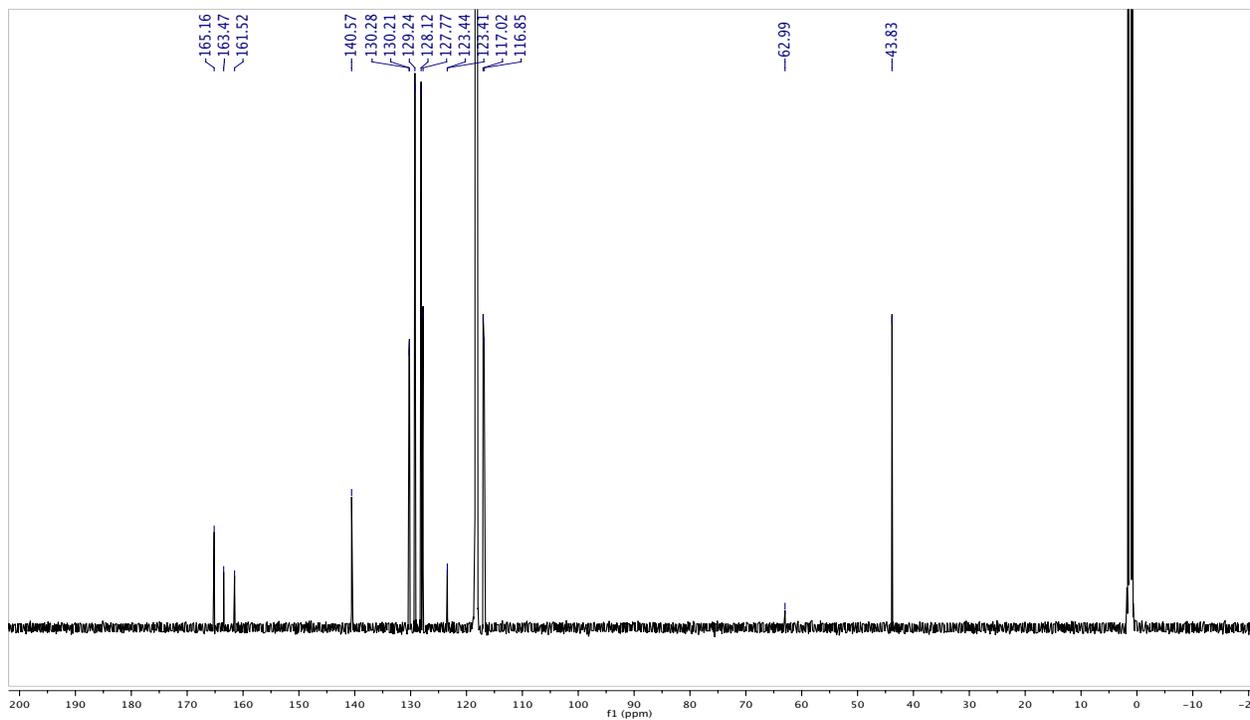
^{13}C NMR of 3 in CD_3CN (125 MHz):



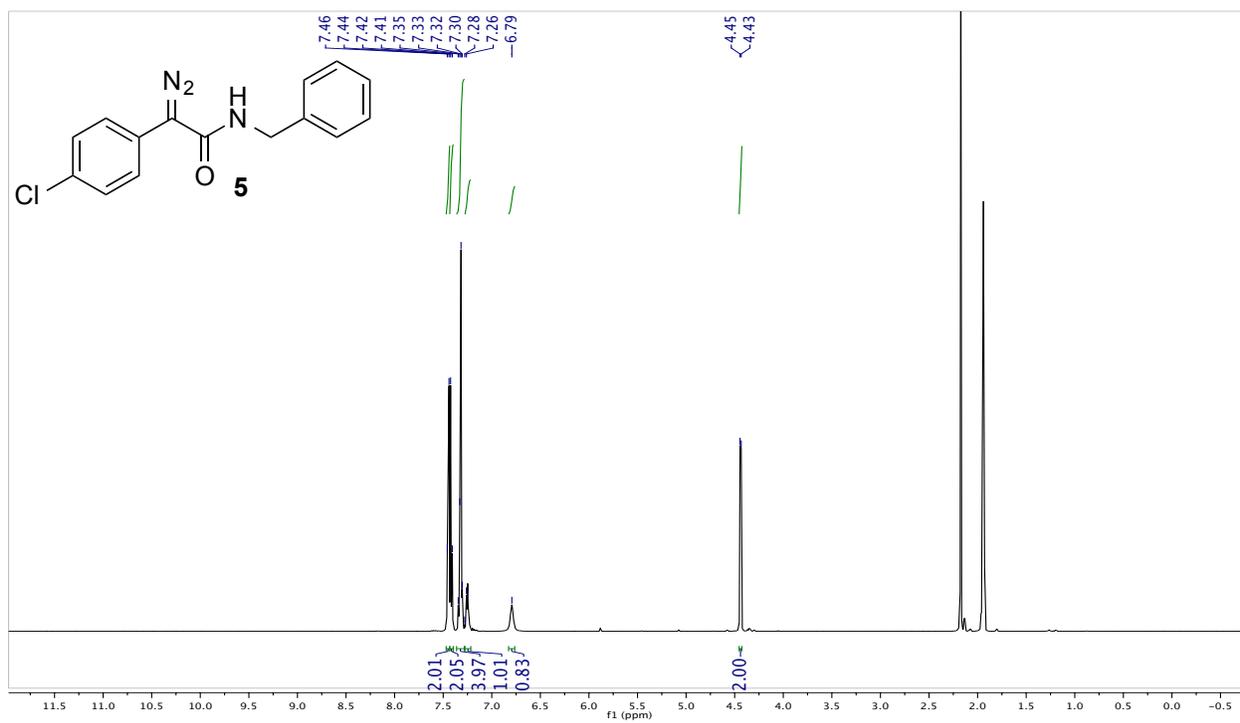
^1H NMR of 4 in CD_3CN (500 MHz):



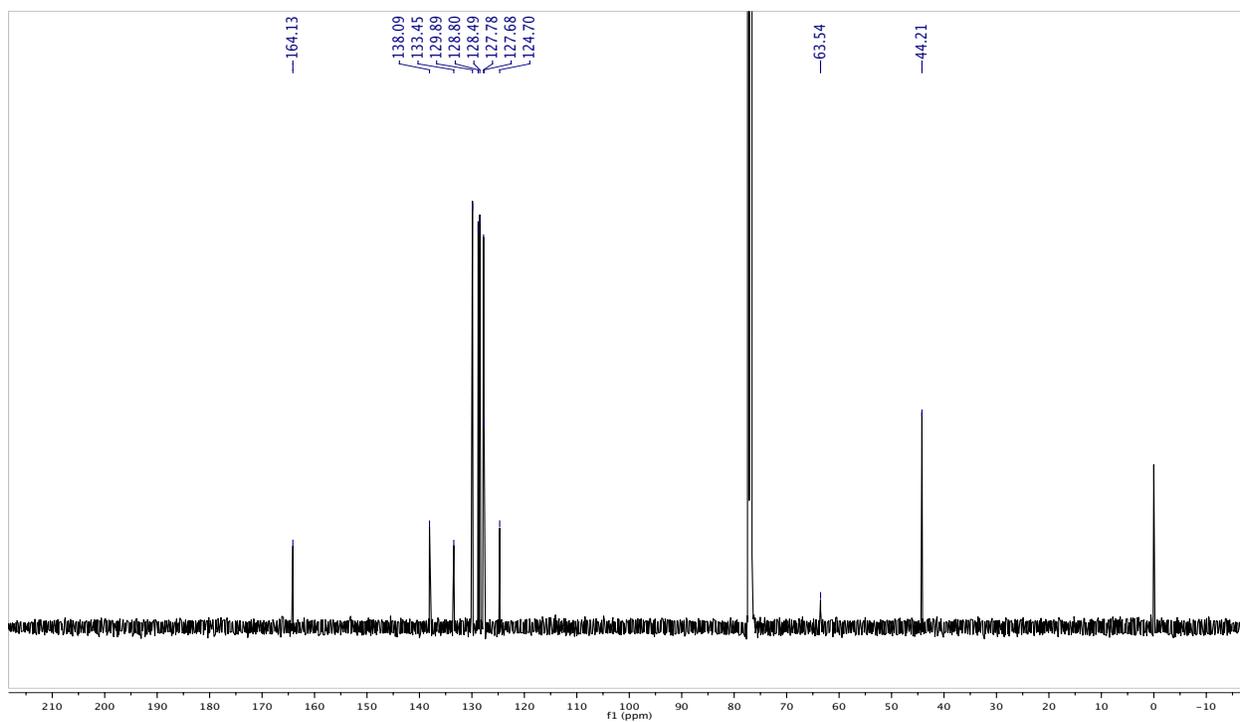
^{13}C NMR of 4 in CD_3CN (125 MHz):



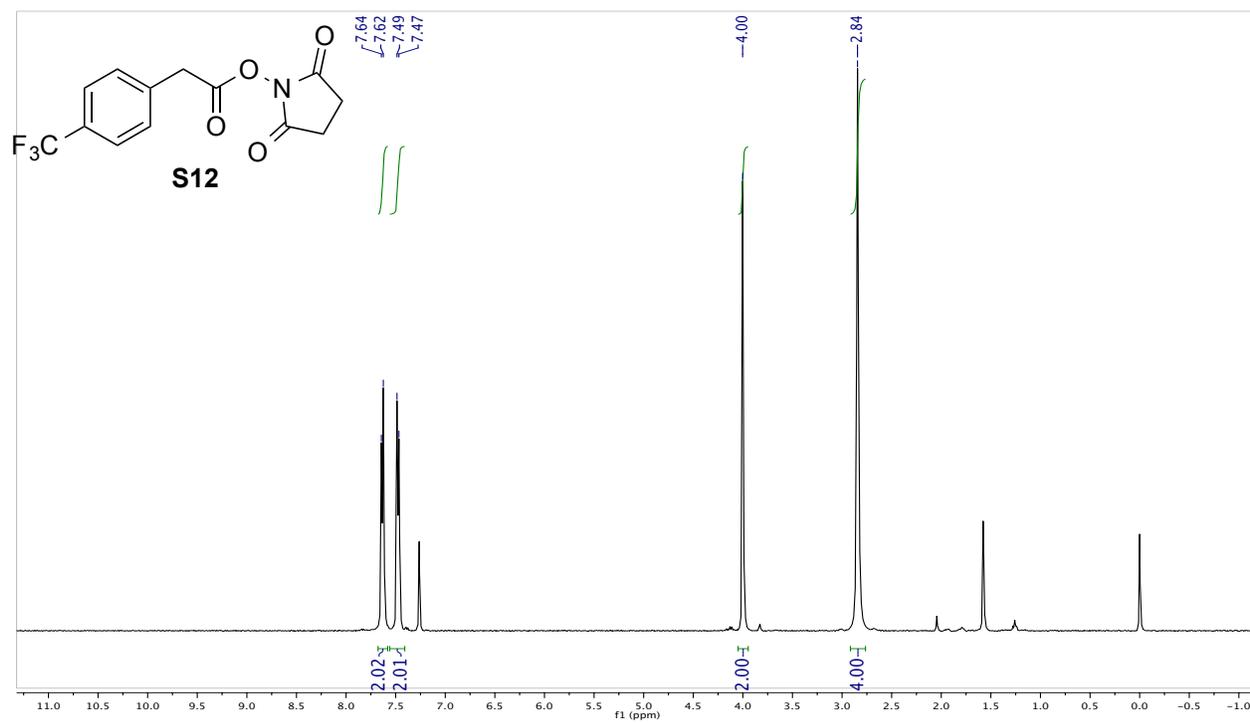
^1H NMR of 5 in CD_3CN (500 MHz):



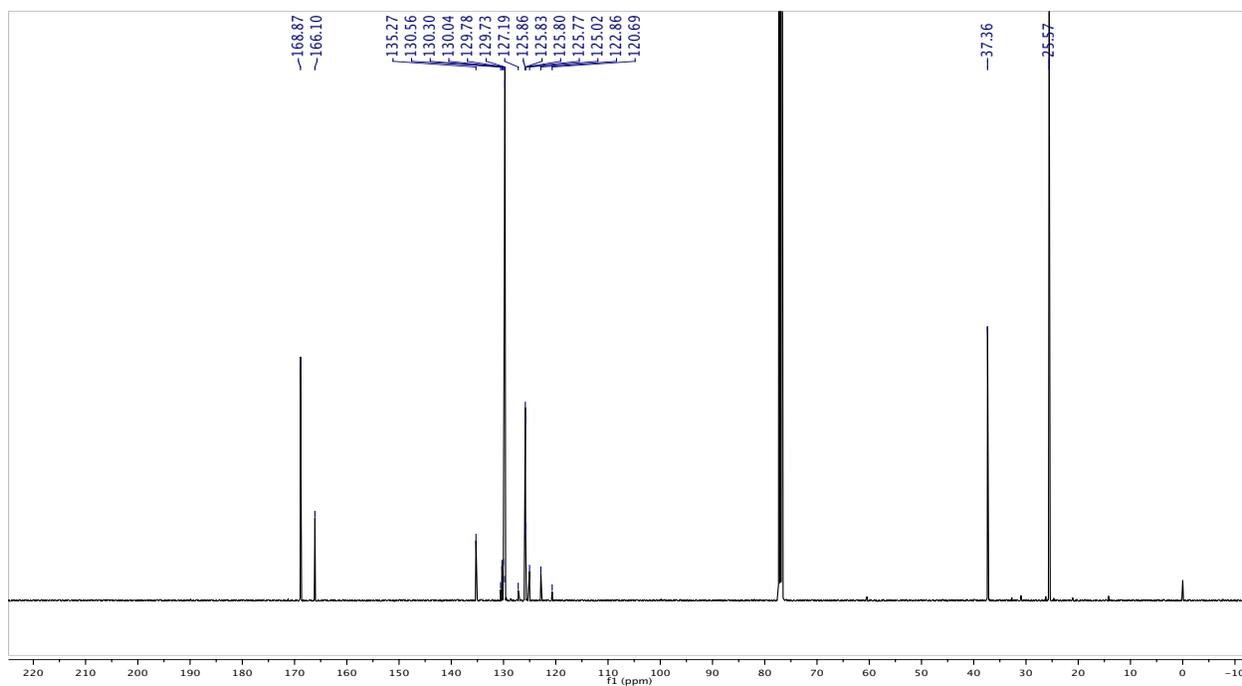
^{13}C NMR of 5 in CDCl_3 (125 MHz):



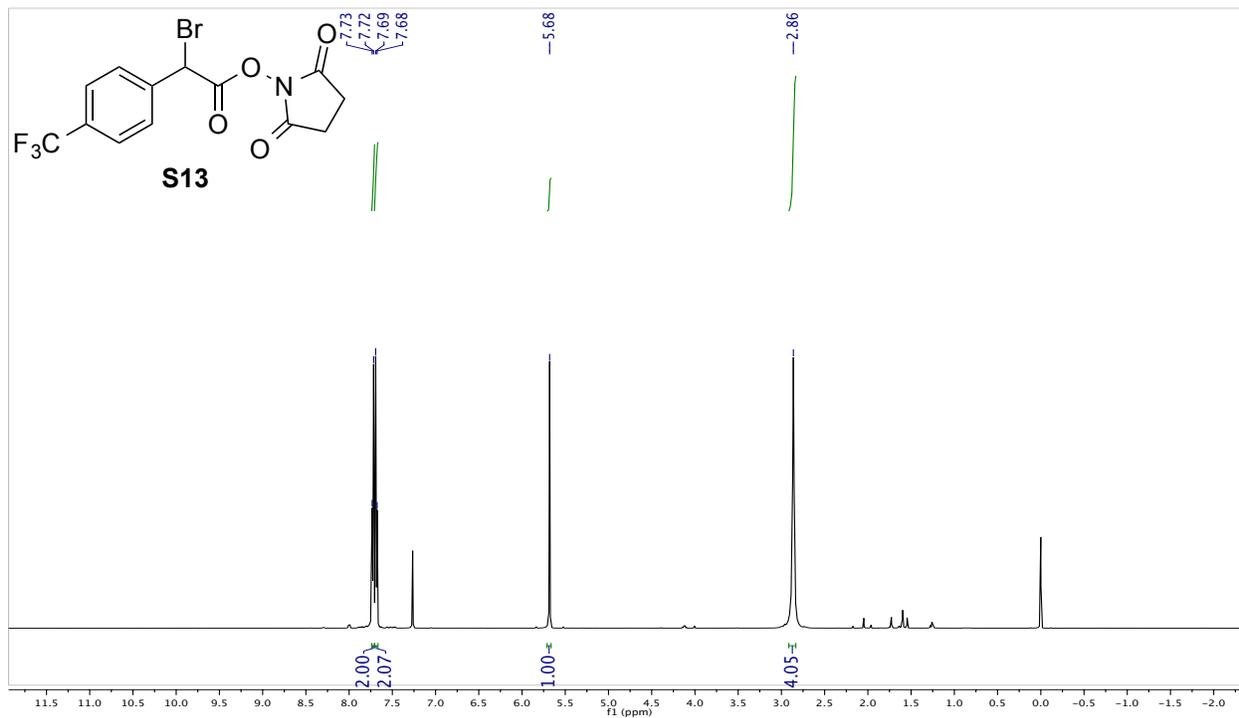
^1H NMR of S12 in CDCl_3 (400 MHz):



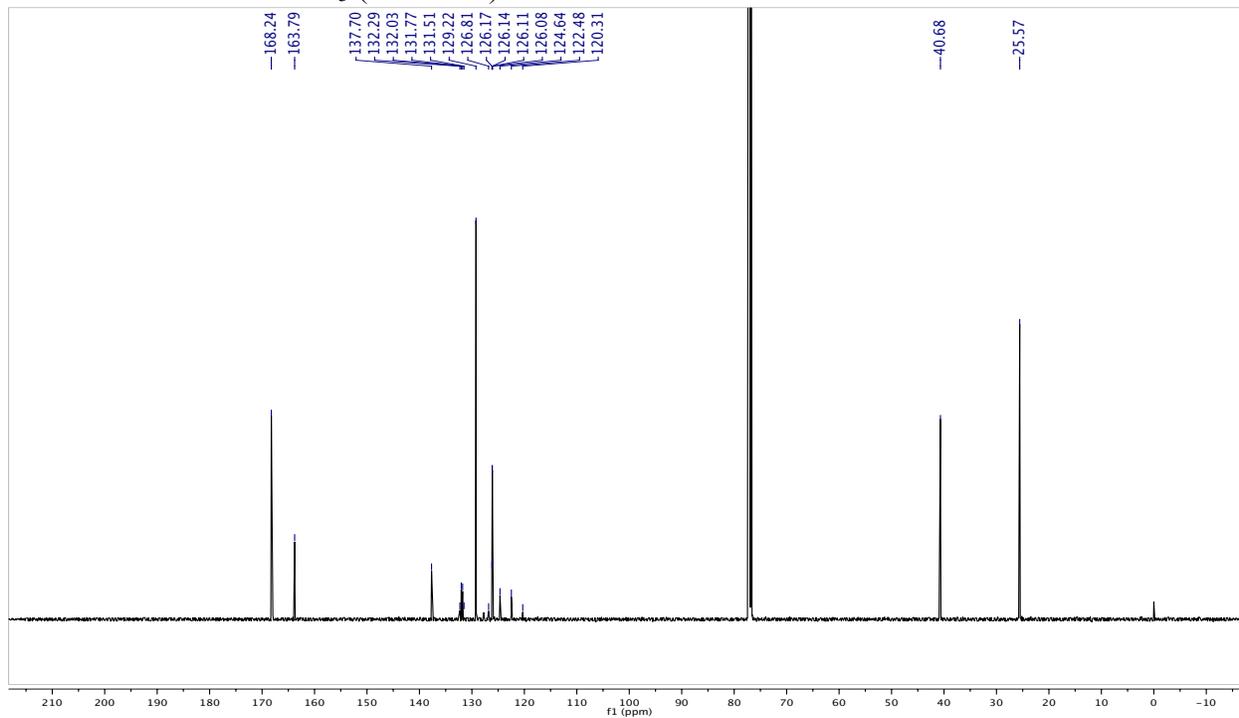
^{13}C NMR of S12 in CDCl_3 (125 MHz):



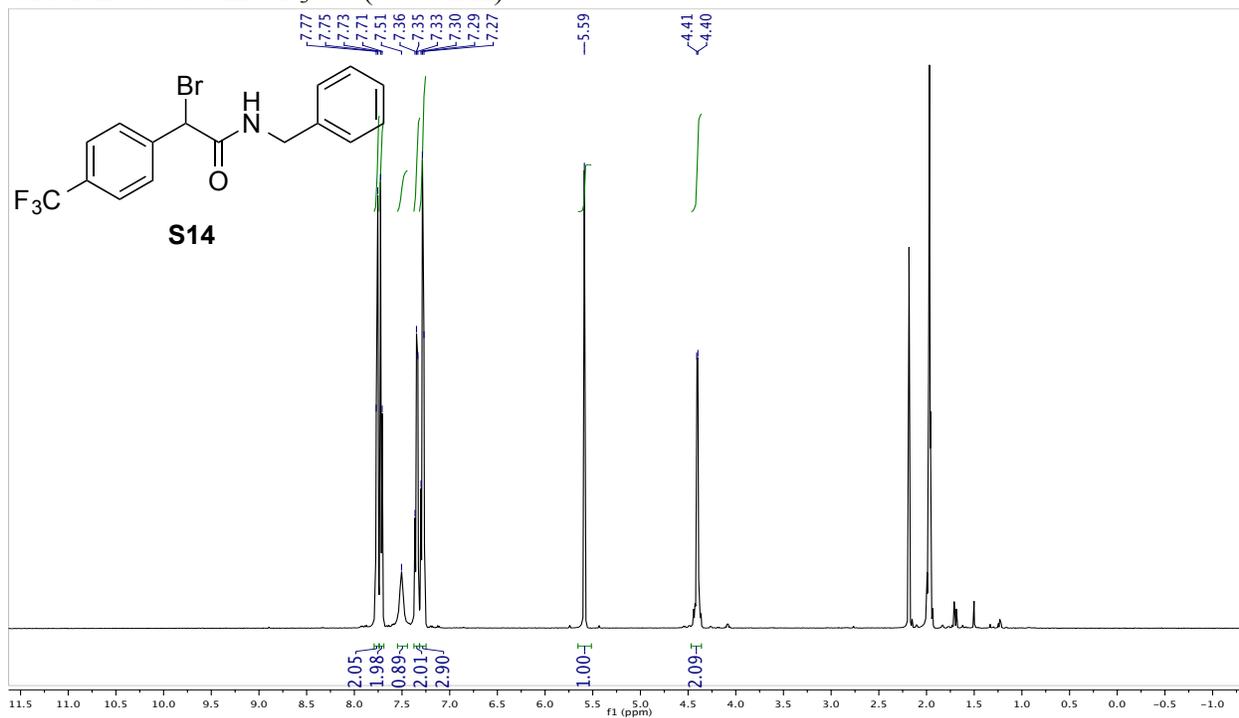
^1H NMR of S13 in CDCl_3 (500 MHz):



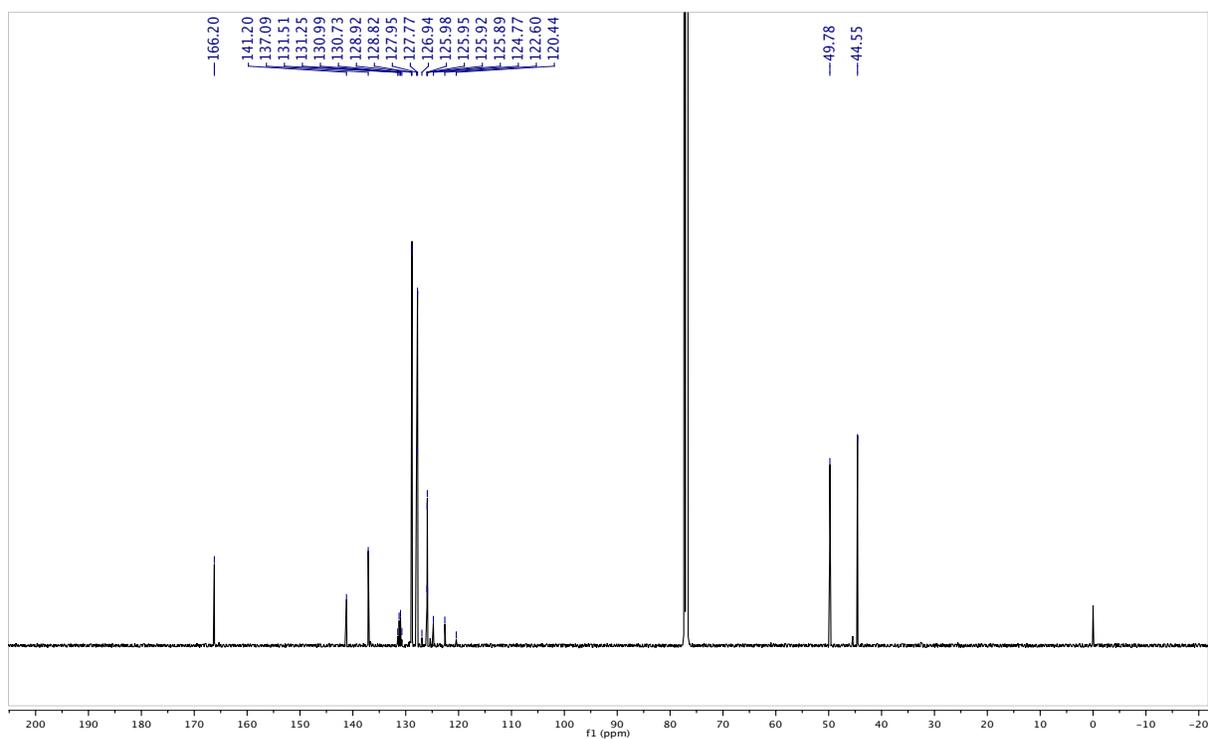
^{13}C NMR of S13 in CDCl_3 (125 MHz):



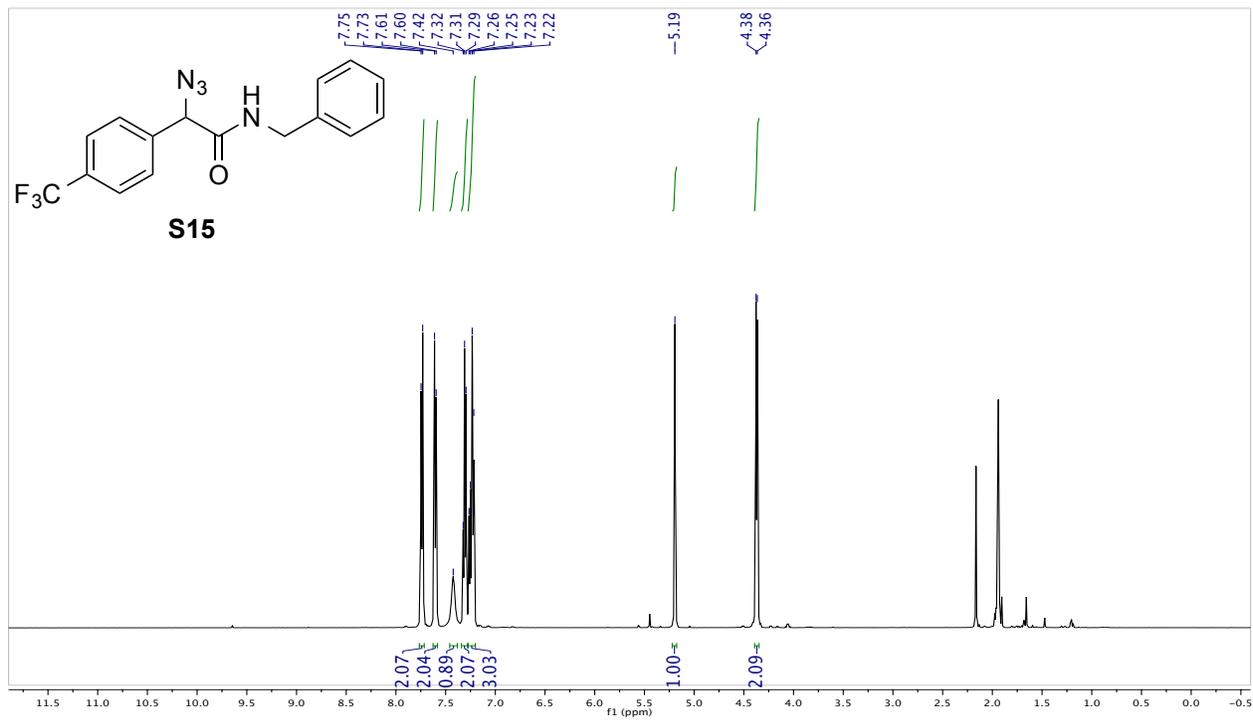
^1H NMR of S14 in CD_3CN (500 MHz):



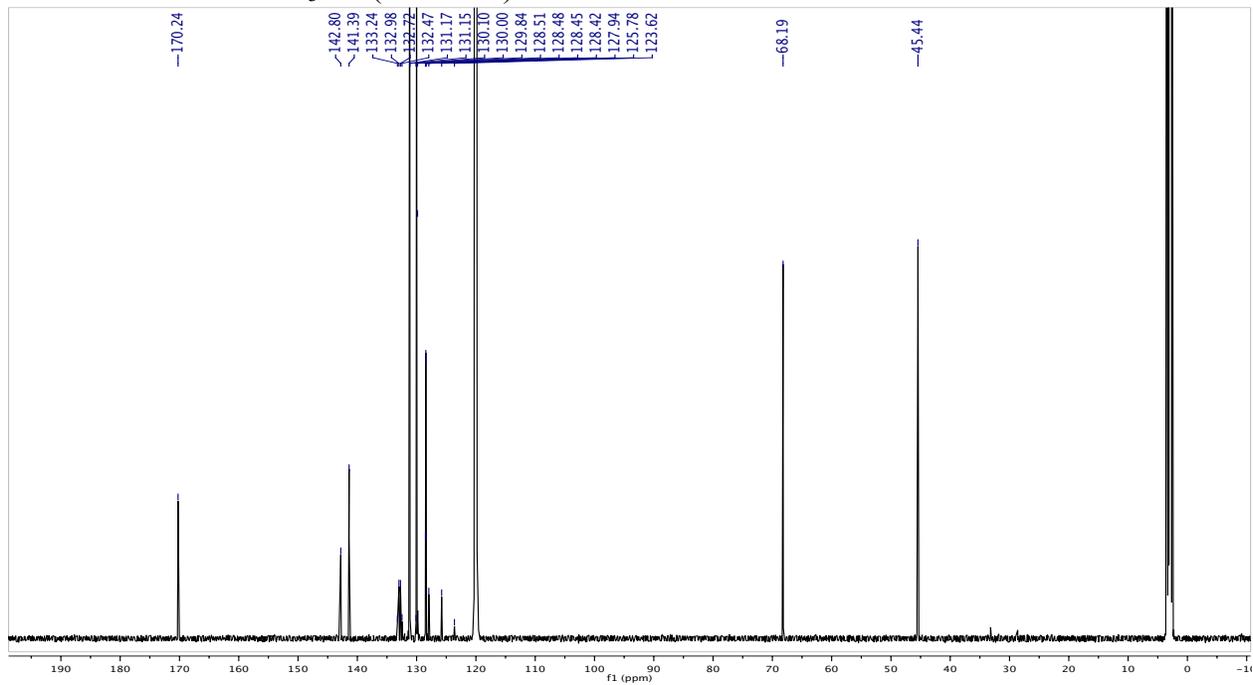
^{13}C NMR of S14 in CDCl_3 (125 MHz):



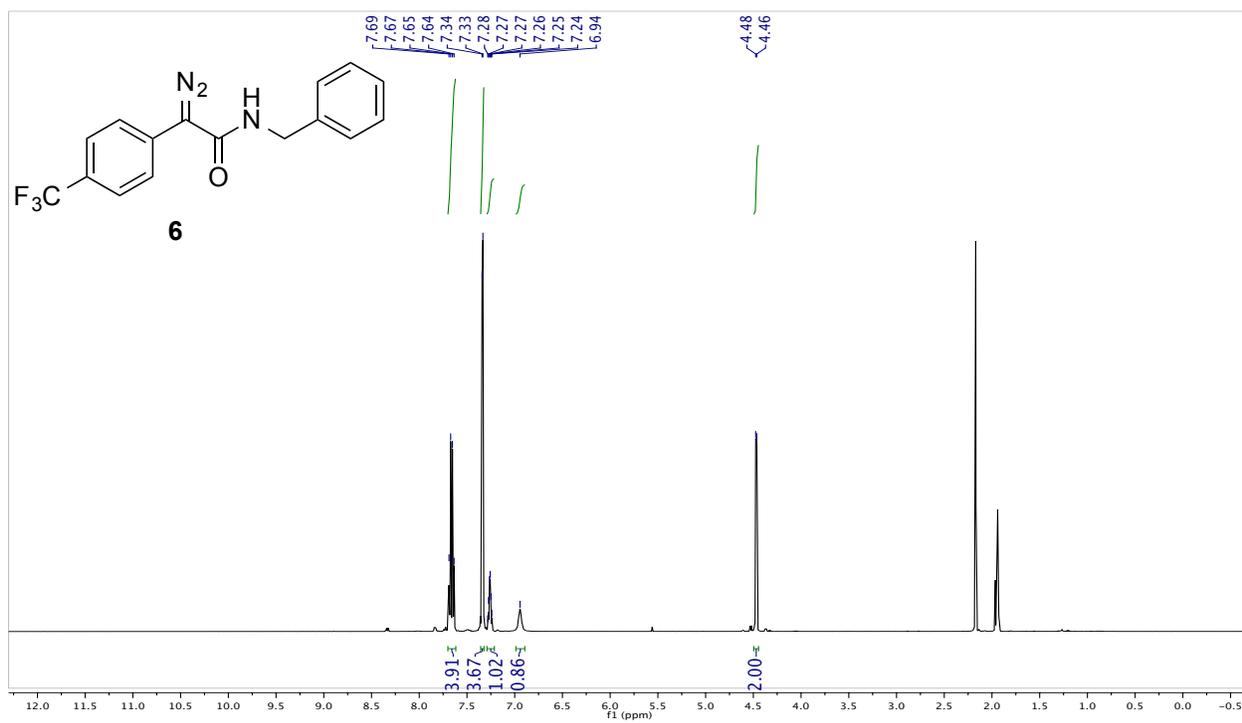
^1H NMR of S15 in CD_3CN (500 MHz):



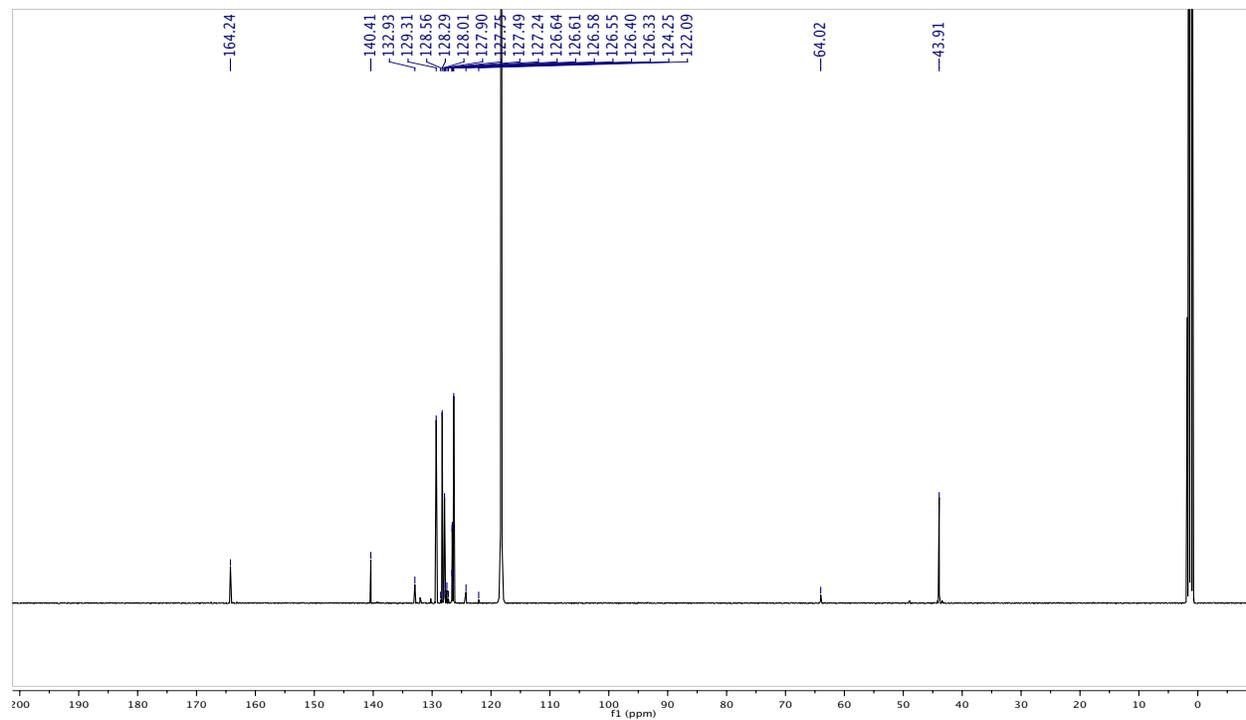
^{13}C NMR of S15 in CD_3CN (125 MHz):



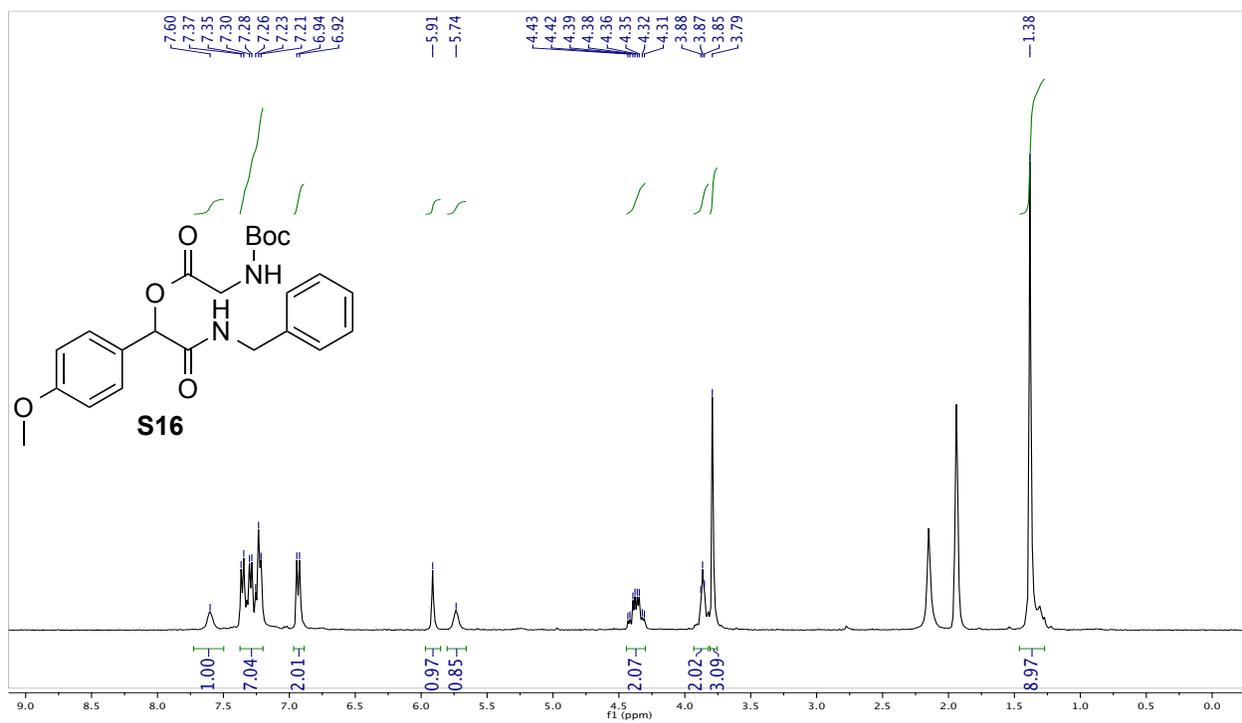
^1H NMR of 6 in CDCl_3 (400 MHz):



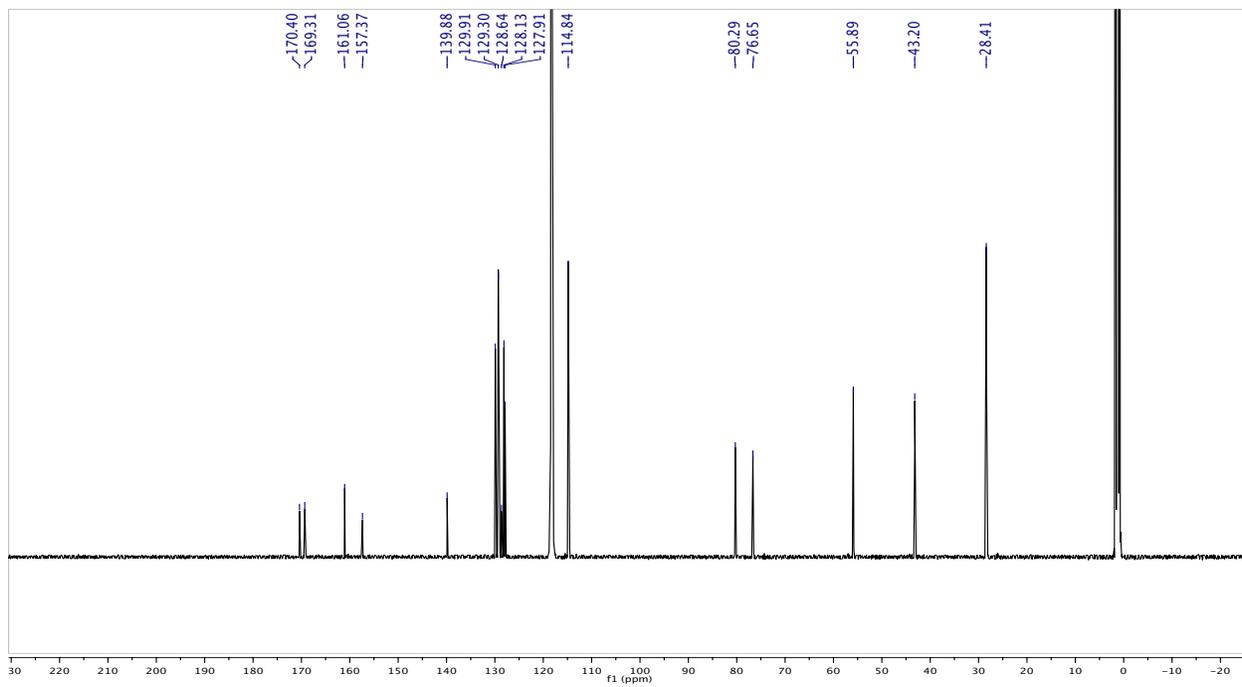
^{13}C NMR of 6 in CD_3CN (125 MHz):



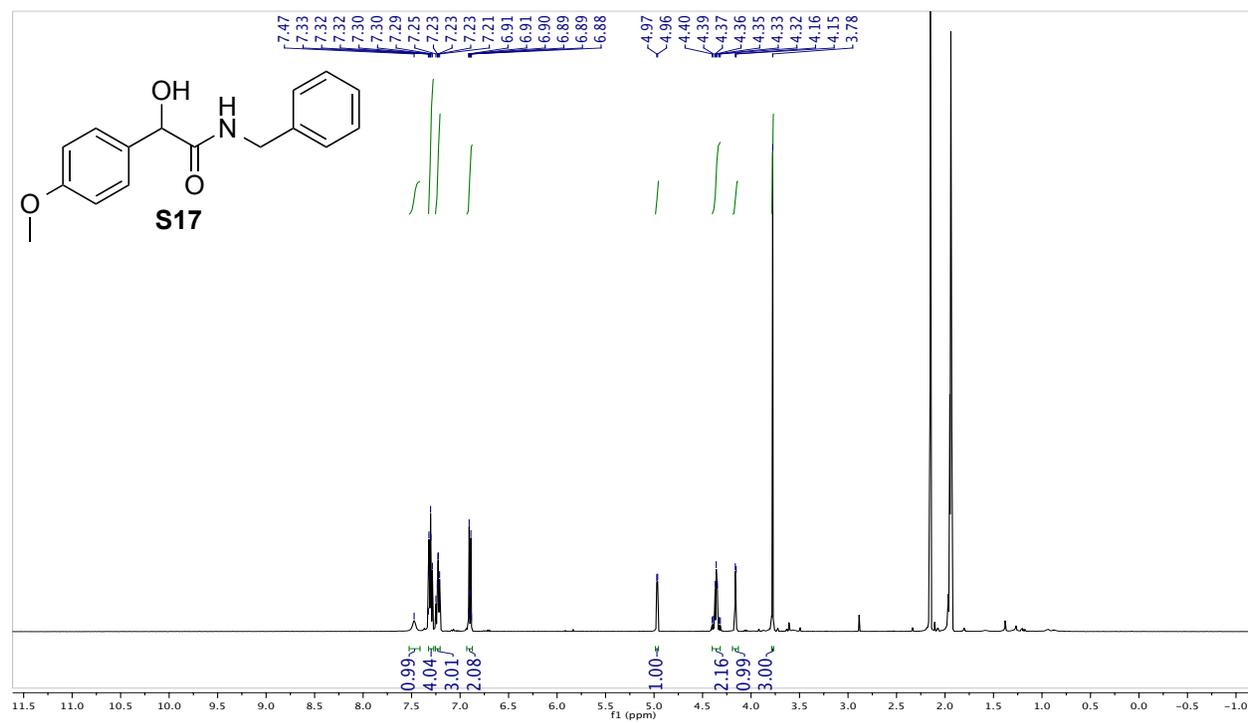
^1H NMR of S16 in CD_3CN (400 MHz):



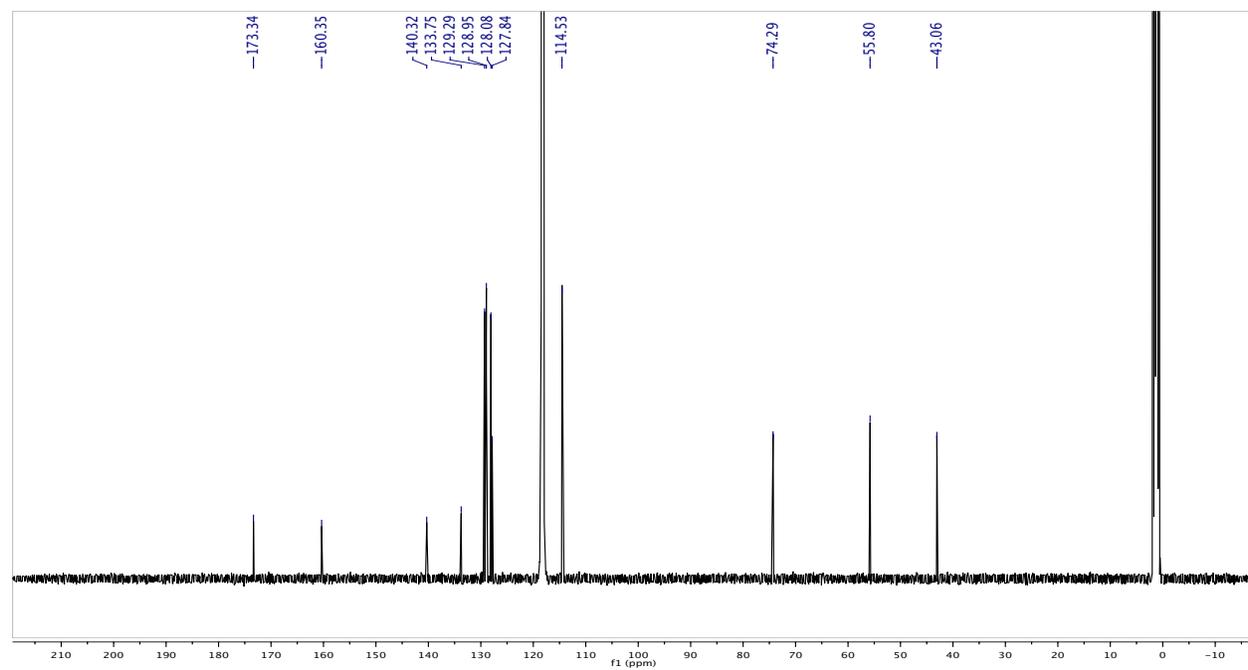
^{13}C NMR of S16 in CD_3CN (100 MHz):



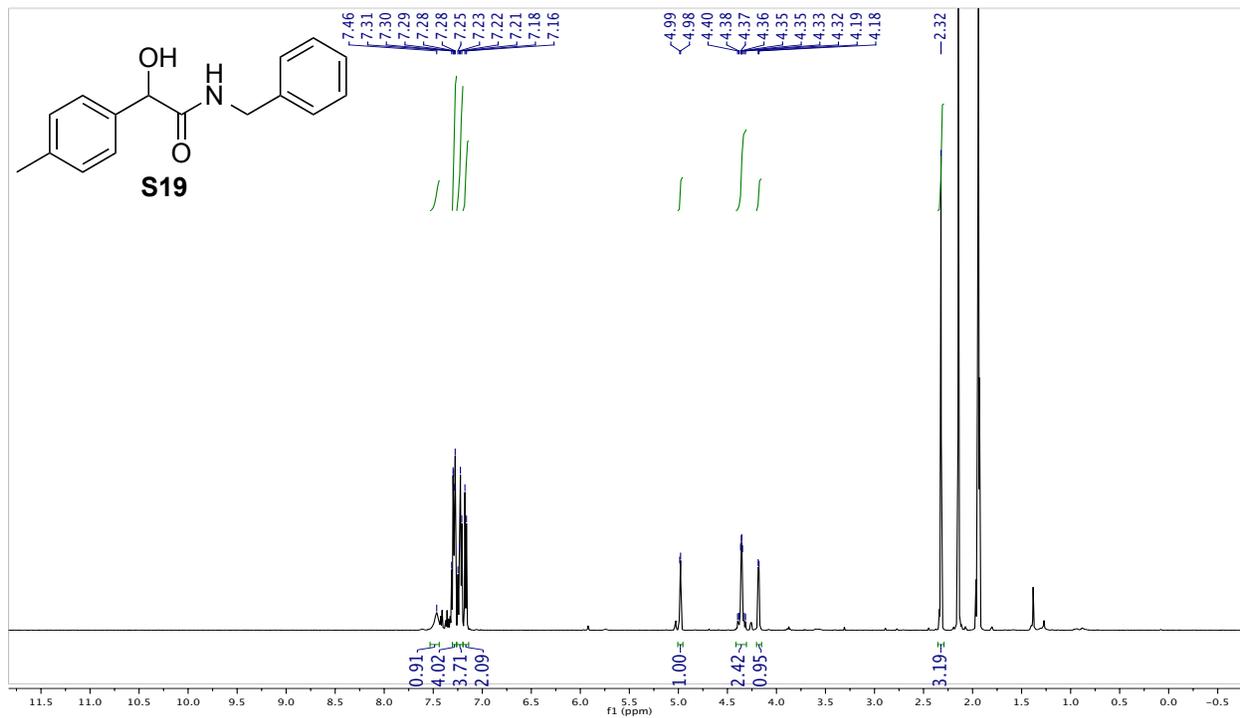
^1H NMR of S17 in CD_3CN (500 MHz):



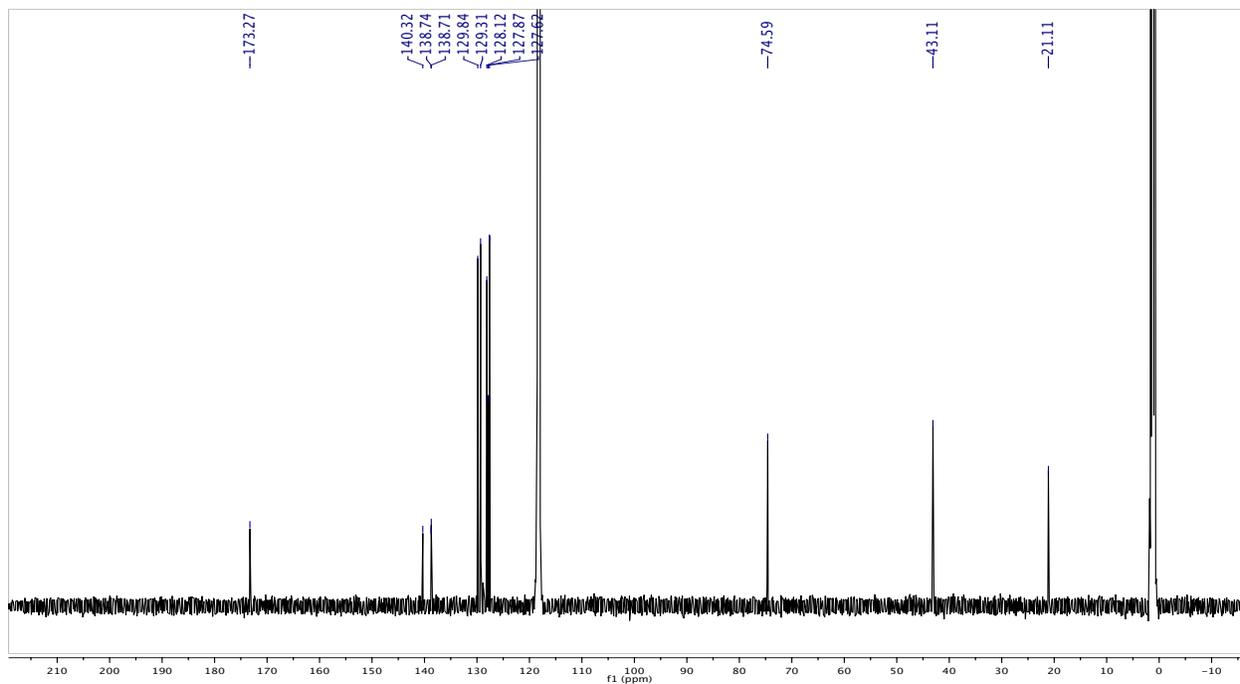
^{13}C NMR of S17 in CD_3CN (125 MHz):



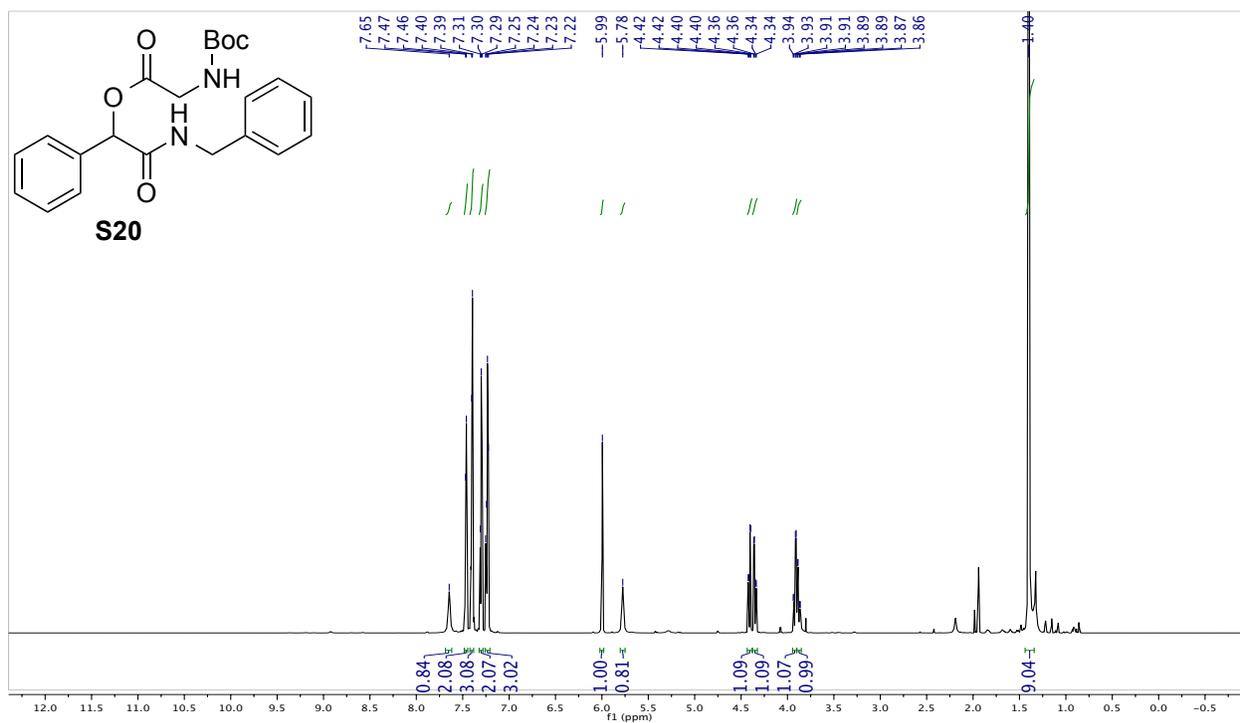
^1H NMR of S19 in CD_3CN (500 MHz):



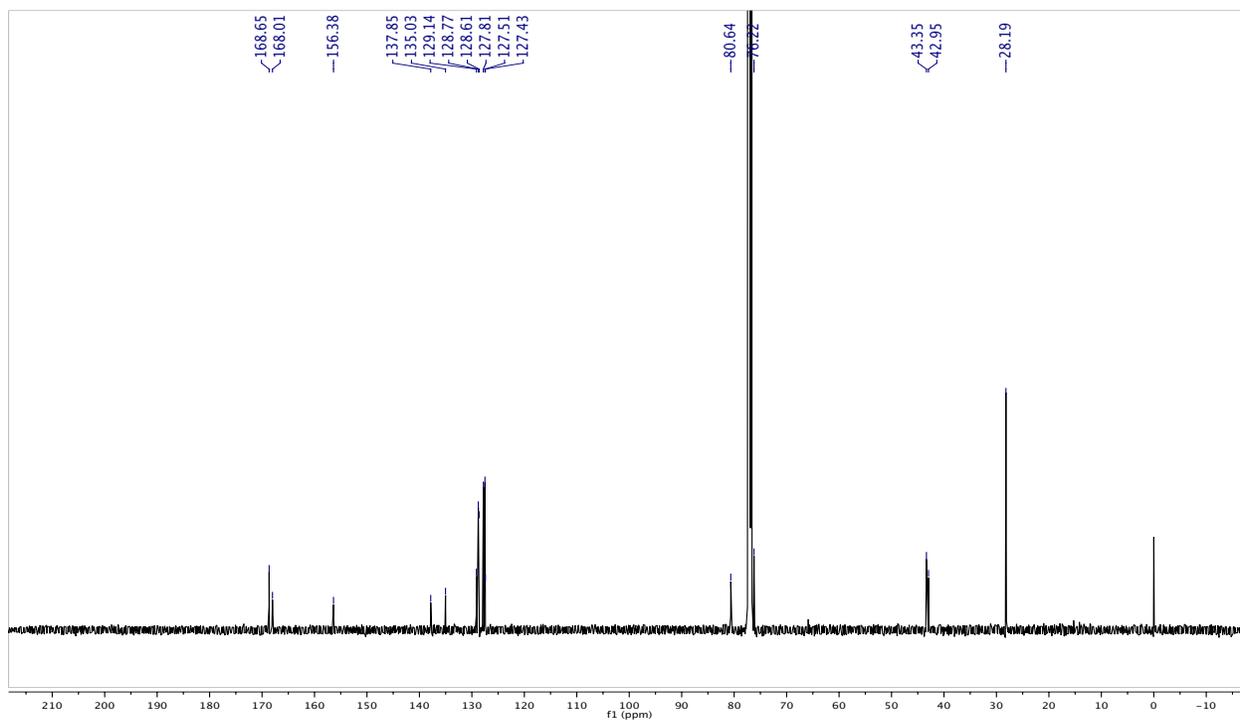
^{13}C NMR of S19 in CD_3CN (125 MHz):



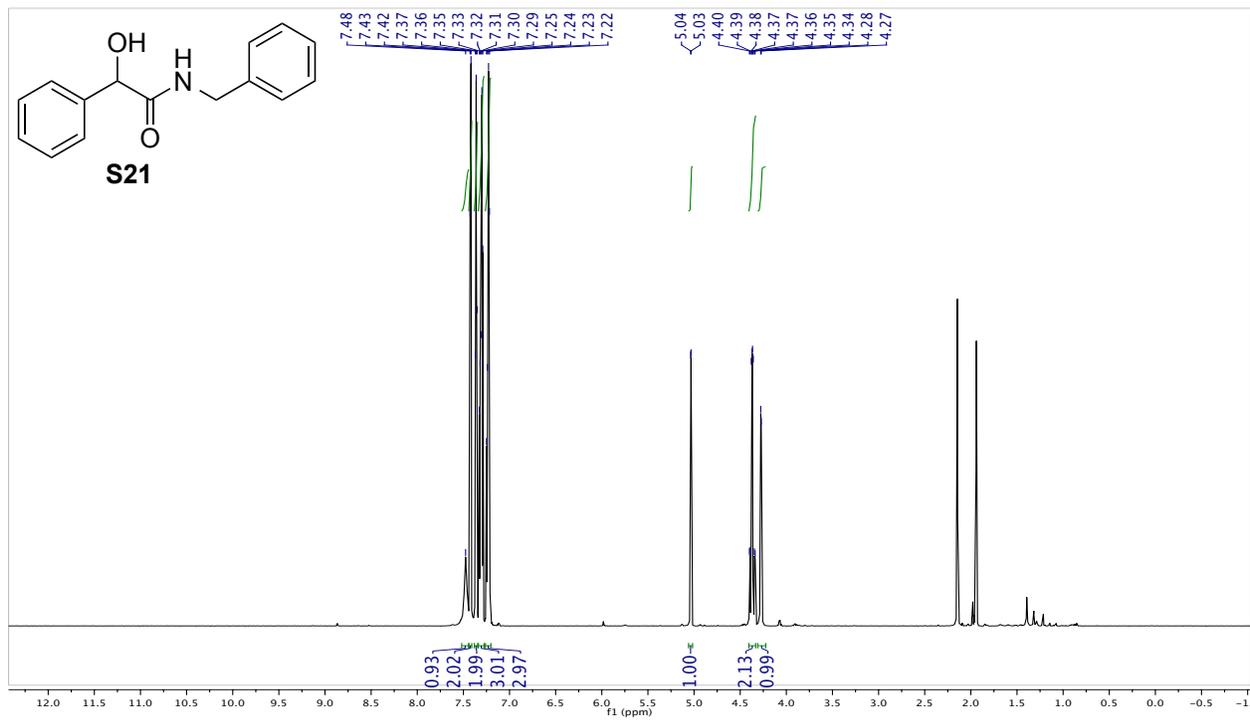
^1H NMR of S20 in CD_3CN (750 MHz):



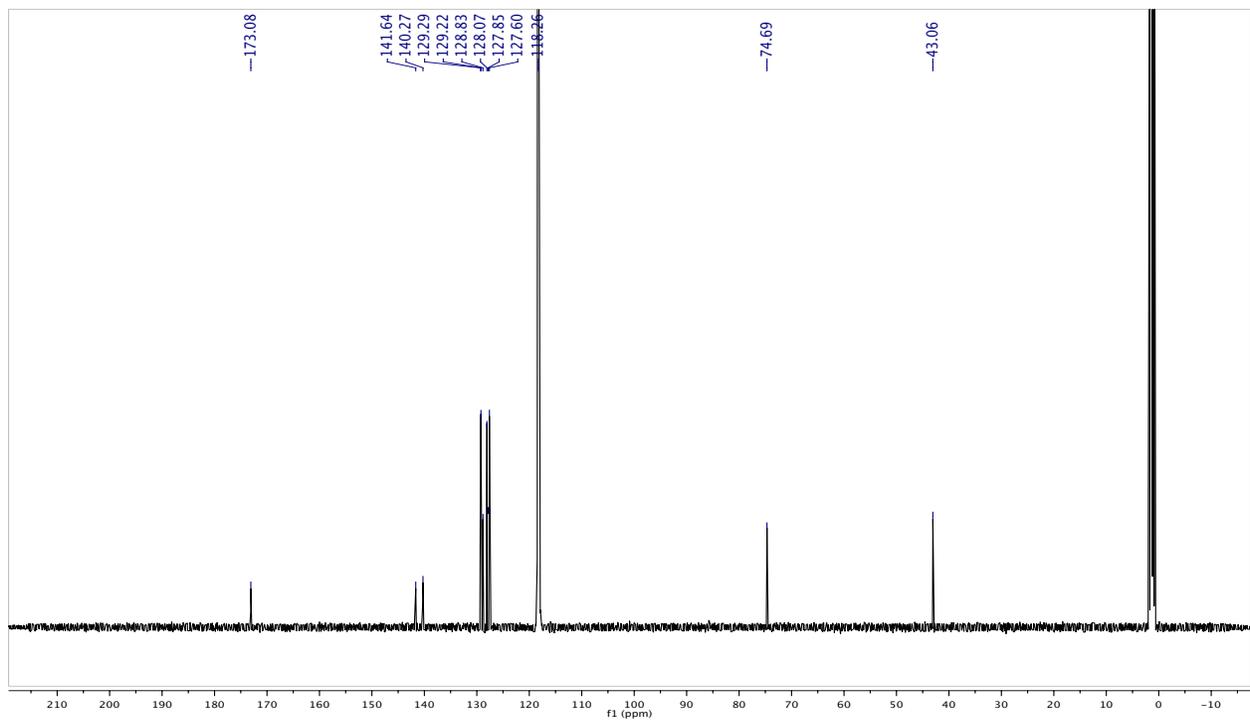
^{13}C NMR of S20 in CDCl_3 (125 MHz):



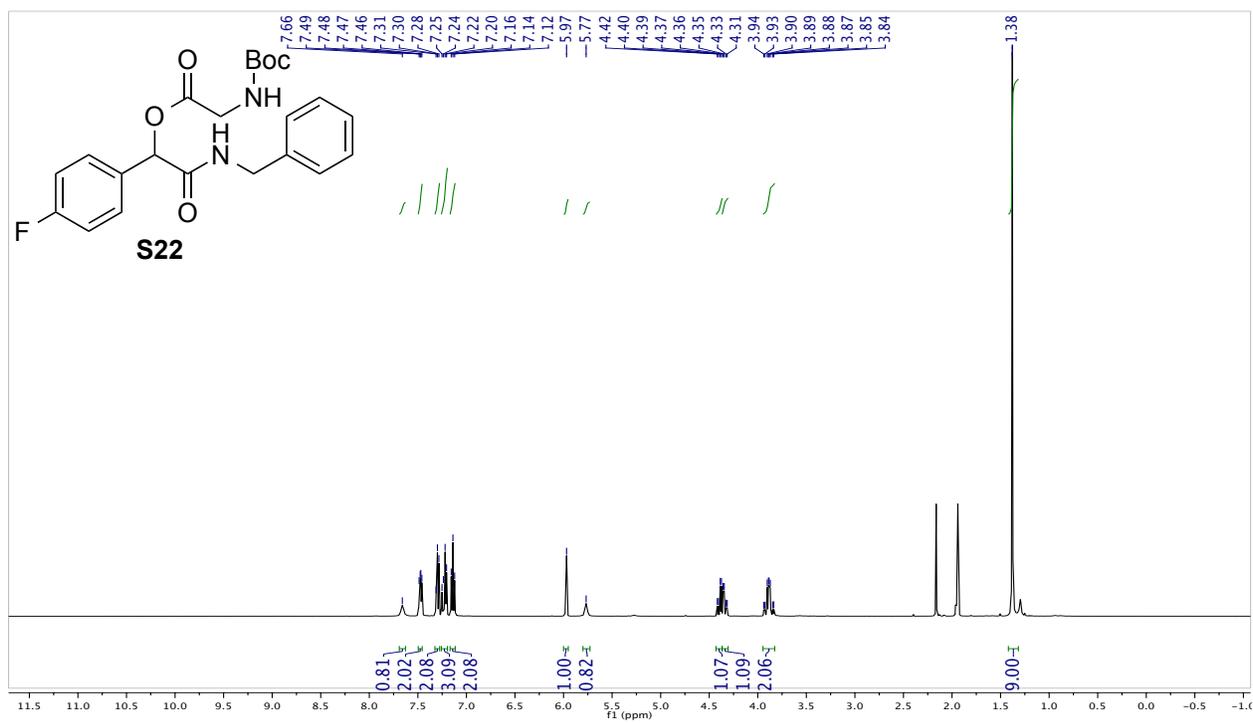
^1H NMR of S21 in CD_3CN (750 MHz):



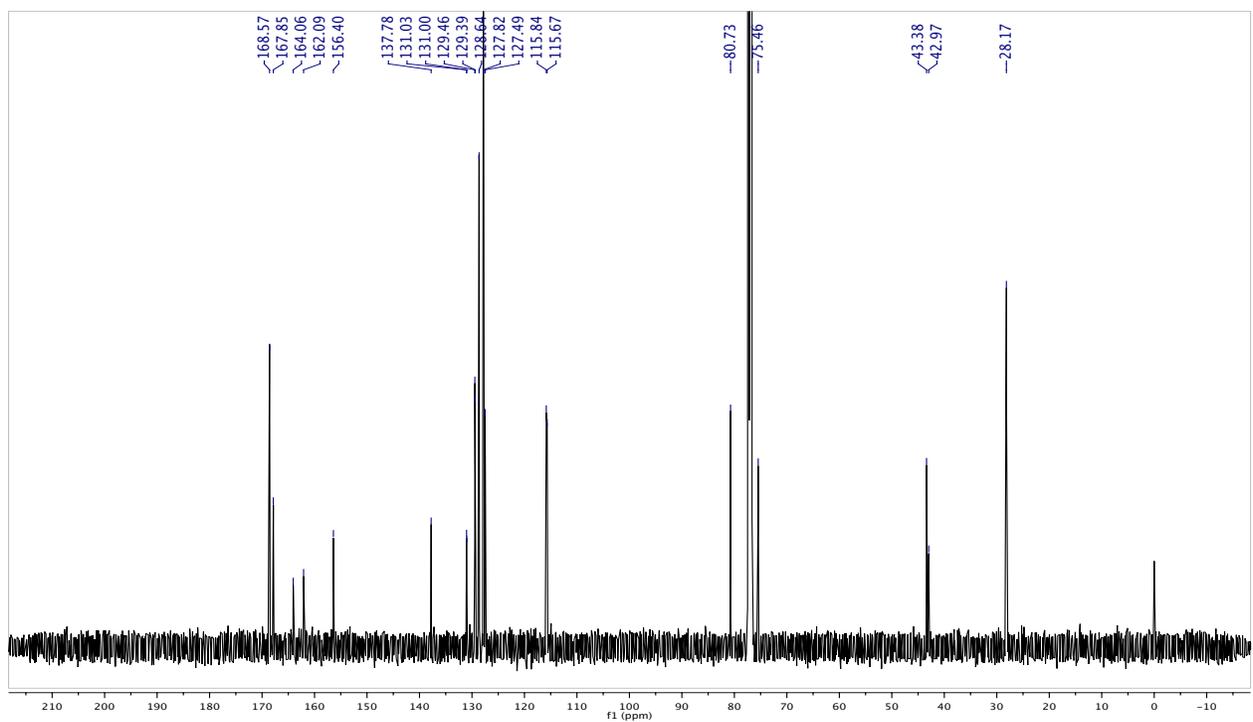
^{13}C NMR of S21 in CD_3CN (125 MHz):



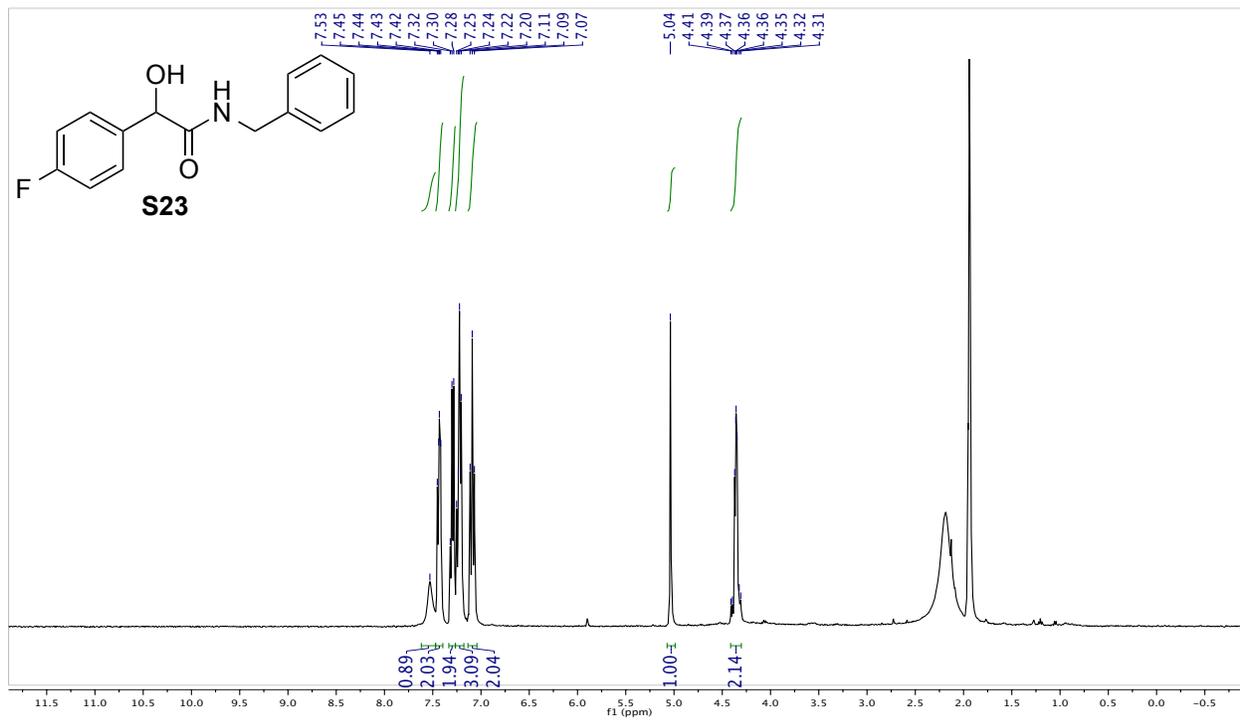
^1H NMR of S22 in CD_3CN (500 MHz):



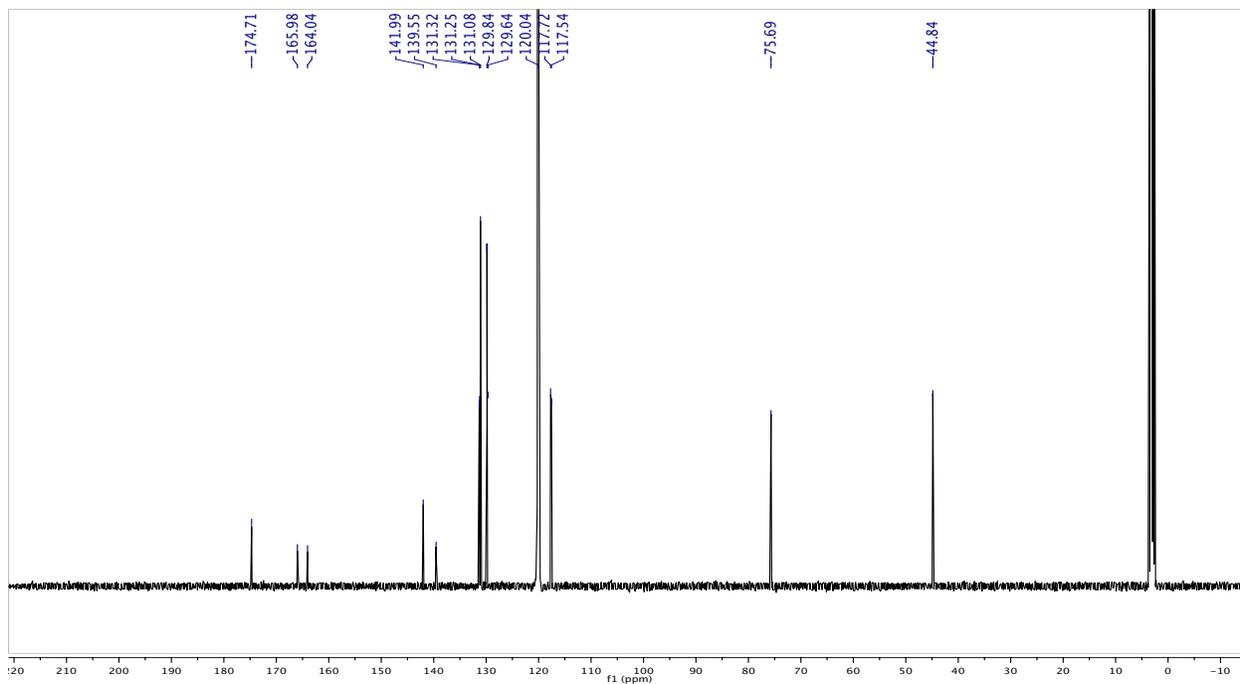
^{13}C NMR of S22 in CDCl_3 (125 MHz):



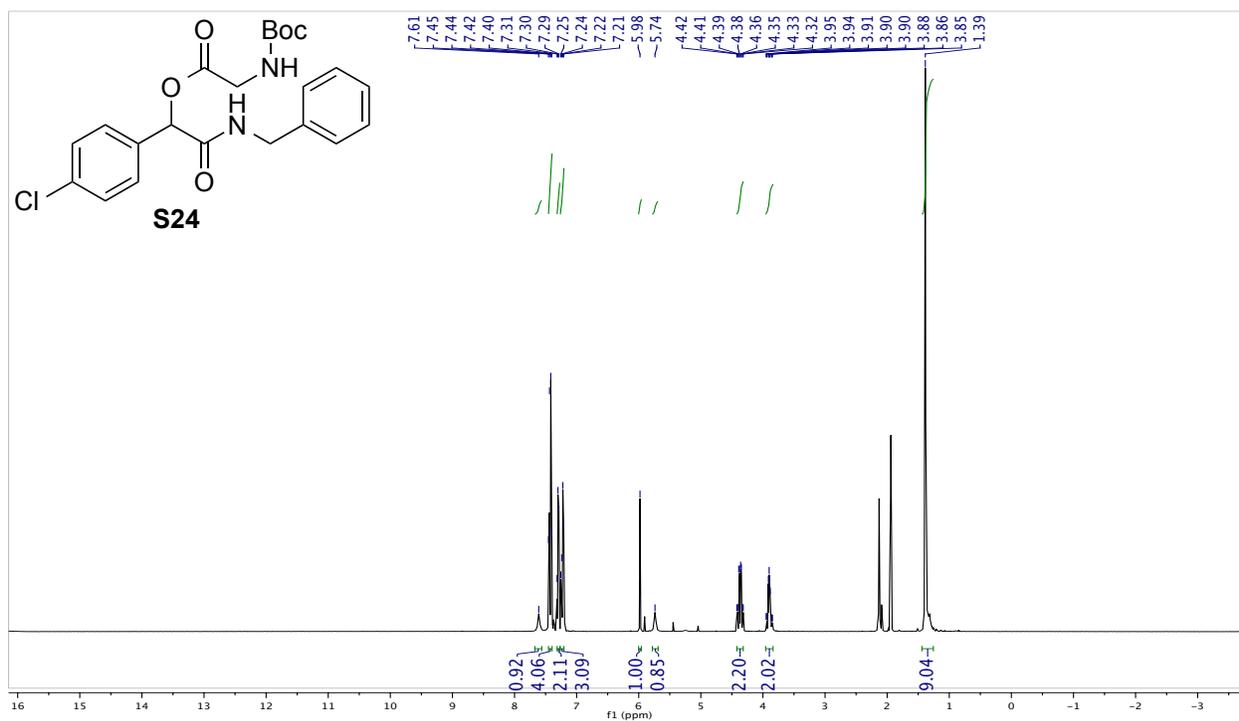
^1H NMR of S23 in CD_3CN (600 MHz):



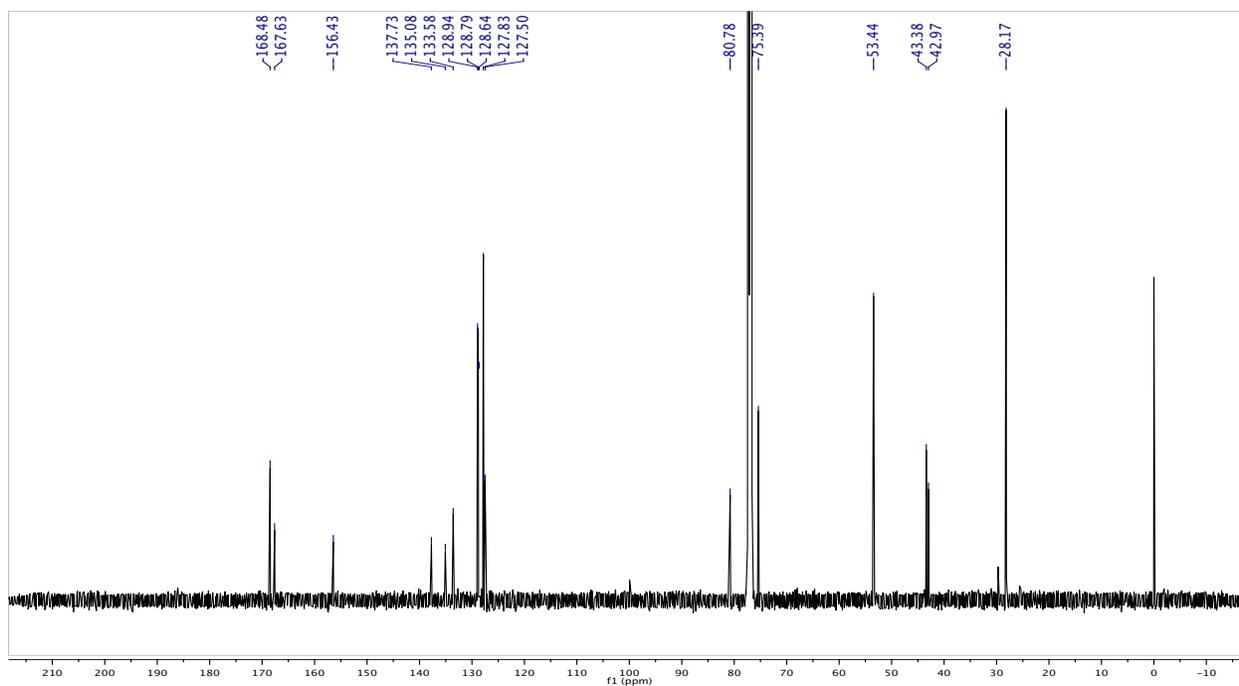
^{13}C NMR of S23 in CD_3CN (125 MHz):



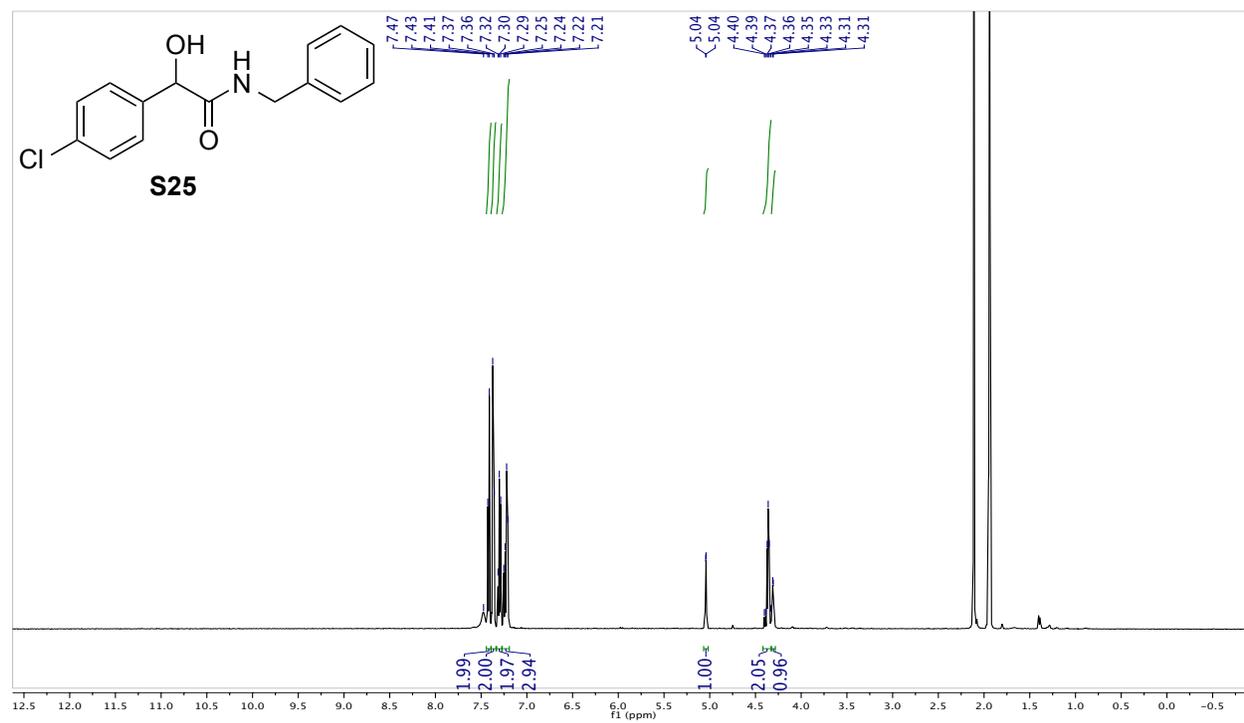
^1H NMR of S24 in CD_3CN (500 MHz):



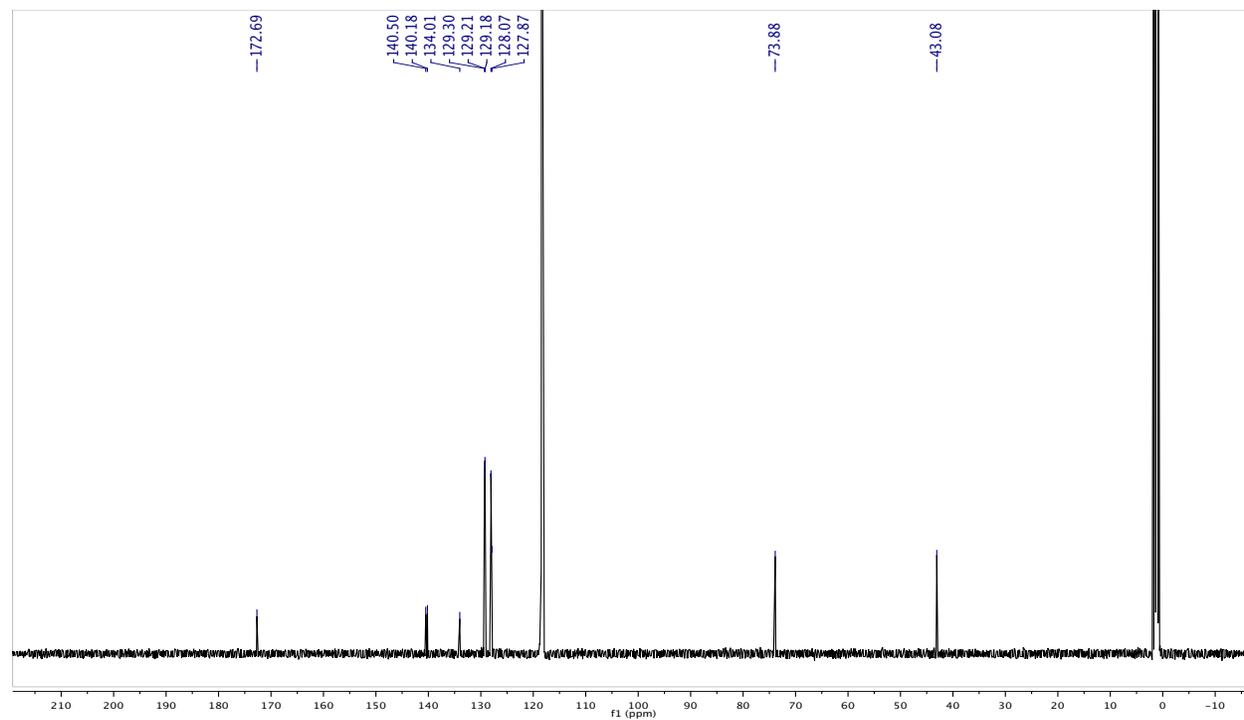
^{13}C NMR of S24/ CH_2Cl_2 in CDCl_3 (125 MHz):



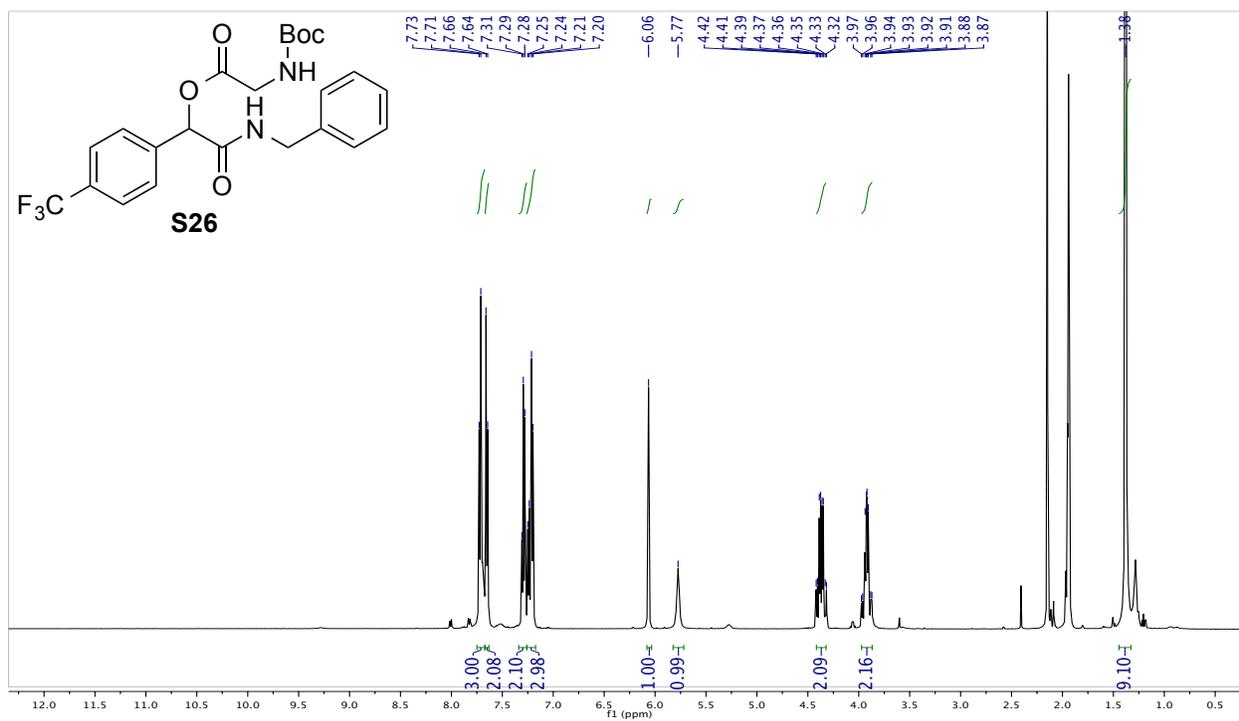
^1H NMR of S25 in CD_3CN (500 MHz):



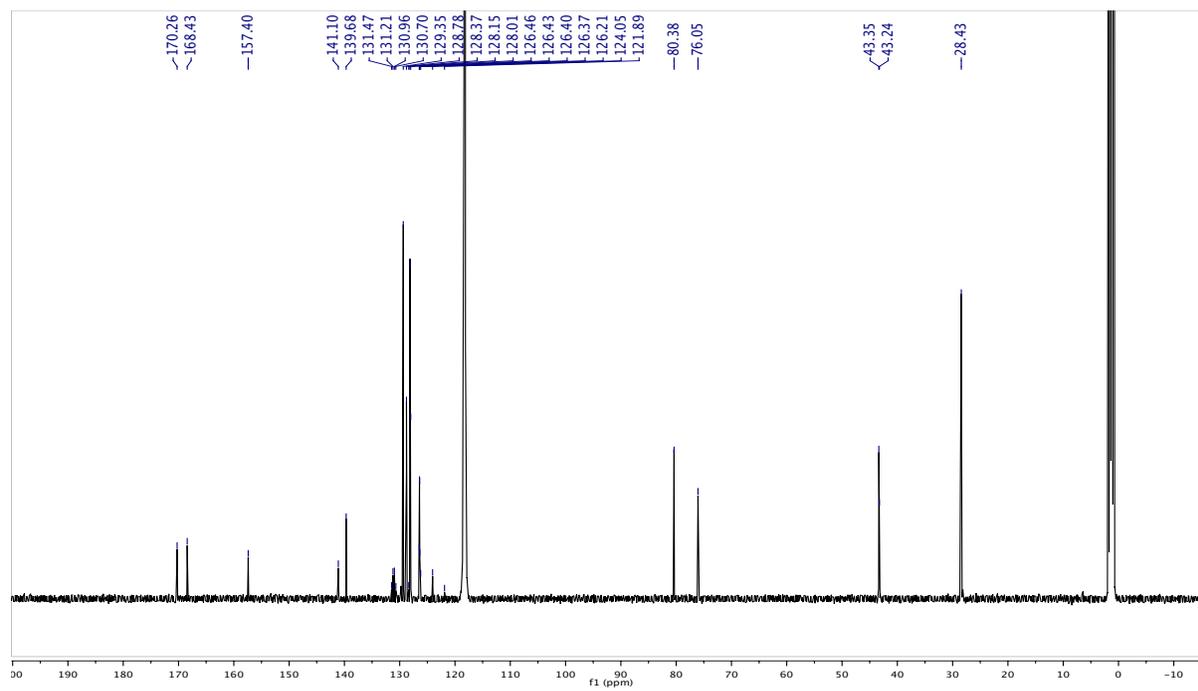
^{13}C NMR of S25 in CDCl_3 (125 MHz):



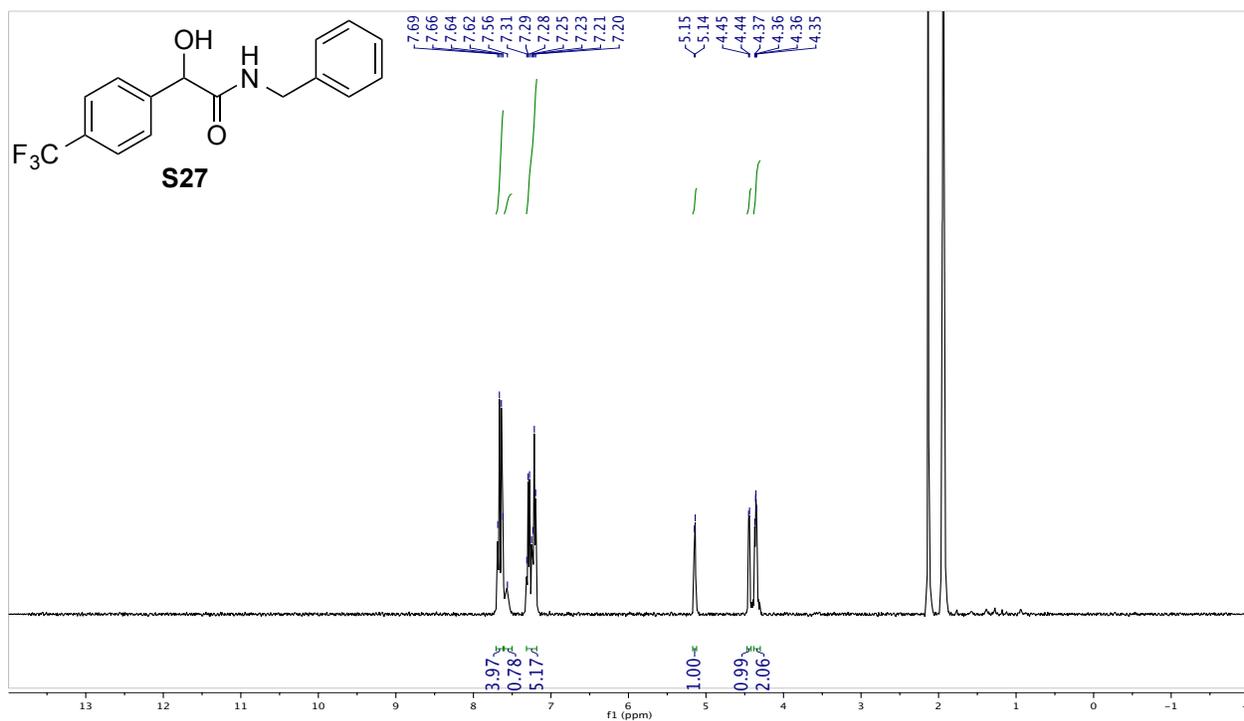
^1H NMR of S26 in CD_3CN (500 MHz):



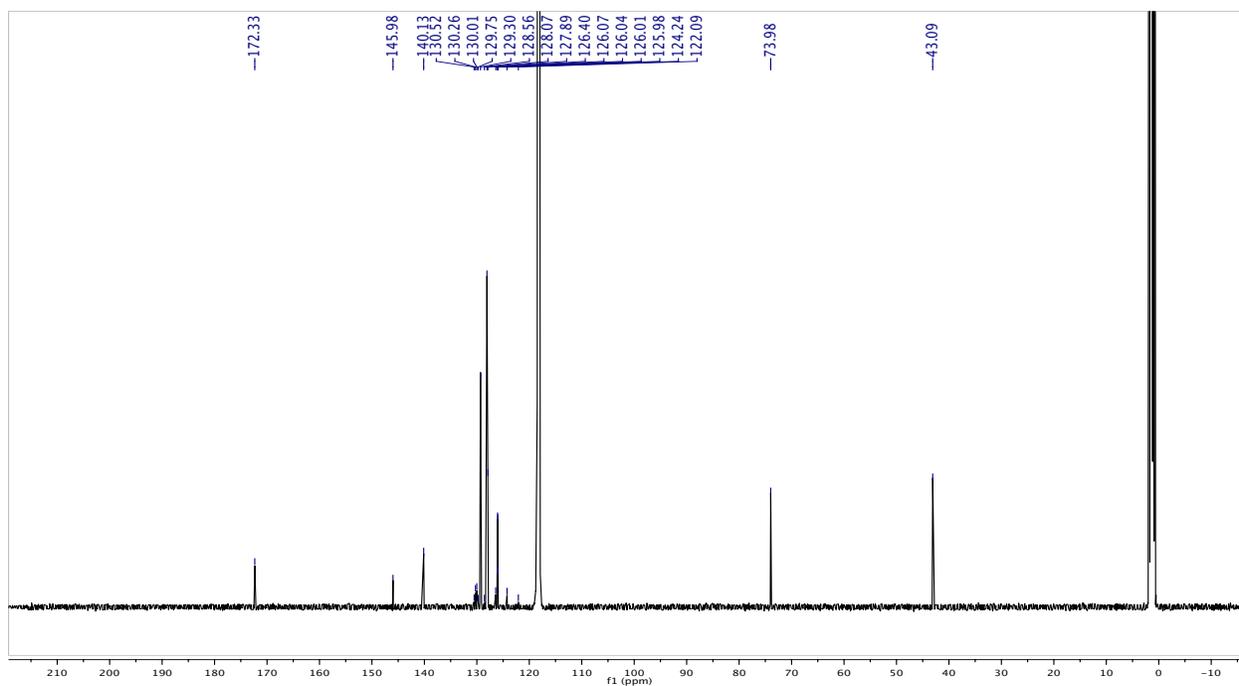
^{13}C NMR of S26 in CD_3CN (125 MHz):



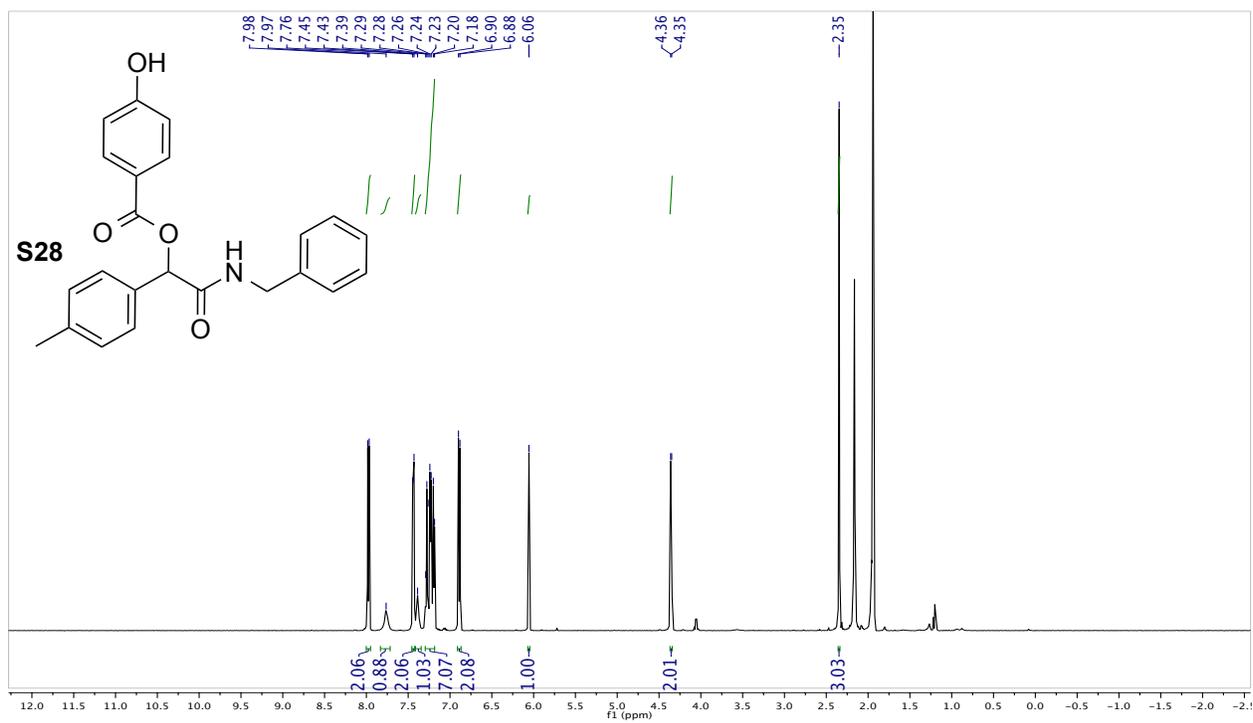
^1H NMR of S27 in CD_3CN (500 MHz):



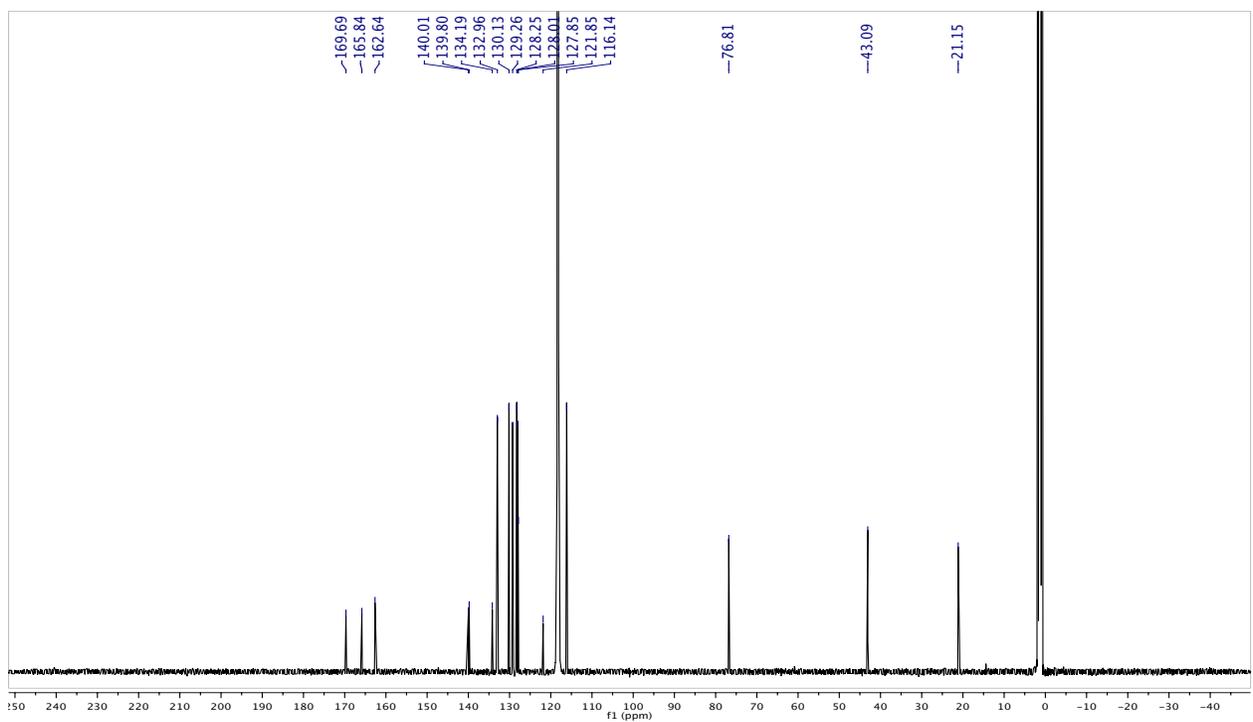
^{13}C NMR of S27 in CD_3CN (125 MHz):



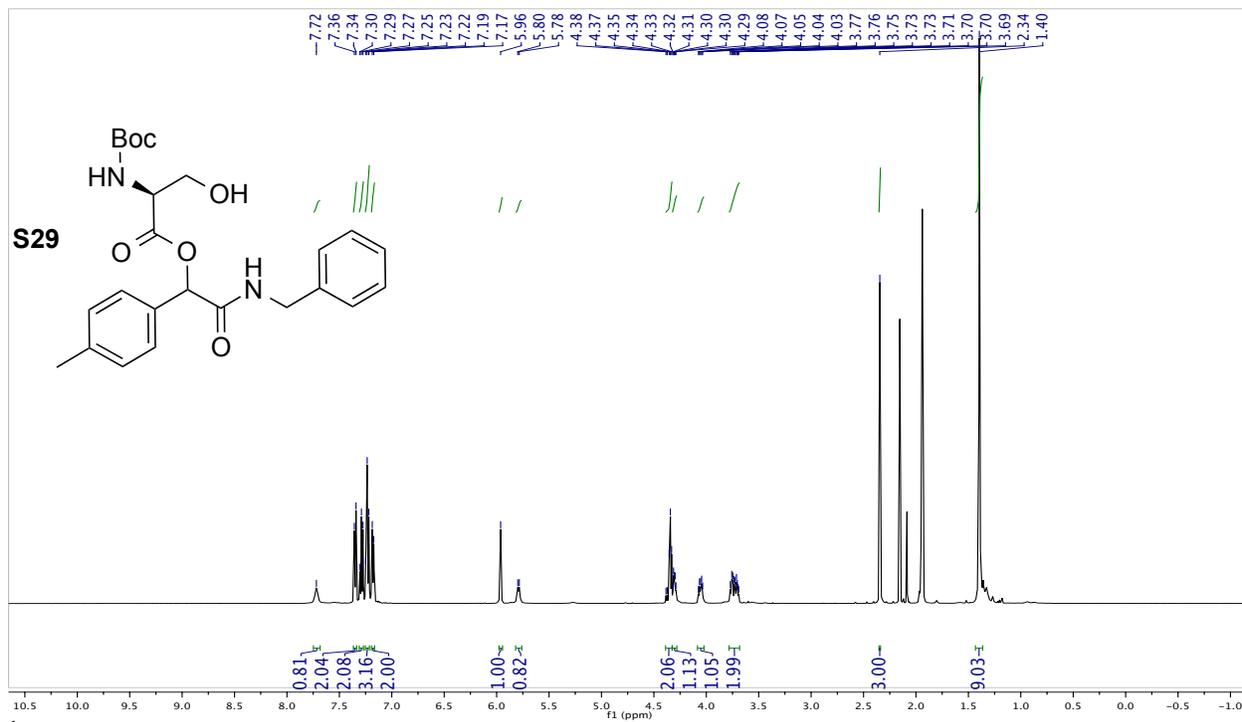
^1H NMR of S28 in CD_3CN (500 MHz):



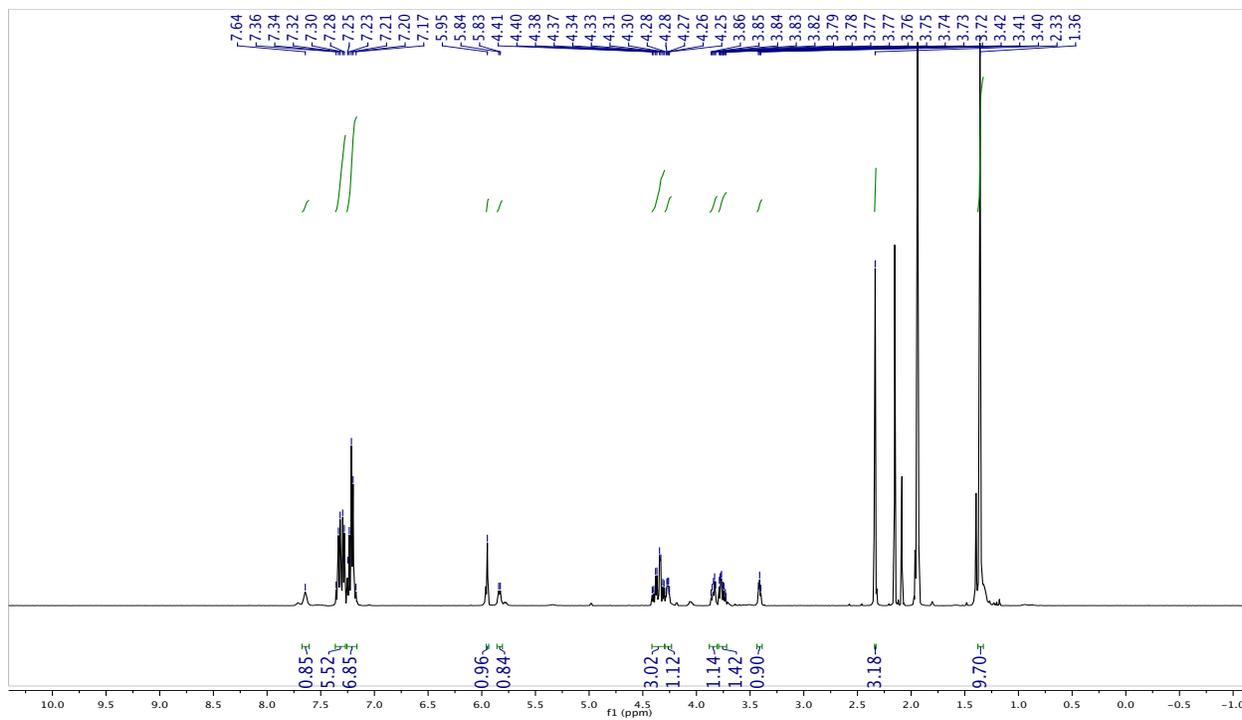
^{13}C NMR of S28 in CD_3CN (125 MHz):



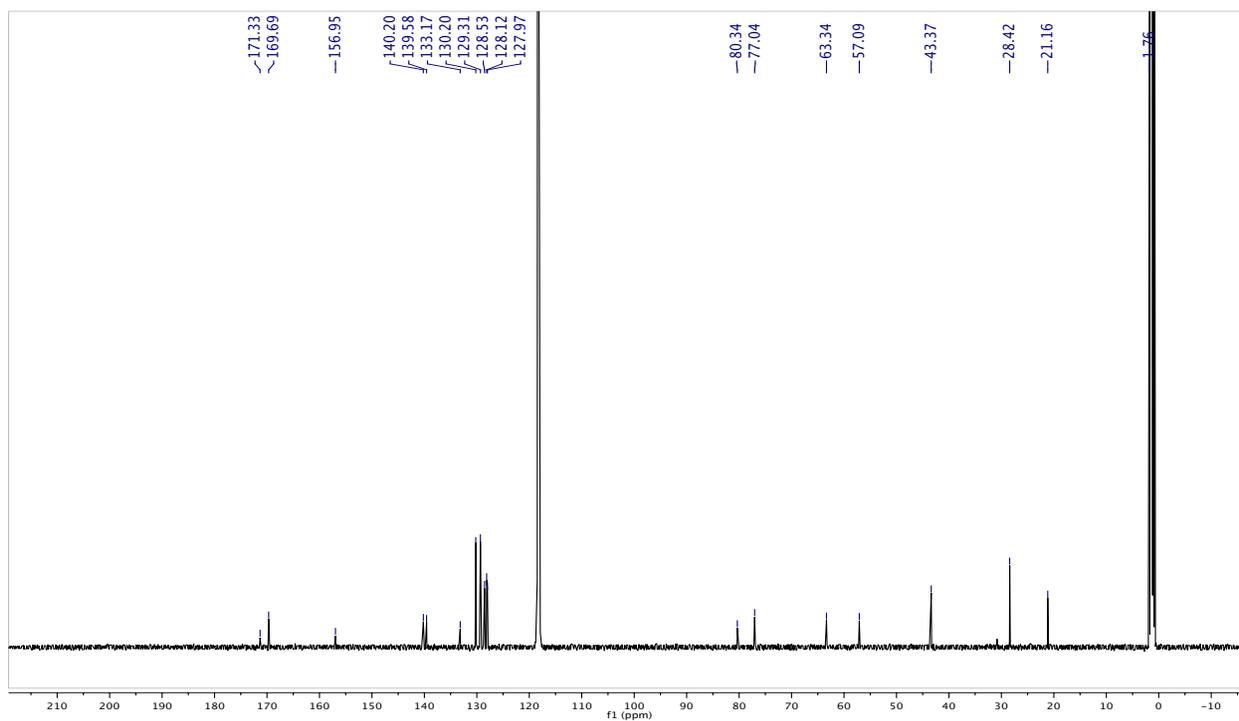
^1H NMR of S29 (Diastereomer A) in CD_3CN (500 MHz):



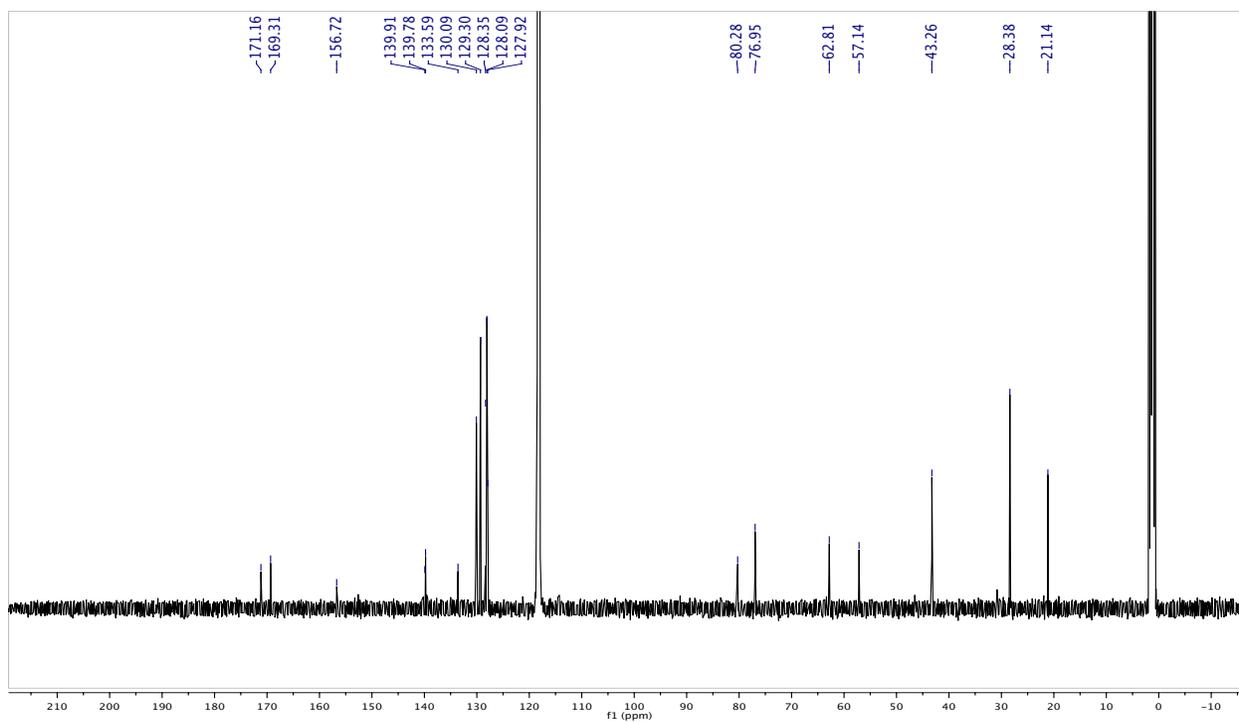
^1H NMR of S29 (Diastereomer B) in CD_3CN (500 MHz):



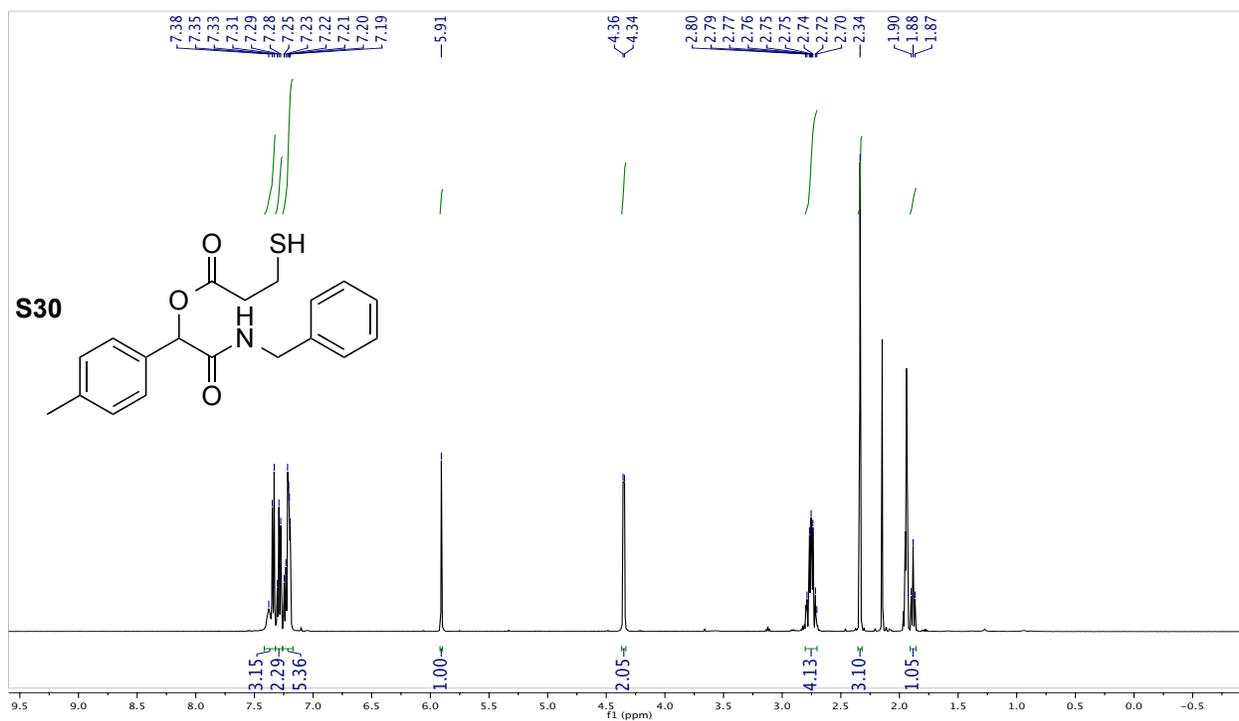
^{13}C NMR of S29 (Diastereomer A) in CD_3CN (100 MHz):



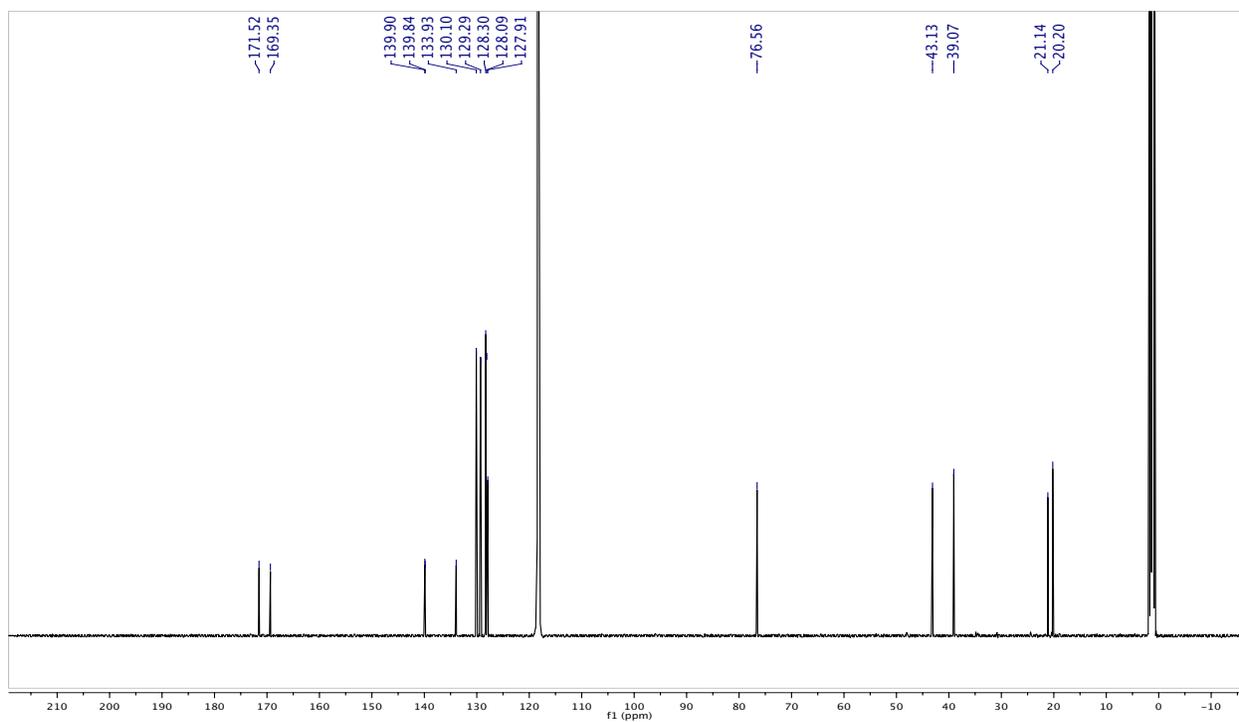
^{13}C NMR of S29 (Diastereomer B) in CD_3CN (100 MHz):



^1H NMR of S30 in CD_3CN (500 MHz):



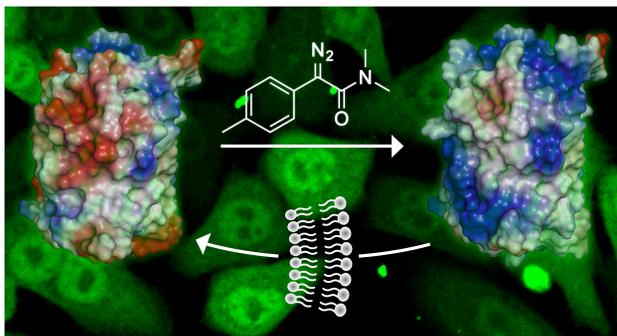
^{13}C NMR of S30 in CD_3CN (125 MHz):



Chapter Three

Cytosolic Delivery of Proteins by Bioreversible Esterification*

*This chapter has been published, in part, under the same title. Reference: Mix, K.A. Lomax, J.E. Raines, R.T. (2017) *J. Am. Chem. Soc.* 139, 14396–14398.



Abstract

Cloaking its carboxyl groups with a hydrophobic moiety is shown to enable a protein to enter the cytosol of a mammalian cell. Diazo compounds derived from (*p*-methylphenyl)glycine were screened for the ability to esterify the green fluorescent protein (GFP) in an aqueous environment. Esterification of GFP with 2-diazo-2-(*p*-methylphenyl)-*N,N*-dimethylacetamide was efficient. The esterified protein entered the cytosol by traversing the plasma membrane directly, like a small-molecule prodrug. As with prodrugs, the nascent esters are substrates for endogenous esterases, which regenerate native protein. Thus, esterification could provide a general means to delivery native proteins to the cytosol.

Author Contributions: Kalie A. Mix synthesized chemical reagents, performed experiments, and analyzed data. Jo E. Lomax constructed the super-charged GFP plasmid. Kalie A. Mix and Ronald T. Raines designed experiments and wrote this chapter.

3.1 Introduction

Approximately 20% of the drugs in today's are proteins.¹⁹³ Essentially all of those proteins act on extracellular targets. This limitation arises from an intrinsic inability of proteins to enter the cytosol.^{194,195} Although viral vectors can be used to deliver DNA that encodes a protein of interest, this genetic approach lacks regulation and can induce stress responses, carcinogenesis, or immunogenicity.^{196,197} In contrast, the direct delivery of proteins into cells would enable temporal control over cellular exposure and minimize deleterious off-target effects.¹⁹⁸

Proteins can be delivered into cells by using site-directed mutagenesis,¹⁹⁹⁻²⁰¹ irreversible chemical modification,^{202,203} conjugation of transduction domains (such as cell-penetrating peptides, CPPs),²⁰⁴⁻²¹⁰ cationic lipid carriers,²¹¹ or electroporation.²¹² Many of these strategies show promise but also pose problems,^{198,194,195} such as inefficient escape from endosomes or inapplicability in an animal.

To cross the plasma membrane, proteins must overcome two barriers: Coulombic repulsion from the anionic glycocalyx and exclusion from the hydrophobic environment of the lipid bilayer.²¹³ Natural and synthetic systems suggest means to overcome these barriers. For example, mammalian ribonucleases are capable of cytosolic entry that is mediated by clusters of positively charged residues.²¹⁴ Cellular uptake can also be enhanced by exogenous hydrophobic moieties.²¹⁵ For example, noncovalent complexation with pyrene butyrate enables the cytosolic delivery of a green fluorescent protein (GFP) conjugate to a cationic CPP.²¹⁶ Additionally, several natural and synthetic protein transduction domains (*e.g.*, penetratin, TP10, and *p*VEC) consist of cationic and hydrophobic residues, which impart an amphipathic character.^{217,218,208,209} Their hydrophobic residues are crucial for mediating membrane translocation.

We envisioned a different strategy—one that invokes a chemoselective reaction that remodels the protein surface to become less anionic and more hydrophobic. The surface of proteins displays cationic groups (*i.e.*, guanidinium, ammonium, and imidazolium) and anionic groups (carboxylates). We hypothesized that the esterification of its carboxyl groups could endow a protein with the ability to access the cytosol. In particular, by cloaking negative charges with a hydrophobic moiety, we might increase the nonpolar surface area while enabling endogenous positive charges to manifest favorable Coulombic interactions with anionic cell-surface components. The ensuing mode-of-action would resemble that of small-molecule prodrugs, which have been in the pharmacopoeia for decades.^{155,219-221}

3.2 Results and Discussion

To effect our strategy, we employed diazo compounds derived from (*p*-methylphenyl)glycine. We had shown previously that the basicity of such diazo compounds enables the efficient esterification of carboxylic acids in an aqueous environment.^{190,129,222} Now, we exploited the modular nature of this scaffold. Specifically, we deimidogenated azide precursors^{78,190,129,222} to access diazo compounds **3.1–3.6**, which span a range of hydrophobicity (Figure 3.1A).

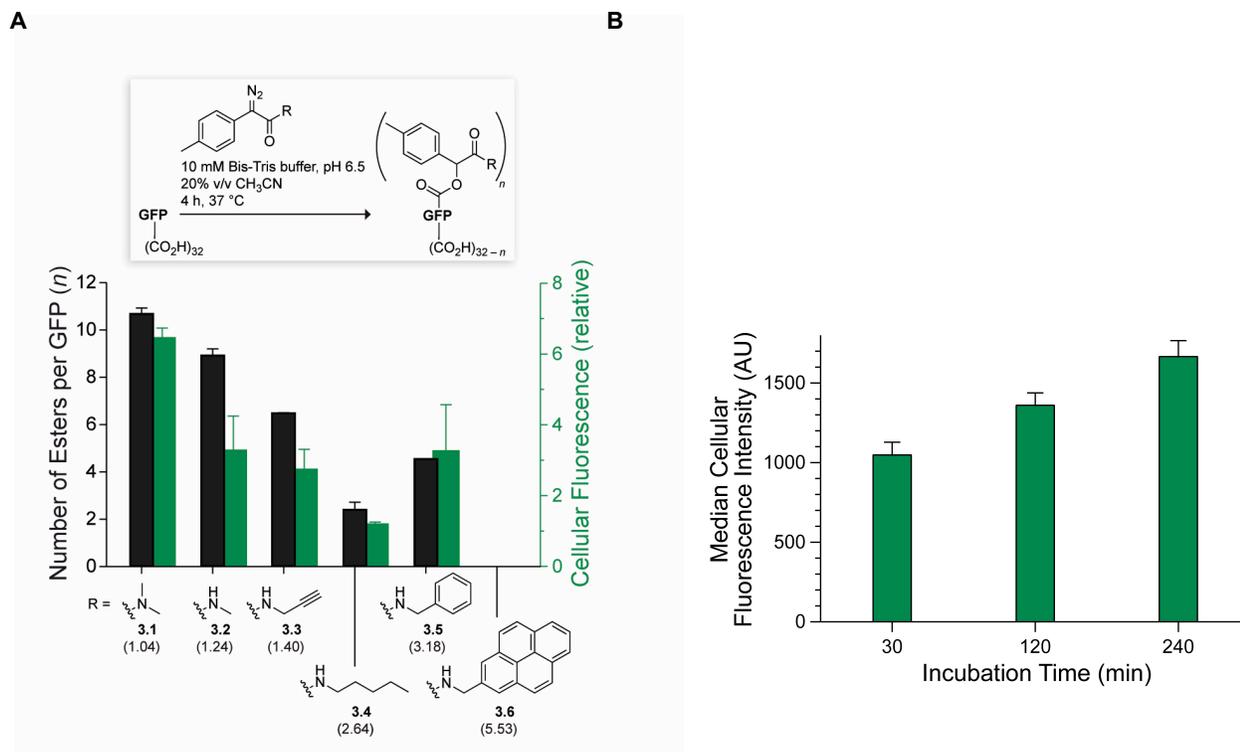


Figure 3.1 (A) Bar graph showing the extent of esterification of a superfolder variant of GFP with diazo compounds **3.1–3.6** (black) and the internalization of the ensuing esterified GFP into CHO-K1 cells (green). Values (\pm SD) were determined by mass spectrometry and flow cytometry, respectively. Parenthetical log*P* values were calculated with software from Molinspiration (Slovensky Grab, Slovak Republic). (B) Time-course for the cellular internalization of GFP–**3.1**. CHO-K1 cells were incubated with GFP–**3.1** (4 μ M) at 37 °C, and internalization was quantified by flow cytometry after 30, 120, and 240 min.

Then, we screened solution conditions for maximal protein esterification by our scaffold. We were aware that the mechanism of esterification requires a protonated carboxyl group,¹⁴⁴ which is encouraged by a low pH and an organic cosolvent. Using GFP and diazo compound **3.3**, we found that an aqueous solution at pH 6.5 that contains 20% v/v acetonitrile gives a high yield of esters (Figure 3.2). These conditions should be tolerable by most proteins.

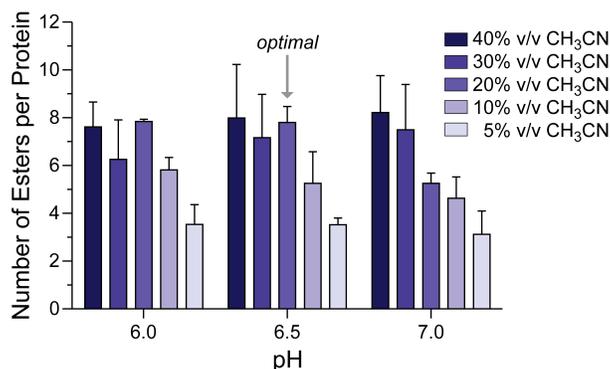


Figure 3.2 Bar graph showing the extent of esterification of GFP with diazo compound **3.3** under different solvent conditions.

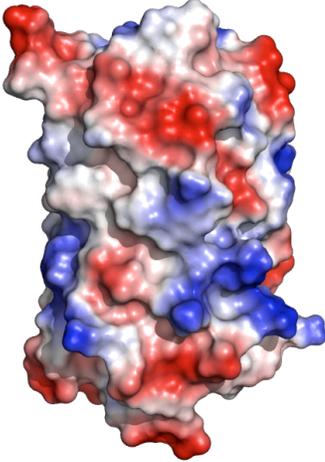
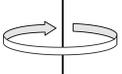
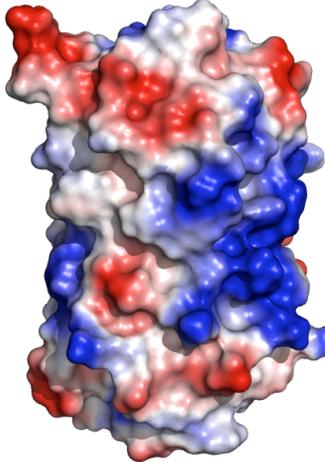
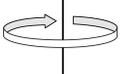
Next, we evaluated diazo compounds **3.1–3.6** for their ability to esterify a protein and facilitate its internalization into a mammalian cell. We found that more polar diazo compounds alkylated more carboxyl groups than did less polar compounds (Figure 3.1A). Then, we treated live cells with esterified proteins and quantified internalization with flow cytometry. We discovered that the level of cellular internalization parallels the number of labels per protein (Figure 3.1A), which suggests that simply masking anionic groups is advantageous. Moreover, cellular fluorescence increases in a time-dependent manner (Figure 3.1B), as expected for a process based on vectorial diffusion from the outside to the inside.

Of the six diazo compounds, compound **3.1** was the most effective in engendering cellular uptake and was selected for further study. On average, 11 of the 32 carboxyl groups in GFP were masked as neutral esters by diazo compound **3.1** (Figure 3.1A). Although the esterification of 11 carboxyl groups in GFP could produce ${}_{32}C_{11} = 1.2 \times 10^8$ different molecules, esters are most likely to form with solvent-accessible carboxyl groups that have a high pK_a value (Table 3.1).¹⁴⁴ That trend was apparent in tandem mass spectrometry data (Figure 3.3). Glutamic acid residues

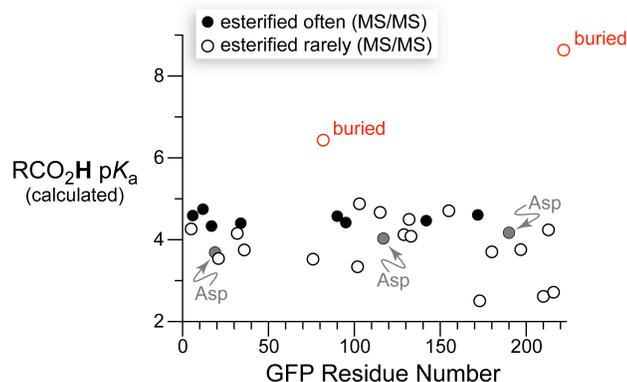
(side chain carboxyl $pK_a = 4.1$) were esterified more often than aspartic acid residues (side chain carboxyl $pK_a = 3.9$). In addition to pK_a , sterics likely play a role in this selectivity since the glutamic acid side chain has one more methylene group than that of aspartic acid, making it more solvent-accessible.

This selectivity has fortuitous consequences. An aspartate or glutamate residue within a hydrophobic patch is a likely target for esterification, which would extend the size of the patch. Clustered anionic residues likewise have high pK_a values, and their esterification would overcome a strong deterrent to cellular uptake. In contrast, an aspartate or glutamate residue within a salt bridge is unlikely to be esterified, but a salt bridge manifests less Coulombic repulsion with anionic cell-surface components than do isolated or clustered anionic residues. The effect of esterification of solvent-accessible carboxyl groups with the highest pK_a values on GFP is modeled in Table 3.1. The 11 aspartic or glutamic acid residues with the highest calculated pK_a values were replaced with phenylalanine to mimic cloaking of the anionic charge. The resulting model demonstrates that large, anionic patches on native GFP are indeed replaced with cationic patches on the esterified GFP model. The tandem mass spectrometry data shows that the esterified residues are in good agreement with those that were predicted to be esterified based on pK_a .

Table 3.1 Notional effect of esterification on the electrostatic surface of GFP.

Residue type	Residue number	Calculated pK_a^a	Esterified? ^b	Calculated electrostatic surfaces ^c	
Lys	140	12.55		GFP (Z = -9)  front  180°	esterified GFP (Z = +2)  front  180°
Lys	45	12.53			
Lys	209	12.47			
Lys	79	12.43			
Lys	3	12.41			
Lys	131	12.31			
Lys	162	12.19			
Lys	113	12.07			
Lys	101	12.06			
Lys	41	12.05			
Lys	105	11.96			
Lys	26	11.93			
Lys	126	11.86			
Lys	107	11.79			
Lys	85	11.10			
Glu	222	8.64			
His	139	7.54			
His	77	7.43			
His	217	7.41			
His	25	7.30			
His	81	7.15			
His	231	7.13			
His	221	7.12			
Asp	82	6.44			
His	148	6.44			
His	199	5.65			
His	181	5.57			
Asp	103	4.88			
Glu	124	4.75			
Asp	155	4.71			
Glu	115	4.67			
Glu	172	4.61			
Glu	6	4.59			
Glu	90	4.58			
Glu	132	4.50			
Glu	142	4.47			
Glu	95	4.42			
Glu	34	4.41			
Glu	17	4.34			
Glu	5	4.27			
Glu	213	4.24			
Asp	190	4.17			
Glu	32	4.16			
Asp	129	4.13			
Asp	133	4.09			
Asp	117	4.03			
His	169	3.76			
Asp	197	3.76			
Asp	36	3.75			
Asp	180	3.71			
Asp	19	3.69			
Asp	21	3.54			
Asp	76	3.53			
Asp	102	3.34			
Asp	216	2.72			
Asp	210	2.62			
Asp	173	2.51			
				back	back

^apK_a values were calculated with the program Depth (Tan, K.P.; Nguyen, T.B.; Patel, S.; Varadarajan, R.; Madhusudhan, M. S. *Nucleic Acids Res.* **2013**, *41*, W314–W321. ^bEsterification with diazo compound **3.1**, as detected by MS/MS analysis. ^cElectrostatic surfaces were calculated with the program PyMOL from Schrödinger (New York, NY). Glu222 and Asp82 (red highlight in table) are not solvent accessible. To generate the “esterified GFP” image, the other eleven aspartate and glutamate residues (green highlight in table) with the highest calculated pK_a values were replaced with phenylalanine residues. Net charge (Z) = Arg + Lys – Asp – Glu.



Peptide	Peptidic Residues	Modified Residue	
		sfGFP Number	2b3p Number
MHHHHHHSSGVDLGTENL	1–18	Glu16	NA
YFQGMVSKGEEL	14–20	Glu29	Glu6
VELDGDVNGHKFSVRGEGEGDATIGKLTLLKF	39–69	Glu40	Glu17
VELDGDVNGHKFSVRGEGEGEGDATIGKLTLLKF	39–69	Asp42	Asp19
VELDGDVNGHKFSVRGEGEGEGDATIGKL	39–65	Glu57	Glu34
KSAMPEGYVQERTISF	108–123	Glu113	Glu90
KSAMPEGYVQERTISF	108–123	Glu118	Glu95
KDDGKYKTRAVVKFEGDTLVNRIEL	124–148	Asp140	Asp117
KDDGKYKTRAVVKFEGDTLVNRIEL	124–148	Glu147	Glu124
KGTFKEDGNILGHKLEYNF	149–168	Glu165	Glu142
TVRHNVEDGSVQL	189–201	Glu195	Glu172
ADHYQQNTPIGDGPVL	202–217	Asp213	Asp190
GMDELYK	255–261	Asp257	NA
HEYVNAAGITLGMDELYK	244–261	Glu258	NA

Figure 3.3 Graph showing the often-esterified carboxyl groups in sfGFP as identified by tandem mass spectrometry. Values of calculated carboxyl group pK_a are for PDB entry 2b3p (Table 3.1). GFP residue number also refers to PDB entry 2b3p.

We used confocal microscopy to visualize the uptake of GFP by live mammalian cells. For calibration, we compared the uptake of GFP with that of a “super-charged” variant in which site-directed mutagenesis was used to replace anionic residues with arginine.¹⁹⁹ Unmodified GFP did not enter cells (Figure 3.4). Super-charged GFP did enter cells, but produced a punctate pattern of fluorescence that is suggestive of endosomal localization. At 4 °C, which is a temperature that

precludes endocytosis,²²³ the fluorescence from supercharged GFP was scant and localized to the plasma membrane.

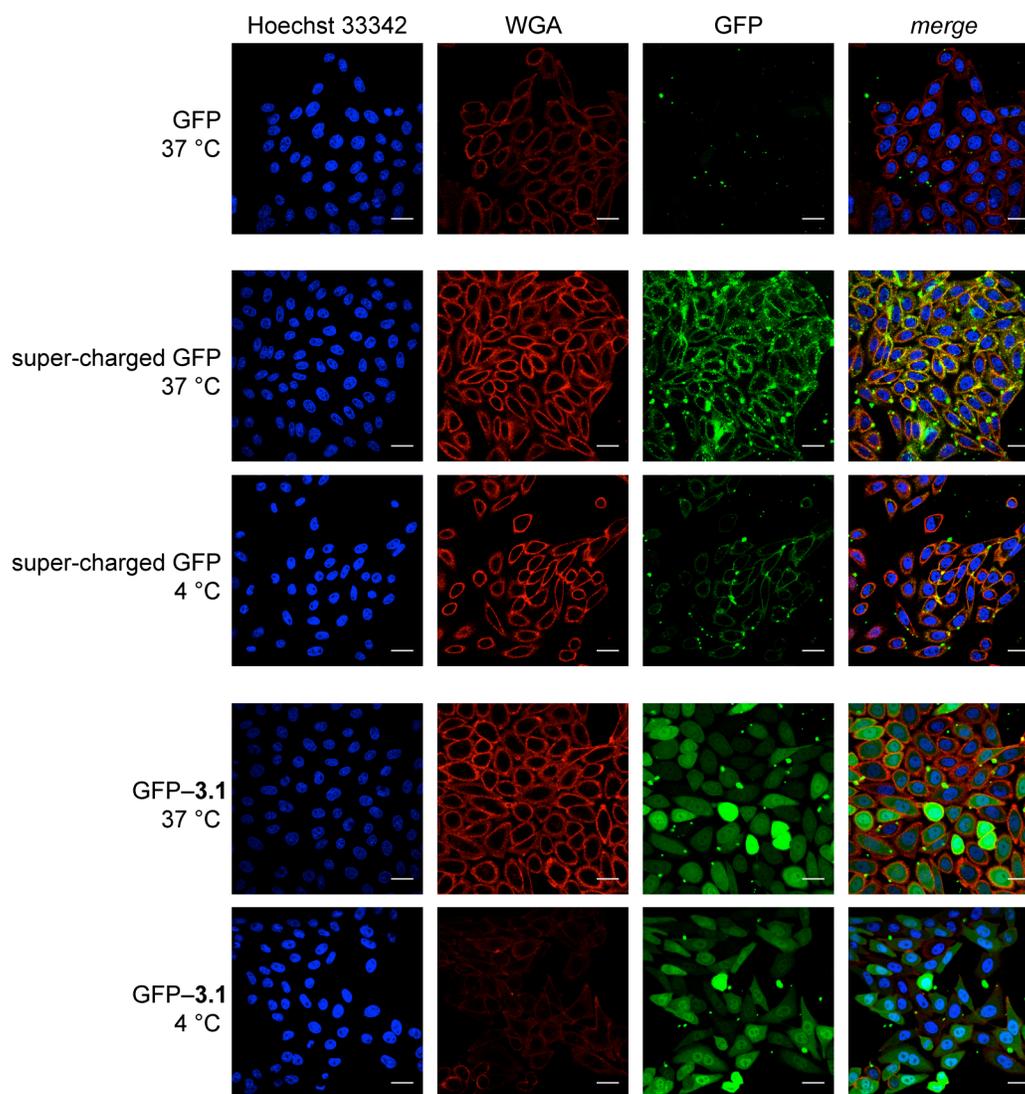


Figure 3.4 Images of the cellular internalization of GFP and its super-charged and esterified variants. CHO-K1 cells were incubated with protein (15 μ M) for 2 h at 37 or 4 $^{\circ}$ C. Cells were then washed, stained with Hoechst 33342 and wheat germ agglutinin (WGA)–Alexa Fluor 647, and imaged by confocal microscopy (Hoechst 33342: ex. 405 nm, em. 450 nm; WGA–Alexa Fluor 647 ex. 647 nm, em. 525 nm; GFP ex. 488 nm, em. 525 nm.) Scale bars; 25 μ m.

Images of cells treated with GFP-**3.1** were in marked contrast to those treated with unmodified GFP or super-charged GFP. At 37 °C, treatment with GFP-**3.1** elicited diffuse fluorescence, suggestive of cytosolic localization (Figure 3.4). Most remarkably, this pattern persisted at 4 °C, indicating that uptake does not rely on endocytosis. In other words, GFP-**1** appears to enter cells by passing directly through the plasma membrane, like a small-molecule prodrug.^{155,219-221}

To enter the nucleus, a protein must pass through the cytosol. To verify cytosolic entry, we reiterated a known GFP variant bearing a nuclear localization signal (nlsGFP)^{224,225} and esterified that variant with compound **3.1**. We then treated live cells with either nlsGFP or esterified nlsGFP (nlsGFP-**3.1**) and visualized the cells with confocal microscopy. In the ensuing images (Figure 3.5), nlsGFP colocalizes with membrane stain (Pearson's $r = 0.21$) and is excluded from the nucleus ($r = -0.12$). This result is expected, as GFP is impermeant but a nuclear localization signal is cationic and can form salt bridges with the anionic glycocalyx. In contrast, nlsGFP-**3.1** not only exhibits diffuse staining like GFP-**3.1** (Figure 3.4), but also colocalizes with a nuclear stain ($r = 0.51$) to an extent expected for this particular variant.^{224,225} These data indicate that nlsGFP-**3.1** accesses the nucleus and, thus, the cytosol.

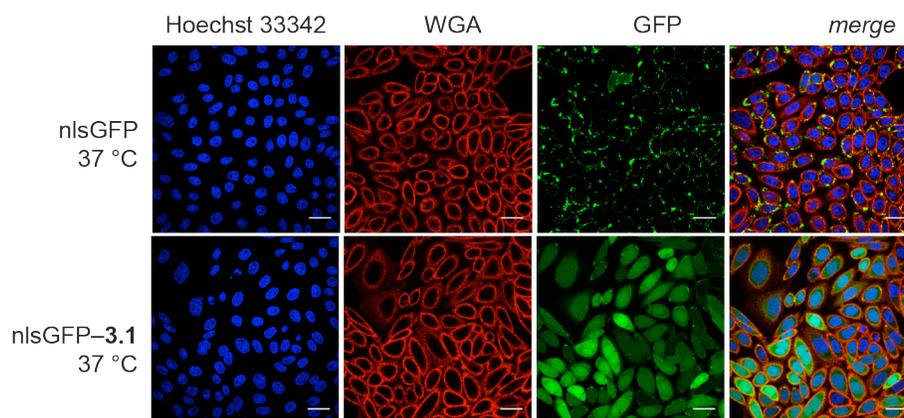


Figure 3.5 Images of the nuclear internalization of a protein that contains a nuclear localization signal and its esterified variant. CHO-K1 cells were incubated with nlsGFP or nlsGFP-3.1 (15 μ M) for 2 h at 37 $^{\circ}$ C. Cells were then washed, stained with Hoechst 33342 and WGA–Alexa Fluor 647, and imaged by confocal microscopy (Hoechst 33342: ex. 405 nm, em. 450 nm; WGA–Alexa Fluor 647: ex. 647 nm, em. 700 nm; GFP: ex. 488 nm, em. 525 nm). Scale bars: 25 μ m

Finally, we investigated the bioreversibility of esterification. Incubation of a model protein esterified with diazo compound **3.1** in a mammalian cell extract resulted in the complete removal of labels (Figure 3.6). This finding is consistent with an inability of de-esterified GFP-3.1 (*i.e.*, GFP) to escape from the cytosol and its accumulation there (Figure 3.1B). Thus, the esters formed upon reaction with **3.1** are substrates for endogenous esterases, like prodrugs.^{155,219-221} Moreover, the alcohol product of the esterase-mediated hydrolysis is benign to mammalian cells (Figure 3.7).

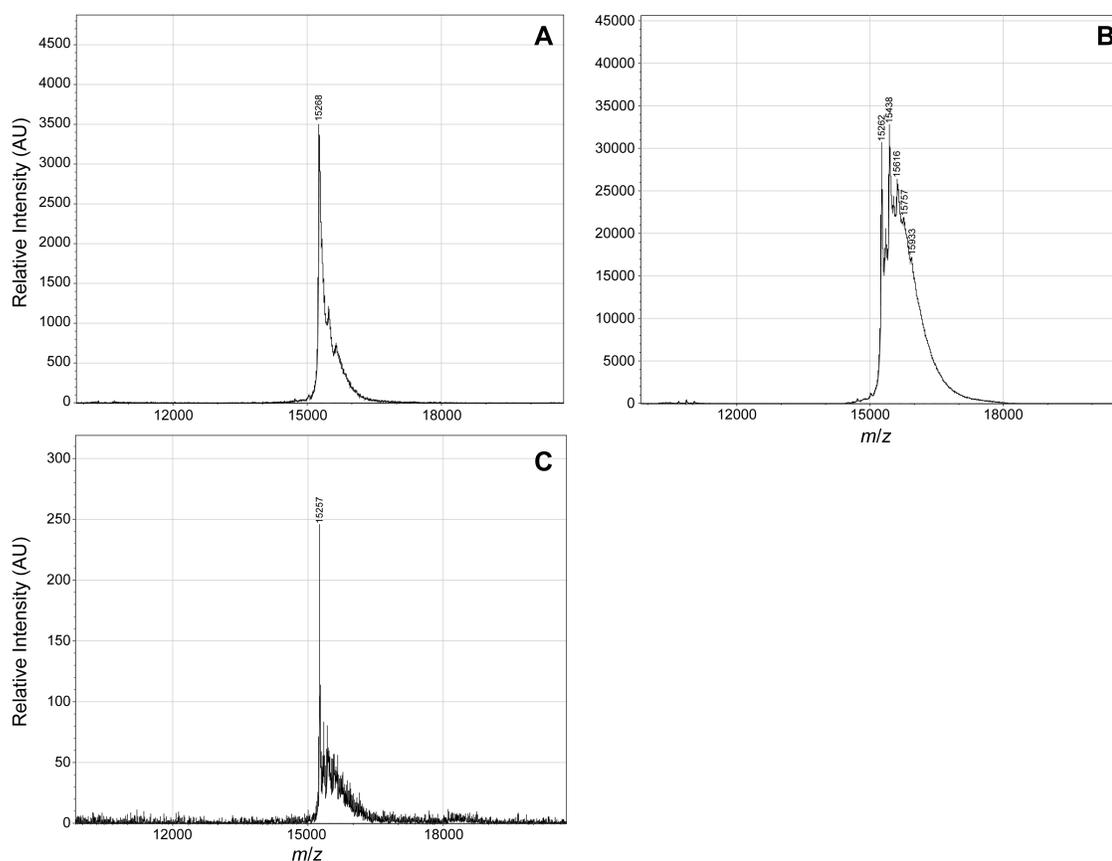


Figure 3.6 MALDI-TOF spectra to assess the reversibility of protein esterification with diazo compound **3.1**. (A) FLAG-angiogenin. (B) FLAG-angiogenin after treatment with diazo compound **3.1**. (C) FLAG-angiogenin after treatment with diazo compound **3.1** and subsequent incubation with a CHO-cell extract. Expected m/z : $15,270 + 175$ per ester group.

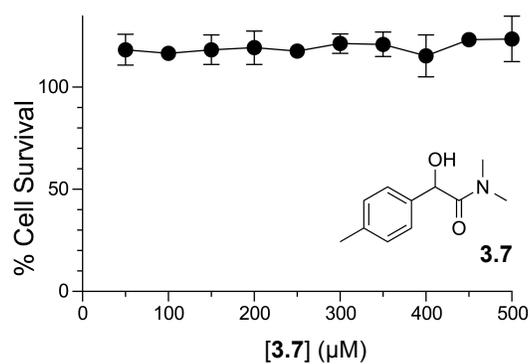


Figure 3.7 Graph of the viability of CHO-K1 cells treated with α -hydroxy dimethylamide **3.7**.

In summary, we have demonstrated that esterification of protein carboxyl groups with a tuned diazo compound can engender delivery of the protein across the plasma membrane as if it were a small molecule. Further, this chemical modification is traceless, being removable by cellular esterases. This delivery strategy provides an unprecedented means to deliver native proteins into cells for applications in the laboratory and, potentially, the clinic.

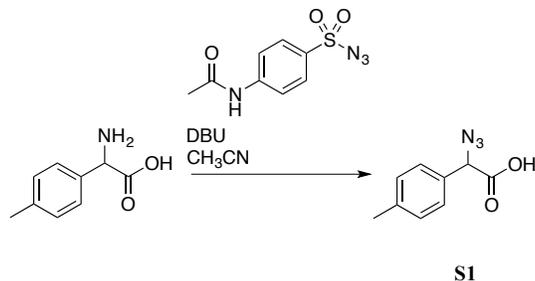
3.3 Acknowledgments

We are grateful to Dr. K. A. Andersen for early observations, Dr. E. K. Grevstad for help with microscopy, L. B. Hyman for technical advice, Dr. T. T. Hoang for supplying FLAG-ANG, and Dr. C. L. Jenkins for contributive discussions. K.A.M. was supported by Molecular Biosciences Training Grant T32 GM007215 (NIH). J.E.L. was supported by a National Science Foundation Graduate Research Fellowship. This work was supported by grant R01 GM044783 (NIH) and made use of the National Magnetic Resonance Facility at Madison, which is supported by grant P41 GM103399 (NIH).

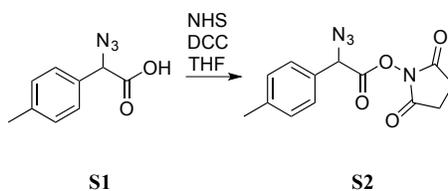
3.4 Materials and Methods

3.4.1 General

Silica gel (40 μm , 230–400 mesh) was from SiliCycle. Reagent chemicals were obtained from commercial sources and used without further purification. Dichloromethane and tetrahydrofuran were dried by passage over a column of alumina. The progress of reactions was monitored by thin-layer chromatography using plates of 250- μm silica 60-F₂₅₄ from EMD Millipore. All procedures were performed in air at ambient temperature (~ 22 °C) and pressure (1.0 atm) unless indicated otherwise. The phrase “concentrated under reduced pressure” refers to the removal of solvents and other volatile materials using a rotary evaporator at water aspirator pressure

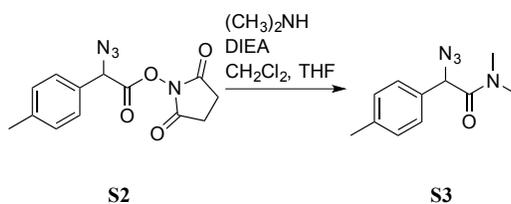


2-Amino-2-(4-methylphenyl)acetic acid (5.0 g, 30.3 mmol) from Matrix Scientific was dissolved in acetonitrile (50 mL). 1,8-Diazabicyclo(5.4.0)undec-7-ene (DBU; 13.8 g, 90.9 mmol) and *p*-aminobenzenesulfonyl azide (8.0 g, 33.3 mmol) were added, and the resulting solution was stirred overnight. The solution was concentrated under reduced pressure. The resulting residue was dissolved in EtOAc (50 mL) and washed twice with 1 M HCl(aq). The organic layer was dried over anhydrous Na₂SO₄(s) and concentrated under reduced pressure to afford α -azido acid S1 (5.2 g, 90%) as a white solid. ¹H NMR (500 MHz, CDCl₃, δ): 7.31 (d, 2H, *J* = 8.1 Hz), 7.24 (d, 2H, *J* = 7.9 Hz), 5.01 (s, 1H), 2.38 (s, 3H). ¹³C NMR (125 MHz, CDCl₃, δ): 173.0, 139.7, 130.2, 129.9, 127.6, 64.8, 21.3. HRMS-ESI (*m/z*): [M - H]⁻ calcd for C₉H₉N₃O₂, 190.0622; found 190.0622.



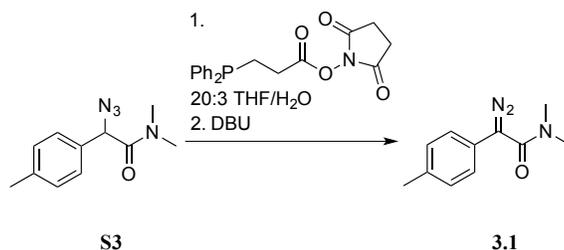
α -Azido acid **S1** (5.2 g, 27.4 mmol) was dissolved in THF (100 mL). *N*-Hydroxysuccinimide (3.1 g, 27.4 mmol) and DCC (6.2 g, 30.1 mmol) were added, and the resulting solution was stirred overnight. The slurry was removed by filtration, and the solution was concentrated under

reduced pressure. The residue was dissolved in EtOAc (50 mL) and washed with saturated aqueous NaHCO₃. The organic layer was dried over anhydrous Na₂SO₄(s) and concentrated under reduced pressure. The residue was purified by chromatography on silica gel, eluting with 3:7 EtOAc/hexanes to afford α -azido *N*-hydroxysuccinimide ester **S2** (6.7 g, 85%) as a white solid. ¹H NMR (500 MHz, CDCl₃, δ): 7.36 (d, 2H, *J* = 8.1 Hz), 7.28 (d, 2H, *J* = 8.0 Hz), 5.25 (s, 1H), 2.83 (s, 4H), 2.38 (s, 3H). ¹³C NMR (125 MHz, CDCl₃, δ): 168.4, 165.2, 140.2, 130.1, 128.8, 127.9, 63.2, 25.6, 21.3. HRMS–ESI (*m/z*): [M + NH₄]⁺ calcd for C₁₃H₁₂N₄O₄, 306.1197; found, 306.1191.

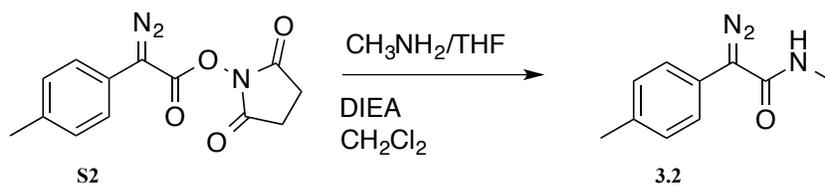


α -Azido *N*-hydroxysuccinimide ester **S2** (417 mg; 1.4 mmol) was dissolved in CH₂Cl₂ (15 mL). A solution of dimethylamine (0.8 mL; 2.0 M in THF) and DIEA (361 mg; 2.8 mmol) were added, and the resulting solution was stirred overnight. The solution was concentrated under reduced pressure. The residue was dissolved in EtOAc and washed twice with 1 M HCl(aq) and saturated aqueous NaHCO₃ (2 \times 10 mL). The organic layer was dried over anhydrous Na₂SO₄(s) and concentrated under reduced pressure. The residue was purified by chromatography on silica gel, eluting with 1:1 EtOAc/hexanes to afford α -azido dimethylamide **S3** (192 mg; 61%) as a white solid. ¹H NMR (500 MHz, CDCl₃, δ): 7.28 (d, 2H, *J* = 8.2 Hz), 7.23 (d, 2H, *J* = 8.0 Hz), 4.91 (s, 1H), 3.01 (s, 3H), 2.81 (s, 3H), 2.37 (s, 3H). ¹³C NMR (125 MHz, CDCl₃, δ): 169.0,

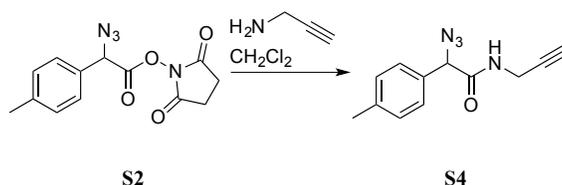
139.2, 130.7, 130.0, 127.9, 63.5, 36.9, 36.1, 21.2. **HRMS-ESI** (m/z): $[M + H]^+$ calcd for $C_{11}H_{14}N_4O$, 219.1240; found, 219.1235.



α -Azido dimethylamide **S3** (100 mg, 0.46 mmol) was dissolved in 20:3 THF/H₂O (4.6 mL). *N*-Succinimidyl 3-(diphenylphosphino)propionate (179 mg, 0.50 mmol) was added, and the resulting solution was stirred for 3 h under N₂(g). 1,8-Diazabicycloundec-7-ene (DBU; 140 mg, 0.92 mmol) was added, and the solution was stirred for 1 h. The solution was diluted with brine (10 mL) and extracted with CH₂Cl₂ (2 × 10 mL). The organic layer was dried over anhydrous Na₂SO₄(s) and concentrated under reduced pressure. The residue was purified by chromatography on silica gel, eluting with 3:7 EtOAc/hexanes to afford α -diazo dimethylamide **3.1** (56 mg, 60%) as an orange solid with mp 57.2–61.6 °C. ¹H NMR (500 MHz, CDCl₃, δ): 7.19 (d, 2H, $J = 8.1$ Hz), 7.11 (d, 2H, $J = 8.2$ Hz), 2.95 (s, 6H), 2.34 (s, 3H). ¹³C NMR (125 MHz, CDCl₃, δ): 166.1, 135.7, 129.9, 124.7, 124.4, 37.7, 21.0. **HRMS-ESI** (m/z): $[M + H]^+$ calcd for $C_{11}H_{13}N_3O$, 204.1131; found, 204.1128.

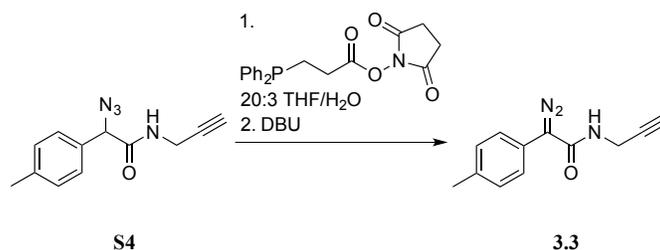


Alpha-diazo *N*-hydroxysuccinimidyl ester **S2** (100 mg, 0.37 mmol) was dissolved in CH₂Cl₂ (37 mL). Methylamine (2.0 M in THF solution; 0.41 mmols, 0.2 mL) and *N,N*-diisopropylethylamine (DIEA; 143 mg, 1.1 mmol) were added, and the solution stirred overnight at ambient temperature. The solution was concentrated under reduced pressure, and the residue was dissolved in EtOAc. The residue was purified by chromatography on silica gel, eluting with 3:7 EtOAc: hexanes to afford **3.2** (34 mg, 49%) as a red solid. ¹H NMR (500 MHz, CDCl₃, δ): 7.26–7.25 (m, 4H), 5.36 (s, 1H), 2.90 (d, 3H, *J* = 4.8 Hz), 2.37 (s, 3H). ¹³C NMR (125 MHz, CDCl₃, δ): 165.6, 138.0, 130.4, 127.9, 123.2, 63.7, 27.0, 21.2. HRMS–ESI *m/z* calcd for C₁₀H₁₁N₄O [M–N₂+H]⁺ 162.0913; found 162.0915.

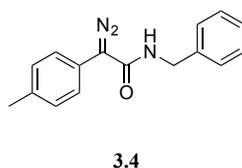


α -Azido *N*-hydroxysuccinimidyl ester **S2** (1.1 g, 3.7 mmol) was dissolved in CH₂Cl₂ (20 mL). Propargylamine (0.2 g, 4.0 mmol) was added, and the resulting solution was stirred overnight. The solution was concentrated under reduced pressure. The resulting residue was dissolved in EtOAc and washed with 1.0 M HCl (2 × 10 mL), followed by saturated aqueous NaHCO₃ (2 × 10 mL). The organic layer was dried with anhydrous Na₂SO₄(s) and concentrated under reduced pressure to afford α -azido propargylamide **S4** (0.6 g, 75%) as an off-white solid. ¹H NMR (500 MHz, CDCl₃, δ): 7.25 (d, *J* = 6.3 Hz, 2H), 7.21 (d, *J* = 8.1 Hz, 2H), 6.64 (s, 1H), 5.03 (s, 1H), 4.08 (dd, *J* = 2.5, 5.25 Hz, 2H), 2.36 (s, 3H), 2.26 (t, *J* = 2.4 Hz, 1H); ¹³C NMR (125 MHz,

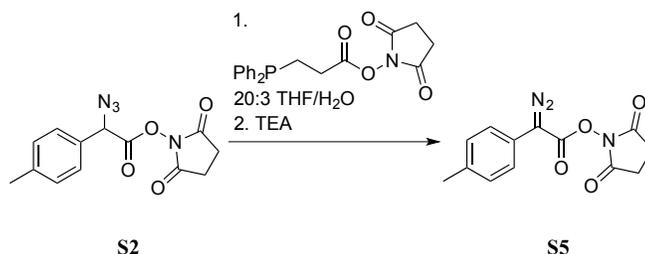
CDCl₃, δ): 167.8, 139.3, 131.6, 129.8, 127.7, 78.8, 72.1, 67.0, 29.4, 21.2; **HRMS-ESI** (m/z): [M-N₂ + H]⁺ calcd for C₁₂H₁₂N₄O, 229.1084; found, 229.1085.



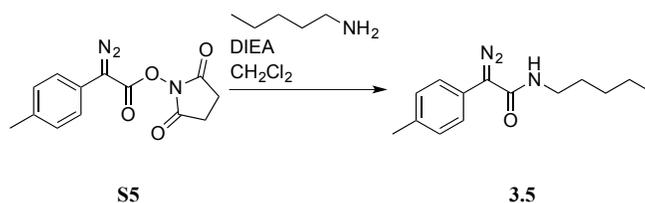
α -Azido propargyl amide **S4** (0.6 g, 2.7 mmol) was dissolved in 16 mL of 20:3 THF/H₂O. *N*-Succinimidyl 3-(diphenylphosphino)propionate (1.1 g, 3.1 mmol) was added under N₂(g), and the resulting solution was stirred for 5 h. 1,8-Diazabicycloundec-7-ene (DBU; 0.8 g, 5.5 mmol) was added, and the solution was stirred for 1 h. The solution was diluted with brine (20 mL) and extracted with CH₂Cl₂ (2 \times 10 mL). The organic layer was dried with anhydrous Na₂SO₄(s) and concentrated under reduced pressure. The residue was purified by chromatography on silica gel, eluting with 3:7 EtOAc/hexanes to afford α -diazo propargylamide **3.3** (0.176 g, 30%) as a red solid. **¹H NMR** (500 MHz, CDCl₃, δ): 7.28–7.24 (m, 4H), 5.52 (s, 1H), 4.15–4.14 (dd, J = 2.5, 5.4 Hz, 2H), 2.38 (s, 3H), 2.23 (s, 1H); **¹³C NMR** (125 MHz, CDCl₃, δ): 164.9, 138.3, 130.5, 128.0, 122.6, 79.6, 71.6, 64.0, 29.7, 21.2; **HRMS-ESI** (m/z): [M + H]⁺ calcd for C₁₂H₁₁N₃O, 214.0975; found, 214.0975.



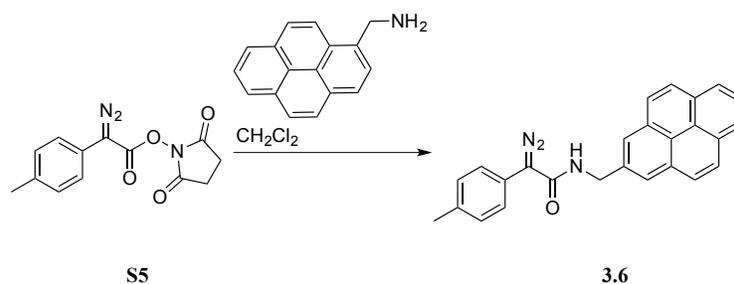
α -Diazo benzylamide **3.4** was prepared as described previously¹²⁹. Yields and spectral data replicated those reported previously.



α -Azido *N*-hydroxysuccinimidyl ester **S2** (3.4 g, 11.6 mmol) was dissolved in 50 mL of 20:3 THF/H₂O. *N*-Succinimidyl 3-(diphenylphosphino)propionate (4.5 g, 12.8 mmol) was added under N₂(g), and the resulting solution was stirred for 5 h. Triethylamine (TEA; 2.3 g, 23.2 mmol) was added, and the solution was stirred for 1 h. The solution was diluted with brine (20 mL) and extracted with CH₂Cl₂ (2 × 10 mL). The organic layer was dried with anhydrous Na₂SO₄(s) and concentrated under reduced pressure. The residue was purified by chromatography on silica gel, eluting with 3:7 EtOAc/hexanes to afford α -diazo *N*-hydroxysuccinimidyl ester **S5** (0.31 g, 10%) as an orange solid. **¹H NMR** (500 MHz, CDCl₃, δ): 7.32 (d, *J* = 8.3 Hz, 2H), 7.22 (d, *J* = 8.1 Hz, 2H), 2.88 (s, 4H), 2.35 (s, 3H); **¹³C NMR** (125 MHz, CDCl₃, δ): 169.4, 160.5, 137.1, 129.9, 124.6, 119.8, 25.6, 21.1; **HRMS–ASAP** (*m/z*): [M–N₂ + H]⁺ calcd for C₁₃H₁₁N₃O₄, 246.0761; found, 246.0764.

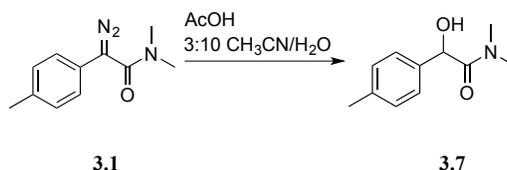


α -Diazo *N*-hydroxysuccinimidyl ester **S2** (100 mg, 0.37 mmol) was dissolved in CH₂Cl₂ (37 mL). *n*-Pentylamine (35.4 mg, 0.41 mmol) and DIEA (143 mg, 1.1 mmol) were added, and the resulting solution was stirred overnight. The solution was concentrated under reduced pressure, and the residue was dissolved in EtOAc. The residue was purified by chromatography on silica gel, eluting with 1:4 EtOAc/hexanes to afford α -diazo pentylamide **3.5** (65 mg, 72%) as a red solid. ¹H NMR (500 MHz, CDCl₃, δ): 7.26–7.23 (m, 4H), 5.37 (s, 1H), 3.36–3.32 (q, *J* = 7.0 Hz, 2H), 2.38 (s, 3H), 1.53–1.49 (m, 2H), 1.33–1.28 (m, 4H), 0.90–0.88 (t, 3H, *J* = 6.9 Hz); ¹³C NMR (125 MHz, CDCl₃, δ): 164.9, 137.9, 130.4, 127.8, 123.3, 63.8, 40.3, 29.6, 29.0, 22.3, 21.2, 14.0; HRMS–ESI (*m/z*): [M–N₂ + H]⁺ calcd for C₁₄H₁₉N₃O, 218.1539; found, 218.1541.



α -Diazo *N*-hydroxysuccinimidyl ester **S5** (55 mg, 0.20 mmol) was dissolved in CH₂Cl₂ (20 mL). 1-Pyrene methylamine (47.0 mg, 0.2 mmol) was added, and the solution was stirred overnight. The solution was concentrated under reduced pressure, and the residue was dissolved in EtOAc. The residue was purified by chromatography on silica gel, eluting with 1:1 EtOAc/hexanes to afford α -diazo pyrenylamide **3.6** (18 mg, 23%) as an orange solid. ¹H NMR (500 MHz, CDCl₃, δ): 8.35 (d, 1H, *J* = 9.2 Hz), 8.28–8.22 (m, 3H), 8.17 (d, 1H, *J* = 7.8 Hz), 8.13–8.06 (m, 3H), 8.00 (d, 1H, *J* = 7.8 Hz), 7.21 (d, 2H, *J* = 8.2 Hz), 7.15 (d, 2H, *J* = 8.1 Hz), 5.78 (t, 1H, *J* = 4.8 Hz), 5.31 (d, 2H, *J* = 5.4 Hz), 2.29 (s, 3H); ¹³C NMR (125 MHz, CDCl₃, δ): 164.7, 137.9, 131.2,

131.17, 130.9, 130.7, 130.4, 129.0, 128.2, 127.7, 127.5, 127.2, 127.1, 126.1, 126.0, 125.4, 125.3, 125.0, 124.8, 124.6, 122.8, 64.0, 42.6, 21.0; **HRMS–ASAP** (m/z): $[M + H]^+$ calcd for $C_{26}H_{19}N_3O$, 390.1601; found, 390.1596.



α -Diazo dimethylamide **3.1** (80 mg, 0.39 mmol) was dissolved in 10 mL of 3:10 $\text{CH}_3\text{CN}/\text{H}_2\text{O}$. Acetic acid (10 μL , 0.17 mmol) was added, and the resulting solution was stirred for 1 h. The solution was concentrated under reduced pressure, and the residue was dissolved in EtOAc. The residue was purified by chromatography on silica gel, eluting with 3:7 EtOAc/hexanes. The eluate was purified further by recrystallization from DCM and hexanes to afford α -hydroxy dimethylamide **3.7** (6 mg, 14%) as a white solid. $^1\text{H NMR}$ (500 MHz, CDCl_3 , δ): 7.21 (d, 2H, $J = 8.2$ Hz), 7.17 (d, 2H, $J = 8.1$ Hz), 5.17 (d, 1H, $J = 6.4$ Hz), 4.71 (d, 1H, $J = 6.4$ Hz), 3.03 (s, 3H), 2.78 (s, 3H), 2.34 (s, 3H); $^{13}\text{C NMR}$ (125 MHz, CDCl_3 , δ): 172.5, 138.3, 136.2, 129.7, 127.4, 71.3, 36.4, 36.3, 21.2; **HRMS–ASAP** (m/z): $[M + H]^+$ calcd for $C_{11}H_{15}NO_2$, 194.1176; found, 194.1176.

3.4.3 Protein Preparation

Preparation of Green Fluorescent Protein (GFP)

The “superfolder” variant of GFP was prepared as described previously²²⁶. The protein was dialyzed into 10 mM Bis-Tris buffer, pH 6.5, prior to esterification.

Preparation of Super-Charged GFP

A gene encoding enhanced GFP was amplified from a mammalian expression vector (Promega) and inserted into a novel vector derived from a pET vector (Novagen). The following substitutions were introduced using site-directed mutagenesis to recapitulate those in the “superfolder”²²⁷ and “cell-penetrating”¹⁹⁹ variants: F64L, S65T, F99S, M153T, V163A, S30R, Y145F, I171V, A106V, Y39I, N105K, I128T, K166T, I167V, S205T, L221H, F223Y, T225N, E17R, D19R, D21R, V111R, and E124R. The expression vector was transformed into BL21(DE3) electrocompetent *E. coli* cells (New England Biolabs) and plated on LB agar containing ampicillin (amp; 200 µg/mL). The resulting plates were incubated overnight at 37 °C. A single colony was added to 50 mL of LB–amp (which contained 200 µg/mL amp) and incubated overnight at 37 °C in a shaking incubator. On the following day, 10 mL of starter culture was added to each of 4 L of Terrific Broth–amp medium (which contained 200 µg/mL amp). Cultures were grown at 37 °C in a shaking incubator until the cell density reached an OD₆₀₀ = 0.6–0.8. Cultures were incubated for 20 min at 20 °C, and then induced by the addition of IPTG (to 1.0 mM). Cells were grown overnight at 20 °C in a shaking incubator.

Cells were harvested by centrifugation at 5,000 rpm for 20 min at 4 °C. Cell pellets were collected and resuspended in 15 mL of lysis buffer per 1 L liquid growth. (Lysis buffer was 50 mM Tris–HCl buffer, pH 7.0, containing 100 mM NaCl, 30 mM imidazole, 1% v/v Triton X-100, and 20% w/v sucrose.) The resuspended cells were stored frozen overnight at –20 °C. On the following day, cells were thawed and lysed by mechanical disruption using a cell disruptor (Constant Systems) at 22 kPsi. The lysate was cleared by centrifugation at 11,000 rpm for 1 h at 4 °C. The supernatant was collected and filtered through a 0.2-µm PES filter (GE Healthcare).

Super-charged GFP in the filtered cell lysate was purified by chromatography with a HisTrap nickel column (GE Healthcare). The binding (wash) buffer was 20 mM sodium phosphate buffer, pH 7.4, containing NaCl (0.50 M) and imidazole (30 mM). The elution buffer was 20 mM sodium phosphate buffer, pH 7.4, containing NaCl (0.50 M) and a linear gradient of imidazole (30 mM–0.50 M). Eluted fractions were pooled and dialyzed against 4 L of 20 mM Tris–HCl buffer, pH 7.4, containing EDTA (1.0 mM).

Dialyzed material was purified further by chromatography with a HiTrap SP HP cation-exchange column (GE Healthcare). The binding (wash) buffer was 20 mM Tris–HCl buffer, pH 7.4, containing EDTA (1.0 mM). The elution buffer was 20 mM Tris–HCl buffer, pH 7.4, containing EDTA (1.0 mM) and NaCl (1.0 M). Upon elution, colored fractions were pooled and dialyzed against PBS overnight and concentrated as needed. *m/z*, 29,547; expected: 29,539 (Figure 3.8A).

Preparation of GFP Containing a Nuclear Localization Sequence (nlsGFP)

A vector containing the gene that encodes “superfolder” GFP gene was reported previously.² This vector was modified to install a nuclear localization sequence³ at the N-terminus of the encoded protein by using the primers:

5'–AAGAAACGCAAGGTACTGGTCCCGGTGGCGACAGTGAGCAAGGGCGAGGAGC–3'

5'–CGGGACCAGTACCTTGCGTTTCTTCTTCGGCATATCTATATCTCCTTCTTAAGGTA–
3'

In addition, the His₆ tag was moved from the N terminus to the C terminus, and the TEV protease recognition sequence was removed. The ensuing nlsGFP has the amino-acid sequence:

MPKKKRKVLVPVATVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDA
 TIGKLTLLKFICTTGKLPVPWPTLVTTLLTYGVQCFSRYPDHMKQHDFFKSA
 MPEGYVQERTISFKDDGKYKTRAVVKFEGDTLVNRIELKGTDFKEDGNIL
 GHKLEYNFNSHNVYITADKQKNGIKANFTVRHNVEDGQSVQLADHYQQNT
 IGDGPVLLPDNHYLSTQTVLSKDPNEKRDHMLHEYVNAAGITLGMDELY
 KAVDKLAAALEHHHHHH

The expression vector was transformed into BL21(DE3) electrocompetent *E. coli* cells (New England Biolabs). The resulting cells were plated on LB agar containing amp (200 µg/mL), and the plates were incubated overnight at 37 °C. A single colony was added to 50 mL of LB–amp (which contained 200 µg/mL amp), and the resulting culture was incubated overnight at 37 °C in a shaking incubator. On the following day, 10 mL of starter culture was added to each of 4 L of Terrific Broth–amp medium (which contained 200 µg/mL amp). Cultures were grown at 37 °C in a shaking incubator until the cell density reached an OD₆₀₀ of 0.6–0.8. Cultures were incubated for 20 min at 20 °C, and then induced by the addition of IPTG (to 1.0 mM). Cells were grown overnight at 20 °C in a shaking incubator.

Cells were harvested by centrifugation at 5,000 rpm for 20 min at 4 °C. Cell pellets were collected and resuspended in 15 mL of lysis buffer per 1 L of liquid growth. (Lysis buffer was 50 mM Tris–HCl buffer, pH 7.0, containing 100 mM NaCl, 30 mM imidazole, 1% v/v Triton X-100, and 20% w/v sucrose.) The resuspended cells were stored frozen overnight at –20 °C. On the following day, cells were thawed and lysed by mechanical disruption using a cell disruptor (Constant Systems) at 22 kPsi. The lysate was cleared by centrifugation for 1 h at 11,000 rpm at 4 °C. The supernatant was collected and filtered through a 0.2-µm PES filter (GE Healthcare), and nlsGFP was purified by chromatography using a HisTrap nickel column, dialysis, and anion-exchange chromatography, as described above *m/z*, 29,945; expected: 29,943 (Figure 3.8B).

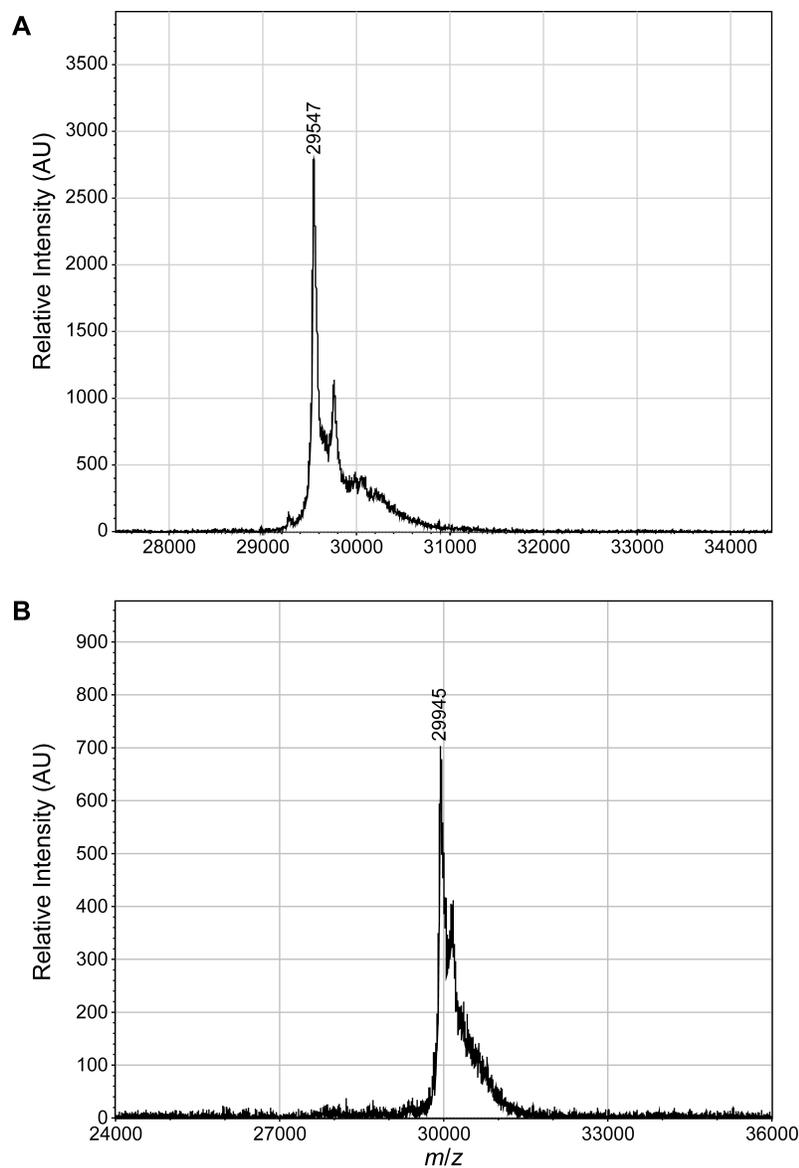
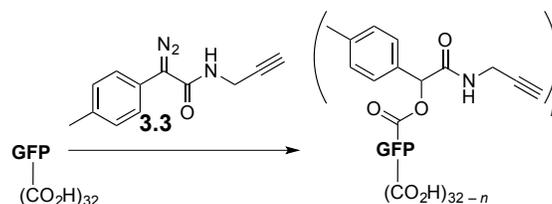


Figure 3.8 MALDI-TOF mass spectra of purified super-charged GFP and nlsGFP. (A) Super-charged GFP. m/z , 29,547; expected: 29,536 without an N-terminal methionine residue. (B) nlsGFP. m/z , 29,945; expected: 29,940 without an N-terminal methionine residue.

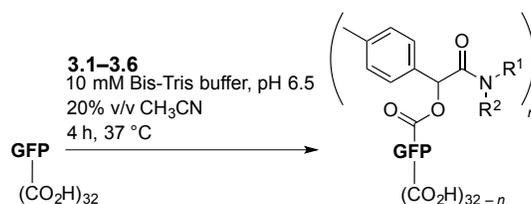
3.4.4 Protein Esterification

Optimization of Solvent Conditions for GFP Esterification



A solution of diazo compound **3.3** (1.2 mg, 0.054 μmol) in acetonitrile was added to a solution of GFP (0.0017 μmol) in 10 mM Bis-Tris buffer (pH 6.0, 6.5, or 7.0). Additional Bis-Tris buffer was added so that the final composition of the solution ranged from 5–40% v/v acetonitrile. The esterification reaction was incubated for 4 h at 37 $^{\circ}\text{C}$. Precipitated protein was removed by filtration through a 0.2- μm PES syringe filter (GE Healthcare), and the number of esters per protein was assigned from the mass of the peak with the highest relative intensity in the MALDI–TOF mass spectrum. The mildest condition that enabled a high level of esterification was 10 mM Bis-Tris buffer, pH 6.5, containing acetonitrile (20% v/v) (Figure 3.2).

Esterification of GFP with Diazo Compounds 3.1–3.6



A solution of diazo compound **3.1–3.6** (34.1 μmol) in acetonitrile (400 μL) was added to a solution of GFP (0.341 μmol) in 1600 μL of 10 mM Bis-Tris buffer, pH 6.5, and incubated for 4 h at 37 $^{\circ}\text{C}$. Precipitated protein was removed by filtration through a 0.2- μm PES syringe filter (GE Healthcare). The number of esters per protein was assigned from the mass of the peak with the highest relative intensity in the MALDI–TOF mass spectrum (Figure 3.9). Protein in each mixture was then purified and exchanged into PBS buffer using PD10 desalting columns (GE Healthcare). Protein was concentrated as needed by centrifugation, and the protein concentration was determined with a BCA assay (Thermo Fisher Scientific).

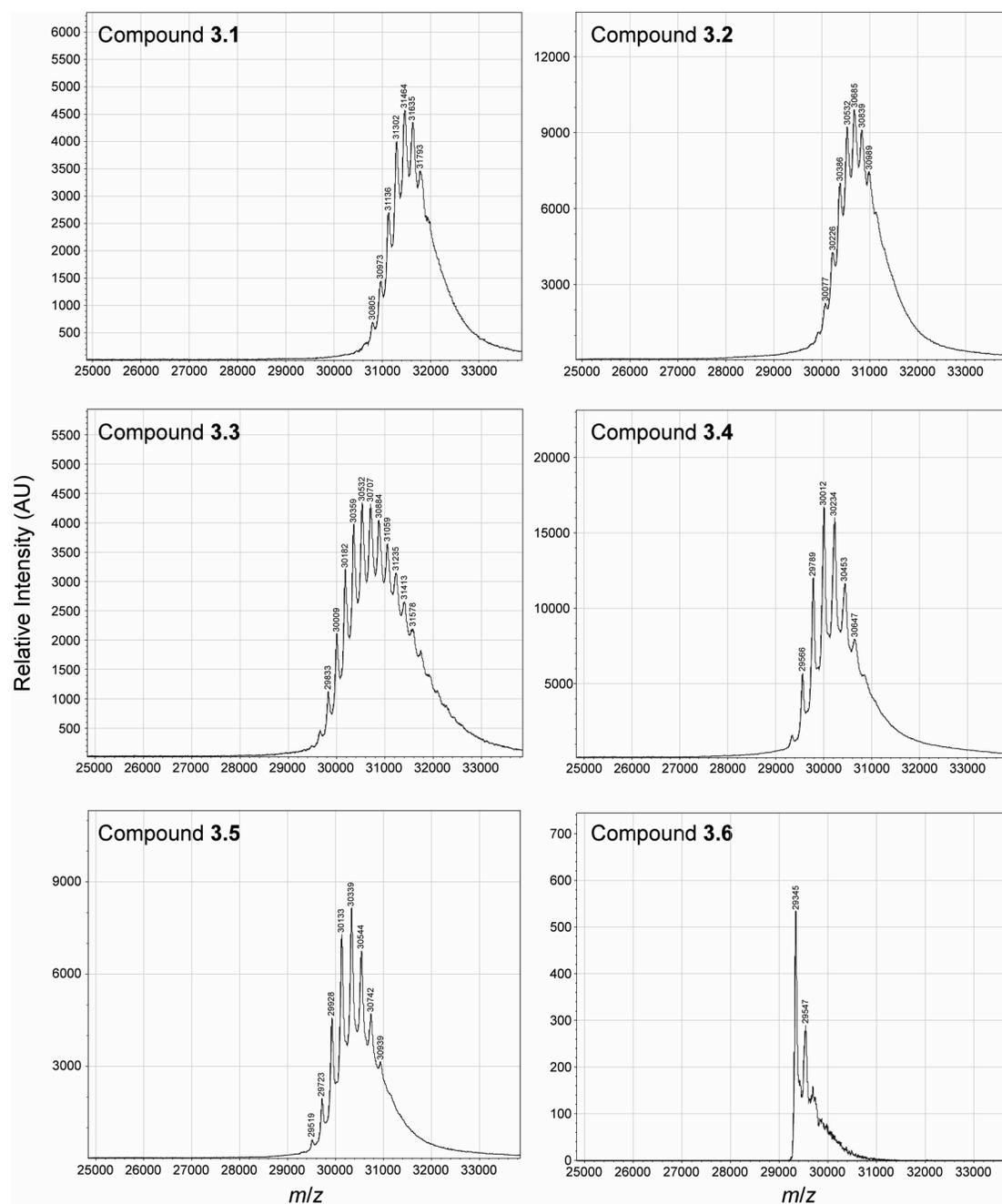


Figure 3.9 Representative MALDI-TOF spectra of GFP esterified with diazo compounds **3.1–3.6** (100 equiv, 3 equiv per carboxyl group) in 10 mM Bis-Tris buffer, pH 6.5, containing CH₃CN (20% v/v). Expected m/z : 29,343 + 175 per ester group.

Identification of GFP Carboxyl Groups Esterified with Diazo Compound 3.1

For each digest, a 10- μ g aliquot of protein solution was diluted with water to 100 μ L, and 1 μ L of 10% v/v aqueous formic acid was added. Immobilized pepsin (Thermo Fisher Scientific product 20343) was washed with 0.1% v/v formic acid according to the manufacturer's instructions and resuspended as a 50% slurry. A 50- μ L aliquot of the immobilized pepsin slurry was added to each sample. Samples were placed on a shaking incubator at 37 °C at 200 rpm and incubated for 2, 5, 10, or 20 min. Upon removal from the incubator, samples were subjected to centrifugation, and the supernatant was removed to quench the digestion.

Data were acquired on an Orbitrap Elite mass spectrometer equipped with a Thermo EasySpray column (15 cm \times 75 μ m, packed with 3- μ m PepMap C18 resin) and eluted over a 45-min gradient using solvents of 0.1% v/v formic acid in water (A) and 0.1% v/v formic acid in acetonitrile (B). A top-20 method was used to acquire MS/MS spectra on the 20 highest abundance precursors in each MS scan with dynamic exclusion of precursors that had been selected already within the preceding 30 s for MS/MS analysis.

Data were searched against an *E.coli* database to which was added the sfGFP sequence, and the +175 modification was allowed as a variable modification. Pepsin was used as the enzyme specificity with up to 4 missed cleavages per peptide. Precursor tolerance was set at 15 ppm, and MS/MS fragment ion tolerance was set to 0.5 Da (MS/MS data collected in the linear ion trap portion of the Orbitrap Elite).

Residues identified as being esterified are indicated in red:

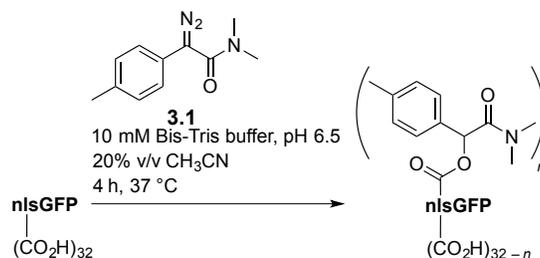
```

1
MHHHHHSSGVLDLGTENLYFQGMVSKGEEELFTGVVPILVLEDGDVNGHKFS
50
VRGEGEGDATIGKLTLLKFICTTGKLPVPWPTLVTTLLTYGVQCFSTRYPDHM
KQHDFFKSAMP EGYVQERTISFKDDGKYKTRAVVKFEGDTLVNRI ELKGT
DFKEDGNILGHKLEYNFNHSHNVYITADKQKNGIKANFTVRHNV EDGSVQL
ADHYQQNTPIGDGPVLLPDNHYLSTQTVLSKDPNEKRDHMLHEYVNAAG
ITLGMDELYK

```

Note: This GFP variant contains additional residues at its N and C termini relative to the protein used to calculate the electrostatic surface in Table 3.1; thus, 3 esterified carboxyl groups are not listed in Table 3.1.

Esterification of nlsGFP with Diazo Compound 3.1



A solution of diazo compound **3.1** (3.5 mg, 17.1 μmol) in acetonitrile (400 μL) was added to a solution of nlsGFP (0.341 μmol) in 1.6 mL of 10 mM Bis-Tris buffer, pH 6.5, and the resulting solution was incubated for 4 h at 37 °C. Precipitated protein was removed by filtration through a 0.2- μm PES syringe filter (GE Healthcare). Protein was purified and exchanged into PBS buffer using a PD10 desalting column (GE Healthcare). The number of esters per protein was assigned from the peak with the highest relative intensity in the MALDI-TOF mass spectrum (Figure 3.10). Protein was concentrated as needed, and the protein concentration was determined with a BCA assay (Thermo Fisher Scientific).

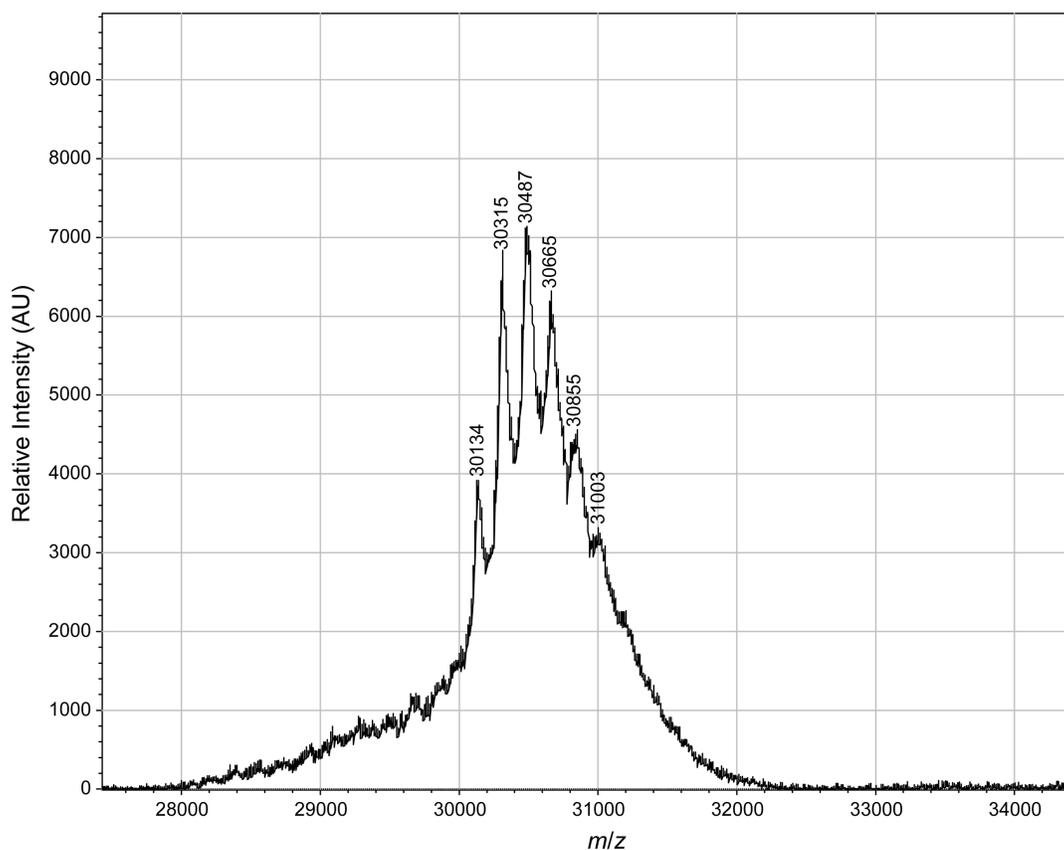


Figure 3.10 MALDI-TOF spectrum of nlsGFP esterified with diazo compound **3.1**. Expected m/z : 29,943 + 175 per ester group.

3.4.5 Mammalian Cell Culture

Chinese hamster ovary- (CHO-) K1 cells were from the American Tissue Culture Collection and cultured according to recommended protocols. Cells were grown in F12K nutrient medium supplemented with fetal bovine serum (10% v/v), penicillin (100 units/mL) and streptomycin (100 μ g/mL). Cells were grown in T75 sterile culture flasks in a cell culture incubator at 37 °C under CO₂ (5% v/v). Cells were counted to determine seeding density using a hemacytometer.

3.4.6 Flow Cytometry

Cells were seeded at a density of 50,000 cell/well in a sterile 8-well dish (Ibidi) 24 h prior to treatment. Cells were incubated with either unmodified GFP or GFP esterified with compounds **3.1–3.6** (15 μ M) in F12K medium supplemented with penicillin (100 units/mL) and streptomycin (100 μ g/mL) for 2 h at 37 °C. Cells were rinsed twice with DPBS, and released from the plate with 250 μ L of 0.25% v/v trypsin–EDTA mix. Trypsin was quenched by the addition of 500 μ L of medium, and cells were then subjected to centrifugation for 5 min at 130g. Cells were resuspended in 300 μ L of DPBS supplemented with fetal bovine serum (10% v/v). 7AAD stain (10 μ L of a 1.0 mg/mL solution) was added to each sample, and cells were kept on ice until the time of analysis. The fluorescence intensity of at least 10,000 events was measured by flow cytometry with an Accuri C6 flow cytometer (BD Biosciences). The median fluorescence intensity of live, single cells is reported.

Time-Dependence of GFP–3.1 Internalization

Cells were seeded at a density of 100,000 cells/well in a sterile 12-well dish (CellStar) 24 h prior to treatment. The cells were then incubated with GFP–**3.1** (4 μ M) in F12K medium supplemented with penicillin (100 units/mL) and streptomycin (100 μ g/mL) for 30, 120, or 240 min at 37 °C. Cells were rinsed with DPBS, and released from the plate with 250 μ L of 0.05% trypsin–EDTA. Trypsin was quenched by the addition of 250 μ L of medium. Propidium iodide was added to each sample (final concentration: 10 μ g/mL), and cells were kept on ice until the time of analysis. The fluorescence intensity of at least 10,000 events was measured by flow cytometry with a FACS Canto II HTS flow cytometer (BD Biosciences). The median fluorescence intensity of live, single cells is reported (Figure 3.1B).

3.4.7 Confocal Microscopy

Internalization of GFP-3.1

Cells were seeded at a density of 50,000 cell/well in a sterile 8-well dish (Ibidi) 24 h prior to treatment. Cells were incubated with either unmodified GFP or GFP-3.1 (15 μ M) in F12K medium supplemented with penicillin (100 units/mL) and streptomycin (100 μ g/mL) for 2 h at 37 °C. Cells were rinsed twice with DPBS, and nuclei were stained by incubation with Hoechst 33342 dye (2 μ g/mL) for 5 min at 37 °C. Cell membranes were stained by incubation with wheat germ agglutinin (WGA)-Alexa Fluor[®] 647 dye (5 μ g/mL) for 15 min on ice. Cells were then washed twice and kept in medium on ice until the time of analysis. Live cells were examined using a Nikon A1R+ scanning confocal microscope. The results are shown in Figure 3.4. Image acquisition and processing settings were maintained between all samples.

Internalization of GFP, Super-Charged GFP, and GFP-3.1 at 37 °C and 4 °C

Cells were seeded at a density of 50,000 cell/well in a sterile 8-well dish (Ibidi) 24 h prior to treatment. Cells were incubated with unmodified GFP, super-charged GFP, or GFP-3.1 (15 μ M) in F12K medium supplemented with penicillin (100 units/mL) and streptomycin (100 μ g/mL) for 2 h at either 37 or 4 °C. Cells were rinsed twice with DPBS, and nuclei were stained by incubation with Hoechst 33342 dye (2 μ g/mL) for 5 min at 37 °C. Cell membranes were stained by incubation with WGA-Alexa Fluor[®] 647 dye (5 μ g/mL) for 15 min on ice. Cells were then washed twice and kept in medium on ice until the time of analysis. Live cells were examined using a Nikon A1R+ scanning confocal microscope. Image acquisition and processing settings were maintained between all samples. The results are shown in Figure 3.4.

Internalization of nlsGFP and nlsGFP-3.1

Cells were seeded at a density of 50,000 cell/well in a sterile 8-well dish (Ibidi) 24 h prior to treatment. Cells were incubated with either unmodified GFP or nlsGFP-3.1 (15 μ M) in F12K medium supplemented with penicillin (100 units/mL) and streptomycin (100 μ g/mL) for 2 h at 37 $^{\circ}$ C. Cells were rinsed twice with DPBS, and nuclei were stained by incubation with Hoechst 33342 dye (2 μ g/mL) for 5 min at 37 $^{\circ}$ C. Cell membranes were stained by incubation with WGA-Alexa Fluor[®] 647 dye (5 μ g/mL) for 15 min on ice. Cells were then washed twice and kept in medium on ice until the time of analysis. Live cells were examined using a Nikon A1R+ scanning confocal microscope. Image acquisition and processing settings were maintained between all samples. The results are shown in Figure 3.5. Pearson's correlation coefficient (r) was calculated with the PSC colocalization plugin in ImageJ software.

3.4.8 Esterification Reversibility

Unlike GFP, human angiogenin is a small protein (15.3 kDa) that maintains its structure after incubation with a detergent-containing cell lysate and produces a well-resolved peak in a MALDI-TOF spectrum. Moreover, a FLAG-tagged variant of angiogenin²²⁸ binds to an anti-FLAG antibody with extremely high affinity, thus enabling high recovery of this protein from a cell lysate. Thus, we used angiogenin for a rigorous assessment of the bioreversibility of protein esterification with diazo compound **3.1**.

Diazo compound **3.1** (437 μ g, 2.2 μ mol) in acetonitrile (40 μ L) was added to a solution of FLAG-angiogenin (0.043 μ mol) in 160 μ L of 10 mM Bis-Tris buffer, pH 6.5. The resulting solution was incubated at 37 $^{\circ}$ C for 4 h, and the number of esters was determined with MALDI-

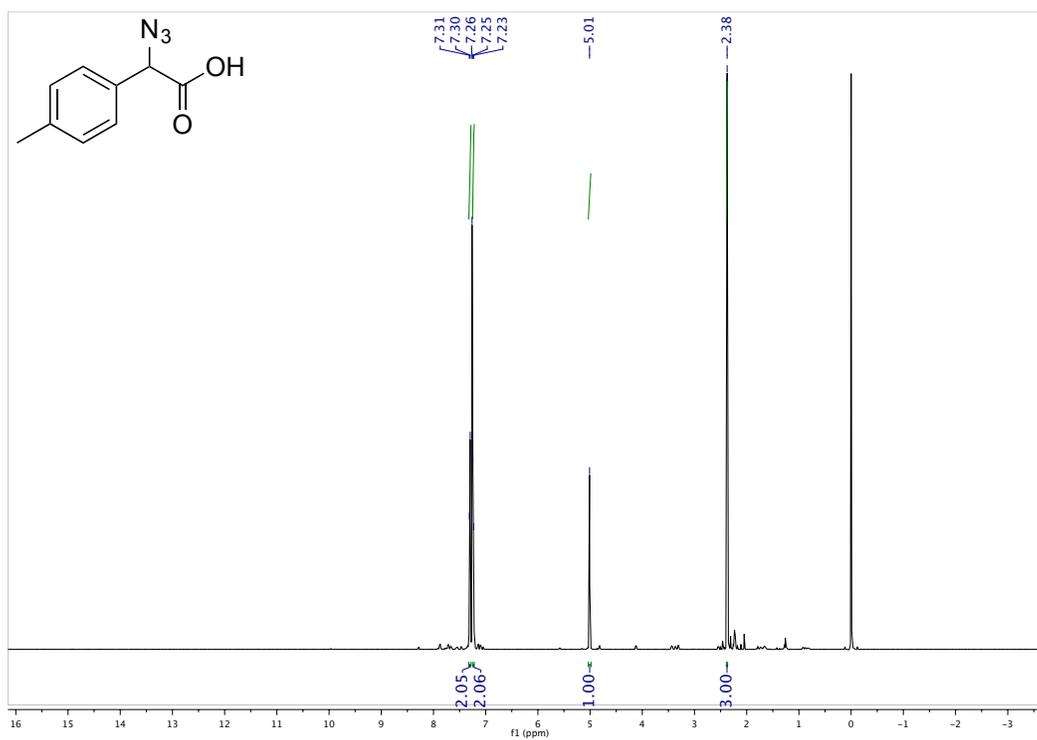
TOF mass spectrometry. Acetonitrile was removed using a Vivaspin filtration column (5,000 MWCO) from GE Life Sciences. The resulting solution was added to a CHO-cell extract (50 μ L), which was prepared by using CellLytic M lysis reagent (Sigma–Aldrich product C2978) supplemented with 1 \times protease inhibitor (Thermo Fisher Scientific product 78430). The solution was incubated at 25 $^{\circ}$ C overnight. FLAG–angiogenin was reisolated using anti-FLAG magnetic beads from Sigma–Aldrich and analyzed again with MALDI–TOF mass spectrometry (Figure 3.6).

3.4.9 Cytotoxicity Assay

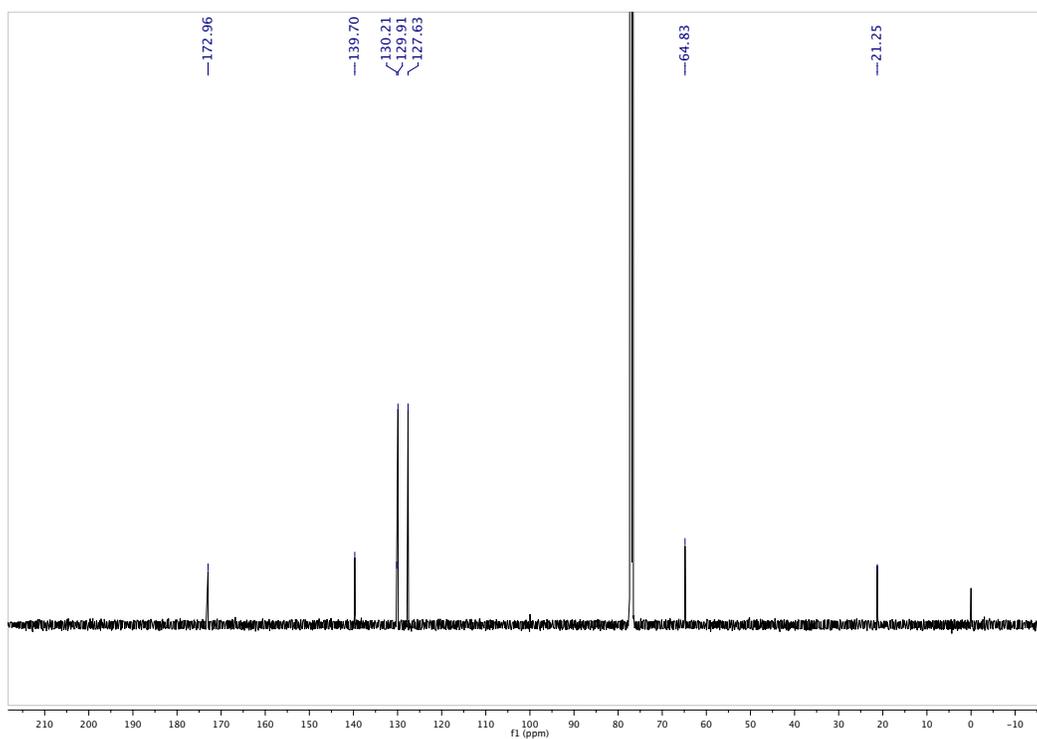
The cytotoxicity of compound **3.7** was measured with a CellTiter96[®] AQueous One Cell Proliferation (MTS) Assay from Promega according to the manufacturer’s instructions. Cells were plated at a density of 50,000 cells/well in a sterile 96-well plate 24 h prior to treatment. Cells were treated with either vehicle (1% v/v DMSO in medium) or compound **3.7** (50–500 μ M in 1% v/v DMSO in medium) for 2½ h. The medium was replaced, and 20 μ L of CellTiter96[®] AQueous One Solution Reagent was added to each well. Cells were incubated for 1 h, and absorbance at 490 nm was measured with a Tecan Infinite M1000 plate reader. Cell viability is expressed relative to vehicle control (Figure 3.7).

3.4.10 NMR Spectra

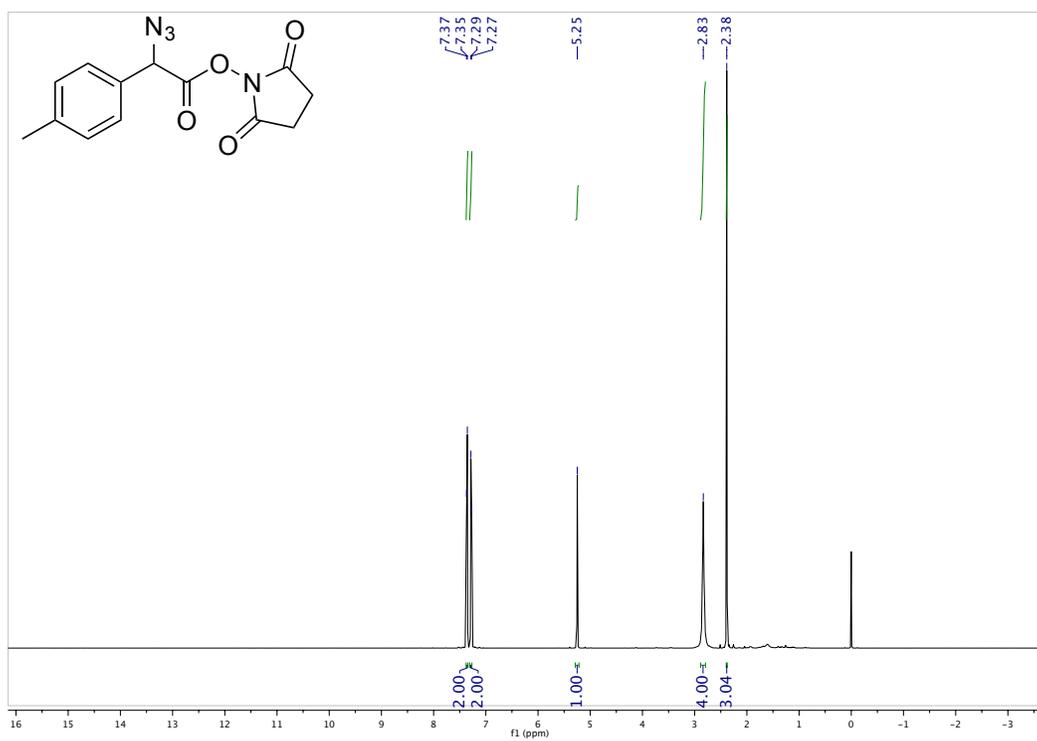
^1H NMR spectrum of compound **S1** in CDCl_3 (500 MHz):



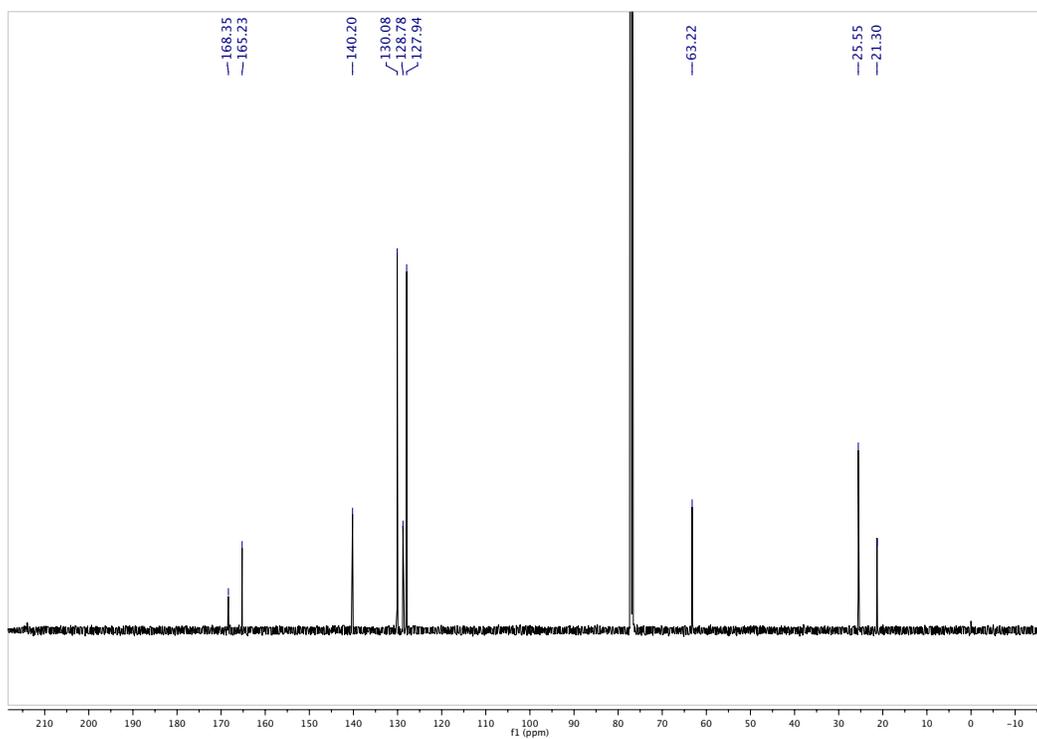
^{13}C NMR spectrum of compound **S1** in CDCl_3 (125 MHz):



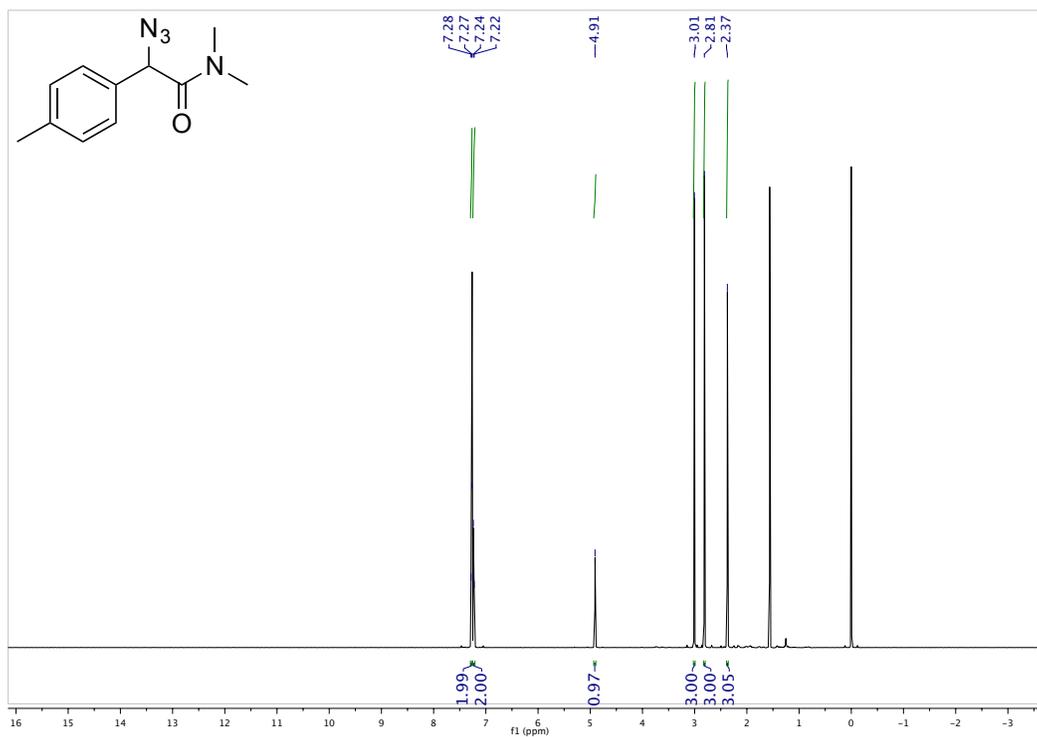
^1H NMR spectrum of compound **S2** in CDCl_3 (500 MHz):



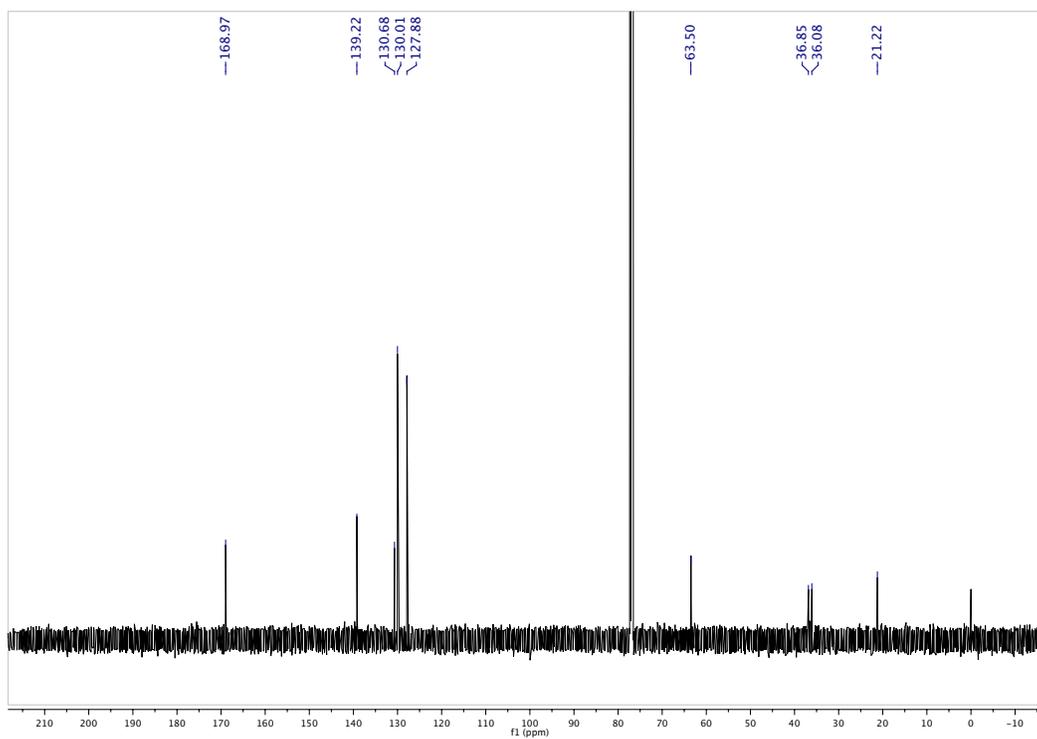
^{13}C NMR spectrum of compound **S2** in CDCl_3 (125 MHz):



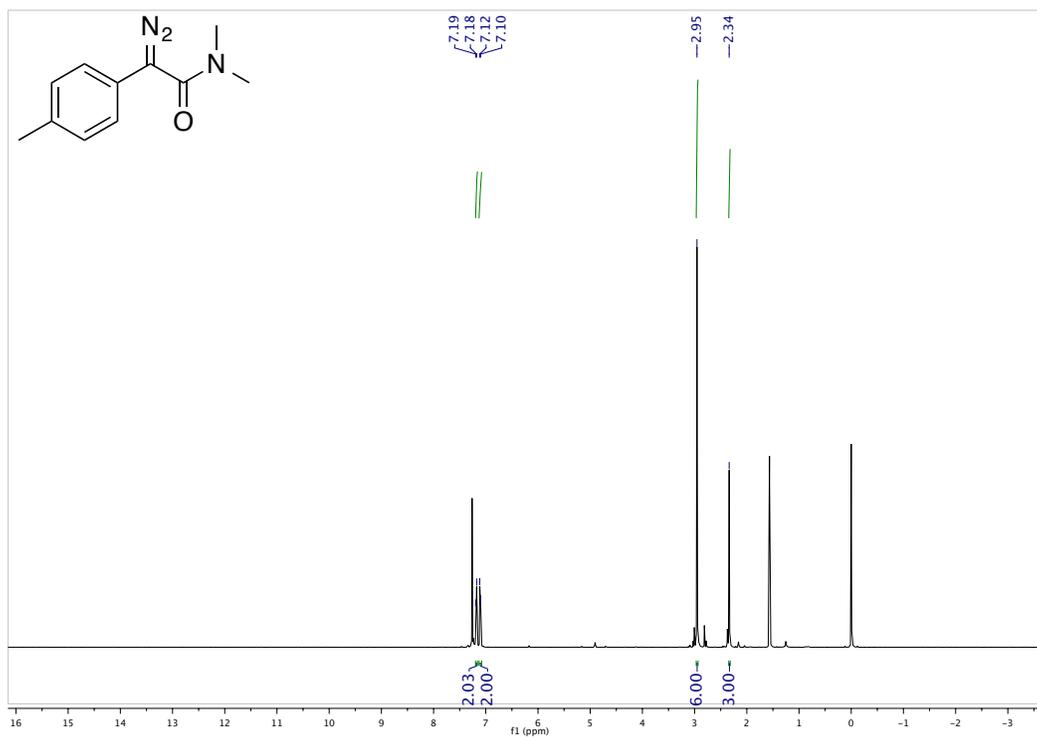
^1H NMR spectrum of compound **S3** in CDCl_3 (500 MHz):



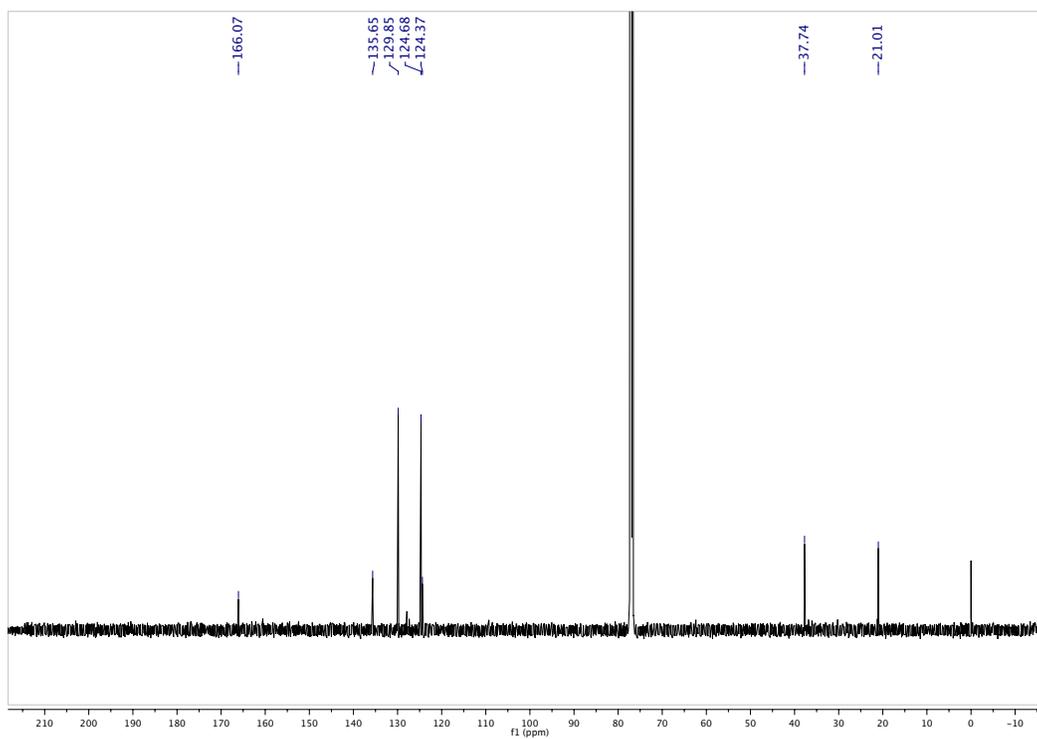
^{13}C NMR spectrum of compound **S3** in CDCl_3 (125 MHz):



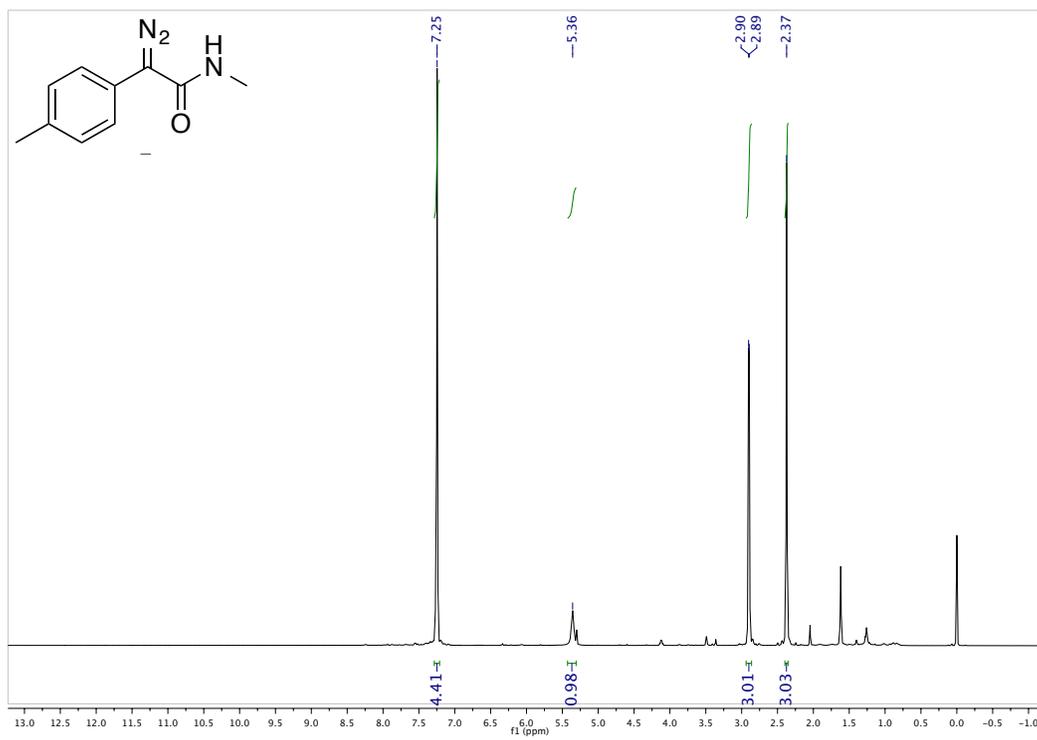
^1H NMR spectrum of compound **3.1** in CDCl_3 (500 MHz):



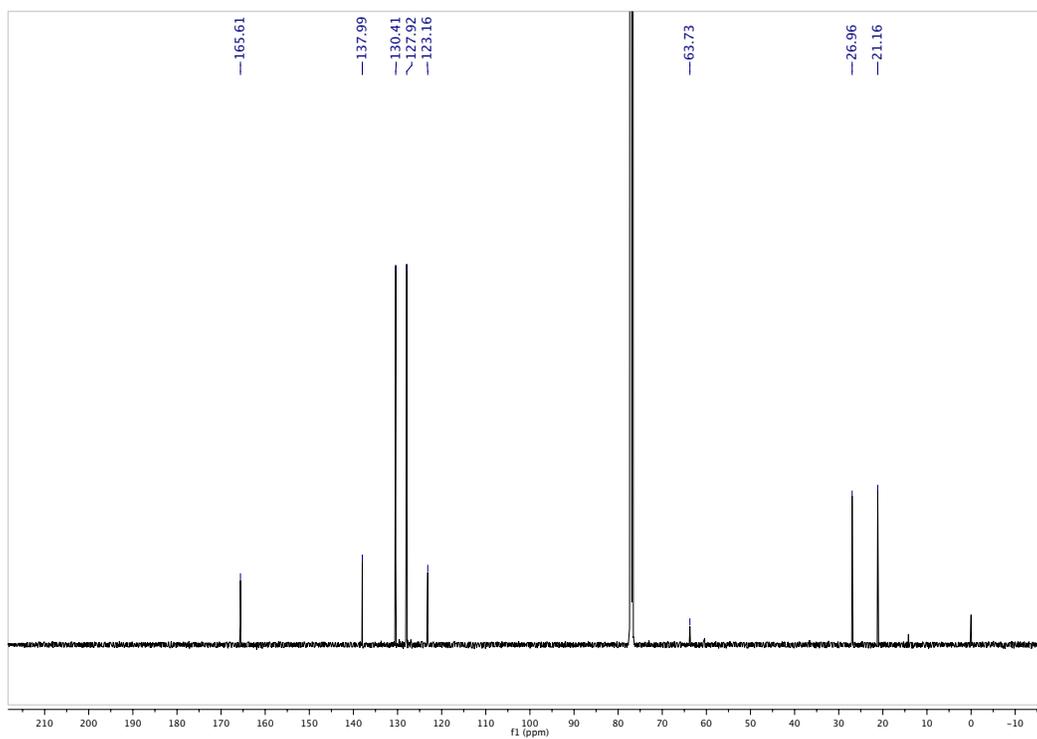
^{13}C NMR spectrum of compound **3.1** in CDCl_3 (125 MHz):



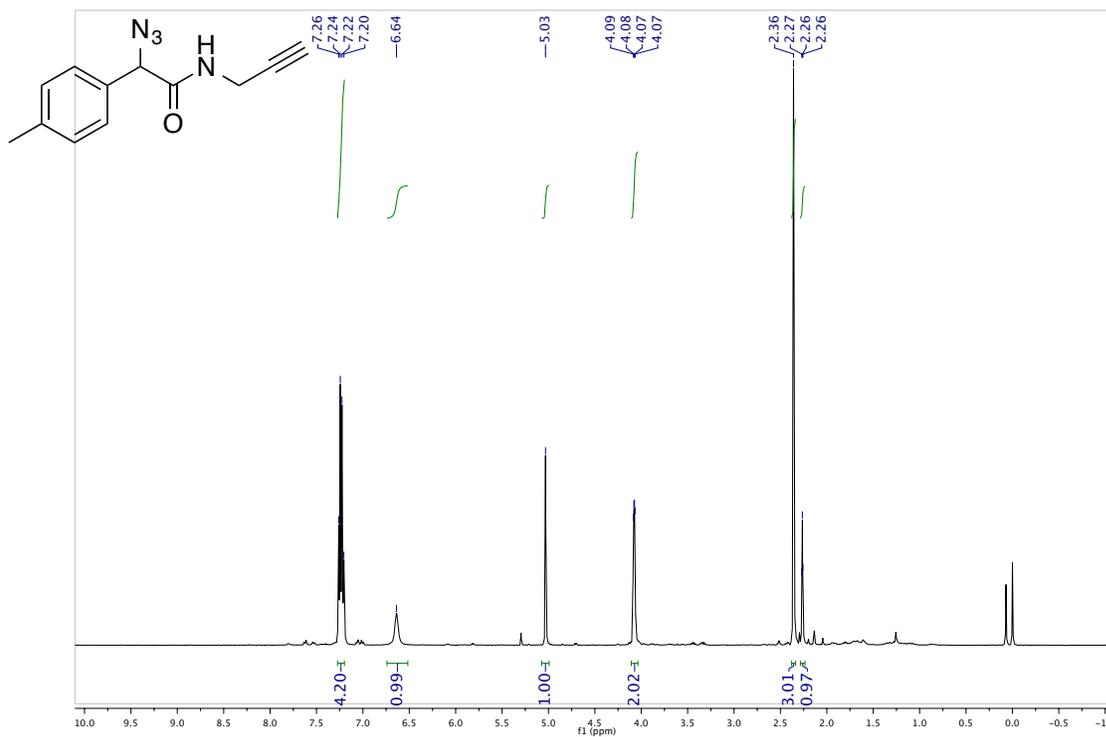
^1H NMR spectrum of compound **3.2** in CDCl_3 (500 MHz):



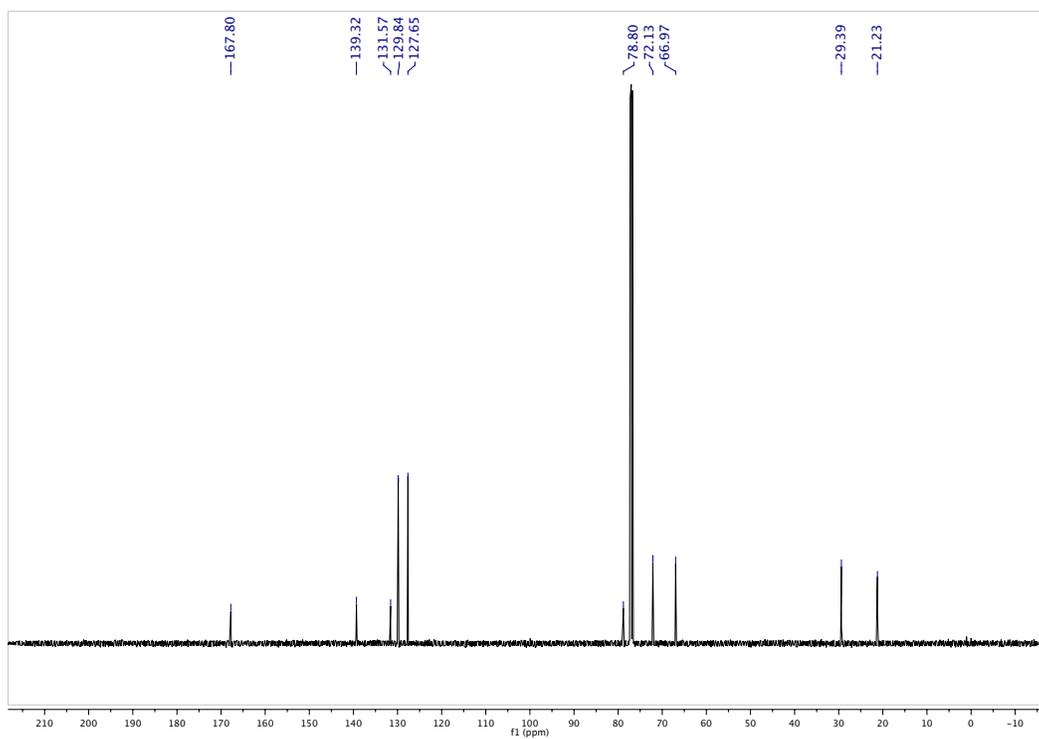
^{13}C NMR spectrum of compound **3.2** in CDCl_3 (125 MHz):



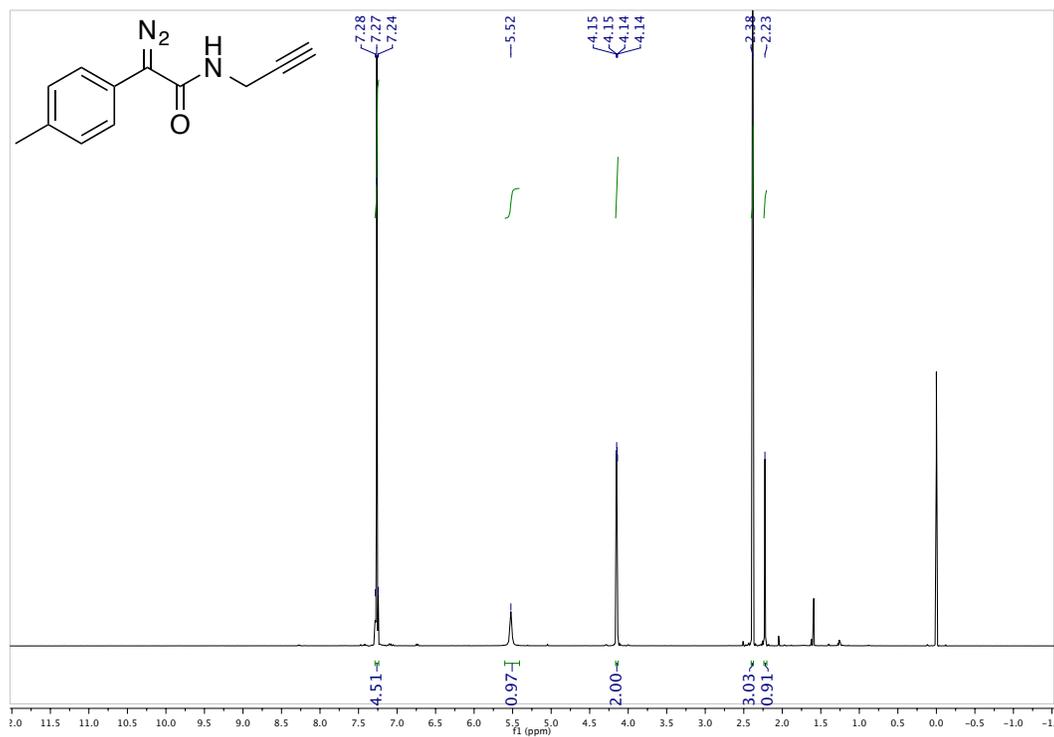
^1H NMR spectrum of compound **S4** in CDCl_3 (500 MHz):



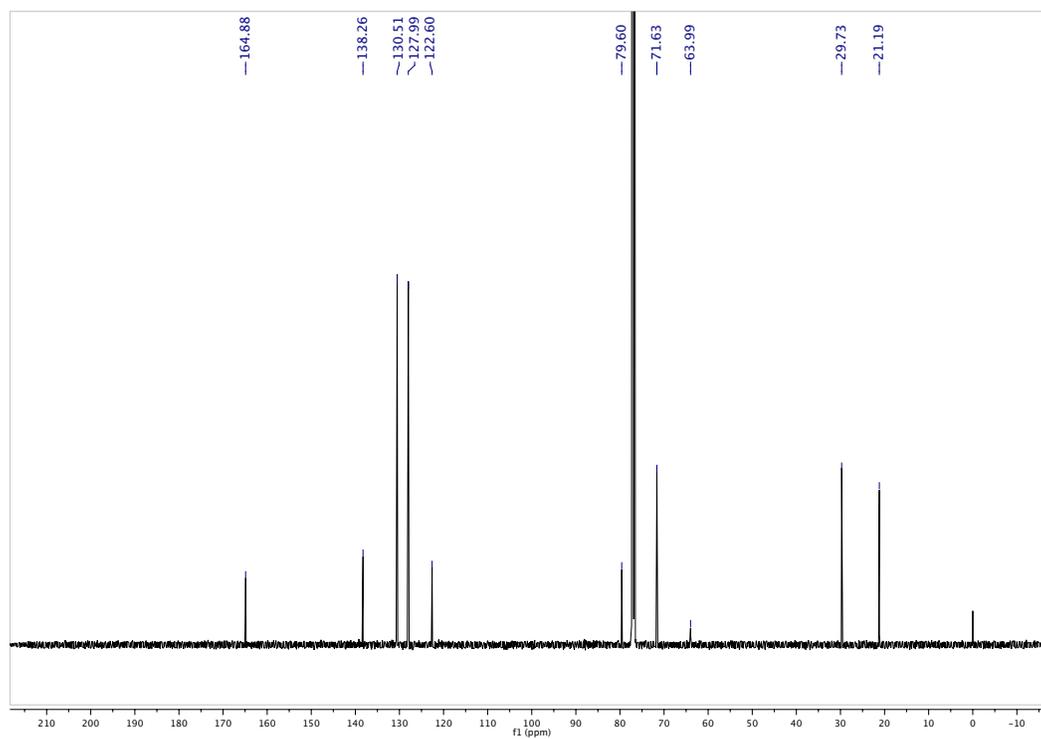
^{13}C NMR spectrum of compound **S4** in CDCl_3 (125 MHz):



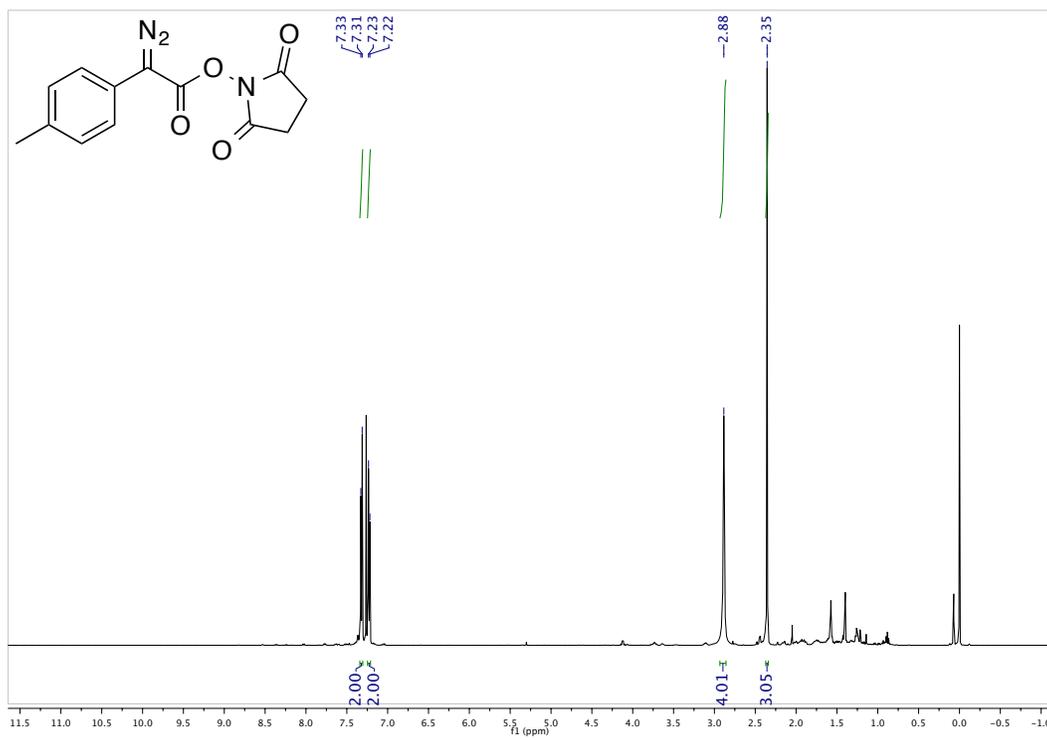
^1H NMR spectrum of compound **3.3** in CDCl_3 (500 MHz):



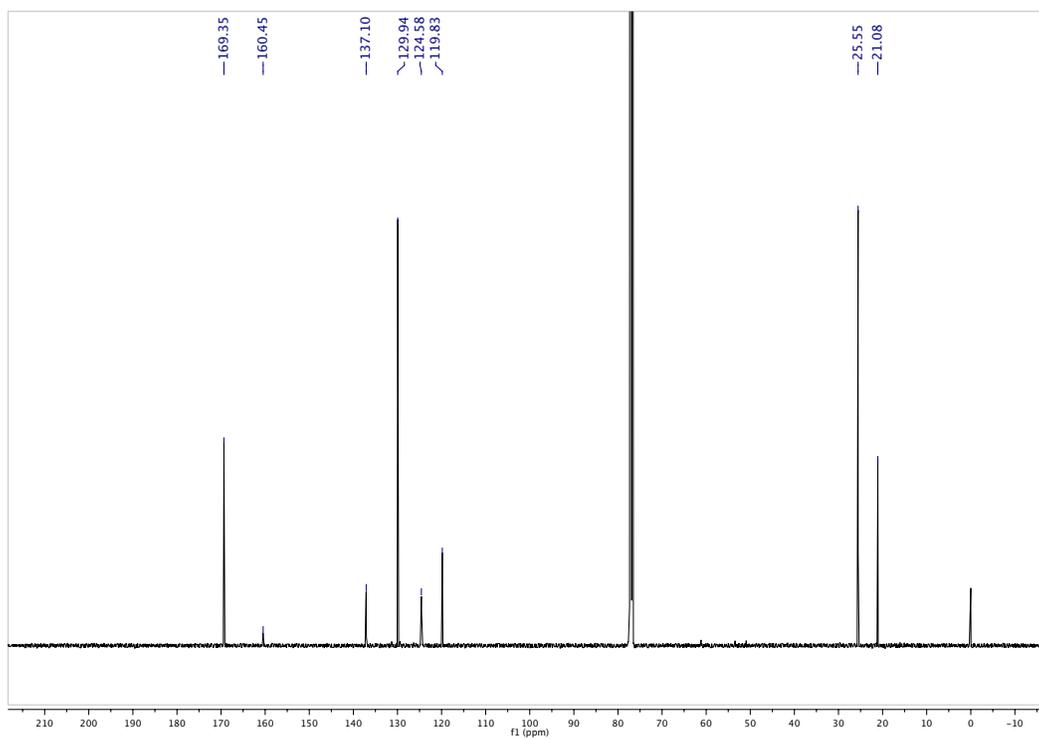
^{13}C NMR spectrum of compound **3.3** in CDCl_3 (125 MHz):



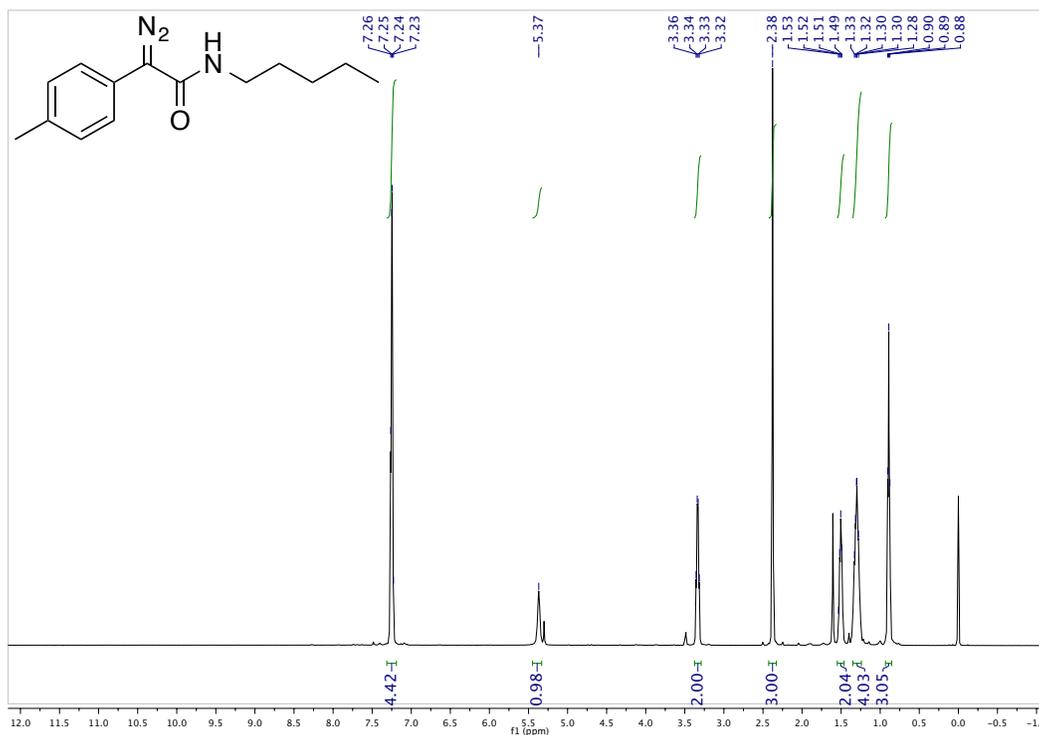
^1H NMR spectrum of compound **S5** in CDCl_3 (500 MHz):



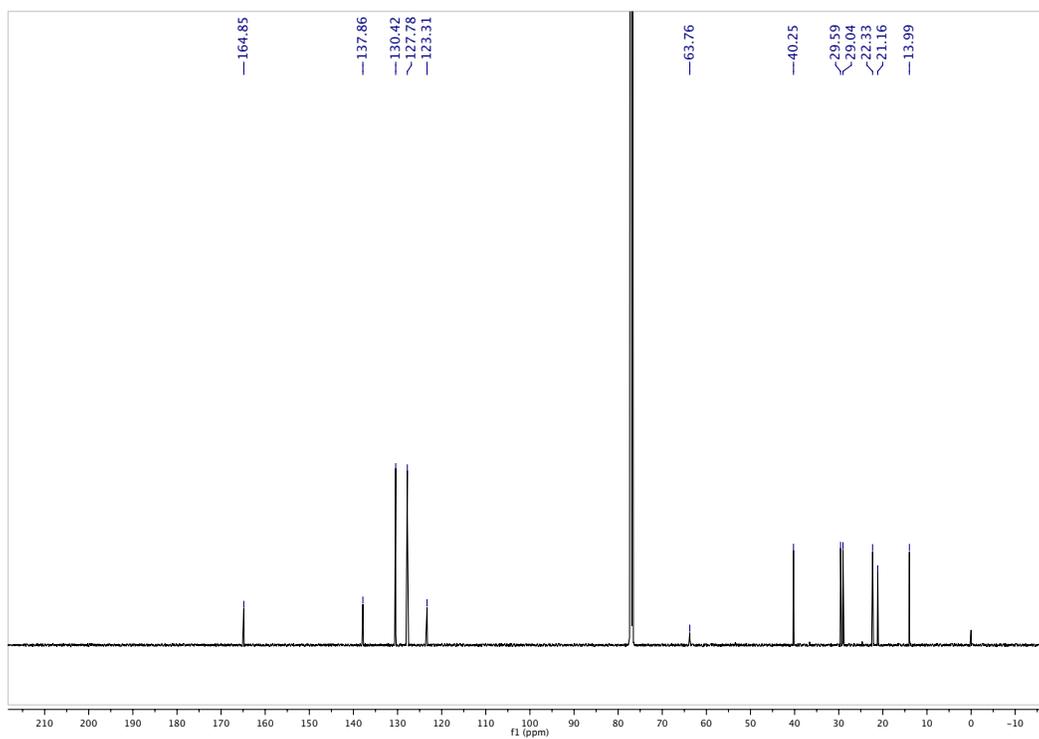
^{13}C NMR spectrum of compound **S5** in CDCl_3 (125 MHz):



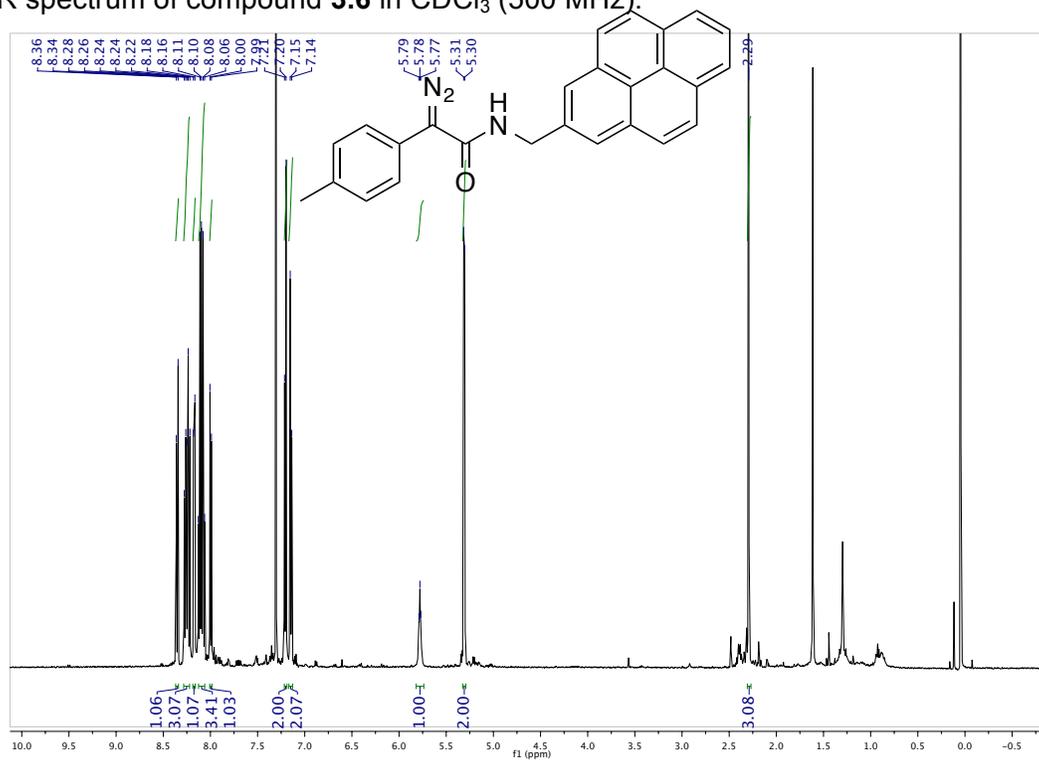
^1H NMR spectrum of compound **3.5** in CDCl_3 (500 MHz):



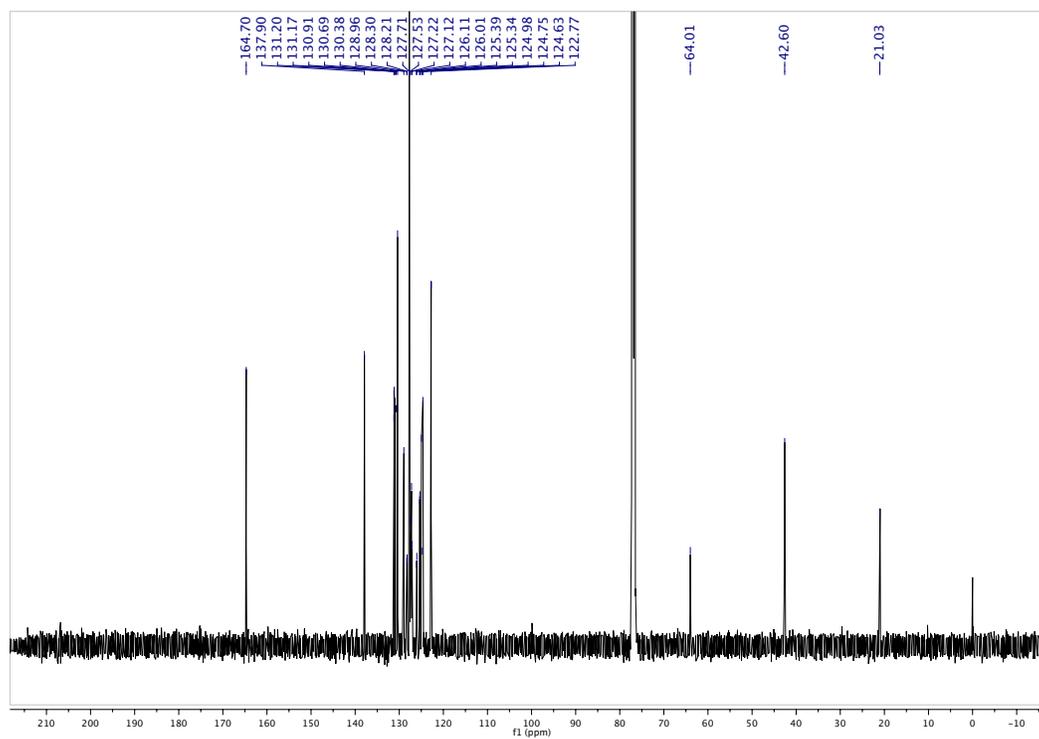
^{13}C NMR spectrum of compound **3.5** in CDCl_3 (125 MHz):



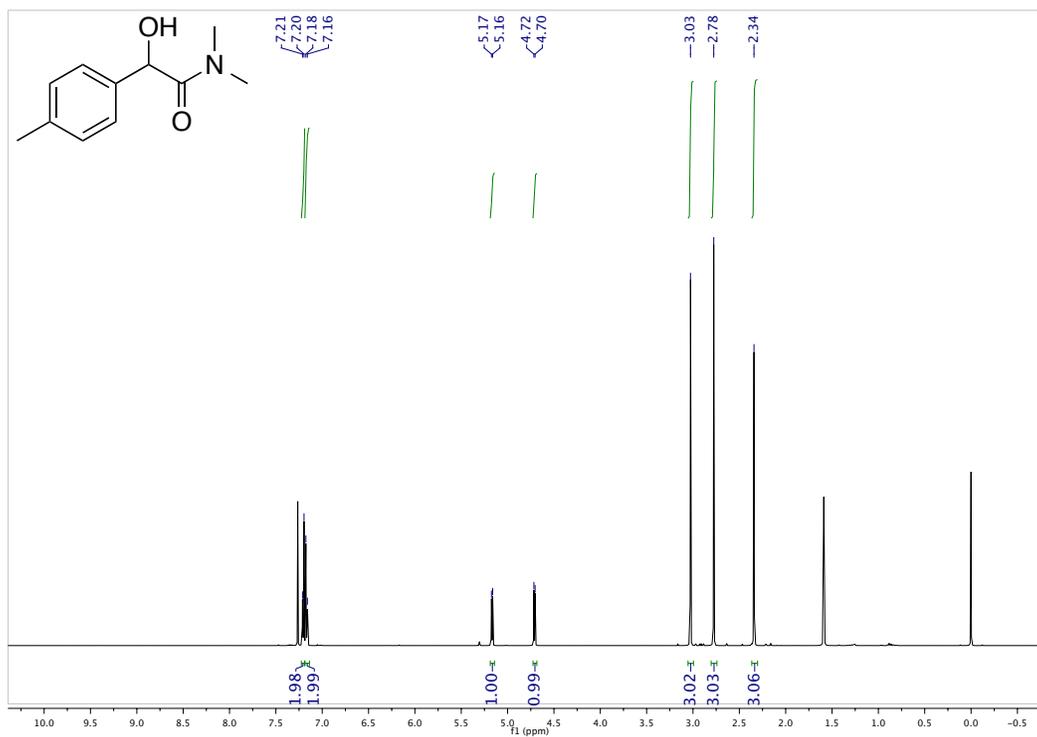
^1H NMR spectrum of compound **3.6** in CDCl_3 (500 MHz):



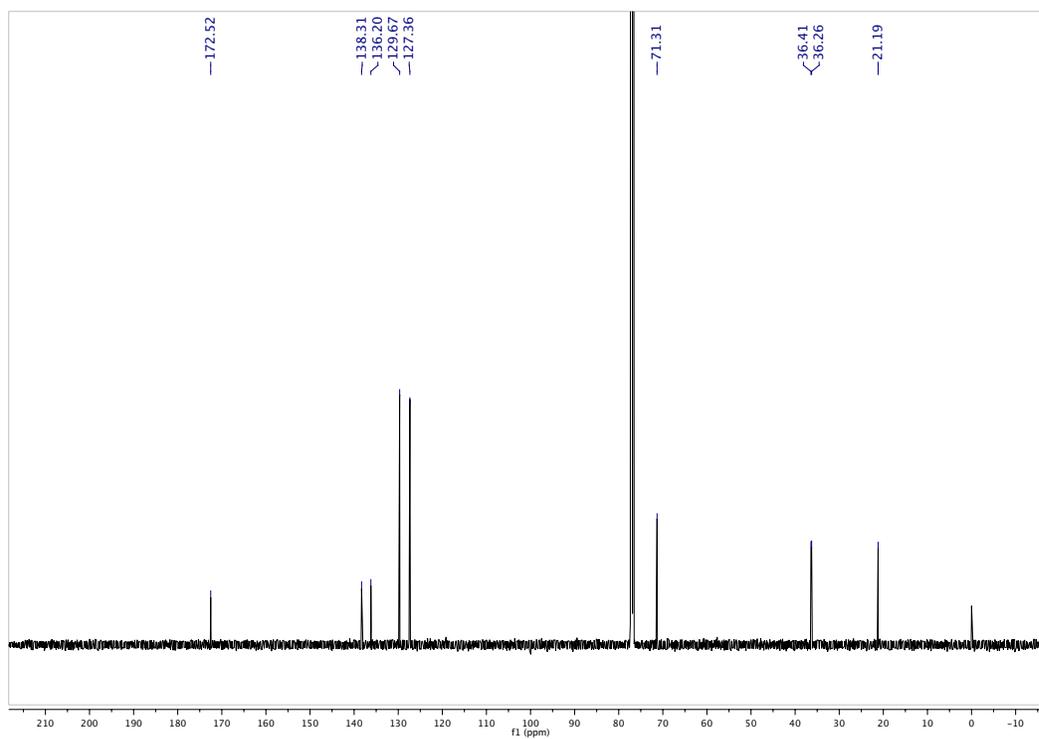
^{13}C NMR spectrum of compound **3.6** in CDCl_3 (125 MHz):



^1H NMR spectrum of compound **3.7** in CDCl_3 (500 MHz):

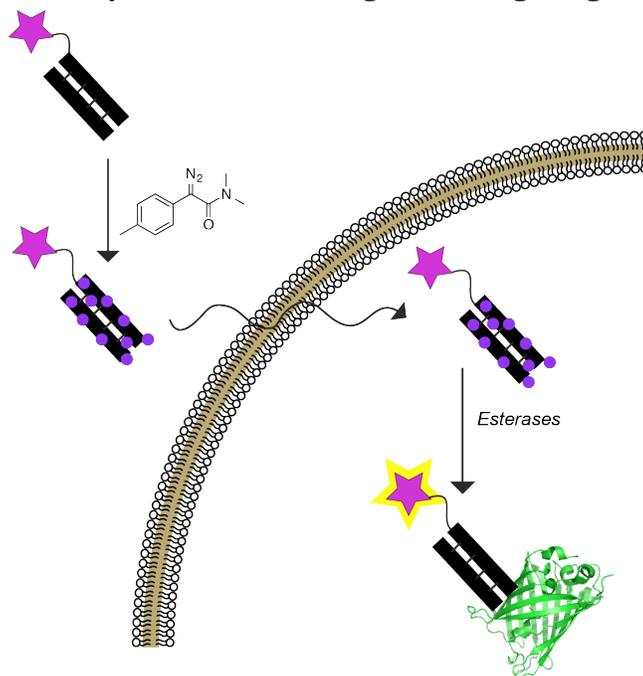


^{13}C NMR spectrum of compound **3.7** in CDCl_3 (125 MHz):



Chapter Four

Cellular Delivery of anti-GFP Antigen-Binding Fragment (Fab)



Abstract

Antibodies are valuable reagents for assaying cellular processes and have been highly impactful in the clinic as therapeutics. However, antibody-based assays are typically performed in a cell lysate or fixed cells, and antibody therapeutics recognize ligands at the cell surface. Here, we demonstrate that a diazo reagent esterifies the carboxyl groups of an anti-GFP antigen-binding fragment (Fab) and enables its cellular uptake in mammalian cells. This method is easily adapted to other Fabs of interest and could enable cellular delivery of therapeutic Fabs.

Author Contributions: Kalie A. Mix synthesized reagents, performed protein ligation, and performed cellular internalization experiments. Amy M. Weeks produced the Fab and performed protein ligation. Nadine Elowe performed mass spectrometry to characterize esterification. Kalie A. Mix, Amy M. Weeks, James A. Wells, and Ronald T. Raines designed experiments. Kalie A. Mix and Ronald T. Raines wrote this chapter.

4.1 Introduction

Antibodies are essential tools for probing cellular processes, and they are a rapidly-growing class of therapeutic agents.²²⁹ Antibodies (and other protein therapeutics) have been successful in the clinic because they can modulate targets that are inaccessible to most small-molecule drugs, such as protein–protein or protein–nucleic acid interactions. They also benefit from higher specificity than small molecules, and as such, typically achieve faster time-to-market.¹⁹⁸

In addition to natural immunoglobulin G (IgG)-derived antibodies, a number of high-affinity antibody-derived proteins have been developed that are endowed with favorable properties such as small size, high stability, and ability to be produced recombinantly in high quantities. These include the camelid-derived nanobodies,²³⁰ IgG-derived antigen-binding fragments (Fabs),²³¹ and IgG-derived single-chain variable fragments (scFv).²³² The vast majority of FDA- and EMA-approved antibodies consists of IgGs and Fabs and chemically or recombinantly linked dimers of these molecules. Fabs are also especially valuable constructs because phage display enables rapid high-throughput screening and production of Fabs that recognize virtually any antigen of interest.²³³

Current technologies are limited by the requirement that the antibody have an extracellular target. For example, therapeutic antibodies recognize a receptor or ligand on the cell surface, and antibodies in biological assays (*e.g.*, for immunoblotting or detection of post-translational modifications), are limited to experiments performed *in vitro* or in a cell lysate. The ability to target antibodies to the cytosol would enable modulation of the many protein–protein interactions and signaling pathways present in the cytosol. Cytosolic antibody delivery would be highly valuable for basic science endeavors that seek to elucidate the role of protein interactions

in a dynamic live-cell environment. This achievement would also enable therapeutic strategies that seek to inhibit or augment these interactions.

A number of strategies to localize antibodies to the cytosol have been attempted with little success.²³⁴ Antibody folding is facilitated by the formation of disulfide bonds and protein secretion through the endoplasmic reticulum (ER), which contains a number of chaperone proteins. Consequently, attempts to express antibodies in the reducing environment of the cytosol that lacks these chaperone proteins have resulted in misfolded or nonfunctional protein, or simply a failure to express.²³⁵ Another approach to localize antibodies to the cytosol is via cellular delivery of the folded recombinant protein. In one study, surface residues of camelid-derived nanobodies were mutated to arginine or lysine in a manner reminiscent of arginine grafting¹⁹⁹ to promote favorable Coulombic interactions with the anionic plasma membrane components.²⁰¹ This resurfacing enabled an anti-GFP nanobody to enter cells. Nevertheless, application of this strategy to other nanobodies that bind to other antigens would require re-engineering of the protein, which is not high-throughput. Additionally, nanobodies are very small proteins (~13 kDa) that have poor pharmacokinetic properties with rapid clearance from the bloodstream.²³⁶ A similar delivery strategy in which the antibody is endowed with cationic character to promote cellular uptake is the fusion of antibodies to cell-penetrating peptides (CPPs), such as polyarginine, TAT peptide, or their derivatives. Though the mechanism of uptake of CPPs and CPP–protein fusions is the subject of debate, many CPP-fusion proteins are internalized primarily via endocytosis, and thus only a small fraction reaches the cytosol when attached to protein cargo.²³⁷ Antibody–CPP fusions have also been relatively unsuccessful, due to low production yields. Most antibodies need to be secreted to express and fold properly, and fusion to a CPP limits secretion efficiency.²³⁸ The production challenge has been partially solved

by the ligation of CPP to intact, recombinant protein using chemical conjugation methods. Still, these constructs require treatment of cells with a high protein concentration to observe protein delivery.^{207,239}

Cellular delivery of antibodies by esterification with a diazo reagent could overcome many of the challenges encountered by previous technologies. In this method, a recombinant antibody is modified, so issues of production and proper folding are circumvented. Because the diazo reagent reacts with carboxyl groups, which are present in all native proteins, no protein engineering is required. Additionally, the ester labels are labile in an intracellular milieu, so it is unlikely that the labels would interfere with antigen binding. We demonstrate here that diazo compound **3.1** esterifies an anti-GFP Fab and enables its delivery into mammalian cells at low (5 μ M) concentration.

4.2 Results and Discussion

The anti-GFP Fab was generated using phage display technology.²³³ An azido group was installed using a new protein ligation chemistry, redox-activated chemical tagging (ReACT).²⁴⁰ In this method, an oxaziridine reagent bearing an azido group reacts chemoselectively with a C-terminal methionine on the protein (Figure 4.1A). This azido group was then used as a handle to conjugate the Cy3 dye to the Fab using strain-promoted azide–alkyne cycloaddition (SPAAC) with a DIBAC–Cy3 conjugate (Figure 4.1A,B).

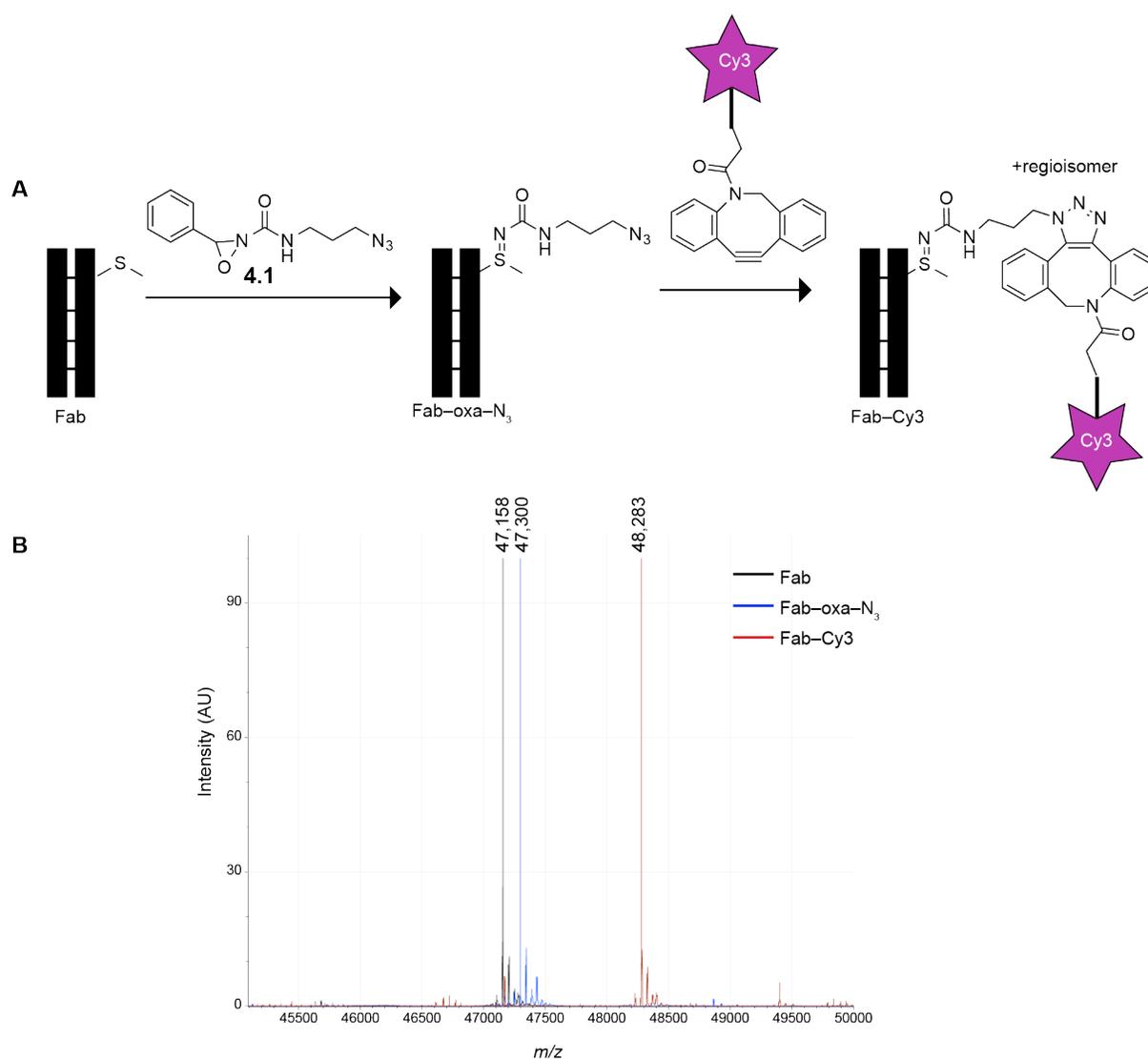


Figure 4.1 Modification of anti-GFP Fab with Cy3 dye. (A) Scheme depicting Cy3 ligation. ReACT chemistry was used to ligate the oxaziridine-azide **4.1** to a C-terminal methionine. Cy3 dye was then incorporated via SPAAC between the azide and a DIBAC-Cy3 reagent (linker structure not shown). (B) Mass spectra of Fab, Fab-oxa- N_3 , and Fab-Cy3. Fab: expected m/z 47,158; found 47,158. Fab-oxa- N_3 : expected m/z 47,299; found 47,300. Fab-Cy3: expected m/z 48,282; found 48,283.

Next, we labeled the Fab carboxyl groups using diazo reagent **3.1** (Figure 4.2A). This protein contains 32 carboxyl groups, which is similar to the superfolder GFP variant described in Chapter Three. This Fab is, however, much more cationic than the GFP variant at physiological pH, which makes esterification less efficient because cationic side chains that neighbor a carboxyl group lower its pK_a . This property is expected to lower the overall efficiency of esterification as carboxyl groups must be protonated to be esterified.¹²⁷ Due to this anticipated reduction in efficiency, we modified the esterification conditions by increasing the reaction time and the equivalents of diazo compound (4 h was extended to 24 h, and 100 equiv was increased to 200 or 400 equiv; Figure 4.2B). The maximum number of labels per protein was measured using mass spectrometry (Figure 4.2B). The highest number of labels (9 per protein) was observed when the Fab was treated with 400 equiv of diazo compound **3.1** for 24 h at 22 °C. These conditions did not result in any measurable protein precipitation, and thus were used for subsequent experiments to produce Fab–Cy3–**3.1** for cell-uptake experiments.

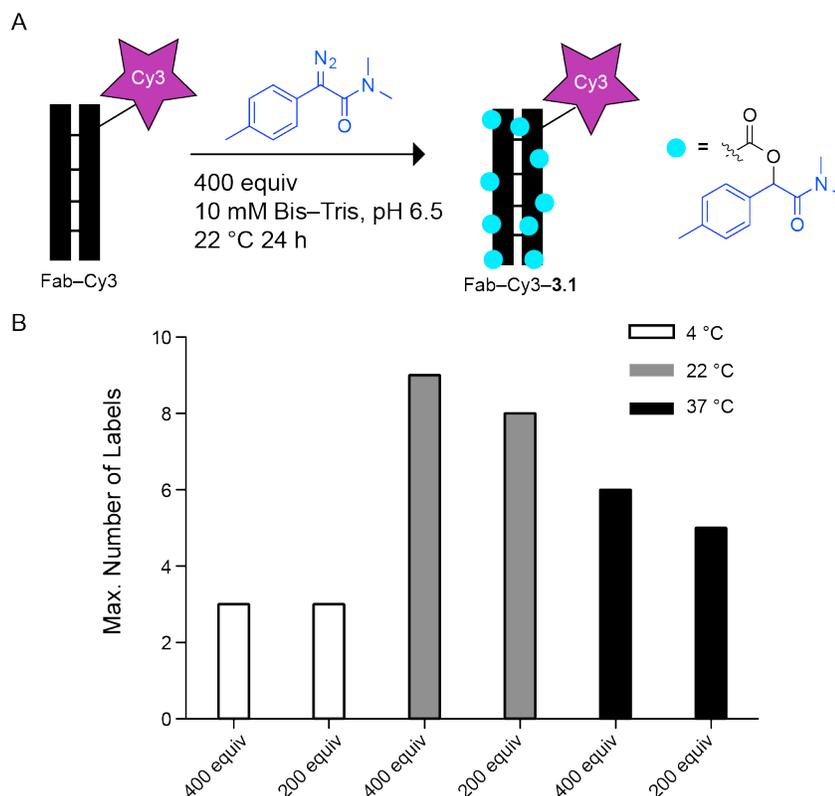


Figure 4.2 (A) Scheme depicting modification of Fab–Cy3 with diazo compound **3.1**. (B) Maximum number of labels added per protein under various esterification conditions. The number of labels was determined by mass spectrometry.

We then used flow cytometry to measure uptake of Fab–Cy3–**3.1** in mammalian cells. Chinese hamster ovary (CHO-K1) cells were treated with Fab–Cy3 or Fab–Cy3–**3.1** (5 μ M), and Cy3 fluorescence of live, single cells was measured (Figure 4.3). Fab–Cy3–**3.1** engendered a 6-fold increase in fluorescence relative to Fab–Cy3, which indicates that the esterified Fab was internalized in cells.

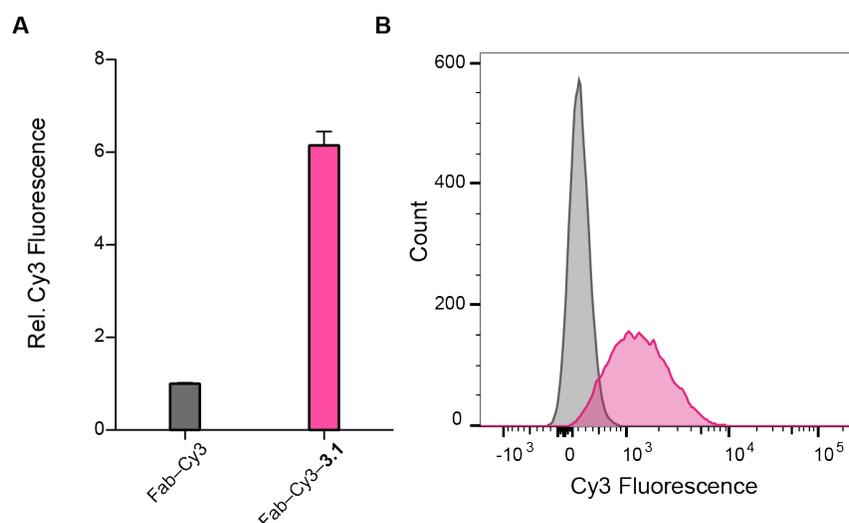


Figure 4.3 Quantification of cell uptake using flow cytometry. CHO-K1 cells were incubated with 5 μ M Fab-Cy3 or Fab-Cy3-**3.1** for 4 h. (A) Relative Cy3 fluorescence of CHO-K1 cells treated with either Fab-Cy3 or Fab-Cy3-**3.1**. The median fluorescence intensity of 10,000 live, single-cell events is reported. (B) Representative histogram of Cy3 fluorescence of cells treated with Fab-Cy3 (gray) or Fab-Cy3-**3.1** (pink).

Next, we used confocal microscopy to visualize uptake of Fab-Cy3-**3.1** in live mammalian cells (Figure 4.4). Cells treated with Fab-Cy3 demonstrate no visible Cy3 fluorescence. In contrast, cells treated with Fab-Cy3-**3.1** display strong pink fluorescence that concentrates primarily in the nucleolus. This finding was initially surprising because nucleolar accumulation was not observed upon cellular delivery of esterified GFP (see Chapter Three).²⁴¹ Previous studies have, however, shown that the propensity of a protein to accumulate in the nucleolus increases with increasing cationicity and isoelectric point.²⁴² This observation is attributable to the primary component of the nucleolus being anionic ribosomal RNA, and cationic proteins or peptides accumulating in the nucleolus due to Coulombic attraction. The calculated isoelectric point of the GFP variant used for experiments in Chapter Three is 6.0, and the calculated isoelectric point of the anti-GFP Fab used in these experiments is 8.7. Esterification of each of

these proteins increases their isoelectric point, and the relative change for each protein is difficult to calculate with certainty because of its dependence on the number and position of the ester labels. Still, the anti-GFP Fab is likely to maintain a higher isoelectric point than GFP, which could endow it with the ability to translocate from the cytoplasm to the nucleolus.

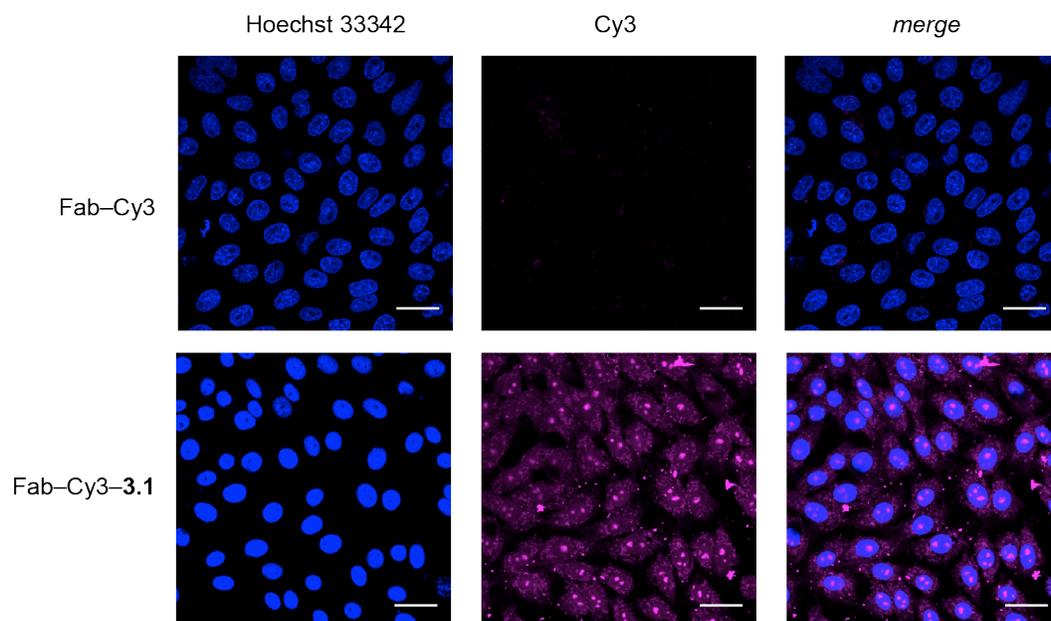


Figure 4.4 Confocal microscopy images of CHO-K1 cells treated with Fab-Cy3 (top) or Fab-Cy3-3.1 (bottom). Scale bars; 25 μm .

These data indicate that esterification of Fab-Cy3 enables its uptake into mammalian cells. Still, a number of follow-up experiments must be performed. A critical criterion for antibody delivery is that the antibody maintains its ability to bind to its antigen after modification and delivery. Confocal microscopy experiments will be repeated using HEK293-GFP cells, which stably express GFP in the cytosol. If the Fab-Cy3-3.1 protein maintains its ability to bind GFP after labeling and cellular delivery, GFP-Cy3 FRET should be observed. A Fab-Cy3 protein that does not bind to GFP will be used as a control to correct for Cy3 fluorescence that results from

coincidental proximity of the FRET pair. If aspartic or glutamic acid residues are required for antigen binding, bioreversibility will be especially important. Experiments that characterize bioreversibility will be performed by incubating Fab–Cy3–**3.1** with cell lysate and measuring ester cleavage by mass spectrometry.

The ability to deliver Fabs into cells would be transformative for protein therapeutics because it could enable disruption of intracellular signaling pathways. One such pathway that would be especially valuable to disrupt is Ras GTPase signaling. Ras plays a critical role in regulation of cell proliferation, and mutations to the gene encoding Ras are pervasive in a number of tumors.²⁴³ In future experiments, we plan to modify an anti-Ras Fab with diazo compound **3.1** and incubate it with various cancerous mammalian cell lines. Effective inhibition of Ras signaling will be measured by examining phosphorylation of proteins downstream in this pathway by immunoblotting.

4.3 Acknowledgments

K.A.M. was supported by Molecular Biosciences Training Grant T32 GM007215 (NIH) and a grant from the Broad Institute. A.M.W. is a Merck Fellow of the Helen Hay Whitney Foundation. This work was supported by grant R01 GM044783 (NIH).

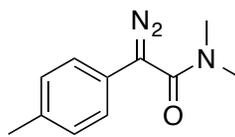
4.4 Materials and Methods

4.4.1 General

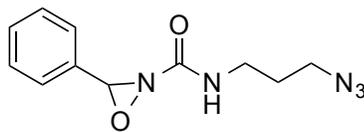
Silica gel (40 μm ; 230–400 mesh) was from SiliCycle. Reagent chemicals were obtained from commercial sources and used without further purification. Dichloromethane and tetrahydrofuran were dried over a column of alumina. Thin-layer chromatography (TLC) was performed on plates of EMD 250 μm silica 60-F₂₅₄. The phrase “concentrated under reduced pressure” refers to

the removal of solvents and other volatile materials using a rotary evaporator at water aspirator pressure (<20 torr) while maintaining a water bath below 40 °C. Residual solvent was removed from samples at high vacuum (<0.1 torr). ¹H and ¹³C NMR spectra for compound **3.1** and its precursors were acquired on Bruker spectrometers in the National Magnetic Resonance Facility at Madison operating at 400 or 500 MHz. ¹H and ¹³C NMR spectra for compound **4.1** and its precursors were acquired on Bruker AV-600, DRX-500, AV-500, AVQ-400, AVB-400 and AV-300 spectrometers. Chemical shift values (δ) are reported in units of ppm relative to an internal standard (residual solvent or TMS). Electrospray ionization (ESI) mass spectrometry for small-molecule characterization was performed with a Micromass LCT at the Mass Spectrometry Facility in the Department of Chemistry at the University of Wisconsin–Madison. Azide and Cy3 modifications were characterized by LC–MS analysis on a Xevo G2-XS mass spectrometer equipped with a LockSpray (ESI) source and Acquity Protein BEH C4 column (2.1 mm inner diameter, 50 mm length, 300 Å pore size, 1.7 μ m particle size) connected to an Acquity I-class liquid chromatography system (Waters). Deconvolution of mass spectra was performed with the maximum entropy (MaxEnt) algorithm in MassLynx 4.1 (Waters). Ester modifications were characterized by LC–MS analysis on a Q Exactive mass spectrometer (Thermo Fisher Scientific) at the Broad Institute. Confocal microscopy was performed using a Zeiss LSM 700 laser scanning confocal microscope with 405 nm and 555 nm excitation and 40X oil objective (W.M. Keck Biological Imaging Facility, Massachusetts Institute of Technology). Flow cytometry was performed using a FACSCanto HTS-II flow cytometer at the Koch Institute for Integrative Cancer Research at Massachusetts Institute of Technology.

4.4.2 Chemical Synthesis

**3.1**

Diazo reagent **3.1** was synthesized as described previously.²⁴¹ Yields and spectral data match those reported previously.

**4.1**

Oxaziridine-azide reagent **4.1** was synthesized as described previously.²⁴⁰ Yields and spectral data match those reported previously.

4.4.3 Production of anti-GFP Fab

Anti-GFP Fab was produced as described previously.^{233,240} Briefly, anti-GFP Fab was produced using phage display methods and its encoding DNA was inserted into a pSVF4 expression vector. Recombinant Fab was produced in C43 (DE3) Pro+ cells grown to an OD₆₀₀ ~0.6 and induced with 0.2 mM IPTG at 30 °C overnight. Fab was purified by protein A-affinity chromatography and dialyzed into PBS buffer.

4.4.4 Modification of Fab with oxaziridine-azide 4.1

Anti-GFP Fab was modified with oxaziridine-azide probe **4.1** as described previously.²⁴⁰ Fab was diluted to 1 mg/mL in PBS buffer, pH 7.4, and treated with with 1.1–10 equiv of oxaziridine

probe **4.1** (100× stock in DMF). The reaction mixture was incubated at room temperature for 10 min with agitation and immediately quenched by desalting twice with a Bio-Spin chromatography column (Bio-Rad).

4.4.5 Modification of Fab-oxa-N₃ with DIBAC-Cy3

Fab-oxa-N₃ was diluted to 1 mg/mL in PBS. Fab was labeled with dibenzocyclooctyne-Cy3 (DIBAC-Cy3) using a strain-promoted azide-alkyne cycloaddition. Anti-GFP-Fab-azide was incubated with 2–10 equiv of DIBAC-Cy3, and reacted for 8 h at room temperature before quenching by protein desalting.

4.4.6 Modification of Fab-Cy3 with diazo compound 3.1

Fab-Cy3 was dialyzed into 10 mM Bis-Tris buffer, pH 6.5, and concentrated to 10 mg/mL using 30-kDa MWCO spin columns (GE Healthcare). Fab (4 μL, 0.83 μmol) was incubated with diazo compound **3.1** (1 μL, 200 or 400 equiv) for 24 h at 4 °C, 22 °C, or 37 °C for 24 h. The reaction mixture was diluted 1:80 with 10 mM Bis-Tris buffer, pH 6.5 and analyzed by LC-MS. The results are displayed in Figure 4.2.

In subsequent experiments, Fab-Cy3-**3.1** was produced by incubating Fab-Cy3 with 400 equiv of diazo compound **3.1** in 80:20 v/v 10 mM Bis-Tris buffer, pH 6.5/acetonitrile at 22 °C for 24 h. The solution was diluted 1:100 with 10 mM Bis-Tris buffer, pH 6.5, and concentrated using 30-kDa MWCO spin columns.

4.4.7 Cell Culture

Chinese hamster ovary-K1 (CHO-K1) cells were from the American Type Culture Collection and cultured according to recommended protocols. Cells were grown in F12K nutrient medium supplemented with fetal bovine serum (10% v/v), penicillin (100 units/mL) and streptomycin

(100 $\mu\text{g}/\text{mL}$). Cells were grown in T75 sterile culture flasks in a cell-culture incubator at 37 °C under CO_2 . Cells were counted to determine seeding density using a Countess automated cell counter (Life Technologies).

4.4.8 Flow Cytometry

Cells were seeded at a density of 25,000 cells/well in a sterile 96-well plate 24 h prior to treatment. Cells were treated with Fab or Fab-**3.1** (5 μM) in F12K medium supplemented with penicillin (100 units/mL) and streptomycin (100 $\mu\text{g}/\text{mL}$) and incubated at 37 °C under CO_2 for 4 h. Cells were rinsed with DPBS (200 μL) and released from the plate with 0.05% trypsin-EDTA (200 μL). Trypsin was quenched by the addition of 200 μL medium containing 1 μL of yellow live/dead stain (Life Technologies cat. no. L34959). The fluorescence intensity (ex. 488 nm, em. 670 LP) of at least 10,000 live cell events was measured by flow cytometry with a FACSCanto-HTS II flow cytometer (BD Biosciences). The median fluorescence intensity of live, single cells is reported.

4.4.9 Confocal Microscopy

Cells were seeded at a density of 50,000 cells/well in an 8-well microscopy dish (Ibidi) 24 h prior to treatment. Cells were incubated with either unmodified Fab or Fab-**3.1** (5 μM) in F12K medium supplemented with penicillin (100 units/mL) and streptomycin (100 $\mu\text{g}/\text{mL}$) for 4 h at 37 °C. Cells were rinsed twice with DPBS, and nuclei were stained by incubation with Hoechst 33342 dye (2 $\mu\text{g}/\text{mL}$) for 5 min at 37 °C. Cells were then kept in medium on ice until the time of analysis. Image acquisition and processing settings were maintained between all samples.

Chapter Five

Site-Specific Antibody Functionalization Using Tetrazine–Styrene Cycloaddition

Abstract: Biologics, ranging from insulin to antibody–drug conjugates, are becoming mainstream therapeutic agents in clinical practice. Consequently, methods to label biologics without disrupting their pharmacological function are essential for identifying, characterizing, and translating candidate biologics from the bench to clinical practice. Here we present a method for labeling antibody single-chain variable fragments (scFv), isolated from the surface of yeast, specifically at the C terminus with a detection probe by combining intein-mediated expressed protein ligation (EPL) with the inverse electron-demand Diels–Alder (IEDDA) cycloaddition. The high concentration of thiols required to trigger EPL present especially challenging conditions for a chemoselective ligation reaction. We overcame this challenge by using a styrene as the dienophile and limiting the exposure of tetrazine reagents to yeast cells and reducing conditions. We demonstrate that a styrene is stable in the presence of high concentrations of thiols and remains functional after scFv modification with EPL. An scFv bearing a styrene handle can react with a tetrazine to generate functionalized scFv. We use the EPL plus IEDDA labeling procedure to functionalize two different scFv’s with fluorescent probes and demonstrate that the ensuing labels do not impede binding to antigen. This means to label a yeast surface-derived scFv rapidly in a site-directed manner could find utility in downstream laboratory and pre-clinical applications.

Author Contributions: Kalie A. Mix synthesized chemical reagents. Benjamin J. Umlauf, Kalie A. Mix, Eric. S. Shusta, and Ronald T. Raines designed experiments and analyzed data.

Benjamin J. Umlauf performed scFv modification and binding experiments. Kalie A. Mix, Benjamin J. Umlauf and Ronald T. Raines wrote this chapter.

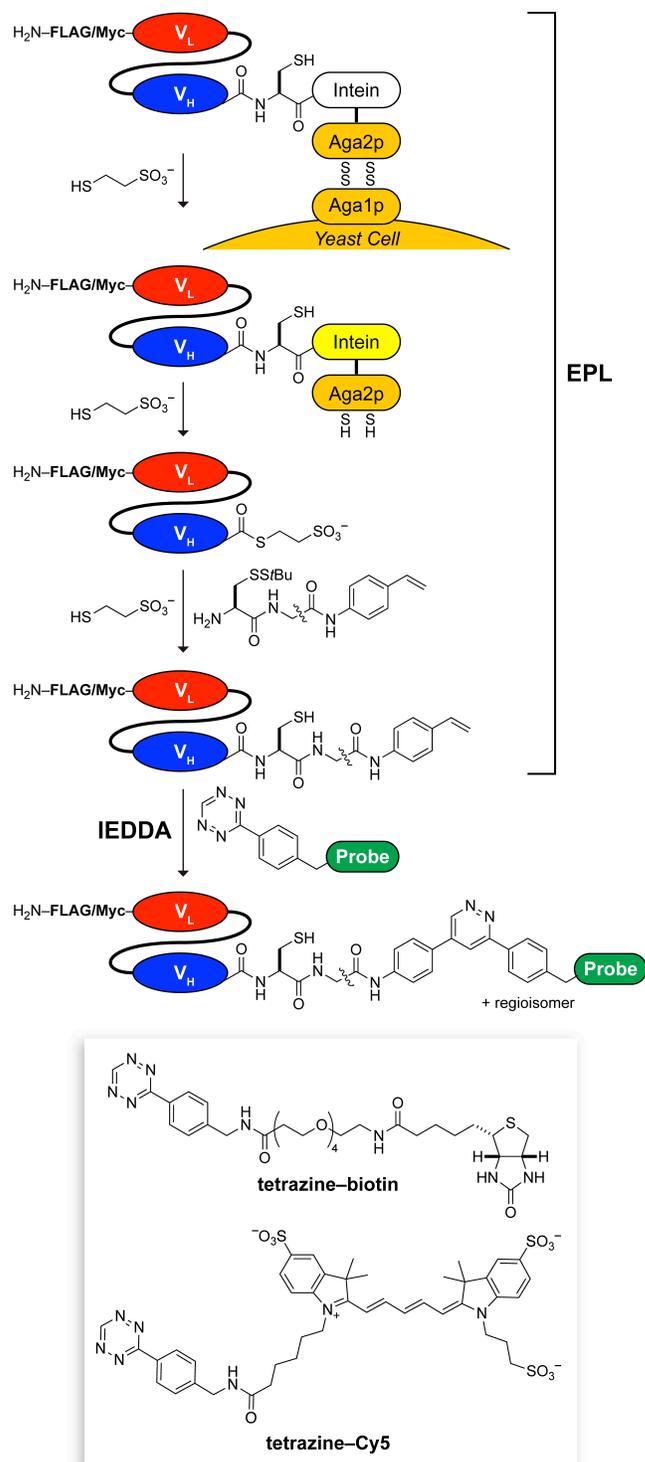
5.1 Introduction

Antibodies are a rapidly growing class of therapeutic agents with significant clinical success and commercial impact. Many technologies such as imaging, diagnostics, and isolation or analysis of biomolecules employ antibodies, due to the high specificity and affinity for their cognate antigen. Functionalization of antibodies with chemical probes such as fluorophores,^{244,245} small-molecule drugs,²⁴⁶⁻²⁴⁸ or other biomolecules²⁴⁹⁻²⁵¹ further increases their utility. Still, a growing number of studies indicate it is essential to append these probes in a site-specific manner that does not disrupt antibody function.^{252,253,246,254,250,255}

We previously developed a method for site-specific antibody modification at the C terminus by employing yeast surface display in combination with expressed protein ligation (EPL).^{256,257} In this system, the C terminus of an scFv is fused to a non-self-cleavable intein, termed 202-08, and expressed on the surface of yeast cells. The scFv fusion is tethered to the yeast surface via two disulfide bonds.²⁵⁸⁻²⁶⁰ Addition of a thiol, 2-mercaptoethanesulfonic acid (MESNA), to the yeast culture reduces the disulfide bonds and liberates the scFv from the yeast surface. MESNA also activates 202-08 intein splicing to undergo transthioesterification, which produces a C-terminal thioester. The soluble scFv bearing a C-terminal thioester reacts with cysteine (Cys) amides to link the scFv to a Cys-modified probe of interest via an amide bond (Scheme 5.1). This system enables protein modification specifically at the C terminus more rapidly than do other technologies that require purification of soluble protein and non-specific functionalization using amino acid or thiol-containing side chains. Thus, the EPL system is ideal

for the rapid, high-throughput functionalization of antibodies. Additionally, non-self-cleaving inteins, such as 202-08, are excised during EPL, resulting in traceless appendage of the probe to the C terminus. Other genetically encoded protein modification domains, such as SNAP tags, are retained after modification with the probe and can alter the antigen-binding ability or specificity of the scFv.

Previous studies have appended a variety of functional groups to proteins, including post-translationally modified peptides,²⁵⁶ non-canonical amino acids,²⁶¹ and biophysical probes²⁶² using EPL. Our group previously appended an azide to the C terminus of an scFv using an EPL reaction between an scFv-intein fusion and a cysteine–azide reagent to install a reactive handle for copper-catalyzed azide–alkyne cycloaddition (CuAAC).²⁵⁷ Although the rapid rate of this reaction makes it useful for *in vitro* applications, CuAAC has limited utility *in vivo* due to the oxidative stress induced by Cu(I) and cross-reactivity of ascorbate with biological nucleophiles.^{263,264} Additionally, the multi-component nature of the reaction (which requires a copper catalyst, activating ligand, and reducing agent in addition to the azide and alkyne reagents) often requires re-optimization to apply the reaction to different molecules and conditions. Further, the high concentration of thiols can impair certain CuAAC reactions.^{265,266}



Scheme 5.1 Route for the two-step site-specific functionalization of a yeast surface-displayed scFv. Functionality is added at the C terminus by using expressed protein ligation (EPL) and inverse electron-demand Diels–Alder (IEDDA) cycloaddition. Inset: structures of two tetrazine probes for IEDDA cycloaddition.

We sought to improve this method by utilizing reagents that enable a simple two-component reaction that is stable, rapid, and directly amenable to *in vitro* and *in vivo* downstream applications. An ideal chemical reaction would be chemoselective in the presence of biological nucleophiles, free of organic solvents or catalysts, and able to be carried out in water. An additional criterion for this EPL-based method is that at least one of the reaction partners must be extremely stable in the presence of a thiol, as the EPL reaction requires ~200 mM of thiol as well as ~5 mM of a cysteine derivative. This criterion is a major challenge because many “bio-orthogonal” reagents are rendered useless at such high thiol concentrations.^{267,268,266}

In this study, we demonstrate that a styrene is inert to millimolar thiol concentration. This stability enables modification of an scFv at the C terminus with styrene handles via EPL. We then functionalize a styrene-modified scFv with tetrazine-containing probes via an inverse electron-demand Diels–Alder (IEDDA) cycloaddition. Finally, we demonstrate that two different probe-functionalized scFv’s retain antigen-binding ability similar to non-functionalized scFv’s.

5.2 Results

A Styrene is Compatible with Both EPL and IEDDA Reaction

The reaction of a tetrazine with *trans*-cyclooctene (TCO) has become a well-established and useful tool in chemical biology due to its rapid rate constant in water and the two-component nature of the reaction.²⁶⁹ Nevertheless, either a tetrazine or a TCO must be inert to high concentrations of thiols for compatibility with yeast surface display EPL. To measure its stability in the presence of free thiols, we incubated *trans*-cyclooctenol with FmocCysOH. After only 4 h, the TCO had isomerized completely to the unreactive *cis* isomer (Figure 5.1B).

The dienophile in an IEDDA cycloaddition can also be activated by electron-donating groups instead of strain.²⁷⁰ We investigated the reactivity and stability of one such activated alkene, 4-aminostyrene.²⁷¹ The amino group serves to both activate the alkene by donating electrons into the aryl system and acts as a handle for derivatization. To test the compatibility of styrene with high concentration of thiols, we incubated 4-aminostyrene with FmocCysOH. Gratifyingly, no degradation of the styrene was detectable after 12 h (Figure 5.1A). We also used NMR spectroscopy to examine the reaction kinetics of 4-acetamidostyrene with a phenyltetrazine (Figure 5.2).²⁶⁹ The second-order rate constant was found to be $k = (4.0 \pm 0.1) \times 10^{-3} \text{ M}^{-1}\text{s}^{-1}$.

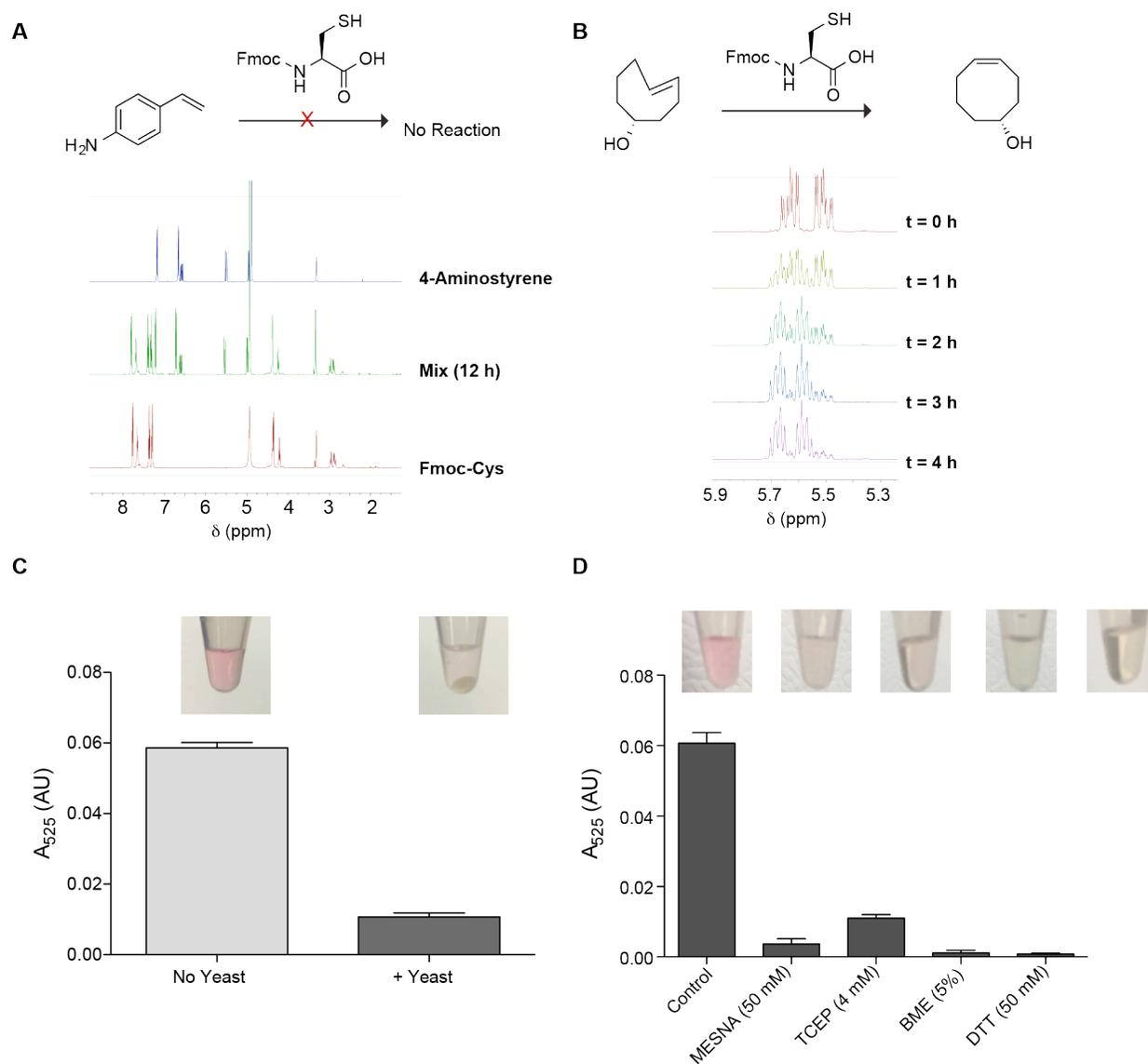


Figure 5.1 Stability of candidate reagents. (A) NMR spectra of 4-aminostyrene, FmocCysOH, and the reaction mixture after a 12-h incubation. (B) NMR spectra of vinyl protons of *trans*-cyclooctenol over the course of 4 h of incubation with FmocCysOH. (C) Absorbance at 525 nm of 5 mM tetrazine-amine with or without 100 μ L of yeast cell culture. Inset: images showing the loss of the characteristic pink color of a tetrazine in the presence of yeast cells. (D) Absorbance at 525 nm of a tetrazine solution containing various reducing agents used for yeast surface cleavage and EPL. Inset: images showing the loss of the characteristic pink color of a tetrazine in the presence of yeast cells.

A Tetrazine is Reduced in the Presence of Yeast Cells or Thiols

We next investigated the possibility of conjugating either cysteine–tetrazine or cysteine–styrene to the scFv via EPL. In theory, a one-pot reaction to achieve both cleavage from the yeast surface

and EPL labeling could be accomplished by simultaneous addition of MESNA and the cysteine-modified chemical handle directly to induced yeast cultures. We identified, however, several constraints to doing so. Incubation of a tetrazine with live yeast cells resulted in a significant decrease of tetrazine absorbance at 525 nm (Figure 5.1C, $p < 0.01$) indicating that the tetrazine is destroyed, sequestered, or reduced by the yeast cells.²⁷² The change is visible to the naked eye by observing a loss of pink hue (Figure 5.1C, insets).

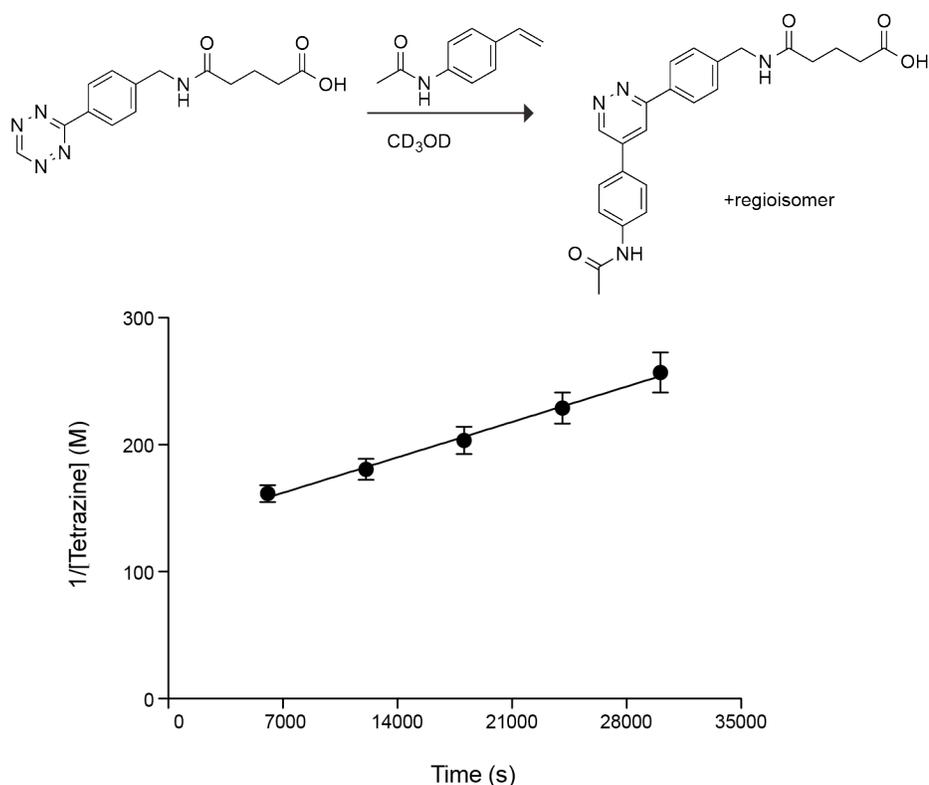


Figure 5.2 Kinetics of the tetrazine–styrene reaction. Percent conversion was monitored by disappearance of the starting tetrazine compound as determined by integration of the peak at 10.3–10.275 ppm.

We also incubated a tetrazine with common reducing agents that are used in intein-mediated EPL (Figure 5.1D). Traditional thiol-based reducing agents, including 2-mercaptoethanol, dithiothreitol, and MESNA, all caused tetrazine reduction as measured by a loss of absorbance at

525 nm ($p < 0.01$). The change in tetrazine oxidation state is also visible to the naked eye by observing the loss of pink hue (Figure 5.1D, insets). Incubation of a tetrazine with tris(2-carboxyethyl)phosphine (TCEP) at only 4 mM also resulted in tetrazine reduction ($p < 0.01$).

Modification of scFv by EPL and then IEDDA Cycloaddition

Based on the results above, we chose to modify the scFv with the styrene reagent using EPL, and then derivatize the modified scFv with a tetrazine probe. We first modified and characterized 4-4-20 scFv, which binds to fluorescein isothiocyanate (FITC). 4-4-20 scFv-intein fusion protein was displayed on the yeast surface and released using 50 mM MESNA. Simultaneously, MESNA triggers the intein to undergo EPL with 5 mM Cys(*St*Bu)-PEG₃-styrene (**5.3**), which was reduced to Cys-PEG₃-styrene in situ, to generate 4-4-20 scFv modified at its C terminus with a styrene handle. Styrene-modified scFv was then incubated with tetrazine-biotin to generate scFv functionalized with biotin at its C terminus (Scheme 5.1). A Western blot with an anti-biotin antibody (Figure 5.3A) demonstrated that functionalization of scFv with biotin was dependent on styrene modification and reaction with tetrazine-biotin. We also characterized the role of MESNA concentration in this reaction and determined that 200 mM MESNA is the optimal concentration for producing an scFv functionalized as in Scheme 1 (Figure 5.3B).

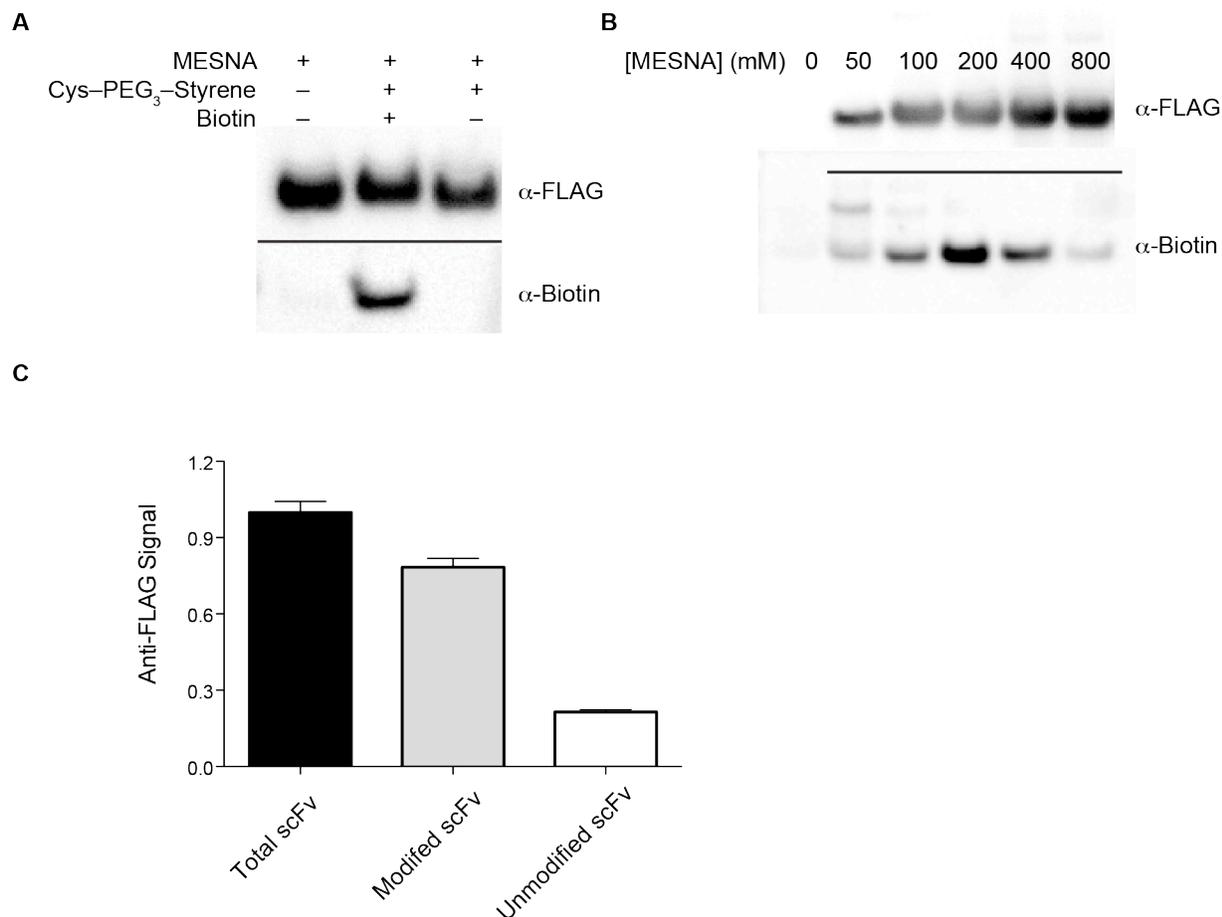


Figure 5.3 Functionalization of 4-4-20 scFv by EPL followed by IEDDA. (A) Western blot showing the modification of a 4-4-20 scFv modified with a styrene-reactive handle and reacted with tetrazine–biotin. (B) Western blot showing EPL reaction yields as a function of MESNA concentration. (C) Bar graph showing the yield of 4-4-20 scFv modification by the method in panel A. Total scFv and modified modified scFv (isolated with streptavidin–magnetic beads) were detected by Western blotting with an anti-FLAG antibody.

To determine the efficiency of the combined EPL and IEDDA reactions, we calculated the percentage of biotin-labeled scFv relative to total scFv released from the yeast surface (Figure 5.3C). Styrene-modified 4-4-20 scFv ± tetrazine–biotin was mixed with streptavidin (SA) magnetic beads. The biotinylated fraction was then isolated from the total pool of scFv and compared to the total scFv population with a Western blot. We observed $83.3 \pm 1.6\%$ of 4-4-20

scFv-bound SA beads, however, $23.1 \pm 6.3\%$ of non-biotin labeled 4-4-20 scFv also bound SA beads. Hence, we estimate that 60–80% of 4-4-40 scFv is labeled using the EPL+IEDDA protocol depicted in Scheme 5.1.

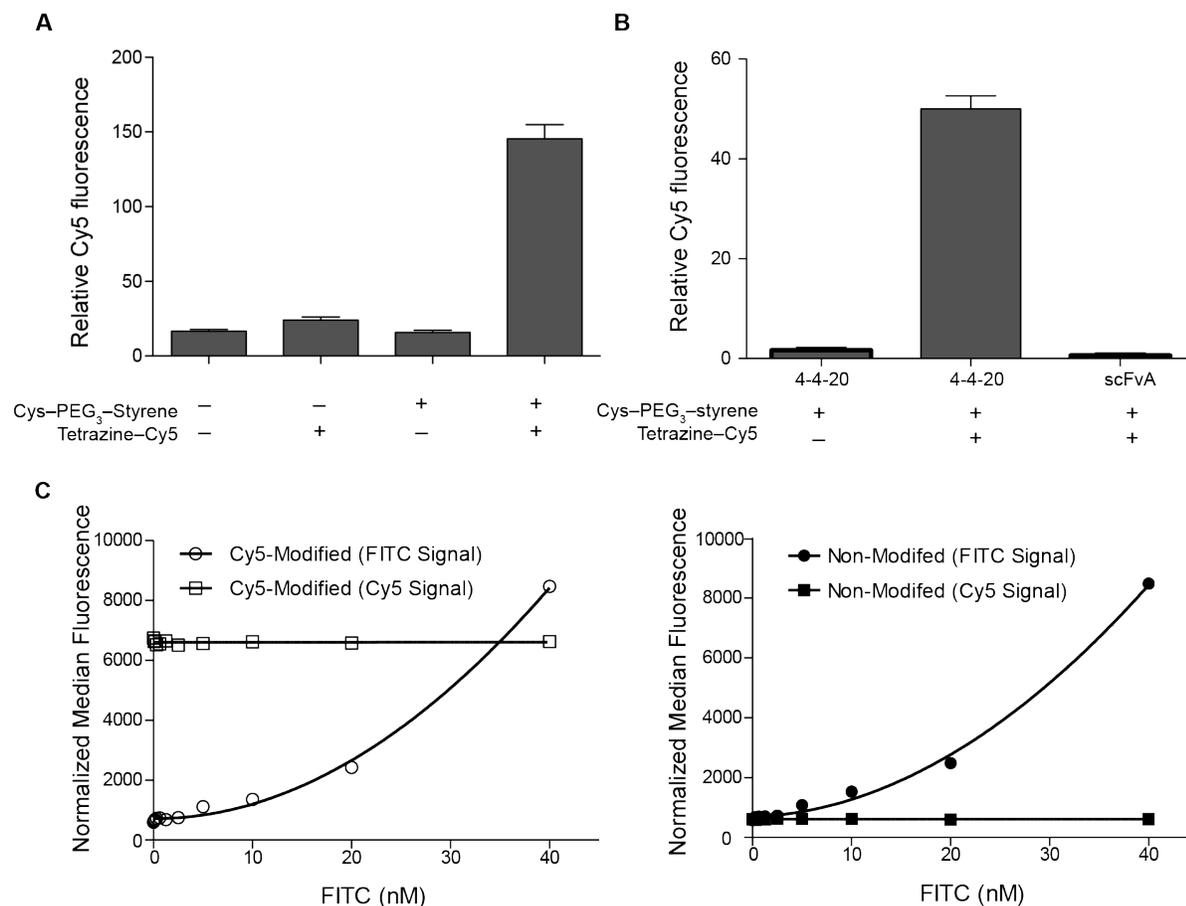


Figure 5.4 4-4-40 scFv maintains function after modification with a styrene and labeling with a tetrazine–Cy5. (A) Cy5 fluorescence signal of 4-4-20 scFv after modification with Cys–PEG₃–styrene and tetrazine–Cy5. (B) Cy5 fluorescence of scFv incubated with immobilized FITC–dextran. (C) Representative plots of 4-4-20 scFv (20 nM) modified with Cy5 (left panel) or unmodified 4-4-20 scFv (right panel) titrated against FITC. Values of K_d were 1.55 ± 0.81 nM and 1.65 ± 0.76 nM, respectively ($p > 0.05$).

4-4-20 scFv Retains Antigen-Binding Ability after Functionalization

Covalent labeling of scFv with fluorescent dyes in a manner that does not impede antigen binding is crucial in chemical, biochemical, and medical applications. We functionalized styrene-modified scFv with tetrazine–Cy5 by incubating styrene-modified (or unmodified) 4-4-20 scFv with 5 mM tetrazine–Cy5. A significant increase in Cy5 fluorescence is apparent only in scFv that is both styrene-modified and incubated with tetrazine–Cy5 (Figure 5.4A, $p < 0.01$).

To test that the scFv antigen-binding region retains function following modification, we incubated 4-4-20 scFv–Cy5 in wells containing the FITC antigen immobilized on dextran. As expected, we observed a significant increase in the fluorescence of 4-4-20 scFv–Cy5 group (Figure 5.4B, $p < 0.05$). Both unmodified 4-4-20 scFv and scFvA–Cy5, an scFv that does not bind FITC, did not generate a Cy5 signal above background. We also quantified the affinity of 4-4-20 scFv and 4-4-20 scFv–Cy5 for FITC using a fluorescence-quenching assay established previously. We measured a K_d value of 1.55 ± 0.81 nM for the complex of 4-4-20 scFv–Cy5 with FITC. This value does not differ significantly than that for unmodified 4-4-20 scFv ($K_d = 1.65 \pm 0.76$ nM, $p > 0.05$, Figure 5.4C).

scFvA Retains Antigen-Binding Activity after Functionalization

To demonstrate the utility of this method for in vitro tissue culture-based assays, we functionalized scFvA with Cy5 using the protocol presented above for 4-4-20 scFv–Cy5. ScFvA recognizes an antigen on the surface of RBE4 cells, a rat brain endothelial cell line. Adherent RBE4 cells were incubated with $0.5 \mu\text{M}$ scFvA–Cy5 or 4-4-20 scFv–Cy5, and internalized scFv was quantified using flow cytometry. We observed that $93.1 \pm 3.3\%$ of RBE4 cells internalized scFvA–Cy5, whereas only 1% of cells internalized 4-4-20 scFv–Cy5 (Figures 5.5A–E, $p < 0.01$).

This discrepancy suggests that scFvA–Cy5 retains the ability to bind native antigen after fluorophore labeling, as anticipated.

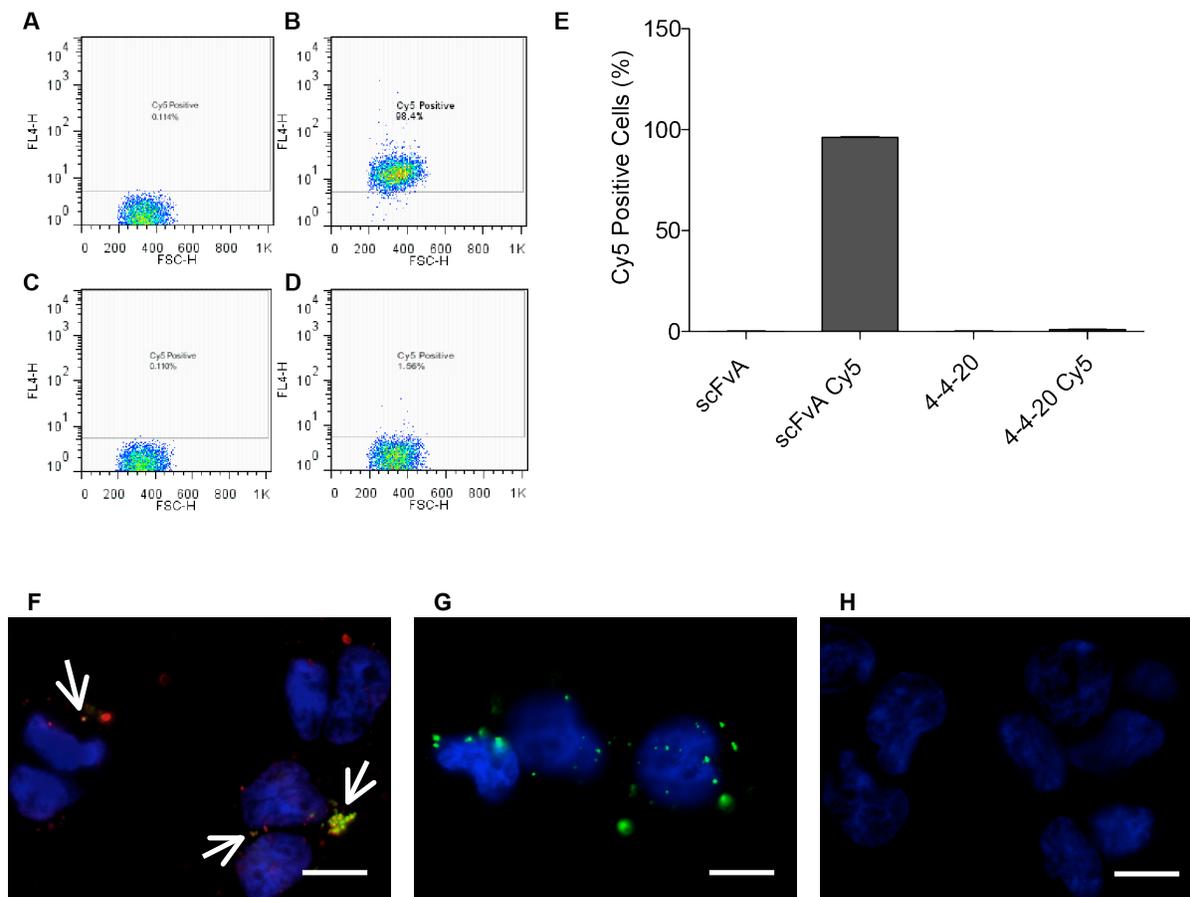


Figure 5.5 Internalization of Cy5-labeled scFv's into rat brain endothelial (RBE4) cells. (A–D) Plots from representative flow cytometry experiments with (A) unmodified scFv, (B) scFvA–Cy5, (C) unmodified 4-4-20 scFv, and (D) 4-4-20 scFv–Cy5 (E) Quantification of Cy5-positive cells across three independent experiments for each group. (F) Fluorescence microscopy images of RBE4 cells incubated with scFvA–Cy5, fixed, permeabilized, and stained for a Myc tag on the scFv. Blue indicates nuclei (Hoechst 33342); green indicates anti-Myc AF488; red indicates scFvA–Cy5. Arrows represent sites of co-localization between Myc and Cy5 signal. (G) RBE4 cells treated with scFvA–styrene that was not reacted with tetrazine–Cy5 and is stained as in panel F. (H) RBE4 cells not treated with an scFv, and stained as in page F. Scale bars; 20 μm.

We also examined the internalization of scFvA using fluorescence microscopy. RBE4 cells were incubated with scFvA–Cy5, scFvA, or no scFv. After 1 h, the cells were washed, fixed, permeabilized, and back-stained with anti-Myc antibody (9E10), goat anti-mouse–AF488, and Hoechst 33342 (Figures 5.5F–H). Colocalization of Cy5 and Myc signal (arrows) indicates intact, internalized scFvA–Cy5. RBE4 cells incubated with unmodified scFvA demonstrate the presence of Myc staining with no Cy5 signal, as expected.

5.3 Discussion

We demonstrate a method for functionalizing scFv's specifically at their C terminus by combining EPL and IEDDA reactions. Modification of scFv displayed on the surface of yeast cells using EPL presents especially challenging conditions due to the high concentration of sulfur nucleophiles required for the EPL reaction.²⁶⁶ We overcame this limitation by using a styrene as the dienophile in the IDEAA reaction.^{270,273} The importance of using a styrene rather than a TCO is highlighted by the isomerization of TCO in the presence of the thiol concentrations required to mediate EPL.

Styrenes are known to undergo cycloaddition with tetrazines.^{270,274,269,273} These dienophiles are, however, used infrequently compared to TCO. To our knowledge, there is only one other report on the use of a styrene–tetrazine cycloaddition for bioconjugation.²⁷³ Further, we are not aware of any studies that use the robust nature of styrene to repel attacks by sulfur nucleophiles as we present herein. This paucity could be due, in part, to the slow rate of reaction between styrene and tetrazine compared to TCO and tetrazine (Figure 5.2). We overcame some of the slow reaction rate by adding an excess of tetrazine probe and raising the reaction temperature.²⁶⁵

Further tuning the electronics of each reagent could also increase the reaction rate and thereby enhance functionalization of low concentrations of styrene-modified proteins.^{275,271}

We also observed several constraints with respect to the tetrazine moiety. Both live yeast cells and common reducing agents, including the non-thiol based TCEP, caused reduction (and subsequent inactivation) of a tetrazine. It is possible that yeast cells are internalizing and degrading the tetrazine in their cytoplasm, that a component of the medium is degrading the tetrazine, or that the tetrazine is being reduced by another mechanism. To circumvent this problem, we used Cys-PEG₃-styrene rather than Cys-PEG₃-tetrazine as the EPL reagent. We note that modification of scFv with Cys-PEG₃-tetrazine could be attainable by removing reducing agents. Alternatively, incubation with methylene blue or exposure to red light could re-oxidize a reduced tetrazine following reduction with biological reduction agents.²⁷² Both of these routes would, however, require additional steps.

Given the relative differences in size between an scFv and styrene (~30 kDa versus ~100 Da), we added a PEG spacer between the cysteine amide used for EPL and the styrene moiety. In our application, the relatively small size of the tetrazine probe is unlikely to cause major steric hindrance. Nevertheless, in applications that seek to modify the scFv-styrene with a tetrazine probe of larger size, the PEG spacer will likely have a more beneficial impact.^{276,277}

Another important factor in the optimization of this method is the concentration of MESNA. Interestingly, we observed an optimal concentration for MESNA at 200 mM. The observation of a maximum efficiency rather than a plateau with increasing concentrations of MESNA was unexpected. The optimal point could be due to high concentrations of MESNA preventing the *S*-to-*N*-acyl switch required for EPL modification with the styrene handle, or competition between MESNA and cysteine for nucleophilic attack on the intein-generated thioester.

We observed 60–80% modification of 4-4-20 scFv using the EPL+IEDDA protocol (Figure 5.3C). We previously observed 70–99% intein cleavage using the 202-08 engineered intein.²⁷⁸ Given that we are coupling two reactions here (EPL and IEDDA) it is unsurprising there is a slight reduction in observed efficiency. Still, our two-step yield is still close to those observed with standard *E. coli*-produced intein-modified proteins.²⁷⁹⁻²⁸¹

Finally, we demonstrate the utility of our antibody modification method using fluorescence-based assays. Modification of 4-4-20 scFv with styrene enabled conjugation to tetrazine–Cy5 without affecting antigen binding.^{282,283} Thus, this protocol represents a distinct advantage over non-specific, amine or thiol-based antibody functionalization protocols that often result in large fractions of inactive antibody because probe functionalization occurs within the antigen-binding domain. We also demonstrate the utility of this method using mammalian tissue culture assays with scFvA functionalized with Cy5.²⁸⁴ The scFvA–Cy5 conjugate maintains its ability to bind to and internalize into RBE4 cells. An scFv that is labeled directly with a fluorophore has numerous potential downstream applications, including multi-time point live imaging studies in which an animal is treated with the antibody, subjected to live imaging techniques, sacrificed, and subjected to fixed tissue imaging. We believe that the modular, site-directed antibody-labeling protocol demonstrated herein is a powerful means to facilitate the development and assessment of antibodies and biologics for laboratory and preclinical use.

5.4 Acknowledgments

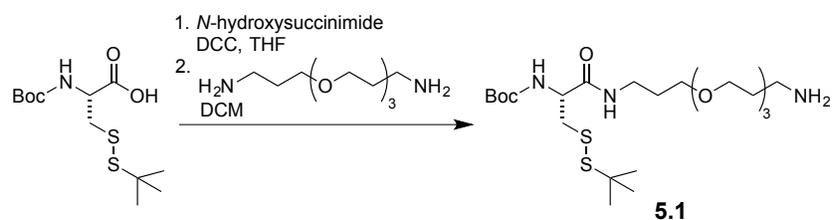
K.A.M. was supported by Molecular Biosciences Training Grant T32 GM007215 (NIH) and a fellowship from the University of Wisconsin–Madison College of Agricultural and Life Sciences. This work was funded by grant CBET 1403350 (NSF).

5.5 Materials and Methods

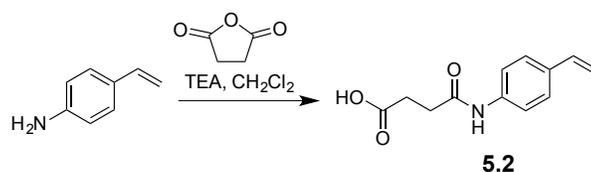
5.5.1 General

All procedures were performed in air at ambient temperature ($\sim 22\text{ }^{\circ}\text{C}$) and pressure (1.0 atm) unless indicated otherwise. Unless noted otherwise, reagents and solvents were from Sigma–Aldrich (Milwaukee, WI) and were used without further purification. Reagent-grade solvents: acetonitrile, dichloromethane (DCM), tetrahydrofuran (THF), and triethylamine (TEA) were dried over a column of alumina and were removed from a dry still under an inert atmosphere. Flash column chromatography was performed with 40–63 Å silica (230–400 mesh) from Silicycle (Québec City, Canada), and thin-layer chromatography was performed with EMD 250 μm silica gel 60 F254 plates. The phrase “concentrated under reduced pressure” refers to the removal of solvents and other volatile materials using a rotary evaporator at water aspirator pressure (<20 torr) while maintaining a water bath below $40\text{ }^{\circ}\text{C}$. Residual solvent was removed from samples at high vacuum (<0.1 torr). ^1H and ^{13}C NMR spectra were acquired on Bruker spectrometers in the National Magnetic Resonance Facility at Madison operating at 500 MHz. Chemical shift values (δ) are reported in units of ppm relative to an internal standard (residual solvent or TMS). Electrospray ionization (ESI) mass spectrometry for small-molecule characterization was performed with a Micromass LCT at the Mass Spectrometry Facility in the Department of Chemistry at the University of Wisconsin–Madison. LC–MS analysis was performed using a Shimadzu LC–MS2020 instrument with a quadrupole mass analyzer.

5.5.2 Chemical Synthesis

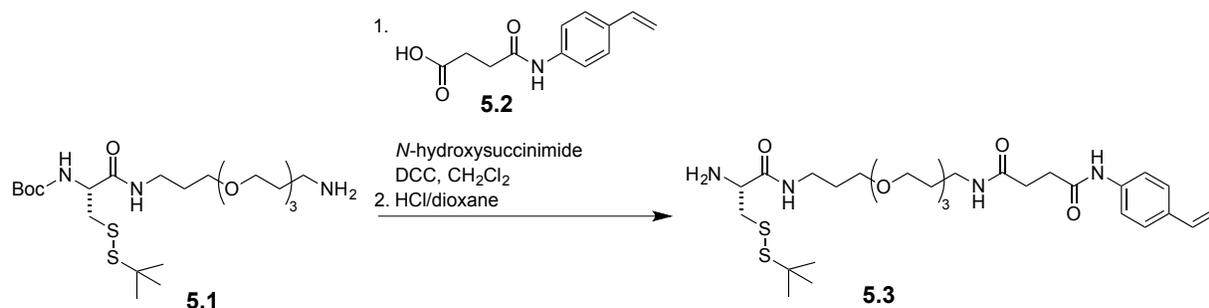


Compound 5.1. Boc-*S-tert*-butylthio-L-cysteine (500 mg, 1.6 mmol) from Chem-Impex International (Wood Dale, IL) was dissolved in THF (5 mL). *N*-Hydroxysuccinimide (186 mg, 1.6 mmol) and *N,N'*-dicyclohexylcarbodiimide (DCC; 363 mg, 1.7 mmol) were added, and the resulting solution was stirred overnight. The reaction mixture was filtered, and the filtrate was concentrated under reduced pressure. The residue was dissolved in DCM (15 mL). 4,7,10-Trioxa-1,13-tridecanediamine (0.9 mL, 4.3 mmol) was added, and the resulting solution was stirred overnight. The reaction mixture was filtered, and the filtrate was concentrated under reduced pressure. The residue was purified by reverse-phase HPLC on a C18 column using a gradient of water–acetonitrile containing trifluoroacetic acid (0.1% v/v) to yield compound **5.1** as a clear oil (64 mg, 10% for 2 steps). LC–MS (ESI⁺) *m/z* calcd for C₂₂H₄₅N₃O₆S₂ [M+H]⁺ 512.27; found 512.40.

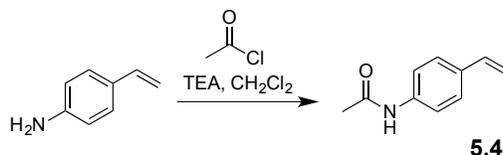


Compound 5.2. 4-Aminostyrene (100 mg, 0.8 mmol) was dissolved in DCM (8 mL). Succinic anhydride (84 mg, 0.84 mmol) and TEA (0.24 mL, 1.7 mmol) were added, and the resulting solution was stirred overnight. The reaction mixture was concentrated under reduced pressure. The residue was dissolved in EtOAc and washed twice with 1 M HCl. The organic layer was dried over Na₂SO₄(s) and concentrated under reduced pressure to yield compound **5.2** as a white solid (99 mg, 54%). ¹H NMR (500 MHz, CD₃OD, δ): 7.52 (d, 2H, *J* = 8.6 Hz), 7.37 (d, 2H, *J* = 8.6 Hz), 6.65–6.71 (dd, 1H, *J* = 10.9 Hz, 17.7 Hz), 5.71 (d, 1H, *J* = 17.6 Hz), 5.15 (d, 1H, *J* =

11.0 Hz), 2.66 (s, 4H). ^{13}C NMR (125 MHz, CD_3OD , δ): 176.4, 172.8, 139.6, 137.6, 137.5, 134.7, 127.6, 120.9, 32.3, 30.0. HRMS–ESI m/z calcd for $\text{C}_{12}\text{H}_{13}\text{NO}_3$ [$\text{M} - \text{H}$] $^-$, 218.0823; found 218.0821.



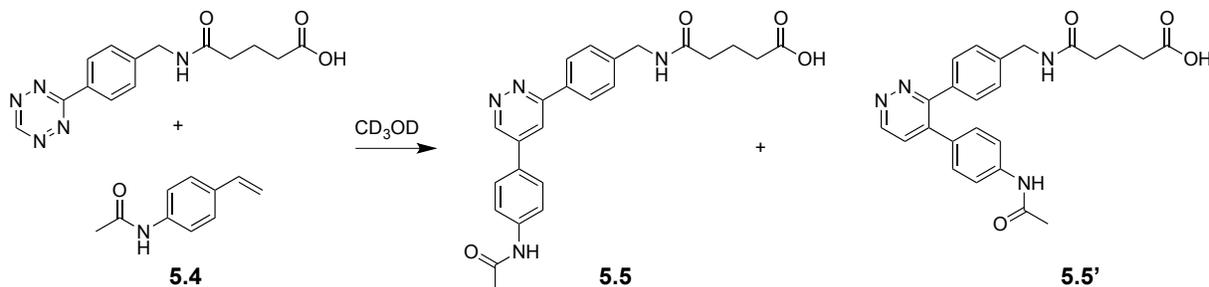
Cys(StBu)–PEG₃–styrene (5.3). Compound **5.1** (361 mg, 0.71 mmol) was dissolved in DCM (7 mL). Compound **5.2** (156 mg, 0.71 mmol), *N*-hydroxysuccinimide (82 mg, 0.71 mmol), and DCC (146 mg, 0.71 mmol) were added, and the resulting solution was stirred overnight. The reaction mixture was filtered, and then concentrated under reduced pressure. The residue was dissolved in acetonitrile and purified by reverse-phase HPLC on a C18 column using a gradient of water–acetonitrile containing trifluoroacetic acid (0.1% v/v). The residue was then dissolved in 4.0 M HCl in dioxane, and the resulting solution was stirred for 1 h. The solution was sparged with $\text{N}_2(\text{g})$ for 10 min to remove HCl and then concentrated under reduced pressure to yield compound **5.3** as a white solid (61 mg, 12% for 2 steps). LC–MS (ESI $^+$) m/z calcd for $\text{C}_{29}\text{H}_{48}\text{N}_4\text{O}_6\text{S}_2$ [$\text{M}+\text{H}$] $^+$ 613.30; found 613.35.



4-Acetamidostyrene (5.4). 4-Aminostyrene (100 mg, 0.84 mmol) was dissolved in DCM (8.4 mL). Acetyl chloride (0.18 mL, 0.84 mmol) and TEA (0.24 mL, 1.68 mmol) were added, and the resulting solution was stirred overnight. The reaction mixture was concentrated under reduced pressure, and the residue was dissolved in EtOAc. The solution was washed twice with 1 M HCl and twice with saturated aqueous NaHCO₃. The organic layer was dried over Na₂SO₄(s), and then concentrated under reduced pressure. The residue was purified further by chromatography on silica gel, eluting with 1:1 EtOAc/hexanes to yield compound **5.4** as a white solid (39 mg, 29%). ¹H NMR (500 MHz, CDCl₃, δ): 7.47 (d, 2H, *J* = 8.6 Hz), 7.37 (d, 2H, *J* = 8.5 Hz), 7.14 (s, 1H), 6.70–6.64 (dd, 1H, *J* = 10.9, 17.6 Hz), 5.68 (d, 2H, *J* = 17.6 Hz), 5.20 (d, 2H, *J* = 10.9 Hz), 2.19 (s, 3H). ¹³C NMR (125 MHz, CDCl₃, δ): 168.1, 137.4, 136.1, 133.7, 126.8, 119.7, 113.1, 24.7. HRMS–ESI⁺ (*m/z*): [M + H]⁺ calcd for C₁₀H₁₁NO, 162.0913; found, 162.0912.

5.5.3 Styrene and *trans*-Cyclooctene Stability

Stock solutions were prepared by dissolving FmocCysOH, 4-acetamidostyrene (**5.4**), and *trans*-cyclooctenol in CD₃OD at a concentration of 200 mM. The solutions were combined in an NMR tube to give an equimolar ratio and mixed, and the tube was inserted immediately into an NMR spectrometer. A 16-scan ¹H NMR spectrum was acquired every 60 min.



5.5.4 Tetrazine–Styrene NMR Kinetics

Stock solutions (6.25 mM in CD₃OD) were prepared of 5-[4-(1,2,4,5-tetrazin-3-yl)benzylamino]-5-oxopentanoic acid and 4-acetamidostyrene (**5.4**). The solutions were combined in an NMR tube at an equimolar ratio and mixed, and the tube was inserted immediately into an NMR spectrometer. A 16-scan ¹H NMR spectrum was acquired every 5 min. Conversion was monitored by disappearance of the tetrazine as determined by integration of the peak at 10.300–10.275 ppm. The integrity of the cycloaddition product (**5.5**) and its regioisomer (**5.5'**) was assessed by LC–MS. The value of the second-order rate constant was determined by linear regression analysis of a plot of 1/[tetrazine] versus time (Figure 5.2).

5.5.5 Yeast Surface Display

Two distinct scFv's were used in this work: 4-4-20 scFv (which binds to FITC and has an N-terminal biotin tag) and scFvA (which binds to an antigen on the surface of RBE4 cells and has an N-terminal Myc tag). Fusions of these scFv's to the 202-08 intein were encoded in the pCTre vector, which was transfected into EBY100 yeast cells as described previously.²⁵⁷ Transfected cells were grown at 30 °C in SD–CAA medium (20.0 g/L dextrose, 6.7 g/L yeast nitrogen base, 5.0 g/L casamino acids, 10.19 g/L Na₂HPO₄·7 H₂O, 8.56 g/L NaH₂PO₄·H₂O) as described previously,^{257,285,278} and induced with SG–CAA medium when the culture reached an OD₆₀₀ of 0.8–0.9. Induction was continued for 48 h at ambient temperature before harvesting an scFv.

5.5.6 EPL and IEDDA Cycloaddition of Yeast Surface-Displayed Proteins

Yeast cultures were pelleted, and the cells were washed two times with 50 mM HEPES buffer, pH 7.2. Cells from a 50-mL culture were re-suspended in 50 mM HEPES buffer, pH 7.2 (800 μL). A 10× MESNA solution was added (final concentration: 200 mM) followed by Cys(S*t*Bu)–

PEG₃-styrene (**5.3**; final concentration: 5 mM). This slurry was incubated for 1 h with gentle shaking. Residual yeast cells were removed by filtration through a 0.22- μ m filter, and the supernatant was harvested. The styrene-modified, scFv-containing supernatant was then exchanged four times into 50 mM HEPES buffer, pH 7.2, using 10,000-Da MWCO filters. The scFv was then concentrated and incubated with a tetrazine probe (final concentration: 5 mM) at 37 °C overnight. Excess tetrazine probe was removed by dialysis against an appropriate downstream buffer using 10,000 Da MWCO filters. Two tetrazine probes were used in this work (Scheme 5.1): tetrazine-biotin (which was product CP-6001 from Conju-Probe, San Diego, CA) and tetrazine-Cy5 (which was product 1189 from Click Chemistry Tools, Scottsdale, AZ).

5.5.7 SDS-PAGE and Immunoblotting of Reacted Proteins

scFv fractions were resolved using 4–12% w/v Bis-Tris acrylamide gels. Proteins were reduced with 5% v/v β -mercaptoethanol and denatured by boiling samples in SDS-containing loading buffer for 10 min. Proteins were transferred from the gel to nitrocellulose membranes. An anti-FLAG M2 monoclonal FLAG antibody (Sigma-Aldrich) and a BTN.1 anti-biotin antibody (NeoMarkers, Inc., Portsmouth, NH) were used to probe membranes for scFv. Anti-mouse HRP secondary antibody (Jackson Laboratory, Bar Harbor, ME) was detected using ECL, and a Bio-Rad imager was used to discern the presence of an scFv. A protein standard that is similar in size and that contained an N-terminal FLAG tag (Sino Biological, Beijing, China) was used to determine concentrations of scFv by immunoblotting.

5.5.8 FITC Titration

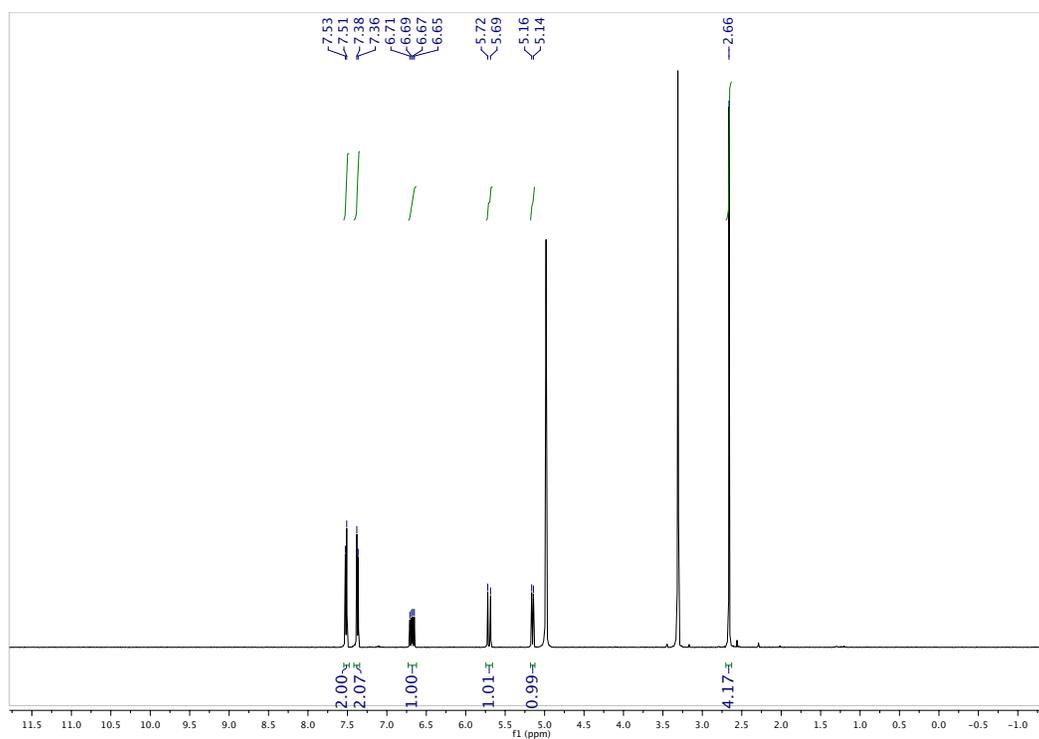
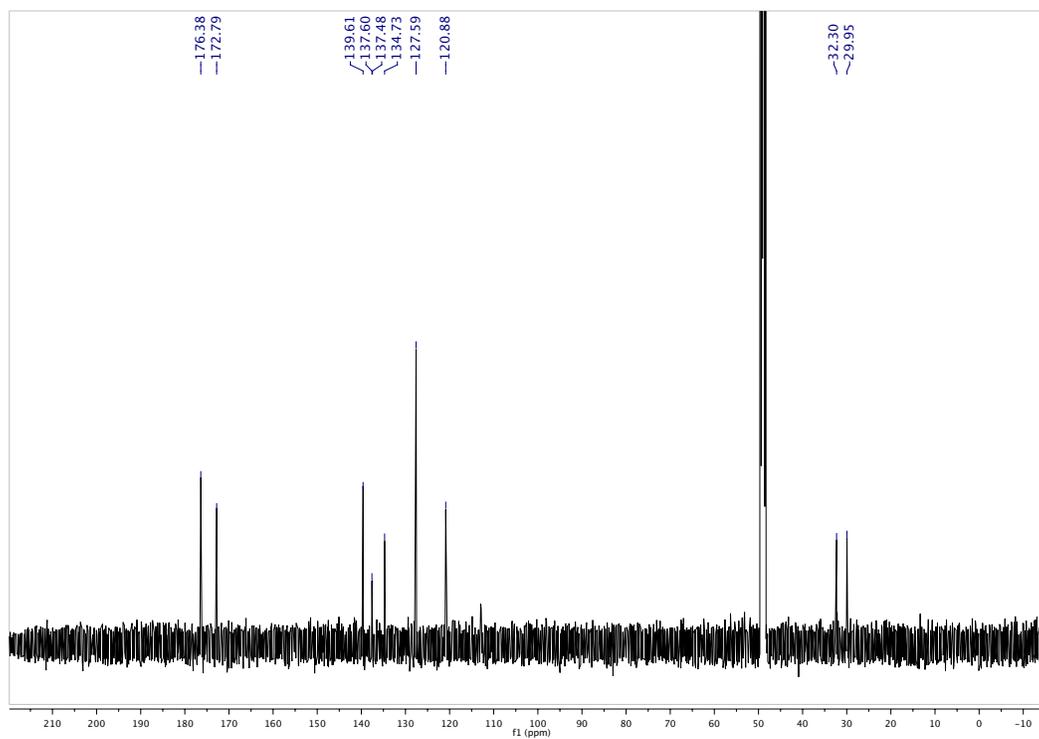
Modified and unmodified 4-4-20 scFv were titrated with FITC as described previously.^{282,286,257}

5.5.9 Flow Cytometry

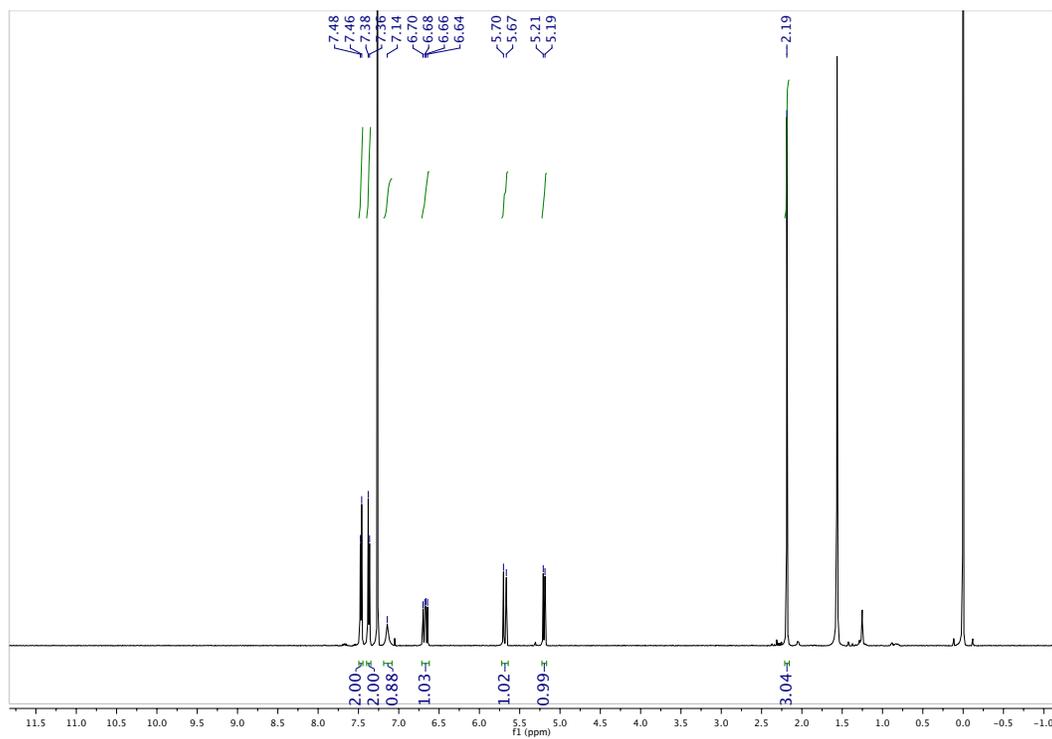
RBE4 cells were cultured on collagen type I-coated tissue culture flasks in 45% v/v Alpha Minimum Essential Medium, 45% v/v Ham's F10 medium, 10% v/v fetal calf serum, 100 mg/L streptomycin, 100,000 units/L penicillin G, 0.3 g/L geneticin, and 1 µg/L basic fibroblast growth factor (bFGF) as described previously.²⁸⁴ Cells were incubated with Cy5-labeled scFvA or unmodified scFvA (0.5 µM) for 1 h at 37 °C and 5% CO₂. Cells were washed three times with PBS (5-min washes) and trypsinized for 5 min at 37 °C and 5% CO₂. Cultures were diluted 1:1 with serum-containing growth medium to quench trypsin, and the cells were pelleted by centrifugation. The cell pellet was resuspended in PBS containing 10 mM EDTA. Fluorescence was measured with a BD FACSCaliber cytometer by quantifying 10,000 events/group using software from FlowJo (Ashland, OR).

5.5.10 Fluorescence Microscopy

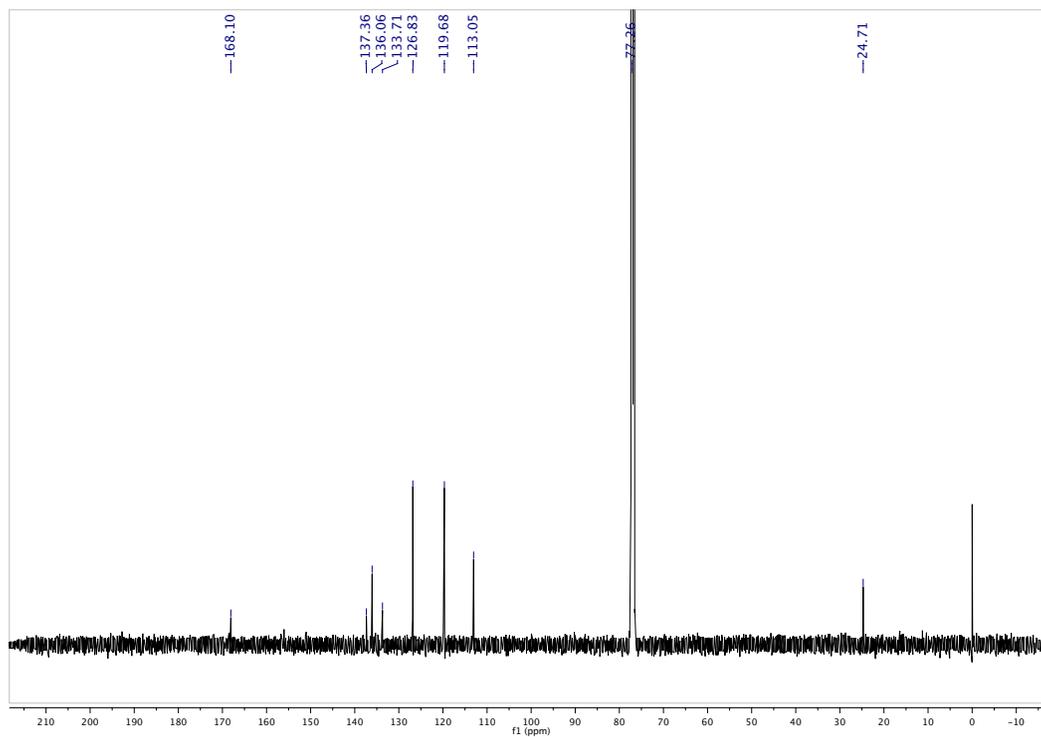
RBE4 cells were cultured on glass coverslips.²⁸⁴ RBE4 cells were incubated with modified or unmodified scFvA (2 µM) for 1 h at 37 °C and 5% CO₂. Cells were then washed three times with PBS, fixed in 4% v/v paraformaldehyde for 10 min, and permeabilized with 0.1% v/v Triton X-100. Permeabilized RBE4 cells were stained with 9E10 anti-Myc antibody (1:200), goat anti-mouse-AF488 antibody (1:200), and Hoechst 33342 (1:800). Slides were washed, mounted, and imaged on a Nikon upright fluorescence microscope.

5.5.11 NMR Spectra: ^1H NMR of 5.2 (CD_3OD , 500 MHz): ^{13}C NMR of 5.2 (CD_3OD , 125 MHz):

^1H NMR of 5.4 (CDCl_3 , 500 MHz):

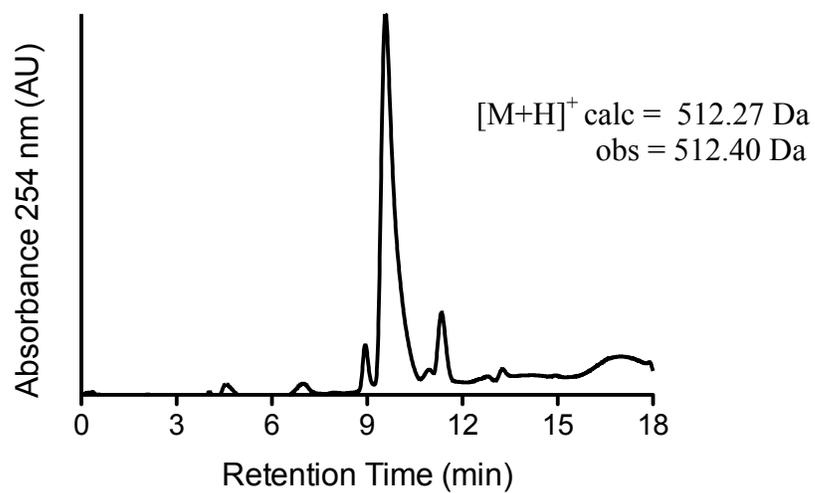


^{13}C NMR of 5.4 (CDCl_3 , 125 MHz):

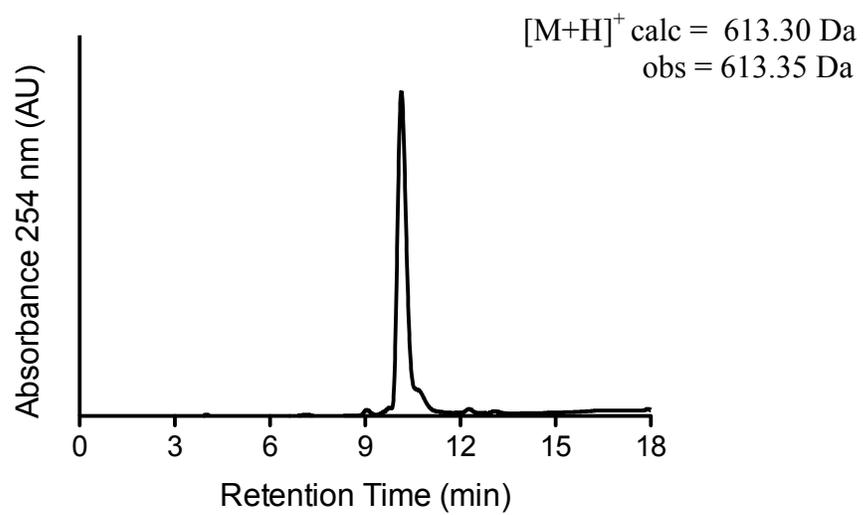


5.5.12 LC-MS Chromatograms

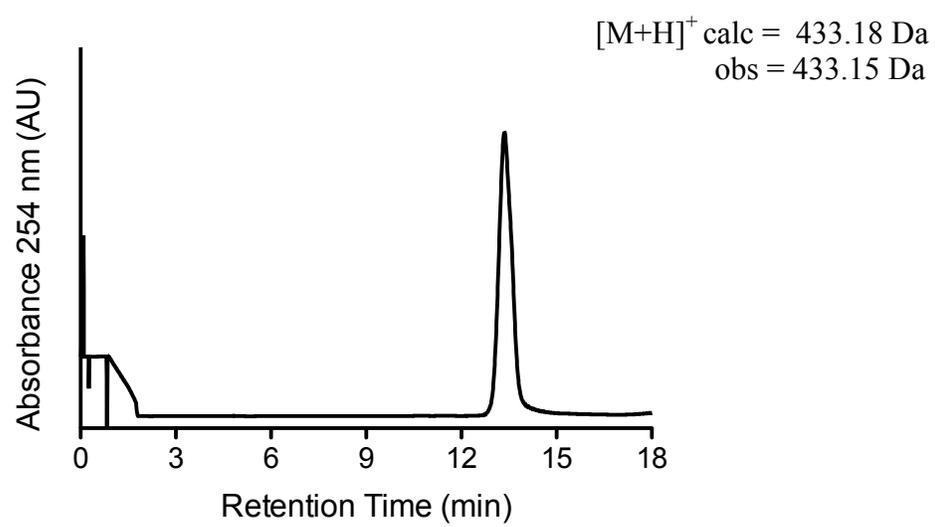
LC-MS Analysis of 5.1:



LC-MS Analysis of 5.3:



LC-MS Analysis of 5.5 and 5.5' (regioisomers):



Chapter Six

Future Directions

6.1 Further Optimization of Esterification Chemistry

In Chapter Two, I describe an optimized diazo amide for selective esterification of carboxyl groups in a solution of buffer and acetonitrile. The efficiency of the esterification reaction with regards to the ester:alcohol product ratio could be further improved further by optimizing the solvent composition. Preliminary experiments have shown that the dielectric constant and polarity of the organic cosolvent affect this product ratio (data not shown), but there was no clear trend. Measurements of the ester:alcohol product ratio that results from esterifying a small-molecule acid under different solvent compositions could produce a rank-ordered list of optimal conditions that engender a high ester:alcohol ratio.

Although this esterification method is theoretically applicable to all native proteins, because all proteins contain carboxyl groups, not all proteins are stable enough to stay properly folded after incubation with organic co-solvents. Previous work has investigated the stability of proteins in various organic solvents.²⁸⁷ The insight from this work could be combined with the rank-ordered list of optimal esterification solvents (above) to guide a search for ideal solvent conditions that would promote both esterification and protein stability.

Another potential method for promoting esterification efficiency would be to develop a reagent that masks cationic amino groups that would otherwise hinder esterification by forming salt bridges with carboxylates. Cationic lysine side chains could be temporarily cloaked during the esterification reaction by using a “catalyst” containing an aldehyde group. Previous work from the Francis group demonstrated that 2-pyridinecarboxaldehyde reacts with the amino group

of a protein N terminus to form an imine, which then undergoes nucleophilic attack by a proximal amide to form an imidazolidinone.²⁸⁸ An aldehyde reagent that reacts with amino groups to form imines but is unable to form the imidazolidinone could temporarily mask the amino groups as neutral imines. The imine could then be hydrolyzed to restore the free lysine amino groups after the esterification reaction is complete.

One potential application of the diazo esterification in medicinal chemistry is to use diazo reagents as covalent inhibitors of enzymes with active-site aspartic and glutamic acid residues. Although there are certainly inhibitors that contain *p*-methylphenylglycinamide groups that resemble diazo amide **2.2**, it would be more useful if the diazo group and adjacent “tuning” functional groups were small and easily incorporated into small-molecule inhibitors of interest. One approach to maintaining the optimized reactivity and selectivity of diazo compound **2.2** while decreasing its size would be to systematically investigate the pK_b of this compound, or the pK_a of the parent acid (which lacks a diazo group). It has been established that the ability of the diazo compound to abstract a proton from the carboxylic acid is a key determinant of esterification efficiency.¹²⁸ Thus, by using a diazo compound with the same basicity as compound **2.2**, it should be possible to retain optimal reactivity and chemoselectivity while minimizing the size. The efficacy of a noncovalent inhibitor could then potentially be improved by utilizing a small diazo group to covalently modify an active-site carboxyl group. Selectivity of the inhibitor for the enzyme of interest would be dictated by design of the small molecule into which the diazo group is introduced.

6.2 Improvement of Cytosolic Delivery Efficiency

In Chapter Three, I describe a method for delivering GFP into the cytosol of mammalian cells using a dimethyl amide diazo compound **3.1**. Compound **3.1**, and the others used in the experiments, were all small, hydrophobic molecules that masked negative charges on GFP. The use of cationic or cell-surface binding moieties in place of hydrophobic ones could prove to be more efficacious for cellular delivery. For example, incorporation of a quaternary amine, guanidinium, or boronic acid group into the diazo reagent (and subsequently, the surface of the esterified protein) could engender cell uptake via favorable interactions with the cell-surface glycocalyx. The effect of these groups on the stability and folding of the esterified protein, as well as on cellular entry, are unknown and warrant further exploration.

Measuring the cytosolic localization of proteins using these reagents would be improved by creating a new GFP construct with better nuclear localization efficiency than the sv40nls-GFP construct described in Chapter Three. Although this nuclear localization signal (NLS) is used frequently, a recent study quantified the nuclear localization efficiency of a variety of NLS sequences and found that the sv40nls is rather poor.²²⁵ The c-Myc NLS engenders approximately triple the nuclear localization efficiency of sv40nls and could enable more rigorous quantification of the cytosolic delivery of a fluorescent protein. A method to assay cytosolic localization using nitroreductase anchored to an intracellular organelle is being developed in the Raines laboratory and could complement the NLS approach.

The esterified GFP produced upon reaction with diazo compound **3.1** actually consists of a heterogeneous mixture of protein bearing varying number of labels in different locations. It is likely that only a subset of these proteins, such as those with the greatest number of labels or those with a certain arrangement of labels, actually enter cells. If these protein species could be

isolated from the heterogeneous mixture by, for example, using affinity chromatography, protein delivery efficacy might be improved.

Another critical aspect of cellular delivery that needs to be investigated is the effect of cell type on diazo-mediated protein delivery. Experiments with CHO cells that are deficient in glycosaminoglycan biosynthetic machinery would provide insight into the mechanism of cell uptake and a role for interaction between esterified protein and cell-surface glycans.²⁸⁹ Cellular delivery experiments should also be performed with both cancerous and non-cancerous human cell lines to determine the relative efficiency of uptake in each cell type, which could inform future applications of this technology.

6.3 Investigation of Ester Stability in Cells and Serum

In Chapter Three, we assay bioreversibility of esterification by labeling FLAG–angiogenin with diazo reagent **3.1**, incubating labeled protein with a cell lysate, and re-isolating protein using the FLAG tag. Attempts to assay bioreversibility using a His-tagged GFP failed due to the inability of GFP to remain folded after incubation with cell lysis detergent and produce a quality mass spectrum. Nonetheless, a His-tagged protein would be better suited for this assay since, unlike a FLAG tag, it does not contain carboxyl groups. There is a chance that the carboxyl groups in the FLAG tag cause the assay to be biased towards selective re-isolation of de-esterified FLAG–angiogenin, even if the mixture contains some esterified FLAG–angiogenin. Future experiments would benefit from use of a His-tagged RNase 1 variant. The RNase 1 would most likely be stable in the presence of cell-lysis buffer, and re-isolation using the His-tag would not be biased towards de-esterified protein.

Investigation of the kinetics of bioreversibility in live cells would also be highly valuable. One experimental approach is to use a FRET pair of fluorophores joined by an esterase-cleavable linker. A cell-penetrating peptide linker in which one fluorophore of the FRET pair is ester-linked would provide a ratiometric means to quantify esterase activity of the internalized peptide over time. The peptide provides an experimentally tractable platform for a FRET experiment due to the ability to carefully control the distance between the FRET pair fluorophores. Because we ultimately want to gain insight into the kinetics of bioreversibility using an esterified protein as a substrate, the experimental setup used for the peptide FRET experiments should then be adapted to investigate bioreversibility with a protein substrate. One approach would be to introduce a cysteine residue into GFP in a position such that a Cy3 dye could be introduced within the Förster radius distance of the GFP chromophore by cysteine–maleimide coupling chemistry. The linkage between the Cy3 dye and the maleimide group must contain an ester bond. Ratiometric measurement of GFP–Cy3 FRET (intact ester) and GFP fluorescence (cleaved ester) over time would enable characterization of esterase-cleavage kinetics.

To be useful in the clinic, an esterified therapeutic protein must be stable in the blood stream for sufficient time to allow the protein to enter cells. Although pharmacokinetic and pharmacodynamics experiments are best done in mice, preliminary in vitro experiments that investigate ester stability in serum would be experimentally straightforward and produce valuable insight. An esterified protein bearing an affinity tag (*e.g.* His, FLAG, or MBP) could be incubated with serum from various organisms and re-isolated from the serum using the affinity tag. The number of intact esters remaining over time could be determined using mass spectrometry.

6.4 Cellular Delivery of Functional Proteins and Enzymes

One key feature of the diazo-mediated protein delivery technology is that it is theoretically applicable to any native protein of interest because all proteins contain carboxyl groups. A few proteins that would be especially interesting to deliver to the cytosol of cells include the following:

1) *Yamanaka factors such as Oct4 and Nanog*: These transcription factors induce pluripotency by reprogramming a differentiated cell to a stem cell.²⁹⁰ Current technologies to induce stem cell pluripotency rely on transfection of DNA that encodes these Yamanaka factors, which is detrimental since the exogenous DNA remains in the cell and can be integrated into the genome.²⁹¹ Delivery of the Yamanaka factors as proteins would be advantageous because this method would be more “traceless:” the proteins would eventually be proteolytically degraded, and the cells would never contain any exogenous DNA.

2) *Anti-CRISPR proteins*: The CRISPR-Cas9 system is a powerful tool to effect a permanent change in the genome of an organism. As such, methods to control CRISPR-Cas9 activity would be especially valuable to ensure safety in clinical applications. Several anti-CRISPR proteins that naturally protect phage from bacterial CRISPR systems have been discovered and characterized.²⁹² The ability to deliver these proteins into mammalian cells could be an effective way to inhibit CRISPR-Cas9 activity.

3) *Anti-Ras antibody fragments (Fab)*: The experiments described in Chapter Four with anti-GFP Fab suggest that the diazo-mediated delivery system could be applied to other Fabs with similar size and structure. For example, delivery of an anti-Ras Fab could inhibit the signaling of this kinase, which is implicated in the progression of many cancer types.²⁴³ Additionally, phage display technology enables production of Fabs against virtually any antigen of interest. The combination of this technology with diazo-mediated cellular delivery of Fabs could enable the modulation of many intracellular signaling pathways of interest.

Appendix 1

Cellular Delivery of Green Fluorescent Protein by Cell-Penetrating Peptides Using Diazo Compound-Mediated Esterification

Abstract

Cell-penetrating peptides (CPPs) enable cellular uptake of a variety of cargo, including peptides, proteins, and nucleic acids. Although linear CPPs such as polyarginine or the HIV TAT peptide can be incorporated easily into recombinant proteins by using standard molecular biology techniques, the conjugation of proteins to CPP derivatives bearing non-peptidic components, such as a fluorophore or cyclic motif, is typically achieved through incorporation of non-canonical amino acids bearing a reactive chemical handle. Here, we present a method for conjugating cyclic and fluorophore-bearing CPPs to native protein via esterification of carboxyl groups using diazo reagents. Conjugation of green fluorescent protein (GFP) to cyclic polyarginine enables efficient uptake into HeLa cells. In contrast to non-canonical amino acid incorporation, this protein ligation method is easily applied to the modification of other native proteins of interest.

Author Contributions: Kalie A. Mix synthesized diazo reagents. Maria Glanz (FMP-Berlin) synthesized peptides and performed protein ligation experiments. Henry D. Herce (Dana Farber Cancer Institute) performed confocal microscopy experiments. Kalie A. Mix, Ronald T. Raines, Christian P.R. Hackenberger (FMP-Berlin), Maria Glanz, and Henry D. Herce designed experiments and analyzed results. Kalie A. Mix and Ronald T. Raines drafted this chapter and figures.

A1.1 Introduction

In 1988, it was serendipitously discovered that the *trans*-activator of transcription (TAT) protein from HIV can cross the plasma membrane of cultured cells.^{293,294} A small cationic peptide derived from HIV TAT can endow other proteins with the ability to enter cells when the peptide is attached as a fusion.²⁰⁴ Further studies of this peptide revealed that the guanidinium groups of the arginine residues were more important for facilitating cellular uptake than were other factors such as cationicity alone or the chirality of the peptide backbone.²⁹⁵ Since these discoveries were made, the TAT peptide, polyarginine (nona-arginine: R9, deca-arginine: R10), and their derivatives have been used to deliver a variety of cargo into cells, including peptides,²⁹⁶ proteins,^{204,216} small-molecule drugs,²⁹⁷ and nucleic acids.²⁹⁸ The mechanism of entry of CPP's such as TAT and R9 is the subject of much debate; depending on concentration,²⁹⁹ cargo,³⁰⁰ and CPP type,²⁹⁹ cellular uptake can occur via either endocytic or non-endocytic pathways.

A number of studies have shown that the spatial organization of the guanidinium groups in polyarginine is especially important for cell uptake. A seminal study demonstrated that decamers containing seven arginines with non-arginine amino acids spaced throughout the peptide facilitated cell uptake to a greater extent than hepta-arginine itself, indicating that the spacing and directionality of the guanidinium groups is important.³⁰¹ Increasing the rigidity of the peptide by using guanidinium-bearing oligoproline peptides was also shown to increase cell uptake of the peptide relative to polyarginine.³⁰² These insights were combined in a third study, in which both the rigidity of the peptide and the space between the guanidinium groups was increased (relative to linear CPP) by cyclization of R10 (cR10) and TAT (cTAT).³⁰³ cR10 and cTAT demonstrated enhanced transduction efficiency and fast uptake kinetics relative to their linear counterparts.

Recently, it was demonstrated that conjugation of cTAT to green fluorescent protein (GFP) endows the protein with the ability to enter the cytosol of cells.²⁰⁷ This delivery method suffers, however, from several drawbacks. First, conjugation of the cTAT peptide to GFP is achieved by copper-catalyzed azide–alkyne cycloaddition (CuAAC) between an azido group on the peptide and an alkynyl non-canonical amino acid on the protein. An ideal protein delivery method could be applied to native proteins and enzymes without the need for non-canonical amino acids. Second, the efficiency of delivery was low: cells needed to be treated with high concentrations of GFP (150 μ M) to observe internalization. The efficiency of cell uptake could potentially be increased by using cR10 instead of cTAT, as the linear R10 is more efficient at cellular transduction than is linear TAT.²⁹⁵ Finally, the triazole linkage between the CPP and GFP is not reversible. Hence, the cyclic CPP has the potential to disrupt intracellular localization or protein–protein interactions of the conjugated protein.

We sought to address these issues and improve upon the protein delivery method by using a diazo reagent for protein conjugation to CPPs. Because diazo compounds react with carboxyl groups, which are present in every protein, the requirement for non-canonical amino acid incorporation would be avoided. An ester linkage between the protein and peptide (instead of a triazole) would engender bioreversibility via cleavage by cellular esterases, and removal of the CPP in cells would prevent disruptions of protein binding and localization. Finally, the use of cR10 in place of cTAT peptide could improve the efficiency of cellular uptake so that the high protein concentrations required for uptake of GFP–cTAT could be avoided. These improvements could make the cyclic CPP delivery method translatable to protein therapeutics.

A1.2 Results and Discussion

Modification of GFP with cyclic cell-penetrating peptides

cR10 peptides bearing an azido or diazo group (Figure A1.1) were constructed using solid-phase peptide synthesis. A Cy3 dye was also incorporated into azido-cR10 to act as a FRET acceptor for GFP fluorescence. The purpose of this construct was to examine cleavage of the ester-linked cR10-Cy3 by cellular esterases.

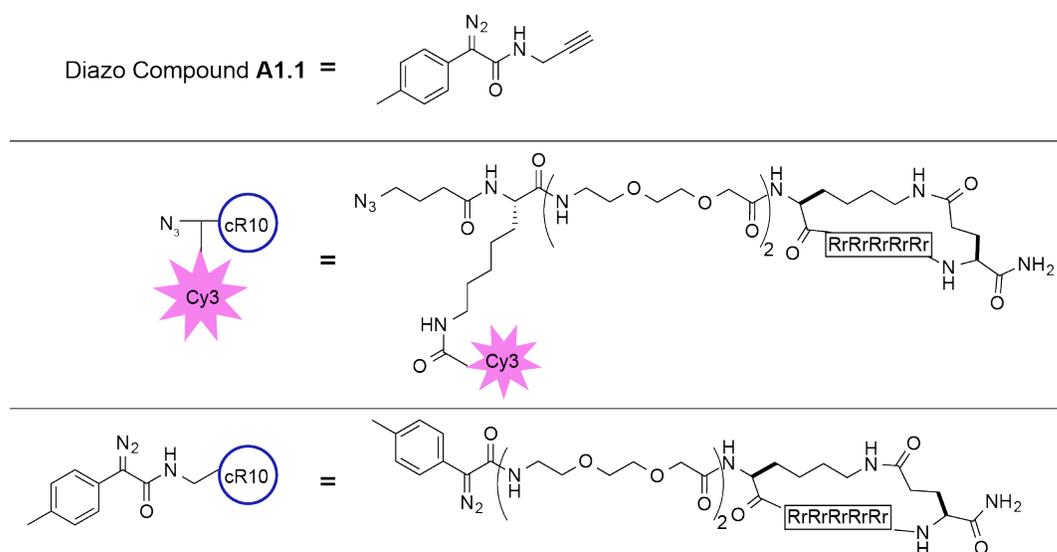


Figure A1.1 Ligation reagents used in this study. Capital letters within amino acid sequences indicate L- amino acids, and lowercase letters indicate D- amino acids.

Each CPP was conjugated to GFP via esterification alone (Figure A1.2A) or esterification followed by CuAAC (Figure A1.2B). Each method produced a mixture of eGFP bearing 0–2 esters per protein (Figure A1.3A,B). Notably, the esterification reaction between GFP and diazo-cR10 was carried out in a completely aqueous environment without the addition of acetonitrile. Acetonitrile was used in prior protein esterification experiments^{190,129} both to solubilize the diazo compound and to increase the rate of esterification by promoting protonation of the protein

carboxyl groups, which is crucial for esterification by diazo compounds.¹⁴⁴ The cationic peptide moiety imparts solubility to the diazo-cR10 reagent in this case, however, the esterification chemistry could still be hindered by the lack of organic cosolvent. The fact that esterification of the protein occurred in a completely aqueous environment (Figure A1.3A) underscores the optimization of the reactivity and selectivity of the *p*-methylphenylglycine-derived diazo compound.

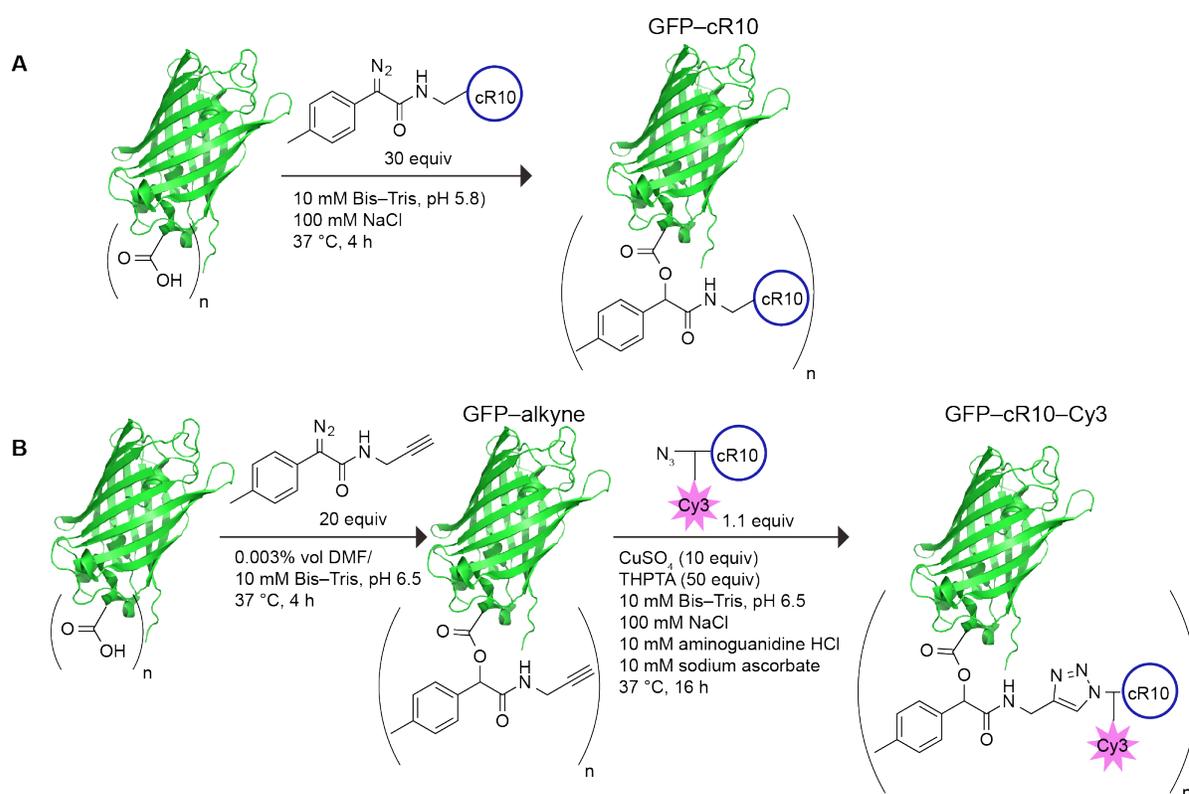


Figure A1.2 Conjugation of cyclic cell-penetrating peptides to GFP by esterification (A) or esterification followed by copper-catalyzed azide-alkyne cycloaddition (B).

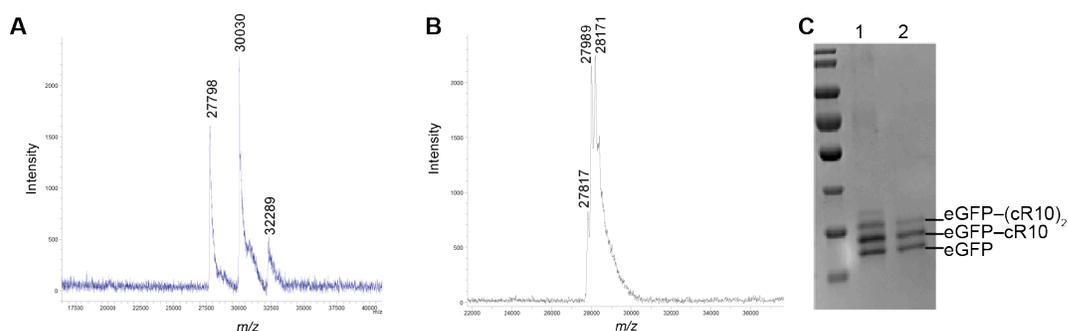


Figure A1.3 Characterization of GFP conjugates. (A) MALDI-TOF mass spectrum of eGFP esterified with diazo-cR10. eGFP $[M+H]^+$ 27793; found 27798, eGFP+1 cR10 ester $[M+H]^+$ 30021; found 30030, eGFP+2 cR10 esters $[M+Na]^+$ 32271; found 32289. (B) MALDI-TOF mass spectrum of GFP esterified with diazo compound A1.1 to form GFP-alkyne. eGFP $[M+Na]^+$ 27815; found 27817, eGFP+ 1 alkynyl ester $[M+Na]^+$ 27990; found 27989, eGFP+ 2 alkynyl esters $[M+Na]^+$ 28165; found 28171. (C) SDS-PAGE of GFP-alkyne modified with azido-cR10-Cy3. Lane 1: crude reaction mixture, Lane 2: reaction mixture after dialysis.

cR10 enables uptake of GFP into HeLa cells

To investigate cell uptake, HeLa cells were treated with either 5 or 10 μM GFP-cR10 (Figure A1.4). Strong green fluorescence was observed in both conditions, in contrast to previous work with cTAT-GFP in which high protein concentration (150 μM) was required to enable cell uptake.²⁰⁷ The efficient cell uptake engendered by cR10 relative to cTAT is in agreement with previous characterization of these peptides which demonstrated that liner R10 traverses the plasma membrane more effectively than does TAT.²⁹⁵

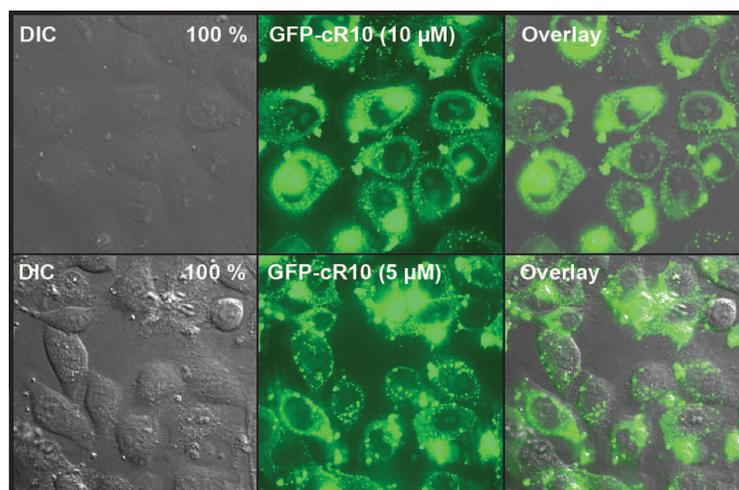


Figure A1.4 Confocal microscopy and DIC images of HeLa cells treated with 5 or 10 μM of GFP-cR10.

Next, bioreversibility was assessed by examining GFP cellular localization and GFP quenching due to Cy3 proximity. cR10 has a high affinity for RNA, which is abundant in the nucleolus. Thus, it is expected that GFP-cR10 would localize to the nucleolus of cells,²³⁹ which is observed in all images (Figure A1.4, Figure A1.5). Since GFP alone does not have affinity for RNA, removal of the cR10 ester by cellular esterases should cause GFP to localize to cellular compartments outside the nucleolus. Yet, nucleolar localization is observed in all images, indicating that the CPP ester is still intact. Bioreversibility was also examined using GFP-cR10-Cy3 construct. Cells treated with GFP-cR10-Cy3 show diminished green fluorescence relative to cells treated with GFP-cR10 alone. This result suggests that the cR10-Cy3 ester is still intact since the Cy3 dye quenches GFP fluorescence.

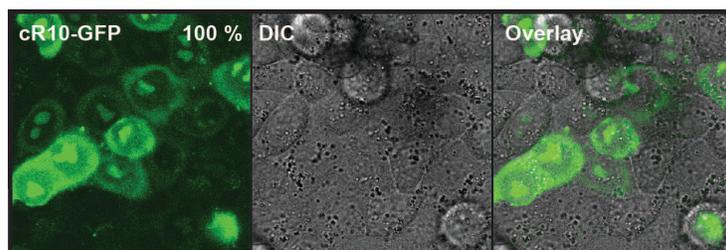


Figure A1.5 Confocal microscopy and DIC images of HeLa cells treated with 10 μ M GFP-cR10-Cy3.

A1.3 Future Directions

These experiments show that diazo-mediated protein esterification is an effective strategy for the ligation of CPPs to proteins in an aqueous environment. Additionally, cR10 endows GFP with the ability to enter cells at low concentrations. Future endeavors will focus on the bioreversibility of these modifications. The diazo-cR10-Cy3 conjugate is a well-poised system for characterization of bioreversibility in live cells. In future experiments, the ratio of GFP-Cy3 FRET (indicating intact esters) to GFP fluorescence (indicating cleaved esters) should be measured rather than qualitatively observing GFP quenching by Cy3 dye. Measurement of fluorescence in a ratiometric manner over time would be valuable because it would distinguish GFP uptake from ester cleavage (unlike measurement of GFP quenching alone) and provide insight into the kinetics of bioreversibility. It is possible that an ester bond between GFP and cR10 is a poor substrate for cellular esterases due to steric hindrance. To investigate this possibility, *in vitro* bioreversibility assays should be performed that measure cleavage of the cR10 ester using recombinant esterases or cell extracts. Cleavage of ester bonds could be measured using LC-MS to monitor the Cy3 dye.

A1.4 Acknowledgments

K.A.M. was supported by Molecular Biosciences Training Grant T32 GM007215 (NIH) and a fellowship from the University of Wisconsin–Madison College of Agricultural and Life Sciences. This work was supported by the Boehringer-Ingelheim Foundation (Plus 3 award) and the Fonds der Chemischen Industrie to C.P.R.H.

A1.5 Materials and Methods

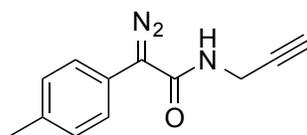
A1.5.1 General

Silica gel (40 μm ; 230–400 mesh) was from SiliCycle. Reagent chemicals were obtained from commercial sources and used without further purification. Dichloromethane and tetrahydrofuran were dried over a column of alumina. Thin-layer chromatography (TLC) was performed on plates of EMD 250 μm silica 60-F₂₅₄. The phrase “concentrated under reduced pressure” refers to the removal of solvents and other volatile materials using a rotary evaporator at water aspirator pressure (<20 torr) while maintaining a water bath below 40 °C. Residual solvent was removed from samples at high vacuum (<0.1 torr). ¹H and ¹³C NMR spectra for all compounds were acquired on Bruker spectrometers in the National Magnetic Resonance Facility at Madison operating at 400 or 500 MHz. Chemical shift values (δ) are reported in units of ppm relative to an internal standard (residual solvent or TMS). Electrospray ionization (ESI) mass spectrometry for small-molecule characterization was performed with a Micromass LCT at the Mass Spectrometry Facility in the Department of Chemistry at the University of Wisconsin–Madison. Matrix-assisted laser desorption-ionization–time-of-flight (MALDI–TOF) mass spectrometry was performed using a Bruker Microflex at FMP–Berlin. Peptides were synthesized with an Activo-P11 Peptide Synthesizer (Activotec SPP Ltd., UK) via standard Fmoc-based conditions

using HOBt/HBTU/DIPEA activation and piperidine Fmoc deprotection in DMF. Confocal microscopy was performed using an UltraVIEWVoX spinning disc system (Perkin–Elmer, United Kingdom) on a Nikon Ti microscope equipped with an oil immersion Plan Apochromat VC x60/1.45 NA using a 488-nm excitation laser and 521-nm emission filter.

A1.5.2 Chemical synthesis

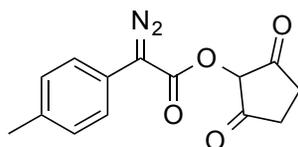
*Synthesis of α -diazo *p*-methylphenylacetic propargyl amide (A1.1)*



A1.1

Compound **A1.1** was synthesized as described previously.²⁴¹ Spectral data and yields match those reported previously.

*Synthesis of α -diazo *p*-methylphenylacetic *N*-hydroxysuccinimidyl ester (A1.2)*



A1.2

Compound **A1.2** was synthesized as described previously.²⁴¹ Spectral data and yields match those reported previously.

A1.5.3 Peptide Synthesis

Synthesis of azido-cR10-Cy3

The cR10 was synthesized on a 0.1-mmol scale on Rink amide resin with a loading of 0.78 mmol/g. The synthesis was carried out on a PTI synthesizer with double couplings of each amino acid (5 equiv of amino acid for 40 min) in DMF using standard Fmoc chemistry protocols. After the final coupling, the Fmoc-protected peptide was treated with Pd(PPh₃)₄ (24 mg, 20 μmol, 20 mol%) and phenylsilane (308 μL, 2.5 mmol) in 4 mL dry DCM for 1 h to cleave the alloc and allyl protecting groups. Cyclization was achieved by overnight incubation with HATU (2 equiv) and 4 equiv DIEA in DMF. The Fmoc group was removed using 20% piperidine in DMF. Fmoc-Lys(Dde)-OH was coupled to the N-terminus by using a standard peptide coupling reaction. After Dde removal with 3% v/v hydrazine in DMF (3 × 3 min), the Cy3-COOH (1.5 equiv) was coupled to the peptide by incubation with HATU (1.5 equiv) and DIPEA (3 equiv) for 6 h. After final Fmoc-deprotection the peptide was coupled with 4-azidobutanoic acid (5 equiv), HATU (5 equiv), and DIEA (10 equiv). The peptide was cleaved from the resin by treatment with 4 mL of 95:2.5:2.5 TFA/TIS/H₂O (for 3 h and precipitated in cold diethyl ether. The crude peptide was purified by preparative reverse-phase C₁₈ HPLC, eluting with a linear gradient of water (containing 0.1% v/v TFA) and acetonitrile.

Data for azido-cR10-Cy3: MALDI-TOF *m/z* calcd for azido-cR10-Cy3 [M+H]⁺ 2787; found 2791.

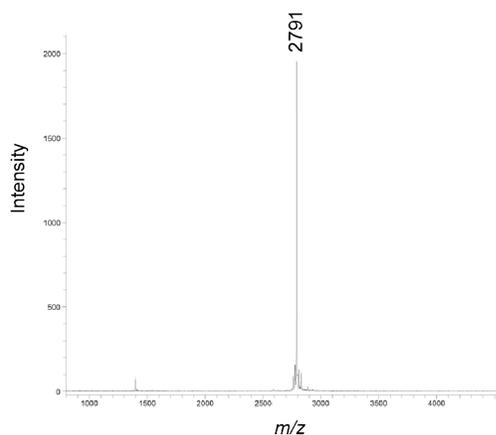


Figure A1.6 MALDI-TOF mass spectrum of azido-cR10-Cy3

Synthesis of amino-cR10

The cR10 was synthesized on a 0.1-mmol scale on a Rink amide resin with a loading of 0.78 mmol/g. The synthesis was carried out on a PTI synthesizer with double couplings of each amino acid (5 equiv amino acid for 40 min) in DMF using standard Fmoc chemistry protocols. After the final building-block coupling the Fmoc-protected peptide was treated with Pd(PPh₃)₄ (24 mg, 20 μmol, 20 mol%) and phenylsilane (308 μl, 2.5 mmol, 2.5 equiv) in 4 mL dry DCM for 1 h to cleave the alloc and allyl protecting groups. Cyclization was achieved by overnight incubation of the peptide with 2 equiv of HATU and 4 equiv of DIEA in DMF. The Fmoc group was then removed using 20% v/v piperidine in DMF. The peptide was cleaved from the resin by treatment with 4 mL of 95:2.5:2.5 TFA/TIS/H₂O for 3 h and precipitated in cold diethyl ether. The crude peptide was purified by preparative reverse-phase C₁₈ HPLC eluting with a linear gradient of water (containing 0.1% v/v TFA) and acetonitrile.

Synthesis of diazo-cR10

Purified amino-cR10 (9 mg, 4.27 μmol) and compound **A1.2** (1.75 mg, 1.5 equiv) were dissolved in 300 μL of DMF. DIEA was added (1.1 μL , 1.5 equiv) to the resulting solution. The reaction mixture was shaken for 4 h, and the peptide was precipitated in diethyl ether.

Data for diazo-cR10: MS (ESI⁺) m/z calcd for diazo-cR10 $[\text{M}+3\text{H}]^{+3}$ 752.8; found 753.4.

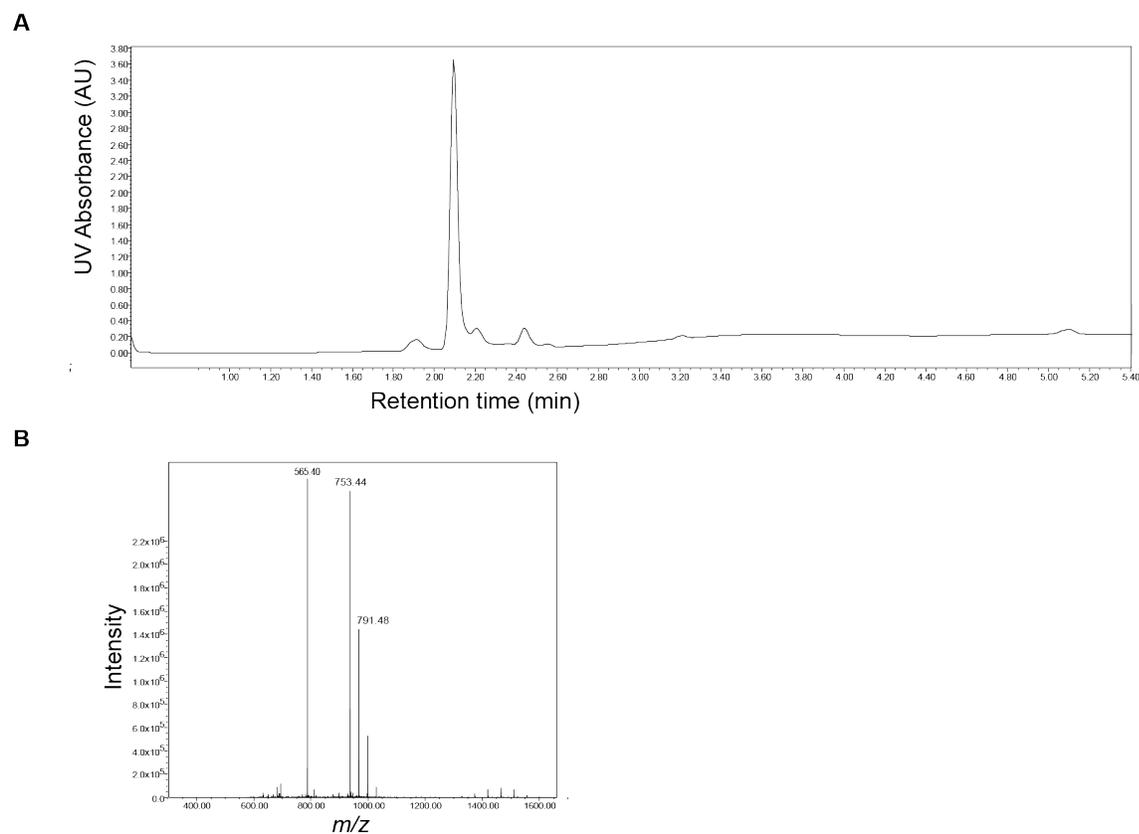


Figure A1.7 A) HPLC chromatogram of diazo-cR10 after precipitation of the peptide in diethyl ether without further purification. B) ESI⁺ mass spectrum of diazo-cR10.

A1.5.4 GFP Labeling

GFP labeling by esterification

eGFP (C70M, S143C) was produced and purified as reported previously.²⁰⁷ eGFP (18 nmol) was dialyzed into 300 μ L of 10 mM Bis-Tris buffer, pH 6.5, containing NaCl (100 mM). Diazo-cR10 (30 equiv) was dissolved in 60 μ L of the same buffer, and the resulting solution was added to the eGFP solution. The reaction mixture was shaken at 37 °C for 4 h. The reaction mixture was desalted and exchanged into PBS by a Zebaspin column (Thermo Fisher) with a MWCO of 7 kDa. Esterification was analyzed by MALDI-TOF mass spectrometry (Figure A1.3A).

GFP labeling by esterification followed by CuAAC

eGFP (16.2 nmol) was dialyzed into 360 μ L of 10 mM Bis-Tris buffer, pH 5.8, and diazo compound **A1.1** (20 equiv) in 1 μ L of DMF was added. The reaction mixture was shaken at 37 °C for 4 h. The reaction mixture was desalted and exchanged into PBS by a Zebaspin column (Thermo Fisher) with a MWCO of 7 kDa. Analysis by MALDI showed 0–2 esters per GFP (Figure A1.3B). To the GFP-alkyne, azido-cR10-Cy3 (1.1 equiv), CuSO₄ (10 equiv) and THPTA (50 equiv) were added. The reaction was carried out in 10 mM Bis-Tris buffer, pH 6.5, containing NaCl (100 mM), aminoguanidine hydrochloride (10 mM), and sodium ascorbate (10 mM). The reaction mixture was shaken at 37 °C for 16 h, purified by dialysis into PBS, and analyzed by SDS-PAGE (Figure A1.3C).

A1.5.5 Cell Culture and Confocal Microscopy

HeLa cells were cultured in DMEM supplemented with 10% v/v FBS and 1× penicillin/streptomycin. Cells were seeded at a subconfluent density 24 h prior to treatment in glass multi-well microscopy chambers (Evotec, Hamburg, Germany). Cells were incubated with 5 or 10 μM GFP-cR10, or 10 μM GFP-cR10-Cy3 for 1 h 37 °C in serum-free medium. Live cells were then washed with DMEM and imaged using a confocal laser-scanning microscope.

Appendix 2

Cellular Delivery of Cas9 for Genome Editing

Abstract

CRISPR–Cas9 technology enables genome-editing or gene-silencing in cells. Use of this technology in mammalian cells now requires introduction of exogenous Cas9 protein and a guide RNA, either via transfection of DNA that encodes these components or via transfection of the protein:RNA complex. Both of these methods require electroporation or cationic lipids, which prohibit in vivo applications due to cytotoxicity. Cellular delivery of the Cas9:RNA complex by esterification of carboxyl groups using a diazo compound could be more easily translated to clinical applications. Here, we show that Cas9 is esterified by a diazo compound, and we present methods for assaying cellular delivery and genome editing.

A2.1. Introduction

CRISPR–Cas9 technology has enabled numerous advances in nearly every field of biological sciences due to the ability to rapidly induce a permanent change in the genome of an organism.^{304,305} In this system, Cas9 nuclease induces a double-strand break in DNA at a locus targeted by a complementary short guide RNA (sgRNA). This double-strand break is either repaired via a non-homologous end-joining mechanism, which silences the gene, or homology-directed repair, which changes the sequence of the target locus using a template DNA strand. The ability to target a gene of interest by simply changing the sequence of the sgRNA is a significant advantage over other genome-editing technologies such as transcription activator-like effector nucleases (TALENs) and zinc finger nucleases (ZFNs), in which the protein itself needs to be re-designed to target different DNA sequences.³⁰⁶

CRISPR–Cas9 presents exciting opportunities in the development of therapeutics. Diseases caused by loss-of-function mutations could be treated by editing the gene sequence that encodes the mutation *in vivo*.³⁰⁷ Alternatively, certain cell types, such as T cells, could be edited *ex vivo* to confer desirable traits such as resistance to HIV infection or enhanced ability to kill cancer cells, before re-introducing the cells into a patient.³⁰⁸

A primary challenge in both strategies is the introduction of exogenous Cas9 and sgRNA into mammalian cells. Transfection of DNA encoding the protein and RNA components results in long-term expression of Cas9 and RNA, which increases the risk of off-target activity.²¹¹ The introduction of Cas9 and sgRNA as a ribonucleoprotein (RNP) complex presents a solution to this problem because proteolytic degradation of Cas9 limits exposure to genome editing activity.^{212,211} In previous work, the RNP complex has been introduced into cells by electroporation²¹² or cationic lipids.²¹¹ Both methods suffer from limitations that prevent

translation into the clinic: electroporation cannot be used for in vivo editing, and cationic lipids are cytotoxic.³⁰⁹

Delivery of the RNP complex into cells using bioreversible esterification could circumvent these issues. Here, we describe a system to assay the cellular internalization and genome-editing ability of esterified Cas9.

A2.2 Results and Discussion

sgRNA sequence design and in vitro DNA cleavage

A number of factors influence the efficiency of CRISPR-mediated genome editing, including the sequence, accessibility, and epigenetic status of the target locus.³¹⁰ A support vector machine model can be used to score the predicted activity of an sgRNA based on these properties of the target locus.³¹⁰ This algorithm was used to design five CRISPR RNAs (crRNAs) that target the gene encoding green fluorescent protein (GFP) and were predicted to have high editing activity at this locus. The designed crRNAs are extended with a sequence that is complementary to a universal trans-activating CRISPR RNA (tracrRNA), which anneals to the crRNA to produce a full-length sgRNA that binds to Cas9. The sgRNA:Cas9 complex is then able to carry out DNA cleavage. When incubated with Cas9 and a plasmid containing the GFP gene, all five sgRNAs enable induction of double-strand DNA breaks (Figure A2.1).

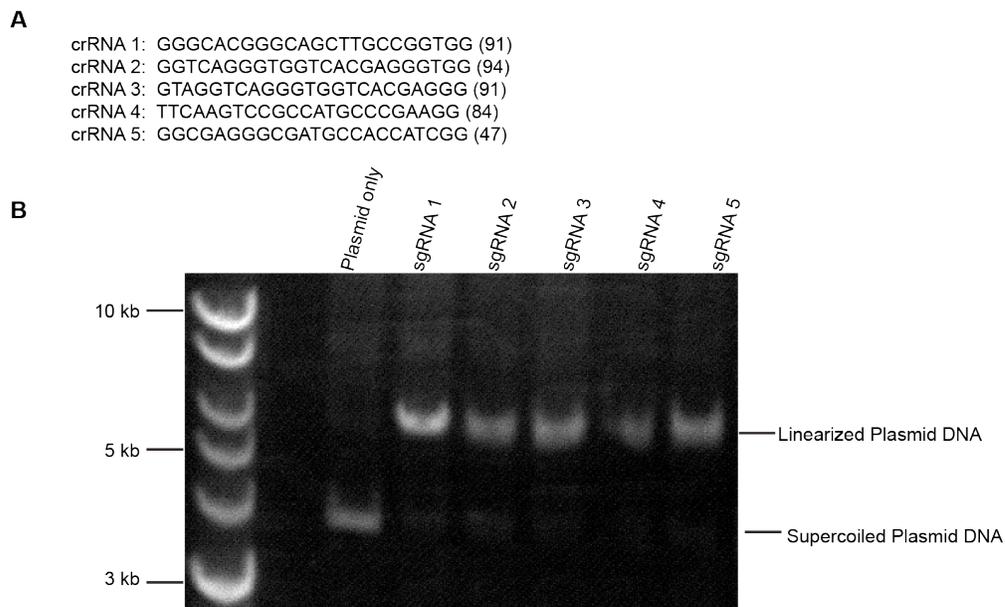


Figure A2.1 (A) Sequences and scores of five crRNA sequences that target the GFP gene. (B) In vitro cleavage of DNA plasmid containing the GFP gene.

Cas9 esterification by diazo compound 3.1

Cas9 is a large, cationic protein (molecular mass: 162 kDa, pI ~9),³¹¹ which makes labeling and characterization especially challenging. Carboxyl groups next to cationic residues generally have relatively low pK_a values, which slows the rate of esterification under the conditions (pH 6.5) used for the diazo compound-mediated esterification reaction, as many carboxyl groups will not be protonated.¹⁴⁴ Additionally, MALDI-TOF mass spectrometry in this size range is unable to resolve peaks separated by 175 Da (which is the mass added to form each ester moiety) to measure the absolute number of labels per protein. We overcame this challenge by using mass spectrometry analysis software to find an average mass of the protein population in comparison to unlabeled protein. The average mass of the protein population increased with increasing

equivalents of diazo compound **3.1**, which indicates that Cas9 was esterified in a dose-dependent fashion (Figure A2.2).

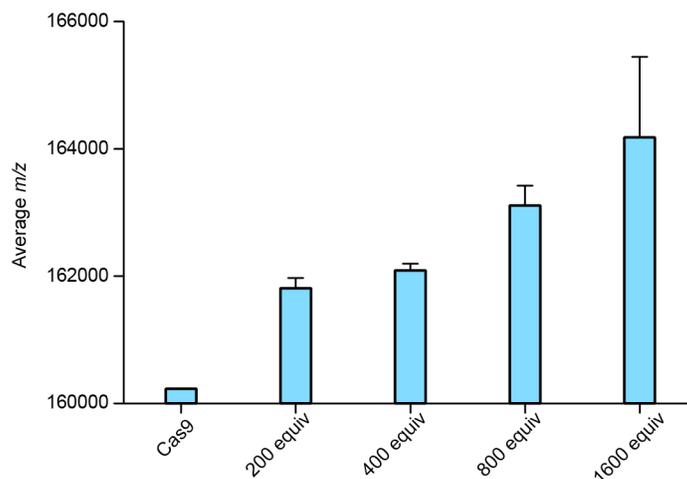


Figure A2.2 Average m/z of Cas9 after treatment with 200–1600 equiv of diazo compound **3.1**.

RNA esterification

Esterification of Cas9 could hinder sgRNA binding, and certain diazo compounds are known to esterify RNA itself.^{166,172} To determine whether the esterification reaction should be carried out on Cas9 alone or the Cas9:sgRNA complex, we investigated the reactivity of compound **3.1** with two short RNA strands: AUGC (which contains only phosphodiester groups) and pAUGC (which contains both phosphodiester groups and a 5' phosphomonoester group). Diazo compound **3.1** was incubated with each of these RNA strands under the same conditions used for protein labeling, and the reaction was analyzed by LC–MS. Analysis of the reaction between AUGC and compound **3.1** reveals a mass peak for singly modified RNA, which suggests that a phosphodiester group was modified; however, the LC chromatogram indicates that this is a very

minor product (Figure A2.3). In contrast, the reaction between pAUGC and compound **3.1** contains mass peaks for pAUGC with one and two modifications, and the LC chromatogram indicates that the singly modified RNA is a major product (Figure A2.4). The pK_a of a phosphodiester group is near 1.5,³¹² making these groups unlikely to be esterified by the diazo compound. This low pK_a is in agreement with the observation that the phosphoryl modification was a minor product in the reaction of **3.1** with AUGC. The pK_a of the 5' phosphoester is, however, ~ 6 ,³¹² which makes it a good substrate for modification by a diazo compound.¹⁷² This high pK_a is in agreement with the observation that a singly modified pAUGC was a major product in the reaction of **3.1** with pAUGC.

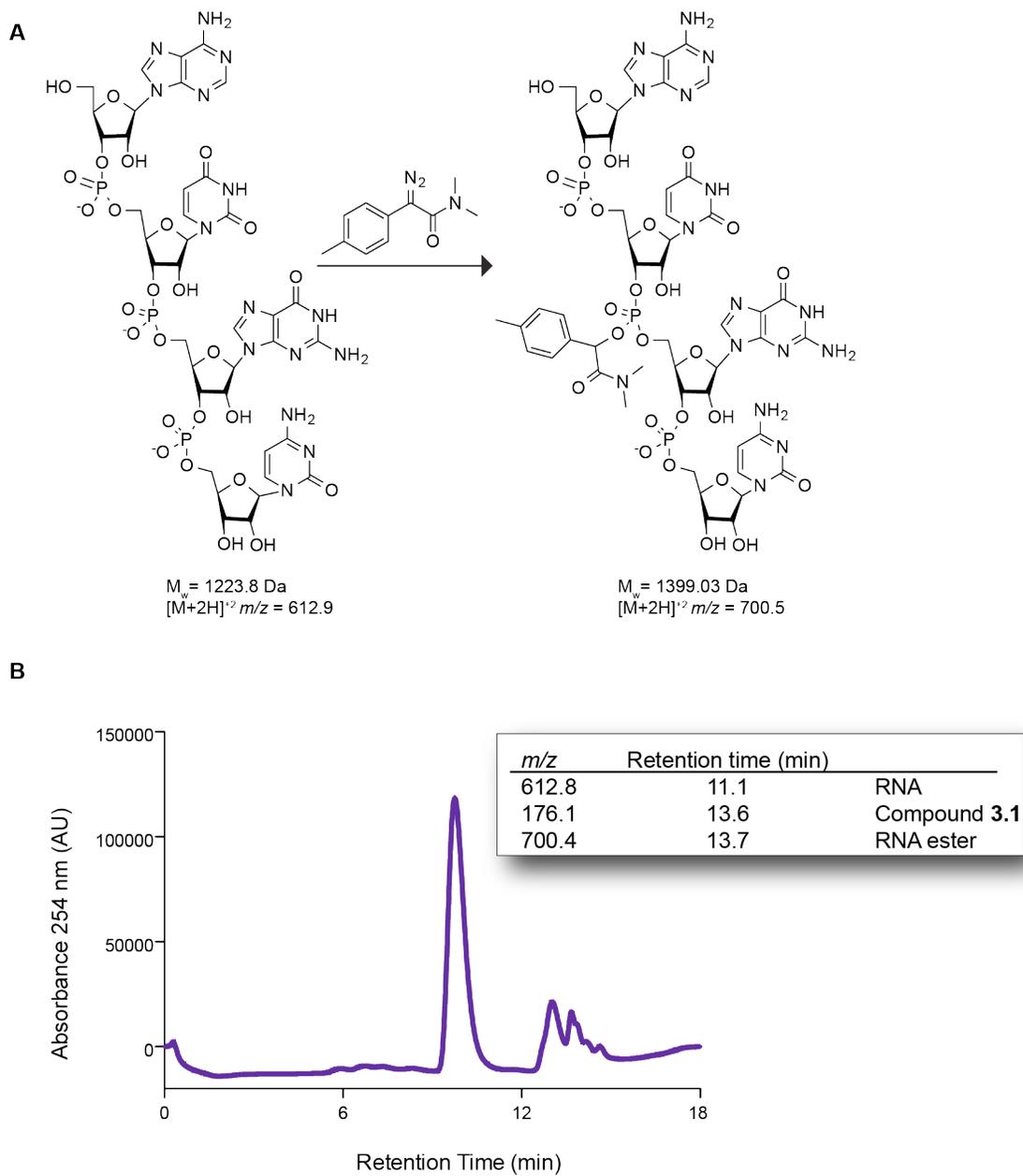


Figure A2.3 (A) Structures of AUGC and AUGC esterified with compound **3.1**. (B) LC chromatogram of the reaction mixture and observed mass peaks (inset).

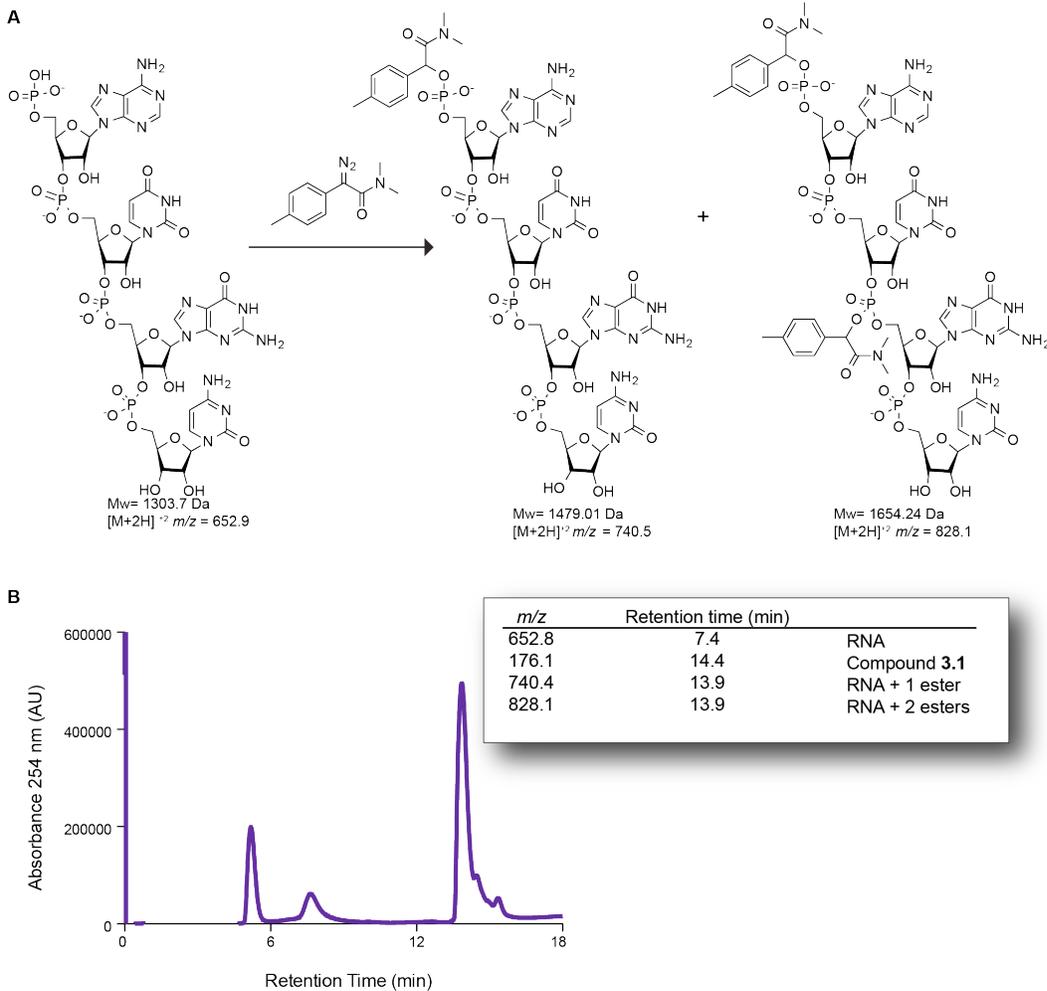


Figure A2.4 (A) Structures of pAUGC and pAUGC esterified with compound **3.1**. (B) LC chromatogram of the reaction between pAUGC and compound **3.1**, with observed mass peaks (inset).

Characterization of genome editing

HEK293T cells that stably express GFP (HEK293T–GFP) were treated with either Cas9 modified with diazo compound **3.1** (Cas9–**3.1**) or Cas9 complexed with cationic lipids (Cas9 + RNAiMAX). Each sample also contained the sgRNA. Successful internalization of the Cas9:sgRNA complex is expected to result in silencing of the GFP gene. The fluorescence of HEK293T–GFP cells after a 48-h incubation with Cas9–**3.1** or Cas9+RNAiMAX was measured (Figure A2.5). No significant decrease in fluorescence (relative to controls) was observed for cells treated with either Cas9–**3.1** or Cas9+RNAiMAX.

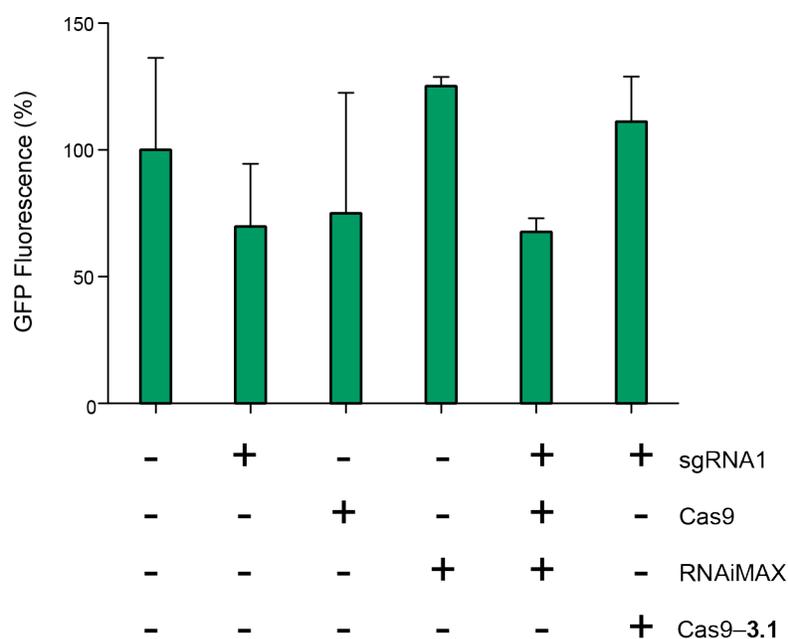


Figure A2.5 GFP fluorescence after 48 h incubation with CRISPR components. Excitation wavelength = 488 nm, emission wavelength = 514 nm.

To characterize DNA cleavage directly, we also performed a T7 endonuclease I (T7E1) assay. In this assay, the GFP gene was amplified from genomic DNA using PCR. If the gene was cleaved successfully by Cas9, then the PCR products will contain random insertions and deletions imparted by the cellular double-strand break repair machinery. PCR products were melted and re-annealed in a thermocycler to form duplexes, which will be mismatched (heteroduplexes) if the gene contained insertions and deletions. These mismatches were recognized and cleaved by T7E1. T7E1 cleavage products were indeed apparent in DNA from cells treated with Cas9+RNAiMAX, but not in DNA from cells treated with Cas9-3.1 (Figure A2.6).

Internalization of sgRNA-Cas9 complexes in HEK293T-GFP cells

To compare the cellular uptake of Cas9-3.1 and Cas9+RNAiMAX, we utilized a tracrRNA-ATTO dye conjugate to visualize sgRNA-Cas9 internalization using confocal microscopy. HEK293T-GFP cells were treated with Cas9-3.1 or Cas9+RNAiMAX, each with the tracrRNA-ATTO dye contained in the sgRNA (Figure A2.7). Cells treated with Cas9+RNAiMAX displayed punctate pink staining, which indicates internalization of the Cas9 and sgRNA-ATTO. In contrast, cells treated with Cas9-3.1 displayed no pink staining.

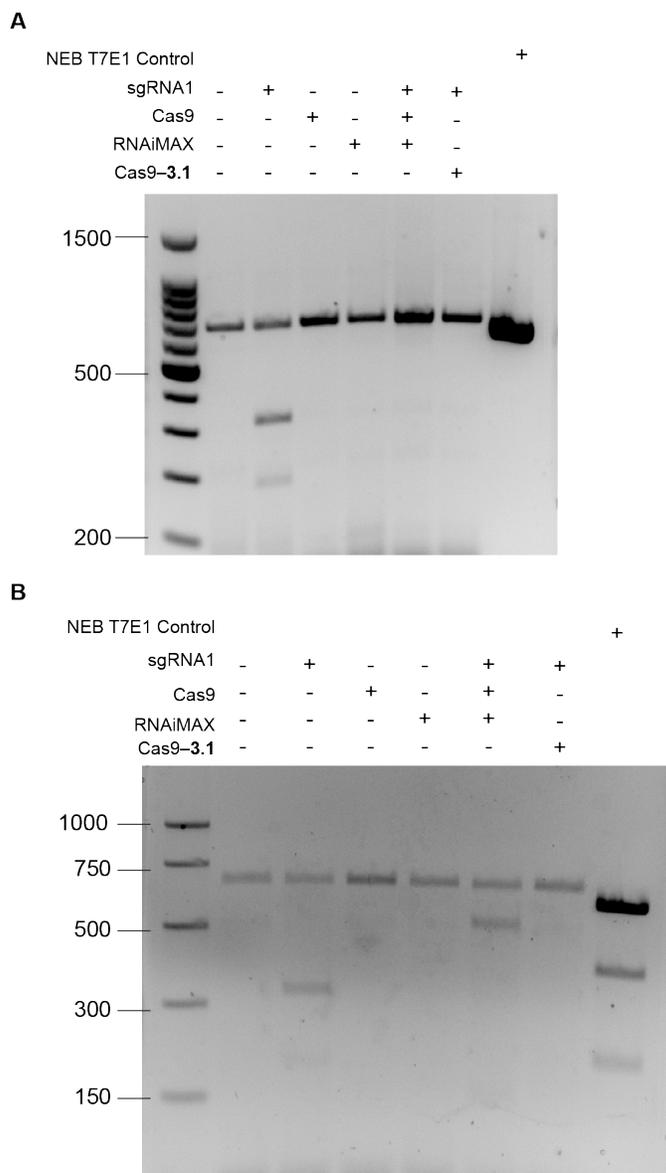


Figure A2.6 (A) DNA after PCR amplification of 659 bp amplicon within the GFP gene. (B) DNA after PCR amplification, duplex re-annealing, and subsequent T7E1 digestion. Expected T7E1 cleavage fragment sizes are ~490 and ~170 bp. NEB T7E1 control is a set of plasmids and primers included in the T7E1 EnGen kit from New England Biolabs, Inc., which is expected to produce a ~600 bp amplicon with ~200 and ~400 bp cleavage fragments.

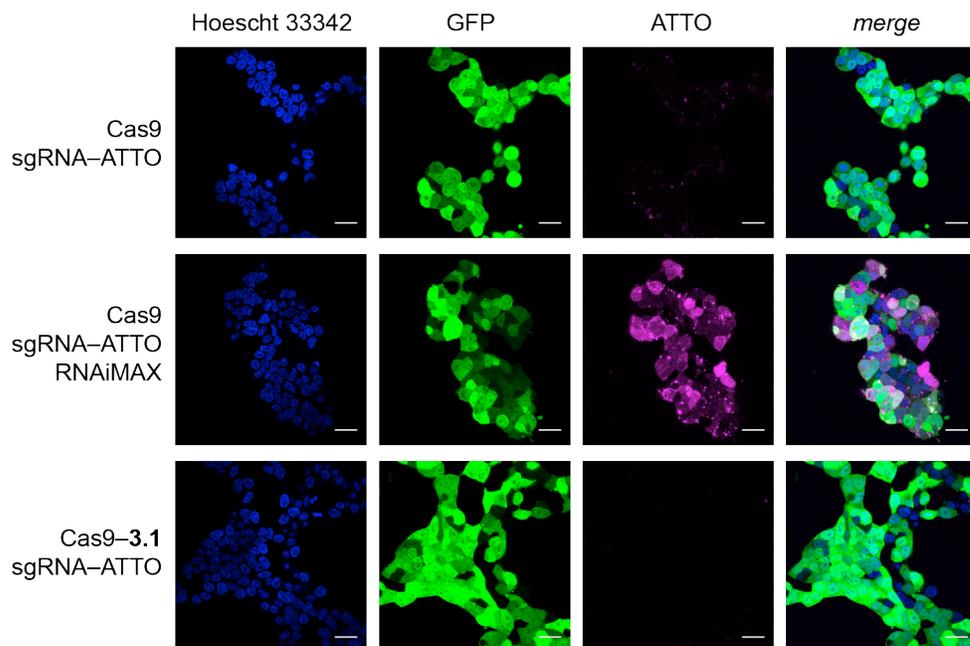


Figure A2.7 Confocal microscopy images of HEK293T–GFP cells treated with Cas9–sgRNA/ATTO complexes. Scale bar: 25 μ m.

A2.3 Future Directions

The lack of DNA cleavage observed in the T7E1 assay and lack of pink fluorescence observed by confocal microscopy suggest that the esterified Cas9 is not internalized into HEK293T–GFP cells. An alternative explanation is that the esterified Cas9 is unable to bind sgRNA, resulting in both an inability to cleave DNA and a lack of pink fluorescence, even if the protein were entering cells.

Future endeavors should aim to optimize the esterification procedure to ensure both maximal protein labeling and maintenance of the ability to bind RNA. First, the absolute number of labels could be measured with ESI–TOF mass spectrometry rather than MALDI–TOF mass spectrometry, as the former ionization method would produce resolved peaks at this high mass

range. This analysis would enable optimization of protein labeling to ensure that a sufficient number of esters are added per protein for cellular entry. Second, the possibility of esterifying the Cas9:sgRNA complex should be explored further. Although diazo compound **3.1** esterifies unbound RNA, the Cas9 protein could hinder attack of the sgRNA by the diazo compound because the sgRNA binds in a groove of the Cas9 protein. Alternatively, if Cas9 is esterified prior to sgRNA binding, the RNA-binding ability of esterified Cas9 should be measured. The K_d values for Cas9 and esterified Cas9 could be measured using an established assay system.³¹³

A second part of this assay system that could be optimized further is the cell line. The HEK293T–GFP cell line was chosen with the goal of using a decrease in GFP fluorescence to characterize silencing of the GFP gene by CRISPR-mediated genome editing. GFP is notorious, however, for its proteolytic stability in mammalian cells.³¹⁴ Thus, genome editing and GFP gene silencing could be occurring without a concurrent loss of GFP fluorescence because GFP can persist in cells for several days prior to proteolytic degradation.³¹⁵ This longevity is apparent in the observation that cells treated with Cas9+RNAiMAX demonstrate no significant decrease in GFP fluorescence but show DNA cleavage by the T7E1 assay. A better choice of cell line would be one that stably expresses GFP with a PEST degradation signal, which would shorten the GFP lifetime to approximately 10 h.³¹⁵

A2.4 Acknowledgments

K.A.M. was supported by Molecular Biosciences Training Grant T32 GM007215 (NIH) and a fellowship from the University of Wisconsin–Madison College of Agricultural and Life Sciences.

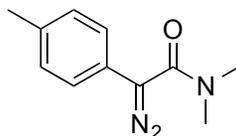
A2.5 Materials and Methods

A2.5.1 General

Silica gel (40 μm 230–400 mesh) was from SiliCycle. Reagent chemicals were obtained from commercial sources and used without further purification. Dichloromethane and tetrahydrofuran were dried over a column of alumina. Thin-layer chromatography (TLC) was performed on plates of EMD 250 μm silica 60-F₂₅₄. The phrase “concentrated under reduced pressure” refers to the removal of solvents and other volatile materials using a rotary evaporator at water aspirator pressure (<20 torr) while maintaining a water bath below 40 °C. Residual solvent was removed from samples at high vacuum (<0.1 torr). ¹H and ¹³C NMR spectra for all compounds were acquired on Bruker spectrometers in the National Magnetic Resonance Facility at Madison operating at 400 or 500 MHz. Chemical shift values (δ) are reported in units of ppm relative to an internal standard (residual solvent or TMS). Electrospray ionization (ESI) mass spectrometry for small-molecule characterization was performed with a Micromass LCT at the Mass Spectrometry Facility in the Department of Chemistry at the University of Wisconsin–Madison. LC–MS analysis of RNA was performed on a Shimadzu LC-MS2020 instrument with a single quadrupole mass analyzer.

A2.5.2 Chemical Synthesis

Synthesis of diazo compound 3.1



Diazo compound **3.1** was synthesized as described previously.²⁴¹ Spectral data and yields match those reported previously.

A2.5.3 Design of crRNA

A support vector machine model can be used to score the predicted activity of a crRNA based on properties of the target locus³¹⁰, and this model has been used to create an online sgRNA design tool (sgRNA Scorer 2.0: <https://crispr.med.harvard.edu/>). This online tool was used to design five sgRNAs that target the GFP gene and were predicted to have high editing activity at this locus.

A2.5.4 In vitro DNA cleavage

Recombinant Cas9 bearing two nuclear localization signals was obtained from Aldevron (Madison, WI). crRNA and tracrRNA were obtained from Integrated DNA Technologies (Coralville, IA). Cas9 (50 nM) was incubated with a plasmid containing the target loci within the GFP gene (600 ng), and crRNA (50 nM) and tracrRNA (50 nM) in NEBuffer 3 (20 µL total reaction volume) for 1 h at 37 °C. Samples were loaded on a 0.8% agarose gel, and migration distance was analyzed.

A2.5.5 RNA esterification

Synthetic RNA containing only phosphodiester (sequence: AUGC) or both phosphomonoesters and phosphodiester (pAUGC) were obtained from IDT (Coralville, IA). Each RNA (4.7 nmol) was dissolved in 200 µL of 10 mM Bis-Tris buffer, pH 6.5. Diazo compound **3.1** (940 nmol) was dissolved in 50 µL of acetonitrile. The RNA solution was mixed with the diazo compound solution and incubated at 37 °C for 4 h in DNA lo-bind tubes. The reaction mixture was analyzed by LC-MS.

A2.5.6 Cas9 protein labeling

Cas9 (29 μM in 10 mM Bis-Tris buffer, pH 6.5) was incubated with 200–1600 equiv of diazo compound **3.1** in acetonitrile (final solution composition: 20% v/v MeCN in Bis-Tris buffer). The reaction mixture was incubated for 4 h at 37 °C in protein lo-bind tubes and analyzed by MALDI–TOF mass spectrometry.

A2.5.7 Delivery of protein in cell culture

sgRNA stock solution was prepared by adding 1 μL of crRNA (100 μM) and 1 μL of tracrRNA (100 μM) to 98 μL of nuclease-free duplex buffer (IDT) and incubating at 95 °C for 5 min. Cas9+RNAiMAX transfection mix was prepared by combining 4.5 μL of Cas9 (1 μM) with 4.5 μL of sgRNA and 3.6 μL of RNAiMAX (Thermo Fisher) in opti-MEM. Cas9–**3.1** transfection mix was prepared by combining 4.5 μL of diazo-modified Cas9 (1 μM) with 4.5 μL of sgRNA in opti-MEM.

HEK293T cells that stably express GFP (Cell Biolabs; San Diego, CA) were cultured in DMEM supplemented with fetal bovine serum (FBS; 10% v/v), penicillin (100 U/mL), streptomycin (100 $\mu\text{g}/\text{mL}$), and GlutaMAX (2 mM). Cells were released from the culturing flask by using 0.05% w/v trypsin–EDTA and added to DMEM containing FBS (10% v/v) and GlutaMax (2 mM). Cells were collected by centrifugation and resuspended in DMEM containing FBS (10% v/v) and GlutaMax (2 mM), and then diluted to a density of 400,000 cells/mL. Cells were reverse-transfected by adding 100 μL of HEK293T–GFP cells to 50 μL aliquots of each transfection mix (or control aliquots containing individual reagents) in the wells of a 96-well

sterile plate. Cells were incubated at 37 °C under 5% CO₂ in a cell-culture incubator for 48 h. GFP fluorescence (ex 488 nm, em 514 nm) was measured with a Tecan plate reader.

A2.5.8 Detection of genomic modifications using T7E1 assay

Following incubation with Cas9–**3.1** or Cas9+RNAiMAX (see above), cells were rinsed with PBS. DNA was extracted from adherent cells using 50 µL of QuickExtract extraction solution (Epicentre). Each sample was then incubated at 65 °C for 10 min, and then 98 °C for 5 min. The cell lysate was diluted 1:5 with TE buffer. PCR mix was prepared by incubating 2.5 µL of diluted cell lysate, 2.5 µL of forward primer (10 µM), 2.5 µL of reverse primer (10 µM), 25 µL of Q5 polymerase Master Mix (New England Biolabs; Ipswich, MA) and nuclease-free water (total reaction volume: 50 µL). PCR was performed using the following cycling conditions:

Step	Temp (°C)	Time (s)
Initial Denaturation	98	30
Denaturation	98	5
Annealing	60	10
Extension	72	20
(35 cycles of step 2–4)		
Final Extension	72	2 min
Hold	4	---

GFP Forward primer: TGAGCAAGGGCGAGGAGCTGTTCA ($T_m = 64.5$ °C)

GFP Reverse primer: AGGACCATGTGATCGCGCTTCTCGT ($T_m = 64.0$ °C)

Target size: 659 bp

Aliquots of the PCR products (5 µL of each sample) were analyzed on a 2% w/v agarose gel. A second aliquot of the PCR products (5 µL of each sample) was mixed with 2 µL of NEBuffer 2 and 12 µL of nuclease-free water. The resulting solution was incubated using the following cycling conditions:

Step	Temp	Ramp Rate	Time
Initial Denaturation	95 °C		5 min
Annealing	95–85 °C	-2 °C/s	
	85–25 °C	-0.1 °C/s	
Hold	4 °C		infinite

EnGen T7 endonuclease (1 μ L) from the EnGen Mutation Detection Kit (New England Biolabs; Ipswich, MA) was added to each sample, and the resulting solution was incubated at 37 °C for 5 min. Proteinase K (1 μ L) from the EnGen Mutation Detection Kit was added to each sample, and the resulting solution was incubated at 37 °C for 5 min. The samples were analyzed on a 2% w/v agarose gel.

A2.5.9 Visualization of Cas9:sgRNA/ATTO internalization

sgRNA–ATTO was prepared by adding 1 μ L of crRNA (100 μ M) and 1 μ L of tracrRNA–ATTO (100 μ M) to 98 μ L of nuclease-free duplex buffer (IDT) and incubating at 95 °C for 5 min. Cas9+RNAiMAX transfection mix was prepared by combining 4.5 μ L of Cas9 (1 μ M) with 4.5 μ L of sgRNA–ATTO and 3.6 μ L of RNAiMAX (Thermo Fisher) in opti-MEM. Cas9–3.1 transfection mix was prepared by combining 4.5 μ L of diazo-modified Cas9 (1 μ M) with 4.5 μ L of sgRNA/ATTO in opti-MEM.

HEK293T cells that stably express GFP (Cell Biolabs; San Diego, CA) were cultured in DMEM supplemented with fetal bovine serum (10% v/v), penicillin (100 U/mL), and streptomycin (100 μ g/mL). Cells were released from the culturing flask by using 0.05% w/v trypsin–EDTA, and added to DMEM containing FBS (10% v/v) and GlutaMax (2 mM). Cells were collected by centrifugation and resuspended in DMEM containing FBS (10% v/v) and GlutaMax (2 mM), and then diluted to a density of 400,000 cells/mL. Cells were reverse transfected by adding 100 μ L of HEK293T–GFP cells to 150 μ L aliquots of transfection mix in

the wells of an 8-well sterile microscopy dish. Cells were incubated at 37 °C under 5% CO₂ in a cell-culture incubator for 48 h. Cells were then rinsed with PBS and stained with Hoescht 33342 dye. Live cells were examined using a Nikon A1R+ scanning confocal microscope. Image acquisition and processing settings were maintained between all samples.

Appendix 3

Synthesis of a New Collagen Mimetic Peptide

Abstract

Collagen mimetic peptides (CMPs) are short, synthetic peptides that can be designed to form either homotrimeric or heterotrimeric triple helices. The ability of CMPs to form a heterotrimeric triple helix presents the opportunity to invade natural collagen at sites of degradation to deliver therapeutic or imaging moieties. Here, we present the synthesis of a novel CMP, (flpHypGly)₇. This peptide has the potential to form an especially stable heterotrimeric triple helix due to favorable stereoelectronic effects imparted by the substitutions on prolines at the 4-position. Additionally, the hydroxyproline residue should endow the peptide with greater aqueous solubility than CMPs used in previous studies due to its hydrogen-bonding ability. These attributes could be valuable in an application in which the CMP invades natural collagen to image and characterize collagen remodeling associated with breast tumor progression.

Author Contributions: Brett S. VanVeller, Patricia J. Keely, and Ronald T. Raines proposed synthesis of the novel peptide and its application to tumor imaging. Kalie A. Mix performed synthesis. Kalie A. Mix and Ronald T. Raines wrote this chapter.

A3.1 Introduction

Collagen mimetic peptides (CMPs) are short, synthetic peptides that consist of derivatives of Xaa-Yaa-Gly repeats, where X and Y are proline residues or derivatives of proline. Modification of proline residues with a functional group such as a halo, methyl, or hydroxy group, can impart desirable attributes such as helical stability, inability to self-anneal, and aqueous solubility.³¹⁶ These peptides can also be conjugated to useful moieties such as wound-healing or imaging agents.³¹⁷ Using this method, CMPs have been developed that invade natural collagen at sites of tissue damage (Figure A3.1).

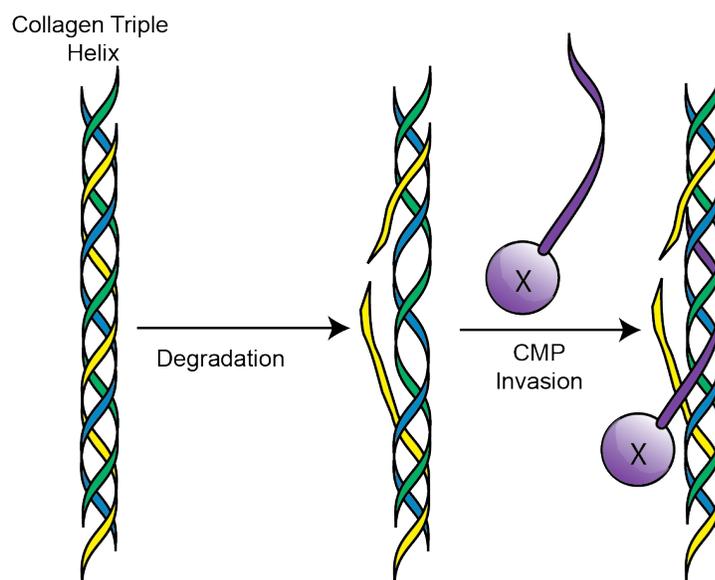


Figure A3.1 CMP invasion of collagen helix at sites of proteolytic or mechanical degradation. X = diagnostic/imaging probe or therapeutic moiety.

One exciting potential application of collagen-invading CMPs is to improve imaging of breast tumors. Tumor interaction with the extracellular matrix (ECM), of which collagen is a primary component, has been implicated as having a key role in cancer progression.³¹⁸ Tumor-associated collagen signatures (TACS) at different stages of disease progression have been characterized *ex vivo* using second harmonic generation imaging (SHG) (Figure A3.2).³¹⁹ The first stage of this sequence, TACS-1, is characterized by an increase in collagen density. At the next stage, TACS-2, collagen morphology undergoes a change and appears straight rather than coiled. Finally, at the TACS-3 stage, collagen aligns radially to the tumor boundary. TACS-3 is especially important because it serves as an independent biomarker for cancer prognosis and has been correlated with poor patient outcome.³²⁰

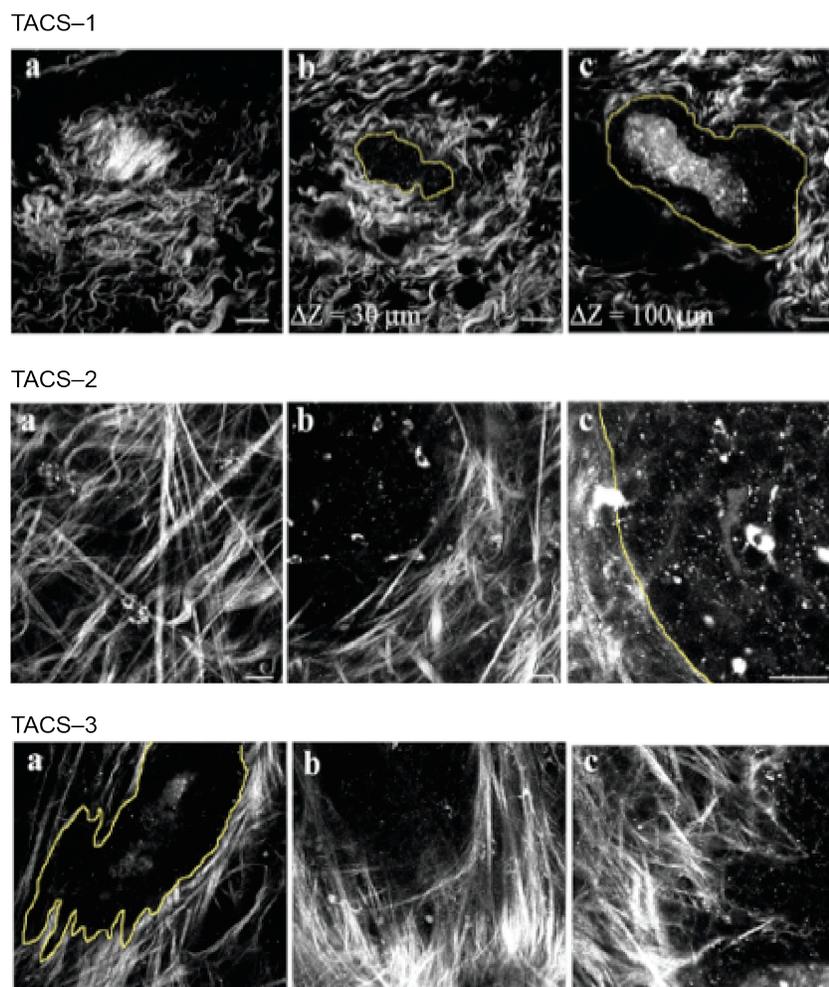


Figure A3.2 Tumor-associated collagen signatures (TACS) imaged by second harmonic generation.³¹⁹ Collagen remodels from the morphology associated with TACS-1 (top) through TACS-2 (middle) to TACS-3 (bottom) as tumor development progresses. Tumor boundary is indicated in yellow, and collagen is shown in white.

Imaging tumor-associated collagen remodeling *in vivo* would be a valuable diagnostic tool. Current clinical practices rely on mammography, an X-ray-based technique that detects changes in tissue density but has a high rate of false positives.³²¹ Using CMPs that are conjugated to MRI contrast agents or PET scan tracers presents an alternative diagnostic method for breast cancer by detecting changes in collagen remodeling (TACS 1-3) during tumor progression.

Previous work by Raines group has examined the effect of proline modifications in many different CMPs,³¹⁶ which has led to the discovery of peptides that could be well-suited for invading and imaging tumor-associated collagen. One candidate is (flpFlpGly)₇ (Figure A3.3A), which is able to form a heterotrimeric triple helix but unable to form a homotrimeric triple helix due to steric clash.³²² This CMP has, however, low solubility in an aqueous solution, which limits its utility in applications under physiological conditions. To overcome this obstacle, a new CMP, (flpHypGly)₇, was designed (Figure A3.3B). Replacement of a fluorine atom with a hydroxyl group in the Yaa position should increase water solubility through additional hydrogen-bonding capability. The inability to form a homotrimeric triple helix should be maintained because the steric clash of the proline 4-position substituents is still present. These biophysical properties make (flpHypGly)₇ an excellent candidate for strand invasion and *in vivo* imaging of collagen remodeling during tumor progression. To create a useful imaging or diagnostic peptide, the (flpHypGly)₇ could be appended with a (Gly-Ser)₃ linker to improve aqueous solubility and a lysine residue for ligation to imaging reagents via amide-bond formation (Figure A3.3C).

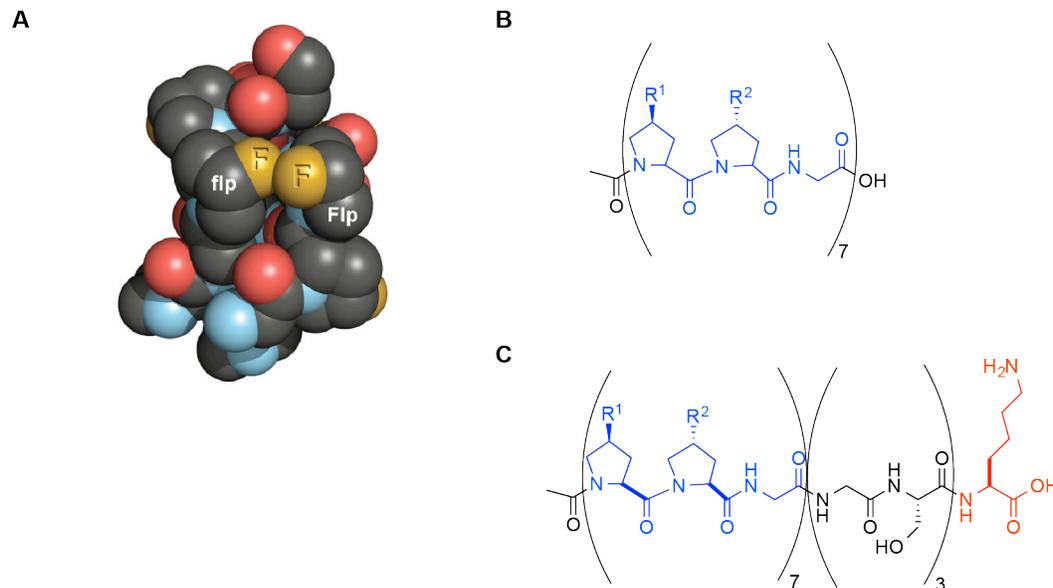
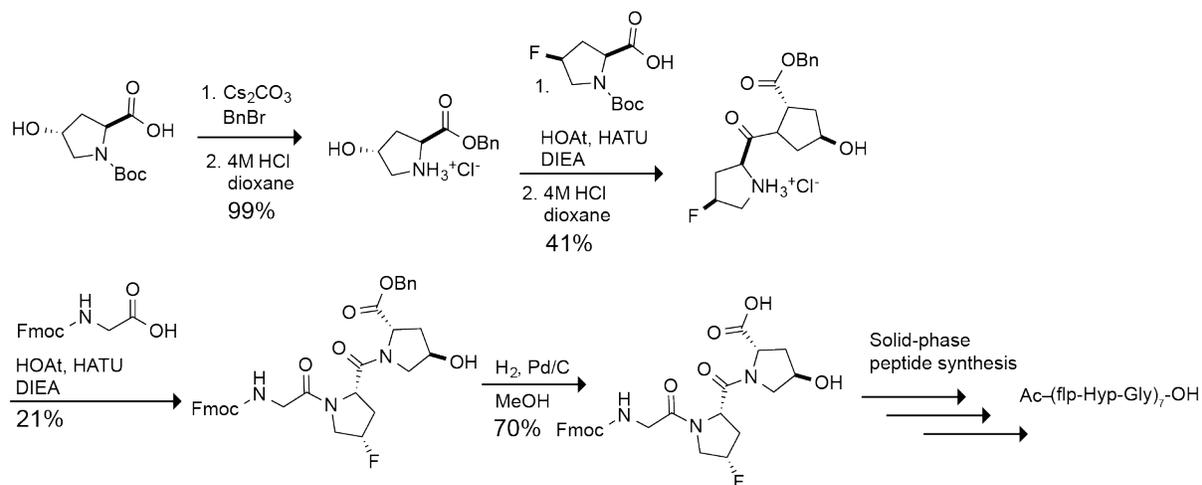


Figure A3.3 (A) Model of a (flpFlpGly)₇ homotrimer in which the X and Y positions from two different strands experience steric clash, preventing formation of a homotrimeric triple helix.³¹⁶ (B) Generic CMP structure. For Ac-(flpHypGly)₇, R1=F, R2=OH. (C) Structure of CMPs with additional solubility spacer region and lysine residue for derivatization. For Ac-(flpHypGly)₇, R1= F, R2= OH

A3.2 Results and Discussion

The acetyl-(flpHypGly)₇ peptide was synthesized by the route shown in Scheme A3.1. The (GlyflpHyp) trimer segment was synthesized in solution and extended to Ac-(flpHypGly)₇ using solid-phase peptide synthesis. The peptide was characterized by MALDI-TOF mass spectrometry (Figure A3.4).



Scheme A3.1 Synthetic route to Ac-(flpHypGly)₇

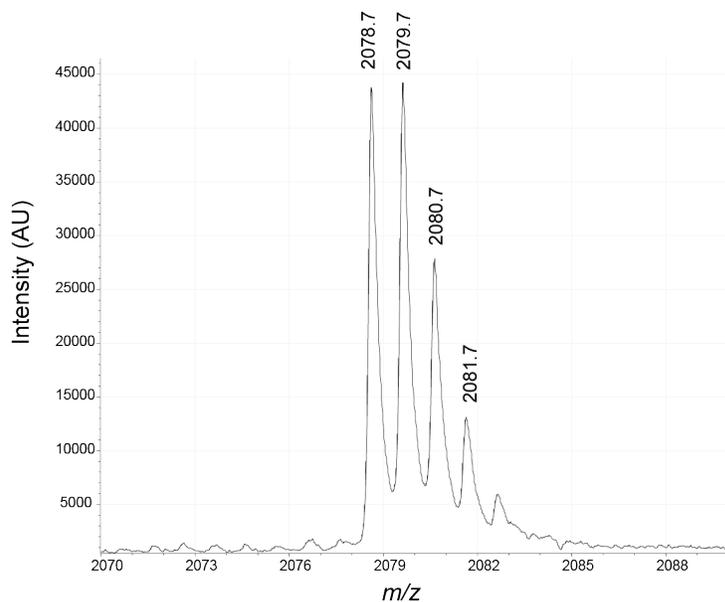


Figure A3.4 MALDI-TOF mass spectrum of Ac-(flpHypGly)₇. *m/z* calcd for C₈₆H₁₁₆F₇N₂₁O₃₀ [M+Na]⁺ 2078.8; found 2078.7

A3.3 Future Directions

Future endeavors should focus on characterizing the biophysical properties of the Ac-(flpHypGly)₇ peptide and compare them to the many CMP variants synthesized by the Raines group and others. The solubility of the peptide, the helical stability of the Ac-(flpHypGly)₇ homotrimeric triple helix, and the helical stability of an Ac-(flpHypGly)₇/(ProProGly)₇ or Ac-(flpHypGly)₇/(ProHypGly)₇ heterotrimer would be especially valuable properties to investigate. If this peptide indeed proves to be superior than other CMPs with regard to its solubility and helical properties, the Ac-(flpHypGly)₇ peptide can be derivatized further to make it suitable for tissue imaging (Figure A3.3C). The amino group of the lysine residue of this peptide can be used as a handle for ligation of a gadolinium chelator for MRI, an IR-dye for imaging in mice, or a positron-emitting radionuclide (tracer) for PET scanning.

A3.4 Materials and Methods

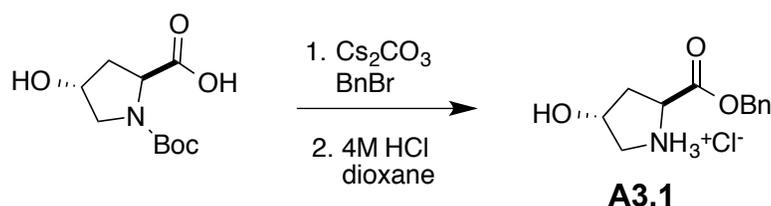
A3.4.1 General

Silica gel (40 μm 230–400 mesh) was from SiliCycle. Reagent chemicals were obtained from commercial sources and used without further purification. Dichloromethane and tetrahydrofuran were dried over a column of alumina. Thin-layer chromatography (TLC) was performed on plates of EMD 250 μm silica 60-F₂₅₄. The phrase “concentrated under reduced pressure” refers to the removal of solvents and other volatile materials using a rotary evaporator at water aspirator pressure (<20 torr) while maintaining a water bath below 40 °C. Residual solvent was removed from samples at high vacuum (<0.1 torr). Peptide synthesis was performed with a Protein Technologies Prelude automated synthesizer in the University of Wisconsin–Madison Biotechnology Center. Peptide purification was accomplished on a Shimadzu LC-20 HPLC.

LC–MS analysis of small molecules was performed using a Shimadzu LCMS-2020 equipped with a single quadrupole mass analyzer. Electrospray ionization (ESI) high-resolution mass spectrometry for small-molecule characterization was performed with a Micromass LCT at the Mass Spectrometry Facility in the Department of Chemistry at the University of Wisconsin–Madison. Matrix-assisted laser desorption-ionization–time-of-flight (MALDI–TOF) mass spectrometry for peptide characterization was performed with a Bruker MicroFlex at Massachusetts Institute of Technology.

A3.4.2 Chemical Synthesis

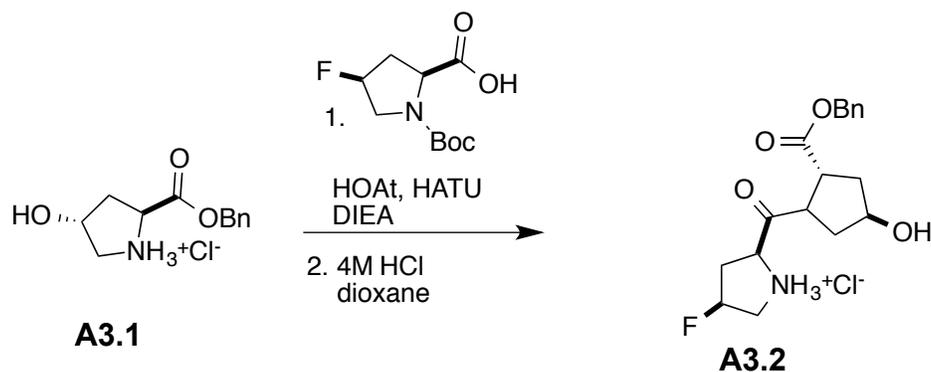
Synthesis of Hyp-OBn (A3.1)



Boc-Hyp-OH (10 g, 43 mmol) was dissolved in MeOH (20 mL). Cesium carbonate (6 g, 43 mmol) was dissolved in water. The two solutions were combined and cooled to 0 °C using an ice bath. Benzyl bromide (7.35 g, 43 mmol) and DMF (200 mL) were added, and the solution stirred for 3 h at ambient temperature. The solution was concentrated under reduced pressure. The residue was dissolved in EtOAc and purified by chromatography, eluting with 1:1 EtOAc/hexanes. The solution was concentrated under reduced pressure. The residue was dissolved in 4 M HCl in dioxanes and stirred at ambient temperature for 1 h. The solution was then sparged with N_{2(g)} and concentrated under reduced pressure to afford compound **A3.1** (9.6 g, 99%) as a white solid.

Data for A3.1: LC–MS (ESI)⁺ m/z calcd for C₁₂H₁₅NO₃ [M+H]⁺ 222; found 222.

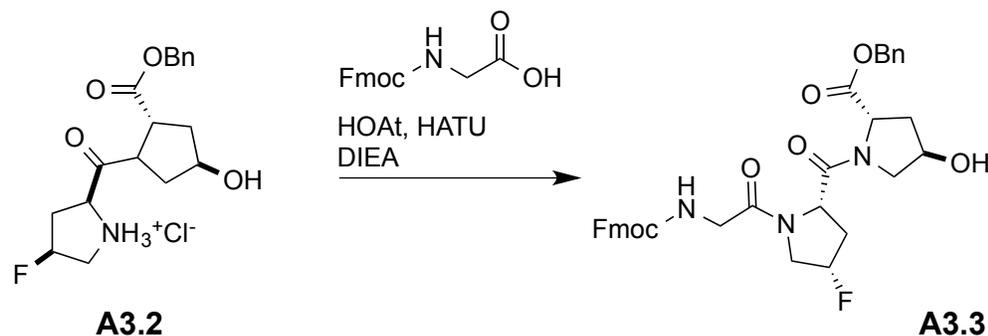
Synthesis of flp-Hyp-OBn (A3.2)



Hyp-OBn (**A3.1**) (9.6 g, 45 mmol) was dissolved in DMF (100 mL). Boc-flp-OH (12.6 g, 54 mmol), HOAt (7.3 g, 54 mmol), HATU (17.1 g, 45 mmol), and DIEA (23 g, 180 mmol) were added, and the solution stirred overnight at ambient temperature. The solution was concentrated under reduced pressure. The residue was dissolved in EtOAc and washed twice with aqueous HCl (1 M) and twice with saturated aqueous Na₂HCO₃. The solution was dried over Na₂SO₄(s) and then purified by chromatography, eluting with EtOAc. The residue was dissolved in 4 M HCl in dioxanes and stirred at ambient temperature for 1 h. The solution was then sparged with N₂(g) and concentrated under reduced pressure to afford compound **A3.2** (6.1 g, 41%) as a white solid.

Data for A3.2: LC–MS (ESI)⁺ m/z calcd for C₁₇H₂₁FN₂O₄ [M+H]⁺ 337; found 337.

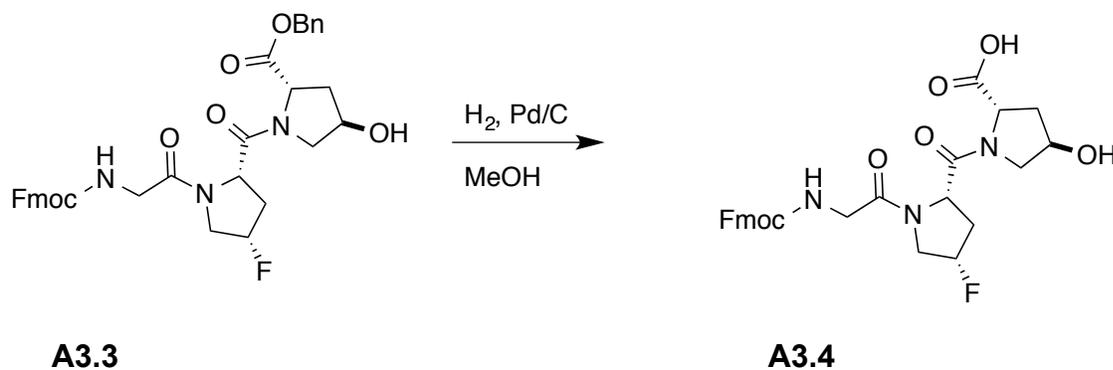
Synthesis of Fmoc-Gly-flp-Hyp-OBn (A3.3)



The compound flp-Hyp-OBn (**A3.2**) (4.9 g, 14 mmol) was dissolved in DMF (140 mL). Fmoc-Gly (4.0 g, 13.3 mmol), HOAt (1.9 g, 14 mmol), HATU (5.3 g, 14 mmol), and DIEA (7.2 g, 56 mmol) were added, and the solution was stirred overnight at ambient temperature. The residue was dissolved in EtOAc and washed twice with aqueous HCl (1 M) and twice with saturated aqueous Na₂HCO₃. The solution was dried over Na₂SO₄(s) and then purified by chromatography, eluting with EtOAc to afford **A3.3**, (1.9 g, 21%) as a white solid.

Data for A3.3: HRMS (ESI⁺) *m/z* calcd for C₃₄H₃₄FN₃O₇ [M+NH₄]⁺ 633.2720; found 633.2712.

Synthesis of Fmoc-gly-flp-Hyp-OH (A3.4)



Fmoc-GlyflpHyp-OBn (**A3.3**) (1.5 g, 2.4 mmol) was dissolved in MeOH. The solution was sparged with N₂(g), and palladium on carbon (250 mg Pd/C, 0.24 mmol Pd) was added. The solution was stirred under an atmosphere of H₂(g) for 5 h. The palladium/carbon was removed by filtration, and the solution was concentrated under reduced pressure. The residue was dissolved in DCM and purified by chromatography, eluting with 4:1 DCM/MeOH to afford compound **A3.4** (0.9 g, 70%) as a white solid.

Data for A3.4: HRMS (ESI⁺) *m/z* calcd for C₂₇H₂₈FN₃O₇ [M+H]⁺ 526.1985; found 526.1976.

3.4.3 Peptide Synthesis

Solid-phase peptide synthesis of Ac-(flpHypGly)₇

The peptide was synthesized on a 25- μ M scale by segment condensation of the corresponding amino acid trimers and monomers on a solid phase using a Prelude peptide synthesizer at the University of Wisconsin–Madison Biotechnology Center. The resin used was NovaSyn Fmoc-Gly TGT resin (0.2 mmol/g loading). Fmoc-deprotection was achieved by treatment with a solution of piperidine (20% v/v) in DMF. The added amino acid (4 equiv) was converted to the active ester by using HCTU and NMM. Each residue was double-coupled between Fmoc-deprotections. All couplings were performed at room temperature. Peptide was cleaved from the resin in 96.5:2.5:1.0 TFA/H₂O/TIPSH (5 mL), precipitated from ethyl ether at 0 °C, and isolated by centrifugation. The isolated peptide was dissolved in 40% (v/v) MeCN/H₂O. The peptide was purified by reversed-phase HPLC using a C₁₈ column, eluting with a linear gradient of water (containing 0.1% v/v TFA) and acetonitrile. The fraction containing the peptide was lyophilized to yield a white powder.

Appendix 4

Synthesis of Demethoxy Q Derivatives for Biochemical Investigation of COQ9 Structure and Function

Abstract

Coenzyme Q (CoQ) plays a vital role in cellular respiration and energy production. A deficiency in this molecule is associated with numerous disease phenotypes. Yet, many aspects of its biosynthesis remain unknown. One biosynthetic enzyme whose role is essential but uncharacterized is COQ9, which is hypothesized to interact with a second biosynthetic enzyme, COQ7. Derivatives of the COQ7 substrate, demethoxy Q (DMQ) were synthesized and used for in vitro biochemical assays to shed light on the structure and function of COQ9 and its interaction with COQ7.

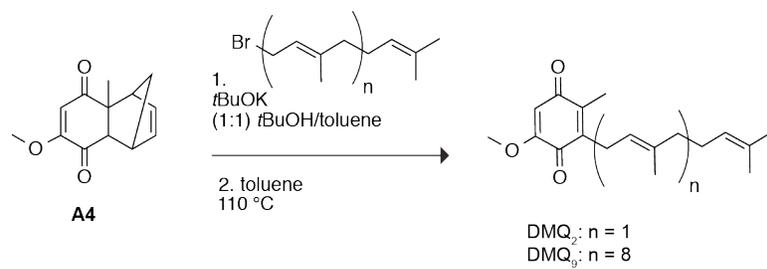
Author Contributions: Kalie A. Mix synthesized reagents. Danielle C. Lohman Matthew S. Stefely, and David J. Pagliarini designed and performed biochemical experiments. Kalie A. Mix, Danielle C. Lohman, and Ronald T. Raines wrote this chapter.

A recently solved crystal structure of *Homo sapiens* COQ9 supports this hypothesis. In this structure, COQ9 co-crystallizes spontaneously with a phospholipid in a hydrophobic binding pocket, though the structure was not well-resolved enough to determine the identity of the lipid.³²⁶ Additionally, COQ7 and COQ9 form a complex in vitro, and mutation of COQ9 at the COQ7–COQ9 binding interface cause accumulation of DMQ₆ in yeast. Collectively, these data suggest that COQ9 might bind a CoQ biosynthetic precursor and present it to COQ7.³²⁶

Further biochemical experiments are needed to support the hypothesis that COQ9 presents DMQ to COQ7. Demonstration of DMQ-binding by COQ9 in vitro, as well as a crystal structure of COQ9 bound to DMQ could support the notion that COQ9 binds DMQ in vivo. Both of these experiments require the use of DMQ derivatives, which are not readily commercially available. Here, we describe the synthesis of DMQ₂ and DMQ₉, as well as preliminary biochemical studies that characterize COQ9–DMQ₉ binding.

A4.2 Results and Discussion

DMQ₂ and DMQ₉ were synthesized as shown in Scheme A4.1. The synthesis of DMQ₂ was reported previously.³²⁷ DMQ₉ was synthesized by a similar route using a solanesyl bromide in place of geranyl bromide. The identity of the product was confirmed by NMR spectroscopy and MALDI–TOF mass spectrometry (Figure A4.2).



Scheme A4.1 Synthetic route to DMQ_2 and DMQ_9

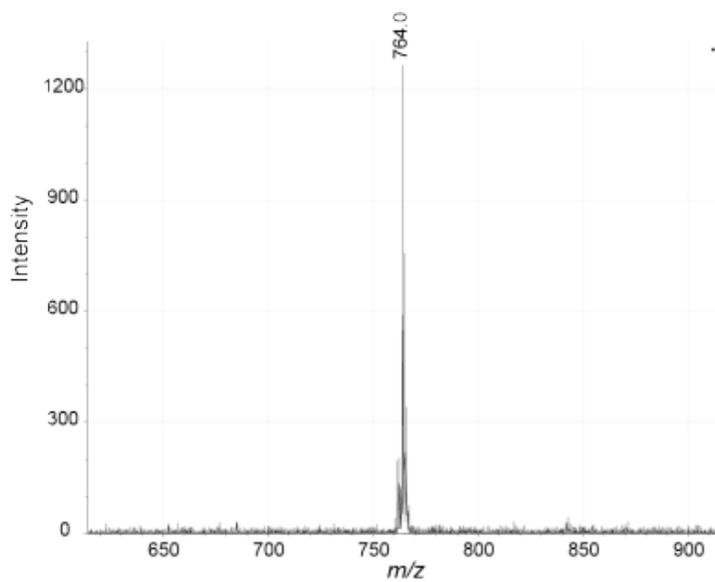


Figure A4.2 MALDI-TOF mass spectrum of DMQ_9 . Expected m/z : 764.5. Measured m/z : 764.0.

To determine whether COQ9 protein associates with DMQ₉ *in vitro*, the protein was incubated with liposomes containing phosphatidylcholine, phosphatidylethanolamine, phosphatidylglycerol (or combinations thereof), DMQ₉, or CoQ (Figure A4.3). COQ9 binds preferentially to CoQ relative to other lipids (Figure A4.3A) and binds to DMQ₉ to an even greater extent than to CoQ (Figure A4.3B,C). These results suggest that COQ9 could bind to CoQ or its biosynthetic intermediates (such as DMQ₉) *in vivo*, which is required if COQ9 does indeed play a role in presentation of substrates to other biosynthetic enzymes (such as COQ7).

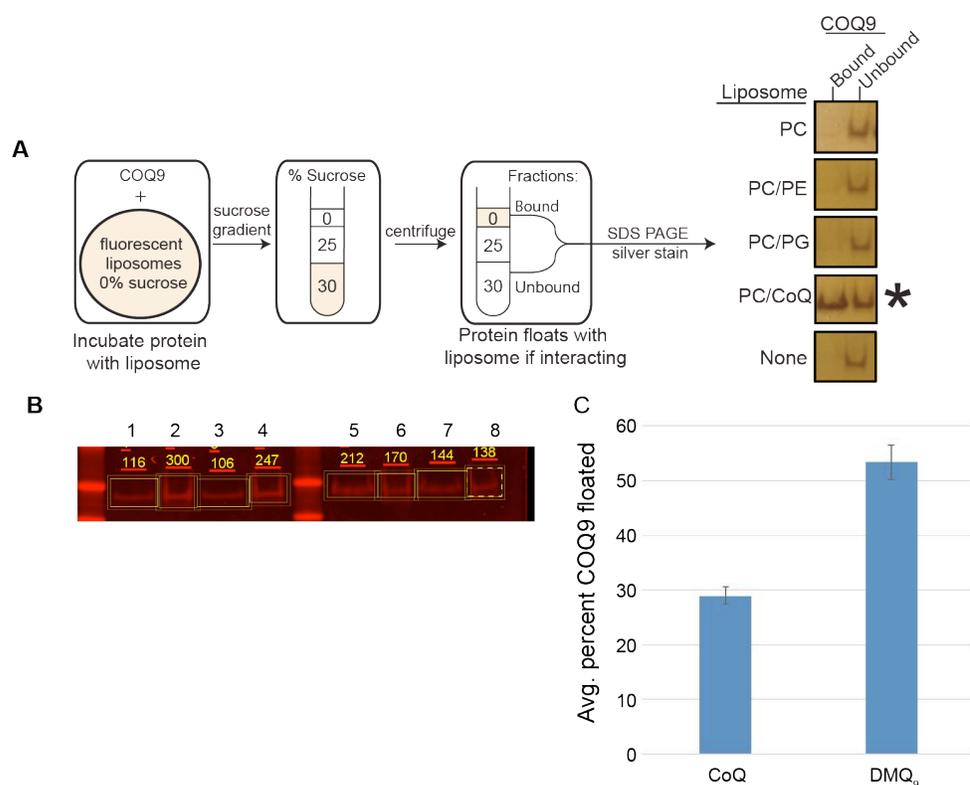


Figure A4.3 Association of COQ9 enzyme with liposomes. (A) Schematic of experimental workflow and silver stained SDS-PAGE of COQ9 after incubation with liposomes and subsequent separation by centrifugation. (B) SDS-PAGE of COQ9 after incubation with liposomes containing CoQ (Lanes 1–4) or DMQ₉ (Lanes 5–8) and subsequent separation by centrifugation. Odd numbered lanes contain top (bound) fraction; even numbered lanes contain lower (unbound) fraction. (C) Quantification of COQ9 associated with CoQ and DMQ₉. The percent of protein floated was calculated as a ratio of intensity of the top fraction to total intensity of both fractions from the gel in (B).

A4.3 Future Directions

Further experimental data is needed to elucidate the role of COQ9 *in vivo*. Experiments are in progress that seek to crystallize COQ9 with DMQ₂ and DMQ₉ to determine if these species reside in the lipid-binding pocket. Additionally, the DMQ molecules will be used for enzymatic activity assays that compare efficiency of COQ7 *in vitro* in the presence and absence of functional COQ9.

A4.4 Acknowledgments

K.A.M. was supported by Molecular Biosciences Training Grant T32 GM007215 (NIH) and a fellowship from the University of Wisconsin–Madison College of Agricultural and Life Sciences. D.C.L. was supported by an NSF Graduate Research Fellowship.

A4.5 Materials and Methods

A4.5.1 General

Silica gel (40 μm , 230–400 mesh) was from SiliCycle. Compound **A4** was from WuXi AppTec (Shanghai, China). Reagent chemicals were from commercial sources and used without further purification. Dichloromethane was dried over a column of alumina. Thin-layer chromatography (TLC) was performed on plates of EMD 250 μm silica 60-F₂₅₄. The phrase “concentrated under reduced pressure” refers to the removal of solvents and other volatile materials using a rotary evaporator at water aspirator pressure (<20 torr) while maintaining a water bath below 40 °C. Residual solvent was removed from samples at high vacuum (<0.1 torr). ¹H and ¹³C NMR spectra for all compounds were acquired on Bruker spectrometers in the National Magnetic Resonance Facility at Madison operating at 500 MHz. Chemical shift values (δ) are reported in

units of ppm relative to an internal standard (residual solvent or TMS). Matrix-assisted laser desorption-ionization–time-of-flight (MALDI–TOF) mass spectrometry for small-molecule characterization was performed with a Bruker Microflex LRF in the Mass Spectrometry Facility in the Department of Chemistry at the University of Wisconsin–Madison.

A4.5.2 Chemical Synthesis

Synthesis of DMQ₂:

DMQ₂ was synthesized as described previously.³²⁷ Spectral data and yields match those reported previously.

Synthesis of DMQ₉:

Compound **A4** (31 mg, 0.14 mmol) was dissolved in 1:1 *t*BuOH/toluene. Solanesyl bromide (100 mg, 0.14 mmol) and *tert*-butoxide (47 mg, 0.42 mmol) were added, and the resulting solution was stirred on ice for 1 h. The solution was diluted with saturated aqueous NH₄Cl and extracted with diethyl ether. The organic layers were combined and concentrated under reduced pressure. The residue was dissolved in diethyl ether and purified by flash chromatography, eluting with 1:1 ether/ligroin to yield the prenylated intermediate as a clear oil. The cyclopentadienyl protecting group was removed by dissolving the residue in toluene (1 mL) and heating the resulting solution at reflux (110 °C) for 1 h. The solution was concentrated under reduced pressure. The residue was dissolved in DCM and purified by flash chromatography, eluting with DCM to yield DMQ₉ (10.7 mg, 61% yield) as a yellow oil.

Data for DMQ₉: ¹H NMR (500 MHz, CDCl₃, δ): 5.88 (s, 1H), 5.11 (t, 7H, J = 6.8 Hz), 5.05 (t, 1H, J = 7.0 Hz), 4.94 (t, 1H, J = 6.7 Hz), 3.79 (s, 3H), 3.22 (d, 2H, J = 7.0 Hz), 2.05 (m, 19H),

1.98 (m, 16H), 1.60 (s, 30H). ^{13}C NMR (125 MHz, CDCl_3 , δ): 190.5, 184.6, 161.0, 144.5, 144.1, 140.4, 138.0, 137.73, 137.69, 137.65, 137.63, 137.60, 133.98, 127.1, 127.0, 126.9, 126.5, 121.4, 109.7, 58.8, 42.5, 42.4, 32.4, 29.5, 29.40, 29.37, 29.2, 28.4, 28.0, 20.4, 19.1, 18.7, 14.9. MALDI-TOF m/z calcd for $\text{C}_{53}\text{H}_{80}\text{O}_3$, 764.6; found, 764.0.

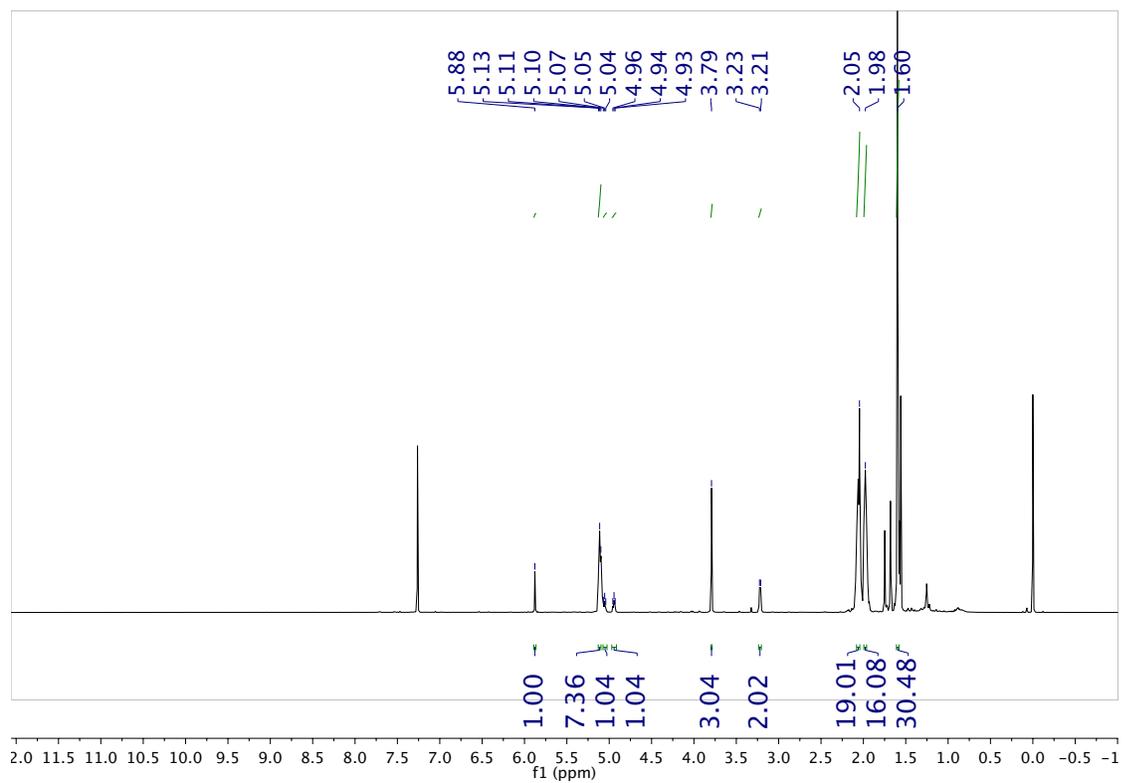
A4.5.3 Liposome floatation assay:

COQ9 protein was expressed and purified as described previously.³²⁶ A liposome floatation assay was adapted from that of Langer and coworkers.³²⁸ Liposomes (100 μL) were incubated with protein (50 μL) (10 min, 20 $^\circ\text{C}$). The final concentration of reagents in HBS were protein: 2.5 μM and liposomes: 6.66 mM. An aqueous solution of sucrose in 2.72 M HEPES buffered saline (HBS; 110 μL) was added, and the resulting solution was transferred (250 μL) to an ultracentrifuge tube (Beckman #343776). A sucrose gradient in HBS was layered as follows: 1.15 M sucrose with reaction (250 μL), 0.86 M sucrose (300 μL), 0.29 M sucrose (250 μL), and 150 μL HBS. After centrifugation (240,000g, 1 h, 4 $^\circ\text{C}$) (Sorvall MX 120 Plus Micro-ultracentrifuge), the top (450 μL) and bottom (450 μL) fractions were removed from top to bottom. Liposomes were quantified by NBD-PE fluorescence (excitation: 460 nm, emission: 535 nm). Proteins in each fraction were precipitated with chloroform/methanol as adapted from Wessel and Flugge.³²⁹ Methanol (1800 μL methanol, 4 volumes) was added, and the samples were vortexed. Chloroform (450 μL , 1 volume) was added, and the solution was mixed. Water (1350 μL , 3 volumes) was added, and the samples were vortexed and subjected to centrifugation (5 min, 4,000g, 20 $^\circ\text{C}$). The majority of the upper aqueous layer was discarded and methanol (1000 μL) was added to the protein disc. After mixing by inversion, samples were subjected to centrifugation (5 min, 20,000g, 20 $^\circ\text{C}$) and all liquid was removed. The precipitated protein

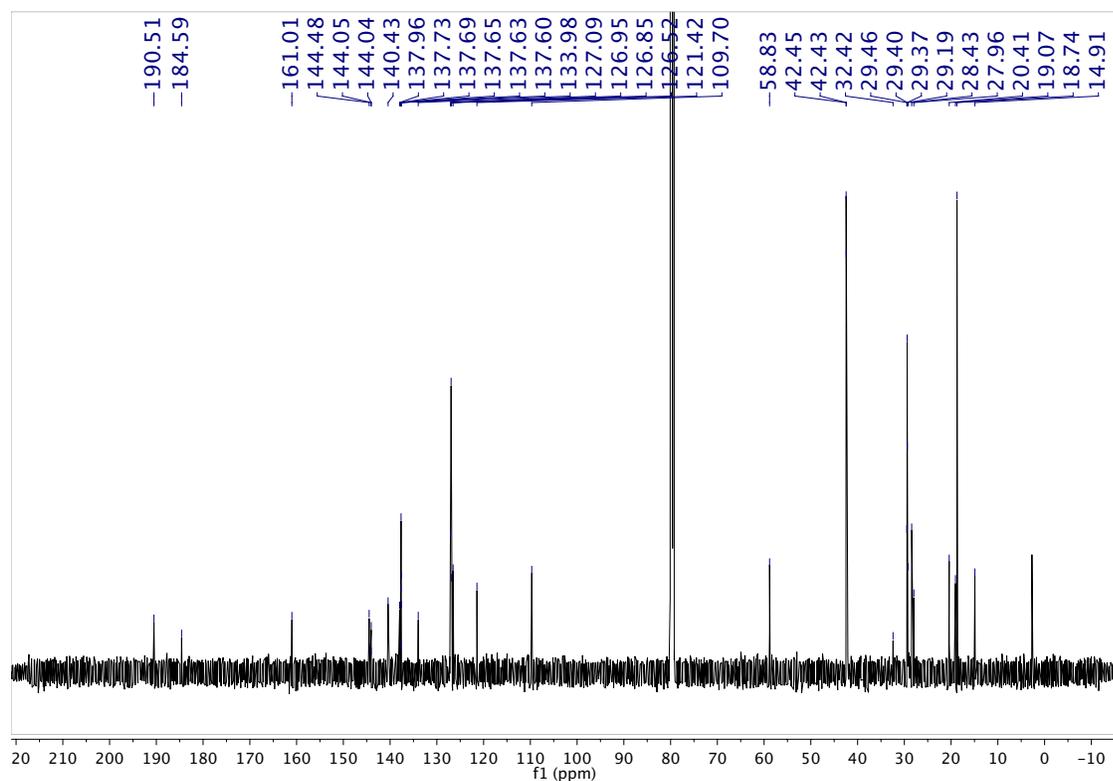
pellet was dried under vacuum (25–30 torr, 30 min, 20 °C), resuspended in lithium dodecyl sulfate containing DTT (10 mM), and analyzed with SDS–PAGE. Protein bands were quantified by densitometry with a LiCOR Odyssey CLx (700 nm) instrument using Image Studio v5.2 software.

A4.5.4 NMR Spectra

^1H NMR of DMQ₉ (500 MHz, CDCl₃):



^{13}C NMR of DMQ₉ (125 MHz, CDCl₃):



References

- (1) Thul, P. J.; Åkesson, L.; Wiking, M.; Mahdessian, D.; Geladaki, A.; Ait Blal, H.; Alm, T.; Asplund, A.; Björk, L.; Breckels, L. M.; Bäckström, A.; Danielsson, F.; Fagerberg, L.; Fall, J.; Gatto, L.; Gnann, C.; Hober, S.; Hjelmare, M.; Johansson, F.; Lee, S.; Lindskog, C.; Mulder, J.; Mulvey, C. M.; Nilsson, P.; Oksvold, P.; Rockberg, J.; Schutten, R.; Schwenk, J. M.; Sivertsson, Å.; Sjöstedt, E.; Skogs, M.; Stadler, C.; Sullivan, D. P.; Tegel, H.; Winsnes, C.; Zhang, C.; Zwahlen, M.; Mardinoglu, A.; Pontén, F.; von Feilitzen, K.; Lilley, K. S.; Uhlén, M.; Lundberg, E. (2017) A subcellular map of the human proteome. *Science* 356, eaal33271.

- (2) *Diazo Compounds: Properties and Synthesis*. Regitz, M.; Maas, G.; Academic Press: London, 1986.

- (3) Doyle, M. P.; McKervey, M. A. (1997) Recent advances in stereoselective synthesis involving diazocarbonyl intermediates. *Chem. Commun.*, 983–989.

- (4) von Pechmann, H. (1894) Ueber Diazomethan. *Ber. Dtsch. Chem. Ges.* 27, 1888–1891.

- (5) von Pechmann, H. (1895) Ueber Diazomethan. *Ber. Dtsch. Chem. Ges.* 28, 855–861.

- (6) Lewinn, E. B. (1949) Diazomethane poisoning: Report of a fatal case with autopsy. *Am. J. Med. Sci.* 218, 556–562.

- (7) Schoental, R. (1960) Carcinogenic action of diazomethane and of nitroso-*N*-methyl urethane. *Nature*, 420–421.

- (8) Lewis, C. E. (1964) Diazomethane poisoning: Report of a case suggesting sensitization reaction. *J. Occup. Environ. Med.* 6, 91–92.

- (9) de Boer, T. J.; Backer, H. J. (1956) Diazomethane. *Org. Syn.* 36, 16–18.

- (10) Diazomethane. Sammakia, T. In *e-EROS Encyclopedia of Reagents for Organic Synthesis*; John Wiley & Sons: New York, NY, 2001.

- (11) We are aware of only one natural azide metabolite, β -azidoalanine, which forms in microbes that are fed inorganic azide (Ciesla, Z., Filutowicz, M., and Klotkowski, T. (1980) Involvement of the L-cysteine biosynthetic pathway in azide-induced mutagenesis in *Salmonella typhimurium*. *Mutat. Res.* 70, 261–268).

- (12) Köpke, T.; Zaleski, J. M. (2008) Diazo-containing molecular constructs as potential anticancer agents: From diazo[b]fluorene natural products to photoactivatable diazo-oxochlorins. *Anticancer Agents Med. Chem.* 8, 292–304.
- (13) Gould, S. J. (1997) Biosynthesis of the kinamycins. *Chem. Rev.* 97, 2499–2509.
- (14) Kersten, R. D.; Lane, A. L.; Nett, M.; Richter, T. K. S.; Duggan, B. M.; Dorrestein, P. C.; Moore, B. S. (2013) Bioactivity-guided genome mining reveals the lomaiviticin biosynthetic gene cluster in *Salinispora tropica*. *ChemBioChem* 14, 955–962.
- (15) Janso, J. E.; Haltli, B. A.; Eustaquio, A. S.; Kulowski, K.; Waldman, A. J.; Zha, L.; Nakamura, H.; Bernan, V. S.; He, H.; Carter, G. T.; Koehn, F. E.; Balskus, E. P. (2014) Discovery of the lomaiviticin biosynthetic gene cluster in *Salinispora pacifica*. *Tetrahedron* 70, 4156–4164.
- (16) Waldman, A. J.; Pechersky, Y.; Wang, P.; Wang, J. X.; Balskus, E. P. (2015) The cremeomycin biosynthetic gene cluster encodes a pathway for diazo formation. *ChemBioChem* 16, 2172–2175.
- (17) Sugai, Y.; Katsuyama, Y.; Ohnishi, Y. (2016) A nitrous acid biosynthetic pathway for diazo group formation in bacteria. *Nat. Chem. Biol.* 12, 73–75.
- (18) Narawat, C. C.; Moody, C. J. (2011) Natural products containing a diazo group. *Nat. Prod. Rep.* 28, 1426–1444.
- (19) Herzon, S. B.; Woo, C. M. (2012) The diazofluorene antitumor antibiotics: Structural elucidation, biosynthetic, synthetic, and chemical biological studies. *Nat. Prod. Rep.* 29, 87–118.
- (20) Ballard, T. E.; Melander, C. (2008) Kinamycin-mediated DNA cleavage under biomimetic conditions. *Tetrahedron Lett.* 49, 3157–3161.
- (21) Woo, C. M.; Li, Z.; Paulson, E. K.; Herzon, S. B. (2015) Structural basis for DNA cleavage by the potent antiproliferative agent (–)-lomaiviticin A. *Proc. Nat. Acad. Sci. USA* 113, 2851–2856.
- (22) Ito, S.; Matsuya, T.; Omura, S.; Otani, M.; Nakagawa, A.; Takeshima, H.; Iwai, Y.; Ohtani, M.; Hata, T. (1970) A new antibiotic, kinamycin. *J. Antibiot.* 23, 315–317.

- (23) Gould, S. J.; Tamayo, N.; Melville, C. R.; Cone, M. C. (1994) Revised structures for the kinamycin antibiotics: 5-Diazobenzo[*b*]fluorenes rather than benzo[*b*]carbazole cyanamides. *J. Am. Chem. Soc.* *116*, 2207–2208.
- (24) Lei, X.; Porco, J. A. (2006) Total synthesis of the diazobenzofluorene antibiotic (–)-kinamycin C. *J. Am. Chem. Soc.* *128*, 14790–14791.
- (25) Kumamoto, T.; Kitani, Y.; Tsuchiya, H.; Yamaguchi, K.; Seki, H.; Ishikawa, T. (2007) Total synthesis of (±)-methyl-kinamycin C. *Tetrahedron* *63*, 5189–5199.
- (26) Nicolau, K. C.; Li, H.; Nold, A. L.; Pappo, D.; Lenzen, A. (2007) Total synthesis of kinamycins C, F, and J. *J. Am. Chem. Soc.* *129*, 10356–10357.
- (27) He, H.; Ding, W.-D.; Bernan, V. S.; Richardson, A. D.; Ireland, C. M.; Greenstein, M.; Ellestad, G. A.; Carter, G. T. (2001) Lomaiviticins A and B, potent antitumor antibiotics from *Micromonospora lomaivitiensis*. *J. Am. Chem. Soc.* *123*, 5362–5363.
- (28) Mulcahy, S. P.; Woo, C. M.; Ding, W.; Ellestad, G. A.; Herzon, S. B. (2012) Characterization of a reductively-activated elimination pathway relevant to the biological chemistry of the kinamycins and lomaiviticins. *Chem. Sci.* *3*, 1070–1074.
- (29) Nicolau, K. C.; Denton, R. M.; Lenzen, A.; Edmonds, D. J.; Li, A.; Milburn, R. R.; Harrison, R. R. (2006) Stereocontrolled synthesis of model core systems of lomaiviticins A and B. *Angew. Chem., Int. Ed.* *45*, 2076–2081.
- (30) Zhang, W.; Baranczak, A.; Sulikowski, G. A. (2008) Stereocontrolled assembly of the C3/C3' dideoxy core of lomaiviticin A/B and congeners. *Org. Lett.* *10*, 1939–1941.
- (31) Nicolau, K. C.; Nold, A. L.; Li, H. (2009) Synthesis of the monomeric unit of the lomaiviticin aglycon. *Angew. Chem., Int. Ed.* *121*, 5974–5977.
- (32) Lee, H. G.; Ahn, J. Y.; Lee, A. S.; Shair, M. D. (2010) Enantioselective synthesis of the lomaiviticin aglycon full carbon skeleton reveals remarkable remote substituent effects during the dimerization event. *Chem. Eur. J.* *16*, 13058–13062.
- (33) Herzon, S. B.; Lu, L.; Woo, C. M.; Gholap, S. L. (2011) 11-Step enantioselective synthesis of (–)-lomaiviticin aglycon. *J. Am. Chem. Soc.* *133*, 7260–7263.

- (34) Arya, D. P.; Jebaratnam, D. J. (1995) DNA cleaving ability of 9-diazofluorenes and diaryl diazomethanes: Implications for the mode of action of the kinamycin antibiotics. *J. Org. Chem.* *60*, 3268–3269.
- (35) Cone, M. C.; Melville, C. R.; Gore, M. P.; Gould, S. J. (1993) Kinafluorenone, a benzo[*b*]fluorene isolated from the kinamycin producer *Streptomyces murayamensis*. *J. Org. Chem.* *58*, 1058–1061.
- (36) Laufer, R. S.; Dmitrienko, G. I. (2002) Diazo group electrophilicity in kinamycins and lomaiviticin A: Potential insights into the molecular mechanism of antibacterial and antitumor activity. *J. Am. Chem. Soc.* *124*, 1854–1855.
- (37) Feldman, K. S.; Eastman, K. J. (2006) Studies on the mechanism of action of prekinamycin, a member of the diazoparaquinone family of natural products: Evidence for both sp^2 radical and orthoquinone methide intermediates. *J. Am. Chem. Soc.* *128*, 12562–12573.
- (38) O'Hara, K. A.; Wu, X.; Patel, D.; Liang, H.; Yalowich, J. C.; Chen, N.; Goodfellow, V.; Adedayo, O.; Dmitrienko, G. I.; Hasinoff, B. B. (2007) Mechanism of the cytotoxicity of the diazoparaquinone antitumor antibiotic kinamycin F. *Free Radic. Biol. Med.* *43*, 1132–1144.
- (39) Colis, L. C.; Woo, C. M.; Hegan, D. C.; Li, Z.; Glazer, P. M.; Herzon, S. B. (2014) The cytotoxicity of (–)-lomaiviticin A arises from induction of double-strand breaks in DNA. *Nat. Chem.* *6*, 504–510.
- (40) Woo, C. M.; Ranjan, N.; Arya, D. P.; Herzon, S. B. (2014) Analysis of diazofluorene DNA binding and damaging activity: DNA cleavage by a synthetic monomeric diazofluorene. *Angew. Chem., Int. Ed.* *53*, 9325–9328.
- (41) Colis, L. C.; Hegan, D. C.; Kaneko, M.; Glazer, P. M.; Herzon, S. B. (2015) Mechanism of action studies of lomaiviticin A and the monomeric lomaiviticin aglycon. Selective and potent activity toward DNA double-strand break repair-deficient cell lines. *J. Am. Chem. Soc.* *137*, 5741–5747.
- (42) Jeggo, P. A.; Pearl, L. H.; Carr, A. M. (2016) DNA repair, genome stability and cancer: A historical perspective. *Nat. Rev. Cancer* *16*, 35–42.
- (43) Dion, H. W., Fusari, S.A., Jakubowski, Z.L., Zora, J.G., Bartz, Q.R. (1956) 6-Diazo-5-oxo-L-norleucine, a new tumor-inhibitory substance. II. Isolation and characterization. *J. Am. Chem. Soc.* *78*, 3075–3077.

- (44) Handschumacher, R. E.; Bates, C. J.; Chang, P. K.; Andrews, A. T.; Fischer, G. A. (1968) 5-Diazo-4-oxo-L-norvaline: Reactive asparagine analog with biological specificity. *Science* *161*, 62–63.
- (45) Pinkus, L. M. (1977) Glutamine binding sites. *Methods Enzymol.* *46*, 414–427.
- (46) Fusari, S. A.; Haskell, T. H.; Frohardt, R. P.; Bartz, Q. R. (1954) Azaserine, a new tumor-inhibitory substance. Structural studies. *J. Am. Chem. Soc.* *76*, 2881–2883.
- (47) Hartman, S. C. (1963) The interaction of 6-diazo-5-oxo-L-norleucine with phosphoribosyl pyrophosphate amidotransferase. *J. Biol. Chem.* *238*, 3036–3047.
- (48) Hartman, S. C.; McGrath, T. F. (1973) Glutaminase A of *Escherichia coli*. *J. Biol. Chem.* *248*, 8506–8510.
- (49) Clark, V. M.; Shapiro, R. A.; Curthoys, N. P. (1982) Comparison of the hydrolysis and the covalent binding of 6-diazo-5-oxo-L-[6-¹⁴C]norleucine by rat renal phosphate-dependent glutaminase. *Arch. Biochem. Biophys* *213*, 232–239.
- (50) Rahman, A.; Smith, F. P.; Luc, P.-V.; Woolley, P. V. (1985) Phase I study and clinical pharmacology of 6-diazo-5-oxo-L-norleucine (DON). *Invest. New Drugs* *3*, 369–374.
- (51) Inoue, M.; Horiuchi, S.; Morino, Y. (1977) Affinity labeling of rat-kidney γ -glutamyl transpeptidase. *Eur. J. Biochem.* *73*, 335–342.
- (52) Tate, S. S.; Meister, A. (1977) Affinity labeling of γ -glutamyl transpeptidase and location of the γ -glutamyl binding site on the light subunit. *Proc. Natl. Acad. Sci. USA* *74*, 931–935.
- (53) Horiuchi, S.; Inoue, M.; Morino, Y. (1978) γ -Glutamyl transpeptidase: Sidedness of its active site on renal brush-border membrane. *Eur. J. Biochem.* *87*, 429–437.
- (54) Holcenberg, J. S.; Ericsson, L.; Roberts, J. (1978) Amino acid sequence of the diazooxonorleucine binding site of *Acinetobacter* and *Pseudomonas 7A* glutaminase–asparaginase enzymes. *Biochemistry* *17*, 411–417.
- (55) Ortlund, E.; Lacount, M. W.; Lewinski, K.; Lebioda, L. (2000) Reactions of *Pseudomonas 7A* glutaminase–asparaginase with diazo analogues of glutamine and asparagine

result in unexpected covalent inhibitions and suggests an unusual catalytic triad Thr-Tyr-Glu. *Biochemistry* 39, 1199–1204.

(56) Peterson, R. G.; Richards, F. F.; Handschumacher, R. E. (1977) Structure of peptide from active site region of *Escherichia coli* L-asparaginase. *J. Biol. Chem.* 252, 2072–2076.

(57) Asselin, B. L.; Lorenson, M. Y.; Whitin, J. C.; Coppola, D. J.; Kende, A. S.; Blakley, R. L.; Cohen, H. J. (1991) Measurement of serum L-asparagine in the presence of L-asparaginase requires the presence of an L-asparaginase inhibitor. *Cancer Res.* 51, 6568–6573.

(58) Regitz, M. (1967) New methods of preparative organic chemistry. *Angew. Chem., Int. Ed.* 6, 733–749.

(59) Baum, J. S.; Shook, D. A.; Davies, H. M. L.; Smith, H. D. (1987) Diazotransfer reactions with *p*-acetamidobenzenesulfonyl azide. *Synth. Commun.* 17, 1709–1716.

(60) Curtius, T. (1883) Ueber die Einwirkung von salpetriger Säure auf salzsauren Glycocolläther. *Ber. Dtsch. Chem. Ges.* 16, 2230–2231.

(61) Womack, E. B.; Nelson, A. B. (1955) Ethyl diazoacetate. *Org. Syn.* 24, 56–57.

(62) Bamford, W. R.; Stevens, T. S. (1952) 924. The decomposition of toluene-*p*-sulphonylhydrazones by alkali. *J. Chem. Soc.*, 4735–4740.

(63) Fulton, J. R.; Aggarwal, V. K.; de Vicente, J. (2005) The use of tosylhydrazone salts as a safe alternative for handling diazo compounds and their applications in organic synthesis. *Eur. J. Org. Chem.* 2005, 1479–1492.

(64) Holton, T. L.; Schechter, H. (1995) Advantageous syntheses of diazo compounds by oxidation of hydrazones with lead tetraacetate in basic environments. *J. Org. Chem.* 60, 4725–4729.

(65) Furrow, M. E.; Myers, A. G. (2004) A general procedure for the esterification of carboxylic acids with diazoalkanes generated *in situ* by the oxidation of *N*-tert-butylidimethylsilylhydrazones with (difluoroiodo)benzene. *J. Am. Chem. Soc.* 126, 12222–12223.

(66) Morandi, B.; Carreira, E. M. (2012) Iron-catalyzed cyclopropanation in 6 M KOH with *in situ* generation of diazomethane. *Science* 335, 1471–1474.

- (67) Baumgarten, R. J. (1967) Preparation of ethyl diazoacetate via a triazene intermediate. *J. Org. Chem.* *32*, 484–485.
- (68) Schroen, M.; Bräse, S. (2005) Polymer-bound diazonium salts for the synthesis of diazoacetic esters. *Tetrahedron* *61*, 12186–12192.
- (69) Fink, J.; Regitz, M. (1985) Electrophilic diazoalkane substitution. *Synthesis* *1985*, 569–585.
- (70) Ye, T.; McKervey, M. A. (1992) Synthesis of chiral *N*-protected α -amino- β -diketones from α -diazoketones derived from natural amino acids. *Tetrahedron* *48*, 8007–8022.
- (71) Zhao, Y.; Wang, J. (2005) Nucleophilic addition to C=O and C=N bonds by nucleophiles containing a diazo group. *Synlett* *2005*, 2886–2892.
- (72) Meyer, M. E.; Ferreira, E. M.; Stoltz, B. M. (2006) 2-Diazoacetoacetic acid, an efficient and convenient reagent for the synthesis of α -diazo- β -ketoesters. *Chem. Commun.*, 1316–1318.
- (73) Liu, Y.; Zhang, Y.; Jee, N.; Doyle, M. P. (2008) Construction of highly functionalized diazoacetoacetates via catalytic Mukaiyama–Michael reactions. *Org. Lett.* *10*, 1605–1608.
- (74) Maas, G. (2009) New syntheses of diazo compounds. *Angew. Chem., Int. Ed.* *48*, 8186–8195.
- (75) Ford, A.; Miel, H.; Ring, A.; Slattery, C. N.; Maguire, A. R.; McKervey, M. A. (2015) Modern organic synthesis with α -diazocarbonyl compounds. *Chem. Rev.* *115*, 9981–10080.
- (76) Antos, J. M.; Francis, M. B. (2004) Selective tryptophan modification with rhodium carbenoids in aqueous solution. *J. Am. Chem. Soc.* *126*, 10256–10257.
- (77) Goddard-Borger, E. D.; Stick, R. V. (2007) An efficient, inexpensive, and shelf-stable diazotransfer reagent: Imidazole-1-sulfonyl azide hydrochloride. *Org. Lett.* *9*, 3797–3800.
- (78) Myers, E. L.; Raines, R. T. (2009) A phosphine-mediated conversion of azides into diazo compounds. *Angew. Chem., Int. Ed.* *48*, 2359–2363.
- (79) Chou, H.; Raines, R. T. (2013) Conversion of azides into diazo compounds in water. *J. Am. Chem. Soc.* *135*, 14936–14939.

- (80) Nilsson, B. L.; Kiessling, L. L.; Raines, R. T. (2000) Staudinger ligation: A peptide from a thioester and azide. *Org. Lett.* 2, 1939–1941.
- (81) Saxon, E.; Bertozzi, C. R. (2000) Cell surface engineering by a modified Staudinger reaction. *Science* 287, 2007–2010.
- (82) Soellner, M. B.; Nilsson, B. L.; Raines, R. T. (2006) Reaction mechanism and kinetics of the traceless Staudinger ligation. *J. Am. Chem. Soc.* 128, 8820–8828.
- (83) Sletten, E. M.; Bertozzi, C. R. (2011) From mechanism to mouse: A tale of two bioorthogonal reactions. *Acc. Chem. Res.* 44, 666–676.
- (84) McGrath, N. A.; Raines, R. T. (2011) Chemoselectivity in chemical biology: Acyl transfer reactions with sulfur and selenium. *Acc. Chem. Res.* 44, 752–761.
- (85) Staudinger, H.; Meyer, J. (1919) Über neue organische Phosphorverbindungen III. Phosphinmethylderivate und Phosphinimine. *Helv. Chim. Acta* 2, 635–646.
- (86) Staudinger, H.; Hauser, E. (1921) Über neue organische Phosphorverbindungen IV Phosphinimine. *Helv. Chim. Acta* 4, 861–886.
- (87) **Caution!** Sodium azide is nearly as toxic to mammals as is sodium cyanide. For example, the LD₅₀ values for acute dermal toxicity in rabbits are 20 mg/kg and 10.4 mg/kg, respectively (MSDS).
- (88) Singh, A.; Thornton, E. R.; Westheimer, F. H. (1962) The photolysis of diazoacetylchymotrypsin. *J. Biol. Chem.* 237, PC3006–PC3008.
- (89) Ouhia, A.; René, L.; Guilhem, J.; Pascard, C.; Badet, B. (1993) A new diazoacylating reagent: Preparation, structure, and use of succinimidyl diazoacetate. *J. Org. Chem.* 58, 1641–1642.
- (90) Doyle, M. P.; Kalinin, A. V. (1996) Highly enantioselective intramolecular cyclopropanation reactions of *N*-allylic-*N*-methyldiazoacetamides catalyzed by chiral dirhodium(II) carboxamidates. *J. Org. Chem.* 61, 2179–2184.
- (91) Fuerst, D. E.; Stoltz, B. M.; Wood, J. L. (2000) Synthesis of C(3) benzofuran-derived bisaryl quaternary centers: Approaches to diazonamide A. *Org. Lett.* 2, 3521–3523.

- (92) Andersen, K. A.; Aronoff, M. R.; McGrath, N. A.; Raines, R. T. (2015) Diazo groups endure metabolism and enable chemoselectivity *in cellulose*. *J. Am. Chem. Soc.* *137*, 2412–2415.
- (93) Josa-Culleré, L.; Wainman, Y. A.; Brindle, K. M.; Leeper, F. J. (2014) Diazo group as a new chemical reporter for bioorthogonal labelling of biomolecules. *RSC Adv.* *4*, 52241–52244.
- (94) Zhou, W.; Hsieh, P.-H.; Xu, Y.; O’Leary, T. R.; Huang, X.; Liu, J. (2015) Design and synthesis of active heparan sulfate-based probes. *Chem. Commun.* *51*, 11019–11021.
- (95) Friscourt, F.; Fahrni, C. J.; Boons, G.-J. (2015) Fluorogenic strain-promoted alkyne–diazo cycloadditions. *Chem. Eur. J.* *21*, 13996–14001.
- (96) Buchner, E. (1888) Einwirkung von Diazoessigäther auf die Aether ungesättigter Säuren. *Ber. Dtsch. Chem. Ges.* *21*, 2637–2647.
- (97) Huisgen, R. (1963) 1,3-Dipolar cycloadditions. Past and future. *Angew. Chem., Int. Ed.* *2*, 565–598.
- (98) Diazoalkanes. Maas, G. In *Synthetic Applications of 1,3-Dipolar Cycloaddition Chemistry Toward Heterocycles and Natural Products*; Padwa, A., Pearson, W. H., Eds.; John Wiley & Sons, Inc.: New York, NY, 2002, p 539–621.
- (99) Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. (2002) A stepwise Huisgen cycloaddition process: Copper(I)-catalyzed regioselective “ligation” of azides and terminal alkynes. *Angew. Chem., Int. Ed.* *41*, 2596–2599.
- (100) Tornøe, C. W.; Christensen, C.; Meldal, M. (2002) Peptidotriazoles on solid phase: [1,2,3]-Triazoles by regiospecific copper(I)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. *J. Org. Chem.* *67*, 3057–3064.
- (101) Wittig, G.; Krebs, A. (1961) Zur Existenz niedergliedriger Cycloalkine, I. *Chem. Ber.* *94*, 3260–3275.
- (102) Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. (2004) A strain-promoted [3 + 2] azide–alkyne cycloaddition for covalent modification of biomolecules in living systems. *J. Am. Chem. Soc.* *126*, 15046–15047.

- (103) 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. (2007) Copper-free click chemistry for dynamic in vivo imaging. *Proc. Natl. Acad. Sci. USA* *104*, 16793–16797.
- (104) Debets, M. F.; van der Doelen, C. W. J.; Rutjes, F. P. J. T.; van Delft, F. L. (2010) Azide: A unique dipole for metal-free bioorthogonal ligations. *ChemBioChem* *11*, 1168–1184.
- (105) Patterson, D. M.; Nazarova, L. A.; Prescher, J. A. (2014) Finding the right (bioorthogonal) chemistry. *ACS Chem. Biol.* *9*, 592–605.
- (106) Moran, J.; McKay, C. S.; Pezacki, J. P. (2010) Strain-promoted 1,3-dipolar cycloadditions of diazo compounds with cyclooctynes. *Can. J. Chem.* *89*, 148–151.
- (107) Sanders, B. C.; Friscourt, F.; Ledin, P. A.; Mbua, N. E.; Arumugam, S.; Guo, J.; Boltje, T. J.; Popik, V. V.; Boons, G.-J. (2011) Metal-free sequential [3 + 2]-dipolar cycloadditions using cyclooctynes and 1,3-dipoles of different reactivity. *J. Am. Chem. Soc.* *133*, 949–957.
- (108) Bihlmaier, W.; Huisgen, R.; Reissig, H.-U.; Voss, S. (1979) Reactivity sequences of dipolarophiles towards diazocarbonyl compounds—MO perturbation treatment. *Tetrahedron Lett.*, 2621–2624.
- (109) McGrath, N. A.; Raines, R. T. (2012) Diazo compounds as highly tunable reactants in 1,3-dipolar cycloaddition reactions with cycloalkynes. *Chem. Sci.* *3*, 3237–3240.
- (110) Aronoff, M. R.; Gold, B.; Raines, R. T. (2016) 1,3-Dipolar cycloadditions of diazo compounds in the presence of azides. *Org. Lett.* *18*, 1538–1541.
- (111) Aronoff, M. R.; Gold, B.; Raines, R. T. (2016) Rapid cycloaddition of a diazo group with an unstrained dipolarophile. *Tetrahedron Lett.* *57*, 2347–2350.
- (112) Gold, B.; Aronoff, M. R.; Raines, R. T. (2016) 1,3-Dipolar cycloaddition with diazo groups: Noncovalent interactions overwhelm strain. *Org. Lett.* *18*, 4466–4469.
- (113) Gold, B.; Aronoff, M. R.; Raines, R. T. (2016) Decreasing distortion energies without strain: Diazo-selective 1,3-dipolar cycloadditions. *J. Org. Chem.* *81*, 5998–6006.
- (114) Bayley, H.; Knowles, J. R. (1977) Photoaffinity labeling. *Methods Enzymol.* *46*, 69–114.

- (115) Chowdhry, V.; Westheimer, F. H. (1979) Photoaffinity labeling of biological systems. *Annu. Rev. Biochem.* 48, 293-325.
- (116) Wolff, L. (1902) Ueber Diazoanhydride. *Justus Liebigs Ann. Chem.* 325, 129-195.
- (117) Kirmse, W. (2002) 100 Years of the Wolff rearrangement. *Eur. J. Org. Chem.*, 2193-2256.
- (118) Converse, C. A.; Richards, F. F. (1969) Two-stage photosensitive label for antibody combining sites. *Biochemistry* 8, 4431-4436.
- (119) Gupta, C. M.; Costello, C. E.; Khorana, H. G. (1979) Sites of intermolecular crosslinking of fatty acyl chains in phospholipids carrying a photoactivable carbene precursor. *Proc. Natl. Acad. Sci. USA* 76, 3139-3143.
- (120) Kale, T. A.; Distefano, M. D. (2003) Diazotrifluoropropionamido-containing prenylcysteines: Syntheses and applications for studying isoprenoid-protein interactions. *Org. Lett.* 5, 609-612.
- (121) Padwa, A.; Weingarten, M. D. (1996) Cascade processes of metallocarbenoids. *Chem. Rev.* 96, 223-269.
- (122) *Modern Catalytic Methods for Organic Synthesis with Diazo Compounds: From Cyclopropanes to Ylides.* Doyle, M. P.; McKervey, M. A.; Ye, T.; John Wiley & Sons: New York, NY, 1998.
- (123) Davies, H. M. L.; Beckwith, R. E. J. (2003) Catalytic enantioselective C-H activation by means of metal-carbenoid-induced C-H insertion. *Chem. Rev.* 103, 2861-2904.
- (124) Candelas, N. R.; Alfonso, C. A. (2009) Developments in the photochemistry of diazo compounds. *Curr. Org. Chem.* 13, 763-787.
- (125) Geake, A.; Nierenstein, M. (1914) The action of diazomethane on caseinogen. *Biochem. J.* 8, 287-292.
- (126) Doscher, M. S.; Wilcox, P. E. (1961) Chemical derivatives of α -chymotrypsinogen: IV. A comparison of the reactions of α -chymotrypsinogen and of simple carboxylic acids with diazoacetamide. *J. Biol. Chem.* 236, 1328-1337.

- (127) Riehm, J. P.; Scheraga, H. A. (1965) Structural studies of ribonuclease. XVII. A reactive carboxyl group in ribonuclease. *Biochemistry* 4, 772–782.
- (128) McGrath, N. A.; Andersen, K. A.; Davis, A. K. F.; Lomax, J. E.; Raines, R. T. (2015) Diazo compounds for the bioreversible esterification of proteins. *Chem. Sci.* 6, 752–755.
- (129) Mix, K. A.; Raines, R. T. (2015) Optimized diazo scaffold for protein esterification. *Org. Lett.* 17, 2359–2361.
- (130) Chibnall, A. C.; Rees, M. W. (1957) Studies on the amide and C-terminal residues in proteins. *Biochem. J.* 68, 105–111.
- (131) Grossberg, A. L.; Pressman, D. (1960) Nature of the combining site of antibody against a hapten bearing a positive charge. *J. Am. Chem. Soc.* 82, 5478–5482.
- (132) Rajagopalan, T. G.; Stein, W. H.; Moore, S. (1966) The inactivation of pepsin by diazoacetyl norleucine methyl ester. *J. Biol. Chem.* 241, 4295–4297.
- (133) Ong, E. B.; Perlmann, G. E. (1967) Specific inactivation of pepsin by benzyloxycarbonyl-L-phenyldiazomethane. *Nature* 215, 1492–1494.
- (134) Bayliss, R. S.; Knowles, J. R. (1968) An active site peptide from pepsin. *Chem. Commun.*, 196–198.
- (135) Takahashi, K.; Mizobe, F.; Chang, W. (1972) Inactivation of acid proteases from *Rhizopus chinensis*, *Aspergillus saitoi* and *Mucor pusillus*, and calf rennin by diazoacetyl norleucine methyl ester. *J. Biochem.* 71, 161–164.
- (136) Kovaleva, G. G.; Shimanskaya, M. P.; Stepanov, V. M. (1972) The site of diazoacetyl inhibitor attachment to acid proteinase of *Aspergillus awamori*—An analog of penicillopepsin and pepsin. *Biochem. Biophys. Res. Commun.* 49, 1075–1081.
- (137) Murao, S.; Oda, K.; Matsushita, Y. (1972) New acid proteases from *Scytalidium lignicolum* M-133. *Agr. Biol. Chem.* 36, 1647–1650.
- (138) Mizobe, F.; Takahashi, K.; Ando, T. (1973) The structure and function of acid proteases. *J. Biochem.* 73, 61–68.

- (139) Johnson, R. L.; Poisner, A. M. (1980) Inactivation of amniotic prorenin by ethyl diazoacetylglucinate. *Biochem. Biophys. Res. Commun.* 95, 1404–1409.
- (140) Delpierre, G. R.; Fruton, J. S. (1966) Specific inactivation of pepsin by a diazo ketone. *Proc. Natl. Acad. Sci. USA* 56, 1817–1822.
- (141) Vohidov, F.; Coughlin, J. M.; Ball, Z. T. (2015) Rhodium(II) metallopeptide catalyst design enables fine control in selective functionalization of natural SH3 domains. *Angew. Chem., Int. Ed.* 54, 4587–4591.
- (142) Bao, Z.; Wang, S.; Shi, W.; Dong, S.; Ma, H. (2007) Selective modification of Trp19 in β -lactoglobulin by a new diazo fluorescence probe. *J. Proteome Res.* 6, 3835–3841.
- (143) Staudinger, H.; Gaule, A. (1917) Diphenyldiazomethan. *Ber. Dtsch. Chem. Ges.* 49, 1951–1960.
- (144) Roberts, J. D.; Watanabe, W.; McMahon, R. E. (1951) The kinetics and mechanism of the reaction of diphenyldiazomethane and benzoic acid in ethanol. *J. Am. Chem. Soc.* 73, 760–765.
- (145) Delpierre, G. R.; Fruton, J. S. (1965) Inactivation of pepsin by diphenyldiazomethane. *Proc. Natl. Acad. Sci. USA* 54, 1161–1167.
- (146) Hamilton, G. A.; Spona, J.; Crowell, L. D. (1967) The inactivation of pepsin by an equimolar amount of 1-diazo-4-phenylbutanone-2. *Biochem. Biophys. Res. Commun.* 26, 193–198.
- (147) Fry, K. T. (1968) A reactive aspartyl residue of pepsin. *Biochem. Biophys. Res. Commun.* 30, 489–495.
- (148) Stepanov, V. M.; Vaganova, T. I. (1968) Identification of the carboxyl group of pepsin reacting with diazoacetamide derivatives. *Biochem. Biophys. Res. Commun.* 31, 825–830.
- (149) Lundblad, R. L.; Stein, W. H. (1969) On the reaction of diazoacetyl compounds with pepsin. *J. Biol. Chem.* 244, 154–160.
- (150) Bayliss, R. S.; Knowles, J. R.; Wybrandt, G. B. (1969) An aspartic acid residue at the active site of pepsin. *Biochem. J.* 113, 377–386.

- (151) Edman, P.; Högfeldt, E.; Sillén, L. G.; Kinell, P. (1950) Method for the determination of the amino acid sequence in peptides. *Acta Chem. Scand.* **4**, 283–293.
- (152) Takahashi, K.; Chang, W. (1973) Specific chemical modifications of acid proteinases in the presence and absence of pepstatin. *J. Biochem.* **73**, 675–677.
- (153) Oda, K.; Sugitani, M.; Fukuhara, K.; Murao, S. (1987) Purification and properties of a pepstatin-insensitive carboxyl proteinase from a Gram-negative bacterium. *Biochem. Biophys. Acta* **923**, 463–469.
- (154) Popp, B. V.; Ball, Z. T. (2010) Structure-selective modification of aromatic side chains with dirhodium metallopeptide catalysts. *J. Am. Chem. Soc.* **132**, 6660–6662.
- (155) Our strategy is reminiscent of the use of small-molecule prodrugs, which have been in the pharmacopoeia for over fifty years. a) A. Albert, *Nature* **1958**, *182*, 421–422; b) B. Testa, J. M. Mayer, *Hydrolysis in Drug and Prodrug Metabolism*, Wiley–VCH, Weinheim, Germany, **2003**; c) B. M. Liederer, R. T. Borchardt, *J. Pharm. Sci.* **2006**, *95*, 1177–1195; d) K. M. Huttunen, H. Raunio, J. Rautio, *Pharmacol. Rev.* **2011**, *63*, 750–771.
- (156) Liederer, B. M.; Borchardt, R. T. (2006) Enzymes involved in the bioconversion of ester-based prodrugs. *J. Pharm. Sci.* **95**, 1177–1195.
- (157) Tian, L.; Yang, Y.; Wysocki, L. M.; Arnold, A. C.; Hu, A.; Ravichandran, B.; Stenerson, S. M.; Looger, L. L.; Lavis, L. D. (2012) Selective esterase–ester pair for targeting small molecules with cellular specificity. *Proc. Natl. Acad. Sci. USA.* **109**, 4756–4761.
- (158) Harris, J. M.; Chess, R. B. (2003) Effect of pegylation on pharmaceuticals. *Nat. Rev. Drug Discov.* **2**, 214–221.
- (159) Szele, I.; Tencer, M.; Zollinger, H. (1983) 163. Reactions of alkenediazonium salts. Part 1. 2,2-Diethoxyethenediazonium hexachloroantimonate: A diazonium, a carbenium, or an oxonium salt? *Helv. Chim. Acta* **66**, 1691–1703.
- (160) Nozaki, H.; Moriuti, S.; Takaya, H.; Noyori, R. (1966) Asymmetric induction in carbenoid reaction by means of a dissymmetric copper chelate. *Tetrahedron Lett.* **7**, 5239–5244.
- (161) Ball, Z. T. (2013) Designing enzyme-like catalysts: A rhodium(II) metallopeptide case study. *Acc. Chem. Res.* **46**, 560–570.

- (162) Ball, Z. T. (2015) Molecular recognition in protein modification with rhodium metallopeptides. *Curr. Opin. Chem. Biol.* 25, 98–102.
- (163) Popp, B. V.; Ball, Z. T. (2011) Proximity-driven metallopeptide catalysis: Remarkable side-chain scope enables modification of the Fos bZip domain. *Chem. Sci.* 2, 690–695.
- (164) Chen, Z.; Popp, B. V.; Bovet, C. L.; Ball, Z. T. (2011) Site-specific protein modification with a dirhodium metallopeptide catalyst. *ACS Chem. Biol.* 6, 920–925.
- (165) Chen, Z. C.; Coughlin, J. M.; Stagg, L. J.; Arold, S. T.; Ladbury, J. E.; Ball, Z. T. (2012) Catalytic protein modification and dirhodium metallopeptides: Specificity in designed and natural systems. *J. Am. Chem. Soc.* 134, 10138–10145.
- (166) Tishinov, K.; Schmidt, K.; Häussinger, D.; Gillingham, D. G. (2012) Structure-selective catalytic alkylation of DNA and RNA. *Angew. Chem. Int. Ed.* 51, 12000–12004.
- (167) Tishinov, K.; Fei, N.; Gillingham, D. (2013) Cu(I)-catalysed N–H insertion in water: A new tool for chemical biology. *Chem. Sci.* 4, 4401–4406.
- (168) Ando, H.; Futara, T.; Tsien, R. Y.; Okamoto, H. (2001) Photo-mediated gene activation using caged RNA/DNA in zebrafish embryos. *Nat. Genet.* 28, 317–325.
- (169) Gillingham, D.; Fei, N. (2013) Catalytic X–H insertion reactions based on carbenoids. *Chem. Soc. Rev.* 42, 4918–4931.
- (170) Shah, S.; Jain, P. K.; Kala, A.; Karunakaran, D.; Friedman, S. H. (2009) Light-activated RNA interference using double-stranded siRNA precursors modified using a remarkable regiospecificity of diazo-based photolabile groups. *Nucleic Acids Res.* 37, 4508–4517.
- (171) Laayoun, A.; Kotera, M.; Sothier, I.; Trévisiol, E.; Bernal-Méndez, E.; Bourget, C.; Menou, L.; Lhomme, J.; Troesch, A. (2003) Aryldiazomethanes for universal labeling of nucleic acids and analysis on DNA chips. *Bioconjugate Chem.* 14, 1298–1306.
- (172) Fei, N.; Sauter, B.; Gillingham, D. (2016) The pK_a of Brønsted acids controls their reactivity with diazo compounds. *Chem. Commun.* 52, 7501–7504.

- (173) **Caution!** Unlike stabilized diazo compounds (*e.g.*, diazo compound **I** in Figure 4A), unstabilized diazo compounds (*e.g.*, diazomethane) are dangerous and should never be used in the context of chemical biology. See: refs. 5–9.
- (174) *Modern Catalytic Methods for Organic Synthesis with Diazo Compounds*. Doyle, M. P.; McKervey, M. A.; Ye, T.; Wiley: New York, NY, 1998.
- (175) Davies, H. M. L.; Beckwith, R. E. J. (2003) Catalytic Enantioselective C-H Activation by Means of Metal-Carbenoid-Induced C-H Insertion. *Chem. Rev.* *103*, 2861–2904.
- (176) Antos, J. M.; McFarland, J. M.; Lavarone, A. T.; Francis, M. B. (2009) Chemoselective Tryptophan Labeling with Rhodium Carbenoids at Mild pH. *J. Am. Chem. Soc.* *131*, 6301–6308.
- (177) *Hydrolysis in Drug and Prodrug Metabolism*. Testa, B.; Mayer, J. M.; Verlag Helvetica Chimica Acta: Zurich, Switzerland, 2003.
- (178) Lavis, L. D. (2008) Ester Bonds in Prodrugs. *ACS Chem. Biol.* *3*, 203–206.
- (179) Tian, L.; Yang, Y.; Wysocki, L. M.; Arnold, A. C.; Hu, A.; Ravichandran, B.; Stenerson, S. M.; Looger, L. L.; Lavis, L. D. (2012) Selective esterase-ester pair for targeting small molecules with cellular specificity. *Proc. Natl. Acad. Sci. USA* *109*, 4756–4761.
- (180) Roberts, J. D.; Watanabe, W.; McMahon, R. E. (1951) *J. Am. Chem. Soc.* *73*, 2521–2523.
- (181) Doscher, M. S.; Wilcox, P. E. (1961) Chemical Derivatives of alpha-Chymotrypsinogen: IV. A Comparison of the Reactions of alpha-Chymotrypsinogen and of Simple Carboxylic Acids with Diazoacetamide. *J. Biol. Chem.* *236*, 1328–1337.
- (182) Riehm, J. P.; Sheraga, H. A. (1965) Structural Studies of Ribonuclease. XVII. A Reactive Carboxyl Group in Ribonuclease. *Biochemistry* *4*, 772–782.
- (183) Hammett, L. P. (1935) *Chem. Rev.* *17*, 125–136.
- (184) Hammett, L. P. (1937) The effect of structure upon the reactions of organic compounds. Benzene derivatives. *J. Am. Chem. Soc.* *59*, 96–103.
- (185) Hammett, L. P. In *Physical Organic Chemistry*; McGraw–Hill: New York, NY, 1940, p 184–228.

- (186) Shorter, J. (2000) *Chem. Listy* 94, 210–214.
- (187) Chou, H.-H.; Raines, R. T. (2013) Conversion of Azides into Diazo Compounds in Water. *J. Am. Chem. Soc.* 135, 14936–14939.
- (188) *Modern Physical Organic Chemistry*. Anslyn, E. V.; Dougherty, D. A.; University Science Books: Sausalito, CA, 2006.
- (189) Hansch, C.; Leo, A.; Taft, R. W. (1991) A survey of Hammett substituent constants and resonance and field parameters. *Chem. Rev.* 91, 165–175.
- (190) McGrath, N. A.; Andersen, K. A.; Davis, A. K. F.; Lomax, J. E.; Raines, R. T. (2014) Diazo compounds for the bioreversible esterification of proteins. *Chem. Sci.* 6, 752–755.
- (191) Raines, R. T. (1998) Ribonuclease A. *Chem. Rev.* 98, 1045–1066.
- (192) Goddard-Borger, E. D., Stick, R.V. (2007) An Efficient, Inexpensive, and Shelf-Stable Diazotransfer Reagent: Imidazole-1-sulfonyl Azide Hydrochloride. *Org. Lett.* 9, 3797–3800.
- (193) Dimitrov, D. S. (2012) Therapeutic proteins. *Methods Mol. Biol.* 899, 1–26.
- (194) Pisal, D. S.; Kosloski, M. P.; Balu-Iyer, S. V. (2010) Delivery of therapeutic proteins. *J. Pharm. Sci.* 99, 2557–2575.
- (195) Fu, A.; Tang, R.; Hardie, J.; Farkas, M. E.; Rotello, V. M. (2014) Promises and pitfalls of intracellular delivery of proteins. *Bioconjugate Chem.* 25, 1602–1608.
- (196) Somia, N.; Verma, I. M. (2000) Gene therapy: Trials and tribulations. *Nat. Rev. Genet.* 1, 91–99.
- (197) Collins, M.; Thrasher, A. (2015) Gene therapy: Progress and predictions. *Proc. Biol. Sci.* 282, 20143003.
- (198) Leader, B.; Baca, Q. J.; Golan, D. E. (2008) Protein therapeutics: A summary and pharmacological classification. *Nat. Rev. Drug Discov.* 7, 21–39.

- (199) Fuchs, S. M.; Raines, R. T. (2007) Arginine grafting to endow cell permeability. *ACS Chem. Biol.* *2*, 167–170.
- (200) McNaughton, B. R.; Cronican, J. J.; Thompson, D. B.; Liu, D. R. (2009) Mammalian cell penetration, siRNA transfection, and DNA transfection by supercharged proteins. *Proc. Natl. Acad. Sci. USA* *106*, 6111–6116.
- (201) Bruce, V. J.; Lopez-Islas, M.; McNaughton, B. R. (2016) Resurfaced cell-penetrating nanobodies: A potentially general scaffold for intracellularly targeted protein discovery. *Protein Sci.* *25*, 1129–1137.
- (202) Futami, J.; Yamada, H. (2008) Design of cytotoxic ribonucleases by cationization to enhance intracellular protein delivery. *Curr. Pharm. Biotechnol.* *9*, 180–184.
- (203) Ellis, G. A.; Palte, M. J.; Raines, R. T. (2012) Boronate-mediated biologic delivery. *J. Am. Chem. Soc.* *134*, 3631–3634.
- (204) Schwarze, S. R.; Ho, A.; Vocero-Akbani, A.; Dowdy, S. F. (1999) *In vivo* protein transduction: Delivery of a biologically active protein into the mouse. *Science* *285*, 1569–1572.
- (205) Fuchs, S. M.; Raines, R. T. (2005) Polyarginine as a multifunctional fusion tag. *Protein Sci.* *14*, 1538–1544.
- (206) Stanzl, E. G.; Trantow, B. M.; Vargas, J. R.; Wender, P. A. (2013) Fifteen years of cell-penetrating, guanidinium-rich molecular transporters: Basic science, research tools, and clinical applications. *Acc. Chem. Res.* *46*, 2944–2954.
- (207) Nischan, N.; Herce, H. D.; Natale, F.; Bohlke, N.; Budisa, N.; Cardoso, M. C.; Hackenberger, C. P. R. (2015) Covalent attachment of cyclic TAT peptides to GFP results in protein delivery into live cells with immediate bioavailability. *Angew. Chem. Int. Ed.* *54*, 1950–1953.
- (208) LaRochelle, J. R.; Cobb, G. B.; Steinauer, A.; Rhoades, E.; Schepartz, A. (2015) Fluorescence correlation spectroscopy reveals highly efficient cytosolic delivery of certain pentamer proteins and stapled peptides. *J. Am. Chem. Soc.* *137*, 2536–2541.
- (209) Qian, Z.; Martyna, A.; Hard, R. L.; Wang, J.; Appiah-Kubi, G.; Coss, C.; Phelps, M. A.; Rossman, J. S.; Pei, D. (2016) Discovery and mechanism of highly efficient cyclic cell-penetrating peptides. *Biochemistry* *55*, 2601–2612.

- (210) Nagel, Y. A.; Raschle, P. S.; Wennemers, H. (2017) Effect of preorganized charge-display on the cell-penetrating properties of cationic peptides. *Angew. Chem. Int. Ed.* *56*, 122–126.
- (211) Zuris, J. A.; Thompson, D. B.; Shu, Y.; Guilinger, J. P.; Bessen, J. L.; Hu, J. H.; Maeder, M. L.; Joung, J. K.; Chen, Z.-Y.; Liu, D. R. (2015) Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing *in vitro* and *in vivo*. *Nat. Biotechnol.* *33*, 73–80.
- (212) Kim, S.; Kim, D.; Cho, S. W.; Kim, J.; Kim, J.-S. (2014) Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* *24*, 1012–1019.
- (213) Palte, M. J.; Raines, R. T. (2012) Interaction of nucleic acids with the glycocalyx. *J. Am. Chem. Soc.* *134*, 6218–6223.
- (214) Turcotte, R. F.; Lavis, L. D.; Raines, R. T. (2009) Onconase cytotoxicity relies on the distribution of its positive charge. *FEBS J.* *276*, 4270–4281.
- (215) Perrett, F.; Nishihara, M.; Takeuchi, T.; Futaki, S.; Lazar, A. N.; Coleman, A. W.; Sakai, N.; Matile, S. (2005) Anionic fullerenes, calixarenes, coronenes, and pyrenes as activators of oligo/polyarginines in model membranes and live cells. *J. Am. Chem. Soc.* *127*, 1114–1115.
- (216) Takeuchi, T.; Kosuge, M.; Tadokoro, A.; Sugiura, Y.; Nishi, M.; Kawata, M.; Sakai, N.; Matile, S.; Futaki, S. (2006) Direct and rapid cytosolic delivery using cell-penetrating peptides mediated by pyrenebutyrate. *ACS Chem. Biol.* *1*, 299–303.
- (217) Elmquist, A.; Hansen, M.; Langel, Ü. (2006) Structure–activity relationship study of the cell-penetrating peptide *pVEC*. *Biochim. Biophys. Acta* *1758*, 721–729.
- (218) Di Pisa, M.; Chassaing, G.; Swiecicki, J.-M. (2015) Translocation mechanism(s) of cell-penetrating peptides: Biophysical studies using artificial membrane bilayers. *Biochemistry* *54*, 194–207.
- (219) Liederer, B. M.; Borchardt, R. T. (2006) *J. Pharm. Sci.*, 1177–1195.
- (220) Huttunen, K. M.; Raunio, H.; Rautio, J. (2011) *Pharmacol. Rev.* *63*, 750–771.

- (221) *Prodrug Design: Perspectives, Approaches and Applications in Medicinal Chemistry*. Redasani, V. K.; Bari, S. B.; Academic Press: New York, NY, 2015.
- (222) Mix, K. A.; Aronoff, M. R.; Raines, R. T. (2016) Diazo compounds: Versatile tools for chemical biology. *ACS Chem. Biol.* *11*, 3233–3244.
- (223) Tomoda, H.; Kishimoto, Y.; Lee, Y. C. (1989) Temperature effect on endocytosis and exocytosis by rabbit alveolar macrophages. *J. Biol. Chem.* *264*, 15445–15450.
- (224) Maier, K.; Wagner, E. (2012) Acid-labile traceless click linker for protein transduction. *J. Am. Chem. Soc.* *134*, 10169–10173.
- (225) Ray, M.; Tang, R.; Jiang, Z.; Rotello, V. M. (2015) Quantitative tracking of protein trafficking to the nucleus using cytosolic protein delivery by nanoparticle-stabilized nanocapsules. *Bioconjugate Chem.* *26*, 1004–1007.
- (226) Andersen, K. A.; Smith, T. P.; Lomax, J. E.; Raines, R. T. (2016) Boronic acid for the traceless delivery of proteins into cells. *ACS Chem. Biol.* *11*, 319–323.
- (227) Pédelacq, J. D.; Cabantous, S.; Tran, T.; Terwilliger, T. C.; Waldo, G. S. (2006) Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* *24*, 79–88.
- (228) Hoang, T. T.; Raines, R. T. (2017) Molecular basis for the autonomous promotion of cell proliferation by angiogenin. *Nucleic Acids. Res.* *45*, 818–831.
- (229) Beck, A.; Wurch, T.; Bailly, C.; Crovaia, N. (2010) Strategies and challenges for the next generation of therapeutic antibodies. *Nat. Rev. Immunol.* *10*, 345–352.
- (230) Mulydermans, S. (2013) Nanobodies: Natural single-domain antibodies. *Annu. Rev. Biochem.* *82*, 775–797.
- (231) Hollinger, P.; Hudson, P. J. (2005) Engineered antibody fragments and the rise of single domains. *Nat. Biotechnol.* *23*, 1126–1136.
- (232) Ahmad, Z. A.; Yeap, S. K.; Ali, A. A.; Ho, W. Y.; Alitheen, N. B. M.; Hamid, M. (2012) scFv antibody: Principles and clinical applications. *Clin. Dev. Immunol.* *2012*, 1–15.

- (233) Hornsby, M.; Paduch, M.; Miersch, S.; Sääf, A.; Matsuguchi, T.; Lee, B.; Wypisniak, K.; Doak, A.; King, D.; Usatyuk, S.; Perry, K.; Lu, V.; Thomas, W.; Luke, J.; Goodman, J.; Hoey, R. J.; Lai, D.; Griffin, C.; Li, Z.; Vizeacoumar, F. J.; Dong, D.; Campbell, E.; Anderson, S.; Zhong, N.; Gräslund, S.; Koide, S.; Moffat, J.; Sidhu, S.; Kossiakoff, A.; Wells, J. A. (2015) A high throughput platform for recombinant antibodies to folded proteins. *Mol. Cell Proteomics* 14, 2833–2847.
- (234) Marschall, A. L. J.; Frenzel, A.; Schirrmann, T.; Schüngel, M.; Dubel, S. (2011) Targeting antibodies to the cytoplasm. *mAbs* 3, 3–16.
- (235) Stocks, M. (2005) Intrabodies as drug discovery tools and therapeutics. *Curr. Opin. Chem. Biol.* 9, 359–365.
- (236) Hoefman, S.; Ottevaere, I.; Baumeister, J.; Sargentini-Maier, M. L. (2015) Pre-clinical intravenous serum pharmacokinetics of albumin binding and non-half-life-extended nanobodies. *Antibodies* 4, 141–156.
- (237) Nakase, I.; Takeuchi, T.; Tanaka, G.; Futaki, S. (2008) Methodological and cellular aspects that govern the internalization mechanisms of arginine-rich cell-penetrating peptides. *Adv. Drug Deliv. Rev.* 60, 598–607.
- (238) Shaw, P. A.; Catchpole, I. R.; Goddard, C. A.; Colledge, W. H. (2008) Comparison of protein transduction domains in mediating cell delivery of a secreted CRE protein. *Biochemistry* 47, 1157–1166.
- (239) Herce, H. D.; Schumacher, D.; Schneider, A. F. L.; Ludwig, A. K.; Mann, F. A.; Fillies, M.; Kasper, M.-A.; Reinke, S.; Krause, E.; Leonhardt, H.; Cardoso, M. C.; Hackenberger, C. P. R. (2017) Cell-permeable nanobodies for targeted immunolabelling and antigen manipulation in living cells. *Nat. Chem.* 9, 762–771.
- (240) Lin, S.; Yang, X.; Jia, S.; Weeks, A. M.; Hornsby, M.; Lee, P. S.; Nichiporuk, R. V.; Iavarone, A. T.; Wells, J. A.; Toste, F. D.; Chang, C. J. (2017) Redox-based reagents for chemoselective methionine bioconjugation. *Science* 355, 597–602.
- (241) Mix, K. A.; Lomax, J. E.; Raines, R. T. (2017) Cytosolic delivery of proteins by bioreversible esterification. *J. Am. Chem. Soc.* 139, 14396–14398.
- (242) Martin, R. M.; Ter-Avetisyan, G.; Herce, H. D.; Ludwig, A. K.; Lättig-Tünnemann, G.; Cardoso, M. C. (2015) Principles of protein targeting to the nucleolus. *Nucleus* 5, 314–325.

- (243) Karnoub, A. E.; Weinberg, R. A. (2008) Ras oncogenes: Split personalities. *Nat. Rev. Mol. Cell Biol.* *9*, 517–531.
- (244) Devaraj, N. K.; Upadhyay, R.; Huan, J. B.; Hilderbrand, S. A.; Weissleder, R. (2009) Fast and sensitive pretargeted labeling of cancer cells through a tetrazine/trans-cyclooctene cycloaddition. *Angew. Chem. Int. Ed.* *48*, 7013–7016.
- (245) Pleiner, T.; Bates, M.; Trakhanov, S.; Lee, C.-T.; Schliep, J. E.; Chug, H.; Böhning, M.; Stark, H.; Urlaub, H.; Görlich, D. (2015) Nanobodies: Site-specific labeling for super-resolution imaging, rapid epitope-mapping and native protein complex isolation. *eLife* *4*, e11349.
- (246) Casi, G.; Huguenin-Dezot, N.; Zuberbühler, K.; Scheuermann, J.; Neri, D. (2012) Site-specific traceless coupling of potent cytotoxic drugs to recombinant antibodies for pharmacodelivery. *J. Am. Chem. Soc.* *134*, 5887–5892.
- (247) Sievers, E. L.; Senter, P. D. (2013) Antibody–drug conjugates in cancer therapy. *Annu. Rev. Med.* *64*, 15–29.
- (248) Parslow, A. C.; Parakh, S.; Lee, F.-T.; Gan, H. K.; Scott, A. M. (2016) Antibody–drug conjugates for cancer therapy. *Biomedicines* *4*, 14–20.
- (249) Schaefer, G., Haber, L., Crocker, L.M., Shia, S., Shao, L., Dowbenko, D., Totpal, K., Wong, A., Lee, C.V., Stawicki, S., Clark, R., Fields, C., Phillips, G.D.L., Prell, R.A., Danilenko, D.M., Franke, Y., Stephan, J.P., Hwang, J., Wu, Y., Bostrom, J., Sliwkowski, M.X., Fuh, G., Eigenbrot, C. (2011) A two-in-one antibody against HER3 and EGFR has superior inhibitory activity compared with monospecific antibodies. *Cancer Cell* *20*, 472–486.
- (250) Bernardes, G. J. L.; Steiner, M.; Hartmann, I.; Neri, D.; Casi, G. (2013) Site-specific chemical modification of antibody fragments using traceless cleavable linkers. *Nat. Protoc.* *8*, 2079–2089.
- (251) Xiao, H.; Woods, E. C.; Vukojicic, P.; Bertozzi, C. R. (2016) Precision glycolyx editing as a strategy for cancer immunotherapy. *Proc. Natl. Acad. Sci. USA.* *113*, 10304–10309.
- (252) Junutula, J. R.; Raab, H.; Clark, S.; Bhakta, S.; Leipold, D. D.; Weir, S.; Chen, Y.; Simpson, M.; Tsai, S. P.; Dennis, M. S.; Lu, Y.; Meng, Y. G.; Ng, C.; Yang, J.; Lee, C. C.; Duenas, E.; Gorrell, J.; Katta, V.; Kim, A.; McDorman, K.; Flagella, K.; Venook, R.; Ross, S.; Spencer, S. D.; Lee Wong, W.; Lowman, H. B.; Vandlen, R.; Sliwkowski, M. X.; Scheller, R. H.; Polakis, P.; Mallet, W. (2008) Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. *Nat. Biotechnol.* *26*, 925–932.

- (253) Mårilind, J.; Kaspar, M.; Trachsel, E.; Somnavilla, R.; Hindle, S.; Bacci, C.; Giovannoni, L.; Neri, D. (2008) Antibody-mediated delivery of interleukin-2 to the stroma of breast cancer strongly enhances the potency of chemotherapy. *Clin. Cancer Res.* *14*, 6515–6524.
- (254) Gutbrodt, K. L.; Schliemann, C.; Giovannoni, L.; Frey, K.; Pabst, T.; Klapper, W.; Berdel, W. E.; Neri, D. (2013) Antibody-based delivery of interleukin-2 to neovasculature has potent activity against acute myeloid leukemia. *Sci. Transl. Med.* *5*, 201ra118.
- (255) 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. (2015) Site-specific conjugation improves therapeutic index of antibody drug conjugates with high drug loading. *Nat. Biotechnol.* *33*, 694–696.
- (256) Muir, T. W.; Sondhi, D.; Cole, P. A. (1998) Expressed protein ligation: A general method for protein engineering. *Proc. Natl. Acad. Sci. USA* *95*, 6705–6710.
- (257) Marshall, C. J.; Agarwal, N.; Kalia, J.; Grosskopf, V. A.; McGrath, N. A.; Abbott, N. L.; Raines, R. T.; Shusta, E. V. (2013) Facile chemical functionalization of proteins through intein-linked yeast display. *Bioconjug. Chem.* *24*, 1634–1644.
- (258) Shusta, E. V.; Holler, P. D.; Kieke, M. C.; Kranz, D. M.; Wittrup, K. D. (2000) Directed evolution of a stable scaffold for T-cell receptor engineering. *Nat. Biotechnol.* *18*, 754–759.
- (259) Huang, D.; Gore, P. R.; Shusta, E. V. (2008) Increasing yeast secretion of heterologous proteins by regulating expression rates and post-secretory loss. *Biotechnol. Bioeng.* *101*, 1264–1275.
- (260) Tillotson, B. J.; Lajoie, J. M.; Shusta, E. V. (2015) Yeast display-based antibody affinity maturation using detergent-solubilized cell lysates. *Methods Mol. Biol.* *1319*, 65–78.
- (261) Valiyaveetil, F. I.; Sekedat, M.; MacKinnon, R.; Muir, T. W. (2004) Glycine as a D-amino acid surrogate in the K⁺-selectivity filter. *Proc. Natl. Acad. Sci. USA* *101*, 17045–17049.
- (262) Flavell, R. R.; Kothari, P.; Bar-Dagan, M.; Synan, M.; Vallabhajosula, S.; Friedman, J. M.; Muir, T. W.; Ceccarini, G. (2008) Site-specific ¹⁸F-labeling of the protein hormone leptin using a general two-step ligation procedure. *J. Am. Chem. Soc.* *130*, 9106–9112.

- (263) Reihl, O.; Lederer, M. O.; Schwack, W. (2004) Characterization and detection of lysine–arginine cross-links derived from dehydroascorbic acid. *Carbohydr. Res.* *339*, 483–491.
- (264) Brewer, G. J. (2010) Risks of copper and iron toxicity during aging in humans. *Chem. Res. Toxicol.* *23*, 319–326.
- (265) Hong, V.; Presolski, S. I.; Ma, C.; Finn, M. G. (2009) Analysis and optimization of copper-catalyzed azide–alkyne cycloaddition for bioconjugation. *Angew. Chem. Int. Ed.* *48*, 9879–9883.
- (266) McKay, Craig S.; Finn, M. G. (2014) Click chemistry in complex mixtures: Bioorthogonal bioconjugation. *Chem. Biol.* *21*, 1075–1101.
- (267) Beatty, K. E.; Fisk, J. D.; Smart, B. P.; Lu, Y. Y.; Szychowski, J.; Hangauer, M. J.; Baskin, J. M.; Bertozzi, C. R.; Tirrell, D. A. (2010) Live-cell imaging of cellular proteins by a strain-promoted azide–alkyne cycloaddition. *ChemBioChem* *11*, 2092–2095.
- (268) Taylor, M. T.; Blackman, M. L.; Dmitrenko, O.; Fox, J. M. (2011) Design and synthesis of highly reactive dienophiles for the tetrazine–*trans*-cyclooctene ligation. *J. Am. Chem. Soc.* *133*, 9646–9649.
- (269) Blackman, M. L.; Royzen, M.; Fox, J. M. (2008) Tetrazine ligation: Fast bioconjugation based on inverse-electron-demand Diels–Alder reactivity. *J. Am. Chem. Soc.* *130*, 13518–13519.
- (270) Wijnen, J. W.; Zavarise, S.; Engberts, J. B. F. N. (1996) Substituent effects on an inverse electron demand hetero Diels–Alder reaction in aqueous solution and organic solvents: Cycloaddition of substituted styrenes to di(2-pyridyl)-1,2,4,5-tetrazine. *J. Org. Chem.* *61*, 2001–2005.
- (271) Knall, A.-C.; Hollauf, M.; Slugovc, C. (2014) Kinetic studies of inverse electron demand Diels–Alder reactions (iEDDA) of norbornenes and 3,6-dipyridin-2-yl-1,2,4,5-tetrazine. *Tetrahedron Lett.* *55*, 4763–4766.
- (272) Zhang, H.; Trout, W. S.; Liu, S.; Andrade, G. A.; Hudson, D. A.; Scinto, S. L.; Dicker, K. T.; Li, Y.; Lazouski, N.; Rosenthal, J.; Thorpe, C.; Jia, X.; Fox, J. M. (2016) Rapid bioorthogonal chemistry turn-on through enzymatic or long wavelength photocatalytic activation of tetrazine ligation. *J. Am. Chem. Soc.* *138*, 5978–5983.

- (273) Rieder, U.; Luedtke, N. W. (2014) Alkene–tetrazine ligation for imaging cellular DNA. *Angew. Chem. Int. Ed.* *126*, 9322–9326.
- (274) Sauer, J., Heldmann, D.K., Hetzenegger, J., Krauthan, J., Sichert, H., Schuster, J. (1998) 1,2,4,5-Tetrazine: Synthesis and reactivity in [4+2] cycloadditions. *Eur. J. Org. Chem.*
- (275) Knall, A.-C.; Slugovc, C. (2013) Inverse electron demand Diels–Alder (iEDDA)-initiated conjugation: A (high) potential click chemistry scheme. *Chem. Soc. Rev.* *42*, 5131–5142.
- (276) Brown, K. C. (2010) Peptidic tumor targeting agents: The road from phage display peptide selections to clinical applications. *Curr. Pharm. Des.* *16*, 1040–1054.
- (277) Li, S.; Gray, B. P.; McGuire, M. J.; Brown, K. C. (2011) Synthesis and biological evaluation of a peptide–paclitaxel conjugate which targets the integrin $\alpha_v\beta_6$. *Bioorg. Med. Chem.* *19*, 5480–5489.
- (278) Marshall, C. J.; Grosskopf, V. A.; Moehling, T. J.; Tillotson, B. J.; Wiepz, G. J.; Abbott, N. L.; Raines, R. T.; Shusta, E. V. (2015) An evolved Mxe GyrA intein for enhanced production of fusion proteins. *ACS Chem. Biol.* *10*, 527–538.
- (279) Sydor, J. R.; Mariano, M.; Sideris, S.; Nock, S. (2002) Establishment of intein-mediated protein ligation under denaturing conditions: C-Terminal labeling of a single-chain antibody for biochip screening. *Bioconjugate Chem.* *13*, 707–712.
- (280) Lin, P.-C.; Ueng, S.-H.; Tseng, M.-C.; Ko, J.-L.; Huang, K.-T.; Yu, S.-C.; Adak, A. K.; Chen, Y.-J.; Lin, C.-C. (2006) Site-specific protein modification through Cu(I)-catalyzed 1,2,3-triazole formation and Its implementation in protein microarray fabrication. *Angew. Chem. Int. Ed.* *118*, 4392–4396.
- (281) Albertsen, L.; Shaw, A. C.; Norrild, J. C.; Strømgaard, K. (2013) Recombinant production of peptide C-terminal α -amides using an engineered intein. *Bioconjugate Chem.* *24*, 1883–1894.
- (282) Boder, E. T.; Midelfort, K. S.; Wittrup, K. D. (2000) Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. *Proc. Natl. Acad. Sci. USA* *97*, 10701–10705.
- (283) Wang, X. X.; Shusta, E. V. (2005) The use of scFv-displaying yeast in mammalian cell surface selections. *J. Immunol. Methods* *304*, 30–42.

- (284) Shindo, A.; Maki, T.; Mandeville, E. T.; Liang, A. C.; Egawa, N.; Itoh, K.; Itoh, N.; Borlongan, M.; Holder, J. C.; Chuang, T. T.; McNeish, J. D.; Tomimoto, H.; Lok, J.; Lo, E. H.; Arai, K. (2016) Astrocyte-derived pentraxin 3 supports blood–brain barrier integrity under acute phase of stroke. *Stroke* 47, 1094–1100.
- (285) Zhang, X.; Wang, X. X.; Shusta, E. V. (2014) Creation and evaluation of a single-chain antibody tetramer that targets brain endothelial cells. *AIChE J.* 60, 1245–1252.
- (286) Tillotson, B. J.; Cho, Y. K.; Shusta, E. V. (2013) Cells and cell lysates: A direct approach for engineering antibodies against membrane proteins using yeast surface display. *Methods* 60, 27–37.
- (287) Klibanov, A. M. (2001) Improving enzymes by using them in organic solvents. *Nature* 409, 241–246.
- (288) MacDonald, J. I.; Munch, H. K.; Moore, T.; Francis, M. B. (2015) One-step site-specific modification of native proteins with 2-pyridinecarboxaldehydes. *Nat. Chem. Biol.* 11, 326–331.
- (289) Lidholt, K.; Weinke, J. L.; Lugenwa, F. N.; Bame, K. J.; Cheifetz, S.; Massagúe, J.; Lindahl, U.; Esko, J. D. (1992) A single mutation affects both N-acetylglucosaminyltransferase and glucuronosyltransferase activities in a Chinese hamster ovary cell mutant defective in heparan sulfate biosynthesis. *Proc. Nat. Acad. Sci. USA* 89, 2267–2271.
- (290) Takahashi, K.; Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663–676.
- (291) Zhou, H.; Wu, S.; Joo, J. Y.; Zhu, S.; Han, D. W.; Lin, T.; Trauger, S.; Bien, G.; Yao, S.; Zhu, Y.; Siuzdak, G.; Schöler, H. R.; Duan, L.; Ding, S. (2009) Generation of induced pluripotent stem cells using recombinant proteins. *Cell Stem Cell* 4, 1–4.
- (292) Bondy-Demomy, J.; Garcia, B.; Strum, S.; Du, M.; Rollins, M. F.; Hidalgo-Reyes, Y.; Wiedenheft, B.; Maxwell, K. L.; Davidson, A. L. (2015) Multiple mechanisms of CRISPR–Cas inhibition by anti-CRISPR proteins. *Nature* 526, 136–139.
- (293) Green, M. M.; Loewenstein, P. M. P. (1988) Autonomous functional domains of chemically synthesized human immunodeficiency virus Tat trans-activator protein. *Cell* 55, 1179–1188.

- (294) Frankel, A. D.; Pabo, C. O. (1988) Cellular uptake of the tat protein from human immunodeficiency virus. *Cell* 55, 1189–1193.
- (295) Wender, P. A.; Mitchell, D. J.; Pattabiraman, K.; Pelkey, E. T.; Stainman, L.; Rothbard, J. B. (2000) The design, synthesis, and evaluation of molecules that enable or enhance cellular uptake: Peptoid molecular transporters. *Proc. Nat. Acad. Sci. USA* 97, 13003–13008.
- (296) Chen, L.; Wright, L. R.; Chen, C.-H.; Oliver, S. F.; Wender, P. A.; Mochly-Rosen, D. (2001) Molecular transporters for peptides: Delivery of a cardioprotective ePKC agonist peptide into cells and intact ischemic heart using a transport system, R7. *Chem. Biol.* 8, 1123–1129.
- (297) Dubikovskaya, E. A.; Thorne, S. H.; Pillow, T. H.; Contag, C. H.; Wender, P. A. (2008) Overcoming multidrug resistance of small-molecule therapeutics through conjugation with releasable octaarginine transporters. *Proc. Nat. Acad. Sci. USA* 105, 12128–12133.
- (298) McKinlay, C. J.; Vargas, J. R.; Blake, T. R.; Hardy, J. W.; Kanada, M.; Contag, C. H.; Wender, P. A.; Waymouth, R. M. (2016) Charge-altering releasable transporters (CARTs) for the delivery and release of mRNA in living animals. *Proc. Nat. Acad. Sci. USA* 114, E448–E456.
- (299) Duchardt, F.; Fotin-Mleczek, M.; Schwarz, H.; Fischer, R.; Brock, R. (2007) A comprehensive model for the cellular uptake of cationic cell-penetrating peptides. *Traffic* 8, 848–866.
- (300) Tünnemann, G.; Martin, R. M.; Haupt, S.; Patsch, C.; Edenhofer, F.; Cardoso, M. C. (2006) Cargo-dependent mode of uptake and bioavailability of TAT-containing proteins and peptides in living cells. *FASEB J.* 20, 1775–1784.
- (301) Rothbard, J. B.; Kreider, E.; VanDeusen, C. L.; Wright, L.; Wylie, B. L.; Wender, P. A. (2002) Arginine-rich molecular transporters for drug delivery: Role of backbone spacing in cellular uptake. *J. Med. Chem.* 45, 3612–3618.
- (302) Nagel, Y. A.; Raschle, P. S.; Wennemers, H. (2017) Effect of preorganized charge display on the cell penetrating properties of cationic peptides. *Angew. Chem. Int. Ed.* 56, 122–126.
- (303) Lättig-Tünnemann, G.; Prinz, M.; Hoffmann, D.; Behlke, J.; Palm-Apergi, C.; Morano, I.; Herce, H. D.; Cardoso, M. C. (2011) Backbone rigidity and static presentation of guanidinium groups increases cellular uptake of arginine-rich cell-penetrating peptides. *Nat. Commun.* 2, 453–464.

- (304) Jinek, M.; Chylinski, K.; Fonfara, I.; Hauer, M.; Doudna, J. A.; Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816–821.
- (305) Cong, L.; Ran, F. A.; Cox, D.; Lin, S.; Barretto, r.; Habib, N.; Hsu, P. D.; Wu, X.; Jiang, W.; Marraffini, L. A.; Zhang, F. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823.
- (306) Gaj, T.; Gersbach, C. A.; Barbas III, C. F. (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol.* 31, 397–405.
- (307) Yin, H.; Xue, W.; Chen, S.; Bogorad, R. L.; Benedetti, E.; Grompe, M.; Kotliansky, V.; Sharp, P. A.; Jacks, T.; Anderson, D. G. (2014) Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nat. Biotechnol.* 32, 551–553.
- (308) Schumann, K.; Lin, S.; Boyer, E.; Simeonov, D. R.; Subramaniam, M.; Gate, R. E.; Haliburton, G. E.; Ye, C. J.; Bluestone, J. A.; Doudna, J. A.; Marson, A. (2015) Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. *Proc. Nat. Acad. Sci. USA* 112, 10437–10442.
- (309) Lv, H.; Zhang, S.; Wang, B.; Cui, S.; Yan, J. (2006) Toxicity of cationic lipids and cationic polymers in gene delivery. *J. Control. Release* 114, 100–109.
- (310) Chari, R.; Mali, P.; Moosburner, M.; Church, G. M. (2015) Unraveling CRISPR–Cas9 genome engineering parameters via a library-on-library approach. *Nat. Methods* 12, 823–826.
- (311) Jinek, M.; Jiang, F.; Taylor, D. W.; Sternberg, S. H.; Kaya, E.; Ma, E.; Anders, C.; Hauer, M.; Zhou, K.; Lin, S.; Kaplan, M.; Iavarone, A. T.; Charpentier, E.; Nogales, E.; Doudna, J. A. (2014) Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* 343, 1247997.
- (312) Kumler, W. D.; Eller, J. J. (1943) The acid strength of mono and diesters of phosphoric acid. The N-alkyl esters from methyl to butyl, the esters of biological importance, and the natural guanidine phosphoric acids. *J. Am. Chem. Soc.* 64, 2355–2361.
- (313) Wright, A. V.; Sternberg, S. H.; Taylor, D. W.; Staahl, B. T.; Bardales, J. A.; Kornfield, J. E.; Doudna, J. A. (2015) Rational design of a split-Cas9 enzyme complex. *Proc. Nat. Acad. Sci. USA* 112, 2984–2989.

- (314) Bokman, S. H.; Ward, W. W. (1981) Renaturation of aequorea green-fluorescent protein. *Biochem. Biophys. Acta* 101, 1372–1380.
- (315) Corish, P.; Tyler-Smith, C. (1999) Attenuation of green fluorescent protein half-life in mammalian cells. *Protein Eng.* 12, 1035–1040.
- (316) Shoulders, M. D.; Raines, R. T. (2009) Collagen structure and stability. *Annu. Rev. Biochem.* 78, 929–958.
- (317) Chattopadhyay, S.; Murphy, C. J.; McAnulty, J. F.; Raines, R. T. (2012) Peptides that anneal to natural collagen *in vitro* and *ex vivo*. *Org. Biomol. Chem.* 10, 5892–5897.
- (318) Provenzano, P. P.; Inman, D. R.; Elicieri, K. W.; Knittel, J. G.; Yan, L.; Rueden, C. T.; White, J. G.; Keely, P. J. (2008) Collagen density promotes mammary tumor initiation and progression. *BMC Med.* 6.
- (319) Provenzano, P. P.; Elicieri, K. W.; Campbell, J. M.; Inman, D. R.; White, J. G.; Keely, P. J. (2006) Collagen reorganization at the tumor-stromal interface facilitates local invasion. *BMC Med.* 4.
- (320) Conklin, M. W.; Eickhoff, J. C.; Riching, K. M.; Pehlke, C. A.; Elicieri, K. W.; Provenzano, P. P.; Friedl, A.; Keely, P. J. (2011) Aligned collagen is a prognostic signature for survival in human breast carcinoma. *Am. J. Pathol.* 178, 1221–1232.
- (321) Elmore, J. G.; Barton, M. B.; Mocerri, V. M.; Polk, S.; Arena, P. J.; Fletcher, S. W. (1998) Ten-year risk of false positive screening mammograms and clinical breast examinations. *N. Engl. J. Med.* 338, 1089–1096.
- (322) Hodges, J. A.; Raines, R. T. (2005) Stereoelectronic and steric effects in the collagen triple helix: Toward a code for strand association. *J. Am. Chem. Soc.* 127, 15923–15932.
- (323) Emmanuele, V.; López, L. C.; Berardo, A.; Naini, A.; Tadesse, S.; Wen, B.; D'Agostino, E.; Solomon, M.; DiMauro, S.; Quinzii, C.; Hirano, M. (2012) Heterogeneity of coenzyme Q10 deficiency. *Arch. Neurol.* 69, 978–983.
- (324) Tran, U. C.; Clarke, C. F. (2007) Endogenous synthesis of coenzyme Q in eukaryotes. *Mitochondrion* 7, S62–S71.

- (325) García-Corzo, L.; Luna-Sánchez, M.; Doerrier, C.; García, J.; Guarás, A.; Acín-Pérez, R.; Bullejos-Peregrín, J.; López, A.; Escames, G.; Acuña-Castroviejo, J. A. E.; López, L. C. (2013) Dysfunctional Coq9 protein causes predominant encephalomyopathy associated with CoQ deficiency. *Hum. Mol. Gen.* 22, 1233–1248.
- (326) Lohman, D. C.; Forouhar, F.; Beebe, E. T.; Stefely, M. S.; Minogue, C. E.; Ulbrich, A.; Stefely, J. A.; Sukumar, S.; Luna-Sánchez, M.; Jochem, A.; Lew, S.; Seetharaman, J.; Xiao, R.; Wang, H.; Westphall, M. S.; Wrobel, R. L.; Everett, J. K.; Mitchell, J. C.; López, L. C.; Coon, J. J.; Tong, L.; Pagliarini, D. J. (2014) Mitochondrial COQ9 is a lipid-binding enzyme that associates with COQ7 to enable coenzyme Q biosynthesis. *Proc. Natl. Acad. Sci. USA* 111, E4697–E4705.
- (327) van der Klei, A.; de Jong, R. L. P.; Lugtenburg, J.; Tielens, A. G. M. (2002) Synthesis and spectroscopic characterization of [1' -¹⁴C] ubiquinone-2, [1' -¹⁴C] 5-demethoxy-5-hydroxyubiquinone, and [1' -¹⁴C]-5-demethoxyubiquinone-2. *Eur. J. Chem.* 2002, 3015-3023.
- (328) Connerth, M.; Tatsuta, T.; Haag, M.; Klecker, T.; Westermann, B.; Langer, T. (2012) Intramitochondrial transport of phosphatidic acid in yeast by a lipid transfer protein. *Science* 338, 815–818.
- (329) Wessel, D.; Flugge, U. I. (1984) A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal. Biochem.* 138, 141–143.