

**CHEMICAL BIOLOGY APPROACHES TO STUDY ANTIBODY-DRUG
CONJUGATE SYNTHESIS, GENETIC CODE EXPANSION AND MAIN
PROTEASE OF SARS-COV-2**

A Dissertation

by

EROL CAN VATANSEVER

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Wenshe R. Liu
Committee Members,	Daniel Singleton
	Frank M. Raushel
	Tadhg P. Begley
Head of Department,	Simon W. North

August 2021

Major Subject: Chemistry

Copyright 2021 Erol Can Vatansever

ABSTRACT

Over the last two decades, the genetic code expansion technique became an important toolkit to study and bring new functionalities to proteins and peptides. Incorporating non-canonical amino acids that are native structures of post-translational modifications (PTMs) or PTM mimics enables researchers to study its function. One of the main limitations of this technique is the low incorporation rates of non-canonical amino acids into proteins. A way to circumvent this problem is by developing efficient aminoacyl-tRNA synthetase (aaRS) non-canonical amino acid pairs. In chapter 3, studies on introducing mutations to *Methanosarcina mazei* pyrrolysyl-tRNA synthetase (*Mm* PylRS) enzyme and studying the mutant enzymes' substrate preference is presented. Our investigation started with testing L-homoarginine (HAr) incorporation using a previously reported mutant PylRS and its cognate tRNA^{Pyl}. To better understand the PylRS enzymatic activity, the mutations were introduced step-wise and the effects of these mutations on PylRS substrate preference were analyzed. During these studies, efficient incorporation of *ortho*-chloro-L-phenylalanine (*o*-ClF) into a mutant PylRS (*o*-ClFRS) is serendipitously discovered. *O*-ClFRS contains three mutations from the PylRS mutant that incorporates HAr. *O*-ClF incorporation was further studied by crystal structure of *o*-ClFRS with *o*-ClF. Next, extending the HAr incorporation system to incorporate methylated derivatives of HAr was aimed. However, the incorporation of HAr derivatives was inefficient due to an uncharacterized modification taking place *in vivo*.

In chapter 2, a kinetics study on ligand-accelerated copper-catalyzed azide-alkyne cycloaddition (CuAAC) reaction and its application to synthesize antibody-drug conjugates (ADCs) is presented. ADCs found widespread use due to offering lower side toxicity in cancer treatment. The current bioconjugation approaches have drawbacks such as plasma instability and conjugation efficiency. Researchers look for new conjugation methods, to prepare stable conjugates in a homogenous form. We started studying CuAAC reaction with a kinetic investigation by using fluorogenic dyes. Next, we adopted the CuAAC reaction to synthesize ADCs by using the Pertuzumab antibody. We tested the effectivity of the conjugated ADCs against cancer cell line MDA-MB-453. The synthesized ADCs performed similar to widely used cysteine-maleimide conjugation and showed lower side toxicity. Then, to improve the conjugation process, we developed a methodology to synthesize ADCs on-resin by immobilizing antibodies on Protein A Sepharose resin using copper-chelating azide modified drug molecule.

In chapters 4 and 5, studies on the main protease of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) are described. The turmoil created by SARS-CoV-2 prompted the study of the main protease (M^{pro}), one of the key enzymes in life cycle of SARS-CoV-2. The studies began by docking selected FDA/EMA (Food Drug Administration/European Medical Agency) drugs to the active site of M^{pro} . Next, the promising drugs were purchased and their IC_{50} values were determined. Then, three of the most promising drugs were tested in live-virus microneutralization assay. Bepridil (an antiginal medicine) showed complete eradication of SARS-CoV-2 replication in low-micromolar concentrations showing promise for animal studies. Next, guided by the

known SARS-CoV M^{pro} inhibitors, 9 aldehyde-based peptide structures were synthesized and their inhibition against M^{pro} was studied. Inhibition constants (K_i) at low-nanomolar level were obtained. Compound MPI8 eradicated SARS-CoV-2 replication in A549/ACE2 cells at 0.3 μ M concentration. Next, attention was turned to engineering the M^{pro} enzyme to improve its catalytic activity. An azide containing ncAA was incorporated into M^{pro} to achieve this goal. Then, two M^{pro} units were cross-linked using a di-cyclooctyne linker and locked the enzyme in its active dimer form. This process yielded a more active M^{pro} derivative that allows measuring inhibition constants at lower enzyme concentrations.

In the final chapter, the findings of the studies presented in this thesis are summarized and the future directions for these studies are discussed.

ACKNOWLEDGEMENTS

First, I would like to express my gratitude to my advisor Dr. Wenshe Liu for encouraging and inspiring me to become a better researcher. His broad knowledge across chemistry and biology enabled me to work on thrilling projects. I would also like to thank my committee members, Dr. Daniel Singleton, Dr. Frank Raushel, Dr. Tadhg Begley for their suggestions and guidance.

In Liu Group, I had the chance to work with Ph.D. students and post-docs who have different attitudes towards research, coming from different scientific backgrounds and cultures. I believe I learned a lot from all these differences. Dr. Zhipeng Wang and Dr. Yan-Jiun Lee inspired me with his strong work ethic at the beginning of my PhD. Dr. Jeffery Tharp set a great example on keeping good lab notebooks and doing research in systematically and elegantly. Dr. Kai S. Yang contributed to my research greatly with his X-ray crystallography expertise and his supportive personality. Besides, the names I mentioned above, I am grateful to all past and present Liu Group members for creating a positive work environment that allowed me to focus on my research.

I consider myself lucky to have worked with knowledgeable and helpful staff scientists of Texas A&M University Department of Chemistry. NMR facility staff scientists Dr. Doug Elliott, Dr. Greg Wylie mass spectrometry facility staff scientists Dr. Yohannes Rezenom and Dr. Klaudia Kocurek who always answered my questions with patience.

Finally, I would like to express my deepest gratitude to my parents Nurhan Vatansever and Ali Vatansever. Their support, encouragement and love helped me to cope with stressful times and stay optimistic.

CONTRIBUTORS AND FUNDING SOURCES

Faculty Committee Recognition: This work was conducted under the direction of Professor Wenshe Liu of the Department of Chemistry and supervised by a thesis committee consisting of Professors Daniel Singleton and Tadhg Begley of the Department of Chemistry and Professor Frank Raushel who holds joint appointment in the Department of Chemistry and the Department of Biochemistry and Biophysics.

Student/Collaborator Contributions: For the work described in Chapter 2, the synthesis of the antibody-drug conjugates were carried out in collaboration with Dr. Sally Ward in Texas A&M Health and Science Department. Antibodies used in the study were expressed and purified by Dr. Sally Ward's student Dr. Jeffrey Kang. Cancer cell line experiment shown in Figure 63 was done by Jeffrey Kang. For the work described in Chapter 3, the crystallization conditions and the crystal structure determination of the expressed and purified N346D, C348S, Y384F PyIRS (185-454) (*o*-CIFRS) protein was done by Dr. Kai S. Yang. The docking study in Chapter 4 was done by Dr. Wenshe R. Liu. IC₅₀ assays and M^{pro} expression and purification described in Chapter 4 are done in collaboration with Dr. Kai S. Yang. For the work in Chapter 5, compounds MPI1-MPI10 were synthesized by Xinyu Ma, Dr. Yugendar Reddy Alugubelli, Dr. Yuying Ma and Zachary Geng. His-SUMO-M^{pro} construct, His-SUMO-M^{pro}97TAG, His-SUMO-M^{pro}277TAG constructs were made by Dr. Kai S. Yang. Purification and expression of His-SUMO-M^{pro} was conducted together with Dr. Kai S. Yang and Lauren R. Blankenship. Crystal structure determination was done by Dr. Kai S. Yang. Synthesis of

the compounds 6mer-Cou-Sub, FRET-Sub1 were done by Kaci C. Kratch. Synthesis of FRET-Sub2 is done by Lauren R. Blankenship.

Funding Sources: NIH R01GM121584, NIH R01GM127575, Welch Foundation Grant A1715, University X Grant

TABLE OF CONTENTS

	Page
ABSTRACT	ii
ACKNOWLEDGEMENTS	v
CONTRIBUTORS AND FUNDING SOURCES.....	vii
TABLE OF CONTENTS	ix
LIST OF FIGURES.....	xii
LIST OF TABLES	xxii
CHAPTER I INTRODUCTION	1
1.1 Genetic Code Expansion	1
1.1.1 Using Genetic Code Expansion to Study Post-Translational Modifications	3
1.1.2 Using Genetic Code Expansion to Improve Functional Group Diversity of Phage-Displayed Peptide Libraries	5
1.1.3 Codon Reassignment for Non-canonical Amino Acid Incorporation	8
1.1.4 A New PylRS variant: Candidatus Methanomethylophilus alvus PylRS	11
1.2 Introduction to Antibody Structure and Function	14
1.2.1 Antibodies in Cancer Treatment.....	15
1.2.2 Structure and Mechanism of Action of ADCs	16
1.2.3 Conjugation Strategies	23
1.3 Main Protease of Severe Acute Respiratory Syndrome Coronavirus 2	30
1.3.1 Life Cycle of Severe Acute Respiratory Syndrome Coronavirus 2	30
1.3.2 M ^{pro} as a Drug Target	33
1.3.3 Structure of M ^{pro}	33
1.3.4 Is M ^{pro} a Chymotrypsin-like Protease or a Papain-like Protease?	34
1.3.5 Amino Acid Preference of M ^{pro} Substrates	35
1.3.6 Substrates Used in M ^{pro} Assays.....	36
1.3.7 Inhibitors of M ^{pro}	40
CHAPTER II OPTIMAL SYNTHESIS OF ANTIBODY-DRUG CONJUGATES	45
2.1 Introduction to CuAAC Reaction.....	45
2.2 Experimental	49
2.2.1 Kinetic Experiments	49
2.2.2 Protein Expression for Protein Kinetics	54
2.2.3 sfGFP134PrK Py-Az-Coumarin Reaction Kinetics	55

2.2.4 Organic Synthesis.....	57
2.2.5 Synthesis of ADCs	79
2.2.6 Cell viability assay with MDA-MB-453	82
2.3 Results & Discussion	83
2.3.1 Kinetics study.....	83
2.3.2 Synthesis of ADCs	94
2.4 Conclusion.....	102
CHAPTER III INCORPORATION OF L-HOMOARGININE AND <i>ORTHO</i> - CHLORO-L-PHENYLALANINE INTO PROTEINS	103
3.1 Introduction	103
3.2 Experimental	105
3.2.1 Protein Sequences.....	105
3.2.2 Primers used	106
3.2.3 Protein Expression & Purification.....	107
3.2.4 Organic Synthesis.....	113
3.3 Results & Discussion	122
3.3.1 Introducing Step-wise Mutations to Study HAr Incorporation.....	122
3.3.2 Organic Synthesis and Incorporation of MeHAr and DiMeHAr	129
3.3.3 Incorporation of <i>o</i> -ClF.....	133
3.4 Conclusion.....	139
CHAPTER IV DRUG REPURPOSING TO TARGET MAIN PROTEASE.....	140
4.1 Introduction	140
4.2 Experimental	142
4.2.1 Docking Study.....	142
4.2.2 Expression of sf-GFP-M ^{P_{TO}} -6xHis	142
4.2.3 Screening Assay	144
4.2.4 Half Maximal Inhibitory Concentration (IC ₅₀) Assays of Drugs.....	145
4.2.5 Live Virus-based Microneutralization Assay.....	145
4.3 Results & Discussion	148
4.3.1 Docking results.....	148
4.3.2 Inhibition Analysis of Selected Drugs.....	150
4.3.3 Live Virus-based Microneutralization Assay Analysis of Pimozide, Ebastine and Bepridil	157
4.3.4 Bepridil as a Potential Treatment Option.....	163
4.4 Conclusion.....	165
CHAPTER V COVALENT INHIBITION & ENGINEERING OF THE MAIN PROTEASE.....	166
5.1 Introduction	166
5.1.1 Aldehyde-based Inhibitors of the Main Protease	166

5.1.2 Challenges in Biochemical Characterization of Main Protease	167
5.2 Experimental	168
5.2.1 Expression & Purification of M ^{pro}	168
5.2.2 Half Maximal Inhibitory Concentration (IC ₅₀) Assays of MPI1-9, GC376, 11a	169
5.2.3 Introducing Non-canonical Amino Acids to M ^{pro}	171
5.2.4 Cross-linking K97p-AzF-M ^{pro} to Form Covalently Dimerized M ^{pro}	173
5.2.5 Biochemical Characterization of M ^{pro} Derivatives	173
5.2.6 Live virus-based microneutralization assay with MPI1-MPI9, GC376, 11a	175
5.3 Results & Discussion	176
5.3.1 Expression & Purification of M ^{pro}	176
5.3.2 Activity of different substrates with M ^{pro}	177
5.3.3 IC ₅₀ assays with MPI1-8, GC376, 11a	181
5.3.4 Crystal Structure Analysis of Interactions Between M ^{pro} and inhibitors	185
5.3.5 Live-virus Microneutralization Analysis of MPI1-9, GC376, 11a	187
5.3.6 Non-canonical Amino Acid Incorporation into M ^{pro}	188
5.3.7 Dimerization of K97p-Azf-M ^{pro} with DBCO and BCN Linkers	190
5.3.8 Testing varying concentrations of wt-M ^{pro} and Dimerized M ^{pro} Against Different Substrates	196
5.4 Conclusion.....	198
CHAPTER VI	199
REFERENCES.....	202
APPENDIX A SPECTRA AND CHROMATOGRAMS.....	215
APPENDIX B	231

LIST OF FIGURES

	Page
Figure 1 A) Summary of ncAA incorporation system B) Molecular mechanism of ncAA incorporation	2
Figure 2 A) NcAAs directly incorporated into POIs B) Installation of acetyllysine derivatives by Aznl C) Installation of Pho-Tyr D) Installation of MeK and Me ₂ K E) Installation of citrulline	5
Figure 3 A) Examples of ncAAs incorporated into phage B) Generation of cyclic peptide phage library with AcrK C) Displaying two ncAAs on phage	8
Figure 4 Alignment of <i>CMa</i> PylRS (cyan color, PDB ID: 6ezd) and <i>Mm</i> PylRS (green color, PDB ID: 2zim) active sites. Three active site residues are shown as sticks (<i>Mm</i> PylRS residues: N346, C348, Y384; <i>CMa</i> PylRS residues: N166, V168, Y206F).....	13
Figure 5 A) Light chain and heavy chain showed in red and green color respectively B) Regions of an antibody C) Disulfide bond patterns of IgG ₁₋₄	15
Figure 6 A) Structure of Brentuximab Vedotin B) Release of free MMAE toxin.....	17
Figure 7 A) Releasing antibody outside the cell with recycling endosome B) Antibody internalization through autophagy C) By-stander effect of toxin	18
Figure 8 A) Chemically Cleavable Linkers B) Enzymatically cleavable linkers	21
Figure 9 A) Synthesis of Trastuzumab emtansine by lysine-NHS conjugation B) Synthesis of brentuximab vedotin by cysteine-maleimide conjugation C) Conjugation with self-hydrolyzing maleimides D) Conjugation by rebridging disulfide bonds.	26
Figure 10 A) CuAAC conjugation B) Tetrazine ligation C) Cu-free conjugation D) Keto-oxime ligation.....	28
Figure 11 A) Transglutaminase Conjugation B) Sortase A Conjugation.....	29
Figure 12 A) Structural and non-structural proteins encoded by SARS-CoV-2 genome B) Cleavage sites for PL ^{pro} and M ^{pro}	30
Figure 13 Life cycle of SARS-CoV-2.....	32

Figure 14 Crystal structure of M ^{Pro} (PDB ID: 6y2e) A) Dimer form M ^{Pro} units are shown in red and green color B) Catalytic residues of M ^{Pro}	34
Figure 15 A) General base mechanism B) Ion-pair mechanism C) Ion-pair formation upon substrate binding	35
Figure 16 A) Residue nomenclature B) Most preferred substrates for each position	36
Figure 17 Reporters used with M ^{Pro} substrates.....	40
Figure 18 Reversible covalent inhibitors A) Ketone inhibitor B) Nitrile inhibitor C) Aldehyde inhibitor D) Bisulfite adduct inhibitor E) Active ester inhibitor	42
Figure 19 Irreversible covalent inhibitors A) Chloroacetamide inhibitor B) Michael acceptor inhibitor C) Epoxy ketone inhibitor D) Phthalic hydrazide leaving group containing inhibitor	43
Figure 20 Examples for non-covalent M ^{Pro} inhibitors.....	44
Figure 21 A) CuAAC Reaction B) Examples of copper-binding ligands used in CuAAC	46
Figure 22 CuAAC mechanism proposed by Bertrand	47
Figure 23 SDS-PAGE analysis of sfGFP134-PrK	55
Figure 24 sfGFP134PrK Py-Az-Coumarin (34) Reaction	55
Figure 25 Synthesis of NHS ester of chloroacetic acid (5)	57
Figure 26 2-azidoethyl-1-amine hydrochloride.....	57
Figure 27 Synthesis of N-(2-azidoethyl)-2-chloroacetamide (6)	58
Figure 28 N-(2-azidoethyl)-2-iodoacetamide (7).....	59
Figure 29 Synthesis of N-(Prop)-L-valine (9).....	60
Figure 30 Synthesis of NHS ester of N-(Prop)-L-valine (10).....	61
Figure 31 Synthesis of Proc-Val-Cit-OH (11)	62
Figure 32 Synthesis of Proc-Val-Cit-PABOH (12).....	63
Figure 33 Synthesis of Proc-Val-Cit-PAB-PNP (13).....	63

Figure 34 Synthesis of Proc-Val-Cit-PAB-MMAE (14).....	64
Figure 35 Synthesis of tert-butyl prop-2-yn-1-yl propane-1,3-diyl dicarbamate (16)	65
Figure 36 Synthesis of prop-2-yn-1-yl (3-aminopropyl)carbamate hydrochloride (17) ..	66
Figure 37 Synthesis of prop-2-yn-1-yl (3-(2-bromoacetamido)propyl)carbamate (Bromoacetyl alkyne) (18).....	66
Figure 38 Synthesis of 6-(hydroxymethyl)nicotinic acid methyl ester (20)	67
Figure 39 Synthesis of 6-(azidomethyl)nicotinic acid methyl ester (21)	68
Figure 40 Synthesis of 6-(azidomethyl)nicotinic acid (22).....	69
Figure 41 Synthesis of 6-(azidomethyl)nicotinate hydroxysuccinimide ester (23)	70
Figure 42 Synthesis of 2,5-dioxopyrrolidin-1-yl (tert-butoxycarbonyl)-L-valinate (25).	71
Figure 43 Synthesis of Boc-Valine-Citrulline-OH (26).....	72
Figure 44 Synthesis of Boc-Valine-Citrulline-PABOH (27)	73
Figure 45 Synthesis of TFA salt of H ₂ N-Valine-Citrulline-PABOH (28)	74
Figure 46 Synthesis of Py-Az-Val-Cit-PABOH (29).....	75
Figure 47 Synthesis of Py-Az-Val-Cit-PAB-PNP Carbonate (30)	76
Figure 48 Synthesis of Py-Az-Val-Cit-PAB-MMAE (31).....	77
Figure 49 Synthesis of 7-((2-aminoethyl)amino)-4-methyl-2H-chromen-2-one hydrobromide (33).....	77
Figure 50 Synthesis of 6-(azidomethyl)-N-(2-((4-methyl-2-oxo-2H-chromen-7- yl)amino)ethyl)nicotinamide (34).....	78
Figure 51 Assay used to track CuAAC reaction trajectory	84
Figure 52 Reactions of AzCou with excessive alkyne concentrations. A) The one- phase exponential increase of product fluorescence when a limiting concentration of AzCou (50 μ M) was reacted with excess concentrations of 2-propyn-1-ol (0.25–4 mM). (B–F) The one-phase exponential increase of product fluorescence when a limiting concentration of AzCou (10 μ M) was reacted with excess concentrations of five phenylacetylene derivatives (0.1– 1 mM). (G) Average apparent reaction rate constants for all tested alkynes. ..	86

Figure 53 A) Limiting EtCou with varying excess concentrations of 2-azidoethylamine B) Limiting 2-azidoethylamine with varying excess concentrations EtCou.....	86
Figure 54 A) Limiting EtCou concentrations with excess PyAz concentration B) Excess EtCou concentrations reacted with PyAz	87
Figure 55 A) Reaction trajectories of varying copper concentrations (10-30 μ M) B) Graph of k_{obs} vs copper concentrations C) $\ln(k_{obs})$ vs $\ln(\text{Cu})$	89
Figure 56 A) Alkynes used in this assay B) Reaction trajectories of 20 μ M of AzCou was used with 400 μ M of alkyne derivative C) Normalized intensity of the reaction trajectories.....	90
Figure 57 A) Reaction trajectories of the reaction between ethoxyacetylene and AzCou B) Normalized intensities of the reaction trajectories	91
Figure 58 Precipitate formed in the reaction between ethoxyacetylene and AzCou.....	92
Figure 59 A) Azide structures B) Reaction trajectories of different azides with EtCou..	93
Figure 60 A) FRET analysis scheme for on-protein kinetics B) Reaction trajectories with varying concentrations of 34 C) k_{obs} vs varying concentrations of 34 ...	94
Figure 61 ADC Conjugation Strategy	94
Figure 62 Synthesis of Linker toxin compound 14	96
Figure 63 Cell-based assay of ADC with MDA-MB-453. P WT (A), P YS (B), P HE (C) are the pertuzmab derivatives with different cell-penetration properties Hulys10 (D) is the control group mentioned in reference [132]. The data points shown with (Click) are the ADCs conjugated according to Figure 61. The ADCs generated by classical maleimide conjugation is shown in Figure 9B with (Maleimide).....	98
Figure 64 ADC synthesis with chelating-azide linker-toxin 31	98
Figure 65 Synthesis scheme of linker-toxin compound 31	99
Figure 66 Synthesizing Per-2C-31 by resin method	101
Figure 67 A) Structures of L-Arginine, MMA, SDMA, ADMA, HAr, MeHAr, DiMeHAr B) Demethylase activity on MMA or ADMA residues C) Demethyliminase activity on MMA or ADMA residues	104

Figure 68 A) SDS-PAGE Analysis of purified sfGFP134 with HAr and sfGFP with MeHAr B) SDS-PAGE analysis of sfGFP134 with DiMeHAr and sfGFP134 with HAr C) ESI-MS analysis of sfGFP134 with HAr (calculated: 27882.42 Da observed: 27881.65 Da) D) ESI-MS analysis of sfGFP134 with MeHAr (calculated: 27896.43 Da observed: 27881.77 Da, 27894.89 Da) E) ESI-MS analysis of sfGFP134 with DiMeHAr (calculated: 27910.45 Da observed: 27750.78 Da, 27881.85 Da)	109
Figure 69 A) SDS-PAGE analysis of sfGFP134- <i>o</i> -ClF B) ESI-MS analysis of <i>o</i> -ClF incorporated sfGFP	110
Figure 70 Purification of N346D, C348S, Y384F PylRS (185-454). Size-exclusion column purification chromatogram and analysis of collected fractions with SDS-PAGE analysis	111
Figure 71 Synthesis of N α -Boc-N ϵ -Cbz-L-Lysine-OtBu Ester (33).....	113
Figure 72 Synthesis of N α -Boc-L-Lysine-OtBu Ester (34).....	113
Figure 73 Synthesis of N α -Boc-N ϵ -CN-L-Lysine-OtBu Ester (35).....	114
Figure 74 Synthesis of N α -Boc-N θ -Methyl-L-Homoarginine-OtBu Ester (36)	115
Figure 75 Synthesis of N θ -Methyl-L-Homoarginine HCl salt (37).....	116
Figure 76 Synthesis of N1-Methyl-N1-Boc-N2-Boc-Thiourea (39).....	116
Figure 77 Synthesis of N α -Boc-N θ -Methyl-N θ ,N θ '-Boc-L-Homoarginine-OtBu Ester (40).....	117
Figure 78 Synthesis of Methyl Homoarginine (MeHAr) (37)	118
Figure 79 Synthesis of O-benzyl thiocyanate (42).....	118
Figure 80 Synthesis of N α -Boc-N ϵ -Thioamide-N θ -Cbz-L-Homoarginine-OtBu Ester (43).....	119
Figure 81 Synthesis of N α -Boc-N θ -Cbz-N θ ,N θ '-Methyl-L-Homoarginine-OtBu Ester (44).....	120
Figure 82 Synthesis of Dimethyl homoarginine (DiMeHAr) (45).....	121
Figure 83 A) Interaction of Asn346 with Pyl side chain (PDB ID: 2ZCE). Oxygen atoms are shown red, nitrogen atoms are shown in blue. B) Structures of Pyl and HAr	123

Figure 84 GFP reporter assay to quantify ncAA incorporation	124
Figure 85 A) Structures of the amino acids tested for incorporation. B) Results of the sfGFP reporter assay. BockK and wt-PylRS are used as positive control. The expressions which contained inducer but no supplemented amino acid serves as the negative control.	125
Figure 86 A) Amino acids used B) Expression levels of sfGFP	125
Figure 87 A) Amino acids used B) Expression levels of sfGFP	126
Figure 88 A) Amino acids used B) Expression levels of sfGFP	127
Figure 89 A) Amino acids used B) Cell pellets visualized under blue light to observe sfGFP C) Quantified expression test	128
Figure 90 A) <i>Mm</i> PylRS (PDB ID: 2Q7H) shown in green, <i>E. coli</i> PheRS (PDB ID: 3PCO) shown in cyan color B) Asn346 of <i>Mm</i> PylRS and Asp117 of <i>E. coli</i> PheRS demonstrated. Nitrogen shown in blue color, oxygen shown in red color.	129
Figure 91 A) First route to synthesis of MeHAr 37 B) Second route to synthesis of MeHAr (37) C) Synthesis of DiMeHAr 45	130
Figure 92 A) Amino acids used B) Expression levels of sfGFP	131
Figure 93 A) Amino acids used B) Expression levels of sfGFP	133
Figure 94 A) Amino acids used B) Expression levels of sfGFP	134
Figure 95 A) Amino acids used B) Expression levels of sfGFP	134
Figure 96 A) Amino acids used B) Expression levels of sfGFP	135
Figure 97 A) Amino acids used B) Expression levels of sfGFP	136
Figure 98 A) 47 as a tool for peptide cyclization B) ESI-MS evidence for cyclization.	136
Figure 99 A) Amino acids used B) Expression levels of sfGFP	137
Figure 100 A) AMP-PNP and <i>o</i> -ClF in sticks, PylRS shown in wheat color B) Close contacts of <i>o</i> -ClF group	138
Figure 101 Structures of 29 FDA/EMA-approved medicines and rupintrivir whose IC ₅₀ values in inhibiting M ^{pro} were determined in the study.....	149

Figure 102 Comparison of assay buffers containing 20% DMSO and 1% DMSO	150
Figure 103 Initial screening of M ^{pro} inhibition by 29 FDA/EMA-approved medicines and rupintrivir	152
Figure 104 IC ₅₀ curves of 18 medicines and rupintrivir.....	154
Figure 105 Pimozide (A), ebastine (B), bepridil (C), and their overlay (D) in the active site of M ^{pro} . The protein surface topography in A, B, and C is presented to show the concaved active site	156
Figure 106 Microscope-recorded cytopathogenic effect (CPE) observation in Vero E6 cells that were infected by SARS-CoV-2 and grown in the presence of different concentration of bepridil or 0.1% DMSO as a positive control.....	160
Figure 107 Microscope-recorded cytopathogenic effect (CPE) observation in A549/ACE2 cells that were infected by SARS-CoV-2 and grown in the presence of different concentration of bepridil or 0.1% DMSO as a positive control	161
Figure 108 Viability of Vero E6 and A549/ACE2 cells at different concentrations of bepridil.....	162
Figure 109 The SARS-CoV-2 inhibition by bepridil in (A) Vero E6 and (B) A549/ACE2 cells. Cells were incubated with different concentrations of bepridil and then infected with 0.5 MOI (multiplicity of infection) of SARS-CoV-2.	163
Figure 110 Structures of the substrates used for M ^{pro} activity assays A) 6mer-Cou-Sub B) VLQ-Cou-Sub C) FRET-Sub1 D) FRET-Sub2 E)FRET-Sub3.....	177
Figure 111 A) Activity of 6mer-Cou-Sub at 1 μM and 50 nM M ^{pro} B) Comparison of the activities of FRET-Sub1 and FRET-Sub3 C) Comparison of activities of 6mer-Cou-Sub and VLQ-Cou-Sub D) Comparison of activities of FRET-Sub1, FRET-Sub2, VLQ-Cou-Sub	178
Figure 112 A) Linear regression analysis of the initial slopes B) K _M calculation for M ^{pro} FRET-Sub3 enzyme substrate pair	180
Figure 113 A) Structures of the inhibitors used in this study B) Inhibition curves of the inhibitors C) Summary of IC ₅₀ and K _i values	183
Figure 114 A) The packing of apo-M ^{pro} in its crystals. An asymmetric unit monomer in the center is colored in red. Its active site is presented as a concave surface. Another monomer that stacks on the active site of the monomer is	

colored in blue. B) A contoured $2F_o - F_c$ map at the 1σ level around MPI3 and C145 in the active site of M^{pro} . A covalent bond between MPI3 and C145 is observable. C) Structure overlay between apo- M^{pro} and the M^{pro} -MPI3 complex. A black arrow points to a region that undergoes structure rearrangement from the apoenzyme to accommodate MPI3 in the M^{pro} -MPI3 complex. D) Occupation of the active-site cavity of M^{pro} by MPI3. The enzyme is shown in its surface presentation mode. E) Extensive hydrogen bonding and van der Waals interactions between M^{pro} and MPI3. The backbone of M^{pro} is colored in blue and side chain carbon atoms in orange. Hydrogen bonds between MPI3 and M^{pro} are depicted as yellow dashed lines. F) Overlay of seven Opal-based inhibitors at the active site of M^{pro} . Color coded for the names is shown in the figure. All images were made by using the program PyMOL. The PDB IDs for M^{pro} in its apo form and complexes with inhibitors are 7JPY (apo), 7JPZ (MPI1), 7JQ0 (MPI3), 7JQ1 (MPI4), 7JQ2 (MPI5), 7JQ3 (MPI6), 7JQ4 (MPI7), and 7JQ5 (MPI8)..... 186

Figure 115 A) AcK, KetoK, PrK, BocK incorporation into His-SUMO- M^{pro} -277TAG. Expected molecular weight is 45 kDa. B) AllocK, ProcK, BocK, *p*-AzF incorporation into His-SUMO- M^{pro} -97TAG Expected molecular weight is 45 kDa. C) AllocK, ProcK, BocK, *p*-AzF incorporation into His-SUMO- M^{pro} -277TAG. Expected molecular weight is 45 kDa. D) NcAAs tried for incorporation E) Purified K97*p*-AzfMpro. Expected molecular weight is 34 kDa..... 190

Figure 116 A) Structures of DBCO linker and BCN linker B) Dimerization of K97*p*-Azf- M^{pro} with DBCO linker to generate DBCO-(M^{pro})₂ C) Dimerization of K97*p*-Azf- M^{pro} with BCN linker BCN-(M^{pro})₂ 192

Figure 117 A) ESI-MS analysis of K97*p*-Azf- M^{pro} B) ESI-MS analysis of DBCO linker reaction after first step C) ESI-MS analysis of DBCO linker reaction after second step D) ESI-MS analysis of BCN linker reaction E) SDS-PAGE analysis of DBCO linker reaction after second step F) SDS-PAGE analysis of BCN linker reaction..... 192

Figure 118 A) Comparison of wt- M^{pro} , KAzf-Mpro, DBCO-(M^{pro})₂ B) Comparison of DBCO-(M^{pro})₂ and BCN-(M^{pro})₂ C) Comparison of wt- M^{pro} and 5 nM BCN-(M^{pro})₂ D) Activity of DBCO-(M^{pro})₂ at various concentrations 194

Figure 119 A) Structure of the inhibitor tested B) Inhibition analysis with 50 nM M^{pro} C) Inhibition analysis with 20 nM M^{pro} Inhibition analysis with 5 nM M^{pro} BCN-(M^{pro})₂..... 195

Figure 120 A) Activity of various concentrations of wt- M^{pro} against 10 μ M of FRET-Sub3 B) Activity of various concentrations of wt- M^{pro} against 10 μ M of

6mer-Cou-Sub C) Activity of various concentrations of wt-M ^{Pro} against 10 μM of VLQ-Cou-Sub D) Activity of various concentrations of DBCO- (M ^{Pro}) ₂ against 10 μM of FRET-Sub3	197
Figure 121 ¹ H NMR spectrum of N-(Prop)-L-valine (9)	216
Figure 122 ¹ H NMR spectrum of N-(Prop)-L-valine-OSu ester (10)	216
Figure 123 ¹ H NMR spectrum of Proc-Val-Cit-OH (11).....	217
Figure 124 ¹ H NMR spectrum of Proc-Val-Cit-PAB-PNP (13).....	217
Figure 125 HPLC chromatogram of Proc-Val-Cit-PAB-MMAE (14)	218
Figure 126 ¹ H NMR spectrum of tert-butyl prop-2-yn-1-yl propane-1,3- diylidicarbamate (16).....	218
Figure 127 ¹³ C NMR spectrum of tert-butyl prop-2-yn-1-yl propane-1,3- diylidicarbamate (16).....	219
Figure 128 ¹ H NMR spectrum of prop-2-yn-1-yl (3-aminopropyl)carbamate hydrochloride (17)	219
Figure 129 ¹ H NMR spectrum of bromoacetyl alkyne (18).....	220
Figure 130 ¹ H NMR spectrum of Py-Az-Val-Cit-PABOH (29)	220
Figure 131 ¹ H NMR spectrum of Py-Az-Val-Cit-PAB-PNP carbonate (30).....	221
Figure 132 LC-MS chromatogram of Py-Az-Val-Cit-PAB-MMAE (31).....	221
Figure 133 ¹ H NMR spectrum of 6-(azidomethyl)-N-(2-((4-methyl-2-oxo-2H- chromen-7-yl)amino)ethyl)nicotinamide (34)	222
Figure 134 LC-MS chromatogram of conjugated Per-4C-31	222
Figure 135 Deconvolution results corresponding to peak at 15.98 min (light chain that was conjugated with 7 only) in Figure 134	223
Figure 136 Deconvolution results corresponding to peak at 18.98 min (light chain reacted with 7 and 14) in Figure 134.....	223
Figure 137 LC-MS chromatogram of Per2C-31 before conjugation	224
Figure 138 Deconvolution of the signal at peak 27.33 min in Figure 137.....	224

Figure 139 LC-MS chromatogram of Per2C-31 after conjugation with 31	225
Figure 140 Deconvolution of mass signals at peak 29.51 mins in Figure 139.....	225
Figure 141 ¹ H NMR spectrum of N _α -Boc-N _ε -CN-L-Lysine-OtBu Ester (35)	226
Figure 142 ¹³ C NMR spectrum of N _α -Boc-N _ε -CN-L-Lysine-OtBu Ester (35).....	226
Figure 143 N _α -Boc-N _θ -Methyl-L-Homoarginine-OtBu Ester (36).....	227
Figure 144 ¹ H NMR spectrum of methyl homoarginine (MeHAr) (37)	227
Figure 145 ¹ H NMR spectrum of N _α -Boc-N _θ -Methyl-N _θ ,N _θ '-Boc-L-Homoarginine- OtBu Ester (40).....	228
Figure 146 ¹ H NMR spectrum of N _α -Boc-N _ε -Thioamide-N _θ -Cbz-L-Homoarginine- OtBu Ester (43).....	228
Figure 147 ¹ H NMR spectrum of N _α -Boc-N _θ -Cbz-N _θ ,N _θ '-Methyl-L-Homoarginine- OtBu Ester (44).....	229
Figure 148 ¹³ C NMR spectrum of N _α -Boc-N _θ -Cbz-N _θ ,N _θ '-Methyl-L-Homoarginine- OtBu Ester (44).....	229
Figure 149 ¹ H NMR spectrum of dimethyl homoarginine (DiMeHAr) (45).....	230

LIST OF TABLES

	Page
Table 1 Properties of M ^{Pro} substrates	39
Table 2 Limiting 3-azido-7-hydroxy-coumarin (AzCou) and excess alkyne.....	50
Table 3 Limiting 7-Ethynylcoumarin (EtCou) and excess 2-azidoethylamine.....	50
Table 4 Limiting 2-azidoethylamine and excess EtCou.....	51
Table 5 Limiting AzPy concentration and excess EtCou concentrations	51
Table 6 EtCou and excess PyAz concentrations	52
Table 7 EtCou and excess PyAz concentrations	52
Table 8 AzCou and alkyne derivatives	52
Table 9 AzCou and different ethoxyacetylene concentrations.....	53
Table 10 Excess EtCou and different limiting azide concentrations	53
Table 11 The 2x stock solution of click reaction mixture (200 mL reaction mix)	80
Table 12 Docking results of small molecule medicines (Compounds whose IC ₅₀ values were tested are asterisked.).....	148
Table 13 IC ₅₀ values and Hill slopes of 18 medicines and rupintrivir	155
Table 14 SARS-CoV-2 induced CPE in (A) Vero E6 and (B) A549/ACE2 cells in the presence of bepridil.....	159
Table 15 Hill slopes of the tested inhibitors.....	185

CHAPTER I

INTRODUCTION*¹

1.1 Genetic Code Expansion

In a landmark study published in 2001,^[1] site-specific incorporation of non-canonical amino acids (ncAA) into proteins by `hijacking` the protein translation machinery in vivo via using an orthogonal tRNA/aaRS pair is demonstrated. As of 2019, this methodology allowed the site-specific installation of more than 150 different ncAAs into proteins.^[2]

In bacterial systems such as (*E. coli*), the most widely-used ncAA incorporation technique requires a two-plasmid system. In this system, one plasmid is constructed to express amino-acyl tRNA synthetase (aaRS) and tRNA pair and the other plasmid is constructed to express the protein of interest (Figure 1A). Site-specific ncAA incorporation mechanism follows the incorporation route similar to the other 22 canonical amino acids except for L-selenocysteine (L-selenocysteine incorporation requires a specific mRNA sequence and multiple enzymatic modifications). Similar to canonical amino acid incorporation, ncAA binds to its specific aaRS and its carboxylate group is modified by adenosine triphosphate (ATP) group to form an activated ester. Then, its cognate tRNA undergoes esterification reaction with activated amino acid ester

¹ Figure 7, Figure 12, Figure 13 are copied from publications DOI: 10.1186/s13046-017-0667-1, DOI: 10.1371/journal.ppat.1009013, DOI:10.1002/cbic.202000047 respectively.

to form the aminoacyl-tRNA (aa-tRNA). The aa-tRNA is recognized by the specific mRNA codon at the aminoacyl (A) site of the ribosome. Next, the amino group of the aa-tRNA reacts with the growing protein chain at the peptidyl (P) site (Figure 1B).

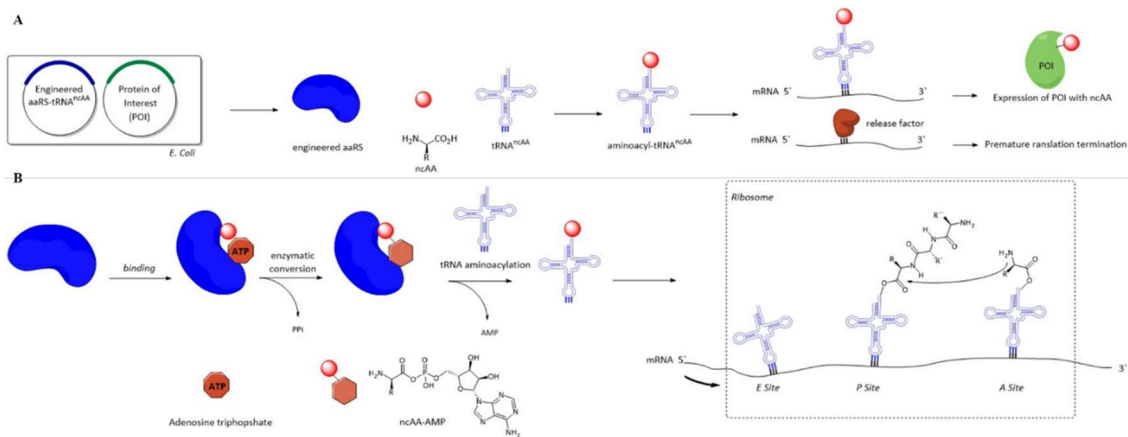


Figure 1 **A)** Summary of ncAA incorporation system **B)** Molecular mechanism of ncAA incorporation

Three main requirements need to be met to achieve site-specific ncAA incorporation system to work effectively. These are: an orthogonal aaRS that will only aminoacylate its cognate tRNA with the provided ncAA and will not aminoacylate endogenous tRNAs in the host organism, the ncAA should have significantly different structure than the canonical amino acids to prevent being recognized by the endogenous aaRS, the codon chosen for ncAA incorporation should be one of the codons that are less frequently used by cellular translation machinery.^[3]

Four orthogonal aaRS-tRNA pairs are frequently used. Tyrosyl-tRNA synthetase (TyrRS)/tRNA_{CUA} from *Methanococcus jannaschii* is orthogonal in bacteria but not in eukaryotic cells. Tyrosyl-tRNA synthetase (EcTyrRS)/tRNA_{CUA} pair, the *E. coli* Leucyl-

tRNA synthetase (*EcLeuRS*)/tRNA_{CUA} pairs are orthogonal in eukaryotic cells but not in bacteria. Pyrrolysyl-tRNA synthetase (*PylRS*)/tRNA_{CUA} pairs from and *Methanosarcina barkeri* (*Mb*), *Methanosarcina mazei* (*Mm*) and *Methanomethylophilus alvus* (*Ma*) are orthogonal in both bacteria and eukaryotic cells.^[4]

1.1.1 Using Genetic Code Expansion to Study Post-Translational Modifications

It is often considered “fascinating” that proteins can undertake a wide-variety of tasks with only 22 amino acid building blocks (most systems use only 20 amino acids). In order to achieve these complex tasks, certain amino acid residues can undergo functional group transformations following protein translation. This phenomenon is called post-translational modifications (PTMs). Some examples of PTMs include methylation of arginines and lysines, acylation of lysines, phosphorylation of tyrosines, hydroxylation of prolines and glycosylation of serines. PTMs often have significant regulatory roles such as turning gene expression, regulating protein-protein interactions, cellular signaling.^[5] Therefore, it is important to site-specifically install PTMs to study their functions. There are mainly four different approaches used to generate proteins with PTMs: using solid-phase peptide synthesis to do semi-synthesis or total synthesis of PTM installed proteins. They are generating PTMs by enzymatic reaction, using cysteine-based approaches to generate mimics or native structure and incorporating PTMs using ncAA incorporation technique. Each of these methods has its unique advantages and disadvantages and choosing which approach to use is decided on a case-by-case basis.^[6] In this section only examples for ncAA-based techniques are given. There have been two different approaches to install PTMs into proteins using the ncAA

incorporation technique: incorporating native structures directly and incorporating a precursor ncAA with a reactive group that can generate different PTMs Previously; phosphoserine (Phospho-S),^[7] 3-methylhistidine (3-MeH),^[8] acetyllysine (AcK),^[9] propionyllysine (PropK), butyryllysine (BuK), crotonyllysine (CrK),^[10] 2-hydroxyisobutyryllysine (2-HIButK)^[11] have been incorporated directly by using site-specific incorporation of ncAAs (Figure 2A).

For the second approach, azidonorleucine (AznL) was incorporated and converted to AcK and succinyllysine (SuK) by Staudinger ligation (Figure 2B).^[12] Phosphoramidate-protected tyrosine (Pha-Tyr) was incorporated into proteins to yield phosphotyrosine (Pho-Tyr) upon HCl deprotection (Figure 2C).^[13] Allylsine was generated by TCEP reduction from its precursor *N*^ε-(4-azidobenzoyl)- δ , ϵ -dehydrolysine (AcdK). Then, allylsine (AlK) underwent reductive amination incorporated to monomethyllysine (MeK) and dimethyllysine (Me₂K) (Figure 2D).^[14] In a more recent example, photocaging citrulline (PC-Cit) was incorporated into proteins which yielded citrulline installed protein (cit-POI) upon irradiation with ultraviolet light (Figure 2E) in mammalian cells.^[15]

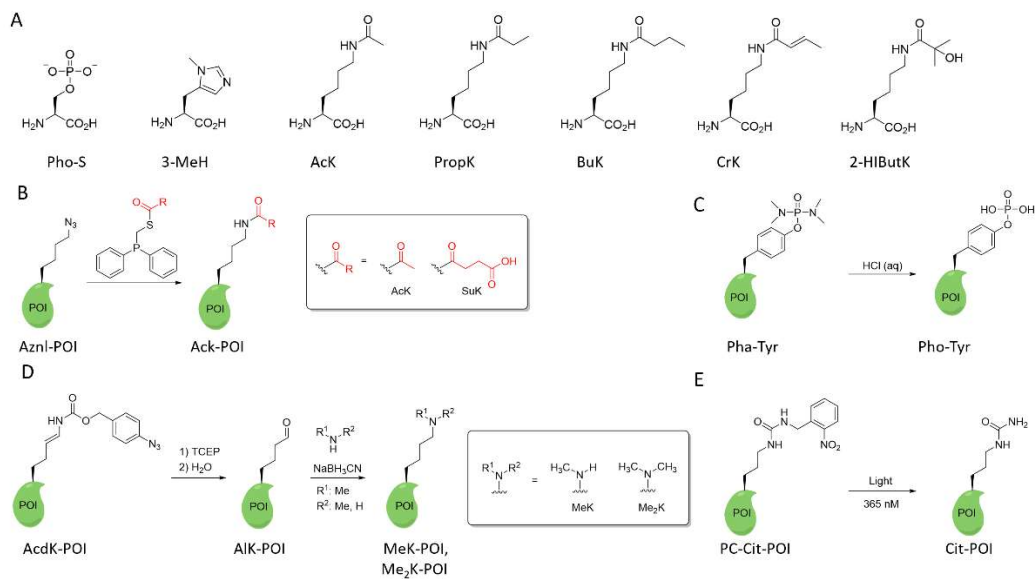


Figure 2 A) NcAAs directly incorporated into POIs B) Installation of acetyllysine derivatives by Aznl C) Installation of Pho-Tyr D) Installation of MeK and Me₂K E) Installation of citrulline

1.1.2 Using Genetic Code Expansion to Improve Functional Group Diversity of Phage-Displayed Peptide Libraries

A phage (bacteriophage) is a nucleic acid (DNA or RNA) encapsulated in a protein coat. It can infect bacteria and replicate by using bacteria as its host system. The amino acid sequence on the protein coat is encoded by the encapsulated nucleic acid of the bacteriophage. By changing the bacteriophage gene sequence, one can change the amino acids displayed on the phage's surface.

By shuffling the part of the nucleic acid sequence that codes for the surface protein, randomized amino acid sequences will be displayed on that region of the protein. This technique is called phage display and it allows rapid screening of peptide libraries with 10^9 - 10^{10} members. The phage-displayed peptide library can be screened against a protein of interest (POI) to identify the binding library members. After

incubating the phage mutants with the immobilized POI, the unbound phages were washed away. The binders are eluted, amplified and then the process is repeated to enrich binding phage mutants. Finally, the amino acid sequence of the binders is identified by gene sequencing.^[16]

This method found widespread use due to its convenience. However, a major limitation is the limited chemical diversity. Only 20 canonical amino acids can be displayed. The genetic code expansion technique is one way to increase the chemical diversity of the phage-displayed peptide libraries.

One means to alleviate the limited functional group diversity of canonical amino acids is to incorporate non-canonical amino acids. In addition to increasing chemical diversity, ncAAs can bring new chemical functionalities (such as metal-binding groups, reactive handles etc.). Some examples of ncAAs incorporated into the M13 phage are shown in Figure 3A. In a proof-of-concept study^[17] *p*-methoxyphenylalanine (*p*-OMeF), *p*-azidophenylalanine (*p*-AzF), *p*-acetylphenylalanine (*p*-AcF), *p*-benzoylphenylalanine (*p*-BzF) and *p*-naphthylalanine (*p*-NapA) were shown to be incorporated into the p3 protein of M13 bacteriophage. Azide functionality on *p*-AzF used covalently reacts with alkyne-containing dyes enabling labeling the phage with a fluorescent dye. *P*-AcF also has a ketone group that can react with active lysines. In a 2008 study,^[18] *p*-sulfotyrosine (*p*-SuY), *p*-boronophenylalanine (*p*-BoF), *p*-bipyridylalanine (*p*-BpyA) were incorporated into phage. These amino acids were chosen due to their characteristic functional groups. *P*-SuY is used to select against a known *p*-SuY binding HIV protein gp-120. *P*-BoF is chosen for its potency against glycosylated proteins and *p*-ByA has a metal-chelating

side chain. In another study,^[19] a TAG-obligate library was constructed and shown to incorporate *o*-chloro-L-phenylalanine (*o*-ClF), *m*-chloro-L-phenylalanine (*m*-ClF), *m*-bromo-L-phenylalanine (*m*-BrF), *m*-trifluoromethyl-L-phenylalanine (*m*-TriFF) and BuK. A known sirtuin 2 (SIRT2) inhibitor contains thiobutyryllysine (tBuK). BuK was used for phage selection to optimize SIRT2 inhibitors. Cyclic peptides are considered more potent binders with superior in vivo stability than their acyclic derivatives. Therefore, screening cyclic peptide phage libraries is a major direction in peptide-based drug discovery. An AcrK containing phage library was shown to cyclize with a neighboring cysteine residue effectively to yield a cyclic peptide phage library. This library was screened against histone deacetylase 8 (HDAC8) protein.^[20] Another study,^[21] showed incorporation of two ncAAs N^ε-[[(2-methylcycloprop-2-en-1-yl)methoxy]carbonyl]-L-lysine (CypK) and propargyltyrosine (PrpY) were incorporated into a displayed single-chain antibody variable fragment (ScFv) on p3 protein of phage. Given the utility of antibody phage display, incorporating ncAAs into antibody fragments displayed on phage surface proteins holds great potential.

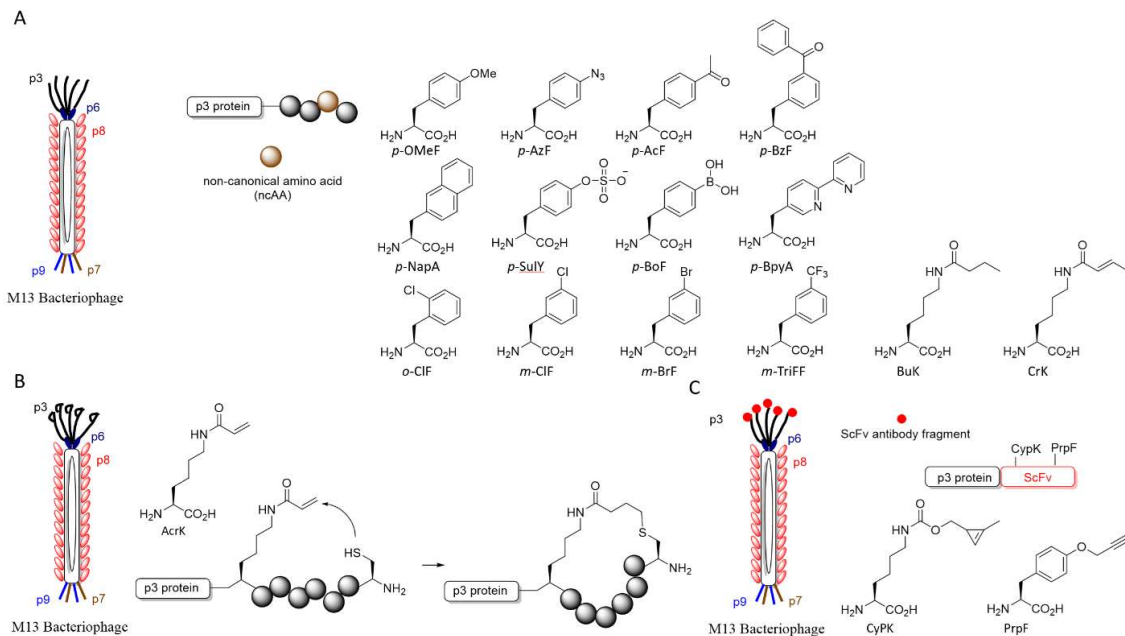


Figure 3 A) Examples of ncAAs incorporated into phage B) Generation of cyclic peptide phage library with AcrK C) Displaying two ncAAs on phage

1.1.3 Codon Reassignment for Non-canonical Amino Acid Incorporation

The four bases that make up the whole genome (A, T, G, C) can form 64 different codon combinations. In most of the organisms; 61, referred to as sense codons, are responsible for coding a canonical amino acid. Three of them are stop codons that do not code any amino acid and are, therefore, classified as nonsense codons. When one of these nonsense codons is reached, a release factor binds to a codon in a similar fashion to a tRNA and terminates the translation. Stop codons are preferred over sense codons for ncAA incorporation due to their lower abundance in the genome and it is easier to outcompete its endogenous competitors that are release factors. For the *E. coli* expression system, the TAG stop codon is the primary choice because, among three stop codons, TAG is the least abundant codon in the *E. coli* genome. To push TAG incorporation further, release factor 1 knockout *E. coli* strain^[22] and a TAG-free *E. coli*

strain^[23] have been generated to increase the ncAA incorporation system. However, having other codons available for ncAA incorporation becomes crucial when more than one ncAA incorporation is desired. Previously, TAG and TAA stop codons were used for incorporating two distinct amino acids in *E. coli* cells for GFP expression.^[24]

Another approach is to use rare sense codons for ncAA incorporation. One example of this is the AGG codon. AGG codes for L-arginine in *E. coli*. However, it has a low frequency (2.1 per thousand codons in a genome) making it the fourth least abundant codon after the three stop codons. It is shown that,^[25] in *E. coli* B-95.ΔA with ArgW and ArgU knockout cell strains, the AGG codon can be successfully suppressed with L-homoarginine (HAr) and L-N⁶-(1-iminoethyl)lysine quantitatively using aaRS/tRNA pair evolved from PylRS. In another study, tRNA^{CCU}Arg (argW) and the gene coding for the enzyme for biosynthesis of L-arginine from L-glutamine is also knocked out from *E. coli* DH10B genome.^[26] This strain is called *E. coli* BS02. By using this strain, *p*-propargyltyrosine (PrpY), *p*-azidophenylalanine (*p*-AzF), *p*-acetylphenylalanine (*p*-AcF), *p*-benzoylphenylalanine (*p*-BzF) were incorporated into endogenous Z-domain protein with their respective aaRS-tRNA pairs. The mass spectrometry analysis showed quantitative incorporation of the desired ncAAs with *E. coli* BS02 strain. By using the BL21-AI strain (without any arginine-related knock-out genes), Liu group reported 92% incorporation of N⁴-(allyloxycarbonyl)-lysine (AllocK) along with 8% L-arginine.^[27] The key to achieve this incorporation efficiency with BL21-AI strain is to induce the gene expression for aaRS/tRNA pair earlier than the reporter gene to allow aaRS/tRNA to reach significant cellular concentrations to

outcompete tRNA^{AccuArg}. NcAAs can be incorporated with 92% percent efficiency in AGG codons.

An alternative for using nonsense codons and rare sense codons is by using quadruplet codons. Quadruplet codons are rare but, are a naturally occurring phenomena. Endogenous tRNAs can form base pairs with quadruplet codons and can read [a+1] frameshifting and 256 different quadruplet codon combinations exist. However, only codons followed by less frequently used sense codons are likely to yield reasonable amounts of ncAA incorporation. As an example, the AGGA codon has been shown to be successful in incorporating ncAA into proteins.^[28] Previously, TAG and AGGA codons were used for double incorporation by Chin group. Triple incorporation of three different ncAAs have also reported in attempt to explore the limits of the ncAA incorporation.^[29] [30], [31]

More out-of-box approaches have been used to make more codons available for ncAA incorporation in the last decade. A promising development is that of Romesberg Lab reporting a non-canonical base pair in 2015 which yielded brand new codons.^[32] NcAA incorporation is also reported by using this system.^[33] Recently, Chin group reported the total synthesis of *E. coli* genome to convert the 61-codon system to a 59-codon system thus freeing up 2 more codons.^[34] These recent developments have made more codons available for incorporating more ncAAs. So far, no updates have been made on using these new codons for ncAA incorporation.

1.1.4 A New PylRS variant: *Candidatus Methanomethylophilus alvus* PylRS

Candidatus Methanomethylophilus alvus (*CMa*) is a methanogen present in the human gut. In 2012, its genome sequence was reported. In that report, in-frame amber stop codons (TAG codons) and genes required for pyrrolysine biosynthesis were observed suggesting the presence of pyrrolysine incorporation machinery.^[35] Two years later, the same research group reported the sequence of pyrrolysine tRNA and pyrrolysine tRNA synthetase present in *CMa*.^[36]

A remarkable feature of this newly discovered PylRS variant was the length of the PylRS enzyme. *CMa* PylRS consists of 275 amino acids whereas, *Methanosarcinaceae* PylRS derivatives such as *Mm* PylRS and *Methanosarcinaceae Barkeri* (*Mb*) PylRS have around 450 amino acids. *Methanosarcinaceae* PylRS contains two domains named as N-terminal domain and C-terminal domain. The N-terminal domain facilitates interaction with tRNA^{pyl} and the C-terminal domain contains the active site. *CMa* PylRS's sequence shows homology with the C-terminal domain only and no gene fragment in the *CMa* genome shows homology with the N-terminal region of *Methanosarcinaceae* PylRS. Although *CMa* PylRS do not have the N-terminal domain which can form additional interactions with its cognate tRNA, it is still orthogonal to *E. coli* and to mammalian cells.^[37]

The lack of the N-terminal domain makes *CMa* PylRS useful for ncAA incorporation. The N-terminal domain of *Methanosarcinaceae* PylRS derivatives causes aggregation of the enzyme in cells hence, its *in vivo* activity is not proportional to its expression level in the cell. Previously, the N-terminal domain deleted PylRS variants

were tested for activity. Although N-terminal deleted PylRS variants can work *in vitro* conditions they did not yield any activity *in vivo*. What makes *CMa* PylRS unique and important is although it lacks the N-terminal domain it is still functional and orthogonal *in vivo*.^[38]

On the other hand, the crystal structure of *CMa* PylRS (PDB ID: 6ezd) aligns very well with the C-terminal domain of *Mm* PylRS (PDB ID: 2zim) (Figure 4). Therefore, ncAA substrates of PylRS variants can be successfully incorporated by transferring the corresponding mutations to *CMa* PylRS. However, some cases exist where the transferring mutations from the *Mm* PylRS variant to *CMa* PylRS did not yield active enzyme.

In a 2018 study, *CMa* PylRS and *CMa* tRNA genes were cloned into pBT5 plasmid, which allowed these genes to be expressed under T5 promoter. T5 promoter is a strong constitutive promoter that allows high expression levels. In this study, *CMa* PylRS seemed to have 14-fold higher incorporation efficiency when compared to its PylRS variant.^[39]

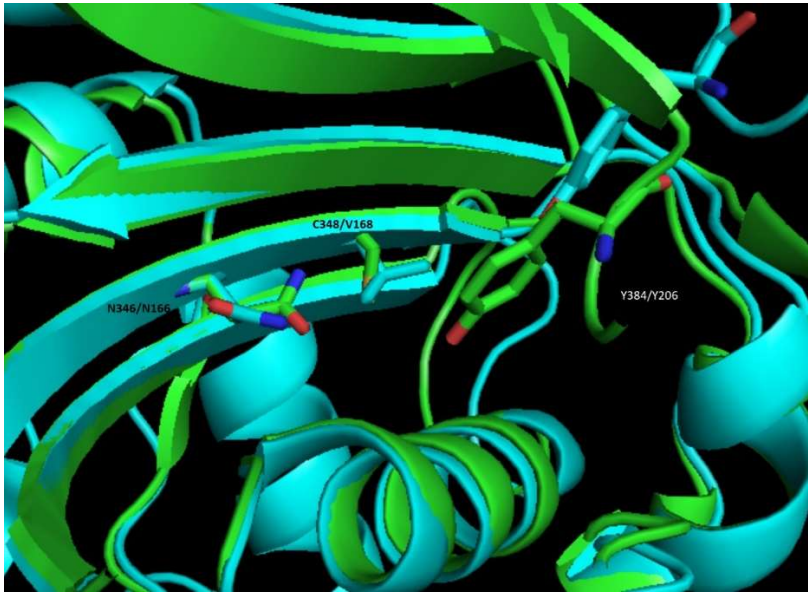


Figure 4 Alignment of *CMa* PylRS (cyan color, PDB ID: 6ezd) and *Mm* PylRS (green color, PDB ID: 2zim) active sites. Three active site residues are shown as sticks (*Mm* PylRS residues: N346, C348, Y384; *CMa* PylRS residues: N166, V168, Y206F)

1.2 Introduction to Antibody Structure and Function

An antibody (immunoglobulin) is a Y-shaped polypeptide complex that is produced by the immune system. Its primary role is to neutralize foreign bacteria and viruses in an organism. Antibodies consist of two sets of polypeptide chains held together by both disulfide bonds and non-covalent interactions. The two polypeptide chains are named as heavy chain and light chain in reference to their molecular weights. Antibodies are classified in to 5 major classes that are IgM, IgD, IgG, IgA and IgE.

In the bloodstream, IgG is the most abundant type of antibody. IgGs have a molecular weight around 150 kDa that are composed of two identical units of heavy chain around 50 kDa and two identical copies of light chain around 25 kDa (Figure 5A). Each of the heavy chain and light chain has a variable region and a constant region. The constant region of the same antibody subtype has the same amino acid sequence in the organism whereas, the variable region changes in amino acid sequence. Variable chains (V_H and V_L) gives the antibody selectivity towards binding to its specific antigen. The region around the N-terminus of the two variable chains is referred as the antigen binding site. Heavy chain`s constant region is further divided in to three regions (CH1, CH2, CH3). CH1 links the variable region with CH2 through a hinge region rich in intermolecular disulfide bonds that hold polypeptides together. CH2 and CH3 regions play a key role in binding to surface antigens on effector cells. Each CH3 domain of heavy chains interact through non-covalent interactions (Figure 5B).^[40] In humans, IgGs are present in 4 different subtypes (IgG₁, IgG₂, IgG₃, IgG₄) characterized by their

distinctive disulfide bond pattern (Figure 5C).^[41]

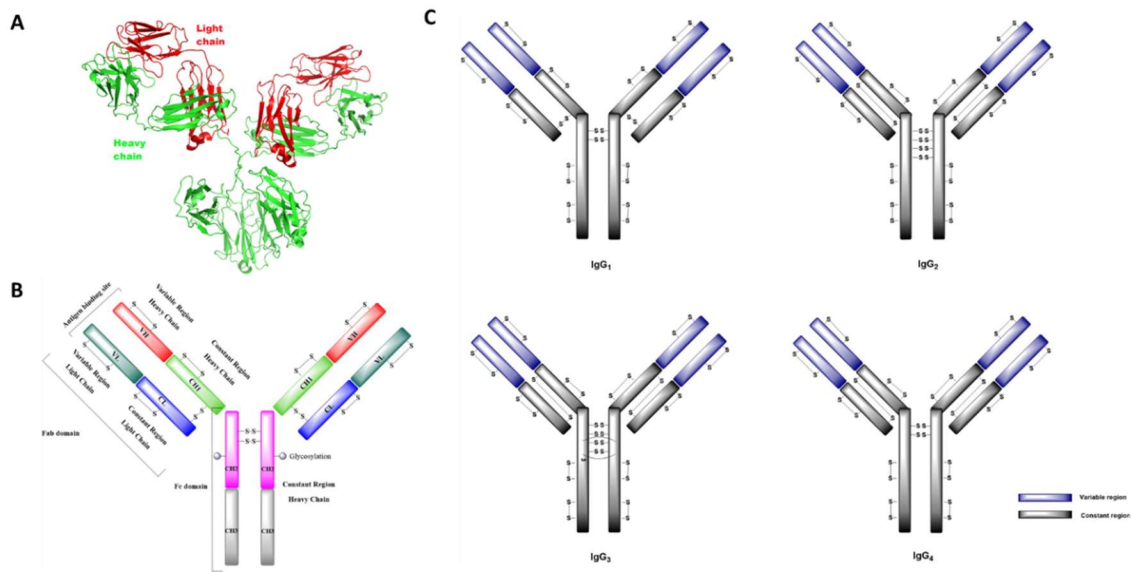


Figure 5 A) Light chain and heavy chain showed in red and green color respectively B) Regions of an antibody C) Disulfide bond patterns of IgG₁₋₄

In organisms, IgGs can serve two purposes: involving in signaling by binding to cell surface receptors through the Fc region and neutralizing foreign particles (bacteria, virus etc.) by binding through antigen binding sites in the organism.^[42] Once an IgG binds to a cell surface receptor it can be internalized through receptor-mediated endocytosis. This binding mode forms the basis behind the antibody-drug conjugate design explained in the next chapter.

1.2.1 Antibodies in Cancer Treatment

The severity of the side effects of untargeted cancer treatments such as cis-platin prompted scientists to search for targeted treatment options.^[43] Antibody-based therapeutics is an example of targeted cancer treatment. A well-studied example is an IgG₁ antibody Trastuzumab. Trastuzumab is an FDA-approved drug used to treat human

epidermal growth factor receptor 2 (HER2)-positive breast cancer. HER2 is a cell surface receptor overexpressed in some aggressive cancer types. In HER2-positive breast cancer, HER2 dimerization induces signaling that promotes tumor growth. Trastuzumab binds to HER2 and helps to slow down tumor growth by blocking the dimerization of HER2. Patients administered with Trastuzumab are shown to have improvements in overall survival time and showed reduced risk of cancer recurrence.^[44]

In addition to their role in signaling pathways, antibodies can be used as cargo delivery agents. An antibody recognizing a specific antigen on the cell surface and being internalized opens up the possibility of using antibodies as targeted drug delivery agents. By using toxin-attached antibodies that target cell overexpressed receptors on a cancer cell's surface, it is possible to deliver the toxin specifically to the cancer cell to destroy it. These toxin-attached antibodies are referred to as antibody-drug conjugates (ADCs).^[45]

1.2.2 Structure and Mechanism of Action of ADCs

ADC design comprises three main parts: antibody, linker and the drug (toxin). The example shown in Figure 6A is FDA-approved ADC Brentuximab Vedotin.^[46] In this example, the antibody is modified by its cysteine residue. Cysteine reactivity is easy to control because cysteines are generally not very abundant in proteins and cysteines have unique nucleophilic properties over other nucleophiles in proteins. The linker-toxin molecule is conjugated to antibody towards cysteine-maleimide reaction. This reaction, is widely used for the preparation of ADCs due to its fast kinetics.^[47] The conjugation moiety is followed by valine-citrulline, a dipeptide cleavable linker. Valine-citrulline group is followed by self-immolating group. A valine-citrulline linker is recognized by

the enzyme Cathepsin-B in the endosome and hydrolyzed at citrulline-anilide amide bond, leaving a free amine group, which decomposes to 4-methylenecyclohexa-2,5-dien-1-imine and carbon dioxide to liberate the free drug Monomethyl auristatin E (MMAE) (Figure 6B).

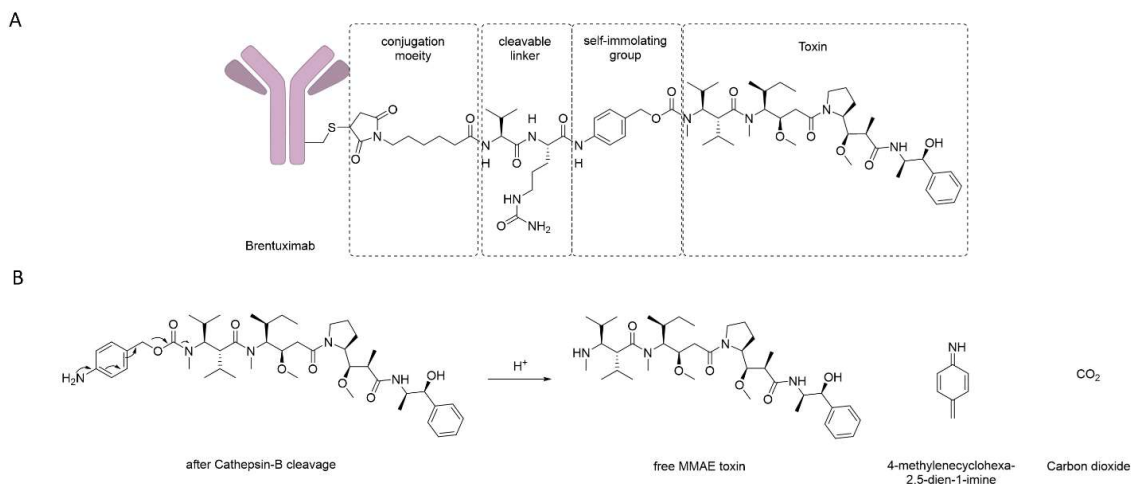


Figure 6 A) Structure of Brentuximab Vedotin **B)** Release of free MMAE toxin

The fate of ADCs in cancer cell is summarized in Figure 7. ADCs enter cells through binding to their specific cell surface antigen. After binding to the antigen, ADCs are internalized in to cells through receptor-mediated endocytosis. In this process, clathrin proteins self-assemble to aid in forming a vesicle with the cell membrane by bending the cell-membrane structure. After the maturation of the vesicle inside the cell clathrin proteins leave to form the endosome. After the endosome, the antibody can be released outside the cell through recycling the endosome (Figure 7a). Another internalization is through autophagy (Figure 7b). If the internalization mechanism proceeds through endosomes the next stage is the lysosome. In this stage, proteases such as Cathepsin B chops off the antibody and the drug is detached from the antibody. Then the drug

imparts its toxicity on the cell. Then, the toxin (in some cases) can get out and enter neighboring cells and continue to kill neighboring cells (Figure 7c). This is called the by-stander killing effect and will be addressed later again.^[48]

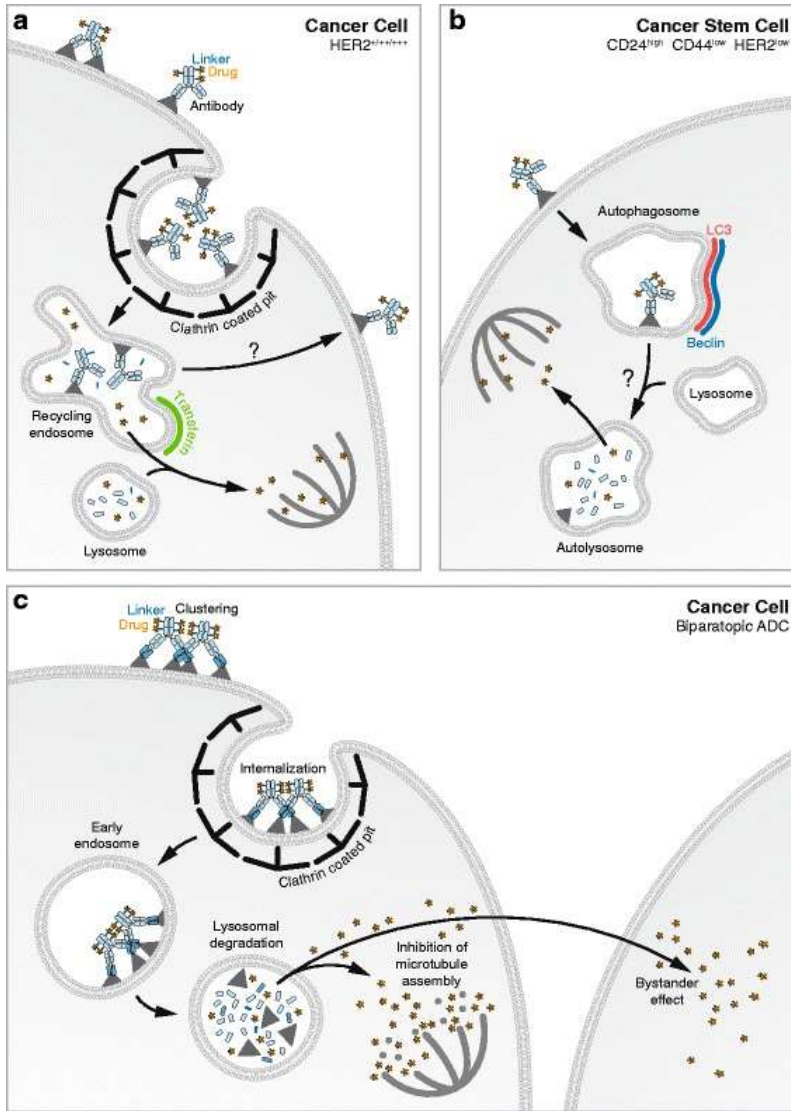


Figure 7 This figure is copied from ref [48] **A)** Releasing antibody outside the cell with recycling endosome **B)** Antibody internalization through autophagy **C)** By-stander effect of toxin

Properties of Antibodies

The antibody part of the ADC should be able to recognize the antigen present on the cancer cell surface specifically. IgG₁ antibodies are mostly used for the preparation of ADCs. They have serum stability of 21 days which is similar to other IgGs. Their main advantage is having a higher binding affinity in the Fc region to bind and activate monocytes and macrophages. For most cases, activating monocytes and macrophages will result in secondary immune functions such as complement-dependent cytotoxicity (CDC), antibody-dependent cell-mediated cytotoxicity (ADCC) and antibody-dependent cell-mediated phagocytosis (ADCP). However, in some cases having ADCs involved in secondary immune functions may decrease their primary effects because, antibodies may not reach cancer cells due to their increased interaction with effector cells. If the secondary immune functions decrease the function of the ADCs, engineering the Fc region to modulate these interactions may improve ADC's overall performance.^[49]

Another concern for ADCs is their low tissue permeability. Only around 1-2% of the administered ADCs can penetrate tumor tissue and reach cancer cells. To circumvent this problem, antibody fragments instead of full-length antibodies were tested. Results indicated that antibody fragments display superior tissue permeability; however, they show faster clearance rates as a downside.^[50]

Properties of Antigens

At the molecular level, there are remarkable differences present between cancer cells and healthy cells. An example of this discrepancy is tumor antigens. Tumor antigens are

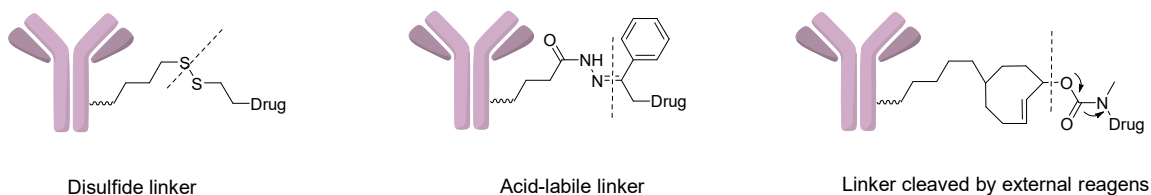
produced in cancer cells that can be recognized by an antibody and/or immune effector cells. There are two subtypes: tumor-specific antigens and tumor-associated antigens. Tumor-specific antigens are expressed only in cancer cells and tumor-associated antigens are expressed mostly in cancer cells and also expressed in healthy cells albeit at a lower level. Ideally, an antigen should be overexpressed on the cancer cell surface with low heterogeneity throughout the tumor tissue. Internalization rate of the antigen and binding affinity of antibody to tumor are some other factors determining the efficiency of the ADCs.^[50]

Properties of Linkers

One of the main goals of an ADC design is to avoid off-target effects as much as possible. Therefore, it is important to have a stable linkage between toxin and antibody that will not release the toxin non-specifically in physiological conditions. Depending on the toxin used, both cleavable and non-cleavable linkers can be used. A non-cleavable linker does not have a specific group involved to undergo self-immolation to liberate free toxin upon being internalized in cancer cells. Kadcylla is an example of a toxin that can be used with non-cleavable linkers in ADCs.^[51] However, toxins need to be in their native (free) form to show potency for most cases. Moreover, the bystander effect is more efficacious with toxins in their native state. Therefore, cleavable linker strategies are generally preferred and in this section covers examples of cleavable linkers are covered.

Cleavable linker strategies can be classified as chemically cleavable and enzymatically cleavable linkers.^[52] Most linker types from both classes rely on the unique chemical and proteomic environment of the lysosome. Subtypes for chemically cleavable linker types can be listed as disulfide-containing linkers that can be cleaved by reducing agents, acid-labile linkers that can be hydrolyzed in the acidic lysosome environment and linkers that are cleaved by external reagents such as tetrazine and palladium (Figure 8A). Examples of enzymatic linkers are glucuronide linkers that can be cleaved by β -Glucuronidase enzyme, dipeptide linkers that can be cleaved by proteases and diphosphate linkers that can be cleaved by diphosphatases (Figure 8B).

A. Chemically cleavable linkers



B. Enzymatically cleavable linkers

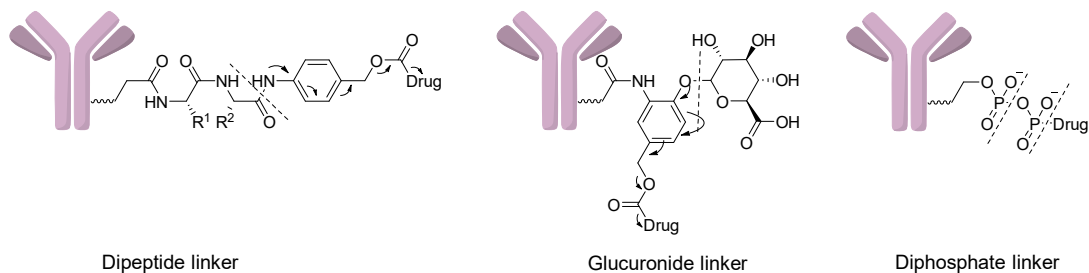


Figure 8 A) Chemically Cleavable Linkers **B)** Enzymatically cleavable linkers

Properties of Toxins

As mentioned above, only a very small fraction of the injected ADCs can penetrate the tissue and reach cancer cells. Therefore, the toxins attached to antibodies need to have

sub-nanomolar concentrations to show efficacy. On the other hand, these toxins are natural products that can be produced by multistep synthesis routes. Thus, making the whole process economically viable is another reason why highly potent toxins are desired. The toxins used in ADCs can be classified under two categories: microtubule-targeting toxins and DNA-damaging toxins. Maytansinoids and auristatins are two classes of toxins that inhibit microtubule formation. Both function by inhibiting microtubule assembly by binding different regions of microtubules. Camptothecins, calicheamicins, doxorubicins, duocarmycins, pyrrolobenzodiazepines are examples of DNA-damaging toxins. These classes of compounds show toxicity by intercalating DNA (camptothecins), alkylating DNA (pyrrolobenzodiazepine and duocarmycin) or breaking DNA double-strands (calicheamicin).^[53]

Drug to Antibody Ratio and Hydrophobicity of ADCs

As discussed above, cysteine-based conjugation methods are preferred due to higher control of the homogeneity of the synthesized ADCs. In an IgG₁ antibody there are 4 mildly reducible disulfide bonds. Reducing and alkylating these cysteines with linker-toxin compounds can lead to 8 drugs per one antibody molecule. At first, having as many toxins as possible on an antibody may seem desirable due to increased potency. However, higher DAR generally comes with poor pharmacokinetics. ADCs with high DAR values become very hydrophobic and tend to form aggregates. Aggregation increases the antibody clearance rate significantly; therefore, it is undesirable. To decrease the aggregation, ADCs with lower DAR are preferred. Lower DAR is achieved by mutating some of the mildly reducible cysteines with serines whenever ADCs with

higher DAR values are desired. Another solution to alleviate this problem is to include long polyethylene glycol (PEG) groups. These PEG groups presumably attract water molecules and decrease hydrophobic interactions and aggregation.^[54]

1.2.3 Conjugation Strategies

The conjugation strategy has direct effects on homogeneity, stability and DAR of the ADCs. Therefore, pharmacokinetics and pharmacodynamics (PK/PD) of ADCs depend heavily on the conjugation strategy chosen. Conjugation methods can be grouped into four different approaches: lysine-based conjugation, cysteine-based conjugation, conjugation by introducing non-canonical functional groups and enzymatic conjugation.

Lysine-based conjugations

Lysine has an amine group in its side chain with a pKa around 9 allowing it to form amide bonds with activated esters. Generating conjugates through amide bond formation offers several advantages. First, antibodies have around 40 accessible lysines on their surface available for amide bond formation to some degree.^[55] Second, amide bond moiety is highly abundant in biological systems so, its fate in cells is more predictable than non-canonical moieties. Lysine and n-hydroxysuccinimide (NHS) ester coupling was used for Trastuzumab emtansine which the FDA approved in 2013. In this case, lysine residues were first reacted with an NHS ester containing a thiol group.^[56] Then, the thiol group linked to the toxin Kadcyla through maleimide chemistry which will be covered in cysteine conjugation methods (Figure 9A). Although this method proved to be useful, it did not yield homogenous ADCs. There are around 10 lysine residues on an

antibody that are modified at significant levels. DAR can be expressed as a statistical distribution and attaining batch-to-batch reproducibility can be challenging. These drawbacks can be alleviated using cysteine-based conjugation, conjugation by introducing non-canonical functional groups and enzymatic conjugation alternatives.

Cysteine-based conjugations

Cysteine is unique among canonical amino acids with its highly nucleophilic thiol side chain. The thiol group in the cysteine side chain has a pKa of 8.1. Therefore, in physiological conditions, 10-20% of the cysteines can be present in their thiolate form. Unlike lysines, cysteines have a low abundance in proteins. An IgG₁ antibody has 32 cysteines. All of them form disulfide bonds so, they do not act as nucleophiles. However, 8 cysteines (forming 4 disulfide bonds with each other) around the hinge region can be reduced under mild conditions unlike the remaining cysteines. Once reduced, 4 disulfide bonds yield 8 thiol groups that are available for conjugation.

Until now, the most widely used conjugation method for ADC generation has been cysteine-maleimide reaction. It has an impressive second order rate constant ($734 \text{ M}^{-1} \text{ s}^{-1}$)^[47] and shows excellent selectivity for cysteines. For example, the FDA-approved ADC brentuximab vedotin^[46] is synthesized by cysteine-maleimide reaction (Figure 9B). However, one significant drawback of this reaction is its reversibility. Retro-michael addition reaction can shift the equilibrium to starting materials and causing linker-toxin moiety to be detached from the antibody.^[57] Another pathway for cysteine-maleimide conjugate is hydrolysis. Once the five membered ring opens up by the hydrolysis of the amide bond of the thiosuccinimide, the conjugate is stabilized (it does not undergo retro-

michael addition). However, at physiological pH, the hydrolysis of the succinimide ring is insufficient to outcompete the retro-michael addition reaction.

Accelerating the cysteine-maleimide conjugate hydrolysis in order to minimize the retro-michael addition reaction has been proven to be an effective solution. Because the amide bond in the maleimide ring is in conjugation with two carbonyl groups and there is ring strain, it is prone to hydrolysis. The treatment of the conjugates with pH 9.2 buffer is sufficient to hydrolyze and open up the maleimide ring.^[58] A more elegant solution is to place a free amine group in the vicinity of the maleimide ring to form self-hydrolyzing maleimides (Figure 9C). These compounds have faster hydrolysis rates after cysteine-maleimide conjugate is formed and yield highly stable ADCs.^[59]

Many other electrophiles in addition to maleimide have been used with success.^[60] A noteworthy example is the disulfide bridging bifunctional electrophiles.^[61] Once, one of the cysteines is modified with the bifunctional electrophile, neighboring cysteine (presumably its former disulfide bond partner) can react with the same molecule. By this, the disulfide bond is 'rebridged' and the final ADC is structurally closer to the naked antibody compared to conjugation methods that do not rebridge the disulfide bond (Figure 9D).

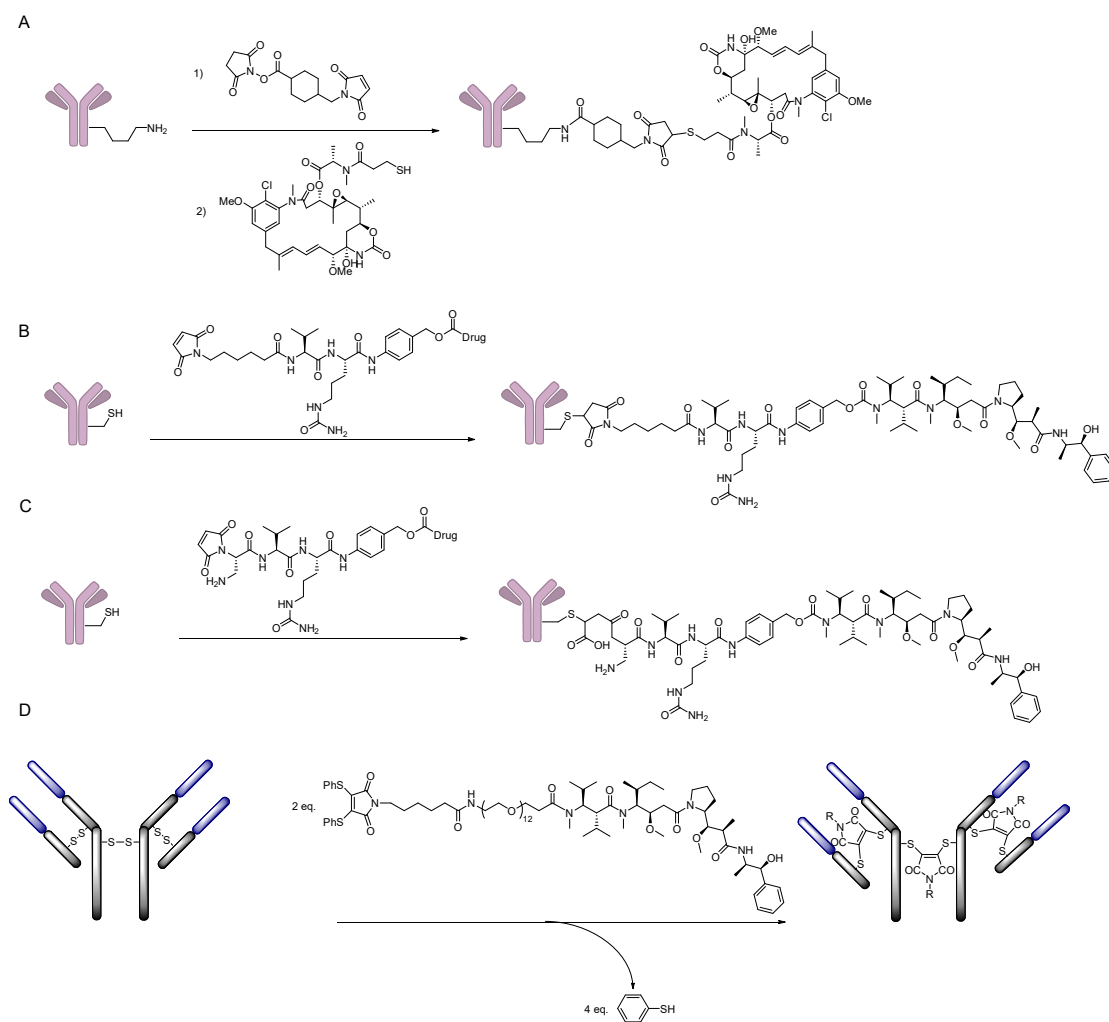


Figure 9 A) Synthesis of Trastuzumab emtansine by lysine-NHS conjugation B) Synthesis of brentuximab vedotin by cysteine-maleimide conjugation C) Conjugation with self-hydrolyzing maleimides D) Conjugation by rebridging disulfide bonds.

Conjugation by non-canonical amino acid incorporation

Although cysteine-based approaches were useful, non-canonical functional group incorporation offers unique benefits. First, by using the ncAA incorporation technique, one can mutate a residue and carry out the conjugation at that desired site while maintaining the native disulfide bond structure. Due to complex disulfide bond structure of the antibodies, mutating residues with cysteines can cause structural problems.

Second, adding new functional groups allow new conjugation chemistries bringing faster conjugation with increased stability. Many conjugation reactions were enabled by incorporating non-canonical functional groups such as azide, cyclopropene, strained alkyne, and ketone. Figure 10 shows examples of antibodies conjugated by CuAAC (copper-catalyzed azide alkyne cycloaddition) reaction (Figure 10A),^[62] tetrazine ligation (Figure 10B),^[63] copper-free click reaction (Figure 10C),^[64] ketone-oxime ligation(Figure 10D)^[65] reactions. CuAAC reaction is one of the first examples of biorthogonal click reactions that have been used for conjugating small molecules to large biomolecules. The reaction creates a triazole group that is fairly stable in cells. Owing to its stability and its geometry, triazole moiety is utilized as bioisostere of amide groups.^[66] The reaction kinetics with non-metal chelating azide and regular alkyne do not allow the conjugation to reach completion in low-micromolar concentrations.²³ On the other hand, the tetrazine-cyclopropene reaction is shown to reach completion in 20h by using 2 equivalent of linker-toxin molecule with respect to antibody.²⁴ Copper-free click reaction eliminates the toxic copper atom from the reaction by using strained alkynes. Copper-free click reaction is ~7 times slower with aromatic azides compared to aliphatic azides.²⁵ Conversely, *para*-azido-L-phenylalanine has a higher incorporation efficiency than *para*-azidomethyl-L-phenylalanine which creates a trade-off between incorporation efficiency and conjugation efficiency. Another example is, conjugating hydroxylamine drug-linker compound with *para*-acetyl-L-phenylalanine incorporated antibody.²⁶ A 30-fold excess of the hydroxylamine-linker-drug compound was incubated at 37 °C for 1-2 days to reach 95% conversion.

conjugation method proven to yield >80% conversion with 64-fold sub-stoichiometric amounts of Sortase A and 5-fold molar excess of linker-toxin compound.

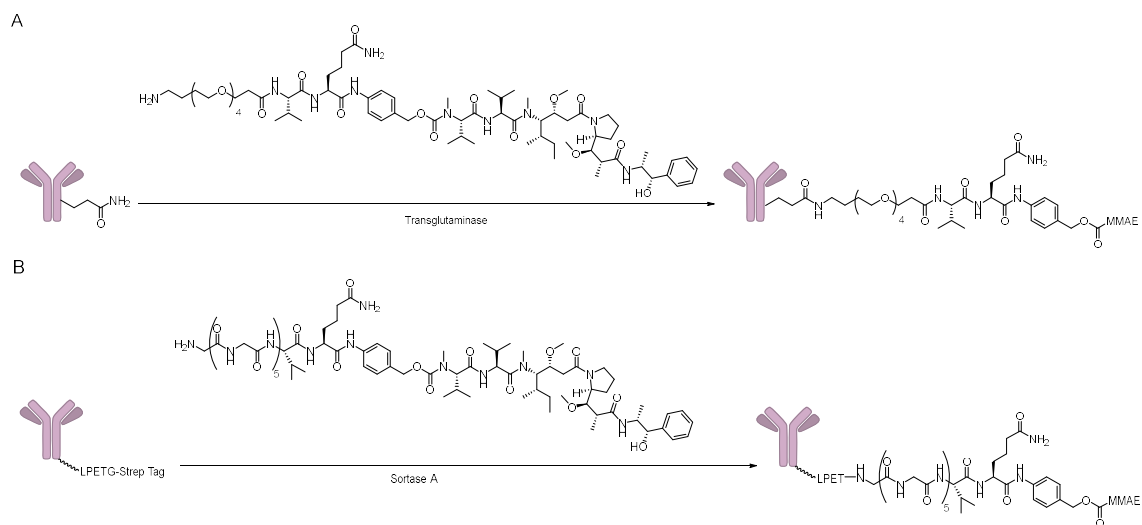


Figure 11 A) Transglutaminase Conjugation B) Sortase A Conjugation

1.3 Main Protease of Severe Acute Respiratory Syndrome Coronavirus 2

1.3.1 Life Cycle of Severe Acute Respiratory Syndrome Coronavirus 2

The pandemic caused by SARS-CoV-2 has emerged in December 2019 and surpassed 2 million global deaths as of February 2021.^[69] SARS-CoV-2 is an mRNA virus. Its genome encodes for four structural proteins (SPs) named as spike (S), envelope (E), membrane (M) and nucleocapsid (N). Non-structural proteins (NSPs) are expressed as two long polypeptides (Figure 12A) and digested by papain-like protease (PL^{pro}) and main protease (M^{pro}) to generate functional proteins (Figure 12B).^[70]

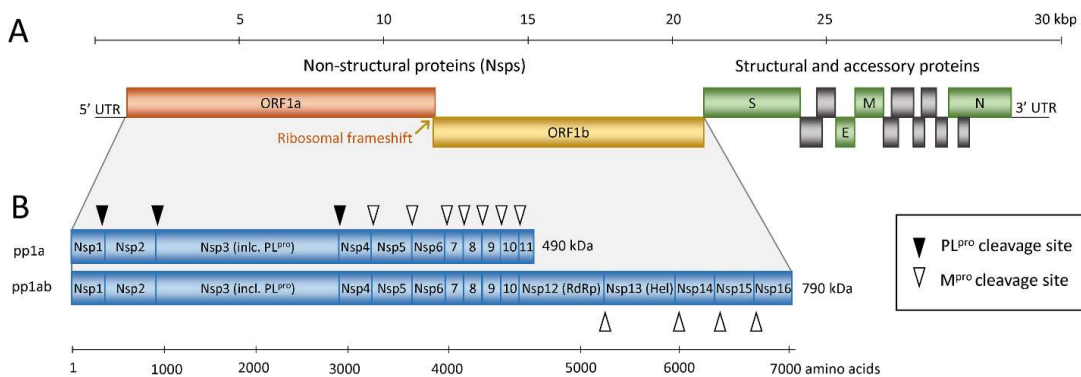


Figure 12 A) Structural and non-structural proteins encoded by SARS-CoV-2 genome **B)** Cleavage sites for PL^{pro} and M^{pro}

The life cycle of SARS-CoV-2 starts by binding to cell surface receptor angiotensin-converting enzyme 2 (ACE2) through spike (Figure 13).^[71] Upon binding, spike is partially digested by cell membrane proteases such as furin and transmembrane serine protease 2 (TMPRSS2) during the adsorption and internalization process. This enables the virus coat to fuse with the cell membrane facilitating endocytosis of the virus. Proteolysis of spike continues in the endosome and lysosome stages primarily by Cathepsin L and Cathepsin B and the viral mRNA is released to the cytosol as a

result.^[72] After the viral mRNA release, large polypeptides that encode for NSPs are expressed. These large polypeptides are digested into 16 proteins by viral proteases M^{pro} and PL^{pro}. M^{pro} is alternatively named as 3-chymotrypsin protease-like protease (3CL^{pro}) and non-structural protein 5 (NSP5) as well and PL^{pro} is also named as non-structural protein 3 (NSP3). Both M^{pro} and PL^{pro} are expressed in the same open reading frame. NSPs and the RNA-dependent polymerase form the replication complex to amplify the viral mRNA and express more SPs and NSPs. Next, SPs and accessory proteins are transferred to the endoplasmic reticulum and together with the Golgi apparatus, the virus particles are assembled and released out of the cell as new virions by exocytosis.^[71]

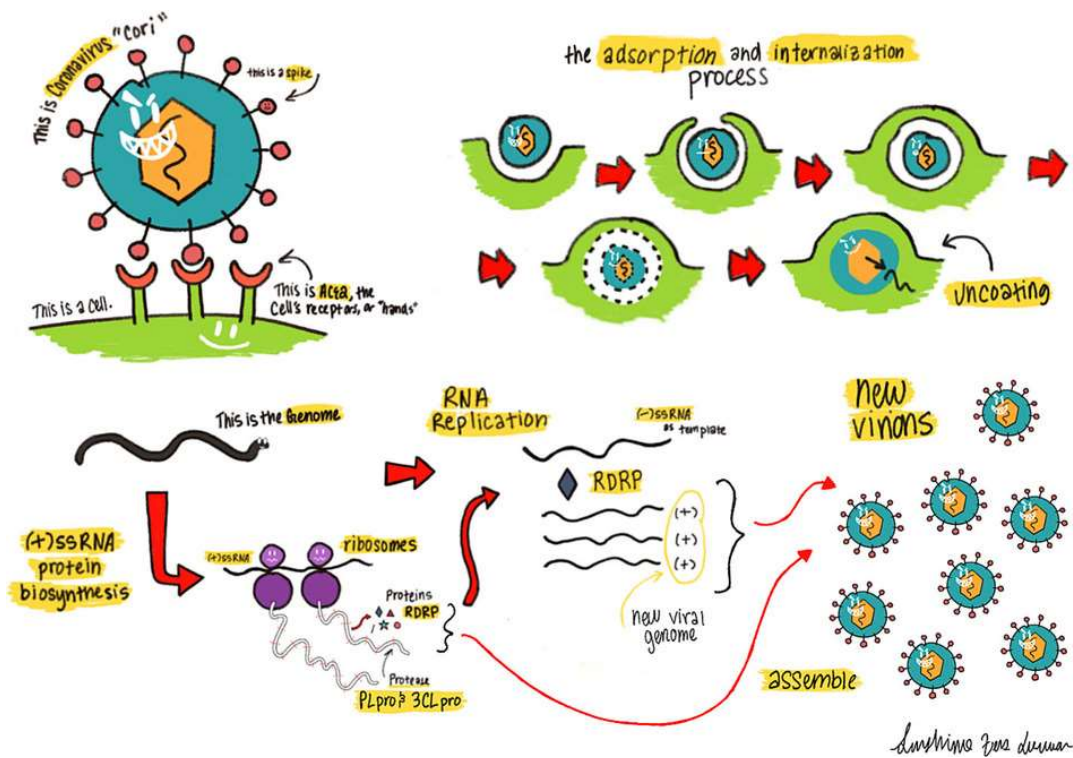


Figure 13 Life cycle of SARS-CoV-2

Many treatment options have been proposed and employed based on the current understanding of the life cycle of SARS-CoV-2. Among the most prominent treatments, FDA-approved Remdesivir inhibits the RNA-dependent polymerase and prevents replication of the mRNA.^[73] Hydroxychloroquine is proposed to be inhibiting endosome formation by raising the endosomal pH and through inactivating M^{pro}.^[74] GC376 is a broad-spectrum protease inhibitor that can inactivate M^{pro} in coronaviruses and disrupts polyprotein digestion into functional proteins for viral mRNA replication.^[75] Favirapir, an antiviral, that is previously known to be effective against influenza, acts as a substrate of viral RNA-dependent RNA polymerase and prevents further expansion of the RNA

strand.^[76] However, Remdesivir, Hydroxychloroquine, and Lopinavir-Ritonavir (previously proposed to inactivate M^{pro}) treatment options did not seem to decrease mortality rates significantly.^[77] Therefore, the hunt for an effective small molecule treatment options is still ongoing.

1.3.2 M^{pro} as a Drug Target

As discussed above, M^{pro} digests the polypeptide chain through 11 sites to generate functional proteins from long polypeptide chains. Therefore, it is an essential enzyme for the life cycle of SARS-CoV-2. The M^{pro} active site contains a cysteine that is essential for its activity. Therefore, this active site cysteine allows researchers to develop highly potent reversible and irreversible covalent inhibitors and non-covalent inhibitors.

Moreover, the substrate specificity of M^{pro} is different from human proteases which paves the way for developing selective M^{pro} inhibitors. There are M^{pro} inhibitors that have showed efficacy in eradicating SARS-CoV-2 replication in cell culture experiments and undergoing pre-clinical trials.^[78]

1.3.3 Structure of M^{pro}

The first crystal structure of M^{pro} was reported 4 months after the release of SARS-CoV-2 genome sequence by Chinese scientists. Sequence data showed 96% similarity suggesting that the structural features, substrate preferences, inhibitors of M^{pro} were likely to be similar to SARS-CoV-M^{pro}.^[71]

Similar to SARS-CoV M^{pro}, M^{pro} also has a dissociation constant (K_d) 2.5 μ M measured by analytical ultracentrifugation.^[79] The dimerization takes place by the N-

terminus residues coming together to form non-covalent interactions. Studies report that M^{Pro} constructs having purification tags on the N-terminus have shown significantly decreased activity.^[80]

M^{Pro} crystal structure shows a tight dimer form (Figure 14A). The M^{Pro} active site consists of a catalytic dyad with Cys145 and His41 and substrate binding pockets (Figure 14B). One significant mutation in M^{Pro} compared to SARS-CoV M^{Pro} is T285A. Residue 285 is at the dimerization interface and mutating the residues S284, T285, I286 with alanine showed a 3.6 fold increase in activity.^[81] This increase is presumably due to increasing the dimerization efficiency by reducing steric crowding at the dimerization interface.

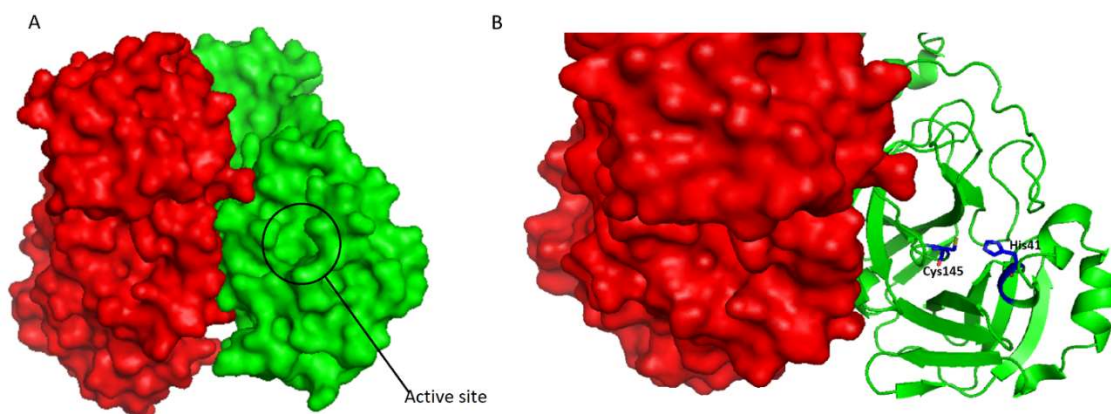


Figure 14 Crystal structure of M^{Pro} (PDB ID: 6y2e) **A)** Dimer form M^{Pro} units are shown in red and green color **B)** Catalytic residues of M^{Pro}

1.3.4 Is M^{Pro} a Chymotrypsin-like Protease or a Papain-like Protease?

Although an alternative name for M^{Pro} is 3-chymotrypsin-like protease, contradicting papers published on its mechanism. There are mainly two proposed mechanisms for the catalytic activity of M^{Pro} based on how deprotonation occurs. These are the general base

mechanism and the ion-pair model. In a 2004 study, pH activity profile and pKa of catalytic dyad residues of the wild-type enzyme and C145S mutation showed evidence of a general base mechanism (Figure 15A).^[82] However; in another study in 2008, the pH profile of iodoacetamide-based inactivation and pre-steady state kinetics showed evidence of an ion-pair mechanism. In this study, the variation in the kinetics of the reaction between iodoacetamide and Cys145 hinted an ion-pair mechanism (Figure 15B).^[80] In a recent neutron crystallography study, an M^{pro} crystal was collected from a crystal at pD 7.0 with 2.5 Å resolution. The structure showed the His41 residue protonated and Cys145 in thiolate form. Surprisingly, among total 12 cysteines, the structure showed C22, C38, C44, C128 cysteines as deprotonated as well.^[83] A computational study proposed the formation of ion-pair upon binding of the substrate (Figure 15C).^[84]

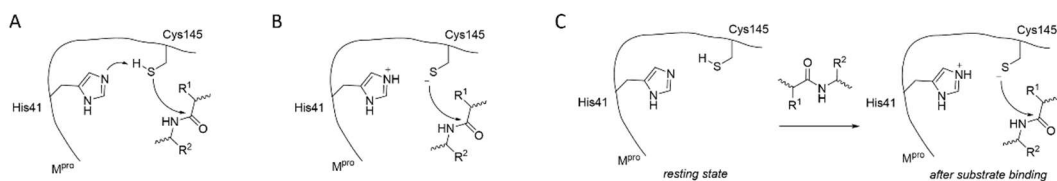


Figure 15 A) General base mechanism B) Ion-pair mechanism C) Ion-pair formation upon substrate binding

1.3.5 Amino Acid Preference of M^{pro} Substrates

Stringent preference on residues P4, P3, P2, P1, P1', P2' was observed in the *in vitro* studies done with SARS-CoV M^{pro} and M^{pro}.^{[85], [86]} The most stringent preference is observed for the P1 position. Stringency decreases moving both N-terminus and C-terminus directions (Figure 16). Uniport orders the degree of selectivity of each position

in the following order $P1 > P2 > P1' > P4 > P3$.^[78] Starting from the N-terminus site, the P4 position favors amino acids with small hydrophobic side chains such as alanine and valine most. P3 position favors amino acids such as threonine, lysine and arginine. The P2 position is generally occupied by hydrophobic residues such as leucine, methionine, phenylalanine with leucine being the most preferred residue. There is a very strong preference for glutamine residue in the P1 position. The only other residue that shows some activity at the P1 position is histidine. P1' prefers residues with small side chains such as serine, alanine, cysteine, glycine, threonine in the given order (serine showing the highest activity). The P2' position favors small side chain amino acids like alanine, glycine, glutamic acid and threonine. SARS-CoV M^{pro} and M^{pro} have very similar sequence preferences.

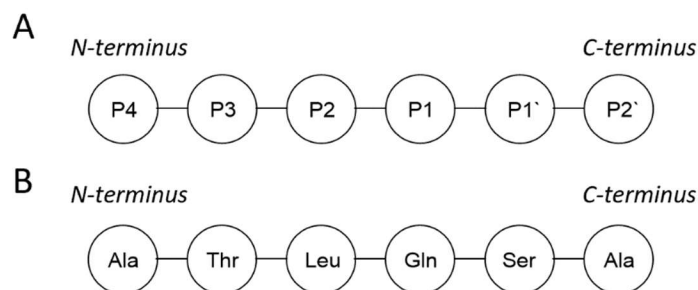


Figure 16 A) Residue nomenclature B) Most preferred substrates for each position

1.3.6 Substrates Used in M^{pro} Assays

Guided by the substrate amino acid preference of SARS-CoV M^{pro}, dozens of substrates were designed for measuring the activity of SARS-CoV-M^{pro} and M^{pro}. The developed substrates can be grouped into three categories: Tagless peptide substrates, P1'-dye

substrates and FRET-pair substrates. Table 1 summarizes the properties of the substrates and Figure 17 shows the structures of the dyes used as reporters.

Tagless Peptide Substrates

In enzyme assays, the structural perturbation brought by tags is generally a point of concern. Therefore, assays without reporters or tags are desirable when applicable. Enzymatic reactions using tagless peptide substrates can be tracked by high pressure liquid chromatography (HPLC). Fluorescence-based methods include using large dye molecules and quantification can be an issue due to quenching. In such cases, tagless peptide substrate counterparts of fluorescent-based substrates can assess the reliability of the later. The major drawback of using tagless peptide substrates is the lengthy HPLC analysis, for instance, when one wants to do inhibition analysis of a range of inhibitor concentrations in triplicates or conducting a screening study. Therefore, tagless peptides are used for perturbation brought by reporter molecules and fluorescence. To meet the demands of doing more analysis in a shorter time, spectrometric analysis that can be done with a plate reader is adapted. An example (entry **19**) for a tagless peptide substrate is shown in Table 1.

P1'-dye Substrates

Fluorogenic assays that rely on turn-on fluorescence are desirable because of the inherent sensitivity of fluorescence. It is common to get clear signals from sub-micromolar concentrations of fluorogenic dyes. For M^{Pro} assays, catalytic action cleaves the amide bond between the P1 and P1' residues. A widely used probe contains 7-amino-

4-carbomylcoumarin (ACC) linked to P1 residue through an amide bond. In this case, only a very weak background fluorescence signal comes from the substrate due to diminished conjugation of the amine with the chromophore. However, upon hydrolysis the liberated amine group makes the chromophore highly fluorescent. However, as the substrate specificity studies demonstrate, M^{pro} is very selective at P1 and P1' residues. Having a rigid bicyclic moiety like ACC at P1' position decreases the enzyme activity significantly, necessitating using higher M^{pro} concentrations. In addition to, fluorogenic compounds, a UV-active compound *p*-nitroamine (*p*-NA) was also used (entries **21-23** in Table 1).

FRET-pair Substrates

FRET-pair substrates have two tags on both ends of the peptide that are the acceptor and the donor. Acceptor can serve only to quench the light emitted from the donor or can be using the light coming from the donor to emit at a different wavelength. FRET-pair substrates are preferred in measuring protease activities because, they allow the tags to be far away from the catalytic residues. In this manner, (different from P1'-dye substrates) they allow the measurement of the cleavage of the native amide bond (entries **20, 29-31** in Table 1). As a result, FRET-pair substrates can show enzyme activity at lower concentrations compared to P1'-dye substrate while maintaining the advantages of being a fluorescence-based assay. However, the kinetic parameters calculated with the FRET-pair substrates show large variation. Probably, due to quenching effects, accurate biochemical characterization with FRET-pairs can be troublesome. Examples of FRET-

pairs substrates and their kinetic parameters (entries **1-18, 24-28**) are summarized in

Table 1.

Table 1 Properties of M^{PT0} substrates

Entry	Reporte r	Sequence														K _{cat} (s ⁻¹)	K _m (μM)	K _{cat} /K _m (s ⁻¹ M ⁻¹)
		P7	P6	P5	P4	P3	P2	P1	P1'	P2'	P3'	P4'	P5'	P6'	P7			
1 ^[87]	E-D	-	E*	S	A	T	L	Q	S	G	L	A	K*	S	-	0.0053 ± 0.00	N/A	N/A
2 ^[79]	E-D	K*	T	S	A	V	L	Q	S	G	F	R	K	M	E*	N/A	N/A	3426.1 ±
3 ^[84]	E-D	K*	T	S	A	V	L	Q	S	G	F	R	K	M	E*	0.159	28.2 ± 3.4	5624
4 ^[89]	E-D	K*	T	S	A	V	L	Q	S	G	F	R	K	M	E*	0.272	170	1600
5 ^[90]	E-D	-	-	-	-	L*	A	Q	A	V	R	S	S	S	R*	0.033 ± 0.001	9 ± 1	3667
6 ^[91]	E-D	-	-	S*	A	V	L	Q	S	G	F	R	K*	-	-	N/A	N/A	29000 ±
7 ^[87]	488-QSY7	C*	E	S	A	T	L	Q	S	G	L	A	K*	-	-	0.0016 ± 0.00	N/A	N/A
8 ^[87]	594-QSY21	C*	E	S	A	T	L	Q	S	G	L	A	K*	-	-	N/A	N/A	N/A
9 ^[92]	Abz-NitroY	-	-	S*	V	T	L	Q	S	G	Y*	R	-	-	-	0.0168 ±	820 ± 130	200 ±
10 ^[93]	Abz-NitroY	-	-	S*	V	T	L	Q	S	G	Y*	R [†]	-	-	-	N/A	190 ± 23	N/A
11 ^[94]	Abz-NitroY	-	-	S*	V	T	L	Q	S	G	Y*	R	-	-	-	N/A	96 ± 31	N/A
12 ^[95]	Abz-dnp	-	-	S*	A	V	L	Q	S	G	F	R	K*	-	-	N/A	16 ± 1.1	N/A
13 ^[96]	Abz-dnp	-	T*	S	A	V	L	Q	S	G	F	R	K*	-	-	N/A	9.3 ± 1.5	N/A
14 ^[97]	Abz-dnp	-	T*	S	A	V	L	Q	S	G	F	R	K*	-	-	0.0107	145	69
15 ^[97]	Abz-dnp	-	S*	G	V	T	F	Q	G	K	F	K	K*	-	-	0.00505	252	20
16 ^[98]	MCA-dnp	-	-	-	A*	V	L	Q	S	G	F	K*	K	-	-	0.14 ± 0.01	129 ± 7	1085
17 ^[99]	MCA-dnp	-	-	-	A*	V	L	Q	S	G	F	K*	K	-	-	N/A	N/A	28500
18 ^[4]	MCA-dnp	-	-	-	A*	V	L	Q	S	G	F	K*	K	-	-	N/A	N/A	25600
19 ^[91]	No tag	-	T	S	A	V	L	Q	S	G	F	R	K	W	-	0.54 ± 0.04	230 ± 60	2400 ±
20 ^[100]	Rho	-	-	-	A	R	L	Q*	-	-	-	-	-	-	-	0.015	306	49
21 ^[101]	pNA	-	T	S	A	V	L	Q*	-	-	-	-	-	-	-	0.30 ± 0.0067	1000 ± 30	300
22 ^[92]	pNA	-	T	S	A	V	L	Q*	-	-	-	-	-	-	-	0.045 ±	690 ± 140	65.2
23 ^[102]	pNA	-	T	S	A	V	L	Q*	-	-	-	-	-	-	-	N/A	6000	N/A
24 ^[103]	MBP, EYFP	-	T*	S	A	V	L	Q	S	G	F	R	K	M*	-	1.349 ± 0.370	5.086 ± 2.046	265200 ±
25 ^[104]	MBP, EYFP	R*	E	G	T	R	V	Q	S	V	E	Q	I	R	E*	0.0033 ±	0.860 ± 0.303	3800 ±
26 ^[96]	ACC, DNP	-	-	G	Abu	Tle	L	Q	S	G	F	R	K*	K	-	N/A	N/A	14748 ±
27 ^[96]	ACC, DNP	-	-	G	Thz	Tle	L	Q	S	G	F	R	K*	K	-	N/A	N/A	19424 ±
28 ^[96]	ACC, DNP	-	-	G	V	K	L	Q	S	G	F	R	K*	K	-	N/A	N/A	6755 ±
29 ^[96]	ACC	-	-	-	Abu	Tle	L	Q*	-	-	-	-	-	-	-	0.178 ± 0.016	207.3 ± 12.0	859 ± 57
30 ^[96]	ACC	-	-	-	Thz	Tle	L	Q*	-	-	-	-	-	-	-	0.144 ± 0.006	189.5 ± 2.7	760 ± 50
31 ^[96]	ACC	-	-	-	V	K	L	Q*	-	-	-	-	-	-	-	0.050 ± 0.002	228.4 ± 9.9	219 ± 3

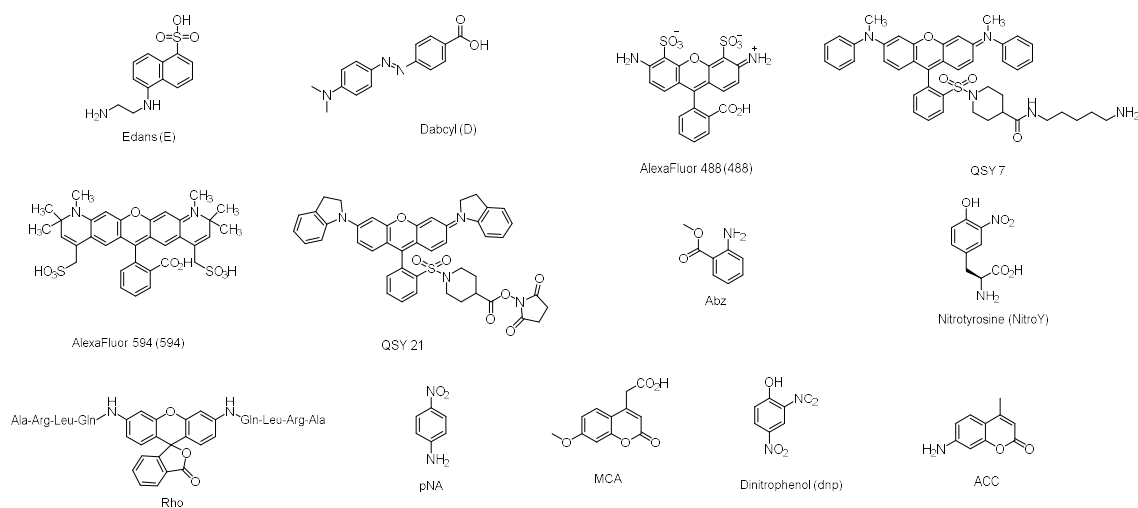


Figure 17 Reporters used with M^{Pro} substrates

1.3.7 Inhibitors of M^{Pro}

In vitro virus inhibition assays show that M^{Pro} inhibitors can inhibit SARS-CoV and SARS-CoV-2 replication with sub-micromolar concentrations. Moreover, M^{Pro} has a wide active site with a cysteine residue as the key moiety for catalytic activity. These all together make M^{Pro} a highly attractive target for inhibitor development. There is a vast amount volume of literature on inhibition of both SARS-CoV M^{Pro} and M^{Pro} . There are also clinical trials ongoing based on M^{Pro} inhibitors that show potency in *in vitro* virus-based assays. The inhibitors developed for M^{Pro} can be grouped into three categories: covalent inhibitors, non-covalent inhibitors and transition-metal based inhibitors.^[78]

Covalent Inhibitors

Covalent inhibitors are divided into two categories: reversible covalent and irreversible covalent inhibitors. Reversible covalent inhibitors include inhibitors with warheads such as ketones (**1a**),^[104] nitriles (**1b**),^[105] aldehydes (**1c**),^[106] bisulfite adducts (**1d** or

GC376),^[107] active esters (**1e**)^[108] that can form covalent adducts in the active site in a reversible fashion (Figure 18 A-E). Irreversible inhibitor examples are chloroacetamide (**2a**),^[109] epoxy ketones (**2b**),^[110] Michael acceptors (**2c**),^[94] phthalic hydrazide leaving group compounds (**2d**)^[111] can be classified as irreversible covalent inhibitors (Figure 19 A-D). These inhibitor designs are guided by substrate preference of M^{Pro}. They are peptide-based compounds that are mostly dipeptides, tripeptides tetrapeptides having side chains that can bind to active site pockets such as P3, P2, P1 active site binding pockets. In addition to the identity of the warhead, the side chain moieties also plays a big role in determining the potency of the inhibitors. These side chains guide the inhibitor to the active site. This class of compounds (especially aldehyde-containing inhibitors) has highly potent inhibitors with IC₅₀ values below 100 nM. However, there may be a trade-off between the higher potency of the covalent inhibitors and the side effects and toxicity. It is also reported that, aldehyde-based inhibitors can target endogenous human proteases such as cathepsins as well.^[107] Therefore, search for non-covalent inhibitors is also major direction in developing inhibitors for M^{Pro}.

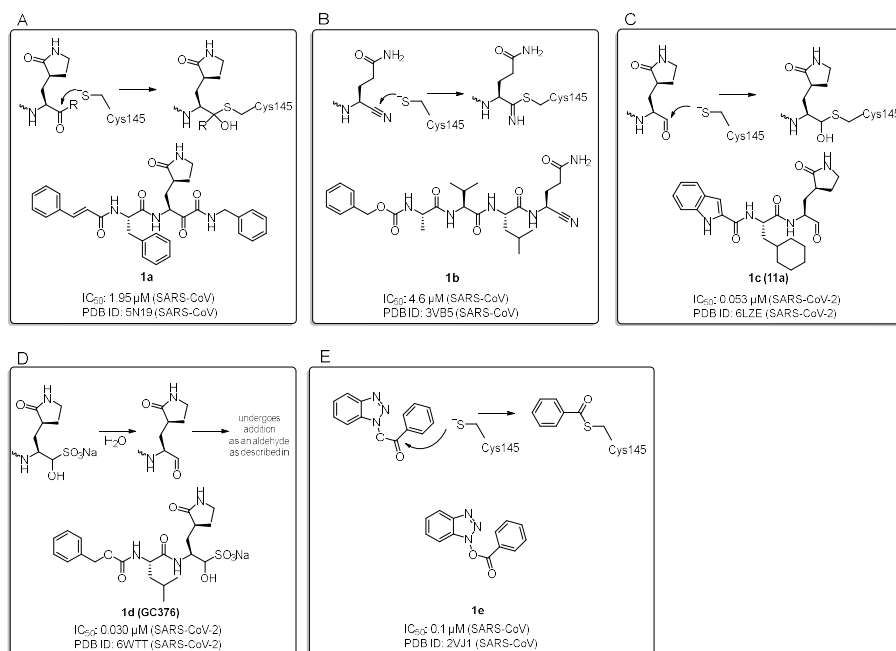


Figure 18 Reversible covalent inhibitors **A)** Ketone inhibitor **B)** Nitrile inhibitor **C)** Aldehyde inhibitor **D)** Bisulfite adduct inhibitor **E)** Active ester inhibitor

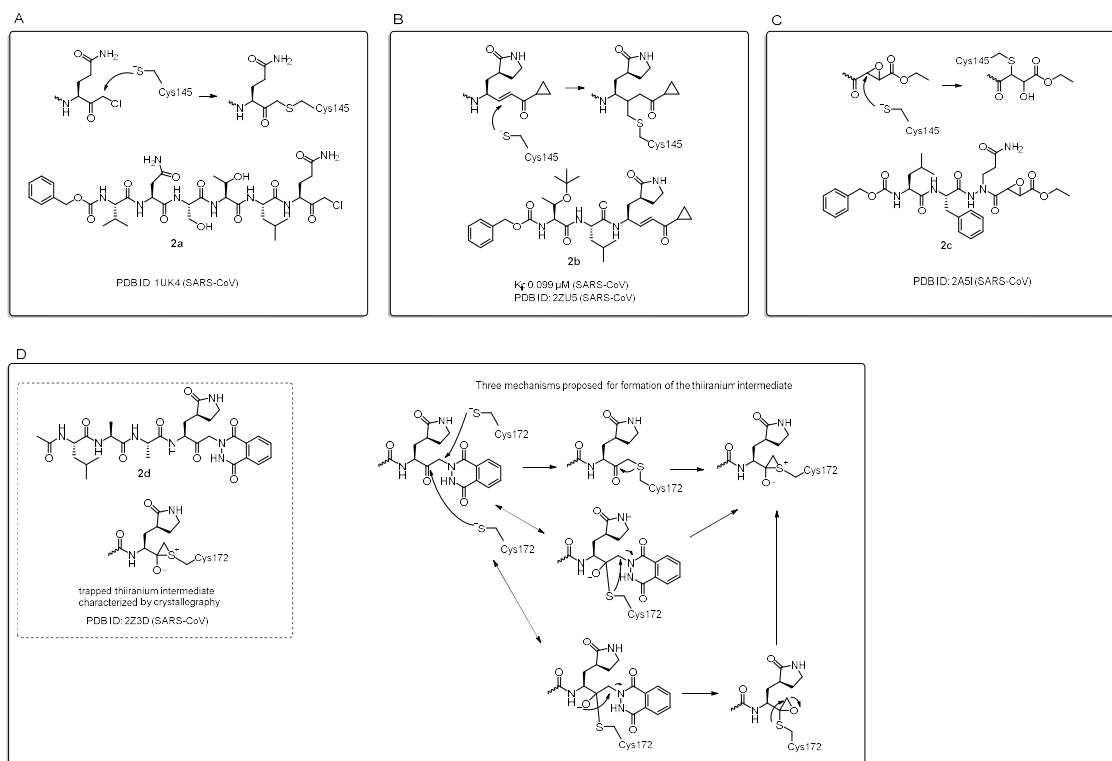


Figure 19 Irreversible covalent inhibitors **A)** Chloroacetamide inhibitor **B)** Michael acceptor inhibitor **C)** Epoxy ketone inhibitor **D)** Phthalic hydrazide leaving group containing inhibitor

Non-covalent Inhibitors

Non-covalent M^{pro} inhibitors are hydrophobic compounds with various different structural motifs. Some examples include benzotriazole (**3a**),^[112] flavone (**3b**),^[113] carmofur (**3c**),^[99] isatin (**3d**),^[114] pyridyl (**3e**)^[115] groups. The IC_{50} or K_i (inhibition constant) values reported for non-covalent inhibitors are generally lower than covalent inhibitors. However, they may offer higher selectivity inside the cell. Moreover, the structural variety of the non-covalent inhibitors is more diverse which may lead to more options for optimizing the structure for improved pharmacokinetics.

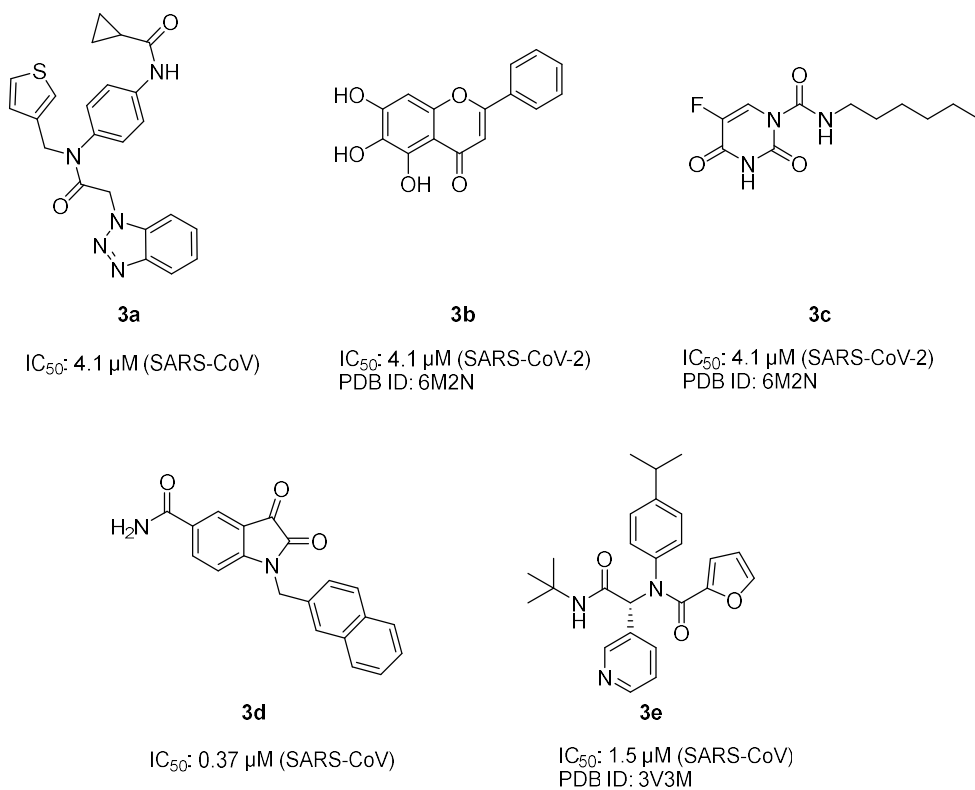


Figure 20 Examples for non-covalent M^{pro} inhibitors

CHAPTER II *2

OPTIMAL SYNTHESIS OF ANTIBODY-DRUG CONJUGATES

2.1 Introduction to CuAAC Reaction

Once published in 2001,^[116] CuAAC reaction (Figure 21A) emerged as a new way of conducting chemistry research. It had a wide-ranging impact from material science to biological sciences. In the 2001 paper, Sharpless referenced Guida's prediction^[117] which estimates the total number of reasonable drug candidates as 10^{62} - 10^{63} different organic compounds thermodynamically stable at room temperature and stable in aqueous solution. Therefore, Sharpless argues that if researchers focus on easy-to-reach scaffolds there will still be plenty of discrete compounds that can have the desired function. This laid the fundamental idea behind 'click chemistry' which is simply the set of reactions that allow researchers to access the easy to reach scaffolds.

CuAAC has been the flagship reaction of click chemistry and is widely studied (Figure 21A). Initially, its relatively slow kinetics and high toxicity (due to the presence of copper) limited its use in biological systems. However, with copper binding ligands the reaction is accelerated and its toxicity became low enough to carry out in live cells

² OPTIMAL SYNTHESIS OF ANTIBODY-DRUG CONJUGATES is readapted from a published report in Bioorganic & Medicinal Chemistry. DOI number: 10.1016/j.bmc.2020.115808.

without imparting an observable toxicity (Figure 21B).^[118]

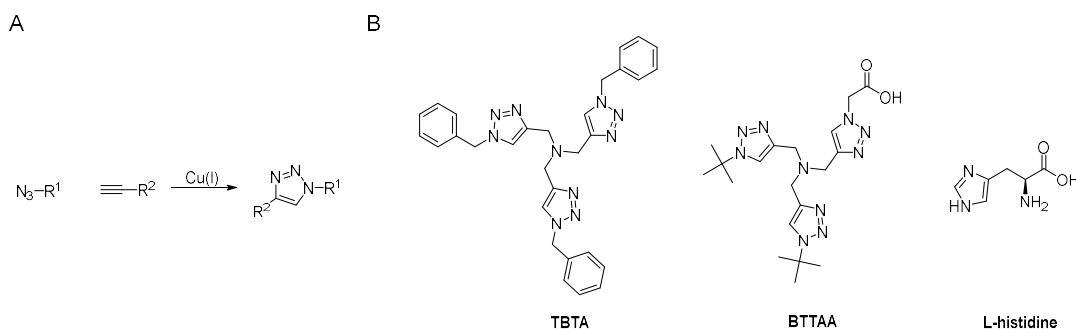


Figure 21 A) CuAAC Reaction B) Examples of copper-binding ligands used in CuAAC

The CuAAC mechanism has been studied extensively over the last two decades.

Although we have learned a lot about this reaction mechanism, our understanding of the CuAAC mechanism is still evolving. In 2013,^[119] an isotope crossover study showed evidence for the involvement of a second copper atom in the alkyne binding. In a later study in 2015,^[120] dicopper intermediate was isolated and proposed the mechanism as shown in Figure 22. According to this mechanism, both mono and dicopper alkyne intermediates are catalytically active while the latter leads to faster product formation. Moreover, it was initially thought the only catalytically active oxidation state of copper is Cu(I); however, a recent cyclic voltammetry study showed the coexistence of Cu(I) and Cu(II) ions.^[121] Another study reported that the addition of Cu(II) ions increased the rate of reaction.^[122] The rate-determining step (RDS) of the CuAAC seems to change with respect to the type of azide and alkyne used. Finn^[123] proposed copper alkyne acetylide formation to be the RDS. A recent study showed that azide-copper binding as

the RDS with using an electron-rich aryl ynamine.^[124]

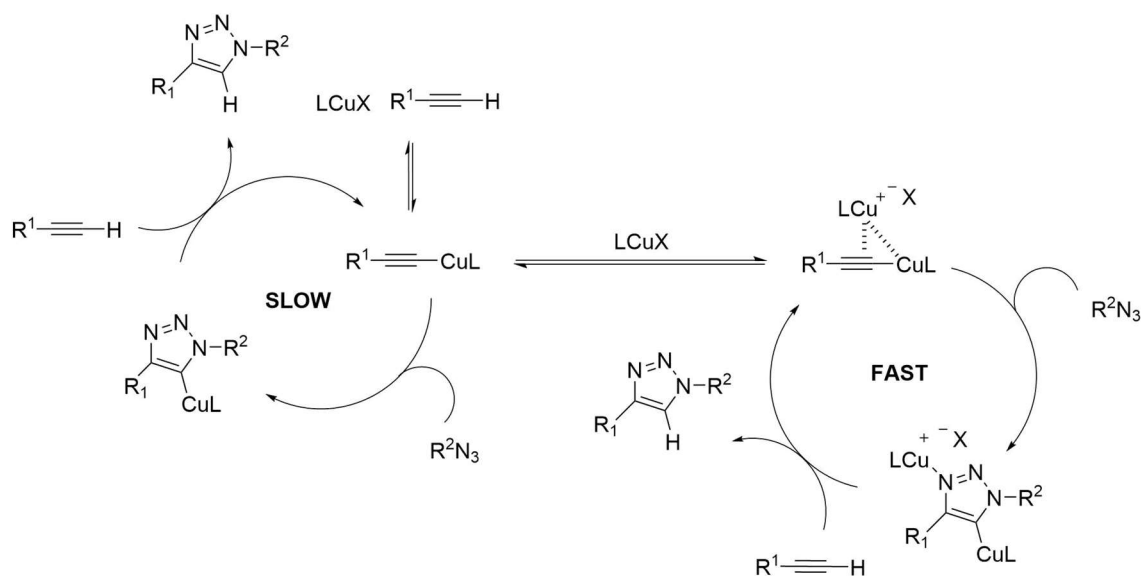


Figure 22 CuAAC mechanism proposed by Bertrand

Site-specific incorporation of azide and alkyne-containing amino acids increased the applicability of CuAAC reaction and Cu-free azide-alkyne cycloaddition reaction for protein conjugation. Several examples were given in Chapter 1.3. These conjugation reactions work fine from a proof-of-concept point of view. However, due to inflexibilities caused by the ncAA incorporation technique, limited room for optimization exists.

As discussed in Chapter 1.3, the antibody to drug ratio (DAR) is an important determinant for pharmacokinetics for ADCs. When using ncAAs, the number of ncAAs that can be incorporated per antibody molecule is limited to 2. This is not a theoretical limitation but, more than 2 drugs per antibody will cause a decrease in antibody production efficiency.

First, this study examines kinetics study to gain more insight into CuAAC. Then,

guided by our small molecule kinetics studies, we synthesized our ADCs and tested them against cancer cell line.

2.2 Experimental

2.2.1 Kinetic Experiments

General Protocol for Kinetics Measurements

Stock solutions for all azide and alkyne compounds were prepared in 100x stock solution in DMSO. BTAA stock solutions were 250x stock solution in DMSO. CuSO₄ stock solutions were 100x stock solution in water. 250 mM sodium ascorbate stock solution in water was used in all analysis.

0.15 M phosphate buffer, pH 7.0, 25% DMSO is added to the reaction well (or cuvette). Next, BTAA, CuSO₄ were added successively followed by azide derivative. After that, ascorbate was added and the mixture was incubated for 60 seconds. Finally, alkyne was added to initiate the click reaction. After addition of each reagent, the reaction mixture was gently mixed by pipetting up and down. Unless otherwise stated, a plate reader was used for all measurements. For 3-azido-7-hydroxyxoumarin (AzCou) excitation at 404 nM and emission at 480 nM was used. For 7-ethynylcoumarin (EtCou) excitation at 320 nM and emission at 400 nM was used. Data were fit to the equation: $Y=Y_0 + (\text{Plateau}-Y_0)*(1-\exp(-K*x))$ by using one-phase association in GraphPad Prism 9.0 software.

Limiting 3-Azido-7-hydroxy-coumarin (AzCou) and Excess Alkyne Reaction

(Conditions in the table below applied except 2-propyn-1-ol where 50 μM CuSO_4 , 300 μM BTAA and 50 μM AzCou, 2.5 mM sodium ascorbate used with 0.25, 0.5, 1, 2, 4 mM propargyl alcohol)

Table 2 Limiting 3-azido-7-hydroxy-coumarin (AzCou) and excess alkyne

Reagent	Final Concentration
CuSO_4	50 μM
BTAA	100 μM
AzCou	10 μM
Sodium ascorbate	2.5 mM
Alkyne Derivative	100, 250, 500, 750, 1000 μM

Limiting 7-Ethynylcoumarin (EtCou) and Excess 2-Azidoethylamine Reaction

Table 3 Limiting 7-Ethynylcoumarin (EtCou) and excess 2-azidoethylamine

Reagent	Final Concentration
CuSO_4	10 μM
BTAA	20 μM
azidoethylamine	100, 200, 300, 400, 500 μM
Sodium ascorbate	2.5 mM in water
EtCou	20 μM

Limiting 2-Azidoethylamine and Excess EtCou Reaction

Table 4 Limiting 2-azidoethylamine and excess EtCou

Reagent	Final Concentration
CuSO ₄	10 μ M
BTAA	20 μ M
azidoethylamine	5 μ M
Sodium ascorbate	2.5 mM in water
EtCou	50, 75, 100, 125, 150 μ M

Limiting Pyridine Azide (AzPy) Concentration and Excess EtCou Reaction

(spectrofluorometer was used)

Table 5 Limiting AzPy concentration and excess EtCou concentrations

Reagent	Final Concentration
CuSO ₄	10 μ M
BTAA	20 μ M
PyAz	5 μ M
Sodium ascorbate	2.5 mM
EtCou	50, 75, 100, 125, 150 μ M

Limiting EtCou and Excess PyAz Reaction

Table 6 EtCou and excess PyAz concentrations

Reagent	Final Concentration
CuSO ₄	10 μ M
BTAA	20 μ M
PyAz	100, 200, 300, 400, 500 μ M
Sodium ascorbate	2.5 mM
EtCou	20 μ M

Limiting EtCou and Excess PyAz Reaction

Table 7 EtCou and excess PyAz concentrations

Reagent	Final Concentration
CuSO ₄	10 μ M
BTAA	20 μ M
PiAz	100, 200, 300, 400, 500 μ M
Sodium ascorbate	2.5 mM
EtCou	20 μ M

Limitng AzCou and Excess Alkyne Derivatives Reaction

Table 8 AzCou and alkyne derivatives

Reagent	Final Concentration
CuSO ₄	50 μ M
BTAA	100 μ M
AzCou	20 μ M
Sodium ascorbate	2.5 mM
Alkyne Derivatives	400 μ M

Limiting AzCou and Different Ethoxyacetylene Reaction

Table 9 AzCou and different ethoxyacetylene concentrations

Reagent	Final Concentration
CuSO ₄	50 μ M
BTAA	100 μ M
AzCou	20 μ M
Sodium ascorbate	2.5 mM
Ethoxyacetylene	100, 200, 300, 400, 500 μ M

Excess EtCou and Different Limiting Azide Reaction

Table 10 Excess EtCou and different limiting azide concentrations

Reagent	Final Concentration
CuSO ₄	50 μ M
BTAA	100 μ M
Azide derivative	20 μ M
Sodium ascorbate	2.5 mM
EtCou	100 μ M

2.2.2 Protein Expression for Protein Kinetics

sfGFP134PrK Expression & Purification:

TOP10 cells were transformed with pBAD-sfGFP134TAG and pEVOL-optmmPyIRS-Y384F plasmids. The cells from an overnight culture were inoculated into 100 mL LB media at 37°C in an orbital shaker. 2 mM PrK (final concentration) was added when the cells reached an OD 600 nm of about 0.5 and the expression was induced by 0.2% (final concentration) L-Arabinose at OD 600 nm. Expression was induced for 8 hours at 37°C with shaking. Then the cells were spun down at 4,000 rpm for 20 mins at 4°C, washed with PBS and pelleted by centrifugation. Cell pellets were stored at -80°C until further use. Cells were thawed on ice and resuspended in 40 mL of 50 mM NaH₂PO₄, 300 mM NaCl, 10 mM imidazole, pH 8 buffer (supplemented with 1 mM PMSF). Next, the cells were sonicated (Branson SFX250 Sonifier W/ 1/2" Horn) in 5 cycles (1 sec on, 0.7 seconds off, 2 minutes at 50% power), 10 minutes rest on ice. Then, cell debris was spun down by centrifuging at 10k rpm for 30 min. Then, the supernatant was loaded on to a column after 1h pre-incubation with 0.5 mL Ni-resin (High affinity Ni charged resin from Genscript). Then, Ni-resin was washed with 40 mL wash buffer (50 mM NaH₂PO₄, 300 mM NaCl, 10 mM imidazole, pH 8 buffer). Then, the target protein was eluted with 10 mL elution buffer (50 mM NaH₂PO₄, 300 mM NaCl, 250 mM imidazole, pH 8 buffer). The protein was dialyzed against PBS buffer at pH 7.4. The final concentration was determined by BCA assay (ThermoFisher) by measuring absorbance at 562 nm.

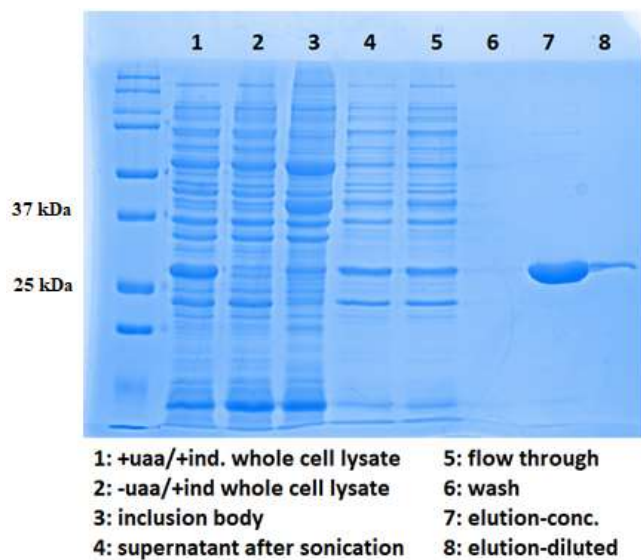


Figure 23 SDS-PAGE analysis of sfGFP134-PrK

2.2.3 sfGFP134PrK Py-Az-Coumarin Reaction Kinetics

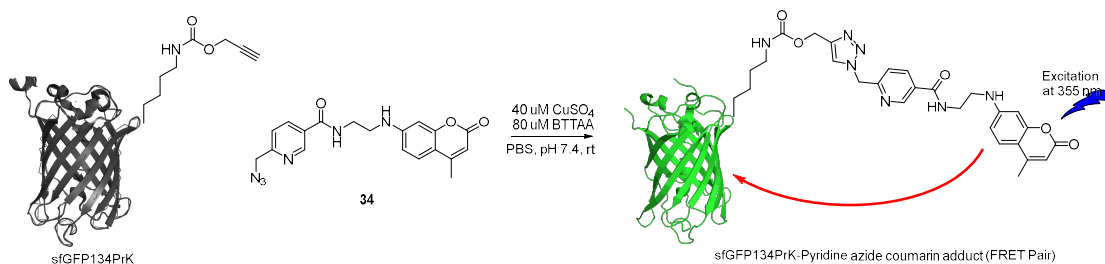


Figure 24 sfGFP134PrK Py-Az-Coumarin (34) Reaction

In order to test the ability of the click reaction to conjugate a protein, site-selective L-propargyl lysine incorporated sfGFP was used as a model system. 1 μ L CuSO₄ (4 mM stock solution in H₂O) and 0.4 μ L BTAA (20 mM stock solution in DMSO) were added to a 70 μ L 1x PBS buffer. Next, 1 μ L (Py-Az coumarin (16), as a 100x stock solution in DMSO) was added. Then, 1 μ L sodium ascorbate (250 mM stock solution in H₂O) was added and reduction reaction carried out for 60 seconds. Then, 30 μ L of sfGFP-134-ProcK (16.78 μ M in PBS) was added, the solution mixed and the kinetics

assay started immediately. The observed rate constant k_{obs} appeared similar as the concentration of the pyridine azide coumarin compound **34** was varied. A fluorometer with excitation at 355 nM and emission at 520 nM was used to monitor the reaction kinetics.

2.2.4 Organic Synthesis

Synthesis of NHS Ester of Chloroacetic Acid (5)

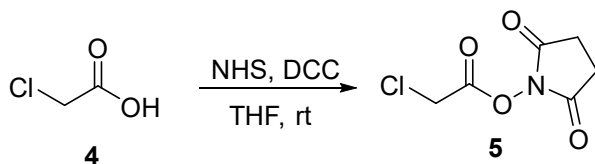


Figure 25 Synthesis of NHS ester of chloroacetic acid (5)

A literature procedure is adapted.^[125] DCC (7.75 g, 37.5 mmol) was dissolved in minimum amount of THF and added dropwise to an ice-cooled solution of NHS (4.35 g, 37.5 mmol) and chloroacetic acid (3.55 g, 37.5 mmol) in 100 mL THF. Reaction stirred for 5 hours, the precipitated urea by-product was filtered off and solid was washed with 25 mL of THF. Then, 25 mL of hexanes were added and filtrate was kept at °C for 12 hours. Additional precipitate filtrated off and filtrate was concentrated in vacuo. Product triturated with 50 mL hexanes and kept for another 2 hours at 4°C. Product is collected by filtration.

¹H NMR (CDCl₃, 400 MHz): δ 2.88 (2H, s), 4.37 (2H, s).

Synthesis of 2-Azidoethyl-1-amine Hydrochloride

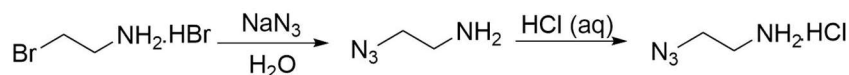


Figure 26 2-azidoethyl-1-amine hydrochloride

A literature protocol^[126] was followed with some modifications. 1.8 g of sodium azide and 2.5 g of 2-bromoethylamine hydrobromide were dissolved in 15 mL water and the solution was heated to 75°C for 12 hours.

Then, the solution is cooled down to ambient temperature and 488 mg NaOH was added. The organic material was extracted with 3 x 25 mL DCM. Combined organic extracts were washed with brine and dried over sodium sulfate.

Following filtration, DCM was evaporated using a rotary evaporator to yield the free amine compound. Then, the free amine compound was stirred in 1 M HCl in dioxane to form the hydrochloride salt. Next, the volatiles were evaporated to yield the target compound.

¹H NMR (D₂O, 400 MHz): 3.19 (2H, s), 3.71 (2H, s)

Synthesis of N-(2-Azidoethyl)-2-chloroacetamide (6)

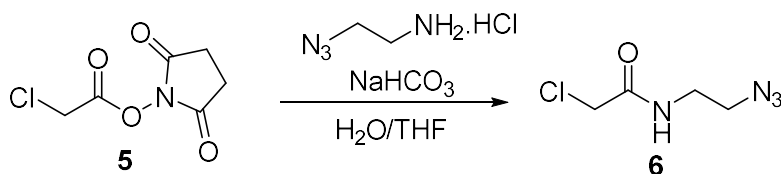


Figure 27 Synthesis of N-(2-azidoethyl)-2-chloroacetamide (6)

A literature procedure is adapted.^[127] 1.0 mmol of 2-azidoethylamine hydrochloride (122 mg, 1.41 mmol) was dissolved in 0.75 mL of water and NaHCO₃ (2.15 mmol, 180 mg) was added. Solution of NHS ester of chloroacetic acid (5, 1.07 mmol, 205 mg) in 3 mL THF was added to mixture and stirred at room temperature for 18 hours. Then, solvents were removed under reduced pressure and diluted to 5 mL with saturated NaHCO₃

solution. The product was extracted with DCM (3 x 10 mL). DCM fractions were combined and washed with 5 mL K₂CO₃ solution at pH 10. Next, the DCM layer was washed with 5 mL 0.1 M HCl solution. Finally, the DCM layer was dried over sodium sulfate and the product (78%) was pure enough to continue to next step.

¹H NMR (CDCl₃, 400 MHz): δ 3.45 (4H, m), 4.04 (2H, s), 6.95 (1H, bs).

Synthesis of N-(2-Azidoethyl)-2-iodoacetamide (7)

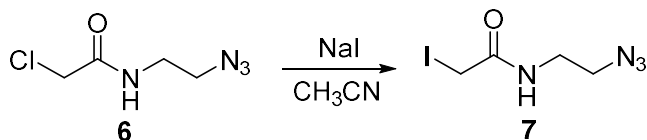


Figure 28 N-(2-azidoethyl)-2-iodoacetamide (7)

A literature protocol was applied.^[128] 1.3 mmol of N-(2-azidoethyl)-2-chloroacetamide (7, 211 mg, 1.28 mmol) in 4 mL ACN was treated with 3 eq. of sodium iodide (584 mg, 3.9 mmol). The reaction ran for 24 hours. After completion of reaction, the solvent was evaporated under reduced pressure. The reaction mixture was suspended in 10 mL ethyl acetate and washed with 10% sodium thiosulfate aqueous solution. Then, the organic layer was washed with water and brine sequentially. Finally, the organic layer was dried on sodium sulfate and evaporated under reduced pressure to yield the final product with 65% yield.

¹H NMR (CDCl₃, 400 MHz): δ 3.48 (4H, m), 3.72 (2H, s), 6.4 (1H, bs).

Synthesis of N-(Prop)-L-valine (9)

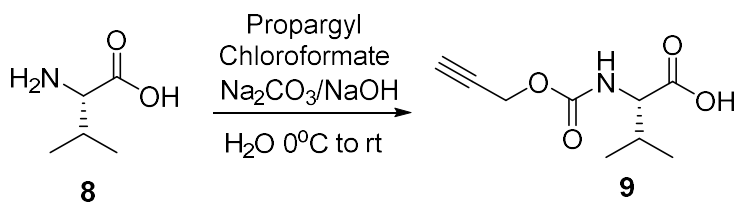


Figure 29 Synthesis of N-(Prop)-L-valine (9)

5 mmol L-Valine was dissolved in 5 mL 1 M NaOH. 2.6 mmol Na₂CO₃ was added after the solution was cooled to 0°C on ice bath. 530 μL propargyl chloroformate was added dropwise to ice-cooled solution. After 10 mins, ice bath is removed and reaction ran for 3 hours at room temperature. When the reaction finished, reaction mixture was poured into separatory funnel and organic material extracted with 10 mL diethyl ether. Then, aqueous layer acidified to pH 2 with addition of 3M HCl dropwise (on ice), white precipitate is formed. Then, the mixture extracted 3 x 10 mL of ethyl acetate. Ethyl acetate fractions combined, dried on sodium sulfate and evaporated to give 830 mg pure compound (81%).

¹H NMR (CDCl₃, 300 MHz): δ 0.97 (6H, dd, *J*=7.17 Hz, *J*=22.15 Hz), 2.24 (1H, m), 2.49 (1H, s), 4.33 (1H, dd, *J*=7.17 Hz, *J*=22.15 Hz), 4.70 (2H, s), 5.28 (1H, d, *J*=8.15 Hz). For NMR spectrum, see Appendix, Figure 121.

Synthesis of NHS Ester of N-(Prop)-L-valine (10)

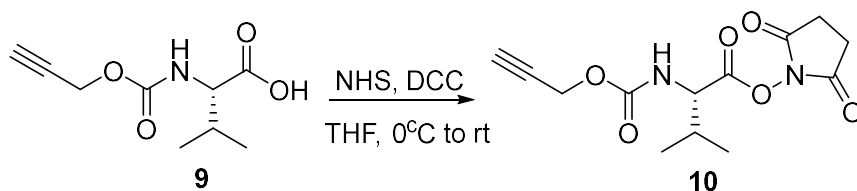


Figure 30 Synthesis of NHS ester of N-(Prop)-L-valine (10)

A literature protocol is adapted.^[129] A round-bottom flask was charged with a stir bar, 995 mg of Proc-Val-OH, 575 mg NHS. Then 12 mL THF was added and mixture was cooled down to 0°C. Then 1032 mg of DCC was added portion-wise. The reaction mixture stirred overnight. The next day, suspension was filtered and filtrate is evaporated under vacuum. Then 10 mL DCM was added and mixture stand 1hr at room temperature. A second filtration was performed and the solvent was evaporated to give 1.35 g compound with minor impurities (91%). Compound used in the next step without further purification.

¹H NMR (CDCl₃, 300 MHz): δ 1.06 (6H, dd, *J* = 6.67 Hz, *J* = 11.82 Hz), 2.29-2.41 (1H, m), 2.50 (1H, t, *J* = 2.45 Hz), 2.85 (4H, s) 4.70 (2H, s), 5.31 (1H, br s). For NMR spectrum, see Appendix, Figure 122.

Synthesis of Proc-Val-Cit-OH (11)

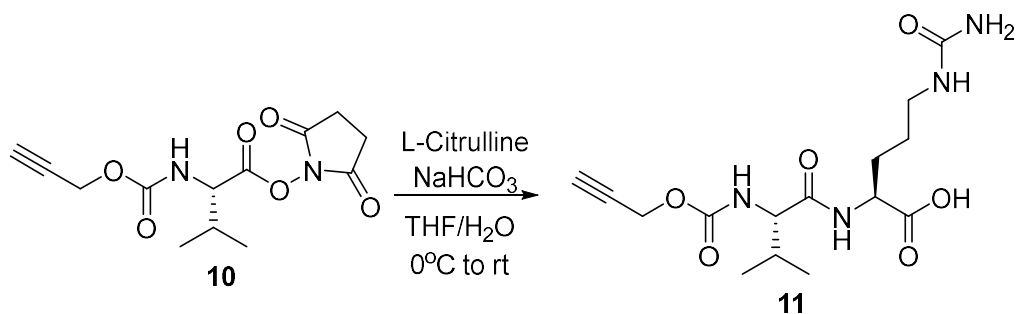


Figure 31 Synthesis of Proc-Val-Cit-OH (11)

A literature protocol is adapted.^[129] To a solution of 1.60 g Proc-Val-OSu in 16 mL THF at 0 C was added the solution of 1.51 g L-Citrulline and 0.73 g NaHCO₃ in 16 mL water. 15 minutes later ice bath was removed and reaction proceeded at room temperature. After 12 hours, 13 mL water and 2 mL saturated aqueous NaHCO₃ solution added. The mixture was washed with 50 mL EtOAc. Next, the aqueous layer was acidified to pH 3 with 10% HCl. The resulting suspension extracted with (5x50 mL) 10% *i*-PrOH/EtOAc to give 760 mg of pure compound with *i*-PrOH impurities (55%).

¹H NMR (CD₃OD, 300 MHz): δ 0.97 (6H, dd, $J = 6.68$ Hz, $J = 11.14$ Hz), 1.49-2.15 (5H, m), 3.13 (2H, t, $J = 7.01$ Hz), 3.97 (1H, br s), 4.40 (1H, br s), 4.66 (2H, br s) For NMR spectrum, see Appendix, Figure 123.

Synthesis of Proc-Val-Cit-PABOH (12)

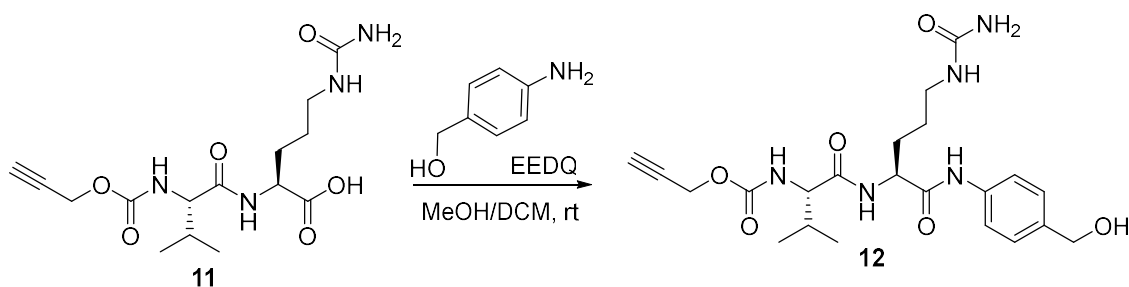


Figure 32 Synthesis of Proc-Val-Cit-PABOH (12)

3.4 mmol of Proc-Val-Cit-OH and 2 eq. alcohol (6.8 mmol, 837 mg) were dissolved in 54 mL DCM/MeOH (36 mL/18 mL) and stirred for 30 minutes. Then EEDQ (6.8 mmol, 1682 mg) was added and stirred for 15 hours. Then, the volatiles were evaporated under vacuum, and the mixture subjected to column chromatography. Target compound was eluted with DCM/MeOH (8/2) to give 1128 mg of pure compound (72%).

Synthesis of Proc-Val-Cit-PAB-PNP (13)

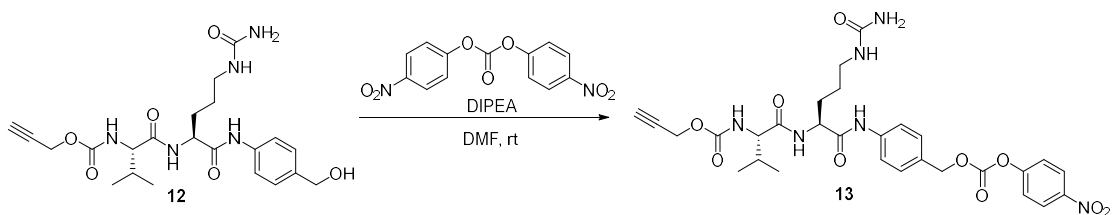


Figure 33 Synthesis of Proc-Val-Cit-PAB-PNP (13)

654 mg of alcohol compound was dissolved in 15 mL of anhydrous DMF. 2 eq. of bis(4-nitrophenyl) carbonate (862 mg) was added to the solution. Then 1.5 eq. DIPEA (194 μ L) was added and reaction was stirred for 12 hours. Next, DMF was evaporated and the compound was dissolved in MeOH and dry loaded to silica. Column chromatography

with MeOH/DCM (gradient from 5% to 12%) applied to obtain 710 mg of pure final compound (80%)

^1H NMR ($\text{C}_2\text{D}_6\text{SO}$, 300 MHz): δ 0.86 (6H, dd, $J = 6.82$ Hz, $J = 13.04$ Hz), 1.29-2.03 (5H, m), 3.91 (1H, br s), 3.97 (1H, br s), 4.40 (1H, br s), 4.63 (2H, s), 5.24 (2H, s), 5.42 (2H, s), 5.76 (1H, s), 7.41 (d, 2H, $J = 8.50$ Hz), 7.57 (d, 2H, $J = 8.89$ Hz), 7.64 (d, 2H, $J = 7.75$ Hz), 8.17 (d, 1H, $J = 8.71$ Hz), 8.31 (d, 2H, $J = 9.42$ Hz), 10.14 (1H, s). For NMR spectrum, see Appendix, Figure 124.

Synthesis of Proc-Val-Cit-PAB-MMAE (14)

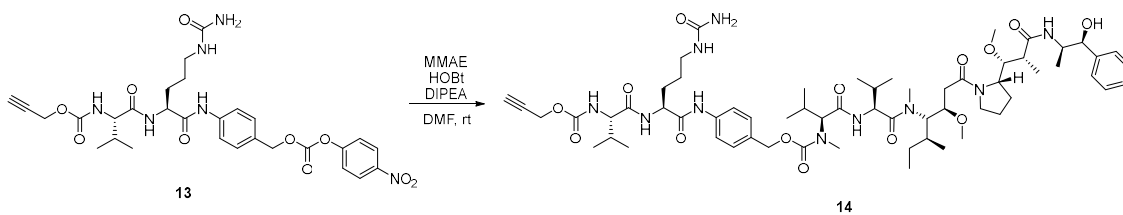


Figure 34 Synthesis of Proc-Val-Cit-PAB-MMAE (14)

4 mg of Proc-Val-Cit-PAB-PNP was dissolved in 500 μL anhydrous DMF. Then, 5 mg of MMAE was added along with 1.1 eq. HOBt and 1.5 eq. of DIPEA. After 16 hours, DMF was evaporated and the material was dissolved in DMSO. Compound was purified on C18 column, flow rate 1 mL/min. Solvent programming: 0-5 mins: 0.1% TFA, 10% ACN in H₂O (isocratic) 5-85 mins: 0.1% TFA, 90% ACN in H₂O (gradient)

1.0 mg target compound was obtained. Calculated $[\text{M}+\text{H}]$: 1205.7185 Da Observed $[\text{M}+\text{H}]$: 1205.7174 Da. For HPLC chromatogram, see Appendix, Figure 125.

Synthesis of Tert-butyl prop-2-yn-1-yl propane-1,3-diyldicarbamate (16)

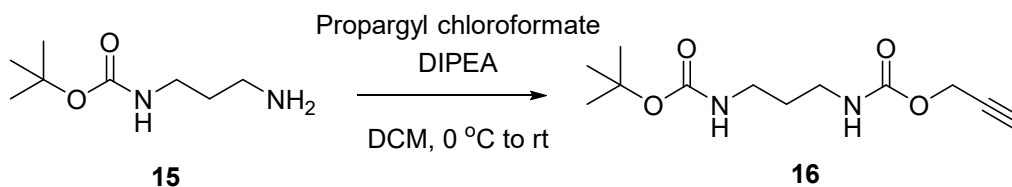


Figure 35 Synthesis of tert-butyl prop-2-yn-1-yl propane-1,3-diyl dicarbamate (**16**)

Amine compound **15** (1300 mg, 7.23 mmol) and DIPEA 1500 μ L, 8.69 mmol) were dissolved in 9 mL DCM and placed on ice. After 10 minutes, propargyl chloroformate (775 μ L in 9 mL DCM) was added to the solution dropwise. Subsequently, the reaction was incubated at room temperature overnight. The next day, the reaction solution was washed with 20% citric acid, water, sat. NaHCO_3 and brine respectively. Then, TLC was used with 40/60 (EtOAc/Hexane) and the target compound was purified by flash chromatography.

^1H NMR (CDCl_3 , 400 MHz): δ 1.42 (9H, s), 1.62 (2H, quint, $J = 6.25$ Hz), 2.45 (1H, s), 3.16 (2H, t, $J = 6.16$ Hz), 3.22 (2H, t, $J = 6.67$ Hz), 4.59 (2H, s)

^{13}C NMR (CDCl_3 , 75 MHz): δ 14.19, 30.48, 37.75, 52.36, 74.54, 78.37, 79.35, 155.81, 156.49

For NMR spectrums, see Appendix, Figures 126, 127.

Synthesis of Prop-2-yn-1-yl (3-aminopropyl)carbamate hydrochloride (17)

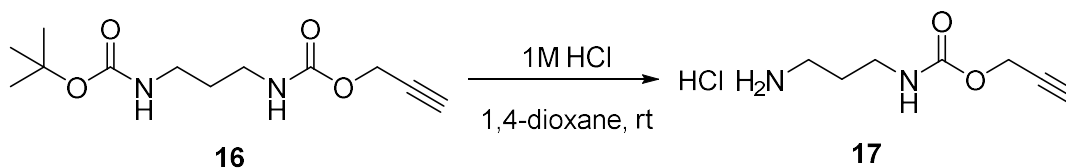


Figure 36 Synthesis of prop-2-yn-1-yl (3-aminopropyl)carbamate hydrochloride (**17**)

Boc-protected compound **16** (595 mg, 2.32 mmol) was dissolved in 9 mL of 1M HCl in 1,4-dioxane was added and the reaction stirred for 2 hours at room temperature. Finally, TLC was used to confirm the completion of the reaction, and the contents were evaporated multiple times to remove residual TFA. Target compound was recovered in quantitative yield.

^1H NMR (D_2O , 400 MHz): δ 1.89 (2H, quint, $J=7.19$ Hz), 2.92 (1H, s), 3.05 (2H, q, $J=7.68$ Hz), 3.26 (2H, q, $J=6.29$ Hz), 4.69 (2H, s) For NMR spectrum, see Appendix, Figure 128.

Synthesis of Prop-2-yn-1-yl (3-(2-bromoacetamido)propyl)carbamate (Bromoacetyl Alkyne) (**18**)

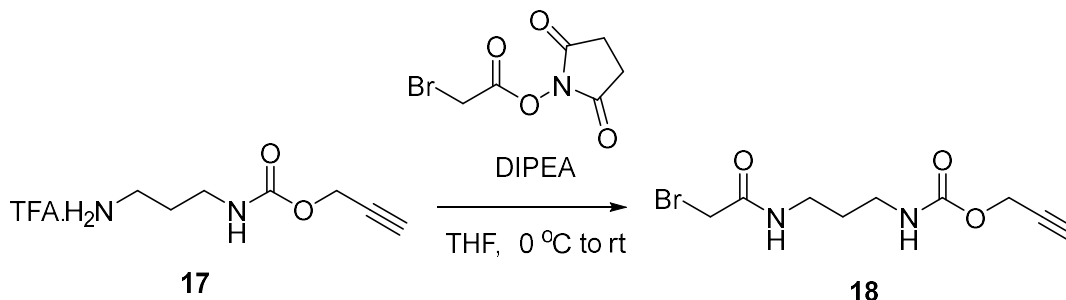


Figure 37 Synthesis of prop-2-yn-1-yl (3-(2-bromoacetamido)propyl)carbamate (Bromoacetyl alkyne) (**18**)

Amine compound **17** (626 mg, 2.32 mmol) was dissolved with 0.8 mL DIPEA (4.64 mmol) in 4 mL THF and placed on ice (DIPEA was added until the litmus paper test gave green color). Then, bromo-OSu ester (821 mg, 3.48 mmol) was added in 4 mL THF solution and the reaction stirred on ice for another 1.5 hr. Then, the reaction was stirred for another 30 minutes at room temperature. Subsequently, TLC did not show signs of additional product formation and the DCM was added to the reaction mixture. Then the organics were washed with 20% citric acid, water, saturated NaHCO₃ solution and brine. Then, the organic phase was dried over sodium sulfate and flash chromatography applied at 70/30 (EtOAc/Hexane) to yield pure compound.

HRMS (m/z): calculated for C₉H₁₃BrN₂O₃ [M+H]⁺ 277.0182, found 277.0183.

¹H NMR (CDCl₃, 400 MHz): δ 1.70 (2H, quint, *J* = 6.26 Hz), 2.47 (1H, s), 3.25 (2H, q, *J* = 6.21 Hz), 3.36 (2H, q, *J* = 6.50 Hz), 3.87 (2H, s), 4.67 (2H, s), 5.26 (1H, s), 6.94 (1H, s). For NMR spectrum, see Appendix, Figure 129.

Synthesis of 6-(Hydroxymethyl)nicotinic acid methyl ester (**20**)

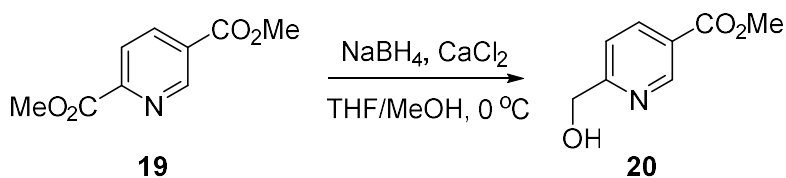


Figure 38 Synthesis of 6-(hydroxymethyl)nicotinic acid methyl ester (**20**)

6-(azidomethyl)nicotinic acid was prepared according to published methods.^[130] For the synthesis of 6-hydroxymethyl-nicotinic acid methyl ester intermediate, NaBH₄ (1.06 g, 27.92 mmol) was added portion-wise to the mixture of 2,5-pyridinedicarboxylic

acid dimethyl ester (2.00 g, 11.16 mmol) and CaCl₂ (4.8 g, 44.64 mmol) in anhydrous tetrahydrofuran/anhydrous methanol (20 mL/40 mL) solvent mixture. The reaction was stirred on ice for 3 hours. Next, unreacted NaBH₄ was quenched by adding 40 mL of ice-cold water to the reaction mixture. The target compound was extracted with chloroform (3 × 80 mL) and organic layers were combined and dried with MgSO₄. Then, the dried solution was concentrated in vacuo. Then, flash chromatography was applied to provide 6-hydroxymethyl-nicotinic acid methyl ester as a white solid (186 mg; 27% yield).

¹H NMR (CDCl₃, 400 MHz): δ 3.96 (3H, s), 4.83 (2H, d, *J* = 4.35 Hz), 7.35 (1H, d, *J* = 8.19 Hz), 8.29 (1H, dt, *J* = 8.11, *J* = 2.0 Hz), 9.17 (1H, s)

Synthesis of 6-(Azidomethyl)nicotinic Acid Methyl Ester (**21**)

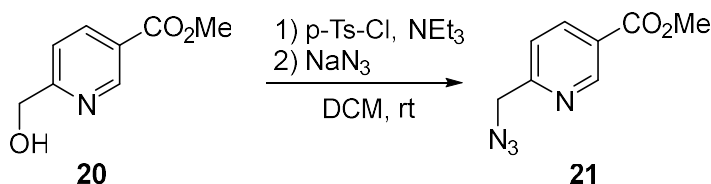


Figure 39 Synthesis of 6-(azidomethyl)nicotinic acid methyl ester (**21**)

Para-toluenesulfonyl chloride (TsCl, 763 mg, 3.96 mmol) and triethylamine (1819 μL, 13.2 mmol) were added to a solution of 6-hydroxymethyl-nicotinic acid methyl ester (440 mg, 2.64 mmol) in dichloromethane (DCM, 60 mL). The reaction was stirred for 2 hours at room temperature. TLC still showed starting alcohol material and *p*-tosyl chloride, and consequently, additional tosyl chloride (381 mg, 1.98 mmol) was added to speed up the reaction. 450 μL of triethylamine was also added. However, it did not seem to affect the reaction rate significantly. Then, the reaction was heated to 40°C and after 2

hours no alcohol starting material was observed. Then, 12 mL DMF was added to the reaction mixture and DCM in reaction flask was evaporated under mild vacuum. Then, sodium azide (1300 mg, 20 mmol) was added. The reaction was further reacted 40°C for 14 hours. Then, reaction mixture was diluted with ethyl acetate and water. The aqueous layer was further extracted with ethyl acetate three times. The combined organic layers were washed with brine, dried over sodium sulfate, and concentrated in vacuo. The residual mixture was purified by flash chromatography to yield azide as light yellow solid (179 mg, 35% yield).

^1H NMR (CDCl_3 , 400 MHz): δ 3.96 (3H, s), 4.57 (2H, s), 7.45 (1H, d, $J = 8.13$ Hz), 8.33 (1H, dt, $J = 8.11$ Hz, $J = 2.1$ Hz), 9.18 (1H, s)

Synthesis of 6-(Azidomethyl)nicotinic Acid (**22**)

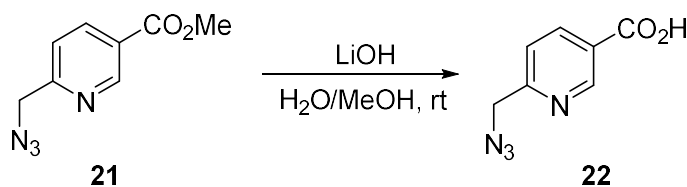


Figure 40 Synthesis of 6-(azidomethyl)nicotinic acid (**22**)

Pyridine ester **21** was dissolved in 4 mL methanol. 1.0 M solution of LiOH in water (2.8 mL, 2.8 mmol) was added to the mixture and stirred for 60 mins. Next, acetic acid (95 μL) was added and the mixture was loaded directly onto a silica gel (60 g of silica used) column equilibrated with ethyl acetate + 1% acetic acid. The desired compound eluted after about 100 mL of eluant addition, and volatile solvents were evaporated to yield the desired compound with some acetic acid (~6%). To remove the remaining acetic acid,

the compound was suspended in ~20 mL water and evaporated again. 179 mg target compound was isolated.

^1H NMR (CDCl_3 , 400 MHz): δ 3.96 (3H, s), 4.57 (2H, s), 7.45 (1H, d, $J = 8.13$ Hz), 8.33 (1H, dt, $J = 8.11$ Hz, $J = 2.1$ Hz), 9.18 (1H, s)

Synthesis of 6-(Azidomethyl)nicotinate Hydroxysuccinimide Ester (**23**)

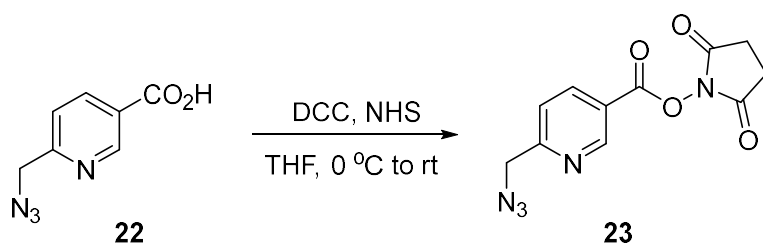


Figure 41 Synthesis of 6-(azidomethyl)nicotinate hydroxysuccinimide ester (**23**)

Nicotinic acid **22** (72 mg, 0.4 mmol), NHS (46 mg, 0.4 mmol) and 3 mL THF were added to a round-bottomed flask. The flask was cooled on ice, DCC (82 mg, 0.4 mmol) was added dropwise in 0.5 mL THF solution and the reaction started. Then, an additional 0.5 mL THF was added and after 30 minutes the reaction was continued for another 2 hours at room temperature. After 2 hours, TLC showed complete conversion and solid appeared to have precipitated out. Then, the suspension was passed through a filter to remove the solid. Then, the flow through was concentrated and evaporated to yield the crude material which was used without further purification in the next step.

^1H NMR (CD_3OD , 400 MHz): δ 4.60 (2H, s), 7.58 (1H, d, $J = 8.17$ Hz), 8.41 (1H, dt, $J = 8.13$ Hz, $J = 2.2$ Hz), 9.12 (1H, d, $J = 1.9$ Hz)

Synthesis of 2,5-Dioxopyrrolidin-1-yl (tert-butoxycarbonyl)-L-valinate (25)

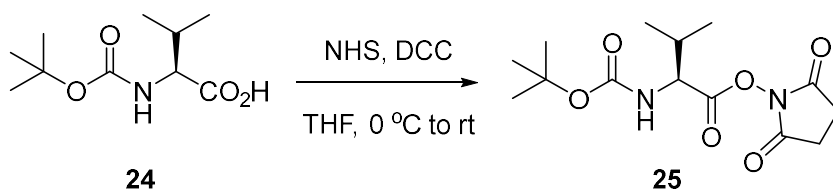


Figure 42 Synthesis of 2,5-dioxopyrrolidin-1-yl (tert-butoxycarbonyl)-L-valinate (25)

A previously described protocol was slightly modified.^[129] A round-bottomed flask was charged with Boc-Val-OH (8.7 g, 40 mmol), NHS (5.6 g, 44 mmol), 80 mL THF and the mixture was cooled on ice with stirring. Then, DCC (8.67 g in 20 mL THF) was added dropwise to reaction solution. After 20 minutes the reaction was warmed to room temperature and carried out for 16 hours. On the following day, the precipitated urea byproduct was filtered off and filtrate was evaporated. The oil was dissolved in 80 mL DCM. The mixture was allowed to incubate at room temperature for 1 hour, and then filtered to remove additional precipitated byproduct. The filtrate was evaporated and ¹H NMR spectrum was analyzed in CDCl₃. The target compound was obtained with a 72% yield.

¹H NMR (CDCl₃, 400 MHz): δ 1.06 (6H, m), 1.46 (9H, s), 2.30 (1H, quint, *J* = 6.06 Hz), 2.84 (4H, s), 4.60 (1H, m), 5.00 (1H, s)

Synthesis of Boc-Valine-Citrulline-OH (**26**)

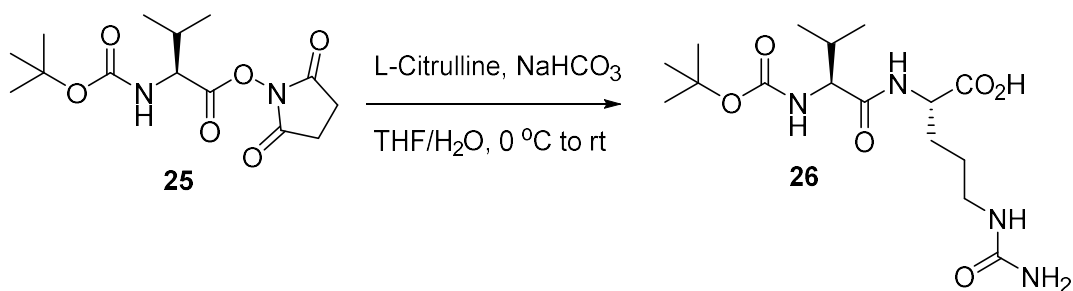


Figure 43 Synthesis of Boc-Valine-Citrulline-OH (**26**)

A solution of L-Citrulline (4.36 g, 24.9 mmol) and NaHCO₃ (2.09 g, 24.9 mmol) in 48 mL water was added to a solution of Boc-Val-OSu (5.2 g, 16.6 mmol) in THF (48 mL) at room temperature, the reaction kept on ice for the 30 minutes and then stirred at room temperature for 16 hours. Then, 4 mL saturated aqueous NaHCO₃ in 40 mL water was added to the solution and the reaction mixture was washed with ethyl acetate once. Then, the pH was adjusted to 2 by addition of 10% HCl, dropwise on ice.

The solution remained clear throughout the pH adjustment and when the desired pH was obtained, the target compound was extracted with 10% *i*-PrOH in 5 x 100 mL ethyl acetate. Then, the organic phases were collected and washed with brine solution at pH 2. Then, the solution was dried over sodium sulfate, evaporated to yield the 5.4 grams of the target compound (87% yield).

¹H NMR (CD₃OD, 400 MHz): δ 0.88-1.01 (6H, m), 1.45 (9H, s), 1.55 – 2.01 (6H, m), 3.07 – 3.15 (2H, m), 3.89 (1H, d, br s), 4.36 (1H, br s)

Synthesis of Boc-Valine-Citrulline-PABOH (27)

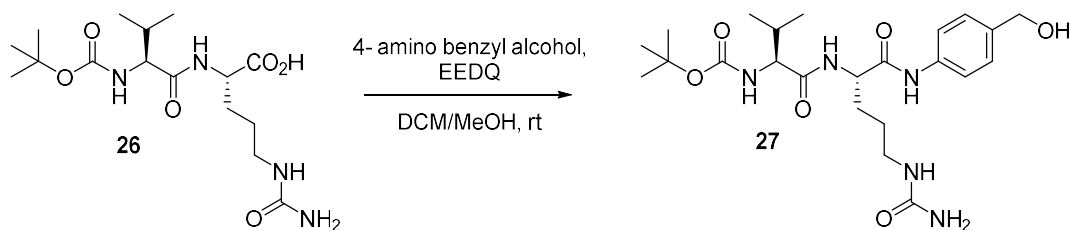


Figure 44 Synthesis of Boc-Valine-Citrulline-PABOH (27)

Boc-Val-Cit-OH (1.0 g, 2.67 mmol) and 4-amino benzyl alcohol (0.658 g, 5.34 mmol) were dissolved in 30 mL DCM/15 mL MeOH at room temperature and stirred for 30 minutes. Then, EEDQ (1.32 g, 5.34 mmol) was added and reaction stirred for 18 hours. The following day, the reaction contents were absorbed on silica and directly loaded onto a silica gel column. Elution started from ethyl acetate and the target compound was eluted using 10% MeOH in ethyl acetate. The amount of the product was 0.8 grams (63% yield).

$^1\text{H NMR}$ (CD_3OD , 400 MHz): δ 0.97 (6H, m), 1.46 (9H, s), 1.52 – 2.14 (6H, m), 3.07 – 3.27 (2H, m), 3.91 (1H, d, $J = 6.82$ Hz), 4.53 (1H, dd, $J = 8.59$ Hz, 5.23 Hz), 4.57, (2H, s), 7.31 (2H, d, $J = 8.59$ Hz), 7.57 (2H, d, $J = 8.69$ Hz)

Synthesis of TFA salt of H₂N-Valine-Citrulline-PABOH (**28**)

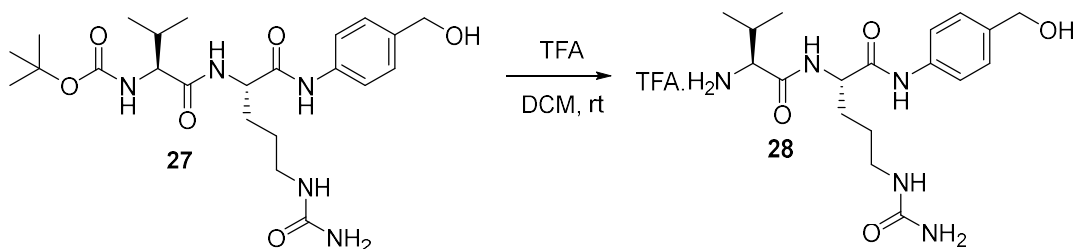


Figure 45 Synthesis of TFA salt of H₂N-Valine-Citrulline-PABOH (**28**)

A round-bottomed flask was charged with Boc-compound **27** (161 mg, 0.34 mmol) and DCM (2.0 mL). Then, 200 μ L TFA was added dropwise and the reaction started at room temperature. Both the starting material and product was not very soluble in the reaction solution. Therefore, sonication was applied every 30 minutes to ensure complete conversion. The reaction was run for 4 hours. Consumption of starting material was confirmed by TLC EtOAc/MeOH (9/1). The product was obtained in quantitative yield.

¹H NMR (CD₃OD, 400 MHz): δ 1.07 (3H, d, $J = 6.92$ Hz), 1.10 (3H, d, $J = 6.97$ Hz), 1.52 – 1.95 (4H, m), 2.24 (1H, m), 3.15 (2H, m), 3.74 (1H, d, $J = 5.82$ Hz), 4.60 (1H, dd, $J = 9.8$ Hz, 4.6 Hz), 4.64, (2H, s), 7.38 (2H, d, $J = 8.54$ Hz), 7.59 (2H, d, $J = 8.14$ Hz)

Synthesis of Py-Az-Val-Cit-PABOH (29)

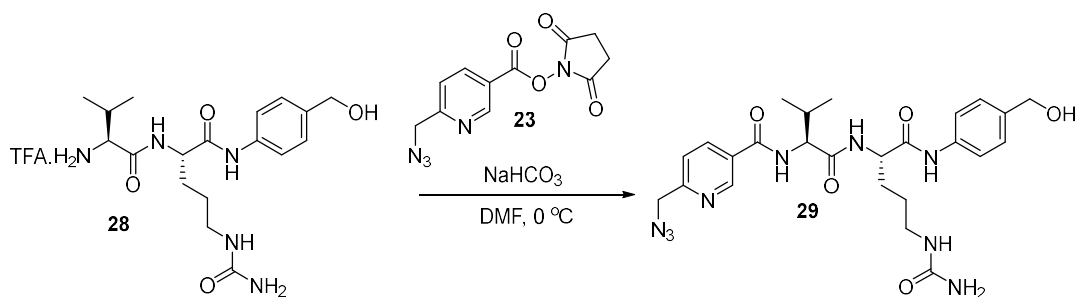


Figure 46 Synthesis of Py-Az-Val-Cit-PABOH (29)

Amine compound 11 (115 mg, 0.277 mmol) was dissolved in 0.5 mL DMF with DIPEA (57 μ L, 0.33 mmol). Then, additional DIPEA added until the solution turned slightly basic (litmus paper). Then OSu ester solution (83 mg) was suspended in 1 mL acetonitrile. After formation of the precipitate, the OSu ester added to the amine compound as a suspension, followed by incubation at room temperature. The reaction progress was analyzed using TLC (EtOAc/MeOH: 9/1) and litmus paper. Additional DIPEA was added as the reaction turned acidic. Then, after several hours the reaction appeared to be complete, and the reaction mixture was absorbed onto silica and added to the column (about 100 g of silica). Following chromatography, pure compound was obtained at a yield of 41%.

¹H NMR (DMSO-d₆, 400 MHz): δ 0.95 (6H, m), 1.31 – 1.76 (4H, m), 2.13 (1H, h, J = 7.2 Hz) 2.88-3.08 (2H, m), 4.32-4.44 (2H, m), 4.60 (2H, s), 4.71 (2H, s), 5.43 (2H, s), 6.00 (1H, t, J = 5.6 Hz), 7.36 (2H, d, J = 8.9 Hz), 7.54 (1H, d, J = 8.1 Hz), 7.60 (2H, d, J = 8.5 Hz), 8.26 (1H, dd, J = 8.1, 3.0 Hz) 8.31 (1H, d, J = 7.3 Hz), 8.60 (1H, d, J = 7.6), 9.02 (1H, d, J = 2.40 Hz), 10.10 (1H, s)

HRMS (m/z): calculated for C₂₅H₃₃N₉O₅ [M+H]⁺ 540.2677, found 540.2673. For NMR spectrum, see Appendix, Figure 130.

Synthesis of Py-Az-Val-Cit-PAB-PNP Carbonate (30)

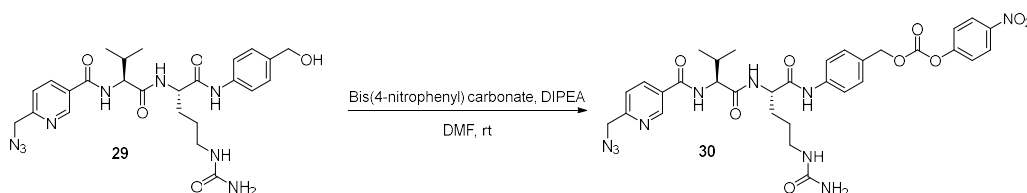


Figure 47 Synthesis of Py-Az-Val-Cit-PAB-PNP Carbonate (30)

Alcohol compound 12 (85 mg, 0.159 mmol) and PNP carbonate (98 mg, 0.318 mmol) were dissolved in 1.3 mL DMF. Then, 46 μ L DIPEA (0.239 mmol) was added and the reaction stirred for 16 hours. Then the contents were evaporated and the reaction contents were loaded to silica and chromatography performed at 9/1 (EtOAc/MeOH) composition. The desired compound was eluted at 15% MeOH in EtOAc. Target product with some NHS contamination obtained in 48% yield.

¹H NMR (DMSO-d₆, 400 MHz): δ 1.05 (6H, d, J = 6.7 Hz), 1.32 – 1.82 (4H, m), 2.14 (1H, h, J = 7.2 Hz), 2.92-3.09 (2H, m), 4.40 (2H, m), 4.59 (2H, s), 5.24 (2H, s), 5.41 (2H, s), 5.98 (1H, t, J = 5.3 Hz) 7.40, (2H, d, J = 8.56 Hz), 7.51-7.60 (2H, m), 7.64 (2H, d, J = 8.38 Hz), 8.24-8.34 (3H, m), 8.57 (1H, J = 7.4 Hz), 9.02 (1H, d, J = 1.7 Hz), 10.10 (1H, s)

HRMS (m/z): calculated for C₃₂H₃₆N₁₀O₉ [M+H]⁺ 705.2739, found 705.2735. For NMR spectrum, see Appendix, Figure 131.

Synthesis of Py-Az-Val-Cit-PAB-MMAE (31)

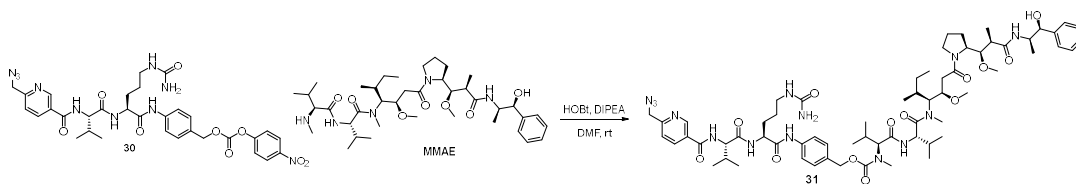


Figure 48 Synthesis of Py-Az-Val-Cit-PAB-MMAE (**31**)

PNP compound (5.9 mg, 8.36 mmol), MMAE (4.4 mg, 5.9 μ mol; purchased from Concoctis), HOBt (0.7 mg, 5.57 μ mol) and 100 μ L DMF were mixed in a 1.5 mL centrifuge tube. Then DIPEA (2 μ L, 11.48 mmol) was added. After 16 hours, 10 mL of this reaction was taken and diluted to 20 mL by DMF and loaded onto a C18 column.

C18 column, flow rate 1 mL/min; Solvent A: Water, 0.1% TFA; Solvent B: ACN, 0.1% TFA; wavelengths 254 nM, 280 nM.

HRMS (m/z): calculated for $C_{65}H_{98}N_{14}O_{13}$ $[M+H]^+$ 1283.7516, found 1283.7492.

For LC-MS chromatogram of Py-Az-Val-Cit-PAB-MMAE (**31**), see Appendix, Figure 132.

Synthesis of 7-((2-Aminoethyl)amino)-4-methyl-2H-chromen-2-one hydrobromide (**33**)

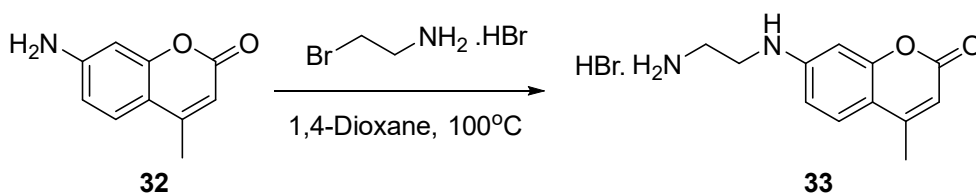


Figure 49 Synthesis of 7-((2-aminoethyl)amino)-4-methyl-2H-chromen-2-one hydrobromide (**33**)

A previously described protocol was applied with modifications.^[131] 2-bromoethylamine hydrobromide (0.11 g, 0.54 mmol) was added to a solution of 7-Amino-4-methylcoumarin (0.36 g, 2.0 mmol) in 3 mL 1,4-dioxane. The mixture was refluxed for

16 h, followed by concentration of the reaction mixture by evaporation. Ethyl acetate was added to the residue, and the solution was extracted with water (200 mL x 2). Then the aqueous extracts were combined and washed with EtOAc to remove remaining 7-amino-4-methyl-coumarin in the aqueous phase. The water was then removed by evaporation. 180 mg of product was isolated in pure form (0.5 mmol, 92% yield).

$^1\text{H NMR}$ (D_2O , 400 MHz): δ 2.42 (3H, s), 3.27 (2H, t, $J = 6.01$ Hz), 3.59 (2H, t, $J = 6.01$ Hz), 6.64 (1H, s), 6.10 (1H, s), 6.68 (1H, d, $J = 8.13$ Hz), 7.61 (1H, d, $J = 8.54$ Hz)

Synthesis of 6-(Azidomethyl)-N-(2-((4-methyl-2-oxo-2H-chromen-7-yl)amino)ethyl)nicotinamide (34)

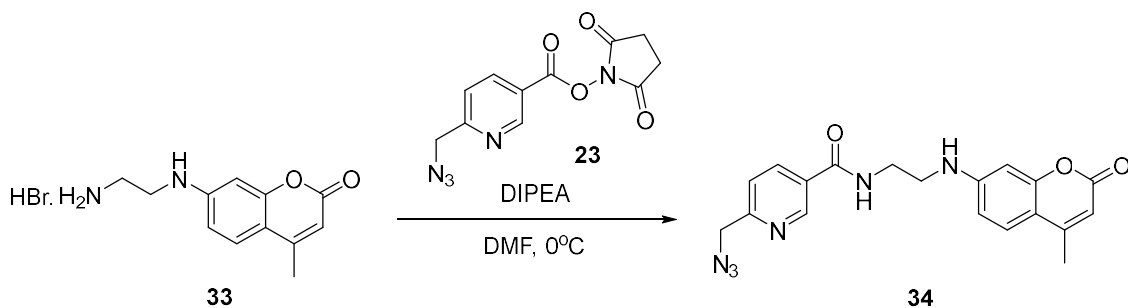


Figure 50 Synthesis of 6-(azidomethyl)-N-(2-((4-methyl-2-oxo-2H-chromen-7-yl)amino)ethyl)nicotinamide (**34**)
 107 mg (0.39 mmol) picolyl succinimide ester **23** was dissolved in 4 mL DMF and cooled on ice. Then, amine compound (87 mg, 0.29 mmol, dissolved in 2 mL DMF) was added to the solution. Finally, 140 μL DIPEA was added and the reaction was stirred for 3 hours on ice. Subsequently, the solution was stored at -20°C until purification. DMF was evaporated under reduced pressure (at 55°C) and then subjected flash chromatography. (TLC was run using 90% EtOAc, 10% Hexane). The product is obtained with 57% yield.

^1H NMR (DMSO- d_6 , 400 MHz): δ 2.30 (3H, s), 4.10 (2H, s), 4.41 (2H, d, $J = 4.2$ Hz), 4.58 (2H, s), 5.93 (1H, s), 6.52 (1H, d, $J = 8.13$ Hz), 6.64 (1H, dd, $J = 8.7$ Hz, 2.2 Hz), 6.78 (1H, t, $J = 6.0$), 7.44 (1H, $J = 8.8$ Hz), 7.53 (1H, $J = 8.1$ Hz), 8.21 (1H, dd, $J = 8.1$ Hz, 2.3 Hz), 8.82 (1H, t, $J = 5.1$ Hz), 8.98 (1H, $J = 1.8$ Hz)

The product was yellow and only slightly soluble in methanol. Samples were prepared in d -DMSO for ^1H NMR.

HRMS (m/z): calculated for $\text{C}_{19}\text{H}_{18}\text{N}_6\text{O}_3$ $[\text{M}+\text{H}]^+$ 379.1519, found 379.1507. For NMR spectrum, see Appendix, Figure 133.

2.2.5 Synthesis of ADCs

ADCs synthesized with solution phase method using non-chelating azide and alkyne linker-toxin molecule 14

The pertuzumab antibodies used in this experiment is Per-2C and Per-4C that have 2 and 4 disulfide bonds that can be reduced under mild conditions respectively.

First, 6 μM of Per-2C and Per-4C in PBS buffer was reduced by 200 μM TCEP at 37°C for 2 hrs. Then, N-(2-azidoethyl)-2-iodoacetamide (**7**) (400 mM stock solution in DMSO) was added to the solution at a final concentration of 4 mM. The reaction was incubated at room temperature for 30 minutes. Then, unreacted **7** were removed from the solution by extensive dialysis. Then, 50 μM CuSO_4 , 100 μM BTAA, 1.2 eq. (per functionalized cysteine group) alkyne linker-toxin compound **14** was added in the given order. Finally, 2.5 mM sodium ascorbate was added and reaction ran for 24 hours at ambient temperature. After 12 hours, the reaction mixture was supplemented with

another 2.5 mM sodium ascorbate to maintain reducing environment. LC-MS analysis of the conjugates were performed.

ADCs synthesized with resin method using chelating azide linker-toxin molecule **31**

The pertuzumab antibody used in this study (Per-2C) has 2 reducible cysteines in its hinge region.³⁰

First, 100 μ L 27 μ M of Per-2C in PBS buffer was reduced by 200 μ M TCEP at 37°C for 2 hrs. Then, bromoacetyl alkyne compound (**18**) (400 mM stock solution in DMSO) was added to the solution at a final concentration of 4 mM. The reaction was incubated at room temperature for 30 minutes. Next, 100 μ L Protein A-Sepharose resin slurry in PBS was added to this reaction mixture. The mixture was incubated on a shaker at room temperature for 2 hours. After that, an empty spin column was used for the washing step. After initial centrifugation, 300 μ L PBS were added to the slurry and the resin washed 5 times with 300 μ L PBS. All the centrifugations were carried out at 1000g for 1 min at room temperature.

After the final centrifugation step, the total slurry volume was brought to 100 μ L.

The 2x stock solution of click reaction mixture (200 mL reaction mix)

Table 11 The 2x stock solution of click reaction mixture (200 mL reaction mix)

2 x Click Reaction mixture
172 μ L PBS
16.7 μ L DMSO
1.6 μ L BTAA (25 mM stock solution in DMSO)
4 μ L Cu(II)SO ₄ (5mM stock solution in sterile water)

1.3 μ L azide (10 mM stock solution in DMSO)
4 μ L sodium ascorbate (250 mM stock solution in sterile water)

The above mixture (2x click reaction mixture) was prepared in the given order and incubated for 1 min after sodium ascorbate addition. Next, 100 μ L of the above solution was added to 100 μ L resin suspension. The tube was covered in aluminum foil and placed on a shaker at room temperature for 2 hours. Next, the slurry was added to a spin column for the final washing. After initial centrifugation, the resin was washed with 3 x 300 μ L PBS containing 1 mM EDTA. For the fourth and fifth washes, 1 mM phosphate buffer, pH 7.4, with 150 mM NaCl used reduce the trace amounts of phosphate in the slurry. ADC was eluted from protein A-Sepharose using acidic 0.1M glycine, 150 mM NaCl, pH 2.8 buffer for ~30 seconds with gentle shaking. Three sequential elutions (90 μ L each) were carried out to elute the ADCs. Each 90 μ L eluent was neutralized immediately by adding 15 μ L 2M Tris, 150 mM NaCl, pH 8.0. The three eluates were combined and EDTA added to a final concentration of 1 mM, followed by dialysis against PBS. The ADCs were characterized by LC-MS.

LC-MS Characterization of ADCs

LC-MS using ThermoFisher Scientific Q Exactive (Slot # 1760) as the detector was used to characterize the ADCs. BIOshell A400Protein C4 (Catalog #66825-U) 10cm x 2.1mm, 3.4 μ m column is used. Both water and ACN (acetonitrile) mobile phases contained 0.1% formic acid, and the following gradient was applied for the analysis:

7.1 min: 25% ACN

37.1 min: 35% ACN

40 min: 98% ACN

Since each antibody heavy chain contains one cysteine, only the heavy chains were modified. This can be seen by the retention times of heavy chains, with a retention time for reduced pertuzumab-2C antibody of 27 mins, whereas, this time is 29 mins for the conjugated antibody. The observed mass difference of 1478 Da is very close to theoretical mass difference of 1479 Da.

2.2.6 Cell viability assay with MDA-MB-453

MTS (3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulfophenyl)-2H-tetrazolium) assay was used to measure cell viability. 96 well plate with 5,000 cells per well (50 μ l) per well is used. Perimeter wells are not used as they dry up faster. One triplicate with only medium (no cells) is used for background control. Drug doses are formulated in the same media as the cells were plated in (it is ok to use phenol-red media). Drug doses should be made at 2 times the desired concentration. 50 μ l of diluted drugs are added to previously plated cells (final volume should be 100 μ l). For the well with no cells, an additional 50 μ l of media are added. Plates were incubated at 37°C + 5% CO₂. Cells were incubated long enough (4 days) for the cells to divide at least twice. MTS reagent (Promega CellTiter 96 AQueous One Solution Cell Proliferation Assay; G3581) is added. 20 μ l is added to each well and plates are incubated in the incubator for 1 – 4 hours. Plate reader is used to measure absorbance at 490 nm wavelength.

2.3 Results & Discussion

2.3.1 Kinetics study

Pseudo-first order kinetics of CuAAC reaction between non-chelating azide and non-chelating alkyne,^[123] chelating azide and non-chelating alkyne,^[122] chelating azide with aromatic ynamines^[124] were previously studied in the literature. These studies were focused on deciphering the CuAAC mechanism and did not include copper-binding ligands (such as TBTA or BTAA) in their kinetic assays. However, for bioconjugation purposes, copper-binding ligands are essential for preventing undesired copper-induced amino acid side chain oxidations and having reasonable reaction kinetics. Click reaction kinetics is studied to find the optimal reaction conditions for biorthogonal chemistry applications. In the kinetic assays, turn-on fluorogenic substrates azido-7-hydroxy-coumarin and 7-ethynylcoumarin with various azide and alkyne substrates were used. Both azido-7-hydroxy-coumarin and 7-ethynylcoumarin become fluorescent after triazole formation enabling direct observation of the product formation (Figure 51).

Tracking product formation by fluorescence signal allowed us to set up reactions simultaneously in series by using plate reader.

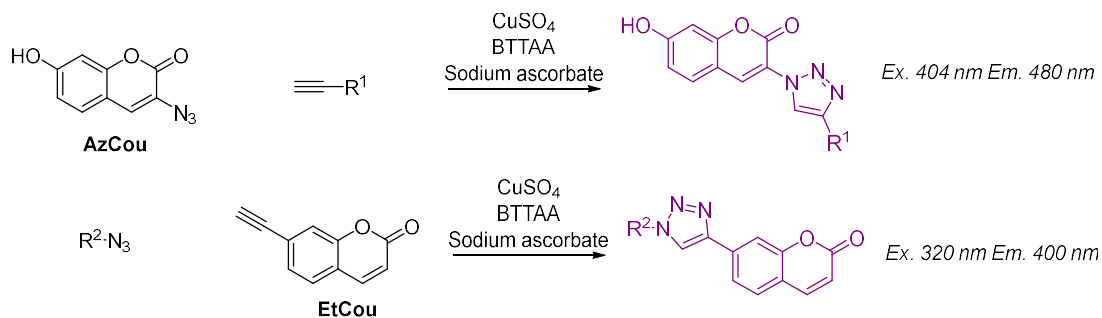


Figure 51 Assay used to track CuAAC reaction trajectory

Reaction Between Propargyl Alcohol and 3-Azido-7-hydroxy-coumarin

Our interest in click reaction kinetics started when Dr. Alfred Tuley (a previous PhD student in our group) observed that reaction is independent of propargyl alcohol with limiting 3-azido-7-hydroxy-coumarin concentrations in pseudo-first order reaction conditions (Figure 52A). That observation agrees with the data reported in literature which was measured in the absence of a copper-binding ligand.^[123] Following that, we tested aromatic alkynes with electron donating and withdrawing groups to see whether, the same phenomenon will apply to alkynes with different electronic properties. Testing five different para-substituted phenyl alkynes with different electron withdrawing and electron donating groups yielded similar results (Figures 52B- 1F). The observed reaction rate was independent of alkyne concentration for all tested alkynes (Figure 52G).

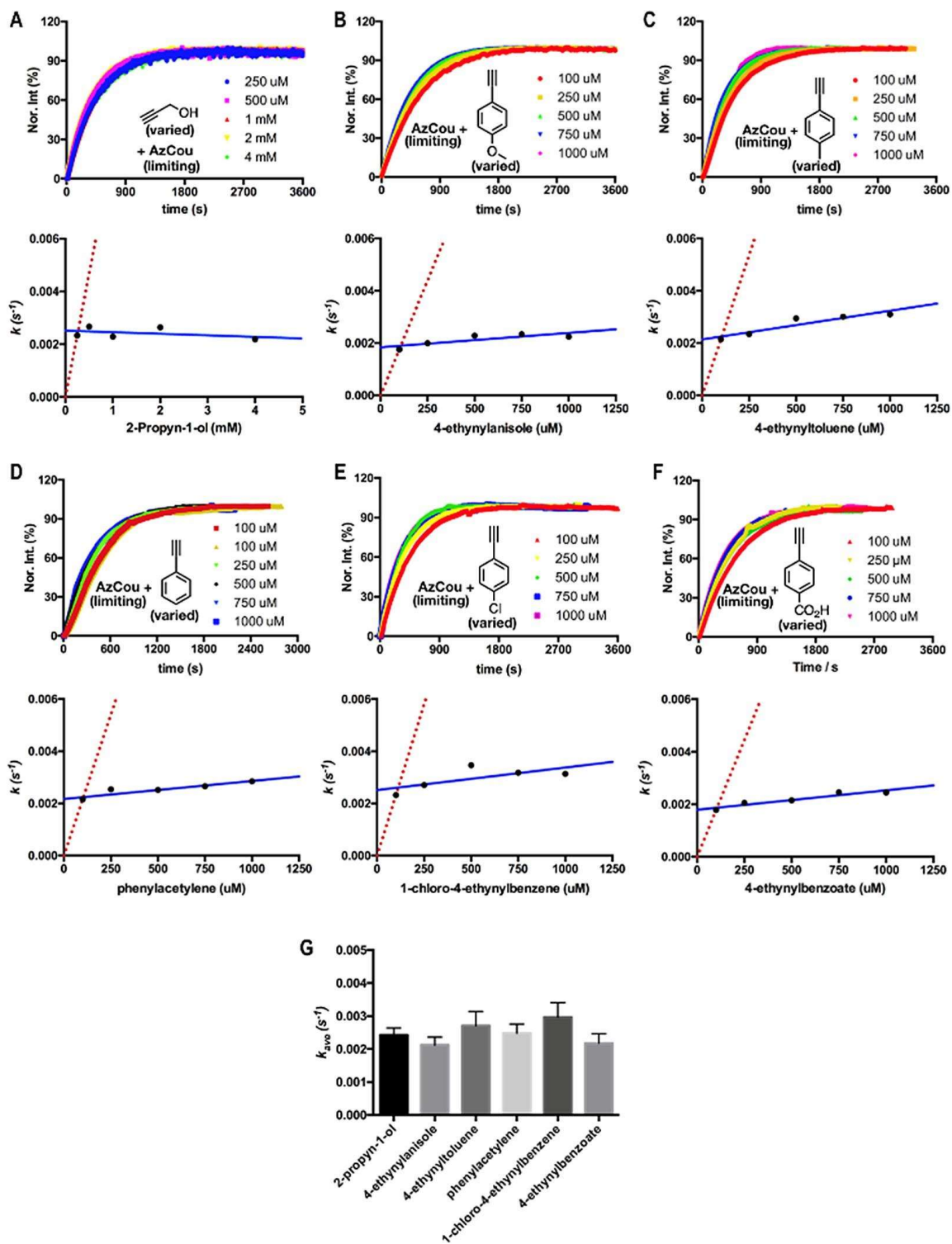


Figure 52 Reactions of AzCou with excessive alkyne concentrations. **A)** The one-phase exponential increase of product fluorescence when a limiting concentration of AzCou (50 μM) was reacted with excess concentrations of 2-propyn-1-ol (0.25–4 mM). **(B–F)** The one-phase exponential increase of product fluorescence when a limiting concentration of AzCou (10 μM) was reacted with excess concentrations of five phenylacetylene derivatives (0.1–1 mM). **(G)** Average apparent reaction rate constants for all tested alkynes.

Reaction of Excess 2-Azidoethylamine and EtCou

Next, we did a similar analysis to check for dependency of observed rate constant (k_{obs}) on azide concentration. When 7-ethynylcoumarin reacted as limiting reagent with varying excess concentrations of 2-azidoethylamine, the reaction showed a linear dependence (Figure 53A) in parallel with the literature report.^[123] When 2-azidoethylamine was used as the limiting reagent and EtCou concentrations (100–500 μM) were varied, minimal change in k_{obs} was observed (Figure 53B) in agreement with our previous observation in Figure 52A–G.

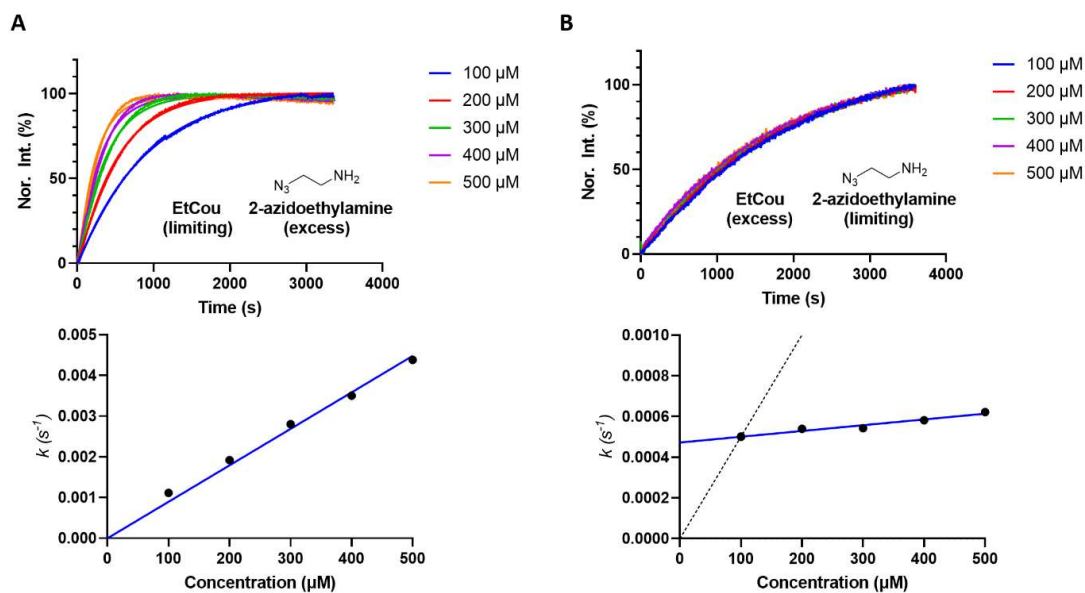


Figure 53 A) Limiting EtCou with varying excess concentrations of 2-azidoethylamine **B)** Limiting 2-azidoethylamine with varying excess concentrations EtCou

Reaction of Excess Pyridine azide and Limiting EtCou

In 2010, pyridine azide (PyAz) was reported to be accelerate the CuAAC reaction more significantly compared to a non-chelating azide.^[132] Then, we tested the dependency of k_{obs} with respect to varying excess PyAz concentrations against limiting EtCou in pseudo-first order conditions. We found the reaction to be zero order for PyAz under this condition (Figure 54A). When we tested reaction order with excess EtCou concentrations against limiting PyAz concentration, a linear relation was observed. (Figure 54B).

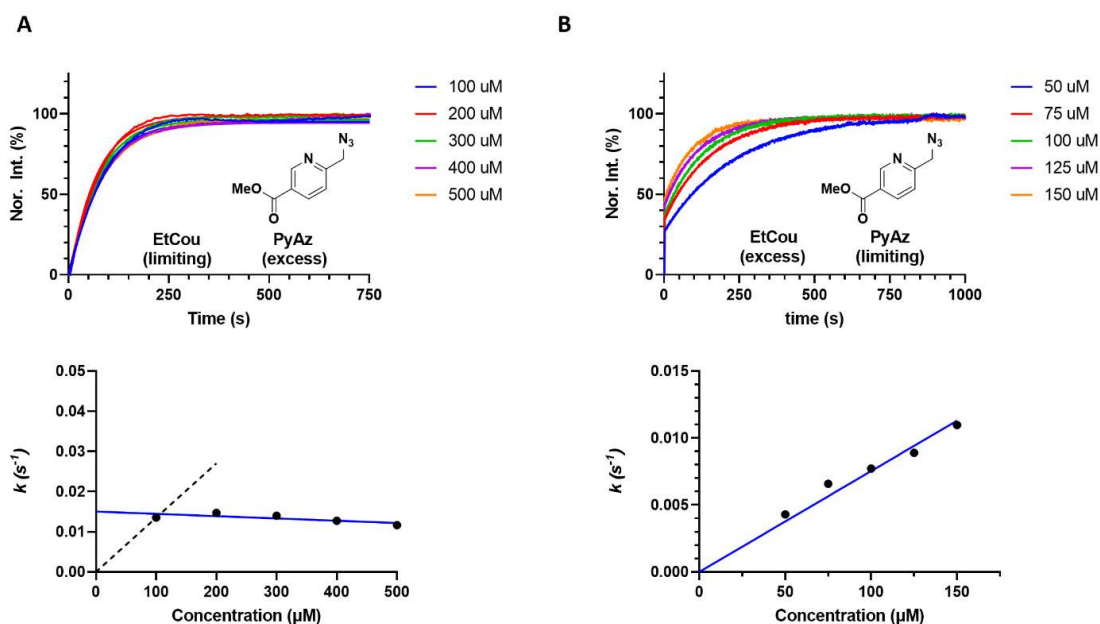


Figure 54 **A)** Limiting EtCou concentrations with excess PyAz concentration **B)** Excess EtCou concentrations reacted with PyAz

These observations suggest pyridine moiety directs copper atom to speed up formation of the azide-copper complex. This makes azide-copper complex formation no longer the

rate-determining step. Instead, the rate-determining-step becomes alkyne dependent as seen by the linear relationship in Figure 54B.

Varying Copper Concentrations with Excess PyAz and Limiting EtCou

Another variable in ligand-accelerated CuAAC is copper. Generally, 50x fold of sodium ascorbate is used to maintain copper as Cu(I). However, Cu(I) forms aggregates in solutions and not likely to have as a homogenous catalyst. The presence of ligands is believed to help break Cu(I) aggregates liberating individual catalytically active Cu(I) atoms.

Previously, isotope exchange study and isolation of stable bis(copper)alkyne intermediates and kinetics studies pointed out having two coppers on alkyne.^[120] However, in these literature conditions, non-chelating alkyne and non-chelating azides were used. We wanted to see the reaction order for copper with limiting amounts of EtCou (10 μ M) and excess PyAz (100 μ M). When we examine the reaction with varying copper concentrations under excess PyAz and limiting EtCou concentration conditions, we expected the copper-alkyne including step to be observable. When we varied copper concentrations between 10-30 μ M, the rate order for copper was observed to be 2.1 (Figure 55 A-C). Our observation showed that in the case of chelating azide PyAz, alkyne is still interacting with two copper atoms.

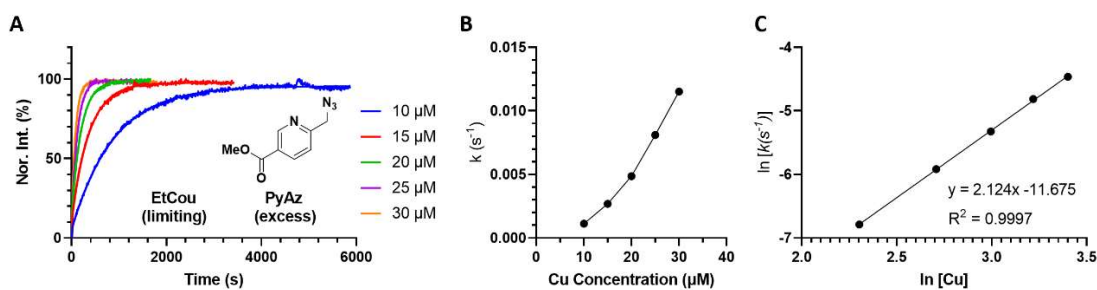


Figure 55 A) Reaction trajectories of varying copper concentrations (10-30 μM) B) Graph of k_{obs} vs copper concentrations C) $\ln(k_{\text{obs}})$ vs $\ln(\text{Cu})$

To sum up, these results show reaction orders determined by pseudo-first-order reaction conditions of ligand-accelerated CuAAC follows a similar trend to reported reaction orders of ligand-free CuAAC reaction in the literature. When non-chelating azide and alkyne are used the azide migration/insertion step is rate-determining step whereas, when a chelating azide (picolyl azide) is used, the copper-acetylide formation becomes the rate-determining-step.

A faster alkyne: Ethoxyacetylene

In 2016 and 2017, aromatic ynamines and alkynes with strong electron-withdrawing groups were introduced as faster reacting alkynes. Alkynes with electron withdrawing groups are unsuitable for bioconjugation due to inevitable side reactions with nucleophilic amino acid side chains. We proposed that ethoxyacetylene (Figure 56A) could be a faster reacting alkyne as well. We purchased ethoxyacetylene, 3-butyne-2-one and compared its reactivity with propargyl alcohol. Ethoxyacetylene and 3-butyne-2-one

were shown to be significantly faster than propargyl alcohol (Figure 56 B, C).

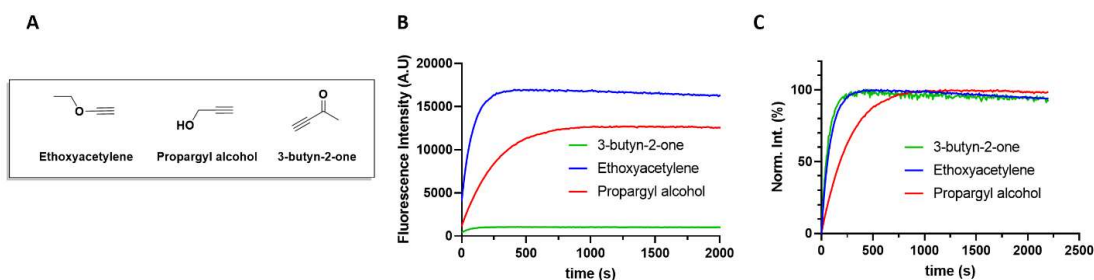


Figure 56 A) Alkynes used in this assay B) Reaction trajectories of 20 μ M of AzCou was used with 400 μ M of alkyne derivative C) Normalized intensity of the reaction trajectories

This observation challenged our understanding of the CuAAC reaction. Since the reaction is independent of alkyne concentration (with a non-chelating azide), a more reactive alkyne should not be able to speed up the reaction. On the contrary, a less reactive alkyne can make the alkyne involving step the rate-determining-step and cause the overall reaction to be slower. However, a more reactive alkyne should perform similar to the other alkynes tested such as propargyl alcohol. The only explanation we could come up with for this observation was that the copper atom(s) in the potential bis(copper) ethoxyacetylene intermediate have a higher tendency to interact with the azide to speed up the azide migratory/insertion step.

Reaction Between Ethoxyacetylene and AzCou

Because ethoxyacetylene and 3-butyn-2-one were significantly more reactive than propargyl alcohol, we evaluated their potential in biorthogonal reactions. 3-butyn-2-one is a reactive electrophile and it is likely to undergo reactions with amino acids having nucleophilic side chains. Therefore, we turned our attention to ethoxyacetylene. We tested excess concentrations of ethoxyacetylene with limiting AzCou concentration.

When 20 μM of coumarin azide reacted with 100, 200, 300, 400, 500 μM of ethoxyacetylene the final fluorescence of each concentration was different in accordance with their initial concentrations (Figure 57A, 57B). Higher ethoxyacetylene concentrations yielded higher final fluorescence values.

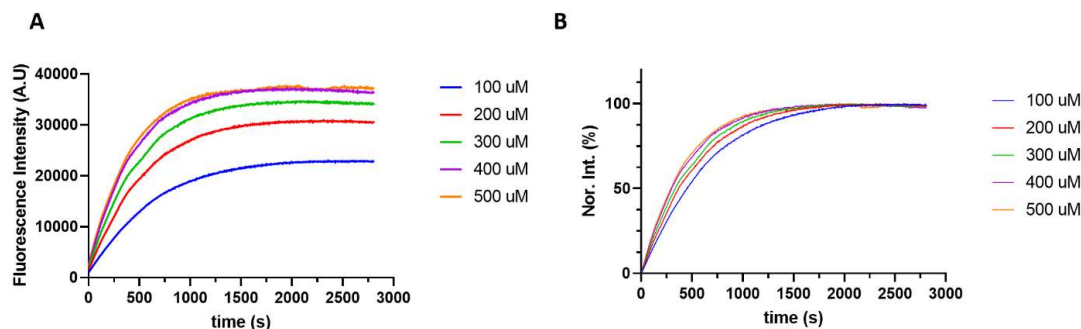


Figure 57 **A)** Reaction trajectories of the reaction between ethoxyacetylene and AzCou **B)** Normalized intensities of the reaction trajectories

We reasoned that ethoxyacetylene was probably undergoing a side reaction under the reaction conditions and had LC-MS and ^1H NMR characterization of the reaction. However, we were unable to identify a side product (data not shown). An interesting observation was when the NMR sample was left in the tube overday, we saw the formation of suspended solid formation which was not observed during the reaction. The two samples (the sample treated with celite and without celite) in CD_3OD both yielded precipitates (Figure 58).



Figure 58 Precipitate formed in the reaction between ethoxyacetylene and AzCou

At this point, ethoxyacetylene did not seem to be a good candidate for use in biorthogonal reactions due to the side reactions it had undergone and we turned our attention back to using aliphatic alkynes without adjacent heteroatoms.

Unfortunately, the alkynes (ethoxyacetylene, 3-butyne-2-one) that improved reaction rate were unsuitable for bioconjugation conditions. Next, we move on to test different azide structures (Figure 59A) effect on reaction rate. We compared four different azides (2-azidoethylamine, phenyl azide, 2-azidoethanol, AzPy) at 20 μM by reacting with 100 μM EtCou. As expected AzPy reacted the fastest among the four. 2-Azidoethylamine reacted faster than 2-azidoethanol probably due to the tendency of the free amine group to direct copper atom. Phenyl azide reacted faster than 2-azidoethanol but, slower than 2-azidoethylamine. Phenyl azide is more reactive than 2-azidoethanol probably due to higher electron density on nitrogen atoms due to withdrawing electrons

from the benzene ring (Figure 59B).

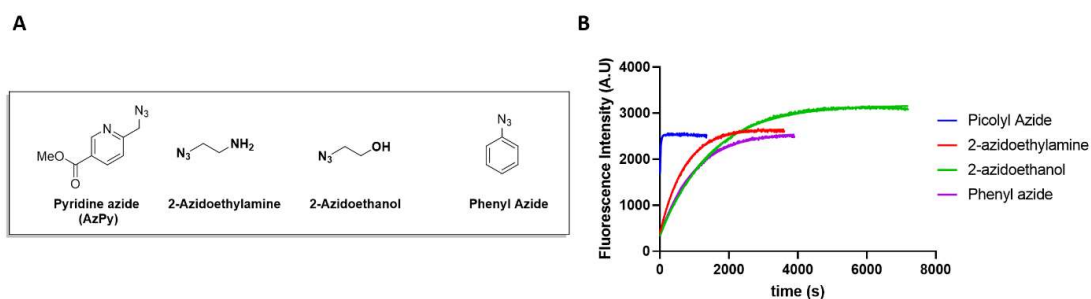


Figure 59 A) Azide structures B) Reaction trajectories of different azides with EtCou

Pyridine azide's (PyAz) fast kinetics made it a good candidate for ADC synthesis. As mentioned in the introduction part, a fast-reacting conjugation reaction is necessary to push all the naked antibodies into homogenous ADCs.

FRET-based assay to study on-protein kinetics

Next, we wanted to verify that the conditions used in small molecule kinetics can be translated to generate protein conjugates. To demonstrate this, we built a model system that allowed us to monitor protein conjugation kinetics by FRET analysis (Figure 60). We site-specifically incorporated L-propargyl-lysine (PrK) into superfolder green fluorescent protein (sfGFP) to generate our model protein (sfGFP134PrK). We synthesized compound **34** which is a metal-chelating azide with coumarin dye moiety. Therefore, the chromophore of sfGFP and coumarin dye can form a FRET pair upon conjugation (Figure 60A). After our first conjugation attempt with 10 μM CuSO_4 and 20 μM BTAA, we did not detect a FRET signal. This result might be due to the presence of 6xHistag on the sfGFP134PrK or another metal-binding motif. Next, we tried the reaction with 40 μM CuSO_4 and 80 μM BTAA and the reaction appeared to reach a

plateau in 10 minutes (Figure 60B, 60C). This result showed us that, from a kinetics point of view, it is reasonable to keep the concentration of CuSO₄ around 50 μM when carrying out click reactions with proteins.

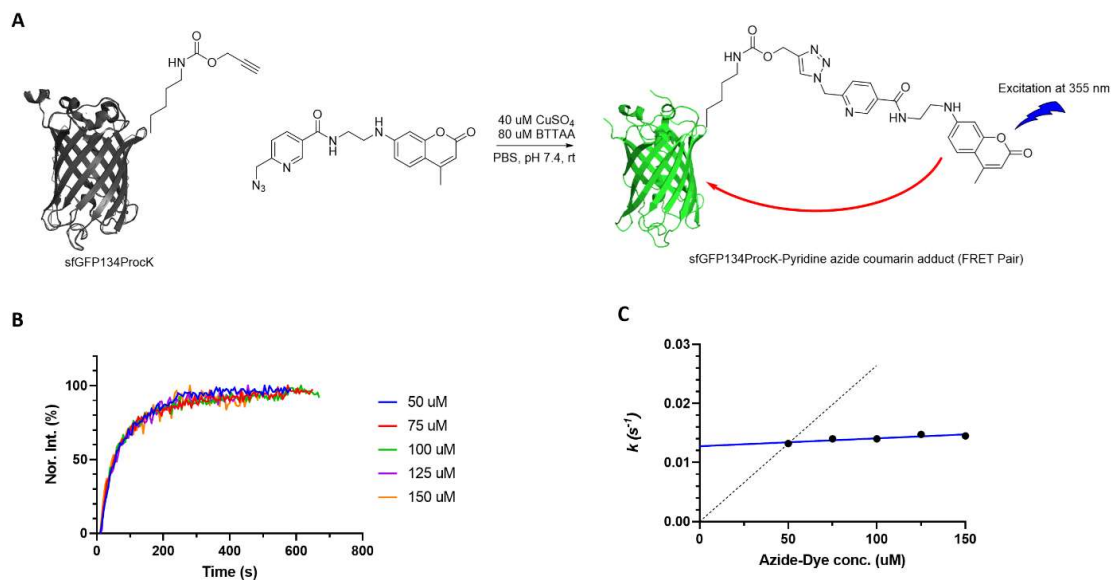


Figure 60 A) FRET analysis scheme for on-protein kinetics B) Reaction trajectories with varying concentrations of **34** C) k_{obs} vs varying concentrations of **34**

2.3.2 Synthesis of ADCs

ADC Synthesis Strategy-I: Solution phase synthesis with non-chelating azide

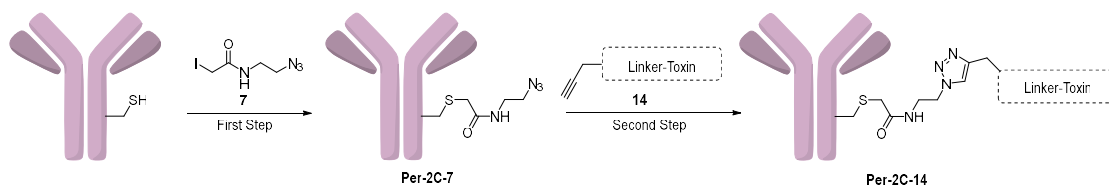


Figure 61 ADC Conjugation Strategy

We devised two different designs to apply click reaction to synthesize ADCs. In the first design, we installed a non-chelating azide moiety to the antibody through cysteine-alkylation. Then, reacted it with an alkyne-containing linker-drug molecule (Figure 61). This design's goal was to put the alkyne on the small molecule to make the linker-drug

synthesis more straight forward. The MMAE toxin we were using is expensive and the purification procedure required reverse phase HPLC which allowed us to work with only a low-miligram scale. To make the synthesis easier, we put the more stable alkyne moiety to the drug-linker compound instead of the light and heat sensitive azide group. Another idea behind this design was to avoid having an alkyne group on the antibody during the click reaction. A known side product of the CuAAC reactions is the homocoupling of the alkyne molecules.^[133] Although, after the first step, the disulfide bond connecting two heavy chains is broken and cysteine residues are alkylated, the two heavy chains are still interacting through non-covalent interactions as described in the introduction section. Having two alkyne groups in close proximity increases the risk of having a homocoupling side product. However, when the alkyne is on the small molecule, even if there is residual homocoupling products, it can be easily removed by using size-exclusion column.

In our synthetic plan (Figure 62), the first step is capping the amino group of l-valine with the propargyl group. For the rest of the synthesis, the propargyl group served as the protecting group and ironically it also served as our reactive handle when attaching the drug-linker molecule to azide-antibody. Next, l-citrulline was coupled to NHS ester N-(Prop) valine **10** to yield the dipeptide **11**. To facilitate a mild coupling condition for relatively slow reacting 4-amino benzyl alcohol, we used EEDQ coupling reagent. Next, primary alcohol **12** was converted to nitrobenzyl carbonate to generate an activated carbonate **13**. Finally, MMAE reacted with **13** to yield the desired compound

14 in 1.0 mg quantity. Characterization was done by ESI-MS and the HPLC chromatogram showed only one peak.

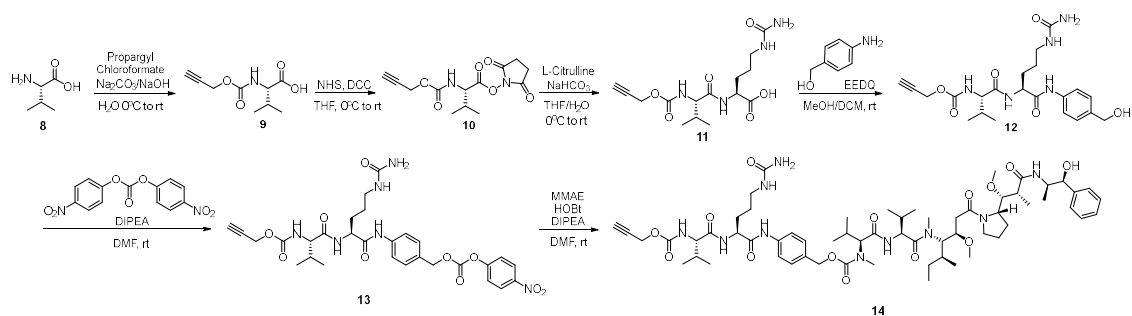


Figure 62 Synthesis of Linker toxin compound **14**

After synthesizing **14** we conjugated the drug to the Pertuzumab antibody. The Pertuzumab antibody is an FDA-approved drug that binds to the extracellular domain of HER2 (human epidermal growth factor receptor 2) and inhibits dimerization of HER2. Pertuzumab is used against HER2+ breast cancer.^[134] We followed our two-step conjugation strategy as in Figure 61. In this experiment, Per4C was used which has 4 reducible disulfide bonds. In this design, 3 disulfide bonds connect two heavy chains and one disulfide bond connects the light chain and the heavy chain. Therefore, in case of 100% conversion, 3 drug molecules will be conjugated to the heavy chain and one drug will be conjugated to the light chain. The first step is a well-known nucleophilic substitution reaction between cysteine and iodoacetamide derivative **7**. Derivative **7** was provided in 4 mM concentration to alkylate the free cysteines on the light and heavy chains. Next, we removed the small-molecule by extensive dialysis and carried out the click reaction to attach the linker-drug molecule. After, reacting for 24 h at room temperature, we still observed some of starting material of light chain that is reacted with

7 but not with **14** (Appendix, Figures 134-136). Because the light chain was modified through one cysteine and the heavy chain was modified by three cysteines, the shift in the retention time of light chain after reacting with **14** revealed the extent of conversion more clearly. It is not possible to estimate the percent conversion precisely without having a standard for the conjugated antibody. By following the same procedure **Per-2C-14** was synthesized. Next, our collaborator tested the activity of the conjugated **Per-2C-14** and maleimide conjugated ADC in parallel against MDA-MB-453 HER2+ breast cancer cell line by using MTT assay. Our collaborator chose to work on **Per-2C-14** instead of **Per-4C-14** because, higher drug to antibody ratios can cause antibody aggregation. The conjugation was tested with four mutants of Pertuzumab antibody with mutations in their light chain that alters their internalization into cell.^[135] Maleimide and Pertuzumab both showed similar efficacy levels for the three of the antibody mutants (Figure 63A-C). However, the Hulys10 antibody showed a difference at the highest ADC concentration. Hulys10 conjugated ADCs are the control group (Figure 63D). Hulys10 cannot penetrate inside the cells. Hence, Hulys10 ADCs are not supposed to show toxicity as a negative control. However, if the drug-linker molecule is detached from the antibody, it can enter cells in a nonspecific fashion as discussed in the introduction section. While **Per-2C-14** did not show any toxicity over the range of concentrations, maleimide conjugated ADC showed toxicity at the highest concentration. The reversibility maleimide conjugates can explain this result. Maleimides can be detached from the antibody and linker-drug molecules can enter the cells to cause non-specific toxicity. In our ADC **Per-2C-14**, triazole linkage is reasonably stable in

biological systems and does not undergo a similar process; hence, no toxicity was observed at the same concentration.

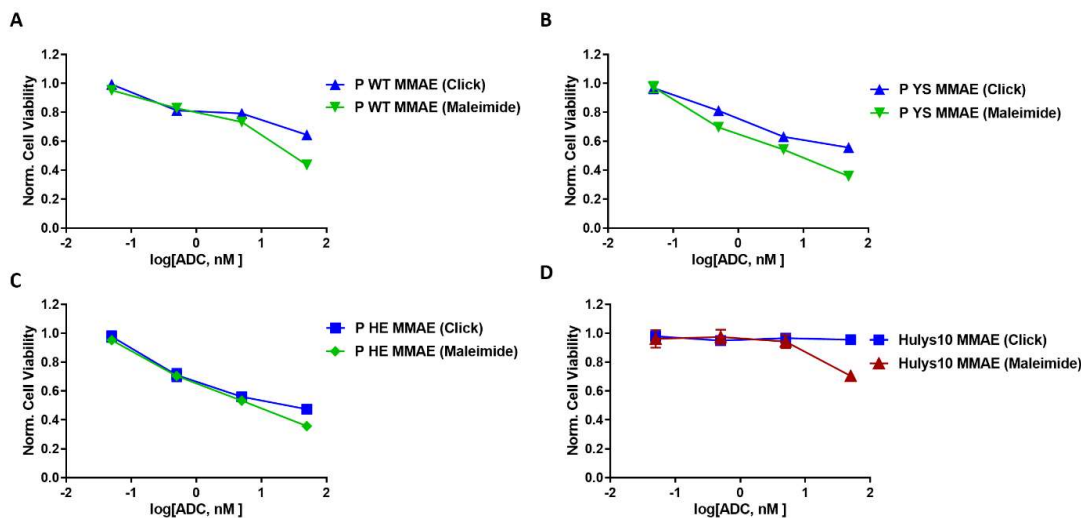


Figure 63 Cell-based assay of ADC with MDA-MB-453. P WT (A), P YS (B), P HE (C) are the Pertuzmab derivatives with different cell-penetration properties Hulys10 (D) is the control group mentioned in reference [132]. The data points shown with (Click) are the ADCs conjugated according to Figure 61. The ADCs generated by classical maleimide conjugation is shown in Figure 9B with (Maleimide).

ADC Synthesis-II: Solid phase synthesis with chelating-azide

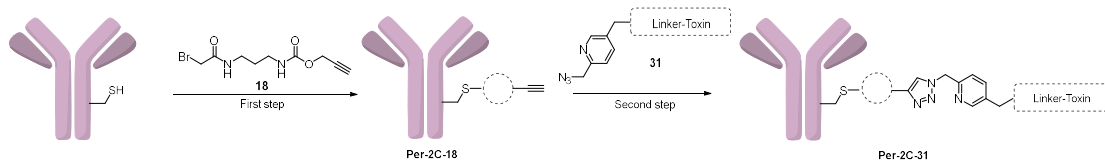


Figure 64 ADC synthesis with chelating-azide linker-toxin 31

To improve the conjugation efficiency of the drug-linker molecule to antibody we decide to use drug molecule having pyridine azide group which would react much faster according to kinetics experiments. To synthesize the desired drug-linker compound 31, we started off with synthesizing NHS ester of boc-l-valine 25 and coupled it to l-citrulline to get the dipeptide 26 (Figure 65). Next, using EEDQ chemistry we coupled

4-aminobenzyl alcohol to boc-l-valine-citrulline dipeptide **26** to synthesize **27**. Then, we deprotected the boc group by using TFA to get **28**. Next, we coupled the pyridine azide group to **23** to yield **29**. Then, similar to above we synthesized the activated carbonate **30** by reacting the alcohol **29** with para-nitrophenyl carbonate. Finally, the reaction with MMAE yielded the desired compound **31**.

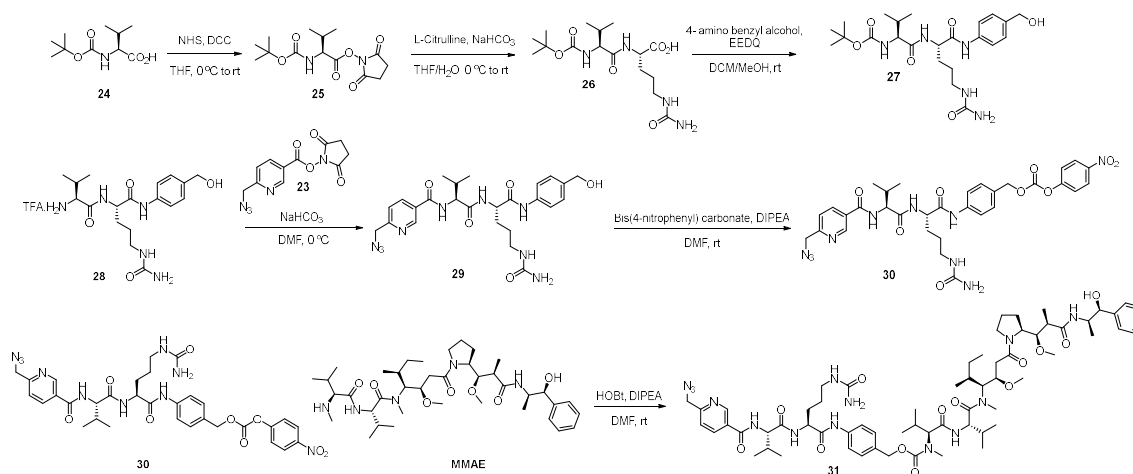


Figure 65 Synthesis scheme of linker-toxin compound **31**

Our conjugation strategy has two steps. It is essential to remove the bromoacetamide alkyne compound **18** completely before moving to the next step which can be achieved by extensive dialysis or gel desalting column. Dialysis is generally the purified technique to eliminate small molecules because, some small molecules may stay bound to the antibody through non-specific interactions and may be eluted with the antibodies. Moreover, complete removal of drug-linker molecules can be a bigger problem because they are above 1 kDa size and very hydrophobic. We reasoned that because our picolyl azide-alkyne reaction is quite efficient in solution, it should perform reasonably well on solid phase too. Besides, it would eliminate the need for dialysis after the first step and

speed up the conjugation process (Figure 66). Therefore, following reacting the free cysteine groups with bromoacetyl alkyne **18** compound, Protein A sepharose resin was added to the reaction mixture. After incubation at ambient temperature for 2 hours, unreacted **18** was removed simply by washing the Protein A Sepharose resin and then CuAAC reaction was performed with chelating-azide **31** on resin for 2 hours. Similarly, the CuAAC reagents were removed easily by washing the resin. Total conjugation time was 9 hours (reduction of the disulfide bonds, first, step, second step, washing steps and final elution). Two different elution buffers were tried: pH 2.8 and pH 11 buffers. Elution with pH 11 buffer yielded a lower amount of **Per-2C-31**. This result might be due to degradation of the antibody due to combination of copper and the basic environment which is known to cause hydrolysis of amide bonds. Therefore elution acidic pH followed by neutralization and dialysis gave the desired conjugate. The LC-MS spectrum from the acidic elution can be seen in Appendix, Figures 137-140. Due to working with Per-2C, only heavy chains get modified and, a clear shift of the retention times of the heavy chains before and after reaction shows near-quantitative conversion of starting material.

The next step was to analyze these conjugates in cell culture the same way. However, around that time our collaboration ended due to the relocation of our collaborator. We set up a similar assay system in our lab. However, our assays did not show the toxicity of **Per-2C-31** on MDA-MB-453 cells after initial trials. Due to the simultaneous spread of COVID-19 we re-evaluated our priorities and decided to terminate the project.

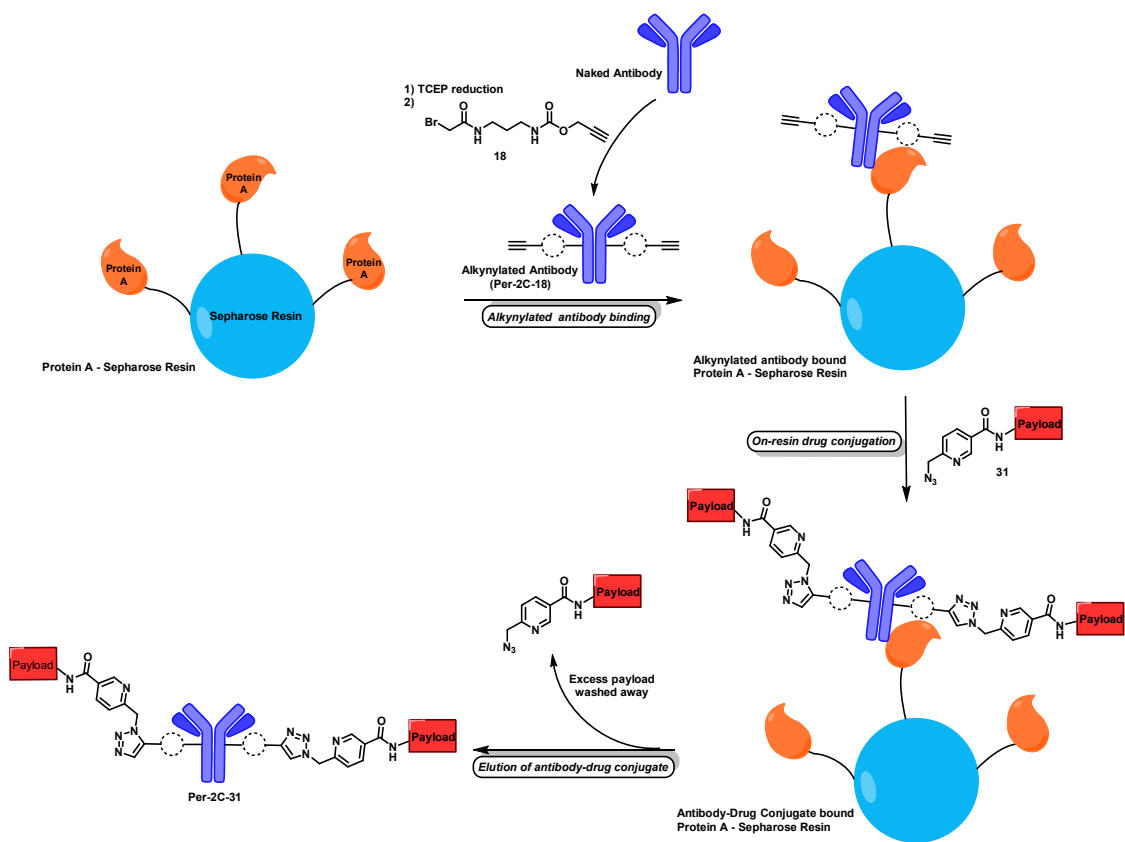


Figure 66 Synthesizing Per-2C-31 by resin method

2.4 Conclusion

In conclusion, by starting from the small molecule kinetic analysis conditions for copper-catalyzed azide-alkyne cycloaddition (CuAAC) reaction were studied and applied to synthesize antibody-drug conjugates. Pseudo-first order kinetics revealed that in our reaction conditions, with non-chelating azide and alkyne reaction rate is independent of alkyne concentration and has a linear relationship with azide concentration. However, when chelating-azide is reacted with an alkyne, the reaction rate becomes dependent on alkyne concentration but, independent of azide concentration. Next, we synthesized ADCs in two different ways: with chelating-azide and non-chelating azide. While ADCs synthesized with non-chelating azides showed toxicity to cancer cell line MDA-MB-453 similar to widely used maleimide-thiol conjugates, ADCs with non-chelating azide showed lower side-toxicity compared to maleimide-thiol conjugates. ADCs with chelating azide synthesis was more efficient and conducted on Protein A attached resin to remove unreacted linker-toxin molecules efficiently. Unfortunately, the cancer cell line tests with chelating-azide were terminated due to COVID-19 surge before a conclusion was reached.

CHAPTER III

INCORPORATION OF L-HOMOARGININE AND *ORTHO*-CHLORO-L-PHENYLALANINE INTO PROTEINS

3.1 Introduction

Proteomics studies show arginine methylation is an abundant post-translational modification (PTM) that includes three methylation motifs: monomethyl-L-arginine (MMA), symmetrical dimethyl-L-arginine (SDMA) and asymmetric dimethyl-L-arginine (ADMA) (Figure 67A).^[136] Converting the N-H group(s) to N-CH₃, arginine methylation results in a decrease in hydrogen bonding ability and an increase in hydrophobicity of the arginine side chain. As a result of this, arginine methylation plays significant regulatory roles in protein-protein interactions.^[137] Histones are arginine-rich proteins that are shown to contain all three types of arginine methylation motifs. Asymmetrically demethylated arginines can involve in crosstalks with histone acetyltransferases to induce acetylations on histone proteins affecting the chromatin state.^[138] A known crosstalk of ADMA residues between p300/CBP-associated factor (PCAF) and histone 3 (H3) can influence the H3K9 and H3K14 levels, affecting gene regulation.^{[139], [140]} Histone 4 can undergo H4R3 methylation which facilitates beta-globin transcription by regulating histone acetyltransferase binding and H3 acetylation Tudor domain-containing protein 3 (TDRD3) and WD repeat-containing protein 5 (WDR5).^{[141], [142]} WDR5 protein recruits mixed-lineage leukemia protein (MLL) complex to nucleosome to repress H3K4 trimethylation which is known to be a gene

promoter. TDRD3 is reported to be linked with cell proliferation of tumor cells and overexpressed in brain cancer. However, a systematic study has not yet been performed to map the binding sites of these proteins on the nucleosome. An important quest in studying epigenetics is identifying physiologically significant demethylases (Figure 67B) or demethyliminases (Figure 67C) of methylated arginine. Although most histone PTMs are reversible, no specific eraser has yet been identified for arginine methylation. There are reports about nine demethylases.^{[143], [144]} and a demethyliminase^[145] shown to be acting on methylated arginine on peptide level and some evidence for *in vivo* activity. However, there are no significant data present *in vivo* conditions to support that these enzymatic modes are playing significant roles in metabolism.

Our first aim was to site-specifically incorporate L-homoarginine (HAr) derivatives monomethyl-L-homoarginine (MeHAr) dimethyl-L-homoarginine (DiMeHAr) and evaluate the potential of MeHAr as MMA and ADMA analogs. Our second aim is to introduce the reported mutations^[25] that facilitates HAr incorporation in a step-wise fashion to gain more insight into the PylRS enzyme.

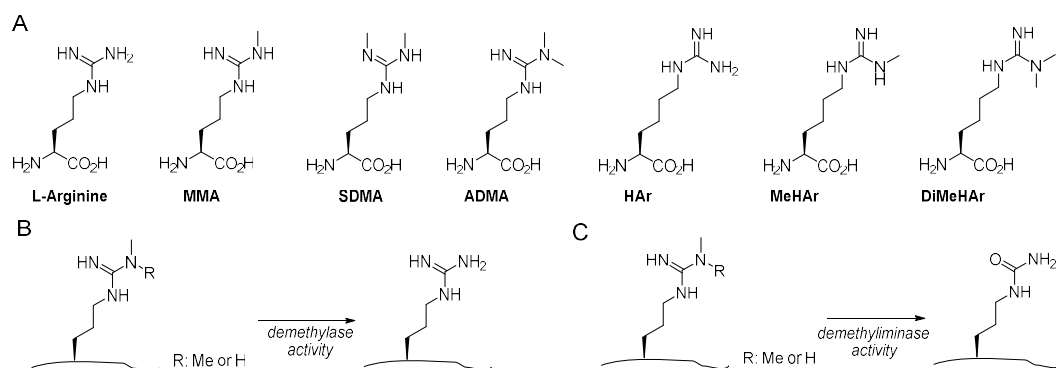


Figure 67 A) Structures of L-Arginine, MMA, SDMA, ADMA, HAr, MeHAr, DiMeHAr B) Demethylase activity on MMA or ADMA residues C) Demethyliminase activity on MMA or ADMA residues

3.2 Experimental

3.2.1 Protein Sequences

sfGFPN134TAG sequence

MVSKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTCLKFICTTGKLV
PVPWPTLVTTLTYGVCFSRYPDHMKRHDFFKSAMPEGYVQERTISFKDDGTY
KTRAEVKFEGDTLVNRIELKGIDFKEXGNILGHKLEYNFNHNVYITADKQKNGI
KANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKR
DHMVLLFEVTAAGITHGMDELKGGSHHHHHH

wt-*Mm*-PylRS sequence

MDKKPLNTLISATGLWMSRTGTIHKIKHHEVSRSKIYIEMACGDHLVVNNSRSS
RTARALRHHKYRKTCKRCRVSEDLNKFLLTKANEDQTSVKVKVVSAPTRTKK
AMPKSVARAPKPLENTEAAQAQPSGSKFSPAIPVSTQESVSVPASVSSISSISTGA
TASALVKGNTNPITSMSAPVQASAPALTKSQTDRLEVLLNPKDEISLNSGKPFRE
LESELLSRRKKDLQQIYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIER
MGIDNDELTSKQIFRVDKNFCLRPMLAPNLNYLRKLDRALPDIKIFEIGPCYR
KESDGKEHLEEFMLNFCQMGSGCTRENLESIITDFLNHLGIDFKIVGDSCMVYGD
DTLDVMHGDLELSSAVVGPIPLDREWIDKPKWIGAGFGLERLLKVKHDFKNIKR
AARSEYYNGISTNL

N-term-opt-*Mm*-PylRS sequence

MDKKPLNTLISATGLWMSHTGTIHKIKHREVSRSKIYIEMACGDHLVVNNSRSS
RTARALRHHKYRKTCKRCRVSEDLNKFLLTKANEDQTSVKVKVVSAPTRTKK
AMPKSVARAPKPLENSEAAQAQPSGSKFSPAIPVSTQESVSVPASVSSISSISTGAT
ASALVKGNTNPITSMSAPVQASAPALTKSQTDRLEVLLNPKDEISLNSGKPFREL
ESELLSRRKKDLQQIYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERM
GIDNDELTSKQIFRVDKNFCLRPMLAPNLNYLRKLDRALPDIKIFEIGPCYRKE
SDGKEHLEEFMLNFCQMGSGCTRENLESIITDFLNHLGIDFKIVGDSCMVYGD
LDVMHGDLELSSAVVGPIPLDREWIDKPKWIGAGFGLERLLKVKHDFKNIKRAA
RSEYYNGISTNL

C_{Ma} PylRS

Chapter IV MTVKYTDAAQIQRLREYNGTYEQKVFEDLASRDAAFSKEMSVASTD
NEKKIKGMIANPSRHGLTQLMNDIADALVAEGFIEVRTPIFISKDALARMTITED
KPLFKQVFWIDEKRALRPMLAPNLYSVMRDLRDHTDGPVKIFEMGSCFRKESH
GMHLEEFMLNLDVMGPRGDATEVLKNYISVVMKAAGLPDYDLVQEESDVYK
ETIDVEINGQEVCSAAVGPYLDAAHDVHEPWSGAGFGLERLLTIREKYSTVKK
GGASISYLNKAKINGS

N346D, C348S, Y384F PylRS (185-454)

MGSSHHHHHHASAPALTKSQTDRLEVLLNPKDEISLNSGKPFRELESELLSRRKK
DLQQIYAEERENYLGKLEREITRFFVDRGFLEIKSPILIPLEYIERMGIDNDTELSK
QIFRVDKNFCLRPMLAPNLYNYLRKLDRALPDPIKIFEIGPCYRKESDGKEHLEEF
TMLDFSQMGSGCTRENLESIITDFLNHLGIDFKIVGDSCMVFGDTLDVMHGDLE
LSSAVVGPIPLDREWIDKWPWIGAGFGLERLLKVKHDFKNIKRAARSESYYNGIS
TNL

3.2.2 Primers used

Primers for site-directed mutagenesis:

N346D-forward: TTTTGTCAGATGGGTTCTGGC

N346D-reverse: GTCCAGCATAGTGAACCTCTCCA

L305H-forward: TACAATTACCTGCGTAAACTGGATC

L305H-reverse: GTGGTTCGGTGCCAGCATC

C348S-forward: CAGATGGGTTCTGGCTGC

C348S-reverse: AGAAAAGTCCAGCATAGTGAACCTC

L309W-forward: CGTAAACTGGATCGTGCACTG

L309W-reverse: CCAGTAATTGTAGTGGTTCGGTGC

S193R-forward: CAGACTGACCGTCTGGAG

S193R-reverse: ACGTTTAGTCAGTGCCGGTG

N203T-forward: CCGAAAGATGAAATCAGCCTG

N203T-reverse: GGTCAGCAGAACCTCCAG

R61K-H63Y-forward: TGCAAGCGCTGTCGCGTGTCTG

R61K-H63Y-reverse: GGTCTTACGGTACTTATAGTGTTTCAGAGCACGTGCAGT

K429M-forward: GTTAAGCACGACTTTAAGAACATCAAAC

K429M-reverse: CATCAGCAGACGTTCCAGACC

431-forward: CAC GAC TTT AAG AAC ATC AAA CGT G

431-reverse: CAT AAC CAT CAG CAG ACG TTC CAG

D433G-forward: TTTAAGAACATCAAACGTGCTGCG

D433G-reverse: GCCGTGCATAACCATCAGCAGA

3.2.3 Protein Expression & Purification

Expression and Quantification of ncAA Incorporated sfGFP134TAG Proteins

TOP10 cells with pBAD-sfGFP134TAG and pEVOL-PylRS variant for each expression test were grown in LB culture (with antibiotics) to OD 0.9-1.1 Then pelleted at 4000 rpm for 10 minutes and washed with PBS for 3 times.

Next, cells were gently resuspended in M9 salts supplemented with 2 mM MgSO₄ and 0.1 mM CaCl₂, 1% LB, 300 μM Leucine and 1 % glycerol. 2.5 mM of N_ε-boc-L-lysine (BocK) and 5 mM of each ncAA derivative unless otherwise stated and induction done by 0.2% arabinose. The expression continued for 12-13 hours at 3 °C. Then the cells, were directly taken to plate reader for fluorescence (ex: 485 nm, em: 528 nm) and absorbance (600 nm) readings. OD normalized fluorescence values (Fluorescence reading/600 nM absorbance reading) were used to compare the incorporation efficiencies of ncAAs.

Expression and Purification of HAr, MeHAr, DiMeHAr Incorporated sfGFP134TAG

TOP10 cells with pBAD-sfGFP134TAG and pEVOL-HArRS (pEVOL-homoargininyl tRNA synthetase) for each expression were grown in LB culture (with antibiotics) to OD 0.9-1.1 Then pelleted at 4000 rpm for 10 minutes and washed with PBS 3 times.

Next, cells were gently resuspended in 5 mL M9 salts supplemented with 2 mM MgSO₄ and 0.1 mM CaCl₂, 1% LB, 300 μM Leucine and 1 % glycerol. 2.5 mM of BocK and 5 mM of each ncAA derivative unless otherwise stated and induction done by 0.2% arabinose. The expression continued for 12-13 hours at 37°C. Next, the cells were

pelleted, washed with PBS and stored at -20°C freezer until purification.

The cell pellets were thawed placing on ice and resuspended in 500 μ L lysis buffer (20 mM Tris, 100 mM NaCl, 10 mM imidazole pH 8.0, 1 mg/mL lysozyme) and incubated on ice for 15 minutes. Then, sonication was applied with 20% power output for 1 sec on, 1sec off for 30 seconds and repeated for 3 rounds. The debris was pelleted by spinning down at 14000 rpm for 30 minutes. Supernatant is kept. Then, 50 μ L of Ni-NTA resin (bed volume) was added. After, incubating 10 mins on ice, Ni-NTA resin was spinned down with 2000 rpm for 1 min and supernatant is poured off. The resin washed with 500 μ L washing buffer (20 mM Tris, 100 mM NaCl, 10 mM imidazole pH 8.0) for 3 times. Finally, the target proteins were obtained by treating resin with 250 μ L elution buffer (20 mM Tris, 100 mM NaCl, 250 mM imidazole pH 8.0). SDS-PAGE gel analysis showed the expected molecular weight of the target protein (Figure 68A, 68B). ESI-MS analysis carried out to confirm incorporation of the ncAA.

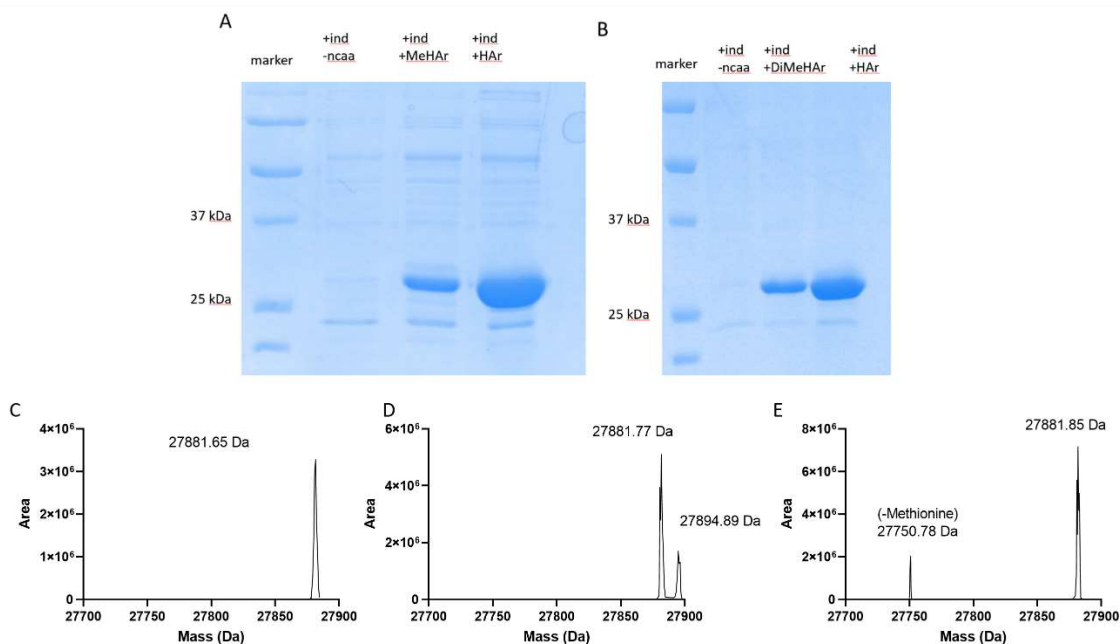


Figure 68 **A)** SDS-PAGE Analysis of purified sfGFP134 with HAr and sfGFP with MeHAr **B)** SDS-PAGE analysis of sfGFP134 with DiMeHAr and sfGFP134 with HAr **C)** ESI-MS analysis of sfGFP134 with HAr (calculated: 27882.42 Da observed: 27881.65 Da) **D)** ESI-MS analysis of sfGFP134 with MeHAr (calculated: 27896.43 Da observed: 27881.77 Da, 27894.89 Da) **E)** ESI-MS analysis of sfGFP134 with DiMeHAr (calculated: 27910.45 Da observed: 27750.78 Da, 27881.85 Da)

Expression and Purification of *o*-ClF Incorporated sfGFP134TAG

TOP10 cells co-transformed with pBAD-sfGFP134TAG and pEVOL-N346D, C348S, Y384F mutations was used. The rest of the procedure is the same as the expression and purification protocol for incorporation of HAr derivatives. The only exception is growth medium contains 5 mM of *o*-ClF instead of HAr derivatives. SDS-PAGE analysis shown below (Figure 69).

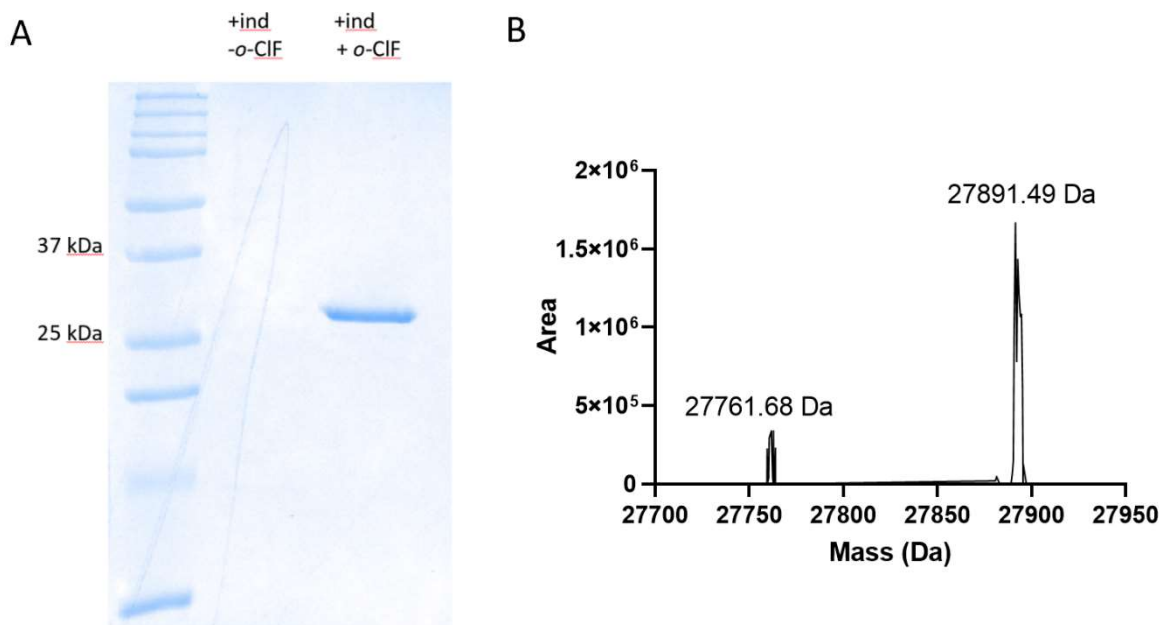


Figure 69 A) SDS-PAGE analysis of sfGFP134-*o*-CIF B) ESI-MS analysis of *o*-CIF incorporated sfGFP

Expression and Purification of N346D, C348S, Y384F PylRS (185-454) for Crystallization

The 40 mL BI21 (DE3) cells containing pETDUET- PylRS (185-454) overnight culture was inoculated in to 2 x 2L Terrific Broth medium. Induction is done with 1 mM IPTG at OD 1.0. The expression went on for 13 hours at 22°C. The cells were washed with PBS buffer and pelleted at 4k rpm for 20 minutes. Then stored at -20°C for a day. Next day, cells from 2L expression were resuspended in 80 mL lysis buffer 10 mM HEPES, 500 mM NaCl, 10 mM βMe, 20 mM imidazole, 1% Tween-20, 10% glycerol (pH 7.3) with 1 mg/mL lysozyme. The cells were incubated for 45 minutes at 4°C. Then the cells were sonicated for 5 cycles 30 seconds each with 10 minutes cooling time. Then cell debris was pelleted at 10k rpm for 30 minutes. The supernatant was separated and incubated with 1 mL Ni-NTA resin for 20 minutes. After eluting the flow through the

resin was washed with 10 mL lysis buffer. Then, the resin washed with 10 mL wash buffer (lysis buffer with no Tween-20, pH 8.0). Finally, the target protein was eluted with 14 mL elution buffer (10 mM HEPES, 500 mM NaCl, 10 mM β Me, 250 mM Imidazole, 10% glycerol, pH 8.0). Next, the protein was further purified on S-100 column which was pre-equilibrated by 10 mM HEPES, 5 mM $MgCl_2$, 300 mM NaCl, 5 mM β Me pH 7.5. Fractions from A8-A20 were collected, concentrated to 15 mg/mL, flash-frozen in liquid nitrogen and stored in $-80^\circ C$ freezer for long-term storage.

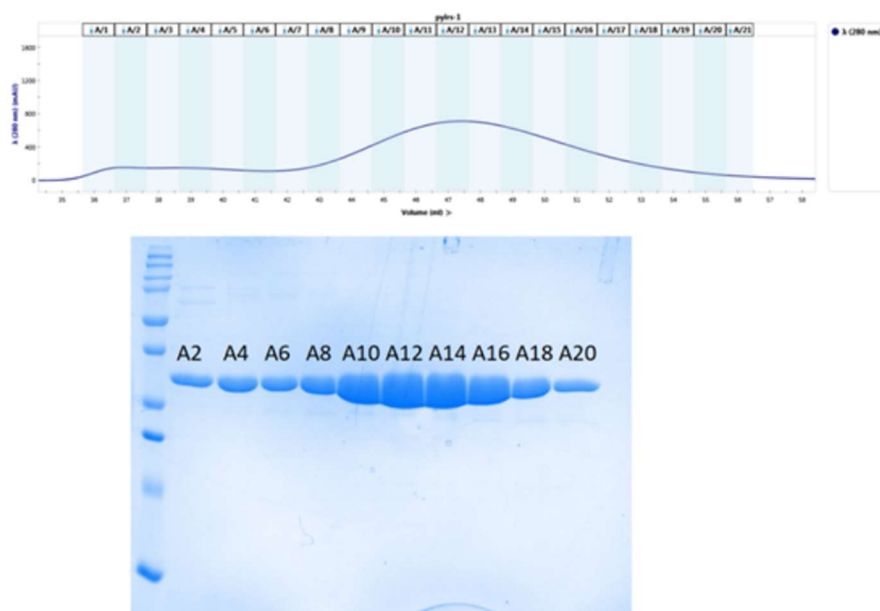


Figure 70 Purification of N346D, C348S, Y384F PyIRS (185-454). Size-exclusion column purification chromatogram and analysis of collected fractions with SDS-PAGE analysis

Crystallization Protocol

Protein stock solution with adenylyl-imidodiphosphate (AMP-PNP) and *o*-ClF prepared as following. A 200 μ L of protein solution was thawed from -80°C stock. Stock solution composition: 10 mg/mL of N346D, C348S, Y384F PylRS (185-454) in 10 mM HEPES, 5 mM MgCl_2 , 300 mM NaCl, 5 mM β Me at pH 7.5. Then, 10 μ L from stock solution of 100 mM HEPES, 3M NaCl, 50 mM MgCl_2 , pH 7.5, 0.7 μ L from 0.7 M solution of β Me were mixed to bring the final buffer composition similar protein stock solution concentration.

Then, 60 μ L of AMP-PNP (50 mM stock in water), 30 μ L (*o*-ClF, 100 mM pH 9.4) The mixture was incubated for 1 h on ice. Then, the solution was concentrated down to 133 μ L. Then, final solution was aliquoted and stored at -80°C as protein stock solution with AMP-PNP and *o*-ClF. By using this protein stock solution, hanging drop method with 1:1 mixture of protein stock solution and crystallization buffer (0.1 M Na citrate tribasic dihydrate pH 5.5, 22% w/v Polyethylene glycol 1000) gave crystals overnight. The crystal structure had P1 space group and had 2.15 \AA resolution. Final refinement yielded Rwork/Rfree ratio as 0.2640/0.3240.

3.2.4 Organic Synthesis

Synthesis of N_α -Boc- N_ϵ -Cbz-L-Lysine- O^t Bu Ester (**33**)

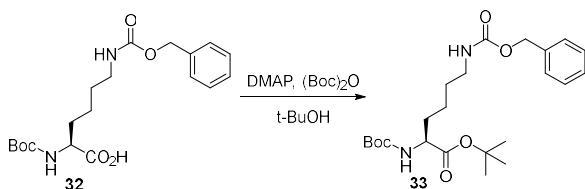


Figure 71 Synthesis of N_α -Boc- N_ϵ -Cbz-L-Lysine- O^t Bu Ester (**33**)

A literature protocol is applied.^[146] To a solution of Boc anhydride (2.0 g, 9.16 mmol) and Lysine-acid compound (2.5 g, 6.65 mmol) in *tert*-butanol (6.5 mL), 4-dimethylaminopyridine (DMAP) (192 mg, 0.3 equiv) in *t*-butanol (1 mL) was added at room temperature. After addition of DMAP excess gas evolution was observed. After stirring for 9 h, the solvent was removed under reduced pressure and the residue was purified by column chromatography (EA/Hexane: 3:7). 2.3g of the product is obtained in 81% yield.

$^1\text{H NMR}$ (CDCl_3 , 300 MHz): δ 1.17-1.23 (6H, m), 1.43 (9H, s) 1.46 (9H, s) 3.19 (2H, q, $J = 6.6$ Hz), 4.15 (1H, m), 5.09 (2H, s), 7.53 (5H, m)

Synthesis of N_α -Boc-L-Lysine- O^t Bu Ester (**34**)

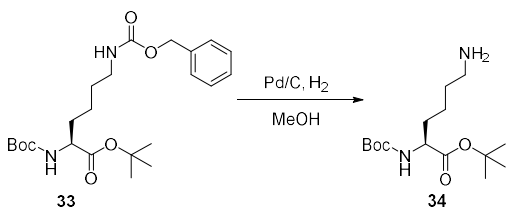


Figure 72 Synthesis of N_α -Boc-L-Lysine- O^t Bu Ester (**34**)

A literature protocol is adopted.^[146] 50 mL RB flask was charged with 530 mg of 10% Pd/C (0.5 mmol) and stir bar was purged with nitrogen for 20 mins. Then solution of 1.9 g of lysine

compound (5 mmol) in 20 mL methanol was added dropwise to RB flask. Then another 20 mins nitrogen purging applied. Finally, hydrogen balloon was added and reaction kept going for 2 hours. TLC analysis indicated the consumption of the starting material. Reaction was filtered through pad of celite. Celite washed couple of times with methanol to yield 1.1 g of target compound in 73% yield.

$^1\text{H NMR}$ (CD_3OD , 300 MHz): δ 1.05-2.25 (4H, m), 1.47 (9H, s), 1.45 (9H, s) 2.79 (2H, br s), 4.16 (1H, br s), 5.04 (2H, br s)

Synthesis of N_α -Boc- N_ϵ -CN-L-Lysine- O^tBu Ester (**35**)

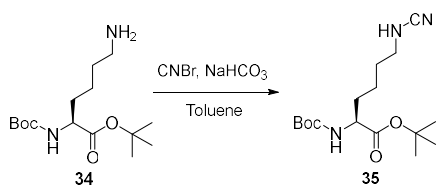


Figure 73 Synthesis of N_α -Boc- N_ϵ -CN-L-Lysine- O^tBu Ester (**35**)

A 10 mL RB flask was charged with lysine compound (631 mg, 2.08 mmol) and NaHCO_3 (612 mg, 7.28 mmol). Then, solution of cyanogen bromide (441 mg, 4.16 mmol) in 4 mL toluene was added to the mixture and reaction started at room temperature. After 2 hours TLC still indicated starting material so, an additional of 305 mg of NaHCO_3 and ~300 mg of cyanogen bromide in 1 mL toluene was added to reaction mixture. The reaction kept going for additional 8 hours. Then, water added to reaction mixture and product was extracted with 3 ethyl acetate washes. Next, ethyl acetate layers were combined, washed with brine and then, dried over magnesium sulfate to yield the crude product. Next, the product was purified over flash column using 40% EtOAc in hexane TLC condition. 368 mg of pure target compound was obtained in 77% yield. This compound was previously reported in literature.^[147]

^1H NMR (CDCl_3 , 300 MHz): δ 1.44 (9H, s), 1.47 (9H, s), 1.57-1.84 (6H, m), 3.04-3.20 (2H, m), 3.77 (1H, br s), 4.17 (1H, br s), 5.10 (1H, d, $J = 6.5$ Hz)

^{13}C NMR (CDCl_3 , 75 MHz): δ 21.88, 27.96, 28.29, 28.89, 32.55, 45.76, 53.51, 79.89, 82.116.60, 155.63, 171.15

Synthesis of N_α -Boc- N_θ -Methyl-L-Homoarginine- O^tBu Ester (**36**)

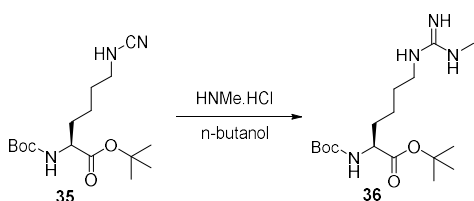


Figure 74 Synthesis of N_α -Boc- N_θ -Methyl-L-Homoarginine- O^tBu Ester (**36**)

Cyano lysine (344 mg, 1.05 mmol), and solvent mixture (3.5 mL Toluene/0.5 mL *tert*-butanol) was charged to RB flask. Then, methylamine hydrochloride (70 mg, 1.05 mmol) was added to RB flask. Then the RB was heated up to 80 degrees. Then, after 6 hours no change in starting material is observed and methylamine hydrochloride does not seem to be dissolved. After that, toluene-*tert*-butanol was evaporated under reduced pressure. Then, 4 mL of n-butanol was added and reaction mixture heated up to 90 degrees and still no product was observed. Next, when the temperature raised to 100 °C a slight image on TLC showed evidence for the formation of the product. Then the temperature raised further to 120 °C and reaction ran overnight. Next, column chromatography applied with 1% NH_4OH , 10% MeOH in 90% DCM. 24 mg of the target compound was obtained in 12% yield.

^1H NMR (CD_3OD , 300 MHz): δ 1.35 (9H, s), 1.39 (9H, s), 1.45-1.81 (4H, m), 2.74 (3H, s), 3.11 (2H, t, $J = 6.71$), 3.90 (1H, t, $J = 7.03$)

Synthesis of N₀-Methyl-L-Homoarginine HCl salt (MeHAr) (37)

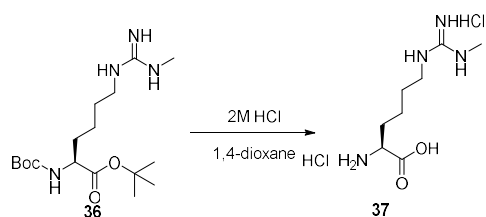


Figure 75 Synthesis of N₀-Methyl-L-Homoarginine HCl salt (37)

An RB flask was charged with 24 mg (0.12 mmol) of the boc-protected compound and 2 mL of 2 M HCl in 1,4-dioxane. The reaction went overnight. Then the volatiles were evaporated under reduced pressure and the target compound was obtained.

¹H NMR (D₂O, 300 MHz): δ 1.27-1.74 (4H, m), 1.75-2.02 (2H, m), 2.74 (3H, s), 3.13 (2H, t, *J* = 6.77 Hz), 4.02 (1H, t, *J* = 6.31)

Synthesis of N1-Methyl-N1-Boc-N2-Boc-Thiourea (39)

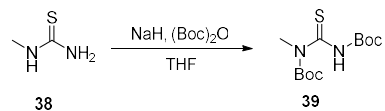


Figure 76 Synthesis of N1-Methyl-N1-Boc-N2-Boc-Thiourea (39)

A literature protocol is adopted.^[148] A 500 mL RB flask is charged with THF (~200 mL) and placed on ice bath until it reaches thermal equilibrium. N-Methyl thiourea (1.69 g, 18.8 mmol, 20 mL solution in THF), NaH (3.0g, 75 mmol) were added successively. The deprotonation reaction went for 30 mins on ice. Then, boc anhydride (8.0g, 36.7 mmol, 30 mL in solution of THF) is added slowly and reaction kept for another 30 mins on ice. Then, the ice bath was removed and reaction kept going for 4 hours at room temperature. After, 4 hours stir bar observed to be stopped (due to electrical connection problem), and the solution became very

viscous. Then, 50 mL NaHCO₃ added to the reaction mixture for quenching. Contents are added to 450 mL water and extracted with 3x150 mL ethyl acetate. Combined organic layers were dried over sodium sulfate and concentrated under vacuum. Column chromatography applied with 10% ethyl acetate in hexanes to purify 1905 mg of the target compound in 35% yield.

¹H NMR (CDCl₃, 400 MHz): δ 1.50 (9H, s), 1.54 (9H, s), 3.58 (3H, s), 12.11 (1H, br s)

Synthesis of N_α-Boc-N_θ-Methyl-N_θ,N_θ'-Boc-L-Homoarginine-O^tBu Ester (40)

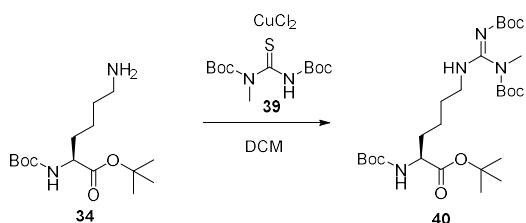


Figure 77 Synthesis of N_α-Boc-N_θ-Methyl-N_θ,N_θ'-Boc-L-Homoarginine-O^tBu Ester (40)

Literature protocol is adapted.^[149] RB flask was charged with Lysine derivative (510 mg, 1.69 mmol), N-methyl diboc thiourea (513 mg, 1.77 mmol), triethylamine (710 uL, 5.07 mmol) and DCM (8.5 mL). To this solution, copper chloride (238 mg, 1.77 mmol) is added and reaction was stirred at room temperature for 40 hours. Then, the compound was absorbed on silica and directly applied to column chromatography without prior treatment. Silica chromatography with pure Hexane/EtOAc (20/80) conducted to yield pure compound (790 mg, 93%).

¹H NMR (CD₃OD, 400 MHz): δ 0.85-1.85 (6H, m), 1.50 (9H, s), 1.49 (18H, br s) 1.46 (9H, s) 3.04 (3H, s), 3.27 (2H, t, *J* = 6.95 Hz), 3.97 (1H, t, *J* = 6.5 Hz)

Synthesis of Methyl Homoarginine (MeHAr) (37)

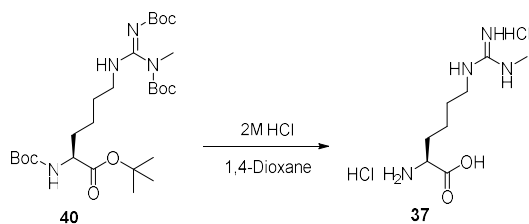


Figure 78 Synthesis of Methyl Homoarginine (MeHAr) (**37**)

Boc compound **40** (740 mg, 1.32 mmol) was dissolved in 5 mL dioxane at room temperature. Then, 5 mL 4M HCl in dioxane added and reaction was stirred overnight (8 hours). Following evaporation of volatiles, 5 mL dioxane and 5 mL 4 M HCl in 1,4-dioxane added again to the compound and reaction kept going for another 6 hours. After evaporation of the volatile components 400 mg of target product was obtained as a yellow compound in %111 yield. The spectra matched with the reported values for synthesis of **37** from **36**.

Synthesis of O-benzyl Thiocyanate (42)

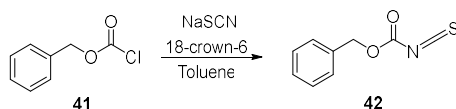


Figure 79 Synthesis of O-benzyl thiocyanate (**42**)

A literature protocol was used.^[150] RB flask was charged with sodium isothiocyanate (419 mg, 5.17 mmol) and the salt was suspended in toluene and evaporated under reduced pressure for two times to remove residual water. Then, 18-crown-6 (60 mg, 0.21 mmol), 11 ml toluene and benzyl chloroformate (0.55 mL, 3.9 mmol) was added to the reaction mixture and stirred at room temp for 15 minutes. Then, the solution was heated to 85 °C with reflux condenser under nitrogen atmosphere. The reaction carried out for

12 hours and then, reaction TLC showed depletion of starting material benzyl chloroformate. The TLC (DCM/hexane 1/4) indicated formation of desired product at Rf 0.3 and an impurity as strong at Rf 0.1. 70 mL hexane added and the reaction mixture was collected by filtration over celite. Then, the solid was washed with DCM/hexane (1/1) and evaporated to yield ~700 mg crude material (yellow oil). The compound kept at -20 freezer until use. Further characterization were not carried out and the compound is used in the next step before further purification.

Synthesis of N_α-Boc-N_ε-Thioamide-N_θ-Cbz-L-Homoarginine-O^tBu Ester (**43**)

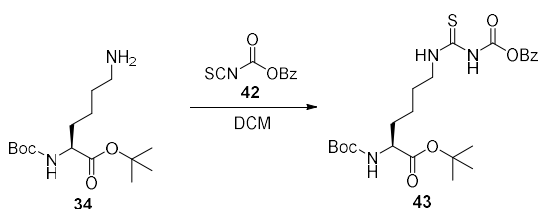


Figure 80 Synthesis of N_α-Boc-N_ε-Thioamide-N_θ-Cbz-L-Homoarginine-O^tBu Ester (**43**)

A literature protocol is adopted.^[151] Free lysine **34** (402 mg, 1.33 mmol) was dissolved in 25 mL DCM on ice bath and crude mixture benzyloxy isothiocyanide (700 mg in 10 mL THF) was added dropwise over 20 mins. Then, ice bath is removed and reaction carried out for 4 hours at ambient temperature. Next, the TLC was done with EtOAc-Hexane (3-7) and concentrated 1-2 mL and applied to flash column directly. 250 mg of product was obtained in 37% yield.

¹H NMR (CDCl₃, 400 MHz): δ 1.33-1.87 (6H, m), 1.46 (9H, s), 1.48 (9H, s), 3.22 (2H, t, *J* = 10 Hz), 4.14-4.23 (1H, m), 5.11 (2H, s), 7.31-7.41 (5H, m)

Synthesis of N_α-Boc-N_θ-Cbz-N_θ,N_θ'-Methyl-L-Homoarginine-O^tBu Ester (44)

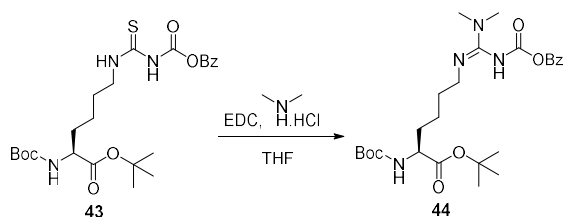


Figure 81 Synthesis of N_α-Boc-N_θ-Cbz-N_θ,N_θ'-Methyl-L-Homoarginine-O^tBu Ester (44)

A literature protocol is adopted.^[151] RB flask was charged with a 2 mL solution of Lysine thiourea (125 mg, 0.25 mmol) and dimethyl amine hydrochloride (31 mg, 0.375 mmol) in THF and cooled down to 0°C on ice bath. Then, DIPEA (130 uL in 0.5 mL THF) was added dropwise. Next, EDCI (72 mg, 0.375 mmol) was added as solid and after few minutes, the ice bath was removed and reaction warmed to room temp. After overnight stirring, The TLC showed starting material mostly did not consume. So, as paper described. More reagents were added. First DIPEA (another 130 uL added), then EDCI (72 mg) and dimethyl amine hydrochloride (31 mg) provided as solids. After few hours TLC did not indicate much change. But after, 3 days the TLC ran with a more polar solvent (10% MeOH in DCM) and showed a polar spot that can be stained by ninhydrin. The purification is done by column chromatography by using the same solvent system used for TLC analysis to yield 78 mg product (61% yield).

¹H NMR (CD₃OD, 400 MHz): δ 1.29-1.79 (6H, m), 1.47 (9H, s), 1.48 (9H, s), 2.99 (6H, s), 3.18 (2H, *J* = 7.16 Hz), 3.94 (1H, t, *J* = 6.93 Hz), 5.10 (2H, s), 7.25-7.44 (5H, m)

¹³C NMR (CD₃OD, 100 MHz): δ 24.06, 28.28, 28.76, 30.40, 32.28, 38.45, 43.80, 55.76, 67.96, 80.34, 82.39, 128.58, 129.32, 139.56, 158.09, 160.18, 163.51, 173.70

Synthesis of Dimethyl Homoarginine (45)

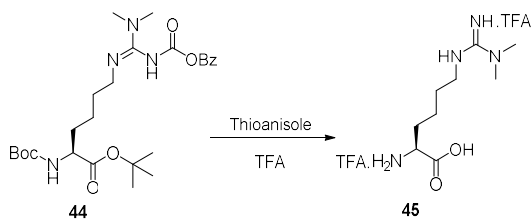


Figure 82 Synthesis of Dimethyl homoarginine (DiMeHAr) (45)

A literature protocol is adopted.^[151] A solution of 4 mL TFA and 1.2 mL thioanisole was added to protected dimethyl HAR (100 mg, 0.2 mmol) and stirred for 1 hour under ambient condition. Then, the volatiles were evaporated under reduced pressure and the residue was dissolved in 6 mL H₂O and extracted with 15 mL Et₂O and two times with 10 mL Et₂O. Next, the water fraction was collected, evaporated and loaded on to C18 column and flash chromatography ran. Ninhydrin positive fractions were collected and dissolved in D₂O for NMR. However, the solution was not homogenous and NMR showed that presence of Cbz group as remaining starting material. Then reaction went on for another 12 hours and final product show only minimal amount of starting material.

¹H NMR (D₂O, 400 MHz): δ 1.36-1.59 (2H, m), 1.65 (2H, p, $J = 7.08$ Hz), 1.83-1.99 (2H, m), 3.00 (6H, s), 3.25 (2H, $J = 6.95$ Hz), 3.85 (1H, t, $J = 6.05$ Hz)

3.3 Results & Discussion

3.3.1 Introducing Step-wise Mutations to Study HAr Incorporation

Step-wise Introduction of Active Site Mutations

L-homoarginine (HAr) has been previously incorporated into proteins by using a PylRS variant with 14 mutations. Only 5 (L305H, L309W, N346D, C348S, Y384F) of these mutations were in the active site of the PylRS enzyme. However, the other 9 mutated positions` (R61K, H63Y, S193R, N203T, L367M, K429M, K431M, D433G and G444E) contribution to enzyme activity were not explained in the paper. In this study, the evolved HAr tRNA synthetase (HArRS) was able to suppress the AGG codon efficiently with only trace amounts of arginine incorporation detected.^[25]

We reasoned that we could introduce these mutations one at a time by starting with the ones that we think are most significant to enzyme activity. By doing so, we might understand which residues are more important for enzyme activity and we could have an enzyme variant with less mutation that can show a similar level of activity. Moreover, we could introduce mono and dimethyl homoarginines (MeHAr, DiMeHAr) and to study them as methylated arginine mimics.

We chose the R3-11 PylRS variant with Y384F mutation a previous group member Dr. Vangmayee Sharma discovered.^[152] This variant contains mutations R19H, H29R, T122S on the N-terminus region of PylRS showing increased incorporation with various ncAAs. The R19H and H29R mutations are on the tRNA binding site and T122S is on the flexible region connecting the N-terminus domain and C-terminus domain.

Therefore, these mutations may be working in two different ways: 1) increasing PylRS-tRNA interaction and 2) altering the flexible linker region of PylRS and increasing in vivo stability of the enzyme.

Asn346 residue was previously denoted as the “gatekeeper” of the PylRS active site.^[3] The asparagine side chain carbonyl group forms a hydrogen bond with the oxygen atom of the l-pyrrolysine (Pyl) side chain (Figure 83A). The analog of the oxygen atom in the side chain of pyrrolysine is the NH moiety in HAr (Figure 83B).

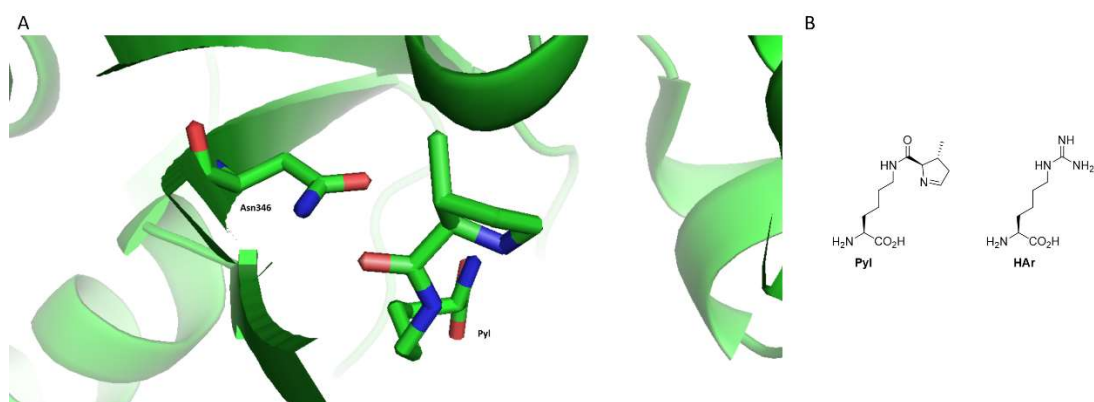


Figure 83 A) Interaction of Asn346 with Pyl side chain (PDB ID: 2ZCE). Oxygen atoms are shown red, nitrogen atoms are shown in blue. B) Structures of Pyl and HAr

Therefore, it is safe to assume that the N346D mutation presumably forming a salt bridge between aspartic acid and guanidino group in HAr is an essential mutation.

To measure fluorescence intensities, we used the GFP reporter system. In this system, TOP10 cells were transformed with mutant PylRS and sfGFP134TAG. The cells were grown in 5 mL culture tubes and the sfGFP expression was induced by the addition of inducer and non-canonical amino acid (ncAA). If the provided ncAA is a good substrate for the mutant PylRS the amino acid incorporation enabled full-length

fluorescent sfGFP formation yielding fluorescent signal. However, for inefficient mutant PyIRS ncAA pairs, the release factor terminated translation to yield truncated protein that was not fluorescent. The fluorescence intensities for positive control were used to assess the extent of ncAA incorporation for the given mutant PyIRS-ncAA pair (Figure 84).

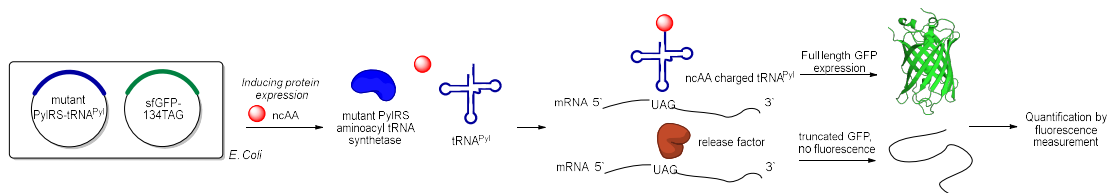


Figure 84 GFP reporter assay to quantify ncAA incorporation

However, when we tested N346D, Y384F mutant with our GFP reporter assay, we observed little to no incorporation of HAR. Next, we chose to do L305H mutation to generate L305H, N346D, Y384F. Although unlikely, a reason for that is, it is possible that aspartic acid at 346 position could be in its protonated form. The mutation of 305 residue (which is in close proximity to the 346 residue) from leucine to histidine might be functioning as a base to remove the hydrogen atom from the N346D residue. Once, the L305H mutation was introduced, the result still showed almost no incorporation (Figure 85).

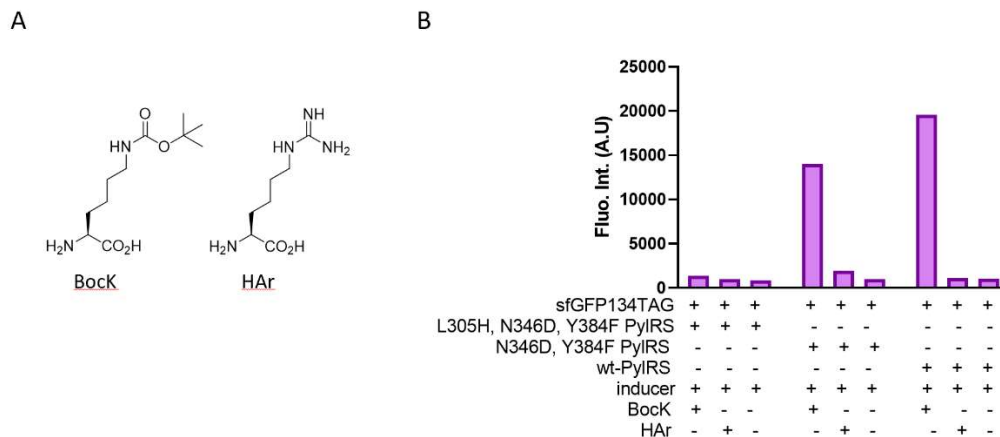


Figure 85 A) Structures of the amino acids tested for incorporation. **B)** Results of the sfGFP reporter assay. BockK and wt-PylRS are used as positive control. The expressions which contained inducer but no supplemented amino acid serves as the negative control.

Next, C348S mutation was done to generate L305H, N346D, C348S, Y384F. Changing a cysteine residue to serine residue could have opened more size in the active site pocket and might have improved incorporation. However, this also yielded a very low incorporation rate (Figure 86). In this assay, phenylalanine incorporation was also tested. Later, the misincorporation of phenylalanine will be discussed in detail.

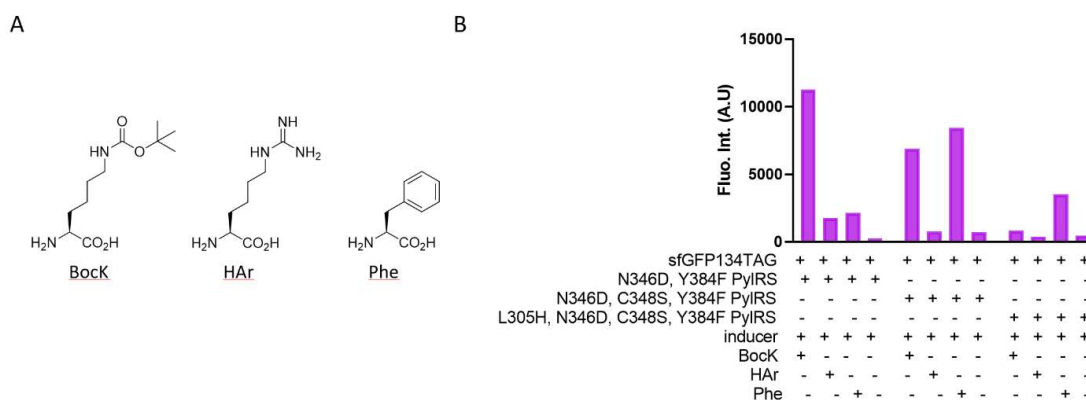


Figure 86 A) Amino acids used **B)** Expression levels of sfGFP

After having 4 out of 5 active site mutations installed, the incorporation efficiency remained very low. The only active site mutation that left was L309W. When this mutation was introduced, the observed increase over fluorescence was around 3-fold only (Figure 87). Given the success of reassigning the AGG codon to HAr this level of incorporation was well below our expectations. Therefore, we decided to examine the 8 other mutations.

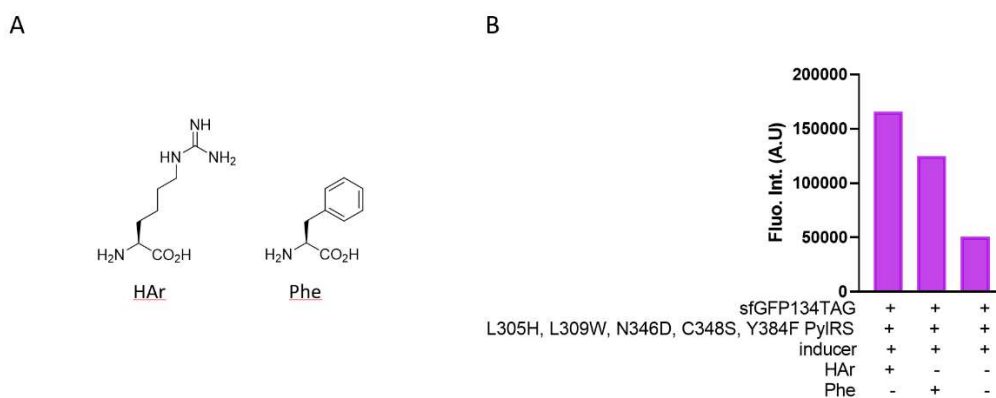


Figure 87 **A)** Amino acids used **B)** Expression levels of sfGFP

Step-wise Incorporation of Mutations Outside the Active Site

We first introduced S193R mutation followed by N203T. Then, we introduced R61K, H63Y mutations. After that, C-terminus mutations (K429M, K431M, D433G) were generated. Once we analyzed the results, we observed the expected high incorporation rates. S193R and N203T mutations seemed to have a significant effect on the incorporation. Introducing R61K, H63Y following S193R and N203T mutations surprisingly decreased the incorporation rate. This decrease might be due to the three additional mutations on the active site, which are also in the tRNA binding domain like

R61K and H63Y. The most significant contribution to incorporation came from the three C-terminus modifications that are K429M, K431M, D433G. These three modifications increased the incorporation level to BocK incorporation with wt-PylRS which is the standard positive control we use in our experiments (Figure 88).

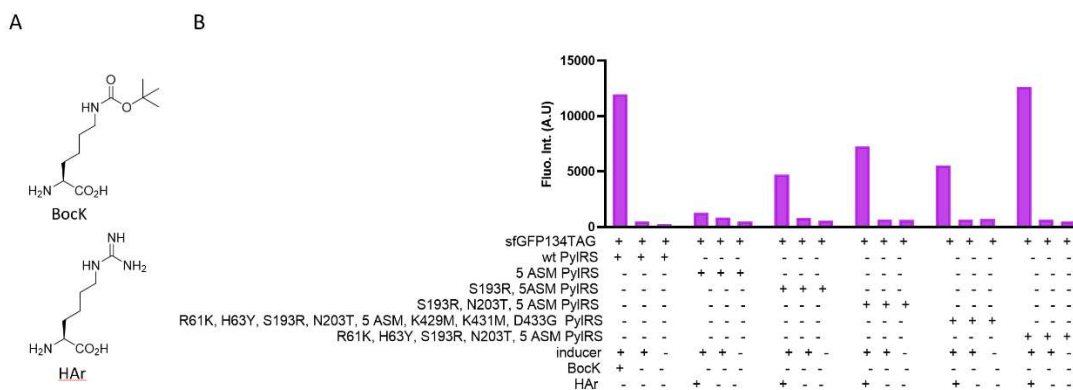


Figure 88 A) Amino acids used B) Expression levels of sfGFP

We finally introduced the final mutation L367M and observed that it increased the incorporation level to be better than BocK incorporation with wt-PylRS (Figure 92). At this point, we were curious whether all the 8 mutations generated outside the active site would have a broad range effect on different PylRS substrates. We introduced the 8 mutations to wt-PylRS to see how the PylRS activity will be affected. The result was a total loss of activity. We know from a literature study that mutations R61K, H63Y, S193R, N203T increase incorporation efficiency on different substrates. When we tried to introduce the mutations R61K, H63Y, S193R, N203T, K429M, K431M, D433G to wt-PylRS we observed total loss of activity. Our current understanding of this enzyme suggests that 429, 431, 433 positions, although not on the active site, probably have long-range effects on the active site structure. However, this hypothesis needs to be

tested by a crystallography study to reach conclusions.

In 2020, Dr. Yane-Shih Wang published a study investigating the function of the other 9 mutations.^[153] In this study, when R61K, H63Y, S193R mutations were introduced to wt-PylRS, the modified enzyme showed a higher incorporation rate with its own substrates. This finding suggests N-terminus modifications do not have long-range effects on the active site but, contribute to PylRS *in vivo* stability or promote its interaction with its cognate tRNA.

Misincorporation of Phenylalanine

N346D, Y384F mutant showed a high background incorporation (Figure 86B and Figure 89B, the fluorescence intensity difference between no inducer and inducer). After testing various canonical amino acids, we found the most likely candidate for misincorporation was l-phenylalanine. Paper, also mentioned that *in vitro* aminoacylation assays show a minimal amount of phenylalanine misincorporation instead of phenylalanine.^[25]

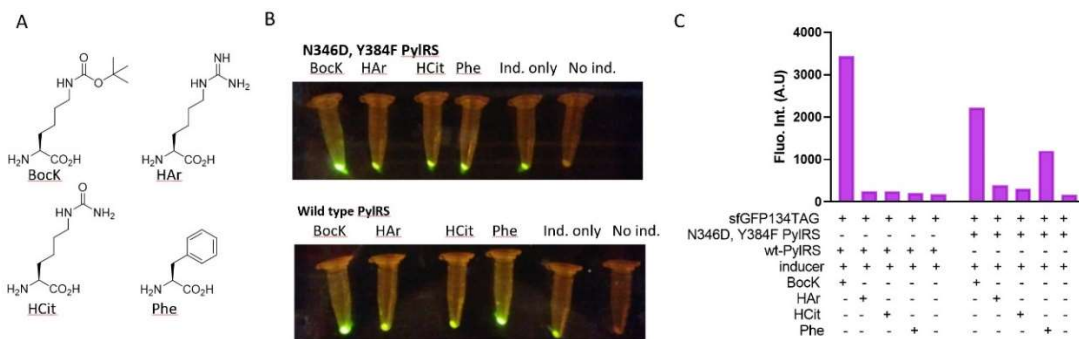


Figure 89 A) Amino acids used B) Cell pellets visualized under blue light to observe sfGFP C) Quantified expression test

Phenylalanine misincorporation is not surprising because, active sites PylRS and PheRS can be aligned (Figure 90A). When PheRS and PylRS are aligned, Asn346 of PylRS

aligns very close to Asp117 residue in PheRS (Figure 90). So, having N346D mutation should be a major contributor to observed phenylalanine incorporation.^[154]

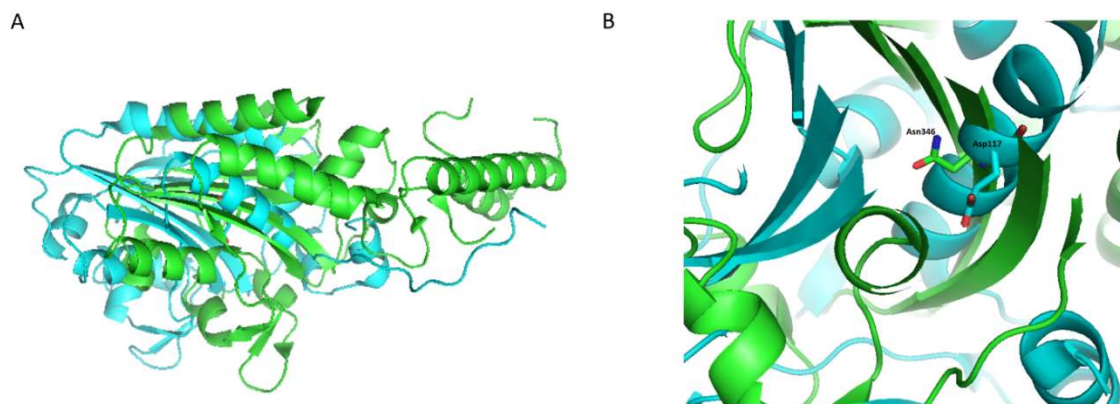


Figure 90 **A)** *Mm* PylRS (PDB ID: 2Q7H) shown in green, *E. coli* PheRS (PDB ID: 3PCO) shown in cyan color **B)** Asn346 of *Mm* PylRS and Asp117 of *E. coli* PheRS demonstrated. Nitrogen shown in blue color, oxygen shown in red color.

3.3.2 Organic Synthesis and Incorporation of MeHAr and DiMeHAr

With an efficient enzyme at hand, we turned our attention to incorporating MeHAr, DiMeHAr. Initially, we tried to synthesize MeHAr via cyano-lysine **35** intermediate (Figure 91A). Reacting **34** with cyanogen bromide gave us the cyano lysine intermediate **35**. Next, we reacted methyl amine hydrochloride at high temperature to obtain the protected MeHAr **36**. Following deprotection, we obtained our desired compound **37**. However, this method was not very convenient because, the intermediate **36** was a free amine and was tedious to purify it via silica gel chromatography. Therefore, we changed our synthesis scheme to work with protected compounds that can be easily purified. We used the lysine thiourea **43** as our key intermediate. We adopted a literature procedure that generates guanidine groups from thiourea by using copper salts.^[149] This route allowed us to obtain our target **37** with high yield (Figure 91B).

However, the same method did not work for synthesizing the DiMeHAr **45**.

Therefore, we adopted another route (Figure 91C) to obtain our target compound.

Reaction of the lysine thiourea derivative **43** with EDC.HCl and dimethyl amine yielded the protected DiMeHAr compound **44**. Following deprotection we obtained our target compound DiMeHAr **45** (Figure 91).

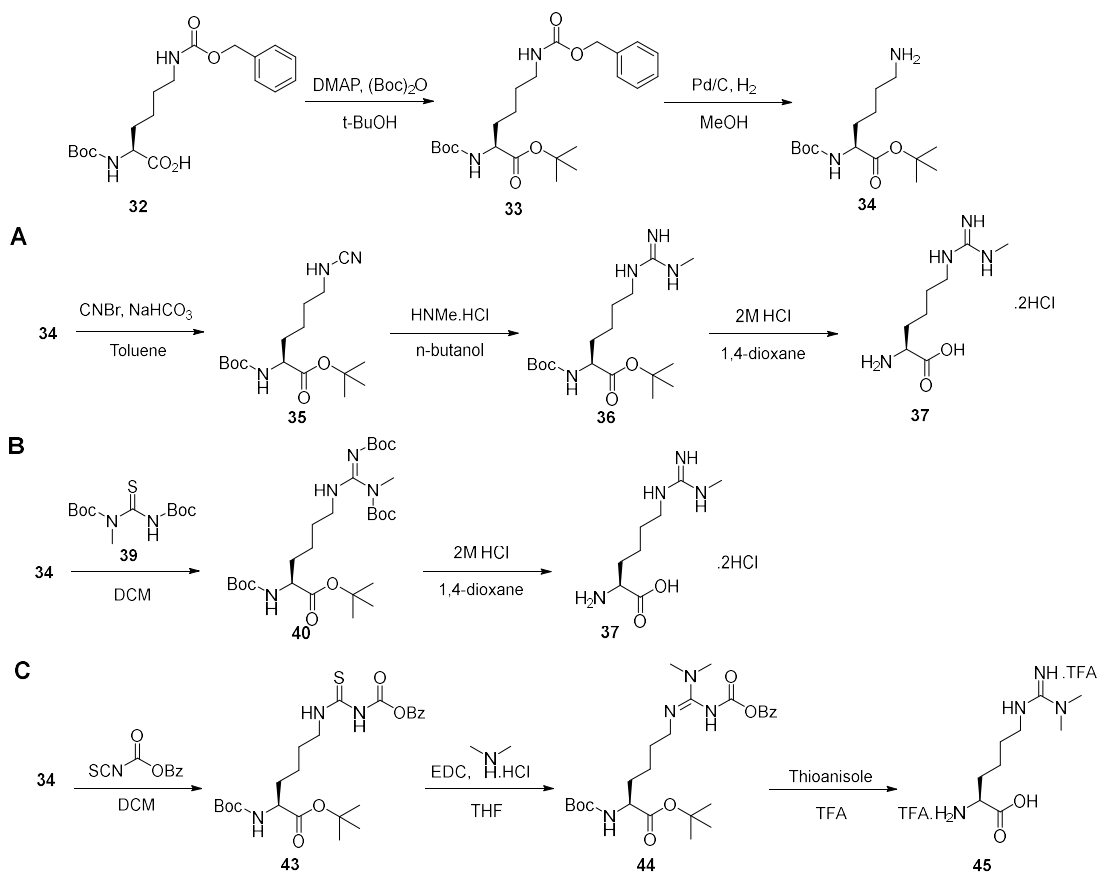


Figure 91 A) First route to synthesis of MeHAr **37** B) Second route to synthesis of MeHAr (**37**) C) Synthesis of DiMeHAr **45**

Incorporation of MeHAr and DiMeHAr

Next, we moved on to check the incorporation of MeHAr and DiMeHAr and compare the incorporation to our positive control BocK with wt-PylRS. The result showed that both MeHAr and DiMeHAr, albeit at a lower level, can be incorporated into proteins (Figure 92).

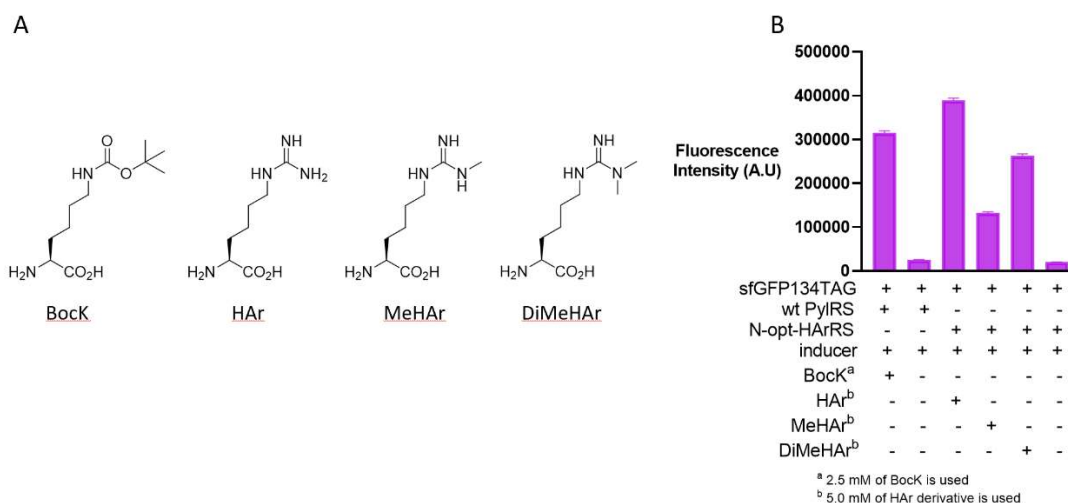


Figure 92 A) Amino acids used B) Expression levels of sfGFP

Then, we wanted to confirm the identity of the ncAAs incorporated by ESI-MS analysis since MeHAr and DiMeHAr have never been introduced to proteins before. First, we verified the incorporation of HAr into sfGFP. ESI-MS analysis of sfGFP134 with HAr showed 27881.65 Da as the most intense peak which is reasonably close to the calculated average mass (27882.42 Da) (Figure 68C). Analysis of sfGFP134 with MeHAr showed two major peaks that are 27881.77 Da, 27894.89 Da (calculated value is 27896.43 Da). This result suggests that in addition to incorporating MeHAr, a demethylation or demethylimination process is also taking place (Figure 68D). Analysis

of sfGFP134 with DiMeHAr did not show a molecular weight value close to the incorporation of DiMeHAr. However, it showed a signal corresponding to demethylation or demethyelimination (calculated: 27910.45 Da observed: 27750.78 Da, 27881.85 Da) (Figure 68E).

An interesting question is the reason for these shifts in molecular weight. Is it due to demethylation or demethyelimination? (Figure 67B, 67C) Is demethylation or demethyelimination taking place before or after the incorporation? We are carrying out MS-MS analysis to identify the exact molecular weight to distinguish between demethylation and demethyelimination. Because, demethyelimination yield homocitrulline instead of homoarginine, a 1 Da difference should be observed.

3.3.3 Incorporation of *o*-ClF

Next, we wanted to test the incorporation of other phenylalanine derivatives. Among the tested phenylalanine derivatives, *ortho*-chloro-l-phenylalanine (*o*-ClF) showed the highest incorporation rate (Figure 93A). It can be seen that, a significant loss of activity occurs as the substituent in the *ortho* position becomes bulkier (Figure 93B).

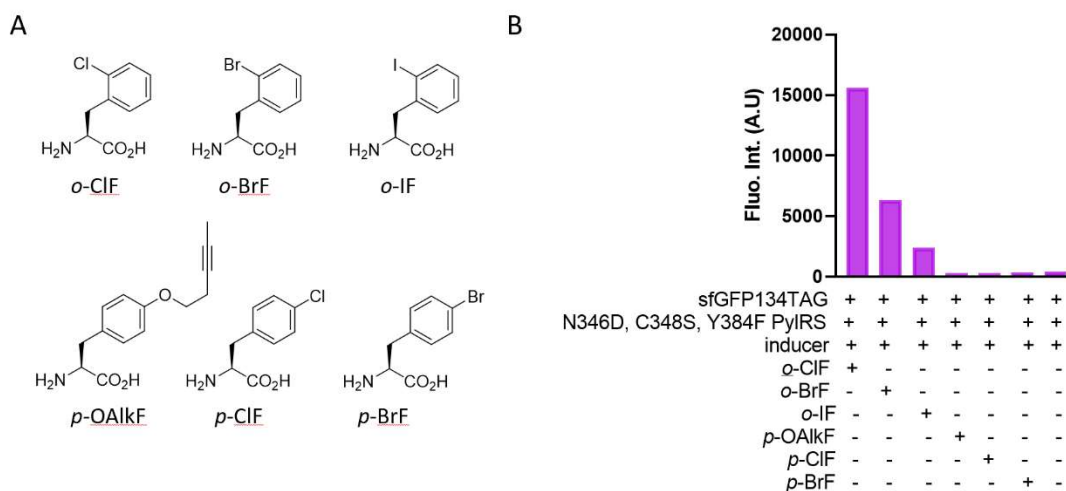


Figure 93 **A)** Amino acids used **B)** Expression levels of sfGFP

Next, we screened all the *ortho*-substituted phenylalanines available in our group to see which substituents will be favored more (Figure 94A). Then, we found that *o*-ClF was the best substrate followed by *ortho*-cyano-l-phenylalanine and *ortho*-methyl-l-phenylalanine (Figure 94B). The incorporation of *o*-ClF was verified by ESI-MS analysis as well (Figure 69B). Although we did not do further study with *ortho*-cyano-l-phenylalanine, the cyano group can be a useful probe for IR spectroscopy analysis of proteins.

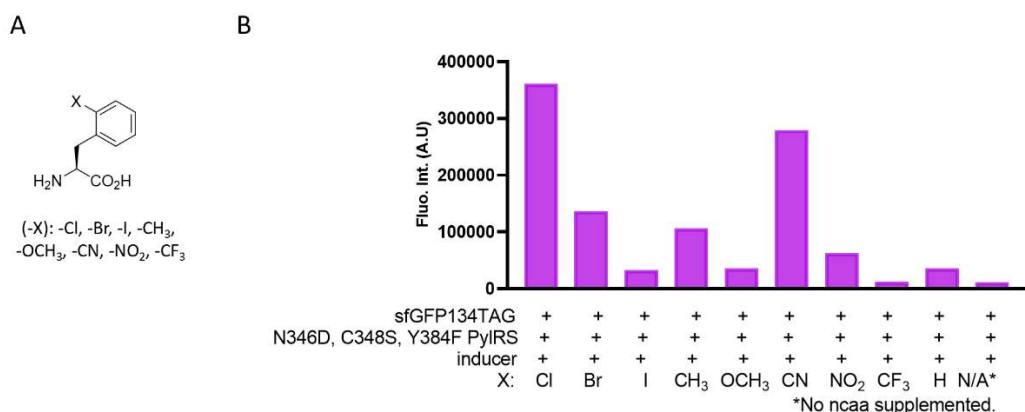


Figure 94 A) Amino acids used **B)** Expression levels of sfGFP

Next, we did further tests with other phenylalanine derivatives (Figure 95A) and confirmed that *ortho*-substituents are the only ones that the N346D, C348S, Y384F-PylRS prefers. Surprisingly *m,p*-F,F-F did not show any incorporation at all (unlike phenylalanine), suggesting the hydrogens of *meta* and/or *para* position of the phenyl ring might involve in an interaction with the enzyme (Figure 95B).

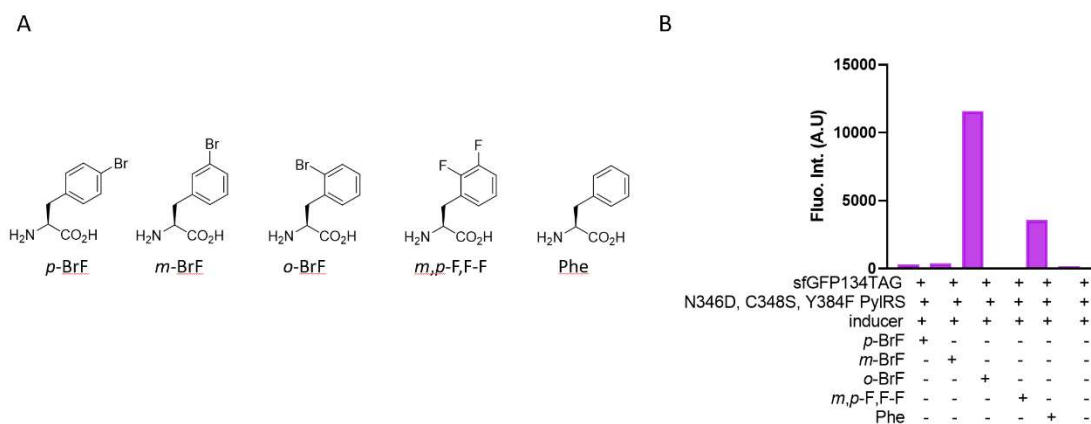


Figure 95 A) Amino acids used **B)** Expression levels of sfGFP

Incorporation of ncAAs with Aromatic Side Chains: L-Histidine and Pyridyl-L-Alanine Derivatives

Then, we further tested N346D, C348S, Y384F PylRS mutant with histidine derivatives (Figure 96A) and found out that, it can incorporate methyl histidine at significant levels as well (Figure 96B). Among the ncAAs tested, 1-methyl-L-histidine (1-MeH) showed significant incorporation. There are two PylRS mutants^{[8], [155]} previously reported incorporating 3-methyl-L-histidine (3-MeH). It is an important structure because, it is a reported post-translational modification. In addition to this, enzymes often use histidine as a base in their active site. Therefore, it would be interesting to incorporate histidine analogs for enzyme engineering.

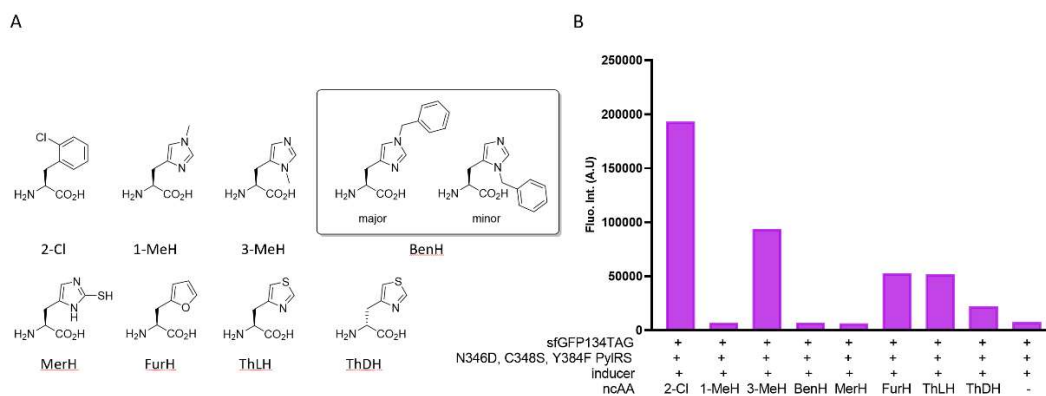


Figure 96 A) Amino acids used **B)** Expression levels of sfGFP

Next, we tried incorporating pyridyl alanine derivatives (Figure 97A). Among the derivatives tested *meta*-pyridyl-L-alanine (*m*-PyAla) showed even higher incorporation than phenylalanine (Figure 97B).

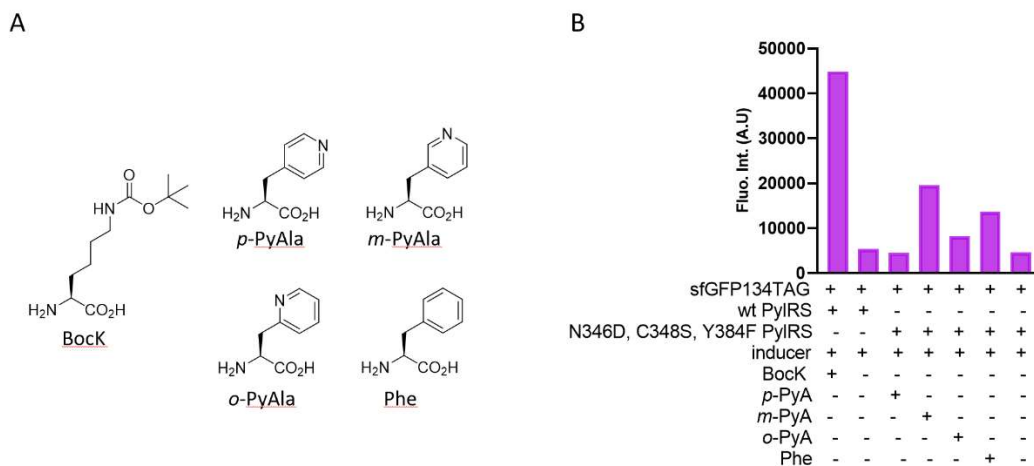


Figure 97 A) Amino acids used B) Expression levels of sfGFP

This result encouraged us to try incorporation with chloro derivative of *m*-PyAla. We tried to synthesize **47** potentially to use for peptide cyclization. In biological systems, cysteines and 2-chloro pyridine derivatives have been shown to react when in proximity.^[156] However, we had trouble synthesizing **47** as a pure compound probably due to intramolecular cyclization (Figure 98). When we tried incorporation with the mixture of **47** and cyclized products, no incorporation was observed (data not shown).

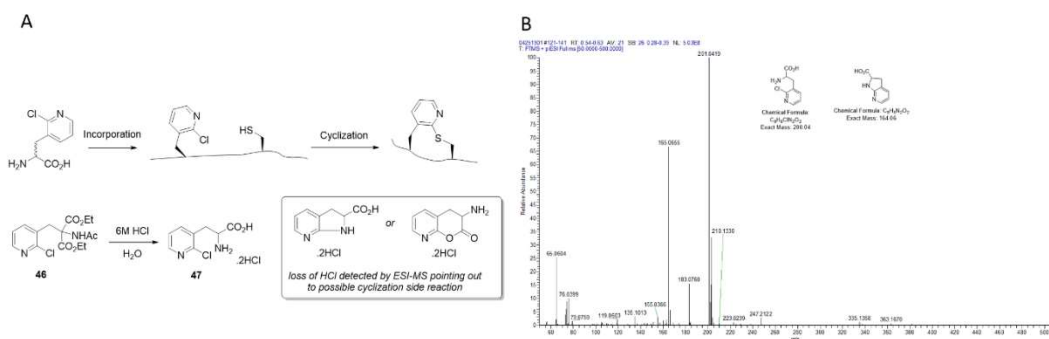


Figure 98 A) **47** as a tool for peptide cyclization B) ESI-MS evidence for cyclization.

The mutations of N346D, C348S, Y384F translated into *Cma* PylRS system as N166D, V168S, Y206F and expression was carried out again which gave similar results as before (Figure 99). If, in the future, the *Cma* PylRS gene is expressed under the control of a stronger promoter such as T5 the expression yield of *o*-ClF is likely to increase.

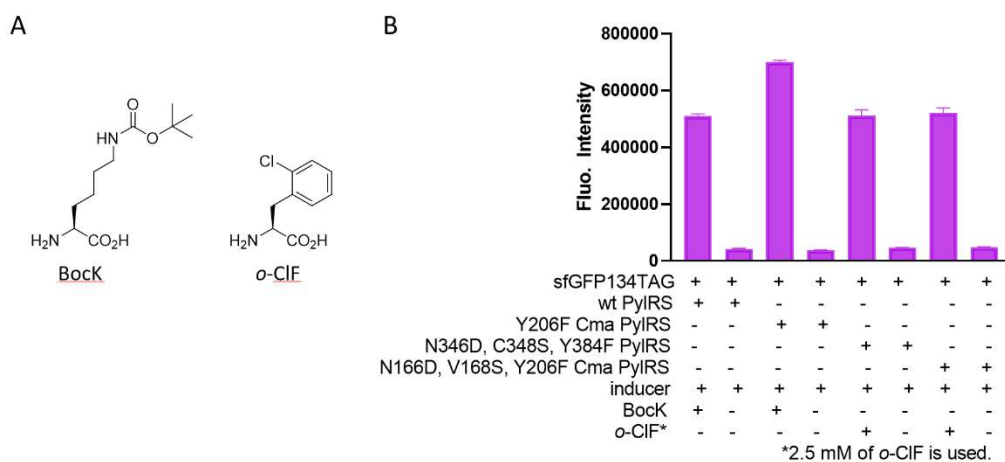


Figure 99 A) Amino acids used B) Expression levels of sfGFP

Crystallization of *o*-Cl-PheRS (N346D, C348S, Y384F-PylRS) with *o*-ClF

We wanted to understand why N346D, C348S, Y384F is showed such a high incorporation rate and decided to obtain the crystal structure of the enzyme-substrate complex. To get the crystal structure, PylRS 185-454 was cloned into pET plasmid and protein is expressed and purified with its C-terminal domain only. The flexible loop between N-terminal and C-terminal domains of PylRS makes it difficult to crystallize it as a whole protein. Because, our aim was to see the active site an *o*-ClF interaction only, we also used the 185-454 PylRS with N346D, C348S, Y384F mutations. Working together with Dr. Kai S. Yang, we crystallized PylRS185-454 N346D, C348S, Y384F

with nonhydrolyzable adenosine triphosphate (ATP) analog adenylyl-imidodiphosphate (AMP-PNP) and *o*-ClF.

In the crystal structure, we observed a clear density for AMP-PNP. After refinement, we were able to fit both AMP-PNP and *o*-ClF structure (Figure 100A). *o*-ClF interacts closely with N346D and C348S residues. The chlorine atom fills in the cavity between N346D and C348S perfectly which seems to be the main reason behind the high incorporation level. The aromatic ring of *o*-ClF is about 3.7 Å away from Trp417 and 3.5 Å away from Tyr306 suggesting the presence of Van-der Waals interactions (Figure 100B). Another close contact of the *o*-ClF group is its carboxylic acid group is 3.3 Å away from Asp346 residue. However, this observation must be confirmed with structures that have better resolution and higher occupancy. The Rwork/Rfree value is 0.2640/0.3240, which shows that more room exists for optimizing the crystallization process.

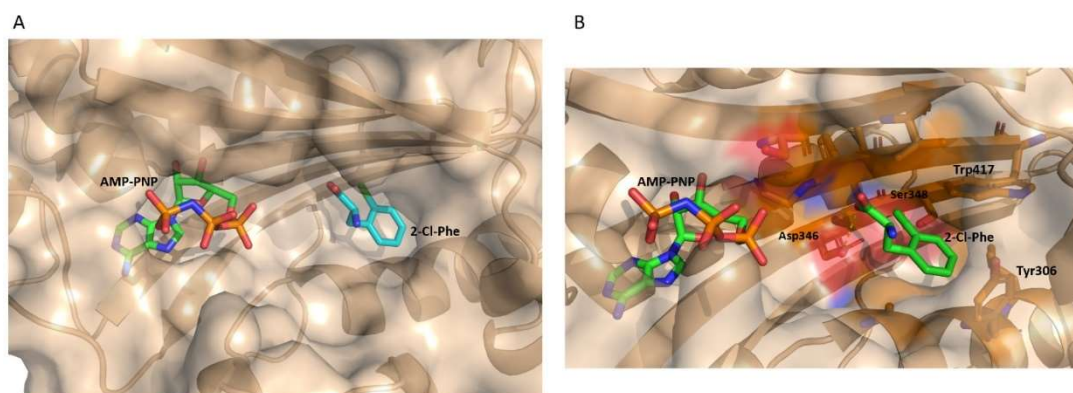


Figure 100 A) AMP-PNP and *o*-ClF in sticks, PyIRS are shown in wheat color **B)** Close contacts of the *o*-ClF group

3.4 Conclusion

The mutations in the PyIRS variant that incorporates L-homoarginine (HAr) were introduced step-wise and the effects of these mutations were studied. Next, MeHAr and DiMeHAr were synthesized and their incorporation into sfGFP134TAG was studied. ESI-MS analysis showed incorporation of only small amounts of MeHAr. The major peak in the ESI-MS spectrum corresponds to a modified MeHAr. For DiMeHAr incorporation, ESI-MS analysis did not show a peak corresponding to desired molecular weight. The major peak (similar to MeHAr) corresponded to a modified DiMeHAr. Identity of this modified residue is currently being investigated via MS/MS studies. During the step-wise introduction of mutations for HAr incorporation, misincorporation of phenylalanine was detected by GFP reporter assay. Further investigation revealed that *ortho*-chloro-L-phenylalanine (*o*-ClF) could be incorporated into proteins as efficiently as boc-lysine incorporation with wt-PyIRS. Crystallography studies were carried out to understand the reason for the high incorporation yield of *o*-ClF. Although, *o*-ClF is not a “privileged” structure that can participate in bioconjugation reactions or bring unique functionalities, it can be used in phage-display system to obtain increased chemical diversity with high expression yields. Moreover, the fate of MeHAr and DiMeHAr incorporation is an interesting developing story. Being able to identify the enzyme responsible for modifying MeHAr and DiMeHAr residues will be valuable information

CHAPTER IV*³

DRUG REPURPOSING TO TARGET MAIN PROTEASE

4.1 Introduction

Finding timely treatment options is of tremendous importance to alleviate catastrophic damages of COVID-19. However, the short time window that is required to contain the disease is extremely challenging for a conventional drug discovery process that requires typically many years to finalize a drug and therefore might not achieve its goal before the pandemic ceases. In this January, we did a comparative biochemical analysis between severe acute respiratory syndrome-coronavirus 2 (SARS-CoV-2), the virus that has caused COVID-19, and SARS-CoV-1 that led to an epidemic in China in 2003 and proposed that remdesivir was a viable choice for the treatment of COVID-19.^[71] We were excited to see that remdesivir was finally approved for emergency use in the United States and for use in Japan for people with severe symptoms. With only one medicine in stock that provides very limited benefits to COVID-19 patients,^[157] the virus may easily evade it, leading us once again with no medicine to use. Given the rapid spread and the high fatality of COVID-19, finding alternative medicines is imperative. Drug repurposing stands out as an attractive option in the current situation. If an approved drug can be identified to treat COVID-19, it can be quickly proceeded to clinical trials

³ DRUG REPURPOSING TO TARGET MAIN PROTEASE is readapted from a published report in PNAS. DOI number: 10.1073/pnas.2012201118

and manufactured at a large scale using its existing good manufacturing practices (GMP) lines. Previously, encouraging results were obtained from repurposing small molecule medicines including teicoplanin, ivermectin, itraconazole, and nitazoxanide.^{[158], [159], [160], [161]} These antimicrobial agents showed antiviral activity against Ebola, Chikungunya, Enterovirus, Influenza viruses respectively.^[162] However, a common drawback of a repurposed drug is its low efficacy level. One way to circumvent this problem is to combine multiple existing medicines to accrue a synergistic effect. To be able to discover such combinations, breaking down the druggable targets of the SARS-CoV-2 to identify drugs that do not cross-act on each other's targets is a promising strategy. For example, a recent study showed that triple combination of interferon β -1b, lopinavir-ritonavir, and ribavirin was safe and superior to lopinavir-ritonavir alone for treating COVID-19 patients.^[163]

Since the release of the first M^{pro} crystal structure, many computational studies have been carried out to screen existing drugs in their inhibition of M^{pro} and many potent leads have been proposed and tested experimentally while this manuscript was being prepared.^{[164], [74]}

We have docked a group of selected FDA/EMA-approved small molecule medicines to the active site of M^{pro} and selected about 30 hit drugs to characterize their inhibition on M^{pro} experimentally. Our results revealed that a number of FDA/EMA-approved small molecule medicines have high potency in inhibiting M^{pro} and bepridil inhibits cytopathogenic effect (CPE) induced by the SARS-CoV-2 virus in Vero E6 and A549/ACE2 cells with low micromolar EC₅₀ values.

4.2 Experimental

4.2.1 Docking Study

Docking studies were done by Dr. Wenshe Liu. 55 FDA/EMA-approved small molecule medicines that have several aromatic or large hydrophobic moieties inter-connected and did a docking analysis of their binding to M^{pro} (PDB ID: 6lu7). Autodock was the program we adopted for the docking analysis. The covalent ligand and non-bonded small molecules in the structure of 6lu7 was removed to prepare the protein structure for docking. Four residues His41, Met49, Asn142, and Gln189 that have shown conformational variations in the SARS-CoV-1 enzyme were set flexible during the docking process. A genetic algorithm method with 100 runs was carried out to dock each small molecule medicine to the enzyme.

4.2.2 Expression of sf-GFP-M^{pro}-6xHis

pBAD-sfGFP-M^{pro} plasmid was constructed from pBAD-sfGFP. The M^{pro} gene was inserted between DNA sequences that coded sfGFP and 6xHis. The overall sfGFP-M^{pro}-6xHis fusion gene was under control of a pBAD promoter. The antibiotic selection marker was ampicillin. To express sfGFP-M^{pro}-6xHis, *E. coli* TOP10 cells were transformed with pBAD-sfGFP-M^{pro}. A single colony was picked and grew in 5 mL LB medium with 100 µg/mL ampicillin overnight. In the next day, we inoculated this starting culture into 5 L 2xYT medium with 100 µg/mL ampicillin in 5 separate flasks at 37 °C. When the OD reached to 0.6, we added L-arabinose (working concentration as 0.2%) to each flask to induce protein expression at 37 °C for 4 h. Then, the cells were

pelleted at 4000 rpm at 4 °C, washed with cold PBS and stored in -80 °C until purification. To purify the expressed protein, we re-suspended frozen cells in a 125 mL buffer containing Tris pH 7.5, 2.5 mM DTT, and 1.25 mg lysozyme. Resuspended cells were sonicated using a Branson 250W sonicator with 1 second on, 4 second off, and a total 5 min 60% power output in two rounds. After sonication, cellular debris was spun down at 16000 rpm for 30 min at 4 °C. Supernatant was collected and recorded the volume. The whole cell lysate analysis showed almost all of the fusion protein was hydrolyzed to two separate proteins sfGFP and M^{Pro}. We were able to obtain an insignificant amount of M^{Pro} when Ni-NTA resins were used for purification. Therefore, ammonium sulfate precipitation was applied. This was done by the addition of a saturated ammonium sulfate solution at 0 °C. The fraction between 30% and 40% of ammonium sulfate was collected. Collected fraction was dissolved in buffer A (20 mM Tris, 10 mM NaCl, and 1 mM DTT at pH 8.0) and the obtained solution was dialyzed against the same buffer to remove ammonium sulfate. Then, this solution was subjected to anion exchange column chromatography using Q sepharose resins. Proteins were eluted from the Q sepharose column by applying a gradient with increasing the concentration of buffer B (20 mM Tris, 1 M NaCl, and 1 mM DTT at pH 8.0). Eluted fractions that contained M^{Pro} were concentrated and subjected to size exclusion chromatography using a HiPrep 16/60 Sephacryl S-100 HR column with a mobile phase containing 10 mM sodium phosphate, 10 mM NaCl, 0.5 mM EDTA and 1 mM DTT at pH 7.8. The final yield of the purified enzyme was 1 mg/L with respect to the original expression medium volume. The concentration of the finally purified M^{Pro} was

determined using the Pierce™ 660nm protein assay and aliquoted 10 μM M^{Pro} in the size exclusion chromatography buffer for storage in -80°C.

M^{Pro} protein sequence

SGFRKMAFPSGKVEGCMVQVTCGTTTLNGLWLDDVVYCPRHVICTSEDMLNP
NYEDLLIRKSNHNFLVQAGNVQLRVIGHSMQNCVLKLVDTANPKTPKYKFVR
IQPGQTFSVLACYNGSPSGVYQCAMRPNFTIKGSFLNGSCGSVGFNIDYDCVSFC
YMHHMELPTGVHAGTDLEGNFYGPFVDRQTAQAAGTDTTITVNVLAWLYAAV
INGDRWFLNRFTTTLNDFNLVAMKYNYEPLTQDHVDILGPLSAQTGIAVLDMC
ASLKELLQNGMNGRTILGSALLEDEFTPFDVVRQCSGVTFQ

4.2.3 Screening Assay

DABCYL-Lys-Thr-Ser-Ala-Val-Leu-Gln-Ser-Gly-Phe-Arg-Lys-Met-Glu-EDANS trifluoroacetate salt was purchased from Bachem (Product #4045664) and used as the substrate. This substrate is referred as FRET-Sub3 in Chapter 5. 5 mM stock solutions of medicines were prepared in DMSO. The final screening assay conditions contained 50 nM M^{Pro}, 10 μM substrate, and 1 mM medicine. Enzyme stock and substrate stock solutions were diluted by using a buffer containing 10 mM sodium phosphate, 10 mM NaCl, and 0.5 mM EDTA at pH 7.6 for reaching desired final concentrations. We ran the assay in triplicates. First, we added 30 μL of a 167 nM M^{Pro} solution to each well in a 96-well plate and then provided 20 μL of 5 mM stock solutions of medicines in DMSO. After a brief shaking, we incubated the plate at 37°C for 30 min. Then 50 μL of a 20 μM substrate was added to the solution to initiate the activity analysis. The EDANS

fluorescence with excitation at 336 nm and emission at 490 nm from the cleaved substrate was detected. The fluorescence was determined by analyzing linear regression of slopes at the initial 5 min and normalized them with respect to the control that had no inhibitor provided.

4.2.4 Half Maximal Inhibitory Concentration (IC₅₀) Assays of Drugs

The final inhibition assay conditions contained 50 nM M^{Pro}, 10 μM substrate, and a varying concentration of an inhibitor. Similar to screening assay, enzyme stock solution was diluted and substrate stock solutions using a buffer containing 10 mM sodium phosphate, 10 mM NaCl, and 0.5 mM EDTA at pH 7.8 for reaching desired final concentrations. The assay was conducted in triplicates. For the inhibition analysis, 30 μL of a 167 nM M^{Pro} solution was added to each well in a 96-well plate and then provided 20 μL of inhibitor solutions with varying concentrations in DMSO. After a brief shaking, the plate was incubated at 37°C for 30 min. Then 50 μL of a 20 μM substrate solution was added to initiate the activity analysis. The fluorescence signal was monitored and processed the initial slopes in the same way described in screening assay part. GraphPad 8.0 was used to analyze the data and used the [Inhibitor] vs. response - Variable slope (four parameters) fitting to determine the values of both IC₅₀ and Hill coefficient.

4.2.5 Live Virus-based Microneutralization Assay

A slightly modified standard live virus-based microneutralization (MN) assay was used as previously described^{[165], [166], [167], [168]} to rapidly evaluate the drug efficacy against

SARS-CoV-2 infection in Vero E6 and A549 cell culture. Briefly, confluent Vero E6 or A549/ACE2 cells grown in 96-wells microtiter plates were pre-treated with serially 2-folds diluted individual drugs in duplicate over eight concentrations for two hours before infection with ~100 and ~500, respectively, infectious SARS-CoV-2 particles in 100 μ L EMEM supplemented with 2% FBS. Vero E6 and A549 cells treated with parallelly diluted dimethyl sulfoxide (DMSO) with or without virus were included as positive and negative controls, respectively. After cultivation at 37 °C for 3 days (Vero E6) or 4 days (A549/ACE2), individual wells were observed under the microcopy for the status of virus-induced formation of CPE and cytotoxicity. The efficacy of individual drugs was calculated and expressed as the lowest concentration capable of completely preventing virus-induced CPE in 100% of the wells with no signs of cytotoxic effect. All compounds were dissolved in 100% DMSO as 10 mM stock solutions and diluted in culture media. The toxicity to the treated cells was assessed by the Cell Cytotoxicity Assay kit (Abcam Cat#ab112118, Cambridge, MA) according to the manufacturer's protocol.

To quantify EC₅₀ values, Vero E6 and A549/ACE2 cells grown in 24-well plates were pre-treated with serially 2-folds diluted drug for two hours before infection with SARS-CoV-2 at an MOI (multiplicity of infection) of 0.5 in 200 μ L EMEM supplemented with 2% FBS. Cells treated with parallelly diluted DMSO with or without virus were included as positive and negative controls, respectively. After incubation for an hour with viral inoculum, cells were washed 3 times with EMEM, and cultivated with fresh medium for 3 days (Vero E6) or 4 days (A549/ACE2). Supernatants from infected

cells were harvested for measuring the infectious virus titers by the tissue culture infective dose (TCID₅₀) assay using Vero E6 cells. Briefly, 50 μ L supernatants from infected cells were serially diluted (10 fold) in EMEM supplemented with 2% FBS; 100 μ L of serially diluted samples were added to VeroE6 cells grown in 96 well plates and cultivated at 37°C for 3 days followed by observation under the microcopy for the status of virus-induced formation of CPE in individual wells. The titers were expressed as log TCID₅₀/mL.

4.3 Results & Discussion

4.3.1 Docking results

The lowest binding energy from the total 100 runs for each small molecule medicine were collected and summarized them in Table 12.

Table 12 Docking results of small molecule medicines (Compounds whose IC₅₀ values were tested are asterisked.)

Name	$\Delta G_{\text{binding}}$	Name	$\Delta G_{\text{binding}}$
Rimonabant*	-11.23	Bepridil*	-8.31
Tipranavir*	-10.74	Isoconazole	-8.15
Ebastine*	-10.62	Econazole	-8.14
Saquinavir*	-10.37	Eluxadoline	-8.12
Zopiclone*	-10.10	(R)-Butoconazole	-8.11
Pimozide*	-10.01	(S)-Butoconazole	-8.10
Pirenzepine*	-9.94	Atazanavir	-8.08
Nelfinavir*	-9.67	Cetirizine	-8.01
Doxapram*	-9.55	Efinaconazole	-8.01
Oxiconazole*	-9.18	Amprenavir	-7.99
Indinavir*	-9.13	Hydroxyzine	-7.99
Sertindole*	-9.04	(R)-Tioconazole	-7.98
Metixene*	-9.01	(R)-Carbinoxamine	-7.96
Fexofenadine*	-8.95	Armodafinil	-7.90
Lopinavir*	-8.91	Desipramine	-7.84
Sertaconazole*	-8.87	Ritonavir	-7.74
Reboxetine*	-8.86	Atomoxetine	-7.73
Ketoconazole*	-8.85	Sulconazole	-7.69
Duloxetine*	-8.79	Clotrimazole	-7.67
Isavuconazole*	-8.77	Dipyridamole	-7.67
Lemborexant*	-8.75	Phentolamine	-7.61
Oxyphencyclimine*	-8.74	(S)-Tioconazole	-7.48
Darunavir*	-8.72	Doxylamine	-7.33
Trihexphenidyl*	-8.72	(S)-Carbinoxamine	-7.21

Pimavanserin*	-8.69	Antazoline	-6.86
Clotiapine*	-8.57	Voriconazole	-6.76
Itraconazole*	-8.44	Fluconazole	-6.41
Clemastine*	-8.36		

Among all 55 small molecule drugs that we used in the docking study, 29 showed a binding energy lower than -8.3 kcal/mol. These molecules were chosen to do further experimental characterizations (Figure 101).

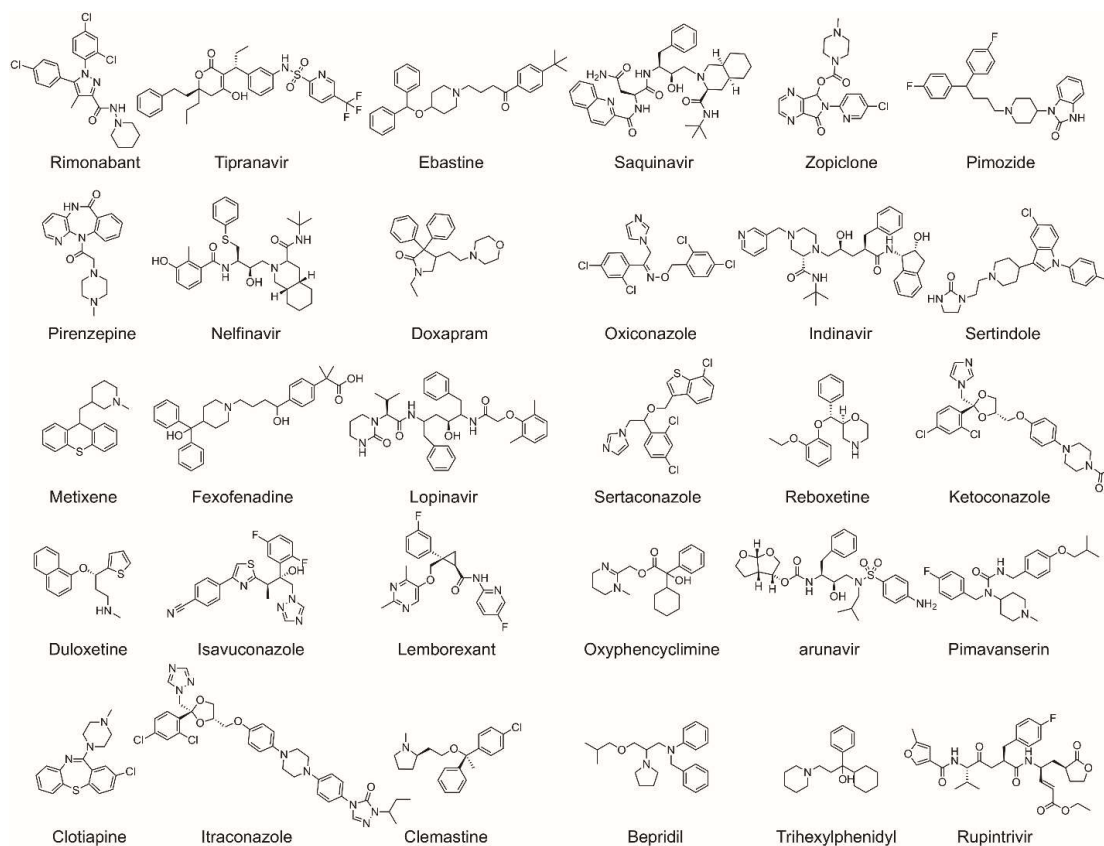


Figure 101 Structures of 29 FDA/EMA-approved medicines and rupintrivir whose IC_{50} values in inhibiting M^{PrO} were determined in the study

4.3.2 Inhibition Analysis of Selected Drugs

All 29 small molecule medicines were purchased from commercial providers and used without further purification and characterization. Rupintrivir is a previously developed 3C protease inhibitor.^[169] It has a key lactone side chain in the P1 residue that has a demonstrated role in tight binding to 3C and 3C-like proteases. Since it has been an investigational antiviral, we purchased it as well with a hope that it could be a potent inhibitor of M^{pro}. Most of the purchased small molecule medicines were dissolved in DMSO to make 5 mM stock solutions and proceeded to use these stock solutions to test inhibition on M^{pro}. Except itraconazole that has low solubility in DMSO, all tested small molecule medicines were diluted to a 1 mM final concentration in the inhibition assay conditions. 20% DMSO were maintained in the final assay condition to prevent small molecule medicines from precipitation. The activity of M^{pro} in 20% DMSO was 29% lower than that in a regular buffer (Figure 102) but satisfied our assay requirement which is to have a fluorescence signal strong enough to produce straight line for reliable and reproducible calculation of initial slopes.

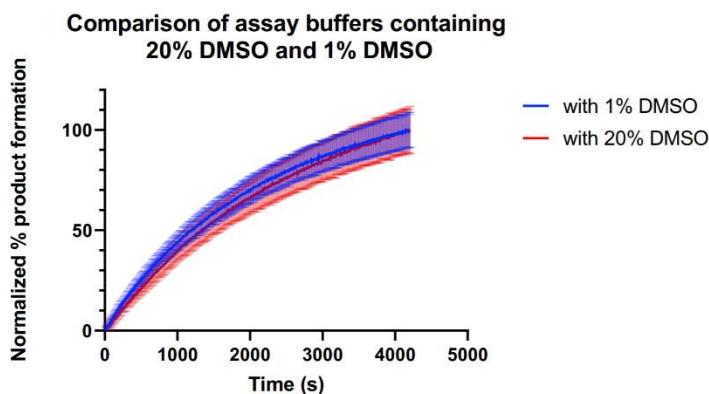


Figure 102 Comparison of assay buffers containing 20% DMSO and 1% DMSO

An M^{pro} activity assay in the absence of a small molecule medicine was set up as a comparison. Triplicate repeats were carried out for all tested small molecules and the control. The results presented in Figure 103 displayed two easily discernable characteristics. First, about half of the tested compounds showed strong inhibition of M^{pro} at the 1 mM concentration level (itraconazole was at 0.14 mM due to its low solubility in DMSO), supporting the practical use of a docking method in guiding the drug repurposing research of COVID-19. Second, several small molecule medicines including fexofenadine, indinavir, pirenzepine, reboxetine, and doxapram clearly activated M^{pro} (> 15%). This was to the contrary of what the docking program predicted. This observation strongly suggests that frontline clinicians need to exhibit caution in repurposing medicines for COVID-19 patients before they are thoroughly investigated on influencing the SARS-CoV-2 biology. A not-well-understood drug might deteriorate the already devastating symptoms in COVID-19 patients. Although it is not the focus of the current study, the observation that M^{pro} can be activated by existing drugs needs to be further investigated.

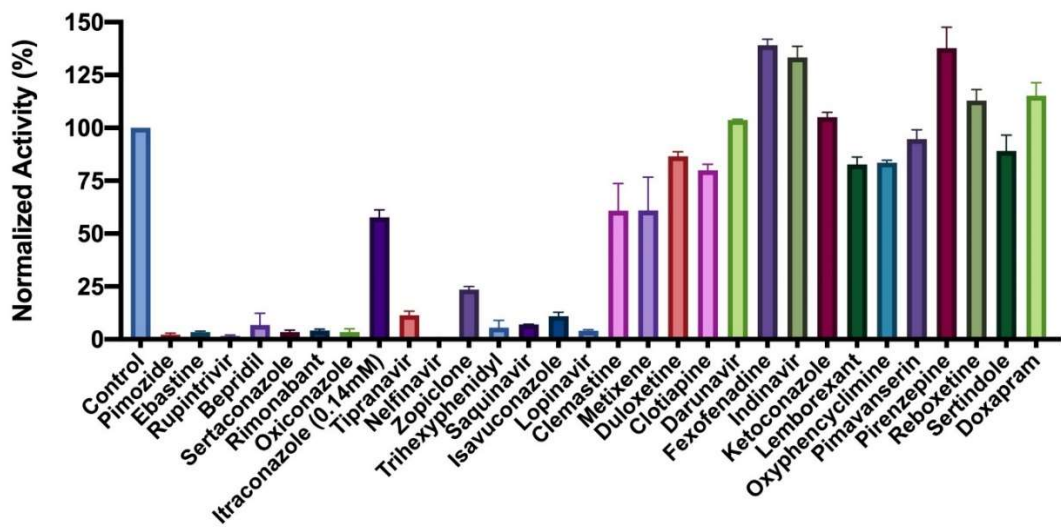


Figure 103 Initial screening of M^{Pro} inhibition by 29 FDA/EMA-approved medicines and rupintrivir

We selected 17 small molecule medicines and rupintrivir that displayed strong inhibition of M^{Pro} to conduct further characterizations of their IC₅₀ values in inhibiting M^{Pro} by varying the small molecule concentration from 1 μM to 10 mM. Results collectively presented in Figure 104 and Table 13 identifies that of the 18 tested compounds, 7 had an IC₅₀ value below 100 μM. These include pimozide, ebastine, rupintrivir, bepridil, sertaconazole, rimonabant, and oxiconazole. There is no strong correlation between calculated binding energies and determined IC₅₀ values. This discrepancy can be explained by limited factors that were involved in the calculation. Pimozide, ebastine, and bepridil were the three most potent FDA/EMA-approved medicines with IC₅₀ values as 42 ± 2, 57 ± 12 and 72 ± 12 μM, respectively. Although rupintrivir is a covalent inhibitor that was developed specifically for 3C and 3C-like proteases, its IC₅₀ value (68 ± 7 μM) is higher than that of pimozide and ebastine. The relatively low activity of rupintrivir in inhibiting M^{Pro} might be due to the change of the

amide bond between the P2 and P3 residues to an methyleneketone. This conversion served to increase the serum stability of rupintrivir but, has likely eliminated a key hydrogen bonding interaction with M^{pro} [99] The repurposing of HIV medicines for the treatment of COVID-19, particularly those targeting HIV1 protease, has been area of much attention. In fact, the cocktail of lopinavir and ritonavir was previously tested in China for the treatment of COVID-19.[170] The IC₅₀ value of lopinavir in inhibiting M^{pro} is about 500 μM, which possibly explains why this anti-HIV viral cocktail demonstrated no significant benefit for treating patients. Nelfinavir was previously shown having high potency in inhibiting the entry of SARS-CoV-2 into mammalian cell hosts. A cell-based study in inhibiting the SARS-CoV-2 entry indicated a 1 μM EC₅₀ value.[171] However, our IC₅₀ determination against M^{pro} resulted in a value of 234 ± 5 μM, highlighting that nelfinavir likely inhibits another key SARS-CoV-2 enzyme or protein or interferes with key cellular processes that are required for the SARS-CoV-2 entry into host cells. These possibilities need to be studied further.

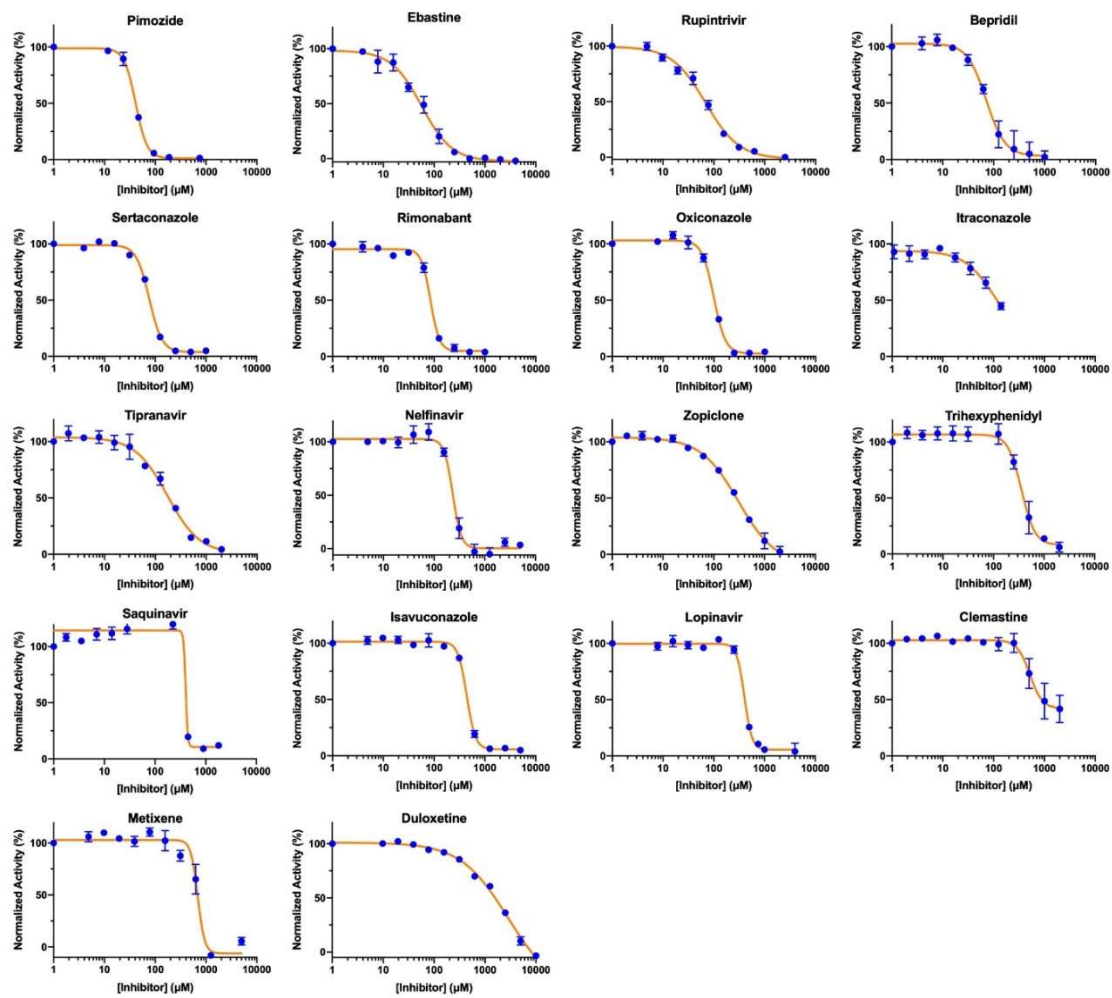


Figure 104 IC₅₀ curves of 18 medicines and rupintrivir

Table 13 IC₅₀ values and Hill slopes of 18 medicines and rupintrivir

Name	IC ₅₀ (μM)	Hill Slope
Pimozide	42 ± 2	3.1 ± 0.4
Ebastine	57 ± 12	1.5 ± 0.2
Rupintrivir	68 ± 7	1.4 ± 0.2
Bepidil	72 ± 3	2.9 ± 1.0
Sertaconazole	76 ± 2	3.5 ± 0.2
Rimonabant	85 ± 3	5.0 ± 0.4
Oxiconazole	99 ± 6	3.8 ± 0.4
Itraconazole	111 ± 35	1.6 ± 0.2
Tipranavir	180 ± 20	1.4 ± 0.2
Nelfinavir	234 ± 15	5.4 ± 1.0
Zopiclone	349 ± 77	1.2 ± 0.2
Trihexyphenidyl	370 ± 53	8.9 ± 6.4
Saquinavir	411 ± 6	26.8 ± 2.6
Isavuconazole	438 ± 11	5.2 ± 0.7
Lopinavir	486 ± 2	29.9 ± 2.4
Clemastine	497 ± 148	11.2 ± 7.3
Metixene	635 ± 43	8.7 ± 5
Duloxetine	3047 ± 634	0.93 ± 0.07

Structurally the two most potent medicines pimozide and ebastine share a same diphenylmethyl moiety. A spatially similar structure moiety N-phenyl-N-benzylamine exists in bepidil. Our docking results suggested a same binding mode for this similar structure moiety in all three drugs (Figure 105). The two aromatic rings occupy the enzyme pockets that associate with the P2 and P4 residues in a substrate.

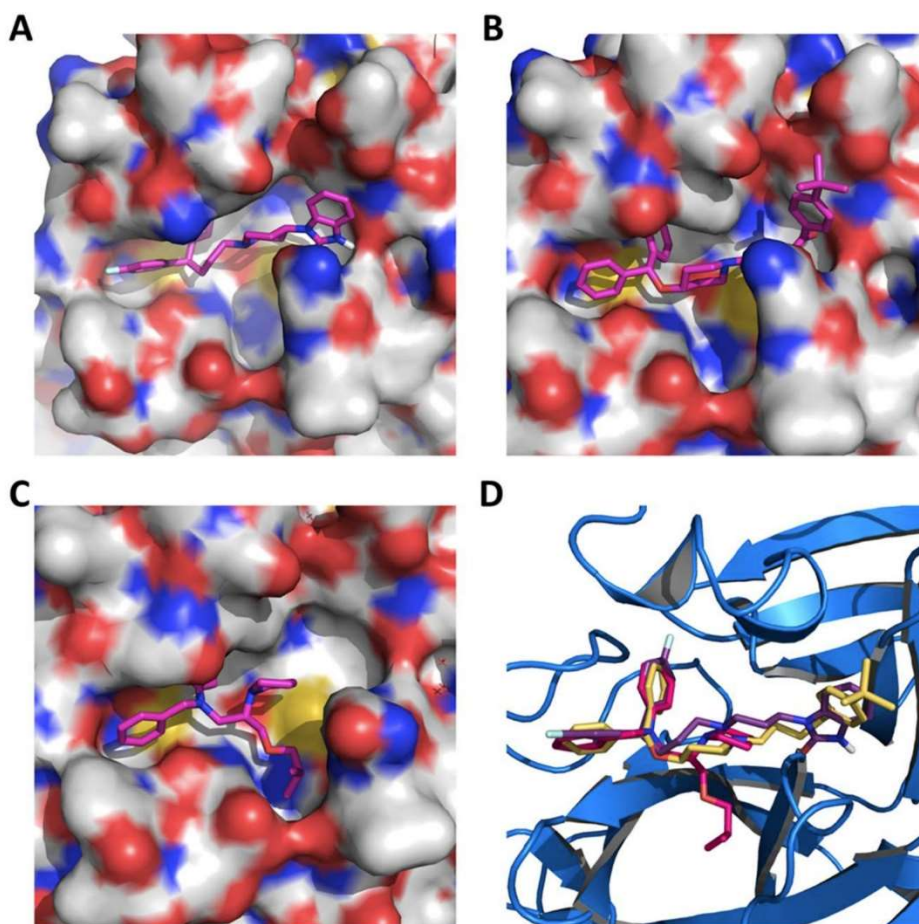


Figure 105 Pimozide (A), ebastine (B), bepridil (C), and their overlay (D) in the active site of M^{pro} . The protein surface topography in A, B, and C is presented to show the concaved active site

This observation is in line with a crystallographic study that showed two aromatic rings with a single methylene linker bound to the active site of the SARS-CoV-1 enzyme.^[172] We believe that the inclusion of the diphenylmethyl moiety in structure-activity relationship studies of M^{pro} -targeting ligands will likely contribute to the identification of both potent and high cell-permeable M^{pro} inhibitors. Figure 104 also revealed large variations in Hill coefficients of IC_{50} curves for different small molecule medicines (IC_{50} values and Hill coefficients are summarized in Table 13). Duloxetine and zopiclone gave the two highest Hill coefficients with a gradual M^{pro} activity

decrease over an increasing inhibitor concentration. On the contrary, saquinavir and lopinavir yielded lowest Hill coefficients with highly steep IC_{50} curves. There are three possible explanations for the large discrepancy in Hill coefficients. It could be attributed to different solubility of tested compounds. It is possible that a high DMSO percentage and a relatively high inhibitor concentration created phase transition for some inhibitors.^[173] A high Hill coefficient may also be due to different ligand-to-enzyme ratios when tested compounds bind to M^{pro} . An additionally possible reason is the co-existence of the M^{pro} dimer and monomer in the assay conditions. A previous report showed a K_d value of the M^{pro} dimerization as $2.5 \mu M$.^[79] In theory, the catalytically active dimer species was present at a very low concentration in our assay conditions, leaving the catalytically inactive monomer species as the major form of M^{pro} . In this situation, the inhibitors that preferentially bind to the M^{pro} dimer and the inhibitors that have a higher affinity to the M^{pro} monomer might yield different Hill coefficients. Although there is no report on the physiological concentration of M^{pro} in infected cells, it is unlikely that it can reach $1 \mu M$ (34 ppm). Even at $1 \mu M$, the majority of M^{pro} is in its inactive monomeric form. Therefore, we believe that our assay conditions mimic physiological states of M^{pro} .

4.3.3 Live Virus-based Microneutralization Assay Analysis of Pimozide, Ebastine and Bepridil

The endocytic pathway has been proposed as a key step for the SARS-CoV-2 entry into host cells.^[174] Strategies that disrupt this process make attractive therapeutic candidates for COVID-19. Based on this concept, hydroxychloroquine that has an ability to raise

endosomal pH ^{[175], [176]} has been clinically investigated for treating COVID-19 albeit with close to no effects.^[177] From chemistry point of view, our three lead compounds pimozone, ebastine, and bepridil share a similarity. They are all basic small molecules that can potentially raise endosomal pH.^[178] Among the three drugs, bepridil can be very interesting because it previously provided 100% protection from Ebola virus infections in mice at a dose of 12 mg/kg.^[179] Bepridil is a calcium channel blocker with a significant anti-anginal activity. For patients with chronic stable angina, recommended daily dose of bepridil is 200-400 mg.^[180] Mice administered with a bepridil dose as high as 300 mg/kg/day did not show alteration in mating behavior and reproductive performance, indicating that bepridil has very low toxicity.^[181] Moreover, a previous study showed that bepridil can increase the pH of acidic endosomes.^[182] Administration of a high dose of bepridil may have dual functions to slow down the virus replication in host cells by both inhibiting M^{Pro} and raising the pH of endosomes. To demonstrate this prospect, we conducted a live virus-based microneutralization assay to evaluate efficacy of pimozone, ebastine and bepridil in their inhibition of SARS-CoV-2 infection in a kidney epithelial cell line isolated from African green monkey (Vero E6) and human A549 cells stably transduced with human ACE2 viral receptor (A549/ACE2). Three medicines were tested in a concentration range from 0.78 to 200 μ M by our collaborator Dr. Tseng Group in University of Texas Medical Branch (UTMB). CPE was clearly observable for pimozone and ebastine at all tested concentrations. However, bepridil prevented completely the SARS-CoV-2-induced CPE in Vero E6 and A549/ACE2 cells when its concentration reached 6.25 μ M (Table 14. For microscopy images of CPE see

Figures 106, 107) with no signs of cytotoxicity observed under the microscopy. A separate cell toxicity assay showed that bepridil was not toxic to Vero E6 and A549/ACE2 cells until its concentration reached above 25 and 50 μM respectively (Figure 108).

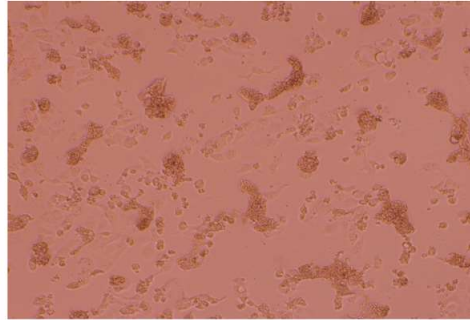
Table 14 SARS-CoV-2 induced CPE in (A) Vero E6 and (B) A549/ACE2 cells in the presence of bepridil

A. Vero E6 cells

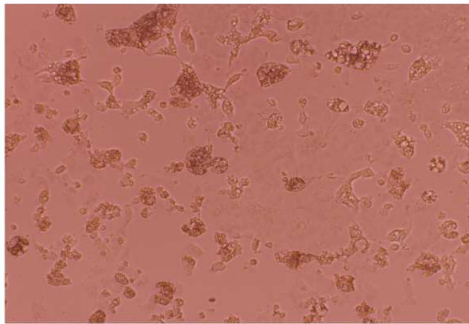
Bepridil (μM)	Repeat #1	Repeat #2	Repeat #3	Repeat #4	Repeat #5	Repeat #6
25	No CPE	No CPE	No CPE	No CPE	No CPE	No CPE
12.5	No CPE	No CPE	No CPE	No CPE	No CPE	No CPE
6.25	No CPE	No CPE	No CPE	No CPE	No CPE	No CPE
3.125	CPE	CPE	CPE	No CPE	CPE	CPE
1.56	CPE	CPE	CPE	CPE	CPE	CPE
0.78	CPE	CPE	CPE	CPE	CPE	CPE

B. A549/ACE2 cells

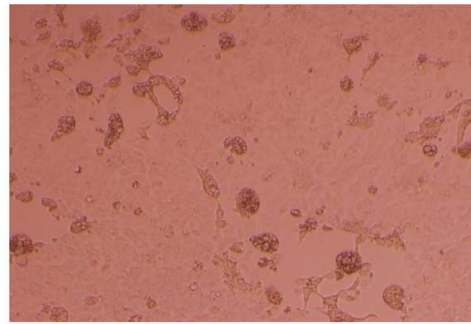
Bepridil (μM)	Repeat #1	Repeat #2	Repeat #3	Repeat #4	Repeat #5	Repeat #6
50	No CPE	No CPE	No CPE	No CPE	No CPE	No CPE
25	No CPE	No CPE	No CPE	No CPE	No CPE	No CPE
12.5	No CPE	No CPE	No CPE	No CPE	No CPE	No CPE
6.25	No CPE	No CPE	No CPE	No CPE	No CPE	No CPE
3.125	CPE	CPE	CPE	CPE	CPE	CPE
1.56	CPE	CPE	CPE	CPE	CPE	CPE
0.78	CPE	CPE	CPE	CPE	CPE	CPE



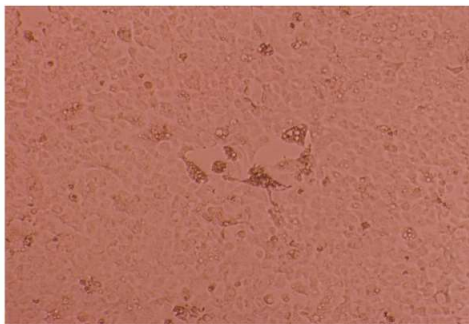
DMSO, Strong CPE



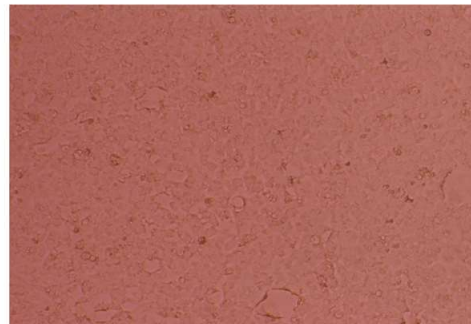
0.78 μM Bepridil, Strong CPE



1.56 μM Bepridil, Strong CPE

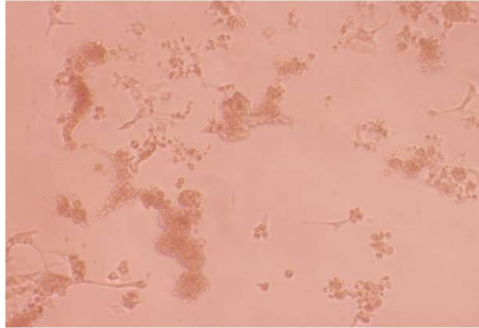


3.12 μM Bepridil, Single CPE

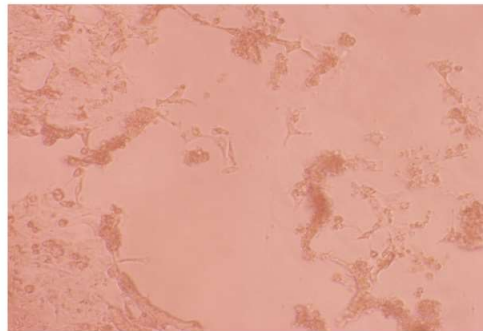


6.25 μM Bepridil, No CPE

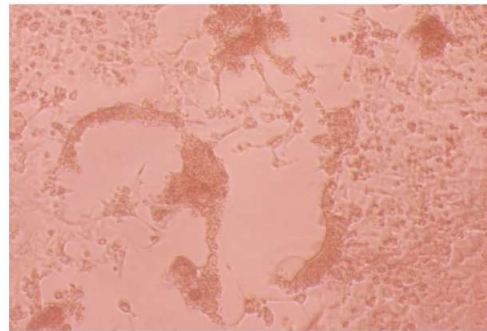
Figure 106 Microscope-recorded cytopathogenic effect (CPE) observation in Vero E6 cells that were infected by SARS-CoV-2 and grown in the presence of different concentration of bepridil or 0.1% DMSO as a positive control.



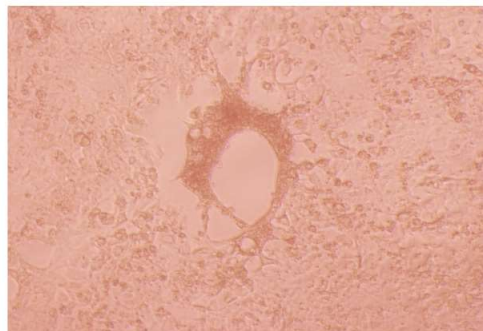
DMSO, Strong CPE



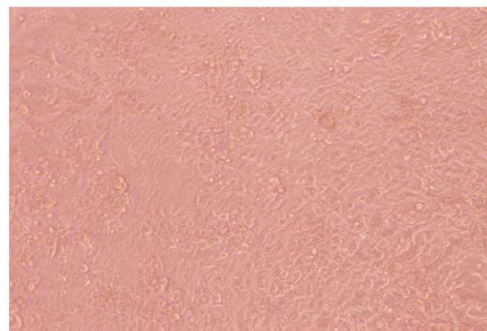
0.78 μM Bepridil, Strong CPE



1.56 μM Bepridil, Strong CPE



3.12 μM Bepridil, Single CPE



6.25 μM Bepridil, Strong CPE

Figure 107 Microscope-recorded cytopathogenic effect (CPE) observation in A549/ACE2 cells that were infected by SARS-CoV-2 and grown in the presence of different concentration of bepridil or 0.1% DMSO as a positive control

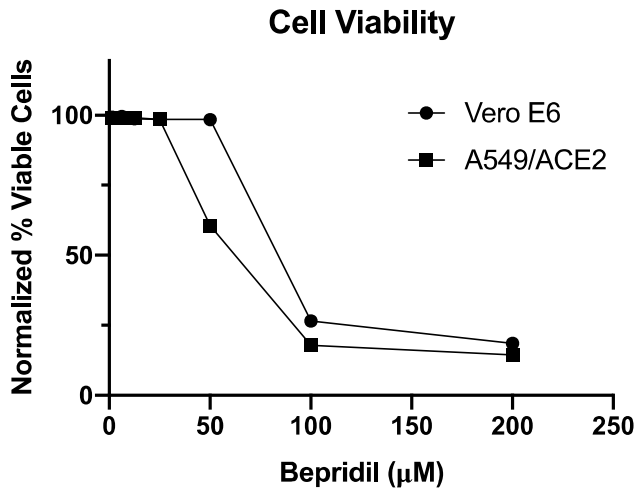


Figure 108 Viability of Vero E6 and A549/ACE2 cells at different concentrations of bepridil.

In order to characterize bepridil EC_{50} values in inhibiting SARS-CoV-2 in the two cell lines, both cells were treated with bepridil at different concentrations and infected with SARS-CoV-2 at an MOI of 0.5. Infected cells were then cultured for three days for Vero E6 and four days for A549/ACE2 cells before assessing the yields of infectious progeny virus. Based on the efficacy to inhibit SARS-CoV-2 infection, EC_{50} values were estimated to be 0.86 and 0.46 μM in Vero E6 (Figure 109A) and A549/ACE2 cells (Figure 109B), respectively.

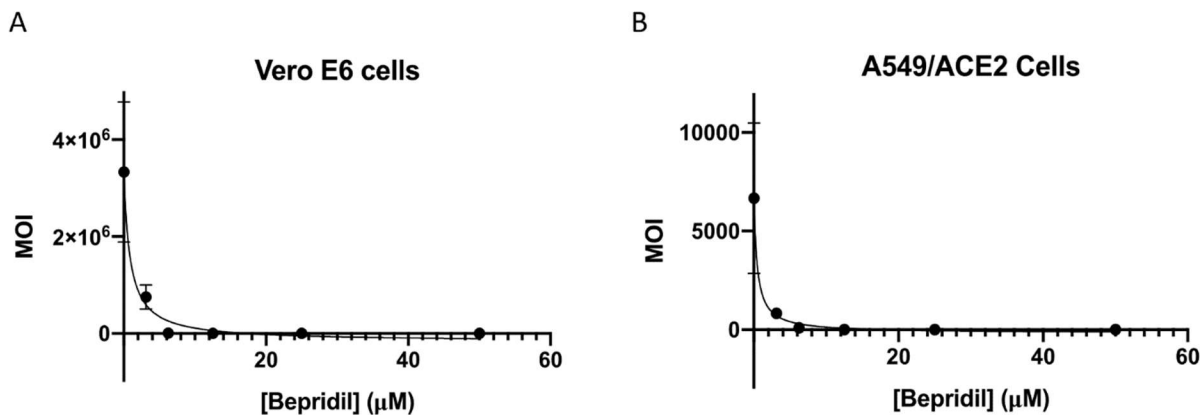


Figure 109 The SARS-CoV-2 inhibition by bepridil in (A) Vero E6 and (B) A549/ACE2 cells. Cells were incubated with different concentrations of bepridil and then infected with 0.5 MOI (multiplicity of infection) of SARS-CoV-2.

4.3.4 Bepridil as a Potential Treatment Option

These values are similar to Remdesivir's reported EC_{50} value of $0.77 \mu\text{M}$ in Vero E6 cells.^[183] The strong inhibition of SARS-CoV-2-induced CPE in Vero E6 and A549 cells by bepridil at a concentration much lower than its IC_{50} value for inhibiting M^{pro} is likely due to the aforementioned dual functions or other cellular effects of bepridil. In patients, bepridil can reach a state C_{max} as $3.72 \mu\text{M}$.^[184] This concentration is effective in inhibiting SARS-CoV-2 based on our virus microneutralization analysis. Collectively, our results indicate that bepridil is an effective medicine in preventing SARS-CoV-2 from entry and replication in mammalian cell hosts. Therefore, we urge the consideration of clinical tests of bepridil in the treatment of COVID-19.

Bepridil was voluntarily withdrawn from US market in 2004 due to its side effects such as QT prolongation.^[185] It has also been implicated that its usage might cause ventricular arrhythmia.^[186] Despite those potential drawbacks, bepridil is still marketed in Japan, China, and France. Several factors such as its in vivo efficacy against

SARS-CoV-2 and cardiovascular effects of COVID-19 need to be assessed before attempting to explore clinical benefits of using it in COVID-19 patients.

4.4 Conclusion

Guided by a computational docking analysis, we experimentally characterized about 30 FDA/EMA-approved drugs on their inhibition of the essential M^{Pro} enzyme of the COVID-19 pathogen SARS-CoV-2. From the study, we identified six FDA/EMA-approved drugs that can potentially inhibit M^{Pro} with an IC₅₀ value lower than 100 μM. One medicine bepridil exhibited strong inhibition of SARS-CoV-2 from entry and replication inside Vero E6 and A549 cells at a low micromolar concentration. Although the IC₅₀ value for Bepridil (72 μM) is relatively high, the better indicative of the potency of an antiviral drug candidate is its potency against live virus. Given that bepridil has been previously demonstrated to show efficacy in Ebola infected mice, we urge a serious consideration of its clinical tests in treating COVID-19.

CHAPTER V*4

COVALENT INHIBITION & ENGINEERING OF THE MAIN PROTEASE

5.1 Introduction

5.1.1 Aldehyde-based Inhibitors of the Main Protease

SARS-CoV-2 M^{Pro} inhibitors having aldehyde as warhead are shown to be effective *in vitro* inhibition analysis and live-virus microneutralization assays. The crystal structure of the 11a compound shows aldehyde undergoes nucleophilic attack by the active site cysteine (Cys 145) and the resulting oxyanion is stabilized by the positively charged His41 residue.^[106] A crystallography study done with SARS-CoV M^{Pro} showed electron density of the aldehyde carbon fits in to an sp² hybridized aldehyde. In the same study Lineweaver–Burk plot analysis suggested aldehyde inhibitor acting as a reversible covalent inhibitor.^[78]

In addition to their promising *in vitro* efficacy, several aldehyde-based inhibitors were shown to be effective for eradicating virus replication by having sub-micromolar EC₅₀ values.^[187] Stability tests showed 83% of MPI8 (Figure 113A) was retained after 120 hours in human plasma.^[188] Three aldehyde-based M^{Pro} inhibitors entered preclinical

⁴ COVALENT INHIBITION & ENGINEERING OF THE MAIN PROTEASE is readapted from a published report in ChemMedChem. DOI number: 10.1002/cmdc.202000924

trials (one of them is GC376). GC376 has also been tested in clinical trials and showed promising results as a broad spectrum coronavirus M^{pro} inhibitor.^[78]

5.1.2 Challenges in Biochemical Characterization of Main Protease

Despite being widely studied, ambiguity exists about the kinetic parameters of M^{pro} and its substrates which is discussed in Chapter 1.3. In the past year, several new studies were released studying inhibition of the SARS-CoV-2 variant of M^{pro}. Some studies used a commercially available FRET-based peptide substrate. (Table 1, entries 2-4)

However, a need for measuring M^{pro} activity at lower M^{pro} concentrations. Reported IC₅₀ values for some β-(S-2-oxopyrrolidin-3-yl)-alaninal (opal)-based aldehyde inhibitors are below 10 nM.^[189] However, currently the M^{pro} substrate displaying the fastest kinetics does not show a detectable signal at sub-nanomolar M^{pro} concentrations. Therefore, to be able to develop and characterize more potent inhibitors, more active substrates are needed. Another way of approaching this problem is to increase the activity of the M^{pro} by bringing minimal structural perturbation.

5.2 Experimental

5.2.1 Expression & Purification of M^{Pro}

Due to the low yields with the previous pBAD-sfGFP-TEV cleavage site-M^{Pro} construct, we generated a new construct: pET28a-His-SUMO-M^{Pro}. The pET28a-His-SUMO-M^{Pro} construct was transformed into *E. coli* strain BL21(DE3). Cells were cultured at 37°C in 6 L 2xYT medium with kanamycin (50 µg/mL) for 3 h and induced with isopropylb-D-1-thiogalactoside (IPTG) at final concentration of 1 mM when the OD600 reached 0.8. After 3 h, cells were harvested by centrifugation at 12,000 rpm, 4°C for 30 min. Cell pellets were resuspended in 150 mL buffer A (20 mM Tris, 100 mM NaCl, 10 mM imidazole, pH 8.0) and then lysed by sonication on ice. The lysate was clarified by centrifugation at 16,000 rpm, 4°C for 30 min. The supernatant was loaded onto a nickel-chelating column with High Affinity Ni-Charged Resin (GenScript) and washed with 10 column volumes of buffer A to remove unspecific binding proteins, followed by elution using buffer B (20 mM Tris, 100 mM NaCl, 250 mM imidazole, pH 8.0). The protein eluates were subjected to buffer exchange with buffer C (20 mM Tris, 10 mM NaCl, 1mM dithiothreitol (DTT), pH 8.0) by using HiPrep 26/10 desalting column (GE Healthcare). The His-SUMO-M^{Pro} proteins were digested with SUMO protease overnight at 4°C. The digested protein was applied to nickel-chelating column again to remove the His-tagged SUMO protease, the His-SUMO tag, and protein with uncleaved His-SUMO tag. The tag-free M^{Pro} protein was loaded onto an anion-exchange column with Q Sepharose, Fast Flow (GE Healthcare) equilibrated with buffer C for further purification. The column was eluted by buffer D (20 mM Tris, 1 M NaCl, 1 mM DTT,

pH 8.0) with a linear gradient ranging from 0 to 500 mM NaCl (10 column volumes buffer). Fractions eluted from the anion-exchange column were condensed and loaded to size exclusion column with HiPrep 16/60 Sephacryl S-100 HR (GE Healthcare) pre-equilibrated with buffer E (20 mM Tris, 100 mM NaCl, 1 mM DTT, 1 mM EDTA, pH 7.8). The eluted M^{pro} protein in buffer E was concentrated to 20 mg/mL and stored in -80°C for further use.

5.2.2 Half Maximal Inhibitory Concentration (IC₅₀) Assays of MPI1-9, GC376, 11a

The assays were carried out with 20 nM enzyme (except for MPI3, for which 10 nM enzyme was used) and 10 μM substrate at 37°C with continuous shaking. All the analyses were carried out in triplicate. The substrate (DABCYL-Lys-Thr-Ser-Ala-Val-Leu-Gln-Ser-Gly-Phe-Arg-Lys-MetGlu-EDANS) was purchased from Bachem and stored as 1 mM solution in 100% DMSO. Enzyme activity was monitored by fluorescence with excitation at 336 nm and emission at 455 nm wavelength. The dilution buffer (used for enzyme and substrate dilution) is 10 mM Na₂HPO₄, 10 mM NaCl, 0.5 mM EDTA, pH 7.6. Final composition of the assay buffer is 10 mM Na₂HPO₄, 10 mM NaCl, 0.5 mM EDTA, 2 μM DTT (coming from enzyme stock solution), pH 7.6 with 1.25% DMSO. All the inhibitors were stored as 10 mM in 100% DMSO solutions in -20°C freezer. For IC₅₀ analysis, the inhibitor was diluted to 400-fold times higher than the highest working concentration to make the secondary stock solution (i.e. if the highest working concentration of inhibitor is 2 μM, then the inhibitor was diluted from its 10 mM stock solution to 800 μM in DMSO). 10 μL from this secondary stock solution was added to the 990 μL of dilution buffer. Serial dilutions were carried out in dilution buffer

containing 1% DMSO to ensure all the inhibitor serial dilutions are at 1% DMSO. 25 μ L of each inhibitor solution were added to 96-well plate with multichannel pipettor. Next, 25 μ L of 80 nM enzyme solution (diluted from 10 μ M enzyme storage solution in 10 mM Na₂HPO₄, 10 mM NaCl, 0.5 mM EDTA, pH 7.6, 1 mM DTT with dilution buffer) were added by multichannel pipettor and mixed by pipetting up and down three times. Then, the enzyme-inhibitor solution was incubated at 37 °C for 30 minutes. During incubation period, 20 μ M of the substrate solution is prepared by diluting from 1 mM stock solution with dilution buffer. When the incubation period is over, 50 μ L of the 20 μ M substrate solution added to each well by multichannel pipettor and the assay started. Data recording were stopped after 30 minutes. Data treatment were done with Graph Pad Prism 8.0 software. First 0-300 seconds were analyzed by linear regression for initial slope analyses. Then, the initial slopes were normalized and IC₅₀ values were determined by inhibitor vs response - Variable slope (four parameters).

Synthesis of Inhibitors MPI1-MPI9 and 11a

Synthesis protocols and characterization data are published. ^[189]

M^{pro} Crystal Structure

Crystallization conditions and data collection parameters for M^{pro} apo crystal structure and M^{pro} crystal structure with inhibitors MPI1-MPI9 are published. ^[189]

5.2.3 Introducing Non-canonical Amino Acids to M^{Pro}

pET28a-His-SUMO-97TAG-M^{Pro} and pET28a-His-SUMO-277TAG-M^{Pro} constructions were generated. In total, three trials were made to test various non-canonical amino acid incorporation. In all three trials, 2xYT growth medium is used and cells were induced at OD 0.5 with 0.2% L-Arabinose, 1 mM IPTG and supplemented with varying concentrations of non-canonical amino acids. Expressions carried out at 37°C for 12-14 hours. Then, the cells were pelleted at 4000 rpm for 10 mins, washed with PBS and pelleted again. Then, cells were lysed by boiling with SDS-PAGE loading buffer for 20 minutes and whole-cell lysates were analyzed to test the incorporation.

For the first trial, BL21 (DE3) cells were co-transformed with following pairs of plasmids: pET28a-His-SUMO-97TAG-M^{Pro} and pEVOL AcKRS, pEVOL wt-cma PylRS, Y206F-cma PylRS. 5 mM Ack, 4 mM KetoK 2 mM PrK 4 mM BocK working concentrations were used to test incorporation. For the second trial, pET28a-His-SUMO-97TAG-M^{Pro} and pEVOL wt-cma PylRS, pEVOL *p*-AzfRS. 4 mM *p*-Azf, 2 mM AllocK 4 mM PrK 2 mM BocK working concentrations were used to test incorporation. For the third trial, pET28a-His-SUMO-277TAG-M^{Pro} and pEVOL wt-cma PylRS, pEVOL *p*-AzfRS. 4 mM *p*-Azf, 2 mM AllocK 4 mM PrK 2 mM BocK working concentrations were used to test incorporation.

Expression & Purification of K97*p*-Azf-M^{Pro}

BL21 (DE3) cells were co-transformed with pEVOL *p*-AzfFRS and pET28a-His-SUMO-K97*p*-Azf-M^{Pro}. 60 mL overnight culture in 2xYT were inoculated in to 6 L

2xYT medium with chloramphenicol (34 mg/L) and kanamycin (50 mg/L) and cells were allowed to grow at 37°C. When OD reached to 0.6, protein expression was induced by addition of 0.2% L-arabinose, 1 mM IPTG, 2 mM *p*-Azf. The expression carried out at 37°C for 6 hours. Then, the cells were pelleted, washed with ice-cold PBS and pelleted again at 4000 rpm for 20 minutes. The pelleted cells were flash-frozen with liquid nitrogen and kept at -80°C freezer until purification. The next day, cells were resuspended in 150 mL ice-cold Buffer A (the same buffer used for purification of SUMO-M^{Pro}). The all subsequent steps were carried out in cold room and the protein solution was protected from exposure to ambient light as much as possible. Next, the cells were splitted in to two portions and lysed by sonication in 2 rounds. Each portion (75 mL) lysed with 60% power output using 1 sec on, 4 sec off, 5 min total program. Next, the lysate was spinned down at 16000 rpm for 30 minutes. Then, the supernatant is decanted. To obtain a clear solution, supernatant was filtered off with 0.45 micrometer filter. Then, the lysate was incubated with of 3 mL (bed volume) of Ni-NTA resin (GenScript). The resin was washed with 50 mL of Buffer A. Then, the protein was eluted with 30 mL of Buffer B. The eluted protein was concentrated down to 2-3 mL and incubated with SUMO protease at 4°C overnight in dark. The sumo cleavage reaction (about 2-3 mL) was diluted to 30 mL to reduce the imidazole concentration to ~25 mM. Then, the diluted solution was ran through the resin 3 times to remove all the his-tagged sumo protein. The flow-through is collected and buffer was exchanged to pH 7.4, 20 mM Tris, 10 mM NaCl. Finally, the protein was concentrated, quantified by using Pierce 660 nM assay kit and flash frozen to store at -80°C.

5.2.4 Cross-linking K97p-AzF-M^{pro} to Form Covalently Dimerized M^{pro}

Dimerization of K97p-Azf-M^{pro} with DBCO (dibenzocyclooctyne) linker

DBCO linker compound was purchased from Conju-Probe (CP-2111). 50 μ M K97p-Azf-M^{pro} in 20 mM Tris pH 7.4, 10 mM NaCl buffer reacted with 200 μ M DBCO linker compound for 10 hours at 25 °C. Then, the unreacted small molecule were removed by spin desalting column. Next, the concentration was determined by Pierce 660 nm assay. Then, equimolar K97p-Azf-M^{pro} is added and the reaction continued for another 10 h at 25°C. The conjugation was confirmed by SDS-PAGE and LC-MS.

Dimerization of K97p-Azf-M^{pro} with BCN (bicyclooctyne) linker

BCN linker compound was purchased from Conju-Probe (CP-6232). 50 μ M K97p-Azf-M^{pro} in 20 mM Tris pH7.4, 10 mM NaCl buffer reacted with 200 μ M BCN linker compound for 10 hours at 25°C. Then, the unreacted small molecule were removed by spin desalting column. Next, the concentration was determined by Pierce 660 nm assay. The conjugation was confirmed by SDS-PAGE and LC-MS.

5.2.5 Biochemical Characterization of M^{pro} Derivatives

K_M determination of M^{pro}

The assays were carried out with 20 nM enzyme and FRET-Sub3 at 37°C. An aliquot was taken out at indicated time points and diluted 10 times to stop the reaction.

Fluorescent intensity was recorded immediately. Data treatment were done with Graph Pad Prism 8.0 software. First 14 min were analyzed by linear regression for initial

reaction rate analyses. The initial reaction rates were used to determine the K_M value by fitting with Michaelis-Menten non-linear regression.

Comparing activities of substrates for M^{pro}

10 μ M of each substrate with indicated M^{pro} concentration. M^{pro} is pre-incubated in assay buffer (10 mM Na₂HPO₄, 10 mM NaCl, 0.5 mM EDTA, pH 7.6) at 37°C for 30 minutes. Then, 20 μ M of substrate solution in 10 mM Na₂HPO₄, 10 mM NaCl, 0.5 mM EDTA, pH 7.6, 2% DMSO was added on to M^{pro} solution and reactions were monitored at 37°C. Data was analyzed by GraphPad Prism 8.0. For coumarin-based substrates (6-mer-Cou-Sub, VLQ-Cou-Sub) 370 nm and 450 nm were used as excitation and emission wavelengths. For FRET-Sub1 and FRET-Sub3, 336 nm and 490 nm, for FRET-Sub2 492 nm and 517 nm were used as excitation and emission wavelengths.

Activity Comparison of wt-M^{pro}, K97p-Azf- M^{pro} and Cross-linked K97p-Azf-M^{pro}

M^{pro} derivatives indicated on the graphs were pre-incubated in assay buffer (10 mM Na₂HPO₄, 10 mM NaCl, 0.5 mM EDTA, pH 7.6) at 37°C for 30 minutes. Then, 20 μ M of FRET-Sub3 solution in 10 mM Na₂HPO₄, 10 mM NaCl, 0.5 mM EDTA, pH 7.6, 2% DMSO was added on to M^{pro} solution and reactions were monitored at 37°C.

Varying enzyme concentration with fixed substrate concentration

Various concentrations of wt-M^{pro} are tested with 10 μ M FRET-Sub3, 6mer-Cou-Sub and VLQ-Cou-Sub. 2x of the working concentration of wt-M^{pro} were pre-incubated in assay buffer (10 mM Na₂HPO₄, 10 mM NaCl, 0.5 mM EDTA, pH 7.6) at 37°C for 30

minutes. Then, 20 μ M of FRET-Sub3, 6mer-Cou-Sub and VLQ-Cou-Sub solution in 10 mM Na₂HPO₄, 10 mM NaCl, 0.5 mM EDTA, pH 7.6, 2% DMSO was added on to M^{Pro} solution and reactions were monitored at 37 °C.

5.2.6 Live virus-based microneutralization assay with MPI1-MPI9, GC376, 11a

A slightly modified cytopathic effect (CPE)-based microneutralization assay was used to evaluate the drug efficacy against SARS-CoV-2 infection. Briefly, confluent African green monkey kidney cells (Vero E6) or human alveolar epithelial A549 cells stably expressing human ACE2 viral receptor, designated A549/hACE2, grown in 96-wells microtiter plates were pretreated with serially 2-folds diluted individual drugs for two hours before infection with 100 or 500 infectious SARS-CoV-2 (US_WA-1 isolate) particles in 100 μ L EMEM supplemented with 2% FBS, respectively. Cells pre-treated with parallelly diluted DMSO with or without virus were included as positive and negative controls, respectively. After cultivation at 37 °C for 3 (Vero E6) or 4 days (A549/hACE2), individual wells were observed under the microcopy for the status of virus-induced formation of CPE. The efficacy of individual drugs was calculated and expressed as the lowest concentration capable of completely preventing virus-induced CPE in 100% (EC₁₀₀) or 50% (EC₅₀) of the wells. All compounds were dissolved in 100% DMSO as 10 mM stock solutions before subjecting to dilutions with culture media.

5.3 Results & Discussion

5.3.1 Expression & Purification of M^{Pro}

For the experiments in Chapter 4, we used our first M^{Pro} expression construct had the following open reading frame: sfGFP-TEV cleavage site-M^{Pro}-Histag. The rationale for choosing this construct is that M^{Pro} is not supposed to be active when its N-terminus is capped because, having extra residues at the N-terminus prevents dimer formation. Moreover, the TEV cleavage site is not recognized by endogenous proteases in *E. coli*. Therefore, we were expecting to obtain full-length protein after expression. However, after the expression most of the protein was in truncated form according to SDS-PAGE analysis. Molecular weights of the fragments suggest the cleavage occurred from the TEV cleavage site. This site can also be recognized by M^{Pro} too. Therefore, it is possible that M^{Pro} may cleave itself intramolecularly or another M^{Pro} may form dimer and cleave at the TEV cleavage site. Once the N-terminus of M^{Pro} is free, the His-tag at C-terminal will be processed next to yield native M^{Pro}. Because the M^{Pro} no longer has the purification tag, we applied ammonium sulfate precipitation to purify the native M^{Pro} which yielded around 1 mg protein per 1L 2xYT culture. However, this amount was not satisfactory for our crystallography studies. To overcome this limitation, Dr. Kai S. Yang made a new construct using M^{Pro} having a SUMO tag at its N-terminus and a 6xhis-tag at its C-terminus. The protein sequence SUMO tag does not contain a sequence motif that can be recognized by M^{Pro}. Therefore, the full-length protein can be purified by carrying out Ni-NTA purification followed by SUMO digestion.

5.3.2 Activity of different substrates with M^{Pro}

To test our inhibitors, Ph.D. candidate Kaci C. Kratch initially synthesized 6mer-Cou-Sub to use as substrate in inhibition assays. In this design, there is a coumarin after glutamine residue. The coumarin has a very low fluorescence in this state and becomes fluorescent when, the amide bond (marked with an arrow in Figure 110A) was cleaved and coumarin with free amine is generated. When we tested this substrate we could only get enzyme activity at high (1 μM) M^{Pro} concentrations (Figure 111A). However, we needed a more active substrate to test our inhibitors with having low-nanomolar IC₅₀ values.

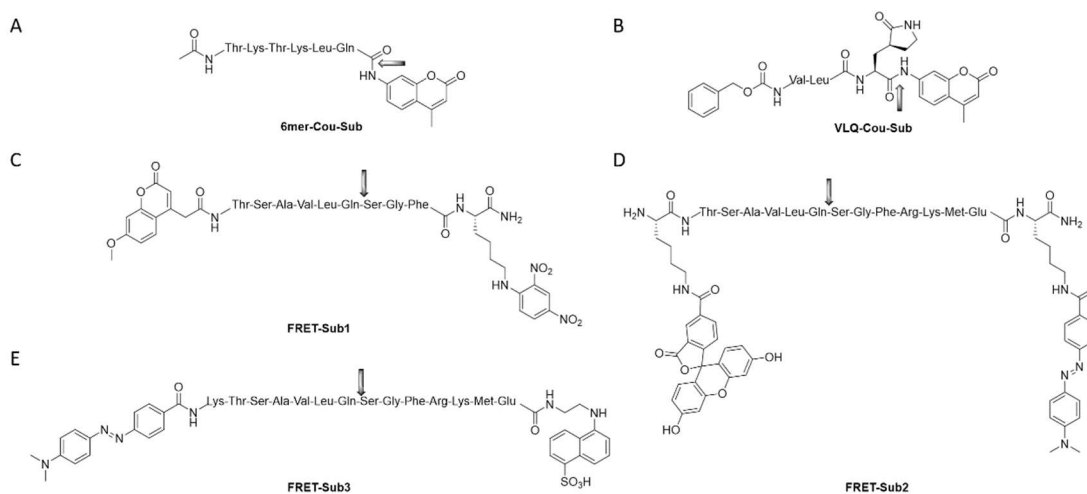


Figure 110 Structures of the substrates used for M^{Pro} activity assays **A)** 6mer-Cou-Sub **B)** VLQ-Cou-Sub **C)** FRET-Sub1 **D)** FRET-Sub2 **E)** FRET-Sub3

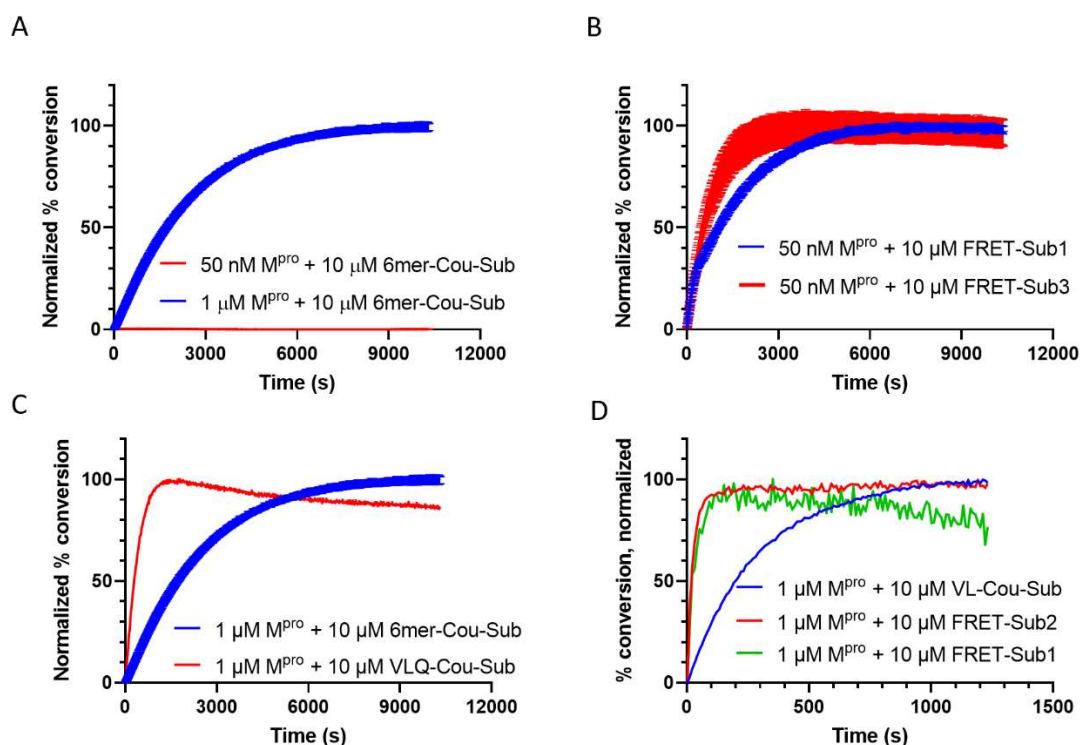


Figure 111 **A)** Activity of 6mer-Cou-Sub at 1 μ M and 50 nM M^{pro} **B)** Comparison of the activities of FRET-Sub1 and FRET-Sub3 **C)** Comparison of activities of 6mer-Cou-Sub and VLQ-Cou-Sub **D)** Comparison of activities of FRET-Sub1, FRET-Sub2, VLQ-Cou-Sub

Therefore, we turned our attention to synthesize FRET-based substrates (Figure 110 C-E) which mimic the native substrates better. In this design, a fluorophore and a quencher are positioned at the two ends of the peptide. In this state, no fluorescence signal was observed due to the proximity of fluorophore and quencher. However, once the M^{pro} cleaves glutamine-serine peptide bond, fluorophore and the quencher are separated and quencher cannot quench the fluorophore's fluorescence as effectively and fluorescence becomes observable. Kaci C. Kratch synthesized FRET-Sub1 and we were able to observe activity with 50 nM M^{pro} (Figure 111B). Then, we decided to purchase FRET-Sub3 (Figure 110E, the substrate used in Chapter 4) because, it is widely used by other

groups working on M^{pro} and allowed us to compare our inhibitors with others reported in the literature. We synthesized another substrate (FRET-Sub2) in order to have a fluorophore that is more red-shifted that would allow us to work with systems that have background fluorescence such as cell lysate. This peptide was synthesized by Ph.D. candidate Lauren R. Blankenship and uses fluorescein as the fluorophore. We aimed to study the effect of cofactors on M^{pro} activity and wanted observe M^{pro} activity in the cell lysate. The results of these experiments are not mentioned in this thesis. Among the FRET substrates synthesized and purchased, the substrate with the fastest kinetics was FRET-Sub3. FRET-Sub2 and FRET-Sub1 seemed to have a similar velocity (Figure 111D).

Finally, we wanted to see whether we could improve the 6-mer-Cou-Sub design based on the structure-activity relationship we acquired during inhibition analysis. Therefore, we designed another substrate, VLQ-Cou-Sub (Figure 110B) which is only a tripeptide and uses an β -(S-2-oxopyrrolidin-3-yl)-alaninal (opal) group instead of glutamine. Post-doc Dr. Yugendar Alugubelli synthesized this compound and which seemed to be much more reactive than the 6mer-Cou-Sub albeit smaller (Figure 111C).

K_M determination of M^{pro}

As mentioned in the introduction, significant differences exist in the reported kinetic properties of M^{pro} . Previously, K_m values of 28 μM , 170 μM reported. The value Dr. Kai S. Yang found for the same substrate and the M^{pro} pair was 422 μM (Figure 112).

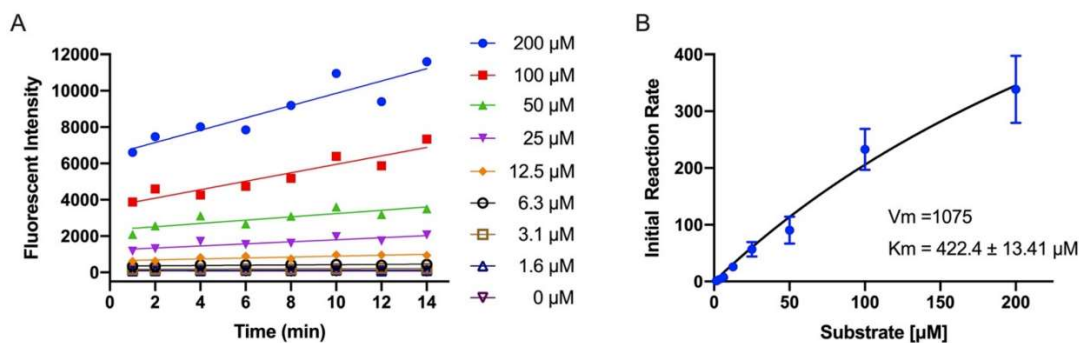


Figure 112 A) Linear regression analysis of the initial slopes B) K_M calculation for M^{Pro} FRET-Sub3 enzyme substrate pair

One drawback of using FRET-based peptides is inner filter effects.^[190] The EDANS dye can be quenched at higher concentrations. Moreover, the presence of the free quencher in the solution will further decrease the observed fluorescence. Because, K_M determination requires high substrate concentrations, fluorescence quickly reaches the saturation level where it has no longer linear relationship with fluorophore concentration. Some papers use inner-filter correction to account for the deviation caused by quenching from the quencher (in this case dabcyI).

We tried diluting the assay solution at given time points to circumvent the problem. However, the curve shows, it did not reach full saturation and the FRET-Sub3 started to precipitate beyond the highest concentration used in our K_M determination assay.

The reproducibility problems in K_M determination are one of the reasons, why we aimed to develop substrates that are more active and more soluble substrates.

5.3.3 IC₅₀ assays with MPI1-8, GC376, 11a

Dipeptidyl and tripeptidyl aldehyde-based compounds (MPI1-MPI19) and 11a were synthesized and their dose-dependent effect on M^{pro} activity was studied. GC376 and 11a are previously reported inhibitors in the literature. We purchased GC376 from BOC Sciences (B0084-007378). In our inhibitor design, we fixed aldehyde and the P1 position to be a 2-pyrrolidone analog of glutamine. The aldehyde group was chosen as the warhead group due to being proven effective as yielding potent inhibitors *in vitro* M^{pro} assays and live-virus microneutralization assays. As discussed in the introduction chapter, M^{pro} has a very strong preference for glutamine at the P1 position. Previous studies have shown the cyclic glutamine analog having 2-pyrrolidone as its side chain shown to have better binding, probably due to the reduced entropic penalty upon binding. Therefore, we fixed the P1 position as to be β -(S-2-oxopyrrolidin-3-yl)-alaninal (Opal) residue as well.

Three different residues in the P2 position were tried: leucine, phenylalanine and cyclohexylalanine. The trend obtained by MPI3, MPI4, MPI5 inhibitors indicate that P2 residues effectiveness follows the trend leucine>phenylalanine>cyclohexylalanine. This structure-activity relationship aligns with the substrate preference of M^{pro} as mentioned in Introduction chapter. Inhibitors MPI3, MPI4, MPI5, MPI6, MPI7, MPI8, MPI9 are tripeptides and contain a P3 residue. Valine residue seems to be the most effective residue in this position. This is also not surprising because, valine is the most preferred canonical amino acid at the P3 position for the M^{pro} substrates. Tert-butyl threonine residues are preferred because, they can maintain the inhibitor's potency to an acceptable

level while bringing extra stability to the tripeptide.

Another varied functional group is the N-terminus protecting group. Three different N-terminus protecting groups have been tried: the CBZ group, the *ortho*-fluoro-*para*-chlorocinnamyl (FCC) group and the indole group. MPI1 and MPI2 allows us to compare CBZ and FCC groups in a dipeptide structure whereas, MPI3 and MPI9 allows comparison of the same groups in a tripeptide structure. For MPI1 and MPI2, the K_i values are very similar (for MPI1 97.9 ± 22.5 nM, for MPI2 100.6 ± 13.7 nM). However; a significant difference existed between MPI3 (8.3 ± 1.5 nM) and MPI9 (54.7 ± 13.7 nM). The literature compound 11a was the only group with an indole protecting group and yielded a K_i value of 30.3 ± 2.9 nM. The structures of the inhibitors (Figure 113A), inhibition curves (Figure 113B) and summarized IC_{50} and K_i values (Figure 113C) are shown.

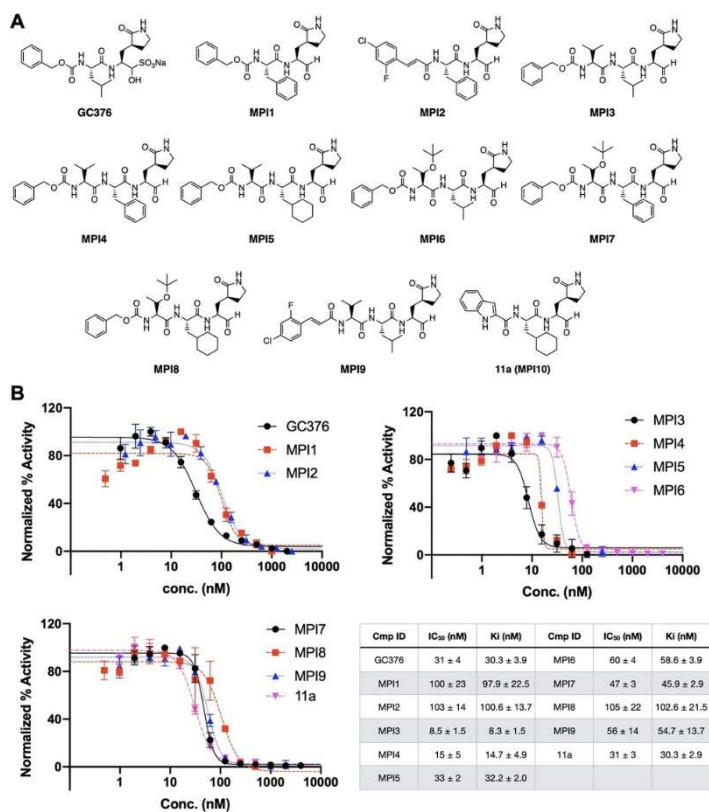


Figure 113 A) Structures of the inhibitors used in this study B) Inhibition curves of the inhibitors C) Summary of IC₅₀ and K_i values

Hill Coefficients of the IC₅₀ Curves

A striking feature of the inhibition curves is the steep Hill slopes. There are various reasons for Hill slopes greater than 1 (Table 15). The first possibility is more than one inhibitors may be binding per enzyme molecule to inactivate it. Because, the inhibitors studied, in this case, are highly potent (having IC₅₀ values below 100 nM) the inhibitor concentrations used are not high enough to cause non-specific binding. Moreover, the difference between enzyme concentration and inhibitor concentration in some cases (IC₅₀ of MPI3 with 10 nM M^{PRO} is 8.5 nM) are so close that, having multiple inhibitors on a single enzyme is unlikely.

The most likely explanation for the unusually high Hill slopes is the effect of inhibitors on the enzyme's dimerization. As previously discussed in the introduction section, M^{pro} has a high dissociation constant (K_d) of 2.5 μ M. The enzyme concentrations we used in our assays (20 nM) are well below the K_d which means M^{pro} is likely to present as a monomer in the assay solution. An interesting phenomenon is that at low concentrations of the inhibitor, the enzyme activity increases with increasing inhibitor concentration until it reaches a peak. MPI3 and MPI4 demonstrate this pattern clearly (Figure 113B). After, reaching the peak there is a sharp decrease in activity. This phenomenon was previously observed with MERS M^{pro} which has about 86% sequence similarity with SARS-CoV-2 M^{pro}. The explanation for this observation is an inhibitor's interaction with an M^{pro} molecule can induce dimerization which will increase the overall activity. From the crystal structure of ligand-bound M^{pro} and apo M^{pro} the inhibitor binding induces conformational change on M^{pro}. However, once the inhibitor increases beyond a certain threshold, the activity starts dropping drastically. A common reason for very steep inhibition curves is the aggregation of proteins.^[164] Because the inhibitor can induce a difference in the oligomerization state of MERS M^{pro} (which is observed in our system as well) it might be causing aggregation of the M^{pro} at higher concentrations.^[191]

Table 15 Hill slopes of the tested inhibitors.

Inhibitor	Hill Slope
GC376	1.8 ± 0.3
MPI1	3.0 ± 0.3
MPI2	2.3 ± 0.1
MPI3	3.8 ± 0.6
MPI4	10.0 ± 3.3
MPI5	7.7 ± 0.3
MPI6	5.2 ± 0.8
MPI7	4.8 ± 1.0
MPI8	3.8 ± 0.6
MPI9	3.8 ± 0.6
11a	2.6 ± 0.4

5.3.4 Crystal Structure Analysis of Interactions Between M^{pro} and inhibitors

The crystal structure of apo-M^{pro} showed that M^{pro} was crystallized as densely packed monomers providing a clear view of the active site (Figure 114A). Inhibitor-bound crystal structures were obtained by soaking apo-M^{pro} crystals with inhibitors. By soaking method crystal structures of MPI1, MPI3-MPI8 were obtained. Crystal structures of MPI2 and MPI9 couldn't obtained due to crystal cracking during soaking.

MPI3 being the most potent inhibitor among the inhibitors tested in this study, showed clear density at P1, P2 sites and the covalent adduct between the aldehyde carbon and Cys145 was observable (Figure 114 B-E). Inhibitor-bound crystal structures for MPI1, MPI3-MPI8 demonstrate the opal residue tightly fits in the P1 pocket of the enzyme by forming three hydrogen bonds with the side chains of E166, H163 and the backbone of F140. For the P2 position leucine fits better than other residues (phenylalanine and cyclohexylalanine) because it can form close range van der waals

interactions with the residues in P2 binding pocket. There is not a specific binding pocket for P3 position but, it is apparent that valine is the preferred residue in this position and it is in close contact with E166 and P168. As can be seen in the overlayed structures in Figure 114F, the CBZ groups do not fit tightly to the P4 binding pocket. P168 loop is positioned differently from the apo-M^{Pro} crystals to accommodate the CBZ group. This implies that, there is room for optimization of the N-terminus capping group.

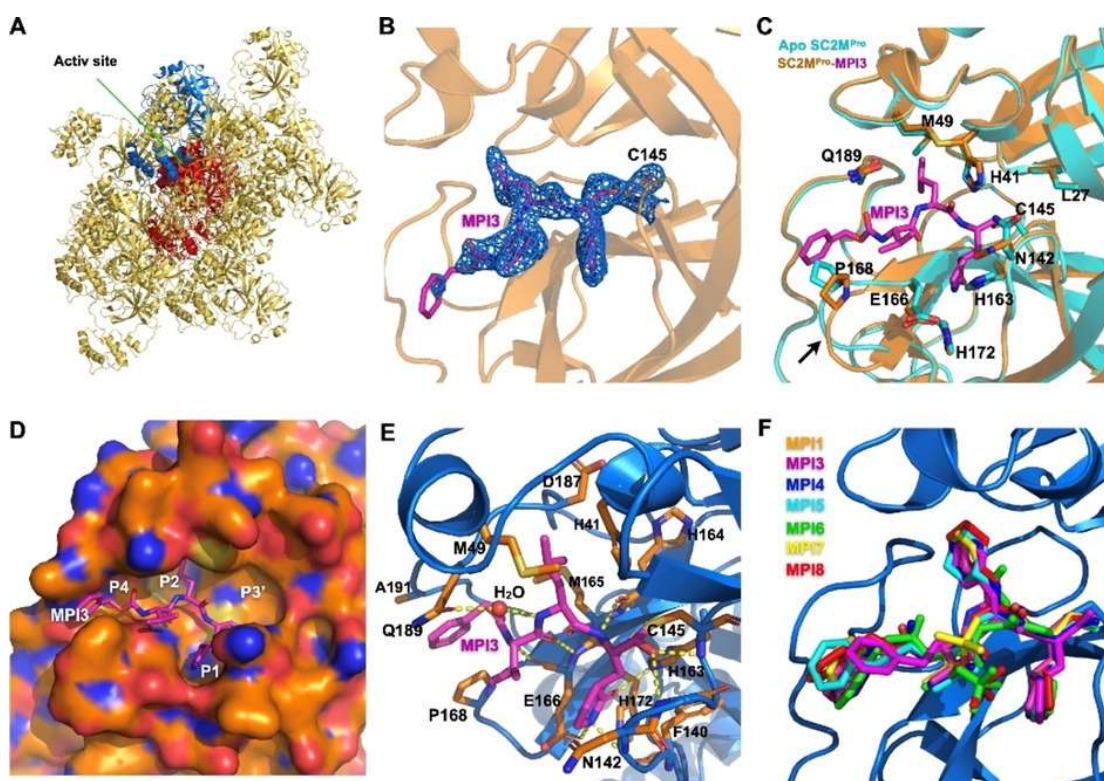


Figure 114 **A)** The packing of apo-M^{Pro} in its crystals. An asymmetric unit monomer in the center is colored in red. Its active site is presented as a concave surface. Another monomer that stacks on the active site of the monomer is colored in blue. **B)** A contoured 2F_o-F_c map at the 1σ level around MPI3 and C145 in the active site of M^{Pro}. A covalent bond between MPI3 and C145 is observable. **C)** Structure overlay between apo-M^{Pro} and the M^{Pro}-MPI3 complex. A black arrow points to a region that undergoes structure rearrangement from the apoenzyme to accommodate MPI3 in the M^{Pro}-MPI3 complex. **D)** Occupation of the active-site cavity of M^{Pro} by MPI3. The enzyme is shown in its surface presentation mode. **E)** Extensive hydrogen bonding and van der Waals interactions between M^{Pro} and MPI3. The backbone of M^{Pro} is colored in blue and side chain carbon atoms in orange. Hydrogen bonds between MPI3 and M^{Pro} are depicted as yellow dashed lines. **F)** Overlay of seven Opal-based inhibitors at the active site of M^{Pro}. Color coded for the names is shown in the figure. All images were made by using the program PyMOL. The PDB IDs for M^{Pro} in its apo form and complexes with inhibitors are 7JPY (apo), 7JPZ (MPI1), 7JQ0 (MPI3), 7JQ1 (MPI4), 7JQ2 (MPI5), 7JQ3 (MPI6), 7JQ4 (MPI7), and 7JQ5 (MPI8).

5.3.5 Live-virus Microneutralization Analysis of MPI1-9, GC376, 11a

Inhibition analysis of MPI1-MPI9 compounds yielded promising results with low-nanomolar IC₅₀ values. Next, we turned our attention to live-virus microneutralization (MN) assays. We conducted these assays in collaboration with Dr. Chien-Te K. Tseng at the University of Texas Medical Branch (UTMB). Two cell lines were used in these assays: Vero E6 cell line (African Green Monkey kidney epithelial cell line) and human alveolar epithelial A549 cells stably expressing human ACE2 viral receptor (A549/hACE2). In the assays we also included the compounds reported in the literature GC376 and 11a. We tested MPI1-MPI9 and reference compounds GC376, 11a up to 10 μM concentration in both cell lines.

The results show that MN assay trends and inhibition assays do not follow similar trends. For instance, the most potent inhibitor in inhibition assays (MPI3) did not show any activity in MN assays up to 10 μM. Ironically, the least potent inhibitor (MPI8) in inhibition assays turned out to be the most potent compound in MN assays. There are several parameters that determine *in vivo* activity of an inhibitor and potency in the inhibition analysis is only one of these parameters. Other important factors are cell membrane permeability, *in vivo* stability and side effects. To eradicate the cytopathogenic effect (CPE) of SARS-CoV-2 in the tested cell lines, it is likely that the *in vivo* stability of these peptides and side effects play a big role. Having a tert-butyl group on threonine (such as in MPI8) will increase the stability of the tripeptide by blocking the interaction with proteases that will degrade the tripeptide.^[192] Moreover, the MPI1-MPI9 inhibitors are potent against endogenous proteases such as cathepsins as

well.^[70] Cathepsins process the spike protein at the endocytosis stage and play a role in internalization. Therefore, inhibiting endogenous proteases that help internalize SARS-CoV-2 such as trans membrane serine protease (TMPRSS2), furin and cathepsins will also cause decreased CPE levels. However, they will bring off-side toxicity which needs to be carefully considered before going on to clinical trials.

Another interesting feature is the difference between the two cell lines. For instance, MPI7 is as potent as MPI8 against Vero E6 cell line. However, MPI7 do not show any potency against A549/hACE2 whereas, MPI8 can fully eradicate CPE at concentrations as low as 0.3 μ M. A similar discrepancy between two cell lines were also observed in their response against Remdesivir. Remdesivir was about 10-fold more potent against eradicating CPE in A549/hACE2 cells than in Vero E6 cells. However, same study found that, the two cell lines showed similar responses against another well-known anti-SARS-CoV-2 agent chloroquine.^[193] This observation reveals once again that in drug discovery process, one of the grand challenges is working with the right model organism or cell line.

5.3.6 Non-canonical Amino Acid Incorporation into M^{pro}

Our M^{pro} crystal structure suggests that residues 97 and 277 are in proximity in the adjacent monomer units. Therefore, we reasoned that if we can incorporate a non-canonical amino acid at this position with a reactive handle, by using a bifunctional small molecule linker, we can covalently link two M^{pro} enzymes. Residues at 97 and 277 positions are lysine and asparagine respectively. We reasoned that a non-canonical amino acid with a flexible side chain would create less structural perturbation.

Therefore, we tried KetoK first. Our group incorporated KetoK successfully in to green fluorescent protein (GFP)^[194] and phage (unpublished data). To our surprise, 97 and 277 positions did not show detectable incorporation of KetoK and we observed the overexpression of the truncated protein only. We then move on to test other non-canonical amino acids. We tested acetyl-l-lysine (AcK), boc-l-lysine (BocK), propargyl-l-lysine (PrK), alloc-l-lysine (AllocK) and para-azido-L-phenylalanine (*p*-AzF) (structures shown in Figure 115D). Among the non-canonical amino acids tested, only BocK, AllocK and *p*-AzF showed a detectable amount of incorporation (Figure 115A). Tests with GFP as reporter protein, showed that propargyl lysine has higher incorporation rates than boc-l-lysine and alloc-l-lysine. Having these three non-canonical amino acids behaving that differently with M^{Pro} has no straightforward explanation. 97 residue was at the loop region next to a proline residue. The mutations at the loop regions can affect protein stability profoundly and are linked with diseases.^[195] However, we did not do further study to investigate this phenomenon, because it is not the focus of this study. Then, we compared the incorporation of *p*-AzF, BocK, AllocK into His-SUMO-97TAG-M^{Pro} (Figure 115B) and His-SUMO-277TAG-M^{Pro} (Figure 115C). Incorporation into His-SUMO-97TAG-M^{Pro} yielded a higher expression level than His-SUMO-277TAG-M^{Pro}. This result was an expected outcome because, TAG mutations closer to N-terminus generally yield higher expression levels. Then, we decided to use *p*-AzF as our reactive non-canonical amino acid. Expression and purification with *p*-AzF gave a yield of 2 mg/L of K97*p*-Azf-M^{Pro} (Figure 115E).

reaction with DBCO linker and BCN linker went smoothly. Dimerized M^{PRO} were prepared by two-step conjugation with DBCO linker (Figure 116B) and by one-step conjugation with BCN linker (Figure 116C). When $K97p\text{-Azf-}M^{\text{PRO}}$ (mass spec before reaction shown in Figure 117A) was reacted with excess DBCO linker, $\text{DBCO-}(M^{\text{PRO}})_2$ (dimerized M^{PRO}) is formed only as minor product (Figure 117B-C). Most of the $K97p\text{-Azf-}M^{\text{PRO}}$ was reacted with DBCO and did not react further with another $K97p\text{-Azf-}M^{\text{PRO}}$. Therefore, after the first step, unreacted DBCO is removed, one equivalent of $K97p\text{-Azf-}M^{\text{PRO}}$ was added to the reaction mixture to generate the desired $\text{DBCO-}(M^{\text{PRO}})_2$ as the major product. Conversely, with the BCN linker, after first step the major product was the dimerized $K97p\text{-Azf-}M^{\text{PRO}}$, $\text{BCN-}(M^{\text{PRO}})_2$ (Figure 117D).

A possible explanation for this result is the size difference between the two linkers. DBCO linker is much bulkier and lengthier than BCN linker. Therefore, after the initial $K97p\text{-Azf-}M^{\text{PRO}}$ reaction, dimer formation between $\text{DBCO-}K97p\text{-Azf-}M^{\text{PRO}}$ conjugate and $K97p\text{-Azf-}M^{\text{PRO}}$ is probably less favored. Due to the smaller size and shorter arm length this effect is probably not as dramatic for the BCN linker. Therefore, the BCN linker can generate $\text{BCN-}(M^{\text{PRO}})_2$ as the major product. Quantification based on SDS-PAGE gel suggested above 70% conversion (Figure 117 E-F). Further purification with the anion exchange column and size-exclusion column did not yield a good separation of monomer and dimer species.

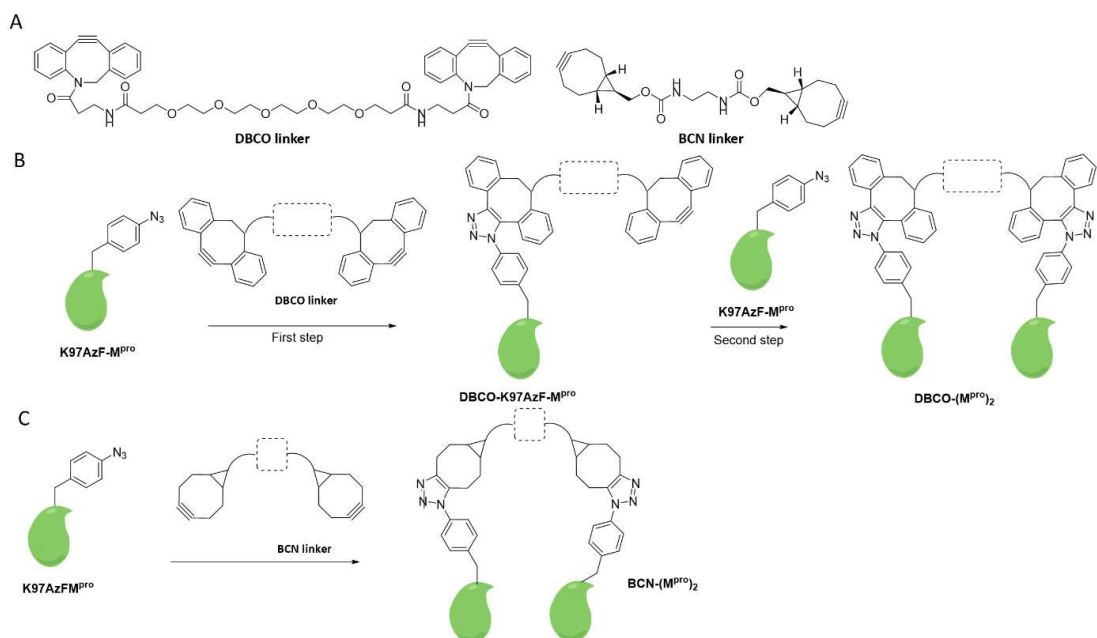


Figure 116 **A)** Structures of DBCO linker and BCN linker **B)** Dimerization of K97*p*-Azf-M^{PRO} with DBCO linker to generate DBCO-(M^{PRO})₂ **C)** Dimerization of K97*p*-Azf-M^{PRO} with BCN linker BCN-(M^{PRO})₂

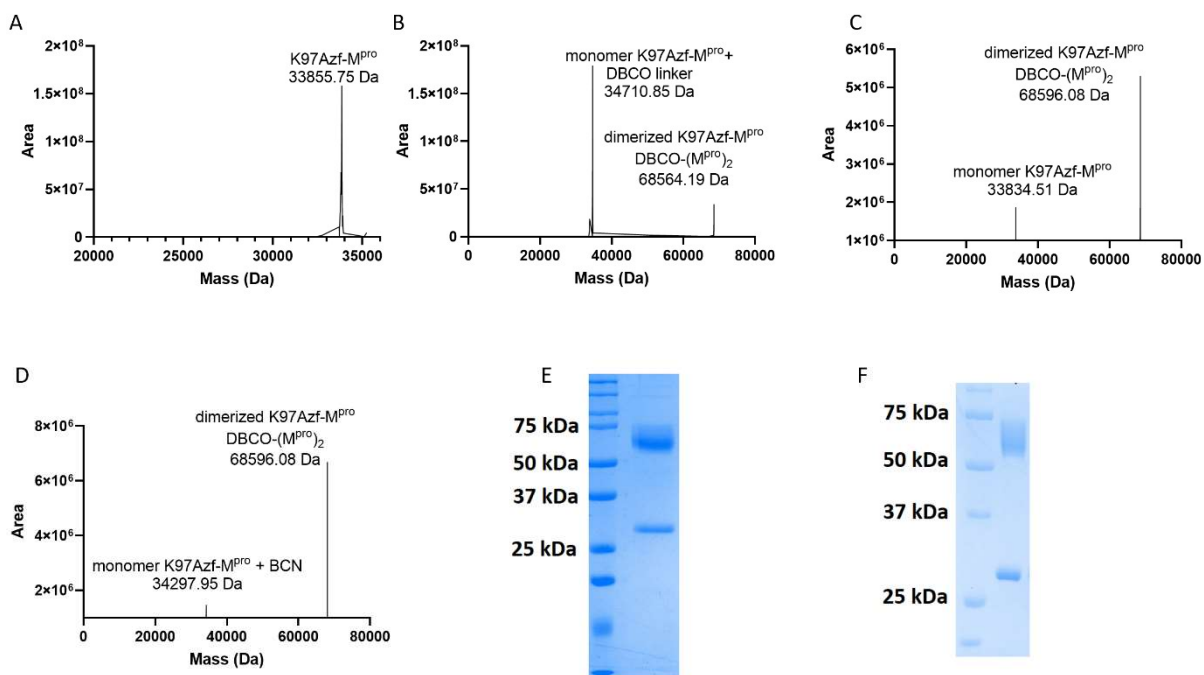


Figure 117 **A)** ESI-MS analysis of K97*p*-Azf-M^{PRO} **B)** ESI-MS analysis of DBCO linker reaction after first step **C)** ESI-MS analysis of DBCO linker reaction after second step **D)** ESI-MS analysis of BCN linker reaction **E)** SDS-PAGE analysis of DBCO linker reaction after second step **F)** SDS-PAGE analysis of BCN linker reaction

Activity Comparison of wt-M^{Pro}, K97p-Azf- M^{Pro} and BCN-(M^{Pro})₂ and DBCO-(M^{Pro})₂

Next, we compared the activities of wt-M^{Pro}, K97p-Azf- M^{Pro}, BCN-(M^{Pro})₂ and DBCO-(M^{Pro})₂ at 50 nM concentration. Figure 118A shows that K97p-Azf- M^{Pro} shows slightly higher activity than wt-M^{Pro}. 25 nM of DBCO-(M^{Pro})₂ was used in the same assay to have a total of 50 nM M^{Pro} units in the assay solution. DBCO-(M^{Pro})₂ was much more active and the reaction curve reached a plateau after ~15 minutes (Figure 118A).

Next, DBCO-conjugated and BCN-(M^{Pro})₂ were compared at 25 nM concentration (Figure 118B). BCN-(M^{Pro})₂ was more active than DBCO-(M^{Pro})₂ in this assay. This observation met with our initial expectation. Because the DBCO group is bulkier than BCN we were expected BCN to induce less structural perturbation on the M^{Pro} structure. Another significant difference between these two linkers is the spacer arm length. Spacer arm length for BCN linker is much shorter than DBCO linker's arm length. Therefore, the short arm length of BCN might also help to hold two M^{Pro} units in proximity. Another reason for higher BCN activity might be the shorter reaction time of BCN conjugate. DBCO-(M^{Pro})₂ were prepared by reacting them at 25 °C for 20 hours whereas, BCN-(M^{Pro})₂ were prepared for only 10 hours. However, being unable to work with pure enzymes refrain us from making certain statements. Both conjugates have around ~70% purity therefore the concentration determination of each conjugate cannot be done accurately. As mentioned above, our attempts to purify conjugates from the monomeric M^{Pro} were not successful.

Finally, we compared 5 nM of BCN-(M^{Pro})₂ with 10 nM of wt-M^{Pro} (Figure 118C). One main reason why we started working on covalently dimerized M^{Pro} was to

use lower enzyme concentrations to characterize inhibitors with sub-nanomolar IC_{50} values if possible. While 10 nM of wt- M^{pro} is barely active, 5 nM of BCN- $(M^{pro})_2$ shows a clear signal.

We also tested the activity of DBCO- $(M^{pro})_2$ at various concentrations (0.63 nM – 10 nM) against 10 μ M FRET-Sub3 to see what is the lowest enzyme concentration that yields an observable signal (Figure 118D). The result showed that DBCO- $(M^{pro})_2$ could show activity at a concentration as low as 1.25 nM (2.5 nM total M^{pro} concentration). For wt- M^{pro} the lowest enzyme concentration that can give an observable signal 10 μ M FRET-Sub3 was 10 nM (Figure 120A).

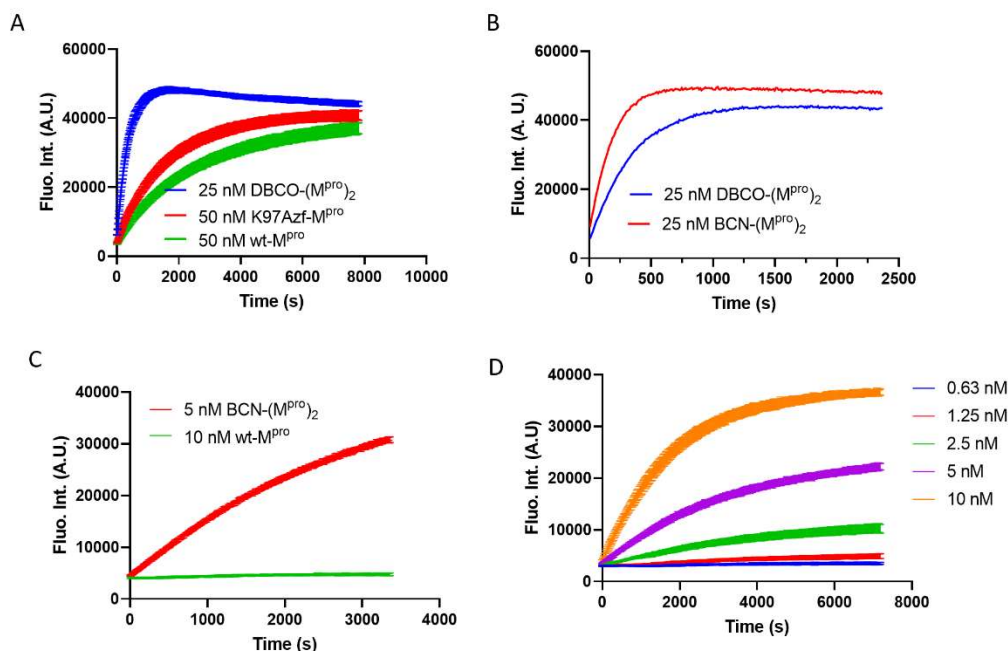


Figure 118 **A)** Comparison of wt- M^{pro} , KAzf- M^{pro} , DBCO- $(M^{pro})_2$ **B)** Comparison of DBCO- $(M^{pro})_2$ and BCN- $(M^{pro})_2$ **C)** Comparison of wt- M^{pro} and 5 nM BCN- $(M^{pro})_2$ **D)** Activity of DBCO- $(M^{pro})_2$ at various concentrations

IC_{50} Analysis of the Inhibitors with BCN- $(M^{pro})_2$

After having our BCN-(M^{Pro})₂ be fairly active at 5 nM concentration, we wanted to try one of our new boceprevir derivative inhibitor with varying concentrations of M^{Pro} and 5 nM of BCN-(M^{Pro})₂. With 50 nM wt-M^{Pro} the IC₅₀ value was 33.4 ± 2.7 nM. With 20 nM wt-M^{Pro} the IC₅₀ value was 19.8 ± 1.8 nM. 5 nM of BCN-(M^{Pro})₂ gave an IC₅₀ value of 9.3 ± 0.8 nM (Figure 119 A-D).

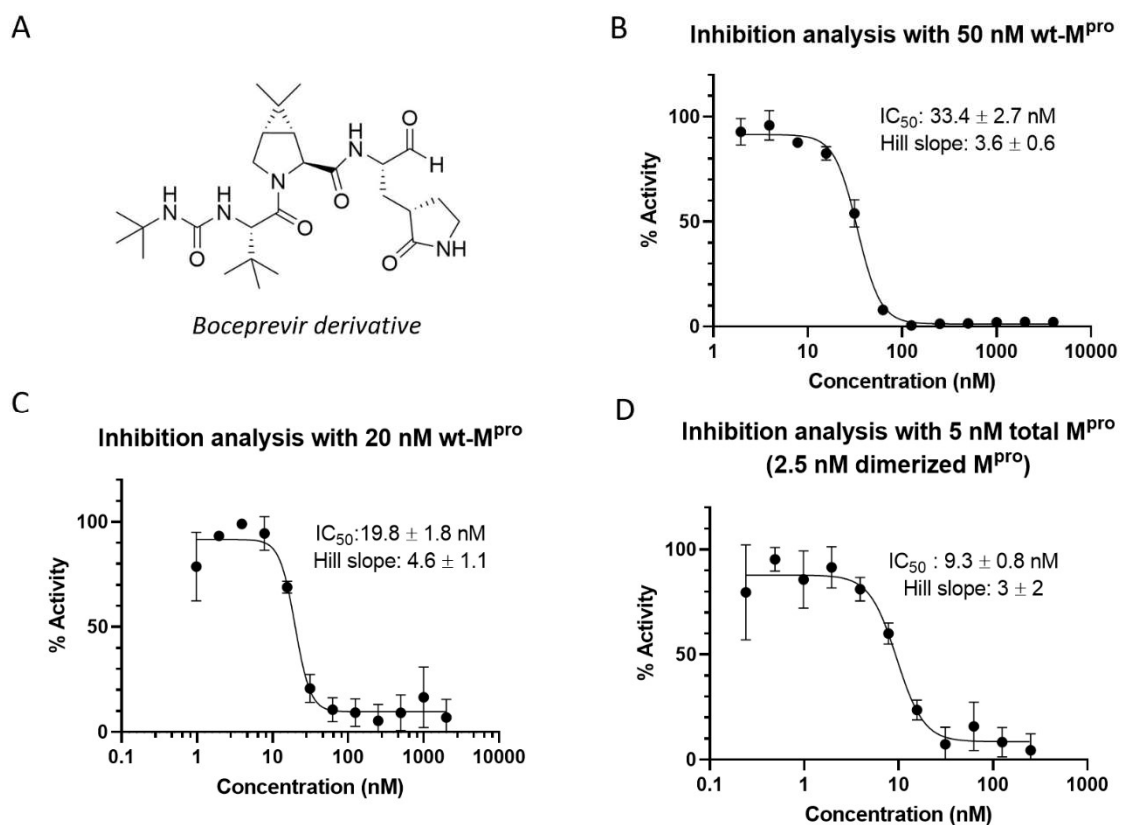


Figure 119 A) Structure of the inhibitor tested B) Inhibition analysis with 50 nM M^{Pro} C) Inhibition analysis with 20 nM M^{Pro} D) Inhibition analysis with 5 nM M^{Pro} BCN-(M^{Pro})₂

5.3.8 Testing varying concentrations of wt-M^{pro} and Dimerized M^{pro} Against Different Substrates

It is widely accepted that M^{pro} shows its activity as a dimer. When we tested different wt-M^{pro} concentrations (10, 20, 30, 40, 50 nM) against 10 μM of FRET-Sub3 we observed a non-linear relationship when initial slopes vs M^{pro} concentrations were plotted. The initial slope roughly quadrupled with doubling M^{pro} concentration (Figure 120A). This result was not surprising because, M^{pro} is believed to be acting as a dimer. However, when a similar experiment was conducted with our coumarin-based substrates 6mer-Cou-Sub (Figure 120B) and VLQ-Cou-Sub (Figure 120C), the result was a linear relationship between M^{pro} concentration and initial slope. It seems like FRET-Sub3 and coumarin-based substrates are processed by a different enzyme mechanism and it is possible that, for coumarin-based substrates dimer is not the active form of the M^{pro}. In order to show that dimerized M^{pro} is active as a monomeric unit, DBCO-(M^{pro})₂ concentrations were varied against constant FRET-Sub3. The resulting concentration-initial slope was not linear (Figure 120D). The initial slope tripled with doubling concentration. For wt-M^{pro} the initial slope quadrupled with doubling the concentration. Therefore, the relationship in the case of DBCO-(M^{pro})₂ seems to be closer to linear. However, for more conclusive data this experiment needs to be repeated with pure dimerized M^{pro}. Having monomeric impurities eliminated can give an initial slope M^{pro} concentration relation closer to be a linear trend.

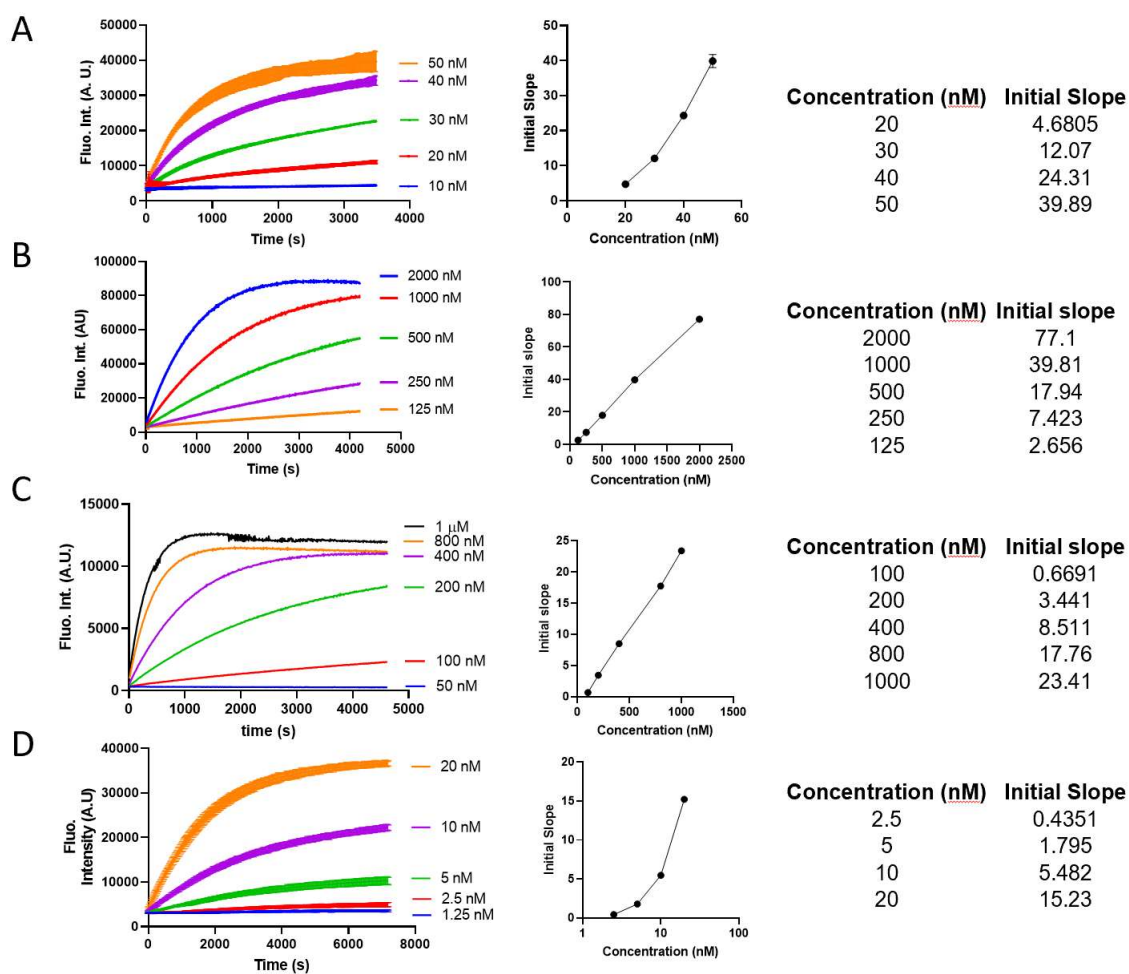


Figure 120 **A**) Activity of various concentrations of wt-M^{Pro} against 10 μ M of FRET-Sub3 **B**) Activity of various concentrations of wt-M^{Pro} against 10 μ M of 6mer-Cou-Sub **C**) Activity of various concentrations of wt-M^{Pro} against 10 μ M of VLQ-Cou-Sub **D**) Activity of various concentrations of DBCO-(M^{Pro})₂ against 10 μ M of FRET-Sub3

5.4 Conclusion

In conclusion, MPI1-MPI9 compounds showed to be potent inhibitors of M^{pro}.

Interaction of compounds MPI1, MPI3-MPI8 with M^{pro} were studied with crystallography. The live-virus microneutralization assays showed the MPI8 compound to be the most potent inhibitor that can inhibit SARS-CoV-2 replication at 0.3 μ M in A549/ACE2 cells. Then, we turned our attention to synthesize new reporter substrates for M^{pro} that will allow a more accurate determination of K_M . In order to lock M^{pro} in its active dimer form, we incorporated *para*-azido-L-phenylalanine (*p*-AzF) into M^{pro}. Then, we cross-linked it with di-cyclooctyne linkers to fix the dimer structure. We observed significantly increased enzyme activity with this covalently dimerized M^{pro}.

Targeting M^{pro} of SARS-CoV-2 continues to be a major research direction in our lab. Synthesis and optimization of aldehyde-based peptide SARS-CoV-2 inhibitors that show potency in both inhibition analysis of M^{pro} and live-virus microneutralization assays are ongoing research directions in our lab. We recently, turned our attention to the synthesis and characterization of proteolysis targeting chimera (PROTAC) compounds based on our knowledge on SARS-CoV-2 and M^{pro} inhibition analysis. New reporter substrates are still being developed and tested for more accurate measurement of K_M .

CHAPTER VI

CONCLUDING REMARKS

This thesis presents findings of my Ph.D. work in three different areas. The first area is the synthesis of antibody-drug conjugates. Our investigation started with examining the kinetics of various azides and alkynes to find an optimal way to apply ligand-accelerated copper-catalyzed azide alkyne cycloaddition reaction (CuAAC) to antibody-drug conjugate synthesis. Kinetics study showed that in the presence of BTAA ligand with non-chelating azide and alkyne, the CuAAC reaction rate depends on the concentration of azide and is independent of the concentration of alkyne. When a chelating-azide is used instead of a non-chelating azide, the CuAAC reaction became independent of the concentration of azide and independent of the concentration of alkyne. We synthesized ADCs both by using chelating-azide and non-chelating azide and found that non-chelating azides provided an efficient synthesis of ADCs on resin. ADCs synthesized using non-chelating azide in the solution phase, showed efficacy similar to widely used cysteine-maleimide ADCs and caused less side toxicity in cancer cell line MDA-MB 453. Then, when we were in the process of obtaining cell-based data for ADCs that are conjugated by chelating azide, COVID-19 surged and we ended the project.

The second area is genetic code expansion. Efforts to increase the yield of non-canonical amino acid incorporation are still a hot research topic. In this study, an already known aminoacyl tRNA synthetase non-canonical amino acid pair was examined.

Mutations were introduced step-wise to gain more insight into the pyrrolysine tRNA synthetase (PylRS) enzyme. During this investigation, serendipitous discovery revealed misincorporation of phenylalanine into a protein of interest. By digging further, we identified 2-chloro-L-phenylalanine (*o*-ClF) to be an efficient substrate for N346D, C348S, Y384F PylRS mutant. The crystallographic study hinted that the reason for the increased incorporation is the better fitting of the chlorine atom into the cavity between Asn346 and Cys348 residues. On the other hand, our quest to incorporate MeHAr did not result in incorporation of desired HAr derivatives. ESI-MS analysis showed that expressed proteins contain a different residue. That residue has not been identified. MS/MS studies are currently being conducted to characterize this residue. Once this residue is identified, it will be an intriguing task to identify the enzyme responsible for that transformation.

The third area is studies conducted on SARS-CoV-2 main protease (M^{pro}). Due to the urgency of the SARS-CoV-2 pandemic, our efforts began with docking of selected FDA/EMA (Food and Drug Administration/European Medical Association) approved drugs to the active site of M^{pro} . 29 of the drugs were chosen based on their $\Delta G_{binding}$ and purchased from commercial sources. IC_{50} analysis of these drugs found show seven drugs show IC_{50} values lower than 100 μ M. The most potent binder was pimozide with 42 μ M. Then, we submitted three of the most potent inhibitors to test against SARS-CoV-2 in a live-virus microneutralization assay to identify Bepridil as a potent drug that inhibited SARS-CoV-2 inhibition at low-micromolar concentrations. Unfortunately, Bepridil was withdrawn from the US market but, it is still in use in Japan. We ordered

good manufacturing practice (GMP) grade Bepridil and looking forward to testing the efficacy of Bepridil in animal studies. We also used peptide-based compound with aldehyde warheads (MPI1-MPI9) to target M^{pro}. The lowest K_i identified was 8.5 ± 1.5 nM for MPI3. Crystal structures of MPI1, MPI3-MPI8 were determined. Live virus-based microneutralization assays showed complete inhibition of SARS-CoV-2 replication by MPI3 at 0.3 μM concentration. Next, we focused our attention on improving the activity of M^{pro} by bringing minimal structural perturbation. Because, M^{pro} requires dimerization and K_d is too high to have activity at low-nanomolar concentrations, we dimerized M^{pro} by incorporating *para*-azido-L-phenylalanine (*p*-AzF) followed by, cross-linking with dicyclooctyne linkers. The dimerized M^{pro} allowed us to get activity concentrations at low-nanomolar concentrations. We are looking forward to optimizing the dimerization approach and extending it to other viral proteases that require dimerization for their activity. In addition, we are still working on finding new peptide-based inhibitors having an aldehyde warhead to target M^{pro}. We are also working on PROTAC compounds that can target M^{pro} based on the potency of our inhibitors MPI1-MPI9.

REFERENCES

1. Wang, L., et al., *Expanding the genetic code of Escherichia coli*. *Science*, 2001. **292**(5516): p. 498-500.
2. Isenegger, P.G. and B.G. Davis, *Concepts of Catalysis in Site-Selective Protein Modifications*. *J Am Chem Soc*, 2019. **141**(20): p. 8005-8013.
3. Wan, W., J.M. Tharp, and W.R. Liu, *Pyrrolysyl-tRNA synthetase: An ordinary enzyme but an outstanding genetic code expansion tool*. *Biochimica Et Biophysica Acta-Proteins and Proteomics*, 2014. **1844**(6): p. 1059-1070.
4. Lang, K. and J.W. Chin, *Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins*. *Chemical Reviews*, 2014. **114**(9): p. 4764-4806.
5. Conibear, A.C., *Deciphering protein post-translational modifications using chemical biology tools*. *Nature Reviews Chemistry*, 2020. **4**(12): p. 674-695.
6. Yang, A., K. Cho, and H.S. Park, *Chemical biology approaches for studying posttranslational modifications*. *Rna Biology*, 2018. **15**(4-5): p. 427-440.
7. Park, H.S., et al., *Expanding the Genetic Code of Escherichia coli with Phosphoserine*. *Science*, 2011. **333**(6046): p. 1151-1154.
8. Xiao, H., et al., *Genetic Incorporation of Histidine Derivatives Using an Engineered Pyrrolysyl-tRNA Synthetase*. *Acs Chemical Biology*, 2014. **9**(5): p. 1092-1096.
9. Neumann, H., S.Y. Peak-Chew, and J.W. Chin, *Genetically encoding N-epsilon-acetyllysine in recombinant proteins*. *Nature Chemical Biology*, 2008. **4**(4): p. 232-234.
10. Lee, Y.J., et al., *A Genetically Encoded Acrylamide Functionality*. *Acs Chemical Biology*, 2013. **8**(8): p. 1664-1670.
11. Xiao, H., et al., *Genetic Incorporation of epsilon-N-2-Hydroxyisobutyryl-lysine into Recombinant Histones*. *Acs Chemical Biology*, 2015. **10**(7): p. 1599-1603.
12. Wang, Z.P.A., et al., *A Versatile Approach for Site-Specific Lysine Acylation in Proteins*. *Angewandte Chemie-International Edition*, 2017. **56**(6): p. 1643-1647.
13. Hoppmann, C., et al., *Site-specific incorporation of phosphotyrosine using an expanded genetic code*. *Nature Chemical Biology*, 2017. **13**(8): p. 842-+.
14. Wang, Z.P.A., et al., *A Genetically Encoded Allylsine for the Synthesis of Proteins with Site-Specific Lysine Dimethylation*. *Angewandte Chemie-International Edition*, 2017. **56**(1): p. 212-216.
15. Mondal, S., et al., *Site-specific incorporation of citrulline into proteins in mammalian cells*. *Nat Commun*, 2021. **12**(1): p. 45.
16. Smith, G.P., *Phage Display: Simple Evolution in a Petri Dish (Nobel Lecture)*. *Angewandte Chemie-International Edition*, 2019. **58**(41): p. 14428-14437.
17. Feng, T., M.L. Tsao, and P.G. Schultz, *A phage display system with unnatural amino acids*. *Journal of the American Chemical Society*, 2004. **126**(49): p. 15962-15963.

18. Liu, C.C., et al., *Protein evolution with an expanded genetic code*. Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(46): p. 17688-17693.
19. Tharp, J.M., et al., *An amber obligate active site-directed ligand evolution technique for phage display*. Nature Communications, 2020. **11**(1).
20. Wang, X.S.S., et al., *A Genetically Encoded, Phage-Displayed Cyclic-Peptide Library*. Angewandte Chemie-International Edition, 2019. **58**(44): p. 15904-15909.
21. Oller-Salvia, B. and J.W. Chin, *Efficient Phage Display with Multiple Distinct Non-Canonical Amino Acids Using Orthogonal Ribosome-Mediated Genetic Code Expansion*. Angewandte Chemie-International Edition, 2019. **58**(32): p. 10844-10848.
22. Johnson, D.B.F., et al., *Release Factor One Is Nonessential in Escherichia coli*. Acs Chemical Biology, 2012. **7**(8): p. 1337-1344.
23. Mukai, T., et al., *Highly reproductive Escherichia coli cells with no specific assignment to the UAG codon*. Scientific Reports, 2015. **5**.
24. Wan, W., et al., *A Facile System for Genetic Incorporation of Two Different Noncanonical Amino Acids into One Protein in Escherichia coli*. Angewandte Chemie-International Edition, 2010. **49**(18): p. 3211-3214.
25. Mukai, T., et al., *Reassignment of a rare sense codon to a non-canonical amino acid in Escherichia coli*. Nucleic Acids Research, 2015. **43**(16): p. 8111-8122.
26. Lee, B.S., et al., *Incorporation of Unnatural Amino Acids in Response to the AGG Codon*. Acs Chemical Biology, 2015. **10**(7): p. 1648-1653.
27. Zeng, Y., W. Wang, and W.S.R. Liu, *Towards Reassigning the Rare AGG Codon in Escherichia coli*. Chembiochem, 2014. **15**(12): p. 1750-1754.
28. Anderson, J.C., et al., *An expanded genetic code with a functional quadruplet codon*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(20): p. 7566-7571.
29. Neumann, H., et al., *Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome*. Nature, 2010. **464**(7287): p. 441-444.
30. Dunkelmann, D.L., et al., *Engineered triply orthogonal pyrrolysyl-tRNA synthetase/tRNA pairs enable the genetic encoding of three distinct non-canonical amino acids*. Nature Chemistry, 2020. **12**(6): p. 535-+.
31. Italia, J.S., et al., *Mutually Orthogonal Nonsense-Suppression Systems and Conjugation Chemistries for Precise Protein Labeling at up to Three Distinct Sites*. Journal of the American Chemical Society, 2019. **141**(15): p. 6204-6212.
32. Malyshev, D.A., et al., *A semi-synthetic organism with an expanded genetic alphabet*. Nature, 2014. **509**(7500): p. 385-+.
33. Fischer, E.C., et al., *New codons for efficient production of unnatural proteins in a semisynthetic organism*. Nature Chemical Biology, 2020. **16**(5): p. 570-+.
34. Fredens, J., et al., *Total synthesis of Escherichia coli with a recoded genome*. Nature, 2019. **569**(7757): p. 514-+.

35. Borrel, G., et al., *Genome Sequence of "Candidatus Methanomethylophilus alvus" Mx1201, a Methanogenic Archaeon from the Human Gut Belonging to a Seventh Order of Methanogens*. Journal of Bacteriology, 2012. **194**(24): p. 6944-6945.
36. Borrel, G., et al., *Comparative genomics highlights the unique biology of Methanomassiliicoccales, a Thermoplasmatales-related seventh order of methanogenic archaea that encodes pyrrolysine*. BMC Genomics, 2014. **15**.
37. Meineke, B., et al., *Methanomethylophilus alvus Mx1201 Provides Basis for Mutual Orthogonal Pyrrolysyl tRNA/Aminoacyl-tRNA Synthetase Pairs in Mammalian Cells*. ACS Chemical Biology, 2018. **13**(11): p. 3087-3096.
38. Herring, S., et al., *The amino-terminal domain of pyrrolysyl-tRNA synthetase is dispensable in vitro but required for in vivo activity*. FEBS Letters, 2007. **581**(17): p. 3197-3203.
39. Yamaguchi, A., et al., *Pyrrolysyl-tRNA Synthetase with a Unique Architecture Enhances the Availability of Lysine Derivatives in Synthetic Genetic Codes*. Molecules, 2018. **23**(10).
40. Chiu, M.L., et al., *Antibody Structure and Function: The Basis for Engineering Therapeutics*. Antibodies (Basel), 2019. **8**(4).
41. Liu, H. and K. May, *Disulfide bond structures of IgG molecules: structural variations, chemical modifications and possible impacts to stability and biological function*. MAbs, 2012. **4**(1): p. 17-23.
42. Bournazos, S., et al., *Signaling by Antibodies: Recent Progress*. Annu Rev Immunol, 2017. **35**: p. 285-311.
43. Ben Ayed, W., et al., *Toxicity, risk factors and management of cisplatin-induced toxicity: A prospective study*. J Oncol Pharm Pract, 2020. **26**(7): p. 1621-1629.
44. Nahta, R., *Molecular Mechanisms of Trastuzumab-Based Treatment in HER2-Overexpressing Breast Cancer*. ISRN Oncol, 2012. **2012**: p. 428062.
45. Dan, N., et al., *Antibody-Drug Conjugates for Cancer Therapy: Chemistry to Clinical Implications*. Pharmaceuticals (Basel), 2018. **11**(2).
46. van de Donk, N.W. and E. Dhimolea, *Brentuximab vedotin*. MAbs, 2012. **4**(4): p. 458-65.
47. Saito, F., H. Noda, and J.W. Bode, *Critical evaluation and rate constants of chemoselective ligation reactions for stoichiometric conjugations in water*. ACS Chem Biol, 2015. **10**(4): p. 1026-33.
48. Chalouni, C. and S. Doll, *Fate of Antibody-Drug Conjugates in Cancer Cells*. J Exp Clin Cancer Res, 2018. **37**(1): p. 20.
49. Kellner, C., et al., *Modulating Cytotoxic Effector Functions by Fc Engineering to Improve Cancer Therapy*. Transfus Med Hemother, 2017. **44**(5): p. 327-336.
50. Lucas, A.T., et al., *Factors Affecting the Pharmacology of Antibody-Drug Conjugates*. Antibodies (Basel), 2018. **7**(1).
51. Lambert, J.M. and R.V. Chari, *Ado-trastuzumab Emtansine (T-DM1): an antibody-drug conjugate (ADC) for HER2-positive breast cancer*. J Med Chem, 2014. **57**(16): p. 6949-64.
52. Bargh, J.D., et al., *Cleavable linkers in antibody-drug conjugates*. Chem Soc Rev, 2019. **48**(16): p. 4361-4374.

53. Nejadmoghammad, M.R., et al., *Antibody-Drug Conjugates: Possibilities and Challenges*. Avicenna J Med Biotechnol, 2019. **11**(1): p. 3-23.
54. Lyon, R.P., et al., *Reducing hydrophobicity of homogeneous antibody-drug conjugates improves pharmacokinetics and therapeutic index*. Nature Biotechnology, 2015. **33**(7): p. 733-735.
55. Agarwal, P. and C.R. Bertozzi, *Site-specific antibody-drug conjugates: the nexus of bioorthogonal chemistry, protein engineering, and drug development*. Bioconjug Chem, 2015. **26**(2): p. 176-92.
56. Lambert, J.M. and R.V.J. Chari, *Ado-trastuzumab Emtansine (T-DM1): An Antibody-Drug Conjugate (ADC) for HER2-Positive Breast Cancer*. Journal of Medicinal Chemistry, 2014. **57**(16): p. 6949-6964.
57. Ravasco, J.M.J.M., et al., *Bioconjugation with Maleimides: A Useful Tool for Chemical Biology*. Chemistry-a European Journal, 2019. **25**(1): p. 43-59.
58. Tumey, L.N., et al., *Mild method for succinimide hydrolysis on ADCs: impact on ADC potency, stability, exposure, and efficacy*. Bioconjug Chem, 2014. **25**(10): p. 1871-80.
59. Lyon, R.P., et al., *Self-hydrolyzing maleimides improve the stability and pharmacological properties of antibody-drug conjugates*. Nat Biotechnol, 2014. **32**(10): p. 1059-62.
60. Khongorzul, P., et al., *Antibody-Drug Conjugates: A Comprehensive Review*. Mol Cancer Res, 2020. **18**(1): p. 3-19.
61. Schumacher, F.F., et al., *In Situ Maleimide Bridging of Disulfides and a New Approach to Protein PEGylation*. Bioconjugate Chemistry, 2011. **22**(2): p. 132-136.
62. VanBrunt, M.P., et al., *Genetically Encoded Azide Containing Amino Acid in Mammalian Cells Enables Site-Specific Antibody-Drug Conjugates Using Click Cycloaddition Chemistry*. Bioconjugate Chemistry, 2015. **26**(11): p. 2249-2260.
63. Oller-Salvia, B., G. Kym, and J.W. Chin, *Rapid and Efficient Generation of Stable Antibody-Drug Conjugates via an Encoded Cyclopropene and an Inverse-Electron-Demand Diels-Alder Reaction*. Angew Chem Int Ed Engl, 2018. **57**(11): p. 2831-2834.
64. Zimmerman, E.S., et al., *Production of Site-Specific Antibody-Drug Conjugates Using Optimized Non-Natural Amino Acids in a Cell-Free Expression System*. Bioconjugate Chemistry, 2014. **25**(2): p. 351-361.
65. Axup, J.Y., et al., *Synthesis of site-specific antibody-drug conjugates using unnatural amino acids*. Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(40): p. 16101-16106.
66. Bonandi, E., et al., *The 1,2,3-triazole ring as a bioisostere in medicinal chemistry*. Drug Discovery Today, 2017. **22**(10): p. 1572-1581.
67. Dennler, P., et al., *Transglutaminase-Based Chemo-Enzymatic Conjugation Approach Yields Homogeneous Antibody-Drug Conjugates*. Bioconjugate Chemistry, 2014. **25**(3): p. 569-578.

68. Beerli, R.R., et al., *Sortase Enzyme-Mediated Generation of Site-Specifically Conjugated Antibody Drug Conjugates with High In Vitro and In Vivo Potency*. Plos One, 2015. **10**(7).
69. [cited 2021 February, 9]; Available from: <https://coronavirus.jhu.edu/map.html>.
70. Pislari, A., et al., *The role of cysteine peptidases in coronavirus cell entry and replication: The therapeutic potential of cathepsin inhibitors*. Plos Pathogens, 2020. **16**(11).
71. Morse, J.S., et al., *Learning from the Past: Possible Urgent Prevention and Treatment Options for Severe Acute Respiratory Infections Caused by 2019-nCoV*. Chembiochem, 2020. **21**(5): p. 730-738.
72. Bestle, D., et al., *TMPRSS2 and furin are both essential for proteolytic activation of SARS-CoV-2 in human airway cells*. Life Science Alliance, 2020. **3**(9).
73. Eastman, R.T., et al., *Remdesivir: A Review of Its Discovery and Development Leading to Emergency Use Authorization for Treatment of COVID-19*. Acs Central Science, 2020. **6**(5): p. 672-683.
74. Li, Z., et al., *Identify potent SARS-CoV-2 main protease inhibitors via accelerated free energy perturbation-based virtual screening of existing drugs*. Proceedings of the National Academy of Sciences of the United States of America, 2020. **117**(44): p. 27381-27387.
75. Vuong, W., et al., *Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication (vol 11, 4282, 2020)*. Nature Communications, 2020. **11**(1).
76. Kaptein, S.J.F., et al., *Favipiravir at high doses has potent antiviral activity in SARS-CoV-2-infected hamsters, whereas hydroxychloroquine lacks activity*. Proceedings of the National Academy of Sciences of the United States of America, 2020. **117**(43): p. 26955-26965.
77. Maciorowski, D., et al., *A Review of the Preclinical and Clinical Efficacy of Remdesivir, Hydroxychloroquine, and Lopinavir-Ritonavir Treatments against COVID-19*. Slas Discovery, 2020. **25**(10): p. 1108-1122.
78. Xiong, M.Y., et al., *What coronavirus 3C-like protease tells us: From structure, substrate selectivity, to inhibitor design*. Medicinal Research Reviews, 2021.
79. Zhang, L.L., et al., *Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved alpha-ketoamide inhibitors*. Science, 2020. **368**(6489): p. 409-+.
80. Solowiej, J., et al., *Steady-state and pre-steady-state kinetic evaluation of severe acute respiratory syndrome coronavirus (SARS-CoV) 3CL(pro) cysteine protease: Development of an ion-pair model for catalysis*. Biochemistry, 2008. **47**(8): p. 2617-2630.
81. Lim, L.Z., et al., *Dynamically-Driven Enhancement of the Catalytic Machinery of the SARS 3C-Like Protease by the S284-T285-I286/A Mutations on the Extra Domain*. Plos One, 2014. **9**(7).
82. Huang, C.K., et al., *3C-like proteinase from SARS coronavirus catalyzes substrate hydrolysis by a general base mechanism*. Biochemistry, 2004. **43**(15): p. 4568-4574.

83. Kneller, D.W., et al., *Unusual zwitterionic catalytic site of SARS-CoV-2 main protease revealed by neutron crystallography*. Journal of Biological Chemistry, 2020. **295**(50): p. 17365-17373.
84. Ramos-Guzman, C.A., J.J. Ruiz-Pernia, and I. Tunon, *Unraveling the SARS-CoV-2 Main Protease Mechanism Using Multiscale Methods*. Acs Catalysis, 2020. **10**(21): p. 12544-12554.
85. Chuck, C.P., et al., *Profiling of Substrate Specificity of SARS-CoV 3CL(pro)*. Plos One, 2010. **5**(10).
86. Rut, W., et al., *SARS-CoV-2 M-pro inhibitors and activity-based probes for patient-sample imaging*. Nature Chemical Biology, 2021. **17**(2): p. 222-+.
87. Grum-Tokars, V., et al., *Evaluating the 3C-like protease activity of SARS-Coronavirus: Recommendations for standardized assays for drug discovery*. Virus Research, 2008. **133**(1): p. 63-73.
88. Ma, C.L., et al., *Boceprevir, GC-376, and calpain inhibitors II, XII inhibit SARS-CoV-2 viral replication by targeting the viral main protease*. Cell Research, 2020. **30**(8): p. 678-692.
89. Kneller, D.W., et al., *Malleability of the SARS-CoV-2 3CL M-pro Active-Site Cavity Facilitates Binding of Clinical Antivirals*. Structure, 2020. **28**(12): p. 1313-+.
90. Bacha, U., et al., *Identification of novel inhibitors of the SARS coronavirus main protease 3CL(pro)*. Biochemistry, 2004. **43**(17): p. 4906-4912.
91. Kao, R.Y., et al., *Characterization of SARS-CoV main protease and identification of biologically active small molecule inhibitors using a continuous fluorescence-based assay*. Febs Letters, 2004. **576**(3): p. 325-330.
92. Blanchard, J.E., et al., *High-throughput screening identifies inhibitors of the SARS coronavirus main proteinase*. Chemistry & Biology, 2004. **11**(10): p. 1445-1453.
93. Kaeppler, U., et al., *A new lead for nonpeptidic active-site-directed inhibitors of the severe acute respiratory syndrome coronavirus main protease discovered by a combination of screening and docking methods*. Journal of Medicinal Chemistry, 2005. **48**(22): p. 6832-6842.
94. Lee, T.W., et al., *Crystal structures of the main peptidase from the SARS coronavirus inhibited by a substrate-like aza-peptide epoxide*. Journal of Molecular Biology, 2005. **353**(5): p. 1137-1151.
95. Liu, Y.C., et al., *Screening of drugs by FRET analysis identifies inhibitors of SARS-CoV 3CL protease*. Biochemical and Biophysical Research Communications, 2005. **333**(1): p. 194-199.
96. Chou, C.Y., et al., *Quaternary structure of the severe acute respiratory syndrome (SARS) coronavirus main protease*. Biochemistry, 2004. **43**(47): p. 14958-14970.
97. Kuang, W.F., et al., *Mutational and inhibitive analysis of SARS coronavirus 3C-like protease by fluorescence resonance energy transfer-based assays*. Biochemical and Biophysical Research Communications, 2005. **331**(4): p. 1554-1559.
98. Yang, H.T., et al., *Design of wide-spectrum inhibitors targeting coronavirus main proteases*. Plos Biology, 2005. **3**(10): p. 1742-1752.

99. Jin, Z.M., et al., *Structure of M-pro from SARS-CoV-2 and discovery of its inhibitors*. Nature, 2020. **582**(7811): p. 289-+.
100. Graziano, V., et al., *Enzymatic activity of the SARS coronavirus main proteinase dimer*. Febs Letters, 2006. **580**(11): p. 2577-2583.
101. Wei, P., et al., *The N-terminal octapeptide acts as a dimerization inhibitor of SARS coronavirus 3C-like proteinase*. Biochemical and Biophysical Research Communications, 2006. **339**(3): p. 865-872.
102. Liu, Z.M., et al., *Virtual screening of novel noncovalent inhibitors for SARS-CoV 3C-like proteinase*. Journal of Chemical Information and Modeling, 2005. **45**(1): p. 10-17.
103. Miczi, M., et al., *Identification of Host Cellular Protein Substrates of SARS-COV-2 Main Protease*. International Journal of Molecular Sciences, 2020. **21**(24).
104. Zhang, L.L., et al., *alpha-Ketoamides as Broad-Spectrum Inhibitors of Coronavirus and Enterovirus Replication: Structure-Based Design, Synthesis, and Activity Assessment*. Journal of Medicinal Chemistry, 2020. **63**(9): p. 4562-4578.
105. Chuck, C.P., et al., *Design, synthesis and crystallographic analysis of nitrile-based broad-spectrum peptidomimetic inhibitors for coronavirus 3C-like proteases*. European Journal of Medicinal Chemistry, 2013. **59**: p. 1-6.
106. Dai, W.H., et al., *Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease*. Science, 2020. **368**(6497): p. 1331-+.
107. Sacco, M.D., et al., *Structure and inhibition of the SARS-CoV-2 main protease reveal strategy for developing dual inhibitors against M-pro and cathepsin L*. Science Advances, 2020. **6**(50).
108. Wu, C.Y., et al., *Stable benzotriazole esters as mechanism-based inactivators of the severe acute respiratory syndrome 3CL protease*. Chemistry & Biology, 2006. **13**(3): p. 261-268.
109. Yang, H.T., et al., *The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor*. Proceedings of the National Academy of Sciences of the United States of America, 2003. **100**(23): p. 13190-13195.
110. Lee, C.C., et al., *Structural basis of inhibition specificities of 3C and 3C-like proteases by zinc-coordinating and peptidomimetic compounds*. J Biol Chem, 2009. **284**(12): p. 7646-55.
111. Yin, J., et al., *A mechanistic view of enzyme inhibition and peptide hydrolysis in the active site of the SARS-CoV 3C-like peptidase*. J Mol Biol, 2007. **371**(4): p. 1060-74.
112. Turlington, M., et al., *Discovery of N-(benzo[1,2,3]triazol-1-yl)-N-(benzyl)acetamido)phenyl carboxamides as severe acute respiratory syndrome coronavirus (SARS-CoV) 3CLpro inhibitors: Identification of ML300 and noncovalent nanomolar inhibitors with an induced-fit binding*. Bioorganic & Medicinal Chemistry Letters, 2013. **23**(22): p. 6172-6177.
113. Su, H.X., et al., *Anti-SARS-CoV-2 activities in vitro of Shuanghuanglian preparations and bioactive ingredients*. Acta Pharmacologica Sinica, 2020. **41**(9): p. 1167-1177.

114. Zhou, L., et al., *Isatin compounds as noncovalent SARS coronavirus 3C-like protease inhibitors*. Journal of Medicinal Chemistry, 2006. **49**(12): p. 3440-3443.
115. Jacobs, J., et al., *Discovery, Synthesis, And Structure-Based Optimization of a Series of N-(tert-Butyl)-2-(N-arylamido)-2-(pyridin-3-yl) Acetamides (ML188) as Potent Noncovalent Small Molecule Inhibitors of the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) 3CL Protease*. Journal of Medicinal Chemistry, 2013. **56**(2): p. 534-546.
116. Kolb, H.C., M.G. Finn, and K.B. Sharpless, *Click Chemistry: Diverse Chemical Function from a Few Good Reactions*. Angew Chem Int Ed Engl, 2001. **40**(11): p. 2004-2021.
117. Bohacek, R.S., C. McMartin, and W.C. Guida, *The art and practice of structure-based drug design: a molecular modeling perspective*. Med Res Rev, 1996. **16**(1): p. 3-50.
118. Yang, M., et al., *Biocompatible click chemistry enabled compartment-specific pH measurement inside E. coli*. Nat Commun, 2014. **5**: p. 4981.
119. Worrell, B.T., J.A. Malik, and V.V. Fokin, *Direct evidence of a dinuclear copper intermediate in Cu(I)-catalyzed azide-alkyne cycloadditions*. Science, 2013. **340**(6131): p. 457-60.
120. Jin, L., et al., *Isolation of bis(copper) key intermediates in Cu-catalyzed azide-alkyne "click reaction"*. Sci Adv, 2015. **1**(5): p. e1500304.
121. Ziegler, M.S., K.V. Lakshmi, and T.D. Tilley, *Dicopper Cu(I)Cu(I) and Cu(I)Cu(II) Complexes in Copper-Catalyzed Azide-Alkyne Cycloaddition*. J Am Chem Soc, 2017. **139**(15): p. 5378-5386.
122. Kuang, G.C., et al., *Experimental investigation on the mechanism of chelation-assisted, copper(II) acetate-accelerated azide-alkyne cycloaddition*. J Am Chem Soc, 2011. **133**(35): p. 13984-4001.
123. Rodionov, V.O., V.V. Fokin, and M.G. Finn, *Mechanism of the ligand-free CuI-catalyzed azide-alkyne cycloaddition reaction*. Angew Chem Int Ed Engl, 2005. **44**(15): p. 2210-5.
124. Seath, C.P., G.A. Burley, and A.J. Watson, *Determining the Origin of Rate-Independent Chemoselectivity in CuAAC Reactions: An Alkyne-Specific Shift in Rate-Determining Step*. Angew Chem Int Ed Engl, 2017. **56**(12): p. 3314-3318.
125. Wojciechowski, F., et al., *A robust and convergent synthesis of Dipeptide-DOTAM conjugates as Chelators for lanthanide ions: New PARACEST MRI agents*. Bioconjugate Chemistry, 2007. **18**(5): p. 1625-1636.
126. Angelos, S., et al., *pH-responsive supramolecular nanovalves based on cucurbit[6]uril pseudorotaxanes*. Angewandte Chemie-International Edition, 2008. **47**(12): p. 2222-2226.
127. Woods, M. and A.D. Sherry, *An improved and versatile synthetic route to 6,7 : 13,14-dibenzo-1,8,4,11-dioxadiazacyclotetradecane*. Inorganica Chimica Acta, 2003. **351**: p. 395-398.
128. Wendeln, C., et al., *Rapid Preparation of Multifunctional Surfaces for Orthogonal Ligation by Microcontact Chemistry*. Chemistry-a European Journal, 2012. **18**(19): p. 5880-5888.

129. Dubowchik, G.M., et al., *Cathepsin B-labile dipeptide linkers for lysosomal release of doxorubicin from internalizing immunoconjugates: Model studies of enzymatic drug release and antigen-specific in vitro anticancer activity*. *Bioconjugate Chemistry*, 2002. **13**(4): p. 855-869.
130. Uttamapinant, C., et al., *Fast, Cell-Compatible Click Chemistry with Copper-Chelating Azides for Biomolecular Labeling*. *Angewandte Chemie-International Edition*, 2012. **51**(24): p. 5852-5856.
131. Shiraishi, Y., S. Sumiya, and T. Hirai, *A coumarin-thiourea conjugate as a fluorescent probe for Hg(II) in aqueous media with a broad pH range 2-12*. *Organic & Biomolecular Chemistry*, 2010. **8**(6): p. 1310-1314.
132. Kuang, G.C., et al., *Chelation-Assisted, Copper(II)-Acetate-Accelerated Azide-Alkyne Cycloaddition*. *Journal of Organic Chemistry*, 2010. **75**(19): p. 6540-6548.
133. Leophairatana, P., et al., *Preventing Alkyne-Alkyne (i.e., Glaser) Coupling Associated with the ATRP Synthesis of Alkyne-Functional Polymers/Macromonomers and for Alkynes under Click (i.e., CuAAC) Reaction Conditions*. *Journal of the American Chemical Society*, 2017. **139**(10): p. 3756-3766.
134. Nami, B., H. Maadi, and Z.X. Wang, *Mechanisms Underlying the Action and Synergism of Trastuzumab and Pertuzumab in Targeting HER2-Positive Breast Cancer*. *Cancers*, 2018. **10**(10).
135. Kang, J.C., et al., *Engineering a HER2-specific antibody-drug conjugate to increase lysosomal delivery and therapeutic efficacy*. *Nature Biotechnology*, 2019. **37**(5): p. 523-+.
136. Bulau, P., et al., *Quantitative assessment of arginine methylation in free versus protein-incorporated amino acids in vitro and in vivo using protein hydrolysis and high-performance liquid chromatography*. *Biotechniques*, 2006. **40**(3): p. 305-310.
137. Fulton, M.D., T. Brown, and Y.G. Zheng, *The Biological Axis of Protein Arginine Methylation and Asymmetric Dimethylarginine*. *International Journal of Molecular Sciences*, 2019. **20**(13).
138. Wang, H.B., et al., *Methylation of histone H4 at arginine 3 facilitating transcriptional activation by nuclear hormone receptor*. *Science*, 2001. **293**(5531): p. 853-857.
139. Karmodiya, K., et al., *H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells*. *Bmc Genomics*, 2012. **13**.
140. Li, X.G., et al., *H4R3 methylation facilitates beta-globin transcription by regulating histone acetyltransferase binding and H3 acetylation*. *Blood*, 2010. **115**(10): p. 2028-2037.
141. Yang, Y.Z., et al., *TDRD3 Is an Effector Molecule for Arginine-Methylated Histone Marks*. *Molecular Cell*, 2010. **40**(6): p. 1016-1023.
142. Iberg, A.N., et al., *Arginine methylation of the histone H3 tail impedes effector binding*. *Journal of Biological Chemistry*, 2008. **283**(6): p. 3006-3010.

143. Chang, B.S., et al., *JMJD6 is a histone arginine demethylase*. *Science*, 2007. **318**(5849): p. 444-447.
144. Walport, L.J., et al., *Arginine demethylation is catalysed by a subset of JmjC histone lysine demethylases*. *Nature Communications*, 2016. **7**.
145. Wang, Y., et al., *Human PAD4 regulates histone arginine methylation levels via demethylimination*. *Science*, 2004. **306**(5694): p. 279-283.
146. Shetty, D., et al., *Synthesis and evaluation of macrocyclic amino acid derivatives for tumor imaging by gallium-68 positron emission tomography*. *Bioorganic & Medicinal Chemistry*, 2010. **18**(21): p. 7338-7347.
147. Moali, C., et al., *Recognition of alpha-amino acids bearing various C=NOH functions by nitric oxide synthase and arginase involves very different structural determinants*. *Biochemistry*, 2000. **39**(28): p. 8208-8218.
148. Ma, Z.D., C.S. Day, and U. Bierbach, *Unexpected reactivity of the 9-aminoacridine chromophore in guanidylolation reactions*. *Journal of Organic Chemistry*, 2007. **72**(14): p. 5387-5390.
149. Kelly, B. and I. Rozas, *Copper(II) chloride promoted transformation of amines into guanidines and asymmetrical N,N '-disubstituted guanidines*. *Tetrahedron Letters*, 2013. **54**(30): p. 3982-3984.
150. Kahvedzic, A., et al., *Aromatic Bis-N-hydroxyguanidinium Derivatives: Synthesis, Biophysical, and Biochemical Evaluations*. *Journal of Medicinal Chemistry*, 2013. **56**(2): p. 451-459.
151. Schade, D., J. Kotthaus, and B. Clement, *Efficient synthesis of optically pure N(omega)-alkylated L-arginines*. *Synthesis-Stuttgart*, 2008(15): p. 2391-2397.
152. Sharma, V., et al., *Evolving the N-Terminal Domain of Pyrrolysyl-tRNA Synthetase for Improved Incorporation of Noncanonical Amino Acids*. *Chembiochem*, 2018. **19**(1): p. 26-30.
153. Jiang, H.K., et al., *Linker and N-Terminal Domain Engineering of Pyrrolysyl-tRNA Synthetase for Substrate Range Shifting and Activity Enhancement*. *Frontiers in Bioengineering and Biotechnology*, 2020. **8**.
154. Lee, Y.J., et al., *Genetically encoded fluorophenylalanines enable insights into the recognition of lysine trimethylation by an epigenetic reader*. *Chemical Communications*, 2016. **52**(85): p. 12606-12609.
155. Sharma, V., Y.S. Wang, and W.S.R. Liu, *Probing the Catalytic Charge-Relay System in Alanine Racemase with Genetically Encoded Histidine Mimetics*. *Acs Chemical Biology*, 2016. **11**(12): p. 3305-3309.
156. Tuley, A. and W. Fast, *The Taxonomy of Covalent Inhibitors*. *Biochemistry*, 2018. **57**(24): p. 3326-3337.
157. McMahon, J.H., A. Udy, and A.Y. Peleg, *Remdesivir for the Treatment of Covid-19-Preliminary Report*. *New England Journal of Medicine*, 2020. **383**(10): p. 992-993.
158. Zhou, N., et al., *Glycopeptide Antibiotics Potently Inhibit Cathepsin L in the Late Endosome/Lysosome and Block the Entry of Ebola Virus, Middle East Respiratory Syndrome Coronavirus (MERS-CoV), and Severe Acute Respiratory Syndrome*

- Coronavirus (SARS-CoV)*. Journal of Biological Chemistry, 2016. **291**(17): p. 9218-9232.
159. Strating, J.R.P.M., et al., *Itraconazole Inhibits Enterovirus Replication by Targeting the Oxysterol-Binding Protein*. Cell Reports, 2015. **10**(4): p. 600-615.
 160. Mastrangelo, E., et al., *Ivermectin is a potent inhibitor of flavivirus replication specifically targeting NS3 helicase activity: new prospects for an old drug*. Journal of Antimicrobial Chemotherapy, 2012. **67**(8): p. 1884-1894.
 161. Rossignol, J.F., *Nitazoxanide: A first-in-class broad-spectrum antiviral agent*. Antiviral Research, 2014. **110**: p. 94-103.
 162. Mercorelli, B., G. Palu, and A. Loregian, *Drug Repurposing for Viral Infectious Diseases: How Far Are We?* Trends in Microbiology, 2018. **26**(10): p. 865-876.
 163. Hung, I.F.N., et al., *Triple combination of interferon beta-1b, lopinavir-ritonavir, and ribavirin in the treatment of patients admitted to hospital with COVID-19: an open-label, randomised, phase 2 trial*. Lancet, 2020. **395**(10238): p. 1695-1704.
 164. Ghahremanpour, M.M., et al., *Identification of 14 Known Drugs as Inhibitors of the Main Protease of SARS-CoV-2*. ACS Medicinal Chemistry Letters, 2020. **11**(12): p. 2526-2533.
 165. Agrawal, A.S., et al., *Passive Transfer of A Germline-like Neutralizing Human Monoclonal Antibody Protects Transgenic Mice Against Lethal Middle East Respiratory Syndrome Coronavirus Infection*. Scientific Reports, 2016. **6**.
 166. Du, L.Y., et al., *A Truncated Receptor-Binding Domain of MERS-CoV Spike Protein Potently Inhibits MERS-CoV Infection and Induces Strong Neutralizing Antibody Responses: Implication for Developing Therapeutics and Vaccines*. Plos One, 2013. **8**(12).
 167. Du, L.Y., et al., *A Conformation-Dependent Neutralizing Monoclonal Antibody Specifically Targeting Receptor-Binding Domain in Middle East Respiratory Syndrome Coronavirus Spike Protein*. Journal of Virology, 2014. **88**(12): p. 7045-7053.
 168. Li, W., et al., *High Potency of a Bivalent Human V-H Domain in SARS-CoV-2 Animal Models*. Cell, 2020. **183**(2): p. 429-+.
 169. Dragovich, P.S., et al., *Structure-based design, synthesis, and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 8. Pharmacological optimization of orally bioavailable 2-pyridone-containing peptidomimetics*. Journal of Medicinal Chemistry, 2003. **46**(21): p. 4572-4585.
 170. Kunz, K.M., *A Trial of Lopinavir-Ritonavir in Covid-19*. New England Journal of Medicine, 2020. **382**(21).
 171. Musarrat, F., et al., *The anti-HIV drug nelfinavir mesylate (Viracept) is a potent inhibitor of cell fusion caused by the SARSCoV-2 spike (S) glycoprotein warranting further evaluation as an antiviral against COVID-19 infections*. Journal of Medical Virology, 2020. **92**(10): p. 2087-2095.
 172. [cited 2021 February 13th]; Available from: <https://www.diamond.ac.uk/covid-19/for-scientists/Main-protease-structure-and-XChem.html>.
 173. Shoichet, B.K., *Interpreting steep dose-response curves in early inhibitor discovery*. Journal of Medicinal Chemistry, 2006. **49**(25): p. 7274-7277.

174. Yang, N.D. and H.M. Shen, *Targeting the Endocytic Pathway and Autophagy Process as a Novel Therapeutic Strategy in COVID-19*. International Journal of Biological Sciences, 2020. **16**(10): p. 1724-1731.
175. Savarino, A., et al., *Effects of chloroquine on viral infections: an old drug against today's diseases?* Lancet Infectious Diseases, 2003. **3**(11): p. 722-727.
176. Geleris, J., et al., *Observational Study of Hydroxychloroquine in Hospitalized Patients with Covid-19*. New England Journal of Medicine, 2020. **382**(25): p. 2411-2418.
177. Choudhary, R., G.K. Bohra, and D. Kumar, *Hydroxychloroquine with or without Azithromycin in Covid-19*. New England Journal of Medicine, 2021. **384**(2): p. 190-191.
178. Ohkuma, S. and B. Poole, *Cytoplasmic Vacuolation of Mouse Peritoneal-Macrophages and the Uptake into Lysosomes of Weakly Basic Substances*. Journal of Cell Biology, 1981. **90**(3): p. 656-664.
179. DeWald, L.E., et al., *The Calcium Channel Blocker Bepridil Demonstrates Efficacy in the Murine Model of Marburg Virus Disease*. Journal of Infectious Diseases, 2018. **218**: p. S588-S591.
180. Shapiro, W., R. Dibianco, and U. Thadani, *Comparative Efficacy of 200, 300 and 400 Mg of Bepridil for Chronic Stable Angina-Pectoris*. American Journal of Cardiology, 1985. **55**(7): p. C36-C42.
181. Wu, W.N., et al., *Disposition of Bepridil in Laboratory-Animals and Man*. Xenobiotica, 1992. **22**(2): p. 153-169.
182. Mitterreiter, S., et al., *Bepridil and Amiodarone Simultaneously Target the Alzheimer's Disease beta- and gamma-Secretase via Distinct Mechanisms*. Journal of Neuroscience, 2010. **30**(26): p. 8974-8983.
183. Wang, M.L., et al., *Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro*. Cell Research, 2020. **30**(3): p. 269-271.
184. Hollingshead, L.M., D. Faulds, and A. Fitton, *Bepridil - a Review of Its Pharmacological Properties and Therapeutic Use in Stable Angina-Pectoris*. Drugs, 1992. **44**(5): p. 835-857.
185. Somberg, J., et al., *Prolongation of Q_t Interval and Antiarrhythmic Action of Bepridil*. American Heart Journal, 1985. **109**(1): p. 19-27.
186. Behr, E.R. and D. Roden, *Drug-induced arrhythmia: pharmacogenomic prescribing?* European Heart Journal, 2013. **34**(2): p. 89-+.
187. Fu, L.F., et al., *Both Boceprevir and GC376 efficaciously inhibit SARS-CoV-2 by targeting its main protease*. Nature Communications, 2020. **11**(1).
188. Yang, S., et al., *Synthesis, crystal structure, structure-activity relationships, and antiviral activity of a potent SARS coronavirus 3CL protease inhibitor*. Journal of Medicinal Chemistry, 2006. **49**(16): p. 4971-4980.
189. Yang, K.S., et al., *A Quick Route to Multiple Highly Potent SARS-CoV-2 Main Protease Inhibitors***. Chemmedchem, 2020.
190. Fonin, A.V., et al., *Fluorescence of Dyes in Solutions with High Absorbance. Inner Filter Effect Correction*. Plos One, 2014. **9**(7).

191. Tomar, S., et al., *Ligand-induced Dimerization of Middle East Respiratory Syndrome (MERS) Coronavirus nsp5 Protease (3CL(pro)) IMPLICATIONS FOR nsp5 REGULATION AND THE DEVELOPMENT OF ANTIVIRALS*. Journal of Biological Chemistry, 2015. **290**(32): p. 19403-19422.
192. Jost, R., *Competitive Peptide-Inhibitors of Leucine Aminopeptidase - Specific Interaction of Threonine (Tertiary Butyl)-Peptides with Enzyme from Swine Kidney and from Bovine Eye Lens*. Febs Letters, 1973. **29**(1): p. 7-9.
193. Xie, X.P., et al., *A nanoluciferase SARS-CoV-2 for rapid neutralization testing and screening of anti-infective drugs for COVID-19*. Nature Communications, 2020. **11**(1).
194. Huang, Y., et al., *Genetic incorporation of an aliphatic keto-containing amino acid into proteins for their site-specific modifications*. Bioorganic & Medicinal Chemistry Letters, 2010. **20**(3): p. 878-880.
195. Vacic, V., et al., *Disease-Associated Mutations Disrupt Functionally Important Regions of Intrinsic Protein Disorder*. Plos Computational Biology, 2012. **8**(10).
196. Lee, C.C., et al., *Structural basis of mercury- and zinc-conjugated complexes as SARS-CoV 3C-like protease inhibitors*. Febs Letters, 2007. **581**(28): p. 5454-5458.
197. Kozak, J.J., H.B. Gray, and R.A. Garza-Lopez, *Structural stability of the SARS-CoV-2 main protease: Can metal ions affect function?* Journal of Inorganic Biochemistry, 2020. **211**.

APPENDIX A
SPECTRA AND CHROMATOGRAMS

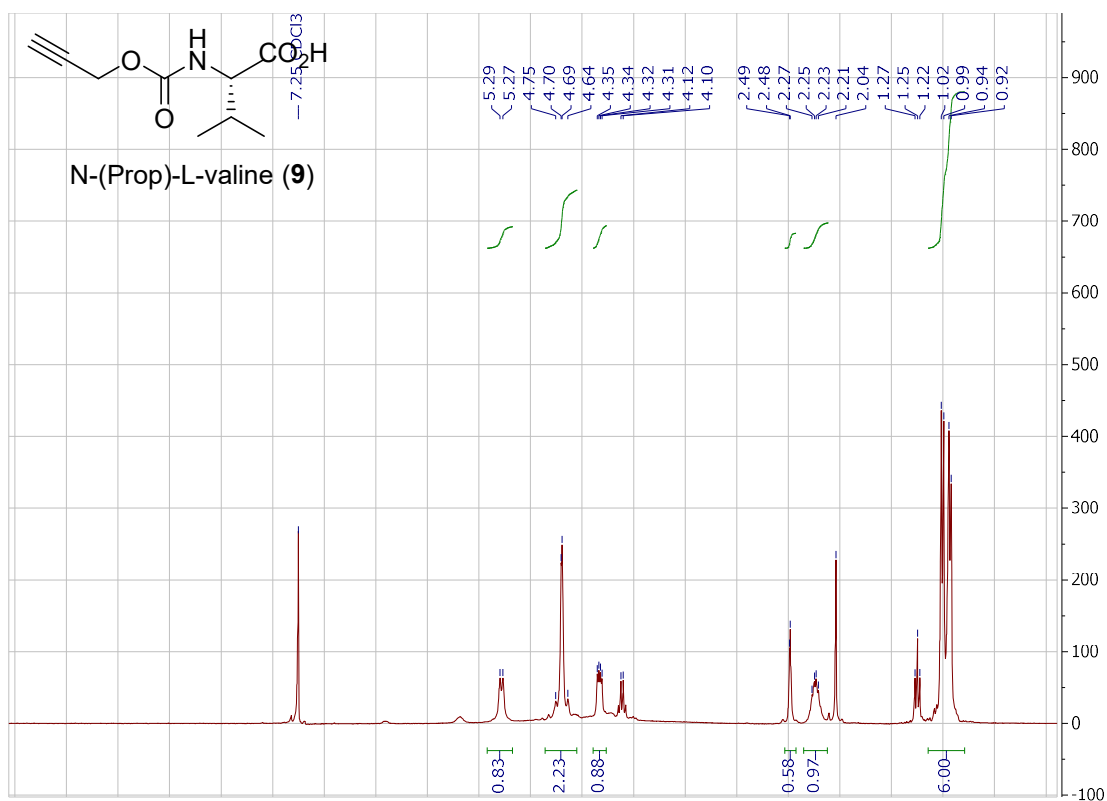


Figure 121 ^1H NMR spectrum of N-(Prop)-L-valine (9)

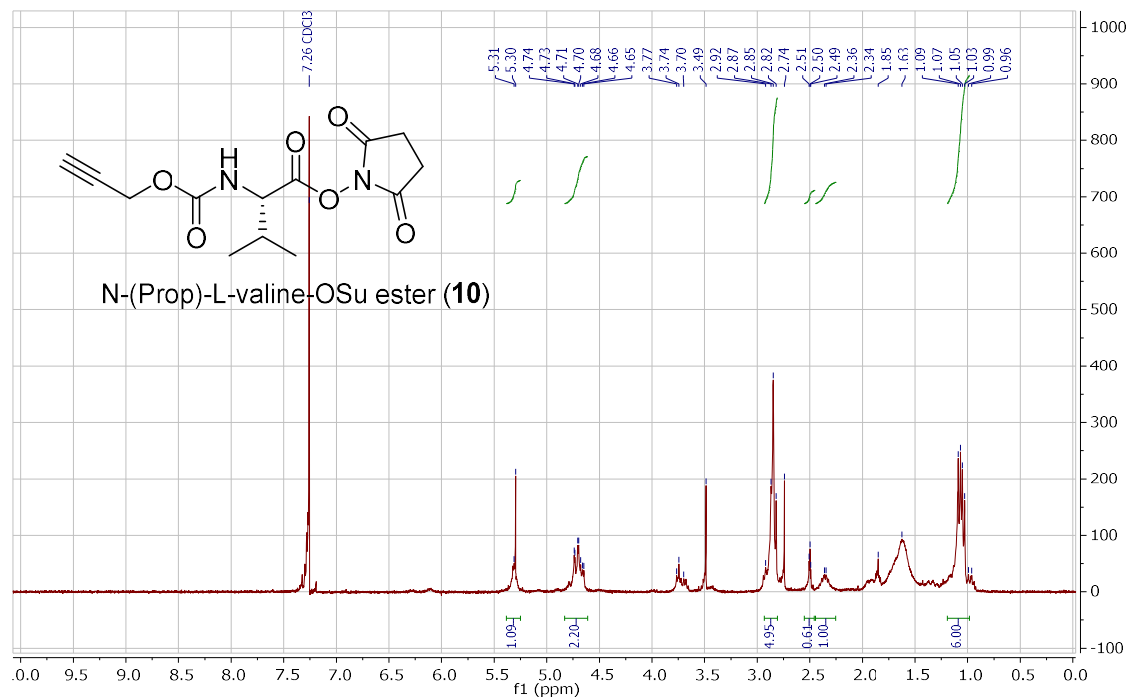


Figure 122 ^1H NMR spectrum of N-(Prop)-L-valine-OSu ester (10)

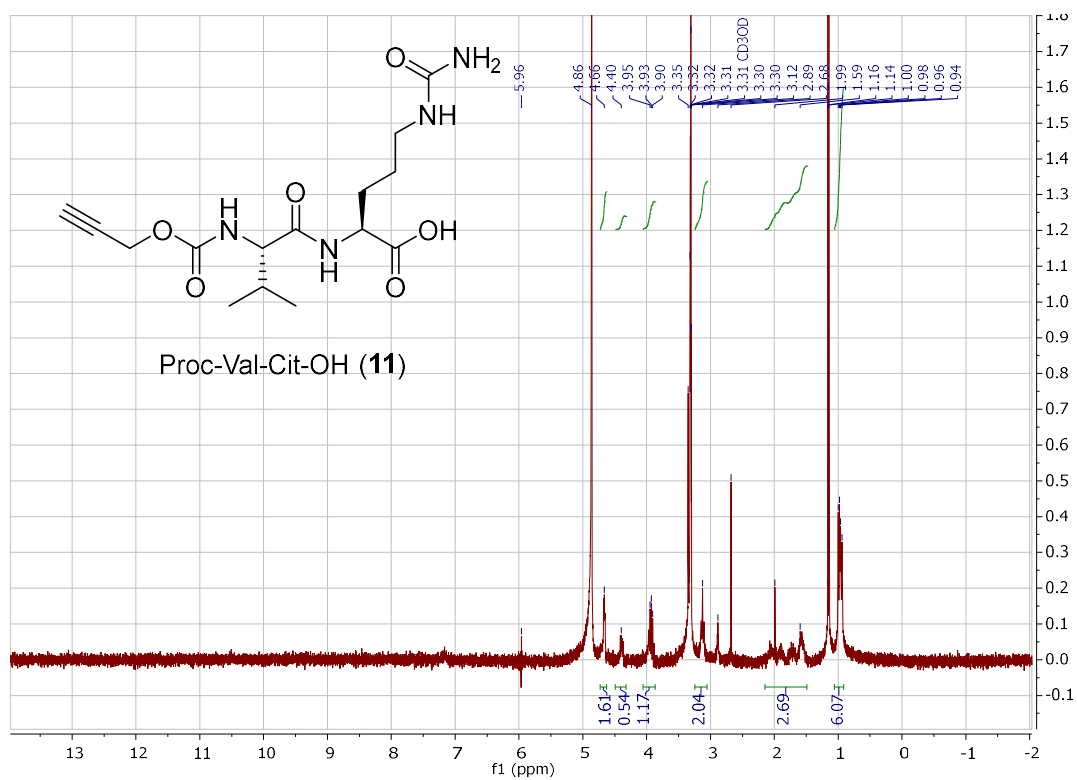


Figure 123 ¹H NMR spectrum of Proc-Val-Cit-OH (11)

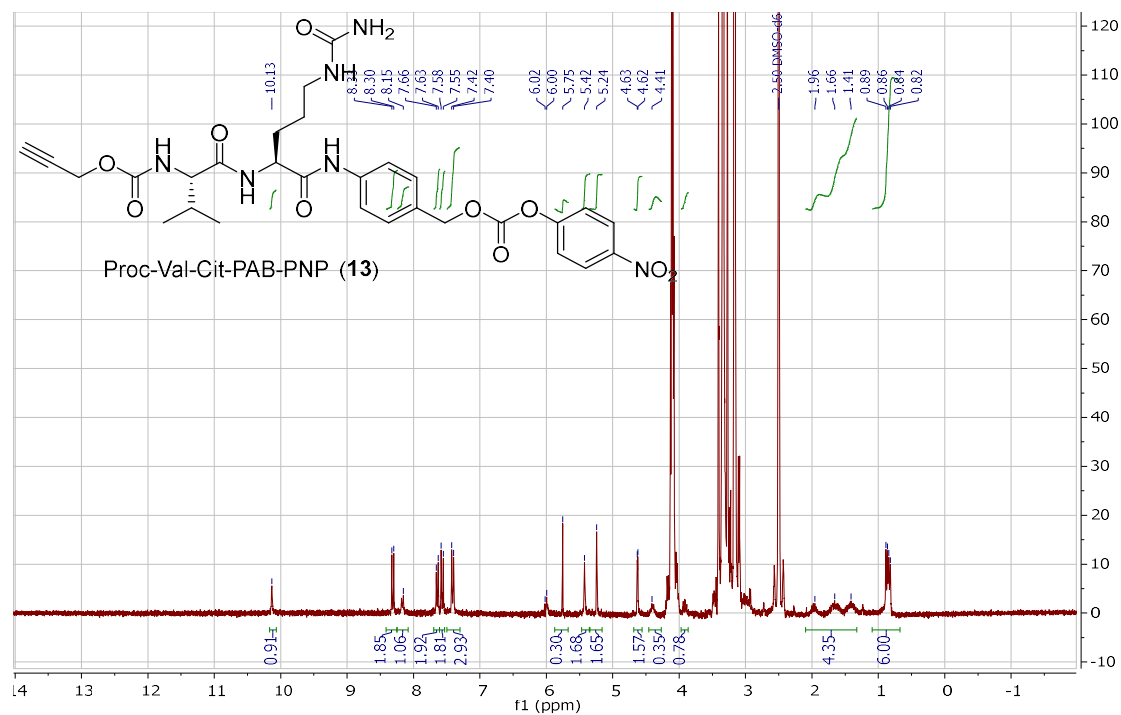


Figure 124 ¹H NMR spectrum of Proc-Val-Cit-PAB-PPN (13)

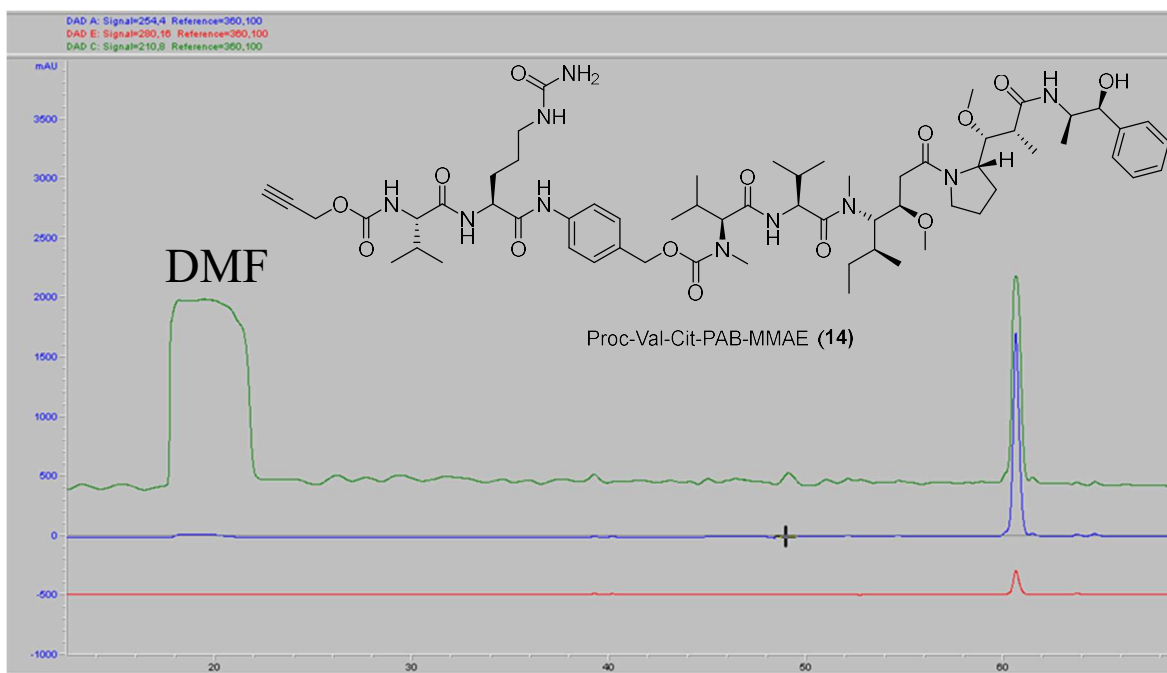


Figure 125 HPLC chromatogram of Proc-Val-Cit-PAB-MMAE (14)

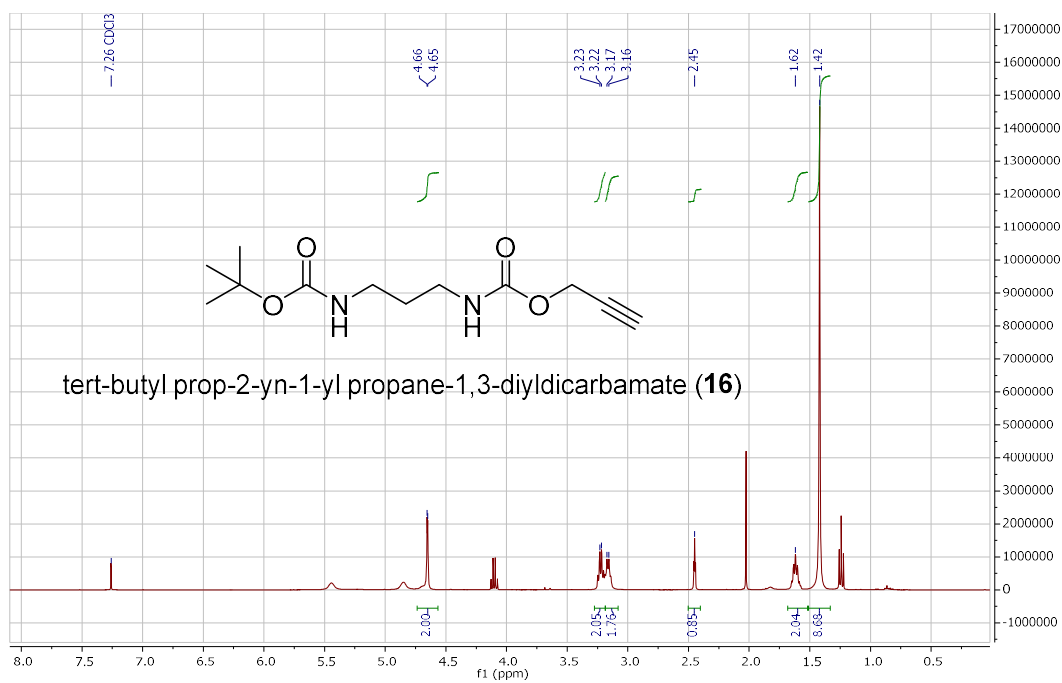


Figure 126 ¹H NMR spectrum of tert-butyl prop-2-yn-1-yl propane-1,3-diylidicarbamate (16)

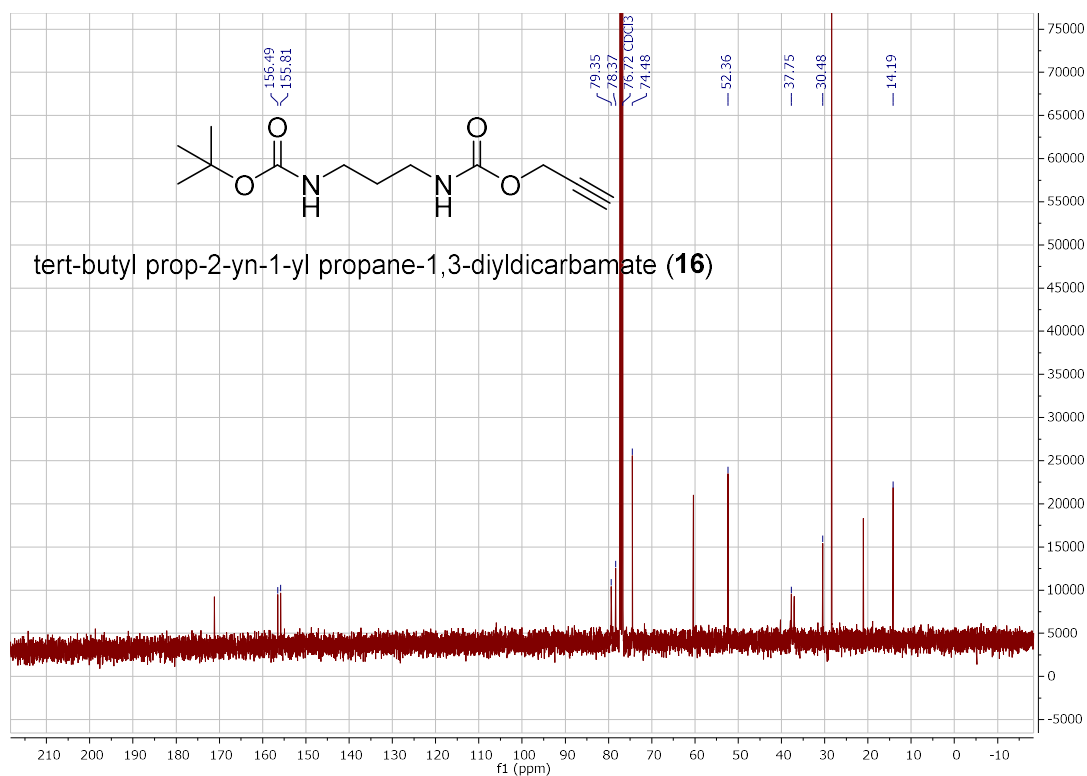


Figure 127 ¹³C NMR spectrum of tert-butyl prop-2-yn-1-yl propane-1,3-diylidicarbamate (16)

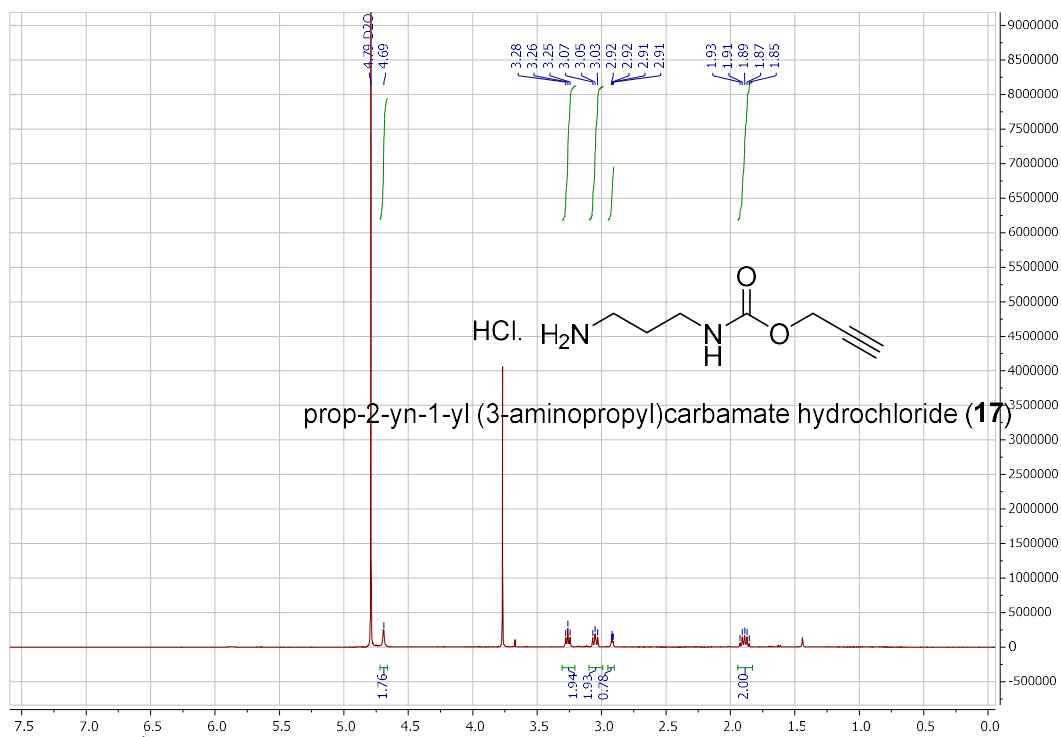


Figure 128 ¹H NMR spectrum of prop-2-yn-1-yl (3-aminopropyl)carbamate hydrochloride (17)

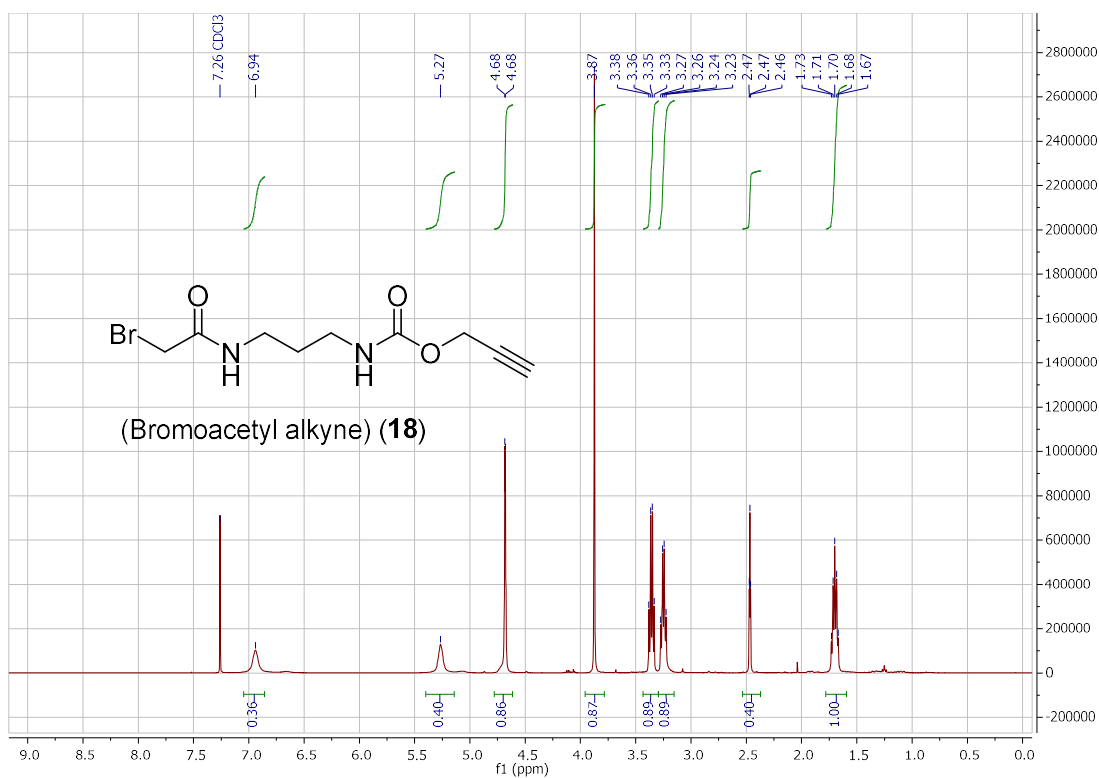


Figure 129 ^1H NMR spectrum of bromoacetyl alkyne (**18**)

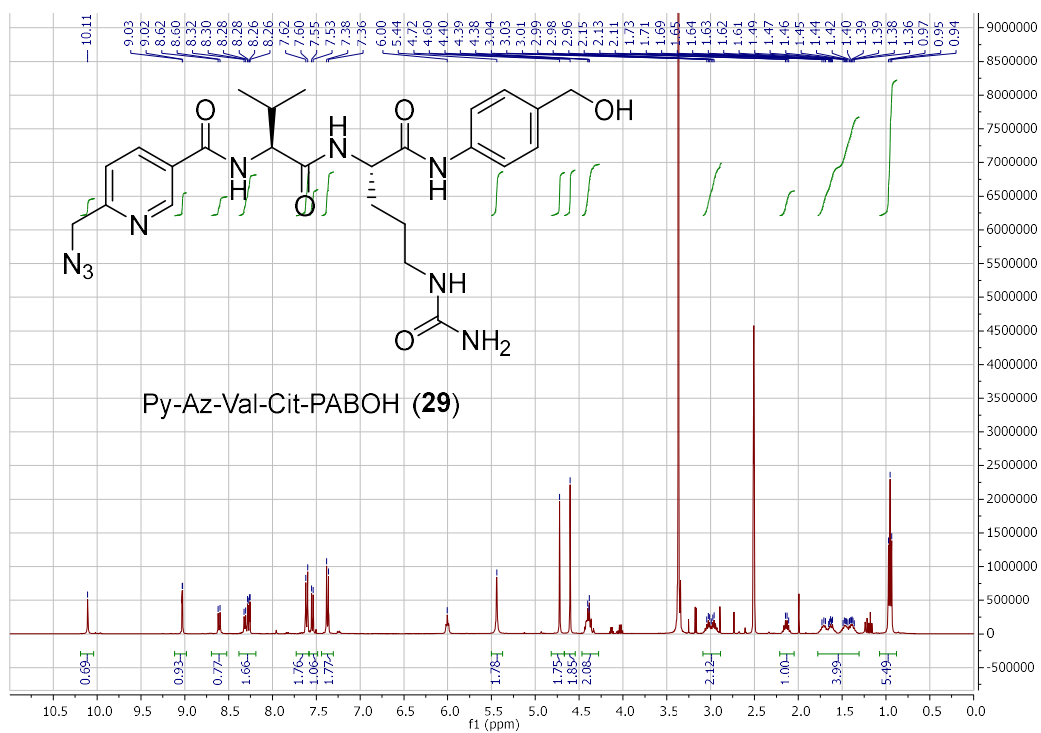


Figure 130 ^1H NMR spectrum of Py-Az-Val-Cit-PABOH (**29**)

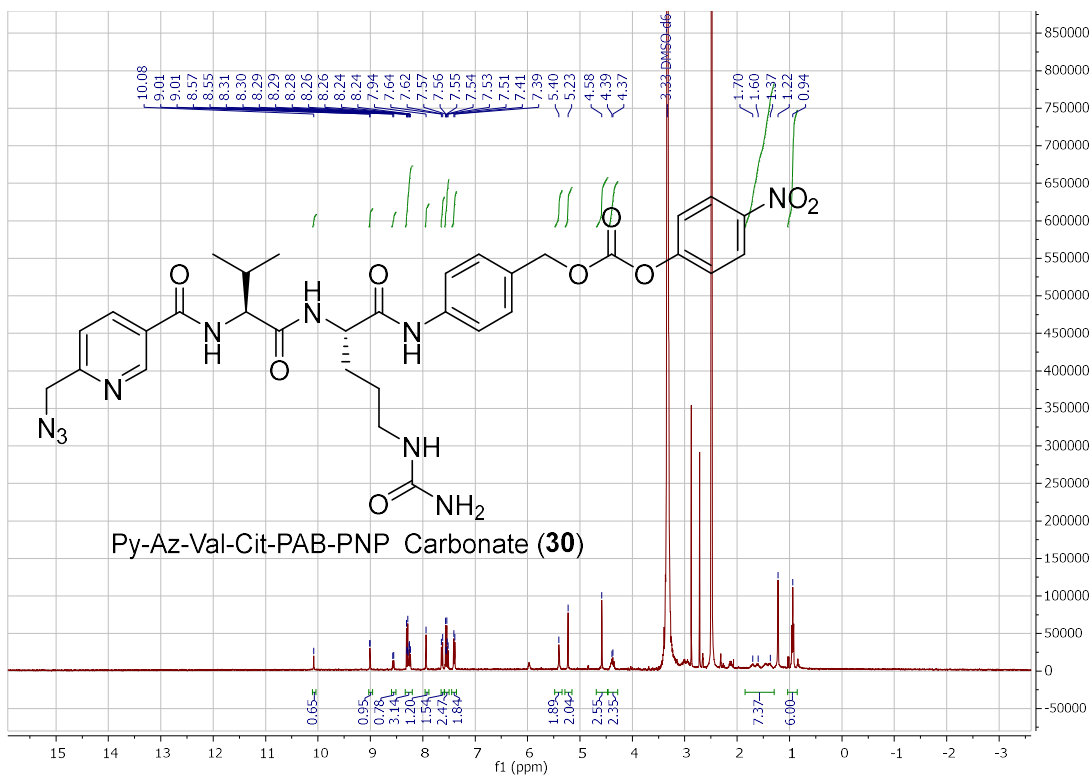


Figure 131 ^1H NMR spectrum of Py-Az-Val-Cit-PAB-PNP carbonate (30)

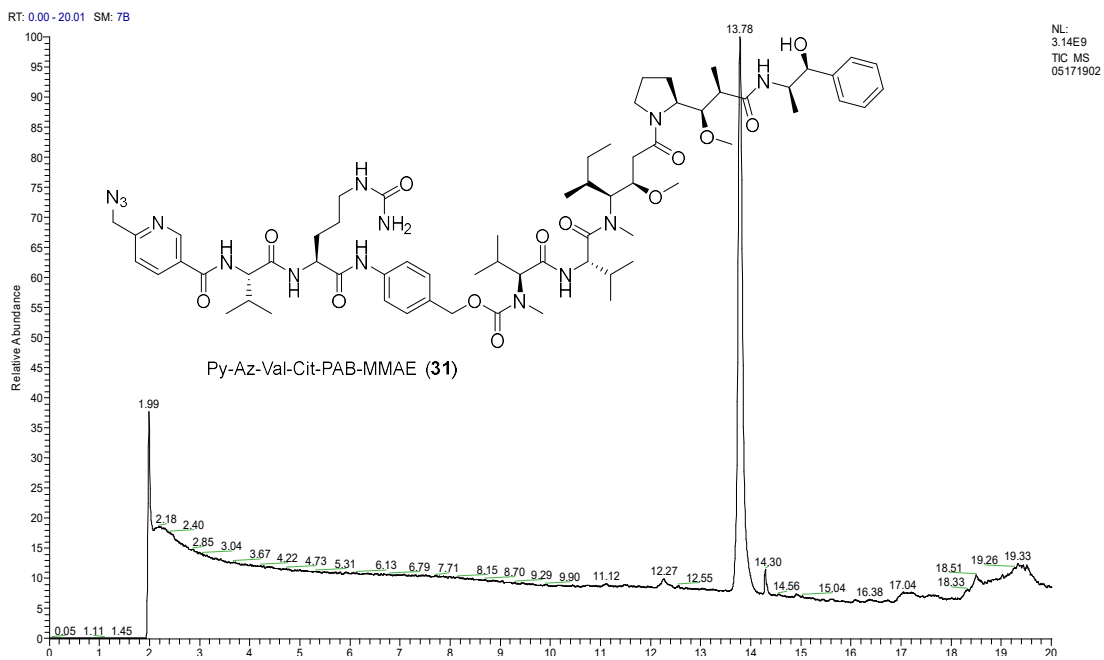


Figure 132 LC-MS chromatogram of Py-Az-Val-Cit-PAB-MMAE (31)

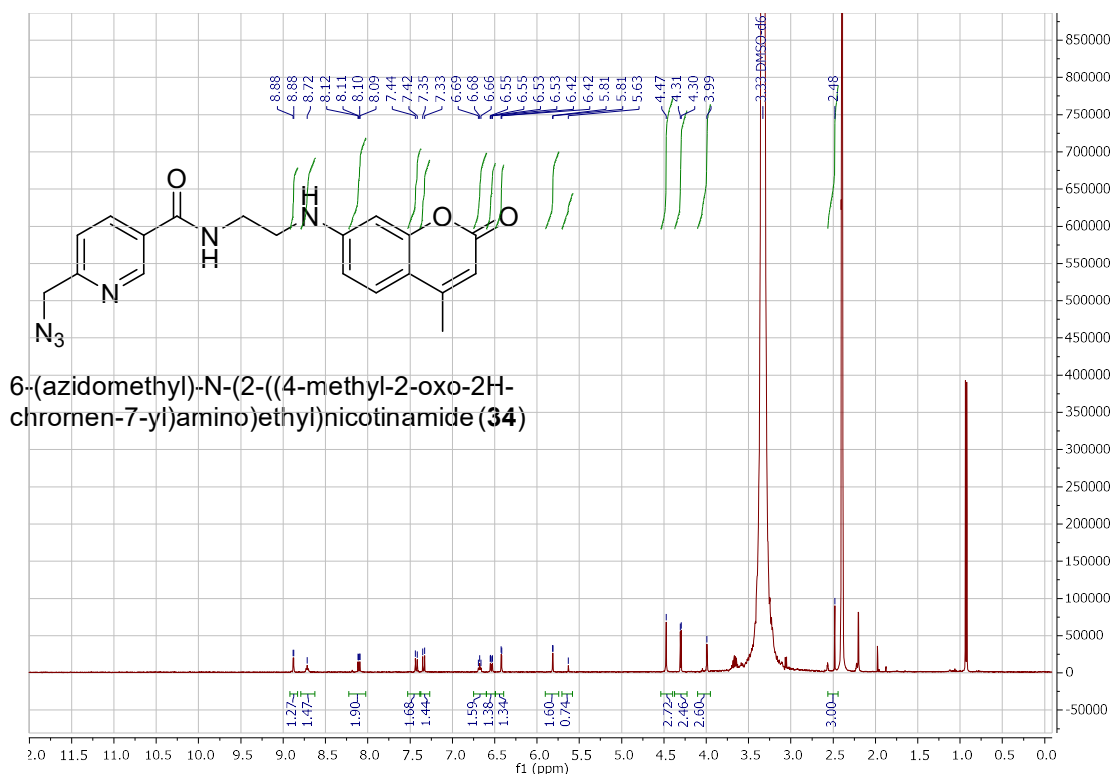


Figure 133 ^1H NMR spectrum of 6-(azidomethyl)-N-(2-((4-methyl-2-oxo-2H-chromen-7-yl)amino)ethyl)nicotinamide (**34**)

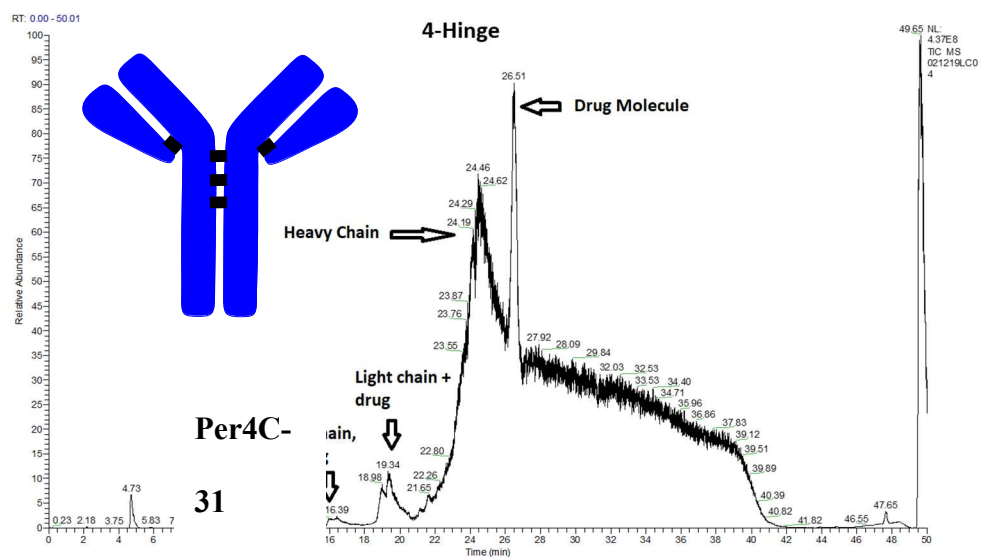


Figure 134 LC-MS chromatogram of conjugated Per4C-31

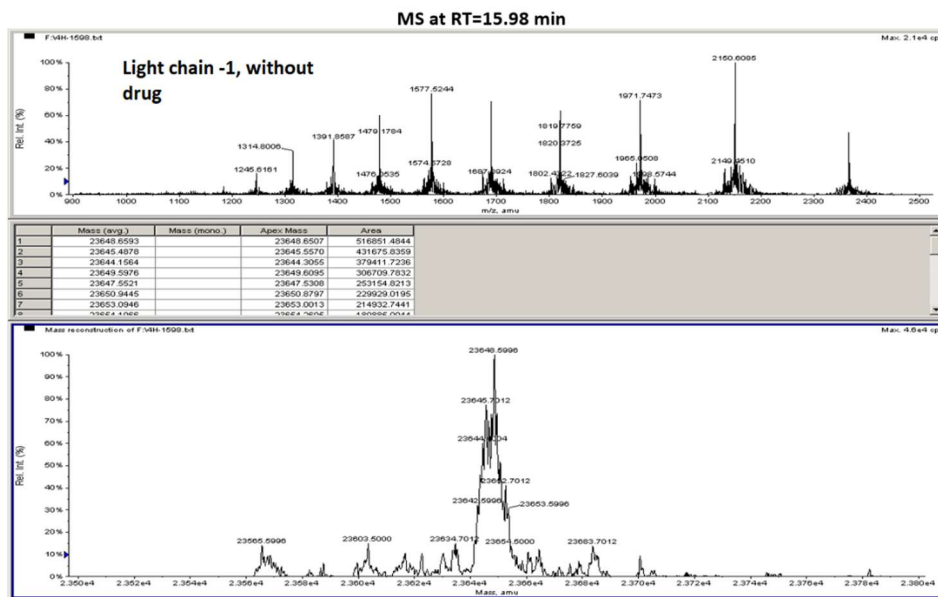


Figure 135 Deconvolution results corresponding to peak at 15.98 min (light chain that was conjugated with 7 only) in Figure 134

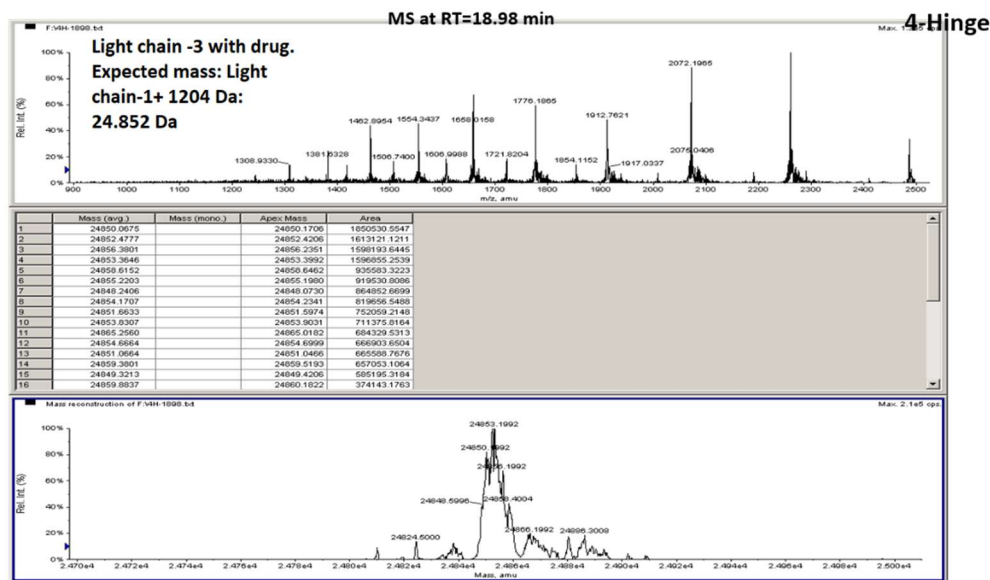


Figure 136 Deconvolution results corresponding to peak at 18.98 min (light chain reacted with 7 and 14) in Figure 134

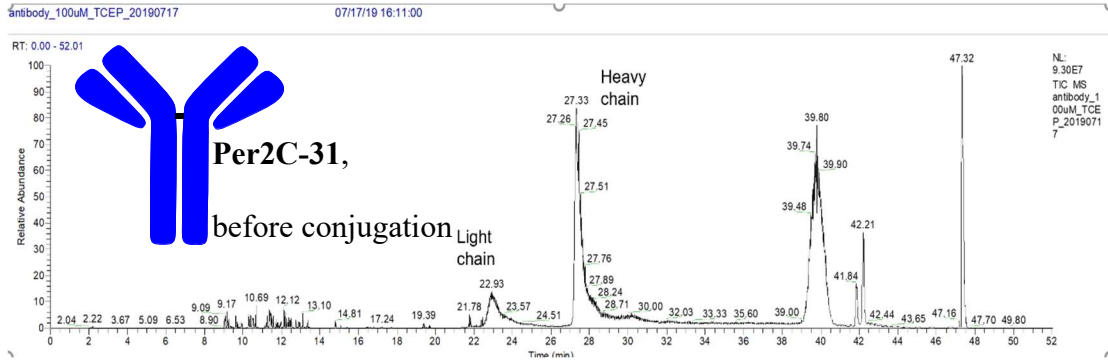


Figure 137 LC-MS chromatogram of Per2C-31 before conjugation

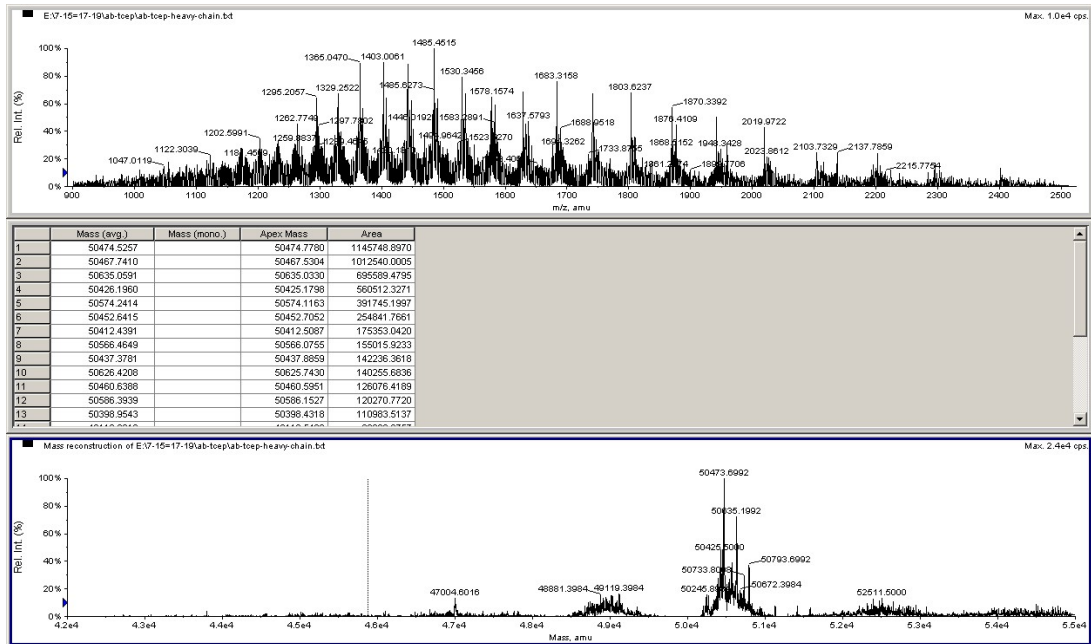


Figure 138 Deconvolution of the signal at peak 27.33 min in Figure 137

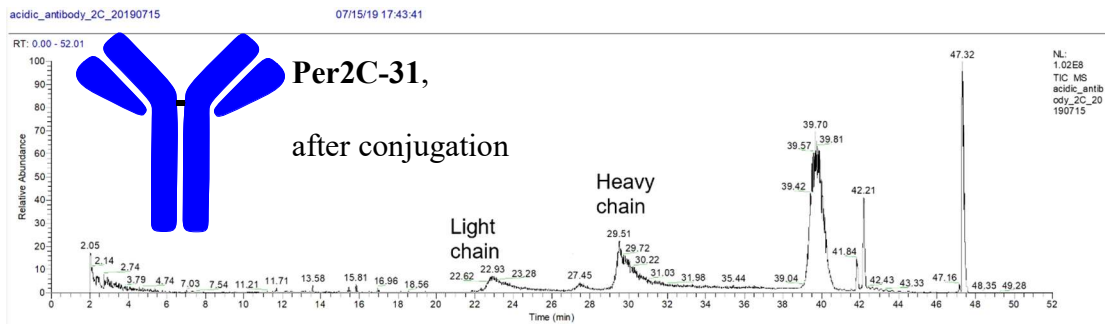


Figure 139 LC-MS chromatogram of Per2C-31 after conjugation with 31

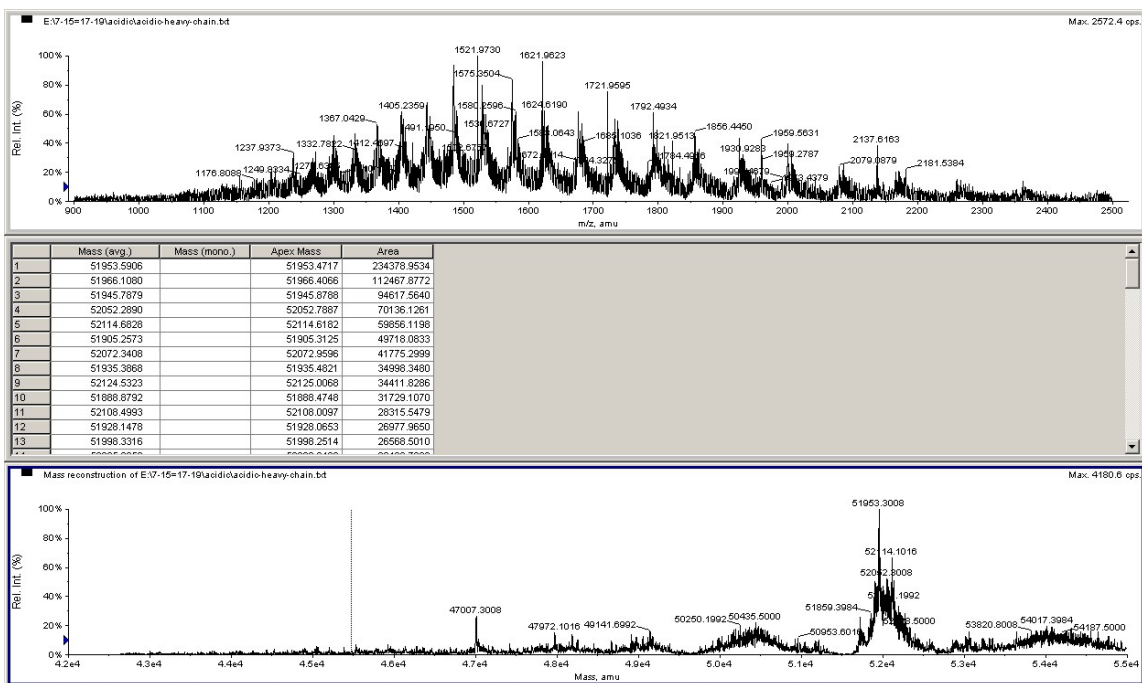


Figure 140 Deconvolution of mass signals at peak 29.51 mins in Figure 139

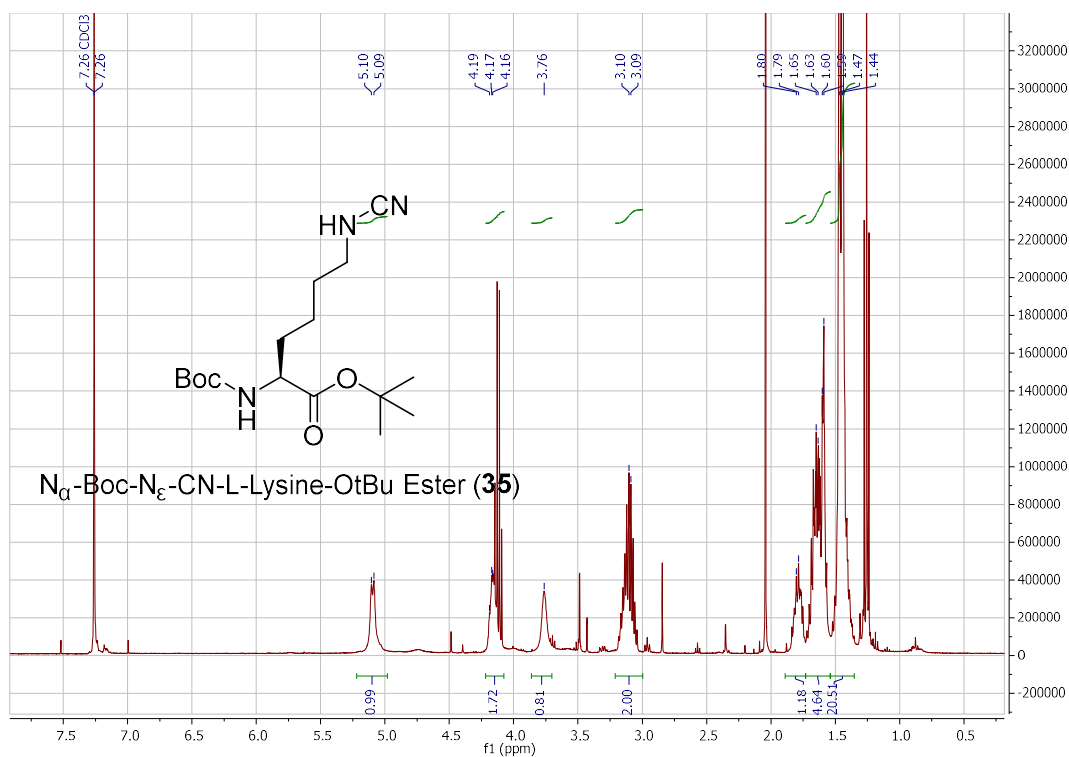


Figure 141 ^1H NMR spectrum of N_{α} -Boc- N_{ϵ} -CN-L-Lysine-OtBu Ester (35)

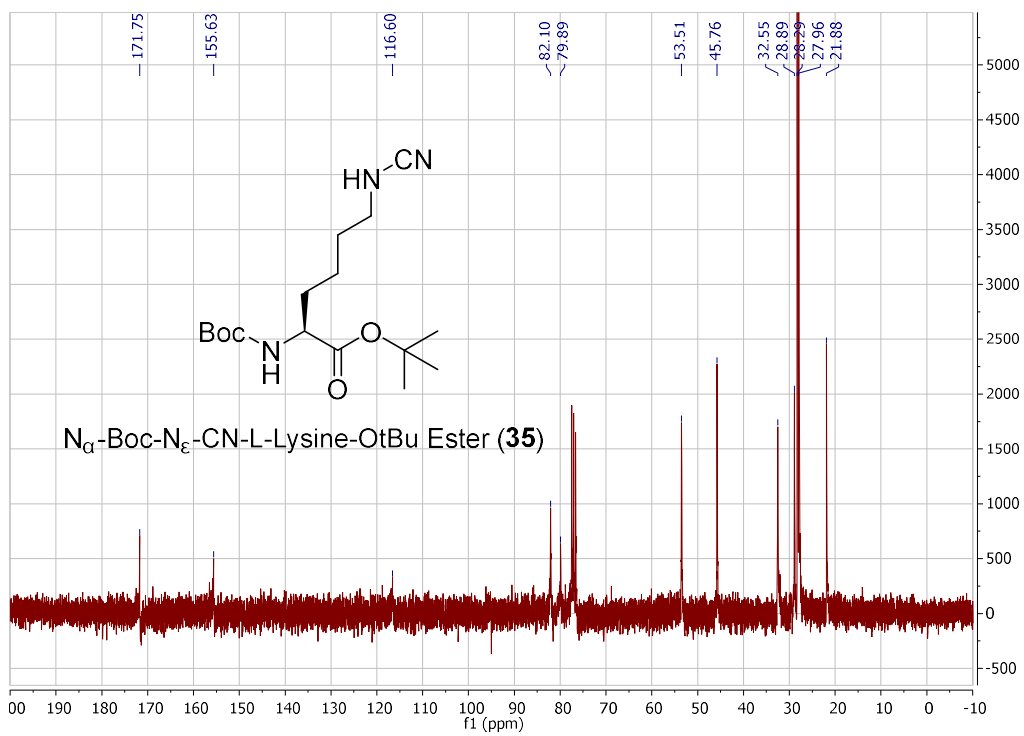


Figure 142 ^{13}C NMR spectrum of N_{α} -Boc- N_{ϵ} -CN-L-Lysine-OtBu Ester (35)

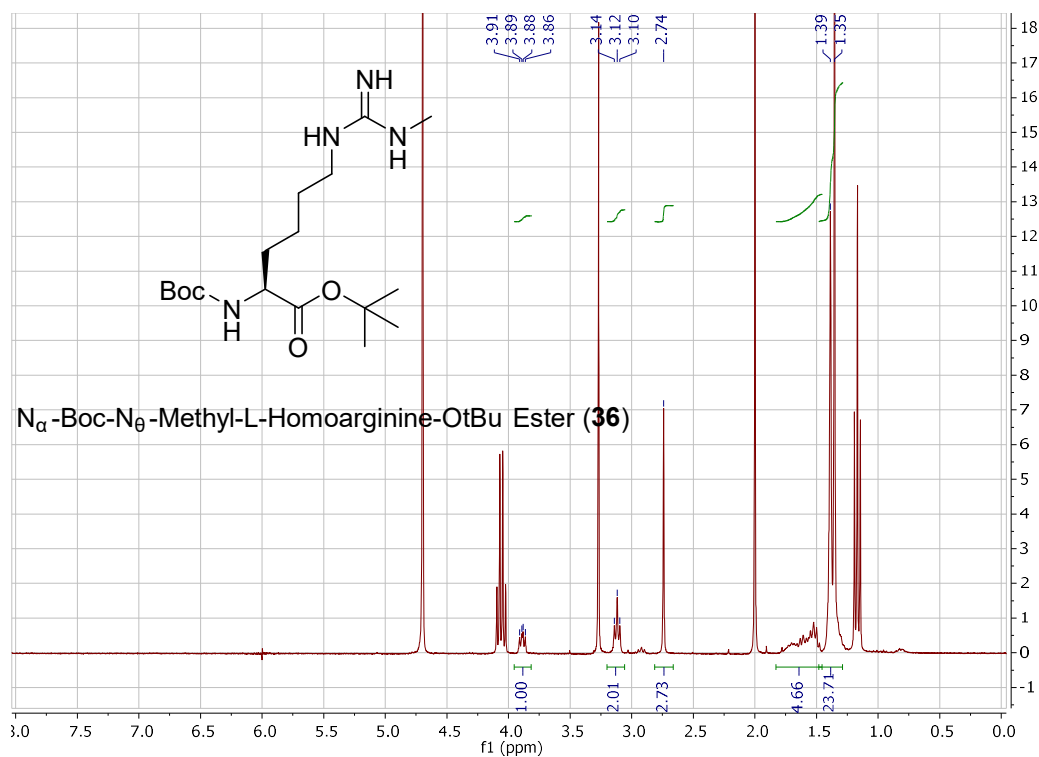


Figure 143 N_α-Boc-N_θ-Methyl-L-Homoarginine-OtBu Ester (36)

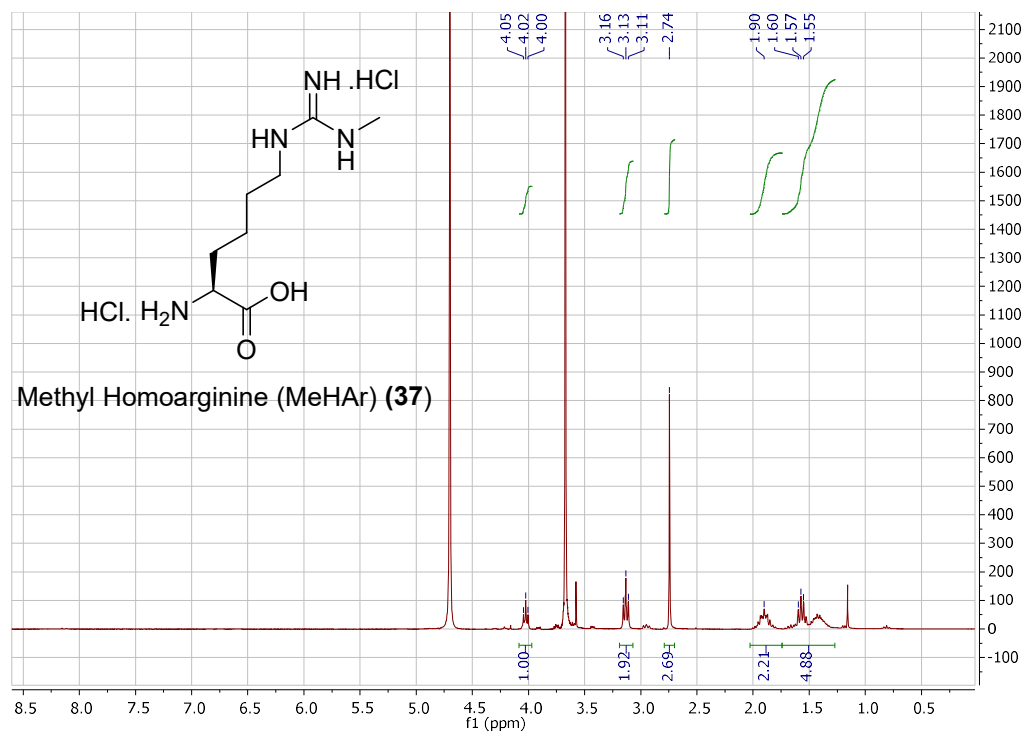


Figure 144 ¹H NMR spectrum of methyl homoarginine (MeHAr) (37)

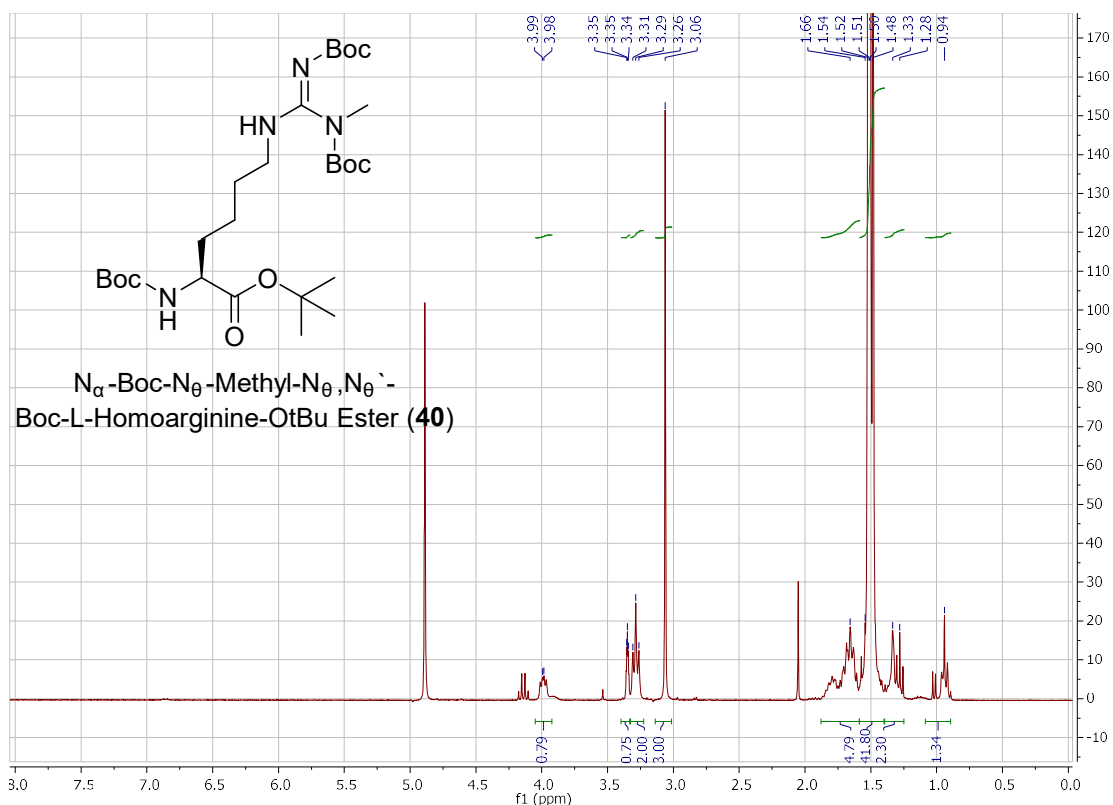


Figure 145 ^1H NMR spectrum of N_{α} -Boc- N_{θ} -Methyl- N_{θ},N_{θ}' -Boc-L-Homoarginine-OtBu Ester (40)

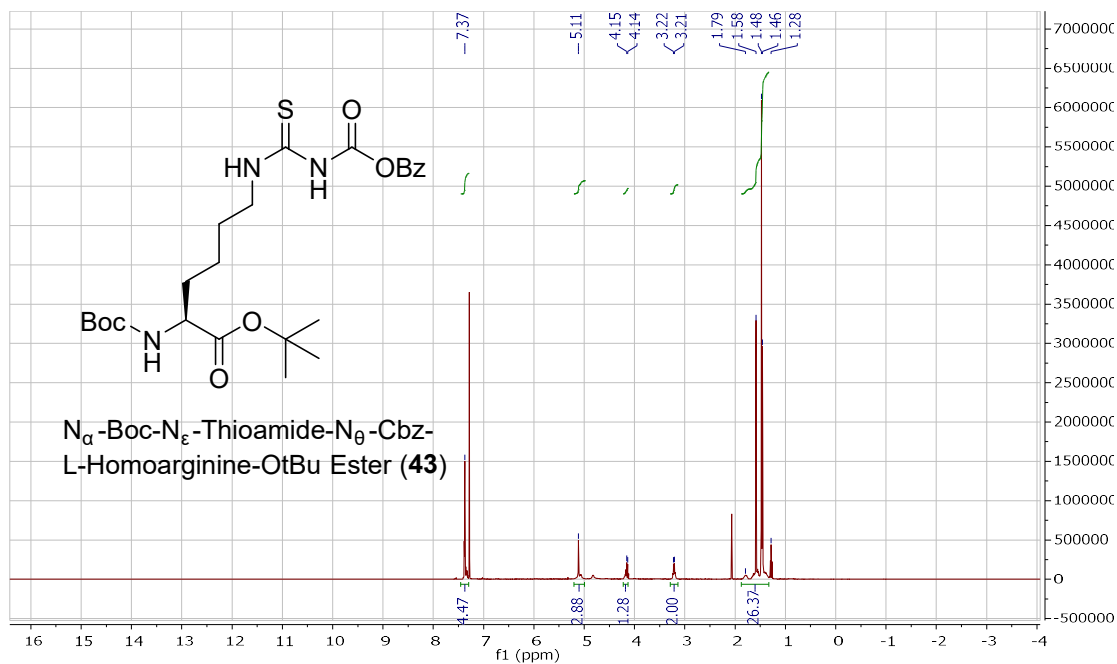


Figure 146 ^1H NMR spectrum of N_{α} -Boc- N_{ϵ} -Thioamide- N_{θ} -Cbz-L-Homoarginine-OtBu Ester (43)

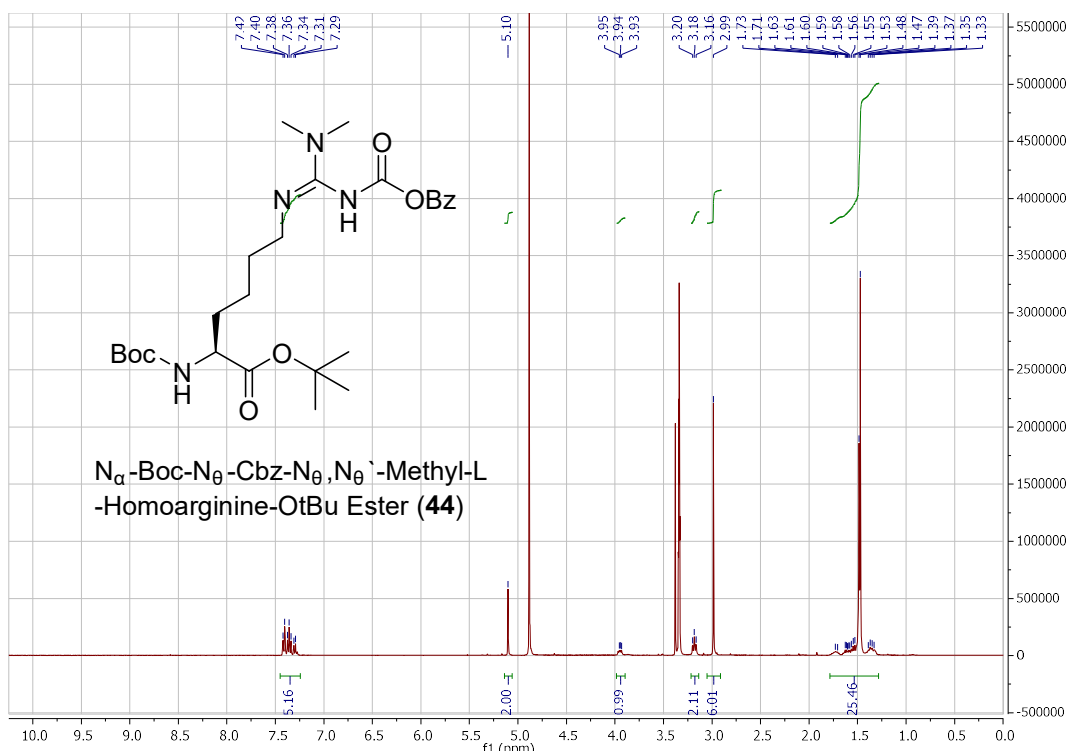


Figure 147 ^1H NMR spectrum of N_{α} -Boc- N_{θ} -Cbz- N_{θ},N_{θ}' -Methyl-L-Homoarginine-OtBu Ester (44)

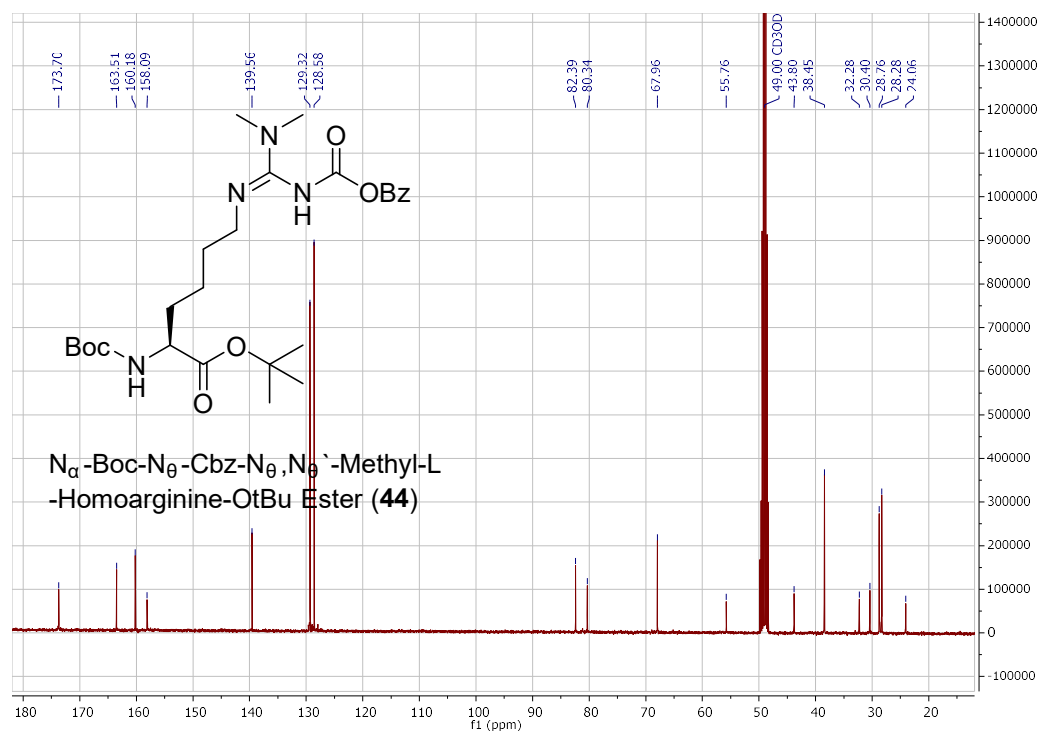


Figure 148 ^{13}C NMR spectrum of N_{α} -Boc- N_{θ} -Cbz- N_{θ},N_{θ}' -Methyl-L-Homoarginine-OtBu Ester (44)

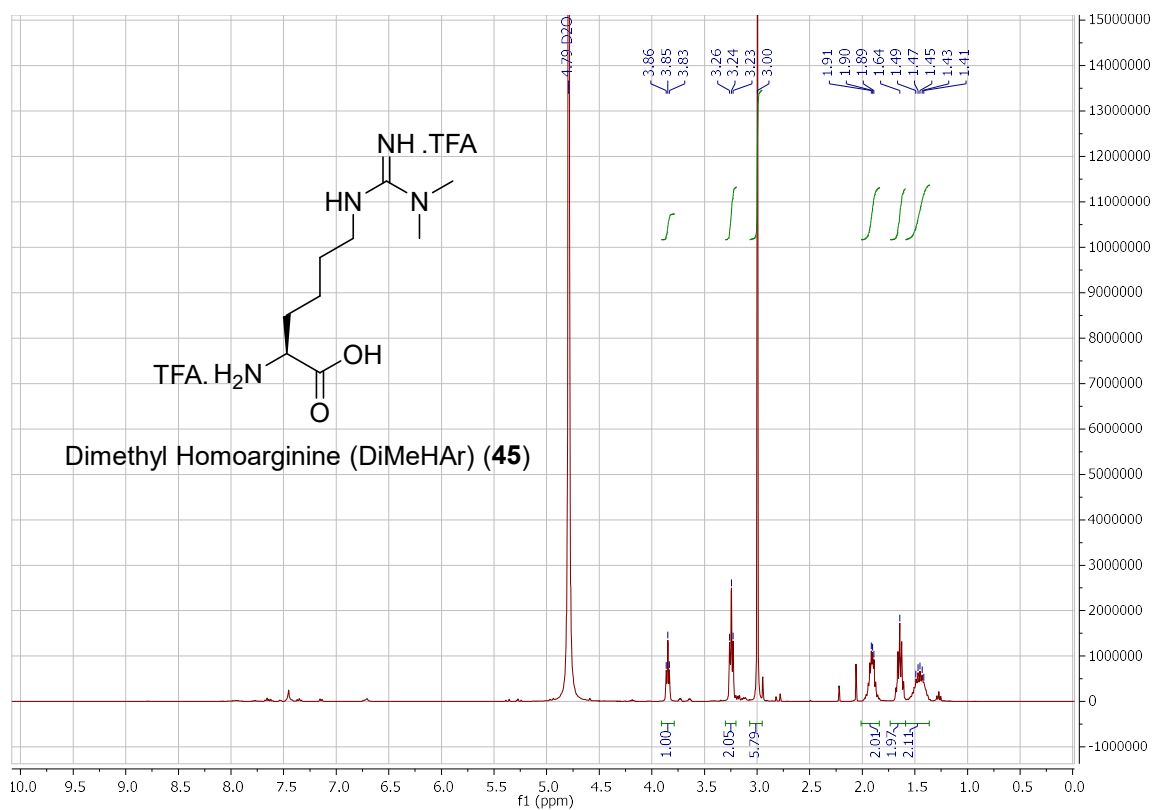


Figure 149 ^1H NMR spectrum of dimethyl homoarginine (DiMeHAr) (45)


APPENDIX B

LIST OF ABBREVIATIONS

aminoacyl-tRNA synthetase	aaRS
angiotensin-converting enzyme 2	ACE2
acetyl-L-lysine	AcK
acetonitrile	ACN
antibody-drug conjugate	ADC
asymmetric dimethyl-L-arginine	ADMA
adenylyl-imidodiphosphate	AMP-PNP
3-azido-7-hydroxycoumarin	AzCou
bicyclononyne	BCN
BCN conjugated main protease dimer	BCN-(M ^{PRO}) ₂
(tert-Butoxycarbonyl)	Boc
N α -(tert-Butoxycarbonyl)-L-lysine	BocK
carboxybenzyl	CBZ
cyano	CN

cytopathogenic effect	CPE
dalton	Da
dibenzocyclooctyne	DBCO
DBCO conjugated main protease dimer	DBCO-(M ^{pro}) ₂
N,N'-Dicyclohexylcarbodiimide	DCC
dichloromethane	DCM
Dimethyl-L-homoarginine	DiMeHAr
N,N-Dimethylformamide	DMF
dimethyl sulfoxide	DMSO
half maximal effective concentration	EC50
1-Ethyl-3-(3-dimethylaminopropyl)carbodiimide	EDCI
Ethylenediaminetetraacetic acid	EDTA
Eagle's minimal essential medium	EMEM
electrospray ionization-mass spectrometry	ESI-MS
7-ethynylcoumarin	EtCou
ethyl acetate	EtOAc

fetal bovine serum	FBS
<i>ortho</i> -fluoro- <i>para</i> -chlorocinnamyl	FCC
L-homoarginine	HAr
hydrogen chloride	HCl
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid	HEPES
high pressure liquid chromatography	HPLC
half maximal inhibitory concentration	IC50
Immunoglobulin	IgG
<i>iso</i> -propanol	<i>i</i> -PrOH
Isopropyl β - d-1-thiogalactopyranoside	IPTG
kilodalton	kDa
keto-lysine	KetoK
Liquid chromatography–mass spectrometry	LC-MS
Methyl-L-homoarginine	MeHAr
methanol	MeOH
monomethyl-L-arginine	MMA

main protease inhibitor	MPI
main protease	M ^{pro}
messenger RNA	mRNA
sodium hydride	NaH
sodium bicarbonate	NaHCO ₃
nuclear magnetic resonance	NMR
<i>ortho</i> -chloro-L-phenylalanine	 -ClF
<i>ortho</i> -chloro-L-phenylalaninyl tRNA synthetase	<i>o</i> -ClFRS
β -(S-2-oxopyrrolidin-3-yl)-alaninal	Opal
para-azido-L-phenylalanine	<i>p</i> -Azf
phosphate-buffered saline	PBS
post-translational modification	PTM
propargyl-L-lysine	PrK
protein data bank	PDB
pyridine azide	PyAz
pyrrolysine	Pyl

pyrrolysyl tRNA synthetase	PyIRS
round-bottom	RB
ribonucleic acid	RNA
symmetric dimethyl-L-arginine	SDMA
sodium dodecyl sulphate–polyacrylamide gel electrophoresis	SDS-PAGE
Small Ubiquitin-like Modifier	SUMO
tert-butyl	tBu
trifluoroacetic acid	TFA
tetrahydrofuran	THF
thin layer chromatography	TLC
tris(hydroxymethyl)aminomethane	Tris
transfer RNA	tRNA