November 2023

REACTIVE CHEMISTRIES FOR PROTEIN LABELING, DEGRADATION, AND STIMULI RESPONSIVE DELIVERY

Myrat Kurbanov

University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/dissertations_2

Part of the Amino Acids, Peptides, and Proteins Commons, Analytical Chemistry Commons, Bioimaging and Biomedical Optics Commons, Biomaterials Commons, Biomedical Devices and Instrumentation Commons, Materials Chemistry Commons, Organic Chemistry Commons, Other Chemicals and Drugs Commons, Polymer Chemistry Commons, and the Structural Biology Commons

Recommended Citation
https://doi.org/10.7275/35806620 https://scholarworks.umass.edu/dissertations_2/3006

This Open Access Dissertation is brought to you for free and open access by the Dissertations and Theses at ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.
REACTIVE CHEMISTRIES FOR PROTEIN LABELING, DEGRADATION, AND STIMULI RESPONSIVE DELIVERY

A Dissertation Presented

by

MYRAT KURBANOV

Submitted to the graduate school of the University of Massachusetts, Amherst In partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2023

Chemistry
REACTIVE CHEMISTRIES FOR PROTEIN LABELING, DEGRADATION, AND STIMULI RESPONSIVE DELIVERY

A Dissertation Presented

by

MYRAT KURBANOV

Approved as to style and content by:

_______________________________
Sankaran Thayumanavan, Chair

_______________________________
Richard Vachet, Member

_______________________________
Mingxu You, Member

_______________________________
Scott Garman, Member

_______________________________
Ricardo Metz, Department Head,
Department of Chemistry
DEDICATION

To my family
ACKNOWLEDGEMENTS

Graduate school was the most arduous journey in my life, and I have learned many lessons through it all. But I would not have made it without professional and social support from my advisors, colleagues, friends, and family. Their support helped me through my journey and inspired me to give back and be a better person.

First, I would like to thank my Advisor, Professor Sankaran Thayumanavan (Thai). Prof. Thai always thinks of new ideas: while walking, in mid-discussions, or conversing. Whenever he learns something new, he always starts thinking of ways to use it and what problems we could solve with it; “when life gives you lemons…” as he would say. That is something that I always appreciated about his mentality, and I tried to emulate his way of thinking in my research. Among the very long list of things I have learned from him are critical thinking, scientific communication, and hard work. He is one of the few people I know who has blurred the lines between work and play. I thank my committee members, Professors Richard Vachet, Mingxu You, and Scott Garman, for their insight, critiques, and input in guiding me through this journey. I had the unique pleasure of collaborating with Prof. Vachet. He always manages to see things from a different angle and gives excellent suggestions, for which I am always grateful. Also, I would like to thank Professor DV for briefly serving on my committee.

This work would not have been possible without notable collaborations with Zachary Kirsch, Jithu Krishna, Ranit Dutta, Yasin Alp, Anirudh Devarajan, Max Bowen, Bin Liu, Wardah Ejaz, Shuai Gong, Mine Canakci and Francesca Anson. I am also thankful to Prof. Cathal Kearney, with whom we have started the hydrogel project. We stand on the shoulders of giants who came before us. A good part of this work is the continuation of the work begun by Dr. Jiaming Zhuang and Bo Zhao. I am grateful to IALS MS and Light Microscopy core facility staff Dr. Stephen
Eyles, Dr. Cedric Bobst, Dr. James Chambers, and Maaya Ikeda for their training, guidance, and hands-on deck assistance. Many thanks to UMass Chemistry staff: JMS, Rebecca David, Vicki Hubby, Kris Knight, Ryan Feyrer, Glenda Pons, and especially to Dennis Glick and Robert Sabola.

I am eternally grateful to Thameez Koyasseril Yehiya. I learned all the fundamentals by working with him side by side and using his fume hood until I knew enough to carry out my experiments independently in my own fume hood. I am thankful to Vikash Kumar and Khushboo Singh for all their guidance. These three remain great mentors and friends from the day I met them. I am grateful for Karen Hakala, who has been the pillar of the Thai lab and supported everyone and myself with compassion going well beyond her responsibilities. And thanks to Cheng Song for accepting Karen’s responsibilities. Thanks to my batchmates Peidong Wu and Emil Samson for beginning and sharing this journey with me.

I had great support from all Thai group members and made great friends with them all, so I thank them all. Special thanks to Jithu Krishna, Ranit Dutta, Hung-Hsun Lu, Yasin Alp, Anirudh Devarajan, Bin Liu, Jiaming Zhuang, Kingshuk Dutta, Ziwen Jiang, Wardah Ejaz, Francesca Anson, Hongxu Liu, Ann Fernandez, Jayashree Bhagabati, Theo Prachyathipsakul, Stephanie Le and Uyen Huynh for all the discussions and help. Last but not least, huge thanks to Pintu Kanjilal and Ritam Das for all their help, discussions, guidance, and support, especially towards the end and friendship beyond the lab.

Outside the lab, I would like to thank my high school senior friend Azady Pirhanov who has supported and helped me with everything during this journey. Thanks to Sylapberdi Bayjanov for his long-distance support from continents away. Finally, I would like to thank my family for their unconditional love and support, which made this journey possible.
ABSTRACT

REACTIVE CHEMISTRIES FOR PROTEIN LABELING,
DEGRADATION, AND STIMULI RESPONSIVE DELIVERY

SEPTEMBER 2023

MYRAT KURBANOV
B.S., KOREA ADVANCED INSTITUTE OF SCIENCE & TECHNOLOGY (KAIST)
Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST
Directed by: Professor Sankaran Thayumanavan

Reactive chemistries for protein chemical modification play an instrumental role in chemical biology, proteomics, and therapeutics. Depending on the application, the selectivity of these modifications can range from precise modification of an amino acid sequence by genetic manipulation of protein expression machinery to a stochastic modification of lysine residues on the protein surface. Ligand-Directed (LD) chemistry is one of the few methods for targeted modification of endogenous proteins without genetic engineering. However, current LD strategies are limited by stringent amino acid selectivity. To bridge this gap, this thesis focuses on the development of highly reactive LD Triggerable Michael Acceptors (LD-TMAcs) that feature rapid protein labeling.

Unlike previous LD approaches, the unique reactivity of LD-TMAcs enables multiple modifications on a single target protein. This capability is attributed to the tunable reactivity of TMAcs that enable the labeling of several amino acid functionalities via a binding-induced
increase in local concentration while remaining fully dormant in the absence of protein binding. We demonstrate the utility of this method by selectively labeling membrane-bound carbonic anhydrase XII in live cells. Then, by using our LD-TMAc platform, we have developed target selective covalent Hydrophobic Tagging (HyT) probes. By target-selectively labeling the membrane CAXII with multiple hydrophobic adamantane tags, we demonstrate ~75% CAXII degradation in MCF7 cells in 30 minutes. We envision that the unique features of LD-TMAcs will find use in a range of applications from target identification, investigation of binding/allosteric sites, studies of membrane proteins, and targeted protein degradation. Finally, this thesis also involves the development of reactive polymers for covalent encapsulation and photo-responsive release of proteins and small molecule therapeutics in hydrogel drug depots. We envision the ability of our polymers to directly conjugate natural proteins without genetic or enzymatic engineering and release the encapsulated cargo in a photo-responsive and traceless manner that will find use in drug-encapsulated hydrogel systems.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>v</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xiii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xiv</td>
</tr>
<tr>
<td>LIST OF SCHEMES</td>
<td>xxiv</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1. INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Background</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Selectivity in protein labeling</td>
<td>2</td>
</tr>
<tr>
<td>1.2.1 Genetic engineering approaches</td>
<td>2</td>
</tr>
<tr>
<td>1.2.2 Amino acid selective approaches</td>
<td>5</td>
</tr>
<tr>
<td>1.2.3 Affinity-based and Proximity-based labeling</td>
<td>6</td>
</tr>
<tr>
<td>1.3 Applications of protein labeling</td>
<td>8</td>
</tr>
<tr>
<td>1.3.1 Proteomics and Drug Discovery</td>
<td>8</td>
</tr>
<tr>
<td>1.3.2 Structural analysis</td>
<td>11</td>
</tr>
<tr>
<td>1.3.3 Targeted protein degradation</td>
<td>13</td>
</tr>
<tr>
<td>1.4 Protein modification chemistries in bioconjugation and drug delivery</td>
<td>14</td>
</tr>
<tr>
<td>1.4.1 Bioconjugation</td>
<td>14</td>
</tr>
<tr>
<td>1.4.2 Drug delivery</td>
<td>16</td>
</tr>
<tr>
<td>1.5 Summary and Overview of the Thesis</td>
<td>17</td>
</tr>
<tr>
<td>1.6 References</td>
<td>19</td>
</tr>
</tbody>
</table>

## 2. MULTISITE LABELING OF PROTEINS USING THE LIGAND-DIRECTED REACTIVITY OF
TRIGGERABLE MICHAEL ACCEPTORS

2.1 Introduction.................................................................................................................. 33
2.2 Results and discussion ................................................................................................. 35
  2.2.1 Molecular design and characterization of carbonic anhydrase (CA) labeling by LD-TMAc reagents .................................................................................................................. 35
  2.2.2 LD-TMAc regiospecific labeling around the ligand binding site .......................... 40
  2.2.3 Target specific labeling of CA in whole cell lysate ................................................. 45
  2.2.4 Live cell labeling of membrane CA ........................................................................ 46
2.3 Conclusions ................................................................................................................... 47
2.4 Acknowledgements ...................................................................................................... 48
2.5 Data availability .......................................................................................................... 48
2.6 Supplementary Information ......................................................................................... 48
  2.6.1 Materials and Instrumentation ............................................................................. 48
  2.6.2 Covalent Labeling and Proteolytic Digestion for LC-MS/MS................................. 49
  2.6.3 Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS) ..................... 49
  2.6.4 Synthesis of compounds 1 to 6 ............................................................................. 51
  2.6.5 Synthesis of compound 2 ................................................................................... 56
  2.6.6 Synthesis of compound 3 ................................................................................... 58
  2.6.7 Synthesis of compound 4 ................................................................................... 60
  2.6.8 Supplementary figures ......................................................................................... 63
  2.6.9 Tandem Mass Spectra of the peptide modifications ............................................... 68
  2.6.10 Nuclear Magnetic Resonance (NMR) of the final compounds ............................ 76
2.7 References .................................................................................................................... 82

3. LIGAND-DIRECTED HYDROPHOBIC TAGGING OF MEMBRANE PROTEINS FOR TARGETED PROTEIN DEGRADATION ...................................................... 88
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.1: LD-TMAc targeted residues on bCAII</td>
<td>42</td>
</tr>
</tbody>
</table>

Table 2.1: LD-TMAc targeted residues on bCAII
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1: Applications of chemical modifications in proteomics (Reproduced with permission,(^3) Copyright 2018, Elsevier)</td>
<td>1</td>
</tr>
<tr>
<td>Figure 1.2: Schematic representation of the site-specific incorporation of unnatural amino acids into proteins by amber suppression (Reproduced with permission,(^23) Copyright 2013, American Chemical Society)</td>
<td>4</td>
</tr>
<tr>
<td>Figure 1.3: Schematic representation of Ligand-Directed (LD) protein labeling (Reproduced with permission,(^32) Copyright 2020, Elsevier)</td>
<td>6</td>
</tr>
<tr>
<td>Figure 1.4: Activity-Based Proteome Labeling. (A) Schematic of conventional active site targeted ABPP. (B) Schematic of site-centric isoTOP-ABPP for identifying a hyper-reactive site. (C) Schematic of site-centric isoTOP-ABPP for fragment discovery (Reproduced with permission,(^32) Copyright 2020, Elsevier)</td>
<td>9</td>
</tr>
<tr>
<td>Figure 1.5: Library of TMAc molecules. (A) Structural variations of A1. (B) Substitution effect on thiol-based trigger-to-release kinetics described using the half-life of thiol. (C) Substitution effect on amine-based trigger-to-release kinetics.(^69)</td>
<td>12</td>
</tr>
<tr>
<td>Figure 1.6: Schematic representation of protein-polymer conjugation, nano-assembly formation, and traceless protein release (Reproduced with permission,(^6) Copyright 2017, American Chemical Society)</td>
<td>15</td>
</tr>
</tbody>
</table>

Figure 2.1: Schematic representation of protein labeling by LD-TMAcs. When the ligand is installed on the ester group of the LD-TMAc it acts as covalent inhibitor (a). If the ligand is installed on the leaving group (b), it is free to leave the active site after the labeling reaction. Then,
an unoccupied active site can bind unreacted LD-TMAcs for additional rounds of labeling. .....

**Figure 2.2:** Structure of LD-TMAc reagents and characterization of in-vitro bCAII labeling. (a) Molecular structure of LD-TMAc reagents 1-3. MALDI-TOF MS of bCAII (5 μM) incubated with (b) 1 (10 μM), PBS buffer, pH 7.4, 37 °C, 2 h, (c) 2 or 3 (10 μM), PBS buffer, pH 7.4, 37 °C, 2 h. Native bCAII has a molecular weight of ~29 kDa. Reagent 1 modification has an adduct of ~368 Da. Reagents 2 and 3 has modification of ~99 Da. * represents the number of labels by 2 or 3.

**Figure 2.3:** Kinetics and tunability of LD-TMAc reagents. (a) Structure of LD-TMAc reagents 1, 5 and 6 with varying leaving groups. (b) Completion rate of bCAII labeling by 1, 5, 6 and in presence of competitive inhibitor EZA. Reaction conditions: 5 μM bCAII, 10 μM LD-TMAc, 100 μM EZA, PBS buffer, pH 7.4, 37 °C. (c) Reaction progress of bCAII labeling with 6 monitored by MALDI-TOF MS. o, native bCAII (MW: ~ 29 kDa); *, 6 labeled bCAII (+ 368 Da).

**Figure 2.4:** Regiospecific labeling of LD-TMAc reagents. (a) Crystal structure of bCAII (PDB: 1V9E) with LD-TMAc reagents 1-3 superimposed on the zinc coordinated sulfonamide binding site. The amino acid residues labeled by LD-TMAC reagents 1, 2, and 3 are shown in their respective colors red, cyan, and magenta. (b) MS-MS spectrum of Met58-Lys75 peptide labeled by reagent 3 on His63. (c) Effect of LD-TMAc modification on carbonic anhydrase activity evaluated by a chromogenic assay.

**Figure 2.5:** Target specific CA labeling of LD-TMAc in whole cell lysate. (a) Structure of alkyne LD-TMAc reagent 4. (b) MALDI-TOF MS of 5 μM bCAII after sequential labeling with 10 μM 4 (30 min) and copper catalyzed Cy3 click (60 min) reactions, compared to unmodified control bCAII. Reaction conditions: PBS buffer, pH 7.4, 37 °C (c) SDS-PAGE in-gel fluorescence analysis of 2 μM bCAII labeling with 4 (30 min) and Cu-catalyzed Cy3 click reaction (60 min) in 750 ug/mL HEK cell lysate in PBS buffer, pH 7.4, 37 °C and (d) Coomassie staining. *, 4 labeled
bCAII (+ 235 Da); *, Cy3 labeled bCAII (+ ~700 Da); ................................................................. 45

Figure 2.6: CLSM images showing the target-specific labeling of CA XII with probe 4 (1 μM) in MCF7 cells in the absence (a) and presence (b) of EZA (100 μM). Fluorescence images of Hoechst, Wheat germ Agglutinin (WGA), 4 + AZDye 647, and their merged images are displayed from left to right of each panel, respectively. Scale bar: 20 μm. ........................................................................ 46

Figure 2.7: Structure of LD-TMAc reagents and characterization of in-vitro bCAII labeling. (A) Molecular structure of LD-TMAc reagents 1-3 and control LD-TMAc. ESI-MS (24+ charge state) of bCAII (5 μM) incubated with (B) 1 (10 μM), 10 mM pH 7.4 MOPS buffer, 37 °C, 1 h, (C) 2, 3, or control TMAc (10 μM), 10 mM pH 7.4 MOPS buffer, 37 °C, 1 h. Native bCAII has molecular weight of ~29 KDa. Reagent 1 modification has an adduct of ~368 Da. Reagents 2, 3, and control TMAc has modification of ~99 Da. Number of modifications by the LD-TMAc reagents are indicated above each new protein peak. ........................................................................ 64

Figure 2.8: MALDI-MS of bCAII (5 μM) incubated with 1 – 6 (10 μM), 100 μM EZA, PBS buffer, pH 7.4, 37 °C, 2 h. No protein modification is observed for 2 – 4. Only 11%, 7% and 28% of bCAII was modified by 1, 5, and 6 respectively. ................................................................. 65

Figure 2.9: MS/MS of LD-TMAc labeled S171-N185 peptide. LD-TMAc labeling on this peptide leads to unambiguous assignment of modification site. ................................................................. 66

Figure 2.10: Crystal structure of bCAII (PDB: 1V9E) with LD-TMAc reagent 3 structure (magenta) superimposed on the zinc coordinated sulfonamide binding site. ................................. 66

Figure 2.11: Solvent accessible surface area (SASA) of Serine, Threonine, Histidine and Lysine on bCAII crystal structure (PDB: 1V9E) calculated using GetArea with a probe radius of 1.4 Å. ........................................................................................................... 67

Figure 2.12: Table of Control TMAc targeted residues on bCAII.................................................... 67
Figure 2.13: H3 modification by LD-TMAc 1 (+367 Da) ......................................................... 68
Figure 2.14: H63 modification by LD-TMAc 1 (+367 Da) ......................................................... 68
Figure 2.15: H3 modification by LD-TMAc 2 (+98 Da) ............................................................. 69
Figure 2.16: H63 modification by LD-TMAc 2 (+98 Da) ............................................................. 69
Figure 2.17: K166 modification by LD-TMAc 2 (+98 Da) ........................................................... 70
Figure 2.18: S182 modification by LD-TMAc 2 (+98 Da) ............................................................ 70
Figure 2.19: T172 modification by LD-TMAc 2 (+98 Da) ............................................................ 71
Figure 2.20: S64 modification by LD-TMAc 3 (+98 Da) ............................................................... 72
Figure 2.21: S72 modification by LD-TMAc 3 (+98 Da) ............................................................... 72
Figure 2.22: S28 modification by LD-TMAc 3 (+98 Da) ............................................................... 73
Figure 2.23: S171 modification by LD-TMAc 3 (+98 Da) ............................................................. 73
Figure 2.24: K44 modification by LD-TMAc 3 (+98 Da) ............................................................. 74
Figure 2.25: K250 modification by LD-TMAc 3 (+98 Da) ............................................................ 74
Figure 2.26: H3 modification by LD-TMAc 3 (+98 Da) ............................................................... 75
Figure 2.27: H2 and H3 modification by LD-TMAc 3 (+98 Da + 98 Da, double modification) 75
Figure 2.28: HNMR of compound 1 .......................................................................................... 76
Figure 2.29: $^{13}$C NMR of compound 1 .................................................................................. 76
Figure 2.30: HNMR of compound 2 .......................................................................................... 77
Figure 2.31: $^{13}$C NMR of compound 2 .................................................................................. 77
Figure 2.32: HNMR of compound 3 .......................................................................................... 78
Figure 2.33: $^{13}$C NMR of compound 3 .................................................................................. 78
Figure 2.34: HNMR of compound 4 .......................................................................................... 79
Figure 2.35: $^{13}$C NMR of compound 4 .................................................................................. 79
Figure 2.36: HNMR of compound 5. ................................................................. 80
Figure 2.37: $^{13}$C NMR of compound 5. ................................................................. 80
Figure 2.38: HNMR of compound 6. ....................................................................... 81
Figure 2.39: $^{13}$C NMR of compound 6. ................................................................. 81

Figure 3.1: A) Hydrophobic tagging induced degradation of proteins; B) PROTAc mediated protein degradation (Reproduced with permission.$^{34}$ Copyright 2017, Elsevier). .......................... 88

Figure 3.2: HyT LD-TMAs’ mechanism of action and possible pathways of protein degradation following hydrophobic tagging. ................................................................. 89

Figure 3.3: a) Structure of HyT LD-TMAc; b) MALDI-MS spectrum of control bCAII and hydrophobically tagged bCAII; c) Western blotting analysis of CAXII and Vinculin in MCF7 cells incubated with increasing concentrations of HyT LD-TMAc for c) 3 hours and d) 6 hours; e) Bar graph of relative CAXII abundance from figures c and d. * Represents the number of hydrophobic adamantane tags by HyT LD-TMAc. ................................................................. 92

Figure 3.4: a) Western blotting analysis of CAXII and vinculin in MCF7 cells incubated with increasing concentrations of HyT LD-TMAs for 12 hours, 3 hours and 0.5 hours; b) Bar graph of relative CAXII abundance from figure a. ................................................................. 93

Figure 3.5: a) Structure of control LD-TMAs with methyl and propargyl ester functionality; Western blotting analysis of CAXII and vinculin in MCF7 cells incubated with increasing concentrations of b) methyl LD-TMAs and d) propargyl LD-TMAs; Bar graph of relative CAXII abundance with increasing concentrations of c) methyl LD-TMAC and e) propargyl LD-TMAC juxtaposed with CAXII abundance when incubated with same concentrations of HyT LD-TMAC (adamantyl probe). ................................................................. 94
Figure 3.6: Western blotting analysis of Vinculin and CAXII to evaluate anti-CAXII IgG recognition post a) HyT, c) methyl and propargyl probe labeling of MCF7 cell lysate; b) Bar graph of anti-CAXII IgG recognition post HyT LD-TMAc labeling of MCF7 cell lysate juxtaposed with anti-CAXII IgG recognition post live-cell labeling; d) Bar graph of anti-CAXII IgG recognition post methyl and propargyl probe labeling of MCF7 cell lysate…………………………………………………………. 96

Figure 3.7: Western blotting analysis of CAXII and Vinculin in MCF7 cells co-incubated with increasing concentrations of HyT LD-TMAc and a) lysosomal inhibitor Bafilomycin or c) proteasomal inhibitor MG132; Bar graph of relative CAXII abundance following co-incubation of increasing HyT LD-TMAc concentrations with b) lysosomal inhibitor Bafilomycin or d) proteasomal inhibitor MG132 juxtaposed with relative CAXII abundance following HyT probe incubation without any inhibitors……………………………………………………………………………………………………. 98

Figure 3.8: a) Western blotting and c) bar graph of relative CAXII abundance in MCF7 cells co- incubated with 5 µM HyT LD-TMAc and increasing concentrations of proteasomal inhibitor MG132; b) gel contents of figure a…………………………………………………………………………………………. 99

Figure 3.9: Structure of coumarin LD-TMAc b) MALDI-MS spectrum of control bCAII and coumarin LD-TMAc labeled bCAII. ………………………………………………………………………………………. 100

Figure 3.10: In-gel analysis of MCF7 cell media and lysate post coumarin LD-TMAc and HyT LD-TMAc labeling. a) Gel contents of figures b-e and LD-TMAc labeling reaction conditions of cells which the samples were collected from. Cell media of figure b and c were collected using method 1 and method 2, respectively. Gels b and c are otherwise identical. b, c) In-gel fluorescence; d, e) SYPRO Ruby total protein staining of gels in figure b and c respectively; f, g) Western blotting of CAII and CAXII respectively. Cell media samples for figures f and g were collected using method 1. Arrow indicates expected position of CAXII based on its molecular
weight. Arrow indicates CAII. Arrow indicates new protein band that does not appear in control media loaded into lane 3. ................................................................. 101

**Figure 3.11:** Effects of ectodomain shedding inhibitors and activators on CAXII degradation. a, b) Western blotting and bar graph quantification of relative CAXII abundance after co-incubation of MCF7 cells with 1 µM or 5 µM HyT LD-TMAc and 20 µM Bat or PMA; c, d) Western blotting and bar graph quantification of relative CAXII abundance after co-incubation of MCF7 with 1 µM or 5 µM HyT LD-TMAc and 20 µM RA, Baf, or MG132. .............................................. 103

**Figure 3.12:** Confocal microscopy workflow for visualization/tracking of a) HyT CAXII; b) control/unmodified CAXII........................................................................................................... 104

**Figure 3.13:** Confocal microscopy tracking of HyT CAXII. Confluent MCF7 cells were incubated with 10 µM HyT LD-TMAc for 30 minutes. Then anti-CAXII FITC IgG was added to cell media to track the HyT CAXII. Hoechst staining, FITC IgG fluorescence, brightfield image and their merged images were obtained following 10, 90 and 120 minutes of FITC IgG incubation. Cells were imaged one more time after 120 minutes’ imaging and washing. Red, pink, and yellow arrows show aggregations of FITC IgG fluorescence outside cells, on cell membrane and inside cells respectively. ........................................................................................................ 106

**Figure 3.14:** Confocal microscopy imaging of CAXII. Anti-CAXII FITC IgG was added to a confluent MCF7 media and cells were stained with Hoechst. Then FITC fluorescence, Hoechst staining and their merged images were obtained after 10 and 120 minutes. ......................... 107

**Figure 3.15:** Confocal microscopy tracking of coumarin LD-TMAc and HyT LD-TMAc labeled CAXII. Coumarin fluorescence, Hoesch staining, brightfield and merged images of MCF7 a) post 30 minutes incubation of 1 µM coumarin LD-TMAc; and b) post 90 and 120 min incubation with 10 µM HyT LD-TMAc following 30 min coumarin labeling. Arrowheads show nuclear membrane
staining. Arrows show perinuclear vesicular staining. Arrows show aggregation of coumarin fluorescence near cell membrane. Arrows show vanished coumarin fluorescence on cell membrane.

Figure 3.16: a) Structure of PDL-1 targeted HyT BMS-1166 LD-TMAc; b) Western blotting analysis of PDL-1 in MDA-MB-231 cells incubated with 5 nM to 5 µM of HyT BMS-1166 LD-TMAc.

Figure 3.17: $^1$H NMR spectrum of compound 7.

Figure 3.18: $^{13}$C NMR spectrum of compound 7.

Figure 3.19: $^1$H NMR spectrum of compound 8.

Figure 3.20: $^{13}$C NMR spectrum of compound 8.

Figure 4.1: Schematic representation and comparison of MDD with parental drug administration. (Reproduced with permission. Copyright 2013, Springer Nature).

Figure 4.2: Schematic representation of protein-polymer conjugation, hydrogel formulation and photo-activated, traceless release of the encapsulated proteins.

Figure 4.3: Schematic representation of polymer-Dexamethasone conjugation, hydrogel formulation and photo-activated, traceless release of the encapsulated Dexamethasone.

Figure 4.4: a) Synthesis of the photo-responsive polymer P1; b) Various polymerizations conditions for optimization of P1 conversion and polydispersity; c) GPC chromatogram of the optimized P1. LC, Low conversion. PDI, polydispersity.

Figure 4.5: a) Synthesis of the photo-responsive polymer P2; b) GPC chromatogram of P2. $M_n = 26$ kDa, PDI = 1.22.

Figure 4.6: Photocleavage of nitrobenzyl carbonate groups in 2 mg/mL PBS solution of P1.
Schematic illustration of P1 photocleavage reaction; b) UV absorption spectrum from different time points of P1 UV exposure................................................................. 132

**Figure 4.7:** SDS-PAGE of P1 protein conjugation reactions from various ratios of a) P1:BSA and b) P1:cytochrome c; c) BSA encapsulation efficiency calculated form the gel on figure a; Surface exposed lysine residues (blue) in the crystal structure of d) BSA (PDB: 3V03) and e) cytochrome c (PDB: 1HRC)..................................................................................................................... 134

**Figure 4.8:** Photo-responsive protein release of P1 hydrogels. a) Cumulative FITC BSA release over 2 hours; b) transformation of solid P1 hydrogel to solution; comparison of hydrogel released c) FITC BSA and d) cytochrome c CD spectra with their control counterparts....................... 135

**Figure 4.9:** a) Conjugation of coumarin with P2; b) HNMR spectrum of the coumarin conjugated P2; GPC chromatogram of P2 polymer c) before (M_n= 26 kDa, PDI = 1.22) and d) after coumarin conjugation (M_n= 27 kDa, PDI = 1.3); e) visual difference of P2 solution after coumarin conjugation................................................................................................................................. 137

**Figure 4.10:** a) Photo-responsive coumarin release of P2 polymer; b) quantification of photo-responsive coumarin release by HNMR spectroscopy. ................................................................. 138

**Figure 4.11:** a) Coumarin fluorescence standard curve; b) slight color transformation of coumarin conjugated non-degradable P2 hydrogels under ~3 hours of UV exposure; photo-responsive coumarin release of P2 hydrogels c) under UV irradiation d) without UV irradiation........ 139

**Figure 4.12:** a) Conjugation of dexamethasone with P2 polymer; b) HNMR spectrum of the dexamethasone conjugated P2; GPC chromatogram of P2 polymer c) before (M_n= 26 kDa, PDI = 1.22) and d) after dexamethasone conjugation (M_n= 29 kDa, PDI = 1.28).............................................. 140

**Figure 4.13:** Mass spectrometry standard curve of Dexamethasone. ......................................................... 141

**Figure 4.14:** HNMR of **PFP monomer**................................................................................................... 150
**Figure 4.15:** $^{13}$C NMR of PFP monomer. ................................................................. 150

**Figure 5.1:** Structure of targeted HyT LD-TMAcs............................................................... 159

**Figure 5.2:** Circadian rhythm enhances physiological responses to chemicals.$^{13}$ ................. 161

**Figure 5.3:** PER2 and BMAL1 expression levels in circadian synchronized and non-synchronized cells.$^{13}$ ................................................................................................................................. 162
## LIST OF SCHEMES

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scheme 2.1:</strong></td>
<td>Synthesis of LD-TMAc 1.</td>
<td>36</td>
</tr>
<tr>
<td><strong>Scheme 2.2:</strong></td>
<td>Synthesis of compound 1.</td>
<td>51</td>
</tr>
<tr>
<td><strong>Scheme 2.3:</strong></td>
<td>Synthesis of compound 2.</td>
<td>56</td>
</tr>
<tr>
<td><strong>Scheme 2.4:</strong></td>
<td>Synthesis of compound 3.</td>
<td>57</td>
</tr>
<tr>
<td><strong>Scheme 2.5:</strong></td>
<td>Synthesis of compound 4.</td>
<td>60</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scheme 3.1:</strong></td>
<td>Synthesis of CA targeted HyT LD-TMAc, compound 7.</td>
<td>114</td>
</tr>
<tr>
<td><strong>Scheme 3.2:</strong></td>
<td>Synthesis of CA targeted coumarin LD-TMAc, compound 8.</td>
<td>116</td>
</tr>
<tr>
<td><strong>Scheme 3.3:</strong></td>
<td>Synthesis of PDL-1 targeted BMS-1166 guided HyT LD-TMAc, compound 9.</td>
<td>117</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Scheme</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Scheme 4.1:</strong></td>
<td>Synthesis of PFP monomer.</td>
<td>147</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

Applications of chemical modifications in proteomics

Figure 1.1: Applications of chemical modifications in proteomics (Reproduced with permission, \(^3\) Copyright 2018, Elsevier).

1.1 Background

Complete sequencing of the human genome brought a wealth of knowledge that led to the discovery of disease targets and markers.\(^1\) This knowledge was the fruit of advanced methods in genomics. However, often more information is necessary for finding therapies for diseases. To come up with treatments, the knowledge needed is to understand the protein function in the complex context of cellular proteomics. Proteomics is the study of a complete set of proteins expressed by a genome, cell, tissue, or organism; and the study of their structure, function, interactions, quantity, and subcellular distribution.\(^2\) A proteome is much more complex than its corresponding genome. The number of different proteins in any proteome is estimated to be ten to hundred times the genome’s genes; this number increases exponentially after the post-translational modifications (PTM).\(^2\) For this reason, proteomics methods could not advance as early as in genomics.

Two of the most used methods in proteomics are two-dimensional gel electrophoresis (2DE) and mass-spectrometry.\(^1\) Low resolving power and its inability to identify low-abundant and membrane proteins limit the use of 2DE, and high throughput analysis using this method is
impractical. Advancements in liquid chromatography (LC) separation and MS instrumentation of proteins and peptides made LC-MS the most used tool for proteomic analysis. Additionally, chemical modification of proteins before LC-MS analysis can add a new dimension to the study. Even the most basic MS proteomic analysis involves tryptic digestion, disulfide reduction, and alkylation chemical modification steps. When followed by chemical modification by designer molecules with tailored reactivities and functionalities, MS proteomic analysis can result in more information than just protein sequence for a range of specialized applications (Figure 1.1).³

Chemical modification of proteins in general, is part of a wide variety of methods in proteomics and biotechnology applications. Coupled with other methods, chemical modification of proteins has been used for inhibitor discovery⁴, antibody-drug conjugates (ADCs),⁵ drug delivery,⁶ amino acid reactivity profiling,⁷ protein structural analysis,⁸ and subcellular localization.⁹ This versatility of chemical modifications fostered a great interest in developing various chemical modification methods, but fundamental challenges have kept back their projected impact.¹⁰ Some of the challenges in chemical modification include: (1) The chemical reaction must tolerate the biological conditions (temperature, pH, aqueous solvent) (2) Kinetics of the chemical modification must be compatible with biologically low protein concentrations (3) Depending on the application, chemical modifications must have varying degrees of selectivity such as chemo-selectivity, regio-selectivity or protein selectivity (4) More challenges are added to that of test tube conditions, when the application is in complex cellular and in-vivo environments, such as permeability, pharmacokinetics, avoiding enzymatic degradation and toxicity.

1.2 Selectivity in protein labeling

1.2.1 Genetic engineering approaches

Selectivity is the most important factor of chemical modifications, and it ultimately decides what
kind of applications the modification can be used for. Therefore, controlling the number and position of chemical modifications is essential. Selectivity in protein labeling can range from selective labeling of specific chemical groups (e.g., cysteines), protein family (e.g., serine proteases), a protein of interest (POI) (e.g., carbonic anhydrase), or a specific region on a protein structure (regio-selective labeling).^10

Genetic engineering approaches for selective protein labeling have been widely used for protein tracking and activity control as they are precisely protein and regio-selective.\textsuperscript{11–13} One of the genetic engineering approaches for protein labeling is the expression of target proteins with self-labeling protein tags.\textsuperscript{14} Self-labeling protein tags are small protein segments (<40 kDa) that can be co-expressed in either the N or C terminus of the POI. Ideally, these protein tags should react with specific chemical warheads selectively and rapidly and have good thermodynamic stability. These warheads are linked to functional groups, such as fluorescent probes for imaging or E3 ligase recruiting units for ubiquitination and proteasomal degradation.\textsuperscript{15,16} Many self-labeling protein tags have been developed: HaloTag, SNAP-tag, CLIP-tag, PYP-tag, DHFR-tag, and RA-tag. Derived from haloalkane dehydrogenase, HaloTag is arguably the most popular of the self-labeling tags. The wild-type haloalkane dehydrogenase hydrolyzes chloroalkanes by its Glu-His-Asp catalytic triad.\textsuperscript{17–19} First, chloride from the haloalkane is displaced by activated glutamate, and then a water molecule activated by the basic histidine hydrolyzes the intermediate ester. In the Halo-tag construct, the catalytic histidine is replaced by inactive phenylalanine, which results in permanent haloalkane labeling of the catalytic glutamate.\textsuperscript{20} Although relatively large when compared to other self-labeling tags, Halo-Tag’s net negative charge allows the tagged protein to remain soluble. While the kinetics of Halo-Tag is almost diffusion-limited for specific rhodamine substrates, it can be slower for non-fluorophore substrates.\textsuperscript{21} SNAP tags can be better
suited for these substrates as they have relatively unbiased reaction kinetics against different substrates.

Genetic code expansion has allowed incorporating various unnatural amino acids with specialty side chains. To express proteins with unnatural amino acids, plasmid constructs for expression of tRNA, aminoacyl tRNA synthase, and the recombinant protein with amber codon genes are transfected to the host cell (Figure 1.2). Together, this construct incorporates the fed unnatural amino acid into a specified sequence of the recombinant protein. Compared to the self-labeling tags, which are segments of a protein that are appended to either the N or C-terminus of the protein, unnatural amino acids can be incorporated anywhere in the protein sequence. Most popular unnatural amino acids for protein labeling applications bear one of the orthogonal reactive
pairs of azide-propargyl/DBCO and tetrazine-TCO/BCN. For example, after the selective incorporation of TCO-modified unnatural amino acids, the recombinant protein can be labeled by a tetrazine dye for live cell imaging.\textsuperscript{24}

Although genetic engineering approaches for protein labeling are a powerful chemical biology tool, by their nature, they are limited to genetically engineered systems. They cannot be used to investigate proteins in their natural state. Genetically engineered protein segments or amino acids may influence protein activity. Furthermore, genetic engineering requires special techniques and laborious procedures to make the genetically engineered protein expression machinery.\textsuperscript{12} Therefore, simple and economical procedures are needed to label natural proteins.

### 1.2.2 Amino acid selective approaches

One of the simplest ways of achieving selectivity in protein labeling is targeting specific amino acid residues, as amino acids have different chemical functionalities such as thiol, amine, imidazole, phenol, or thioether. Cysteine has the most nucleophilic side chain among natural amino acids and low genomic abundance (2.3 \% of the genome).\textsuperscript{25} Therefore, it makes an excellent target for chemical modifications, and it has been modified by many electrophiles such as maleimides, \(\alpha\)-halocarbonyls, and others. When the number of cysteines is not enough, or the protein lacks cysteines, the \(\varepsilon\)-amino group of lysine is another good target, and it has been modified by activated esters, sulfonyl chlorides, isocyanates, isothiocyanates, and others.\textsuperscript{10} However, the high abundance of lysine residues makes controlling the number of modifications and reaction yield challenging. Among others, there are aromatic amino acids, N and C terminal residue targeted chemical modifications. Although these types of chemical modifications have some degree of chemical and regioselectivity, they do not have any protein selectivity. Therefore, they are primarily used for modification of pure protein samples in test tube conditions such as ADC preparation and
attachment of PEG residues on therapeutic proteins.\textsuperscript{26,27} These reagents have also been broadly used for covalent labeling applications such as in-vivo imaging,\textsuperscript{28} site-selective dual labeling,\textsuperscript{29} and protein immobilization.\textsuperscript{30,31}

1.2.3 Affinity-based and Proximity-based labeling

Affinity-based labeling is a site-specific and target-selective chemical modification method that uses the ligand of the targeted POI.\textsuperscript{10} Although a few different platforms can be broadly categorized as affinity-based, each platform and specific approach have limitations. Ligand-Directed (LD) protein labeling is an affinity-based labeling approach effective in test tube and live-cell conditions.\textsuperscript{10,32} LD probes are comprised of a covalently linked ligand, a reactive warhead, and a functional handle such as a fluorophore, biotin, or bio-orthogonal reactive group (Figure 1.3).\textsuperscript{32} Ligand binding increases the reactive warhead’s local concentration, allowing the nearby surface exposed nucleophilic amino acids to react by proximity effect. Various reactive warheads such as tosylates,\textsuperscript{33} acyl imidazoles,\textsuperscript{34} sulfones,\textsuperscript{35} epoxides,\textsuperscript{36} and Michael acceptors\textsuperscript{37} have been used in LD probes for numerous applications: development of fluorescent biosensors, caged enzymes, covalent inhibitors, pulse-chase analysis of membrane proteins and studying protein-protein interactions.\textsuperscript{10,33,35,37–40} However, some LD probes have shortcomings, such as
slow reaction kinetics, low yield, and inhibition of protein activity after protein labeling. More importantly, they all have different nonoverlapping amino acid preferences necessitating an informed choice among them based on the compatibility of the proximal nucleophiles around the binding pocket of the targeted protein. Therefore expanding the scope of targetable amino acids by the LD approach remains imperative.

Affinity-guided catalyst is another affinity-based approach that labels the target protein in two steps: 1) binding of ligand-guided catalyst to the target protein and 2) protein labeling by probe-appended catalyst substrate. In this three-component reaction, target proteins are labeled by short-lived reactive species generated when the catalyst-substrate finds the protein-bound ligand-catalyst. Unlike the LD approach, where the ligand-bound reactive groups react intramolecularly with the target protein, the reaction of probe-appended catalyst, substrate, and the target protein is a conventional intermolecular reaction. Therefore, affinity-guided catalyst approaches suffer from slow reaction kinetics.

Photoaffinity-based probes are ligands equipped with photoreactive groups and functional handles. In a typical procedure, photoaffinity-based probes are first added to a live-cell culture or extracted proteome. Then, photoreactive groups generate radicals upon UV activation that react with CH/XH (X = N, O, S) bonds on the ligand-bound proteins. Since this reaction is not dependent on the ligand binding state, photoreactive groups can result in off-target protein labeling, causing a significant drawback for photoaffinity-based probes. Furthermore, nonspecific CH/XH reactions result in high background signal, and the radicals’ short half-life can result in low yields. Nevertheless, photoaffinity-based probes have been instrumental in identifying target proteins and binding sites.
Proximity-based labeling maps the interactome of a POI by genetically attaching an enzyme that generates short-lived reactive biotinylating intermediates. Subsequently, the interactome can be enriched with avidin affinity and identified with MS. Some proximity labeling approaches use peroxidase-based enzymes such as ascorbate peroxidase (APEX) with hydrogen peroxide and biotin-phenol substrate to generate reactive phenoxy radicals that primarily react with tyrosines. Other methods, such as biotin ligase proximity-dependent biotinylation identification (BioID) and TurboID, use biotin ligase with adenosine triphosphate (ATP) and biotin substrate to generate lysine reactive biotin–adenosine monophosphate (biotin–5'-AMP) intermediate.

1.3 Applications of protein labeling
The most widely used application of target selective protein labeling is the localization and tracking of the POIs. All target selective approaches: self-labeling protein tags, unnatural amino acid incorporation, affinity-guided catalysts, LD approaches and photoaffinity labeling have been demonstrated effective for protein tracking applications.

1.3.1 Proteomics and Drug Discovery
Protein and enzyme activity levels depend on multiple factors such as protein quantity, health, and post-translational modifications. Disturbances in the activity levels of proteins can result in substantial cellular and systemic complications. For example, while telomerase and glyceraldehyde-3-phosphate dehydrogenase (GAPDH) is upregulated, fructose-bisphosphatase 1 (FBP1) and succinate dehydrogenase (SDH) is downregulated in cancer cells. Therefore, quantifying the activity levels of enzymes in unhealthy cells can reveal which enzymes contribute to a particular phenotype. Additionally, quantifying the activity level of known enzyme biomarkers can help monitor disease progression. Activity-based probes (ABP) are covalently linked warheads that react with active sites of enzymes and fluorescent or affinity
Since APB warheads are mechanism-based inhibitors, they only target active enzymes and do not react with inactive enzymes. Activity-Based Protein Profiling (ABPP) analyzes and quantifies enzymatic activity in native biological systems. Although ABPs are specific to family or families of proteins, they are not specific to a protein of interest. A wide range of warheads such as phosphonate, E-64, ketone, and HxBP-Rh probes have been developed to target serine hydrolases, cathepsins, caspasas,
metalloproteinases, and others, respectively.\textsuperscript{52}

The application of ABPP has been extended to amino acid reactivity profiling and the discovery of covalent inhibitors (Fragment discovery) (Figure 1.4 b, c). PTMs and other functional sites, such as active sites, allosteric sites and protein interaction surfaces often depend on hyperreactive amino acid side chains. Amino acid side chain reactivity can range on protein tertiary structure by several orders of magnitude.\textsuperscript{7} Due to their prevalence in functional sites, hyperreactive amino acids are protein functional hotspots.\textsuperscript{7} isoTOP-ABPP (isotopic tandem orthogonal proteolysis–ABPP) is a method that quantifies amino acid side chain reactivity in a whole proteome (Figure 1.4 b).\textsuperscript{53} In this method, nucleophilic amino acids of two proteome samples are labeled with low and high concentrations of site-reactive electrophilic probes with an alkyne group. While hyperreactive amino acids react faster and can be labeled efficiently with both low and high probe concentrations, less reactive amino acids show concentration-dependent labeling. Alkynes on the site-reactive probes are then clicked with either isotopically light or heavy tags/enrichment handles. Two protein samples are mixed, enriched, digested with trypsin, and subjected to MS analysis. After LC separation, each peptide is identified by MS fragmentation, and the reactivity of the labeled amino acid is quantified as the ratio of light and heavy peptide MS intensities.

Hyperreactive amino acids are also targets for covalent drugs.\textsuperscript{54} Most initial covalent drugs targeted hyperreactive catalytic residues, but recent covalent drugs have also targeted noncatalytic residues of nuclear export factors, GTPases and kinases. The isoTOP-ABPP method can also discover covalent fragment electrophiles that specifically target hyperreactive amino acids (Figure 1.4 c).\textsuperscript{53} First, one of the two proteome samples is treated with a potential covalent inhibitor (fragment electrophile), then both samples are treated with equal concentrations of the site-reactive
probe and are subjected to the rest of the isoTOP-ABPP procedure. Competitive displacement of the site-reactive probe by the fragment electrophile leads to high heavy/light channel ratios indicating reactivity of the fragment electrophile with a specific residue.

1.3.2 Structural analysis

Cross-linking mass spectrometry (XL-MS) is a method of mapping protein structure with fixed-length chemical cross-linkers.\textsuperscript{55,56} Although this method does not give a complete picture of the three-dimensional structure, it has advantages over other popular protein structure analysis techniques such as Nuclear Magnetic Resonance (NMR) spectroscopy, X-ray crystallography, and cryo-Electron Microscopy (EM) techniques. The XL-MS method analyzes protein samples in solution, enabling the study of flexible and dynamic parts of the protein structure. The XL-MS method has been used to solve the topology of protein complexes. High-density crosslinking of model peptides has been used to guide software that predicts protein folding.\textsuperscript{57} XL-MS is also used to monitor conformational changes in protein structure.\textsuperscript{57} Furthermore, the entire protein interactome can be mapped by directly applying chemical crosslinkers to live cells.\textsuperscript{57} Reactive groups are the most critical functionality of the XL-MS crosslinker probes and have varying degrees of selectivity.\textsuperscript{58} Zero-length crosslinkers such as carbodiimides crosslink only very close amine and carboxylic acids. Maleimides and NHS-esters crosslink a limited number of thiols and primary amines, respectively. On the other hand, non-specific photo-activatable cross-linkers generate carbenes or nitrenes that crosslink pretty much anywhere on the protein structure.

The availability of an amino acid residue for covalent labeling depends on its surface exposure. Therefore, covalent labeling extents of amino acids on protein structure can reveal surface topology and higher order structure information in different conditions, such as while the protein is bound or unbound. Various reagents such as hydroxyl radicals,\textsuperscript{59} carbenes,\textsuperscript{60,61} N-
hydroxysuccinimide (NHS), dimethyl(2-hydroxy-5-nitrobenzyl)sulfonium bromide (HNSB),
1-ethyl-3-(3-(dimethylamino)propyl)-carbodiimide/glycine ethyl ester (EDC/GEE),
iodoacetamide and more have been used for covalent labeling to understand protein
structure. However, while radicals and carbenes generate many modification products
resulting in complex data challenging to analyze, other reagents such as iodoacetamide and NHS
can only target limited residues such as cysteine and lysine, respectively. Diethylpyrocarbo
nate (DEPC) is a compound that can modify Lys, His, Tyr, Ser, and Thr, and its modification products
are limited to a constant mass adduct with all these amino acids. This comprehensive amino
acid coverage translates into high structural resolution while the data analysis is still simple.

In a recent work by Zhuang et al., a library of activated Triggerable Michael Acceptors
(TMAs) with a good leaving group at β position with reactive tunability of over six orders of
magnitude was synthesized (Figure 1.5). TMAs can react with sulfhydryl, amine, and
hydroxyl nucleophiles in a two-step mechanism initiated by a nucleophilic attack on the α
position, which releases the leaving group on the β position. The precise tunability of the TMAcs was further demonstrated by chemo-selective modification of cysteines on β lactoglobulin (bLGb) by a relatively less reactive compound 8. On the other hand, more reactive compound 1 has modified multiple other residues on the same protein, including lysine, tyrosine, and threonine. Using these molecules, Zhao et al. demonstrated the application of TMAcs in protein covalent labeling to elucidate protein structure and protein-protein interaction. This study showed TMAcs could label up to 13 different amino acids: Cys, Lys, His, Ser, Thr, Tyr, Asp, Glu, Arg, Trp, Asn, Gln, and Met, with a constant mass adduct keeping the data analysis simple. Furthermore, the modular structure of the TMAcs allowed the synthesis of isotopic TMAcs for facilitated MS² analysis and multiplexing applications.

1.3.3 Targeted protein degradation

After the emergence of PROteolysis TArgeting Chimeras (PROTACs) as the first targeted protein degradation (TPD) platforms, new TPD platforms are emerging yearly, and their interest keeps soaring. PROTACs are covalently linked ligands for the target protein and E3 ligase, forming a ternary complex with E3 ligase and the target protein. This complex results in ubiquitination and subsequent proteolytic degradation of the targeted protein. Although they do not have therapeutic potential as they are employed in genetically engineered systems, Halo tags have been demonstrated as an effective chemical biology tool for TPD. Target proteins were efficiently degraded by the proteasome after covalently labeling the Halo-tagged target protein with E3 ligase VHL ligand.

Interestingly, labeling of Halo-tagged target protein with hydrophobic moieties such as adamantane and Boc₃Arg has been shown to induce protein degradation. Since only 2% of ~600 E3 ligases in the human genome can be employed by PROTACs, Hydrophobic tagging
(HyT) is an excellent addition to TPD strategies, especially in a system where E3 ligases cannot be used. Exemplified with Fulvestrant, even noncovalent binding of ligands with hydrophobic groups can induce protein degradation. To the best of our knowledge, a comparison of covalent and noncovalent hydrophobic labeling of the same target protein and ligand combination has not been shown. Still, it is easy to guess that permanent covalent labeling would result in more effective TPD. For instance, covalent and noncovalent Boc₃Arg labeling of glutathione S-transferase (GST) and *Escherichia coli* dihydrofolate reductase (eDHFR) led to degradation of 80% and 30%, respectively, in whole cells. Therefore, it will be interesting to see other target-specific protein labeling approaches, such as the Ligand-Directed approach employed in HyT-induced TPD. As the LD approach results in a covalent label, it could increase the efficiency of hydrophobic ligands/drugs, especially if they have a lower affinity. Additionally, some of the reactive groups used in LD approaches have leaving groups that could be further functionalized to improve the solubility and nonspecific association concerns of HyT probes. As the leaving groups are not bound to the protein after protein labeling, only the hydrophobic portion of the HyT probe would remain on the targeted protein.

1.4 Protein modification chemistries in bioconjugation and drug delivery

1.4.1 Bioconjugation

Although small molecules are still dominating the pharmaceutical market, the approval of biologics is on the rise. Along with many other factors, post-translational modification of proteins to improve pharmacokinetics and immunogenicity has significantly pushed biologics into drug markets. One such modification is an attachment of polyethylene glycol (PEG) chains. PEG can increase the hydrodynamic radius of protein, resulting in less renal filtration. Additionally, PEG can shield the proteins from proteases and hide them from immune detection.
Currently, most of the approved PEG bioconjugates were modified with non-selective chemistries. For instance, Adagen was developed by nonspecific conjugation of NHS PEG on lysine, serine, tyrosine, and histidine residues of adenosine deaminase. However, such non-selective conjugation can reduce the activity of protein therapeutics. Therefore, site-selective approaches were developed that modify N-terminal residues with 2-pyridinecarboxyaldehyde or generate aldehyde at the N-terminus for subsequent selective modifications. Another approach is the incorporation of unnatural amino acid para-acetylphenylalanine at pre-determined sites for modification with aminoxy-PEG. Antibody Drug Conjugates (ADCs) that target specifically deliver cytotoxic payloads are also generated by covalent modification of proteins. There are multiple strategies for antibody drug conjugation, such as stochastic modification of lysine or cysteines with non-selective maleimide or NHS chemistries. More advanced approaches for
selective modification include incorporating engineered cysteines, unnatural amino acids, and disulfide re-bridging.

1.4.2 Drug delivery

Beyond making protein conjugates, protein modification chemistries are used to conjugate proteins with macromolecular delivery vehicles. Towards this end, our group has developed random poly(methyl methacrylate) copolymers with PEG and Nitrophenyl Carbonate (NPC) repeating units (Figure 1.6). Under slightly basic conditions, NPC units react with surface-exposed lysine residues to form polymer-protein conjugates, which can be crosslinked with diamines on the unreacted NPC units to form a protein encapsulated nano-assembly. NPC units are connected to the polymer backbone via self-immolative disulfide linkers, which release the proteins without trace under reducing cytoplasmic conditions. This release can be accomplished by reducing agents such as DTT or glutathione (GSH) in cellular conditions (Figure 1.6). In other works by or group, proteins were modified by bifunctional NPC-boronic acid linker molecules. While NPC reacts with surface-exposed lysine residues, boronic acid conjugates the modified polymers with salicylic hydroxamate polymers or positively charged silica nanoparticles. Notably, the linkers are designed to be self-immolative to release the protein cargo tracelessly in the presence of redox, pH, or GSH triggers.

Non-covalent protein encapsulation into hydrogels has been widely used for insulin delivery and controlled release of growth factors. Covalent conjugation of proteins into hydrogels has enabled more applications, such as advanced cell culture and tissue engineering. With the combination of genetic and enzymatic engineering, Shadish et al. have developed various proteins with photo-responsive azide handle that can be covalently conjugated to hydrogel network via Strain-Promoted Azide-Alkyne Cycloaddition (SPAAC) reaction.
have generated precisely controlled patterns of covalently conjugated proteins in hydrogels using laser-scanning lithographic strategies. Their cell culture experiments have demonstrated that growth factor patterned hydrogels can guide cell proliferation. Covalent protein conjugation and light or enzyme-responsive protein release to guide cell proliferation in advanced cell culture settings have been demonstrated in multiple works. However, while protein genetic engineering has previously discussed shortcomings, proteins released in these works often have chemical modification traces which can result in unexpected effects on protein activity. In a rare example of traceless protein release in hydrogels, Grim et al. have demonstrated photo-responsive thiol-ene bioconjugation and release of TGF-β1 for guided cell growth. However, this method can be applied only for proteins with surface-available cysteines.

1.5 Summary and Overview of the Thesis

In this chapter, we have briefly discussed chemical modification of proteins and their applications in chemical biology, proteomics, and therapeutics. We have also discussed the limitations of different protein labeling platforms and approaches. Although various reactive groups have been employed in Ligand-directed (LD) target-specific protein labeling of endogenous proteins, most current strategies can target only one or two amino acid residues. Therefore, prior knowledge of the ligand binding pocket and surrounding residues is a prerequisite to determine if a particular reactive group employed in the LD approach could successfully label the targeted POI. Reactive groups such as TMAcs with a broad scope of amino acid targets will not only be able to target most proteins, but they could also install multiple residues around the binding pocket.

Although TPD is currently a hot topic, to the best of our knowledge, only HaloTags have been employed for covalent hydrophobic tagging-induced TPD; and the effects of extracellular hydrophobic tagging on membrane proteins remain unexplored. To bridge this gap, chapters 2 and
3 of this thesis focus on developing an LD approach with a broader amino acid scope and its application on TPD of membrane proteins. Additionally, chapter 4 will focus on the development of polymers for covalent conjugation and photo-responsive release of therapeutics in hydrogel drug depots.

1) In Chapter 2, we have developed highly reactive Ligand-Directed Triggerable Michael Acceptors (LD-TMAcs) that feature rapid protein labeling. Unlike previous LD approaches, the unique reactivity and broad amino acid scope of LD-TMAcs enable multiple modifications on a single target protein, effectively mapping the ligand binding site. Moreover, the number of modifications on the target protein can be controlled by the linker length between the reactive TMAc and the ligand. LD-TMAcs are target selective, and we demonstrate the utility of this method by selectively labeling membrane-bound carbonic anhydrase XII in live cells.

2) In Chapter 3, we have extended the application of LD-TMAcs by functionalizing them into HyT targeted protein degraders. HyT LD-TMAcs are functionalized with hydrophobic adamantane group on their ester functionality. In-vitro studies show that HyT LD-TMAcs can install multiple hydrophobic adamantane units on a single target protein. We hypothesize that multiple hydrophobic tags concentrated around the ligand binding pocket collectively inflict much more significant structural damage than the simple binding of adamantane conjugated ligand. Our HyT LD-TMAcs showed 75% targeted degradation of CAXII under 3 hours on MCF7 cells.

3) In Chapter 4, we have designed reactive polymers that can be covalently conjugated with protein or small molecule therapeutics. After covalent conjugation, we crosslinked these polymers to make hydrogel drug depots. The covalent bonds
between polymer and therapeutics can be photo-cleaved on-demand to release active therapeutics tracelessly. We envision our hydrogel drop cast on devices will find applications in tissue engineering and Macroscale Drug Delivery (MDD).96

1.6 References


Bachovchin, D. A.; Mowen, K.; Baker, D.; Cravatt, B. F. Quantitative Reactivity Profiling

**2003**, *125* (9), 2416–2425.

(9) Yasui, R.; Miki, T.; Ojida, A.; Hamachi, I. Ligand-Directed Acyl Imidazole Chemistry for
3961–3964.

(10) Tamura, T.; Hamachi, I. Chemistry for Covalent Modification of Endogenous/Native
(7), 2782–2799.

(11) Oliveira, B. L.; Guo, Z.; Bernardes, G. J. L. Inverse Electron Demand Diels–Alder

(12) Lang, K.; Chin, J. W. Cellular Incorporation of Unnatural Amino Acids and

(13) Sletten, E. M.; Bertozzi, C. R.; Bertozzi, C. R.; Sletten, E. M. Bioorthogonal Chemistry:
Fishing for Selectivity in a Sea of Functionality.

(14) Hoelzel, C. A.; Zhang, X. Visualizing and Manipulating Biological Processes by Using


(42) Herner, A.; Marjanovic, J.; Lewandowski, T. M.; Marin, V.; Patterson, M.; Miesbauer, L.;


(49) Jafri, M. A.; Ansari, S. A.; Alqahtani, M. H.; Shay, J. W. Roles of Telomeres and


(79) Dutta, K.; Kanjilal, P.; Das, R.; Thayumanavan, S. Synergistic Interplay of Covalent and


CHAPTER 2

MULTISITE LABELING OF PROTEINS USING THE LIGAND-DIRECTED
REACTIVITY OF TRIGGERABLE MICHAEL ACCEPTORS


2.1 Introduction

Chemical modification of proteins has been of great interest because of potential utility in a variety of applications including therapeutic biologics, quantification, enrichment, imaging, and structural analysis.1–8 Modification of proteins typically involves a protein family, specific amino acid, single protein of interest, or specific region on the protein of interest.7,9–13 Selective modification of proteins can be accomplished by genetically engineering cells to express a protein bearing unnatural amino acids with reactive handles that are amenable for further bio-orthogonal functionalization.14,15 This method can be exquisitely specific but is labor intensive. To overcome this challenge, photoaffinity-based probes have been developed where photoreactive appendages generate reactive radicals that are rapidly captured by the neighboring protein. While the unrestrained reactivity of the photoaffinity based probes enables them to label virtually any protein regardless of their amino acid composition it results in off-target labeling, as this process is not dependent on the bound state of the ligand.16

Ligand-directed labeling approaches use ligand-protein interactions, where a protein-selective ligand is linked to a reactive warhead and a labeling functional group.9,17–23 The reactivity of the warhead with surface functionalities on the protein is substantially enhanced in the bound state, because of the higher local concentration of the reactive functionality. Many such approaches have
been developed and utilized in a variety of applications including biosensing, enzyme caging, covalent inhibition, pulse-chase analysis, and for studying protein-protein interactions.\textsuperscript{17,20,22,24–27}

The reactive warheads in these cases are primarily designed to attach a single label on the protein using the reactivity to a one type of amino acid, most commonly lysine or cysteine. This strict amino acid preference requires the presence of proximal nucleophiles around the ligand binding pocket. This means that a particular warhead cannot be used to label the target protein if there is no compatible amino acid around the ligand binding site. While the single labeling strategy offers advantages in many applications, a complementary approach with wider amino acid scope that is capable of installing multiple labels on specific protein surfaces would greatly expand the repertoire of ligand-directed protein labeling, such as in the context of protein surface mapping. Such a possibility is schematically illustrated in Figure 2.1. As shown in Figure 2.1b, incorporation of multiple labels offers a much better opportunity to map the ligand binding site, compared to a single site labeling.

**Figure 2.1:** Schematic representation of protein labeling by LD-TMAcs. When the ligand is installed on the ester group of the LD-TMAc it acts as covalent inhibitor (a). If the ligand is installed on the leaving group (b), it is free to leave the active site after the labeling reaction. Then, an unoccupied active site can bind unreacted LD-TMAcs for additional rounds of labeling.
For multi-labeling of specific proteins using the ligand-directed strategy, two criteria must be satisfied: (i) although the reactivity of the warhead must be high to satisfy the first condition, it must be relatively dormant when not bound to the protein such that the labeling occurs only on the surface proximal to the ligand binding pocket; (ii) the reactive warhead should be able to label many amino acid types on a protein surface. Recently, we reported a library of triggerable/tunable Michael acceptors (TMAcs) that exhibit a range of reactivity with nucleophilic functionalities.\textsuperscript{27} In the context of functional groups that are presented on protein surfaces, TMAcs have been shown to label 13 different amino acids,\textsuperscript{28} making them valuable for covalent labeling mass spectrometry (MS) experiments.\textsuperscript{29} The reactive tunability of TMAcs over six orders of magnitude offer the unique potential to design molecules that could afford ligand-directed labeling, while also incorporating multiple surface functionalities on proteins. In this manuscript, we report on ligand-directed TMAcs (LD-TMAcs) that label specific proteins but on multiple sites. In addition to demonstrating this possibility with soluble proteins, we also show that the strategy can be used to specifically label proteins in cell lysates and specific membrane proteins on live cells.

2.2 Results and discussion

2.2.1 Molecular design and characterization of carbonic anhydrase (CA) labeling by LD-TMAc reagents

Two different types of LD-TMACs, \textit{viz.} the ones that can label the protein surface at multiple sites and ones that functionalize at a single site, have been designed. Ligand-directed functionalization of proteins is achieved when the ligand binds to a specific site in the protein, which in turn brings the reactive Michael acceptor moiety in close proximity to the nucleophilic functionalities in the protein. This increased local concentration causes the reactive moiety to attach to an accessible and complementary surface functionality on the protein surface. In LD-
TMACs reported here, the reactive functionality is an α,β-unsaturated carbonyl molecule that acts as a Michael acceptor, the ligand is based on benzenesulfonamide that is known to be a nM binder to the model protein carbonic anhydrase (CA)\(^{30}\), and the leaving group is a pyridinium moiety. Representative LD-TMAc structures are shown in 2.1. When the leaving group is separate from the ligand, the reaction between the protein surface and the TMAc functionality affords covalent attachment of the ligand to the protein at an optimal location; \textit{i.e.}, this reaction results in the generation of a covalent inhibitor. This possibility is illustrated with LD-TMAc in Figure 2.1a. On the other hand, if the leaving group is part of the linker that connects the TMAc functionality and the ligand, then the reaction with the protein surface functionality would disconnect the ligand moiety from the LD-TMAc, while concurrently labeling the protein surface. As the product sulfonamide ligand in this process is non-covalently bound to the protein, it can be displaced by another LD-TMAc molecule in solution and cause the protein surface labeling to occur again. This opens up the possibility of multiple labeling of protein surfaces, as illustrated with LD-TMAc in Figure 2.1b.

**Scheme 2.1:** Synthesis of LD-TMAc 1.
The design hypotheses mentioned above were tested by reacting bovine carbonic anhydrase II (bCAII) with LD-TMAcs 1-3 (Figure 2.2a). Syntheses of LD-TMAcs is exemplified in Scheme 2.1 with the synthesis of 1, while experimental details and characterization of all LD-TMACs are provided in the Supporting Information (SI). To test the ligand-directed functionalization of proteins, bCAII (5 μM) was incubated with 10 μM LD-TMAcs for 2 hours in PBS buffer at 37 °C. Extent of protein labeling was evaluated using matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF-MS) (2. 2 b, c). When treated with 1, the bCAII spectrum shifts by 368 Da (Figure 2.2b). This shift corresponds to a single adduct from 1 that includes the sulfonamide ligand, with no discernible evidence for multiple adducts in this experiment. On the other hand, under the same concentration and conditions, bCAII is labeled by multiple reagent
molecules when treated with LD-TMAcs 2 or 3 (Figure 2.2c). The SN2’ reaction of molecules 2 and 3 would incorporate a label with the molecular weight of 99 Da. Indeed, multiple new peaks appear with a separation of 99 Da. Up to six labels are observed with LD-TMAc molecules 2 and 3. Interestingly, the extent of multiple label incorporation is higher for 3 relative to 2, as discerned by the relative abundance changes in the adducts. This higher labeling extent is attributed to the longer linker length in 3 that allows better reach for the protein surface functionalities. The linker length between the electrophilic site in the TMAc to the sulfur atom on the sulfonamide in 1, 2, and 3 are approximately 18, 16, and 24 Å, respectively. Labeling results were also evaluated using electrospray ionization MS (ESI-MS) (Figure 2.7). These results reveal that bCAII labeling with 1
can result in doubly labeled proteins. This could either be due to a second labeling event while the covalently tethered ligand from the first label is in its unbound state, or a non-ligand directed labeling. bCAII labeling by a control TMAc without benzenesulfonamide ligand can also result in some non-ligand directed labeling, but also reveals that such a labeling is minimal (Figure 2.7).

To further test whether the observed labeling is indeed due to ligand-directed increases in effective concentration, we carried out competitive inhibition experiments. Ethoxzolamide (EZA) is a strong binder with bCAII with the \( K_d \) of \(~0.2\) nM.\(^{31}\) When 100 \( \mu \)M of EZA was used as a competitive inhibitor during the treatment of bCAII with the LD-TMAcs, no protein modification is observed for 2 and 3, and only 11% modification for 1 confirming that the observed results are indeed a ligand-directed functionalization process (Figure 2.8).

A potential concern would be whether the reactivity of 1 would be substantially different from 2 and 3, because of the substituents at the para-position of the latter molecules. To test this possibility, we synthesized LD-TMAcs 5 and 6 containing a strong electron-donating methoxy and a strong electron-withdrawing carbonitrile moiety, respectively (Figure 2.3a). Protein labeling ability of 5 and 6 was compared with that of 1. As anticipated, we find 5 to be less reactive than 1, as the electron-donating methoxy unit resonance stabilizes the pyridinium cation causing it to be less effective as a leaving group. Similarly, the electron-withdrawing carbonitrile causes 6 to be more reactive than 1. Reaction rates were estimated from the ratio of MALDI-MS abundances of modified and unmodified bCAII with time (Figure 2.3b,c). In the presence of 100 \( \mu \)M EZA as a competitive inhibitor, 1, 5 and 6 have completion of only 11%, 7% and 28% respectively in two hours, supporting the ligand-directed labeling mechanism (Figure 2.8). Perhaps most importantly, only single labeling was observed in all three pyridinium-based LD-TMAcs, suggesting that subtle reactivity differences between 1 and 2 (or 3) do not explain the observed variations about the
number of labels that are added upon reacting with 2 and 3. Moreover, it is crucial to be able to tune the kinetics of LD labeling as it depends on the affinity of the ligand being used.\textsuperscript{9,20,32} These results show that the kinetics of LD-TMAcs can be easily tuned by changing the leaving group, as was observed previously by our group.\textsuperscript{27}

2.2.2 LD-TMAc regiospecific labeling around the ligand binding site

Next, we were interested in testing if the multi-labeling features of the LD-TMAcs could map the region around the ligand binding site. To identify the labeling sites on the protein surface, bCAII was subjected to a tryptic digest following the labeling reaction with the LD-TMAcs. The product peptides were then analyzed using LC-MS/MS to identify the specific residues that were modified upon reaction with the LD-TMAcs (Table 1). Tandem mass spectra of the peptide modifications are provided in the Supplementary Information.

On average, bCAII treated with 1 has 1.2 labels per protein as determined by measurements of the intact protein. The labeling occurs at one of the two sites: His3 (95 ± 4\% ) and His63 (50 ± 20 \%) (Figure 2.4a, left) (Figure 2.8), as determined after proteolytic digestion and LC-MS/MS analysis (see Supplemental Information for description of covalent labeling-MS method). These residues line the wall of the sulfonamide binding site and are within the 18 Å reach of LD-TMAc warhead. bCAII treated with 2 has an average of 2.2 labels per protein with the labeling spread across five sites: His3, His63, Ser181, Ser171/Thr172, and Lys166 (Figure 2.4a, middle). His3 and His63 dominate the labeling of the protein by 2, with modification extents of 70 ± 20\% and 60 ± 30\%, respectively. Like with reagent 1, these residues are close to the binding site and within the 16 Å reach of 2. Ser181, Ser171/Thr172, and Lys 166 are minimally labeled with modification extents of 0.3 ± 0.2\%, 0.09 ± 0.08\%, and 0.06 ± 0.07\%, respectively. Ser171 and Thr172 neighbor each other on the same tryptic peptide, and either residue could be labeled as a result of ambiguous
assignments from the MS/MS data (Figure 2.9). Ser171 and Thr172 are on the fringes of probe 2’s reach, whereas Lys166 and Ser181 are farther away. It is likely that Ser181 and Lys166 labeling occurs in a non-ligand directed manner. Under the same conditions, there is only negligible extent of bCAII labeling with a control TMAc molecule without the sulfonamide ligand (Figure 2.7). It should be noted that Lys166 labeling is only statistically significant at 90% confidence, unlike the other residues that are significantly labeled by 2 at a 95% confidence. bCAII treated with 3 has an average of 2.7 labels per protein with the labeling occurring at His3, His63, Ser72, Ser171/Thr172, Lys250, and Lys44 (Figure 2.4a, right). Like with reagents 1 and 2, His3 and His63 dominate the labeling of the protein quantitatively, with modification extents of 80 ± 30% and 70 ± 10%,
Table 1. LD-TMAc targeted residues on bCAII

<table>
<thead>
<tr>
<th>Probe</th>
<th>Label Site</th>
<th>Modification %</th>
<th>SASA %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1†</td>
<td>His3</td>
<td>95 ± 4</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>His63</td>
<td>50 ± 20</td>
<td>17.3</td>
</tr>
<tr>
<td>2†</td>
<td>His3</td>
<td>70 ± 20</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>His63</td>
<td>60 ± 30</td>
<td>17.3</td>
</tr>
<tr>
<td></td>
<td>Ser181‡</td>
<td>0.3 ± 0.2</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Ser171-Thr172</td>
<td>0.09 ± 0.08</td>
<td>31.5$</td>
</tr>
<tr>
<td></td>
<td>Lys166*</td>
<td>0.06 ± 0.07</td>
<td>29.3</td>
</tr>
<tr>
<td>3‡</td>
<td>His3</td>
<td>80 ± 30</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>His63</td>
<td>70 ± 10</td>
<td>17.3</td>
</tr>
<tr>
<td></td>
<td>Ser72‡</td>
<td>30 ± 1</td>
<td>74.8</td>
</tr>
<tr>
<td></td>
<td>Lys250</td>
<td>0.2 ± 0.03</td>
<td>29.2</td>
</tr>
<tr>
<td></td>
<td>Lys44</td>
<td>0.1 ± 0.04</td>
<td>66.3</td>
</tr>
</tbody>
</table>

† Based on 6 replicates. ‡ Based on 3 replicates. § Ser171 and Thr172 average. # An outlier replicate removed at 95% confidence. * Statistically significant at 90% confidence. The rest of the data is statistically significant at 95% confidence.

<table>
<thead>
<tr>
<th>Probe</th>
<th>Label Site</th>
<th>Modification %</th>
<th>SASA %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1†</td>
<td>His3</td>
<td>95 ± 4</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>His63</td>
<td>50 ± 20</td>
<td>17.3</td>
</tr>
<tr>
<td>2†</td>
<td>His3</td>
<td>70 ± 20</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>His63</td>
<td>60 ± 30</td>
<td>17.3</td>
</tr>
<tr>
<td></td>
<td>Ser181‡</td>
<td>0.3 ± 0.2</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td>Ser171-Thr172</td>
<td>0.09 ± 0.08</td>
<td>31.5$</td>
</tr>
<tr>
<td></td>
<td>Lys166*</td>
<td>0.06 ± 0.07</td>
<td>29.3</td>
</tr>
<tr>
<td>3‡</td>
<td>His3</td>
<td>80 ± 30</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>His63</td>
<td>70 ± 10</td>
<td>17.3</td>
</tr>
<tr>
<td></td>
<td>Ser72‡</td>
<td>30 ± 1</td>
<td>74.8</td>
</tr>
<tr>
<td></td>
<td>Lys250</td>
<td>0.2 ± 0.03</td>
<td>29.2</td>
</tr>
<tr>
<td></td>
<td>Lys44</td>
<td>0.1 ± 0.04</td>
<td>66.3</td>
</tr>
</tbody>
</table>

† Based on 6 replicates. ‡ Based on 3 replicates. § Ser171 and Thr172 average. # An outlier replicate removed at 95% confidence. * Statistically significant at 90% confidence. The rest of the data is statistically significant at 95% confidence.

Table 2.1: LD-TMAc targeted residues on bCAII.

respectively (Table 1) (Figure 2.4b). Ser72 and Ser171/Thr172 are labeled with modification extents of 30 ± 1% and 0.3 ± 0.1%, respectively. These residues are somewhat distant from the sulfonamide binding site, but the ~24 Å long reach of 3 may allow their labeling. Lys250 and Lys44 are both minimally labeled (~0.1% labeling) and are likely too far away from the binding site to be labeled when the sulfonamide is bound (Figure 2.10). Thus, the labeling of these residues is likely indicative of non-ligand directed labeling. To give an idea on their reach, LD-TMAc reagents were superimposed on the crystal structure of sulfonamide bound bCAII (Figure 2.4a). Although statistically significant only at 90% confidence, His2, Ser28, and Ser64 are minimally labeled with modification extents of 2 ± 2%, 1 ± 0.5% and 0.5 ± 0.4%, respectively, and are all within reach of 3.

To better understand the modification extents of each amino acid, we calculated solvent accessible surface area (SASA) of bCAII crystal structure (PDB: 1V9E) using GetArea$ with a probe radius of 1.4 Å (Table 2.1). After closely examining the SASA of each residue modified by probe 3, it is possible to deduce that higher solvent accessibility results in higher modification extent. However, this trend is observed only for the proximal residues that are within the reach of the LD-TMAc probes. His, Lys, Ser, and Thr residues with the highest SASA in bCAII are not
labeled, highlighting proximity to the ligand binding site as a prerequisite (Figure 2.11). Owing to their binding site proximity, His3 and His63 are labeled by all three probes, and His3 has a higher modification extent in all three cases, despite the closer proximity of His63. This can be explained by the higher SASA of His3. The relatively high labeling extents of His63 and Ser72 can also be explained by their high SASA values. His2 with low solvent exposure faces competition with His3, which results in a low modification extent. Although Ser28 and Ser64 have little to no solvent exposure, their proximity to the binding might allow their minimal labeling. Additionally, although SASA values calculated using the crystal structure is useful to understand the protein structure, these values are not ideal to infer conclusions about the dynamic structure. Protein structural dynamics and LD-TMAc modifications on the protein structure might have facilitated the minimal labeling of these residues. Finally, despite being out of reach of the reagent, Lys250 and Lys44 have moderate to high solvent accessibility, allowing them to be minimally labeled in a non-ligand directed manner.

While there are other residues in the vicinity of sulfonamide binding pocket, factors such as LD-TMAc reagents’ steric, length, and flexibility may not allow their labeling. In addition, low nucleophilicity, SASA, presence of other nearby residues with higher nucleophilicity, better positioning, and orientation may add to these factors. Among these residues, Thr197 and Thr198 lie right next to the Zn ion in the active site. However, they are not solvent accessible (SASA < 20%), and because the LD-TMAc’s electrophilic site is in the opposite end of sulfonamide ligand, the reactive site is therefore not conformationally accessible to these residues. No labels are found on Tyr6, Tyr69, and Lys170, but they are in close vicinity of His3, Ser72, and Ser171, respectively, which were labeled by 3. No labels are found on Thr131 despite having 56.1% SASA, being within the reach of 3 and having no nearby labeled residues. Finally, analysis of residues targeted by the
minimal non-ligand directed labeling of the control TMAc molecule reveals that these residues minimally overlap with the residues targeted by ligand-directed labeling (Figure 2.7, 2.12). Therefore, the kinetically preferred residues by the TMAcs differ from the residues targeted by ligand-directed labeling.

These results show that LD-TMAcs are capable of comprehensive regiospecific labeling to afford ample information to elucidate the ligand/drug binding site. Affinity guided methods such as photoaffinity labeling and other LD approaches have been used to identify binding sites and allosteric sites,\(^{34-41}\) and the LD-TMAcs’ capability to label up to seven different residues around the binding pocket could make them a great complement to these methods.

To evaluate the effect of LD-TMAc modification on the enzymatic activity of bCAII, we conducted a chromogenic assay (Figure 2.4c). First, 5 μM bCAII was incubated with 10 μM 1, or 2 in PBS buffer, pH 7.4, 37 °C for two hours. After that, buffer exchange was carried out with fresh PBS to get rid of excess reagents and sulfonamide ligands. Then, 1 μM labeled bCAII was incubated with 2 mM nitrophenyl acetate, and absorbance at 400 nm was monitored over 90 seconds. Hydrolysis of nitrophenyl acetate by bCAII generates nitrophenol, which has a UV absorption peak at 400 nm. Compared to the steady increase of UV absorbance from the unmodified control bCAII activity, there is no significant change in absorbance for the reaction mixture of 1 modified bCAII, demonstrating the covalent inhibition by 1 as the result of sulfonamide occupying the active site. To our surprise, 2 modified bCAII in a way that caused the enzyme to lose most of its activity despite having an unoccupied active site. This could be attributed to the fact that ~64% of His63 is modified after the bCAII labeling reaction with 2; His63 plays an important role as a proton-transfer group in the catalytic mechanism of CA.\(^{42}\)
2.2.3 Target specific labeling of CA in whole cell lysate

To evaluate the labeling specificity in a complex mixture, we synthesized LD-TMAc reagent 4 that contains a pendant alkyne moiety (Figure 2.5a). This functionality can be used as a bio-orthogonal handle to conjugate fluorophores to specific protein(s) onto which the LD-TMAc is attached. To test this possibility, we first evaluated the fidelity of the protein functionalization and subsequent bio-orthogonal conjugation with soluble bCAII. Accordingly, after treating 5 μM of bCAII with 10 μM 4 for 30 min, 25 μM of an azido-Cy3 was added to the mixture for 60 min in the presence of a copper catalyst. MALDI-MS of the reaction mixture after each step indicates multi-site modification of bCAII with 4 and the corresponding shift in molecular weight following the copper-catalyzed conjugation of Cy3 (Figure 2.5b).

For specific labeling in a complex mixture, 2 μM of bCAII in human embryonic kidney (HEK) cell lysate was sequentially treated with 4 and azido-Cy3. The product mixture was analyzed using
SDS-PAGE and in-gel fluorescence (Figure 2.5c). With 5 μM probe (4), there is a single fluorescent band at ~30 kDa that corresponds to bCAII, with no other discernible fluorescent bands, which demonstrates the target specificity of 4. Also, no fluorescence is observed in the gel in three different control experiments: (i) absence of 4; (ii) absence of azido-Cy3; and (iii) in the presence of 20 μM competitive inhibitor EZA. To illustrate that a complex mixture of proteins is present in the cell lysate, Coomassie staining of the same gel was carried out, which shows all the protein bands in the cell lysate (Figure 2.5d). Together, these results demonstrate that the labeling is ligand-directed and can be achieved specifically in the target protein even in a complex mixture such as the cell lysate.

2.2.4 Live cell labeling of membrane CA.
Following the demonstration of labeling in cell lysates, we investigated the possibility of using
LD-TMAcs to label membrane proteins in live cells. For this purpose, MCF-7 cells that are known
to have membrane-bound human carbonic anhydrase XII (hCAXII) were used. Cells were cultured
and treated with 4 (1 μM) for 30 min in serum-free media. Then, a bio-orthogonal click reaction
was performed with AZDye647 for 10 min, followed by incubation with wheat germ agglutinin
(WGA) for membrane staining. Hoechst was used to stain the nucleus prior to imaging. Confocal
laser scanning microscopy (CLSM) images show that the cellular membrane is labeled with
AZDye647, as discerned by the overlap in fluorescence with WGA fluorescence (Figure 2.6a). To
further demonstrate that the membrane labeling is indeed due to the ligand-directed reaction with
the membrane hCA, the cell culture media was pre-incubated with the competitive inhibitor EZA.
In the presence of EZA, no discernible labeling of the membrane with AZDye647 is observed
(Figure 2.6b). These results show that LD-TMAcs can target specifically label CA XII on live cells
rapidly and efficiently.

2.3 Conclusions

In summary, we have developed fast and efficient LD-TMAc probes for target specific protein
labeling. We show here that: (i) unlike previously reported ligand directed methods, the highly
reactive nature of LD-TMAcs and the resultant ability to react with a diverse set of residues allow
labeling of multiple residues around the binding pocket; (ii) strategic placement of the ligand
moiety relative to the leaving group of the TMAc functionality determines whether the molecule
acts as a covalent inhibitor or a multi-site labeling reagent; (iii) the modular design of the LD-
TMAcs allows us to fine-tune their reaction kinetics; (iv) the labeling site is primarily determined
by conformational accessibility and proximity of the reactive warhead in the bound state; (v) the
specificity of LD-TMAc based labeling strategy is translated even in a complex mixture, such as in cell lysates; and (vi) the approach can be used to label specific membrane proteins in live cells.

Ligand-directed protein modification approaches have been used in target identification of bioactive compounds. Relaxing stringent amino acid selectivity of this strategy expands their applicability to a much broader range of proteins. Overall, we envision LD-TMAc will expand the repertoire of ligand-directed labeling approaches in a variety of applications, including in target identification, understanding of protein binding sites and investigation of membrane proteins.

2.4 Acknowledgements

LC-MS experiments were conducted by Zachary Kirsch.

This work was supported in part by the National Institutes of Health (NIH) under grant R35 GM136395 to ST R35 GM145272 to RWV. The authors wish to thank the UMass Amherst Institute of Applied Life Sciences Mass Spectrometry Core Facility (RRID:SCR_019063) for access to the Thermo Scientific Orbitrap Fusion mass spectrometer, which was funded by a National Institutes of Health grant S10OD010645. We thank UMass Amherst Institute of Applied Life Sciences Mass Spectrometry Core Facility staff Dr. Stephen Eyles and Dr. Cedric Bobst. We thank UMass Amherst Institute of Applied Life Sciences Light Microscopy Facility staff Dr. James Chambers and Maaya Ikeda.

2.5 Data availability

The raw mass spectrometric data can be accessed in Massive (MassIVE MSV000091614).

2.6 Supplementary Information

2.6.1 Materials and Instrumentation

All reagents were used as received from commercial sources unless otherwise mentioned. IH-
NMR, 13C-NMR spectra were recorded on a Bruker 400 MHz NMR spectrometer. UV-Vis spectra were recorded on PerkinElmer Lambda 35 UV/Vis spectrometer. Protein labeling and quantification were carried out on Biodrop microvolume UV-Vis spectrophotometer. MALDI-MS studies were carried out using Bruker UltraflleXtreme MALDI-TOF/TOF. All optical images were captured using a CrestV2-2xTIRF confocal microscope.

2.6.2 Covalent Labeling and Proteolytic Digestion for LC-MS/MS

Bovine carbonic anhydrase II (bCAII) was prepared in 10 mM pH 7.4 3-(N-morpholino)propanesulfonic acid (MOPS) buffer. LD-TMAc labeling of bCAII (5 μM) was initiated by adding LD-TMAc in 2:1 molar excess and the reaction was held at 37 °C for 1 hour. Stock solutions of LD-TMAc were prepared in dimethyl sulfoxide (DMSO) and the final volume of DMSO in the reaction solutions was less than 1% v/v in all experiments. Unreacted reagent was removed, and the protein was preconcentrated using 10 kDa molecular weight cutoff (MWCO) filters. The resulting solutions were incubated with 8 M urea for 20 minutes at room temperature to denature the protein. After unfolding, the urea concentration was lowered below 1.6 M by dilution with MOPS and the solution was preconcentrated using 10 kDa MWCO filters. The protein was then digested overnight at 37 °C using 50 μL of immobilized trypsin. To remove the trypsin and collect the digested peptides, the sample was centrifuged at 12,000 rpm and the supernatant was collected and flash frozen with liquid N₂ until analysis.

2.6.3 Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS)

For on-line LC-MS/MS analyses, 5 μL of digested protein was injected into a Thermo Scientific UltiMate 3000 RSLC capillary LC system (Waltham, MA). Separation of peptides was performed using a Thermo Scientific Acclaim™ PepMap™ RSLC C18 capillary column (15 cm x 300 μm, 2 μm particle size) with a flow rate of 4 μL/min. HPLC-grade water (solvent A) and acetonitrile (solvent B), each containing 0.1% formic acid, were used as mobile phase. A linear gradient of 5% B to 50% B over 50 minutes with a final wash of 95% B for 5 minutes was used. The column
compartment was held at 40°C.

Mass spectra were acquired on a Thermo Scientific Orbitrap Fusion mass spectrometer (Waltham, MA). The electrospray ionization source was operated in the positive mode with a needle voltage of 4000 V. Mass spectra were acquired on an Orbitrap analyzer with a resolution of 60,000 and tandem mass spectra were acquired using a linear quadrupole ion trap using collision induced dissociation with a normalized collision energy of 35%.

A custom software pipeline described previously and specifically designed for protein covalent labeling-mass spectrometry was utilized to quantify peak areas and determine sites of labeling. Search parameters included a precursor mass tolerance of 10 ppm and LD-TMAc modification at His, Lys, Ser, Thr, and Tyr as a variable modification. Residue labeling extent was calculated from the chromatographic peak areas of modified and unmodified peptides applying eq 1.

\[
\% \text{ labeling} = \frac{\sum_{i=1}^{n} \sum_{z=1}^{m} A_{i,z}^{\text{modified}}}{\sum_{i=1}^{n} \sum_{z=1}^{m} A_{i,z}^{\text{modified}} + \sum_{i=1}^{n} \sum_{z=1}^{m} A_{i,z}^{\text{unmodified}}} \times 100
\]

(1)

Where \( A_{i,z} \) is the peptide peak area from a given peptide, \( i \), that contains the residue of interest and includes all charge states detected, \( z \), for that peptide. Labeling extents are representative of the ratio of the chromatographic peak areas of peptides containing a given residue that are labeled at that residue to the total peak area of all peptides, labeled or unlabeled at that residue, that contain that site. 1-sample t-tests with a 95% confidence interval were used to assess if a given residue’s modification was statistically significant.

For intact labeling measurements using LC-MS, 1 µL of the labeled protein was injected into Thermo Scientific UltiMate 3000 RSLC capillary LC system (Waltham, MA) with a 5 µL bed volume Optimize Technologies OPTI-TRAP cartridge (Oregon City, OR). Mass spectra were acquired on a Bruker MicrOTOF mass spectrometer (Billerica, MA). The electrospray ionization
source was operated in the positive mode with a needle voltage of 4000 V. Mass spectra were viewed and analyzed using Bruker DataAnalysis.

2.6.4 Synthesis of compounds 1 to 6

Synthesis of compound 1a

Scheme 2.2: Synthesis of compound 1.

1,6-Hexanediol (20.00 g, 169.25 mmol) and Triethylamine (TEA) (17.13 g, 169.25 mmol) was dissolved in 400 mL Dichloromethane (DCM) in a round bottom flask in ice bath. Tert-Butyldimethylsilyl chloride (TBDMS) (12.75 g, 84.62 mmol) was added to the stirring reaction mixture dropwise, and the reaction was stirred overnight at room temperature. The reaction mixture was washed by water and brine sequentially in a separatory funnel. The organic layer was
collected, dried over anhydrous sodium sulfate, and subjected to purification by flash chromatography on silica gel with gradient of hexanes and ethyl acetate. Yield: 17.14 g, 22%. $^1$H NMR (400 MHz, CDCl$_3$) δ 3.65, 3.63-3.61 (t, 2H), 3.61-3.58 (t, 2H), 1.60-1.49 (m, 4H), 1.38-1.33 (m, 4H), 0.88 (s, 9H), 0.04 (s, 6H). $^{13}$C NMR (100 MHz, CDCl$_3$) δ 63.18, 62.95, 32.79, 32.77, 25.98, 25.62, 25.54, 18.37, -5.27. MS (m/z): [M+Na]$^+$ expected for C$_{12}$H$_{28}$O$_2$Si, 255.19; found, 255.1009 for [M+Na]$^+$.

*Synthesis of compound 1b*

4-Sulfamoylbenzoic acid (14.84 g, 73.74 mmol) and N-Ethyl-N’-(3-dimethylaminopropyl)carbodiimide hydrochloride (EDC HCl) (11.91 g, 92.17 mmol) was dissolved in 300 mL of dimethylformamide (DMF) in a round bottom flask. Then compound 1a (17.14 g, 73.74 mmol) and 4-Dimethylaminopyridine (4.5 g, 36.87 mmol) was added to the stirring reaction mixture sequentially. The reaction mixture was stirred overnight at room temperature. The reaction mixture was transferred to double neck round bottom flask and DMF was evaporated overnight by passing stream of air through the stirring reaction mixture. The reaction mixture was dissolved in ethyl acetate and washed by water and brine sequentially in a separatory funnel. The organic layer was collected, dried over anhydrous sodium sulfate, and subjected to purification by flash chromatography on silica gel with gradient of hexanes and ethyl acetate. Yield: 13.71 g, 45%. $^1$H NMR (400 MHz, CDCl$_3$) δ 8.16-8.13 (td, 2H), 7.99-7.96 (td, 2H), 5.24 (s, 2H), 4.36-4.33 (t, 2H), 3.62-3.59 (t, 2H), 1.82-1.75 (tt, 2H), 1.57-1.49 (m, 2H), 1.48-1.38 (m, 4H), 0.88 (s, 9H), 0.04 (6H). $^{13}$C NMR (100 MHz, CDCl$_3$) δ 165.22, 145.71, 134.24, 130.32, 126.44, 65.87, 63.07, 32.68, 28.64, 25.98, 25.83, 25.52, 18.37, -5.26. MS (m/z): [M+Na]$^+$ expected for C$_{19}$H$_{33}$NO$_5$Si, 438.18; found, 438.1173 for [M+Na]$^+$.

*Synthesis of compound 1c*
Compound 1b (13.71 g, 32.99 mmol) was dissolved in 200 mL of tetrahydrofuran (THF) in a round bottom flask, and 1 M Tetrabutylammonium fluoride (TBAF) in THF (39.58 mL, 39.58 mmol) was added to the stirring reaction mixture dropwise. The reaction mixture was stirred overnight at room temperature. After evaporation of THF, the reaction mixture was dissolved in ethyl acetate and washed by water and brine sequentially in a separatory funnel. The organic layer was collected, dried over anhydrous sodium sulfate, and subjected to purification by flash chromatography on silica gel with gradient of hexanes and ethyl acetate. Yield: 6.10 g, 61%. $^1$H NMR (400 MHz, CD$_3$OD) δ 8.19-8.15 (td, 2H), 8.03-8.00 (td, 2H), 4.39-4.36 (t, 2H), 3.59-3.56 (t, 2H), 1.86-1.79 (tt, 2H), 1.62-1.55 (tt, 2H), 1.55-1.42 (m, 4H). $^{13}$C NMR (100 MHz, CD$_3$OD) δ 165.37, 147.70, 133.35, 129.66, 125.96, 65.32, 61.43, 32.10, 28.33, 25.56, 25.21. MS (m/z): [M+Na]$^+$ expected for C$_{13}$H$_{19}$NO$_5$S, 324.1; found, 324.0092 for [M+Na]$^+$. 

Synthesis of compound 1e

Compound 1d (6.89 g, 20.24 mmol) was dissolved in 100 mL of DCM in a round bottom flask. Then oxalyl chloride (5.14 g, 40.48 mmol) and 3-5 drops of catalytic DMF were added to the stirring reaction mixture dropwise and sequentially. After 2 hours, reaction completion was monitored, and the solvent was removed under reduced pressure. In a separate round bottom flask, compound 1c (6.10 g, 20.24 mmol) and triethylamine (3.07 g, 30.36 mmol) were dissolved in 100 mL of THF. To the stirring reaction mixture in the ice bath, acyl chloride of compound 1c was added dropwise at 0 °C. The reaction mixture was stirred overnight at room temperature. After evaporation of THF, the reaction mixture was dissolved in ethyl acetate and washed with water and brine sequentially in a separatory funnel. The organic layer was collected, dried over anhydrous sodium sulfate, and subjected to purification by flash chromatography on silica gel with a gradient of hexanes and ethyl acetate. Yield: 4.28 g, 34%. $^1$H NMR (400 MHz, CDCl$_3$) δ 8.17-
8.14 (td, 2H), 8.00-7.97 (td, 2H), 7.68-7.65 (td, 4H), 7.45-7.35 (m, 6H), 6.32-6.30 (q, 1H), 6.10-
6.09 (q, 1H), 5.04 (s, 2H), 4.42-4.41 (t, 2H), 4.35-4.32 (t, 2H), 4.13-4.10 (t, 2H), 1.80-1.73 (p, 2H),
1.70-1.63 (p, 2H), 1.50-1.38 (m, 4H), 1.07 (s, 9H). $^1$H NMR (400 MHz, acetone-D6) δ 8.19-8.16
(td, 2H), 8.03-8.00 (td, 2H), 6.78 (s, 1H), 6.17-6.16 (q, 1H), 5.88-5.86 (q, 1H), 4.38-4.35 (t, 2H),
4.29-4.26 (m, 2H), 4.18-4.14 (t, 2H), 1.86-1.78 (tt, 2H), 1.75-1.69 (tt, 2H), 1.57-1.47 (m, 4H). $^{13}$C
NMR (100 MHz, acetone-D6) δ 165.55, 164.87, 148.07, 141.33, 133.45, 129.88, 126.26, 122.65,
65.14, 64.03, 61.44, 60.39, 32.72, 25.71, 25.47. MS (m/z): [M+Na]$^+$ expected for C$_{17}$H$_{23}$NO$_7$S,
408.12; found, 407.3480 for [M+Na]$^+$.

**Synthesis of compound 1f**

Compound 1e (4.28 g, 6.86 mmol) was dissolved in 75 mL of tetrahydrofuran (THF) in a round
bottom flask, and 1 M TBAF in THF (8.23 mL, 8.23 mmol) was added to the stirring reaction
mixture dropwise. The reaction mixture was stirred overnight at room temperature. After
evaporation of THF, the reaction mixture was dissolved in ethyl acetate and washed by water and
brine sequentially in a separatory funnel. The organic layer was collected, dried over anhydrous
sodium sulfate, and subjected to purification by flash chromatography on silica gel with gradient
of hexanes and ethyl acetate. **Yield:** 1.54 g, 58%. $^1$H NMR (400 MHz, acetone-D6) δ 8.19-8.16
(td, 2H), 8.03-8.00 (td, 2H), 6.78 (s, 1H), 6.17-6.16 (q, 1H), 5.88-5.86 (q, 1H), 4.38-4.35 (t, 2H),
4.29-4.26 (m, 2H), 4.18-4.14 (t, 2H), 1.86-1.78 (tt, 2H), 1.75-1.69 (tt, 2H), 1.57-1.47 (m, 4H). $^{13}$C
NMR (100 MHz, acetone-D6) δ 165.55, 164.87, 148.07, 141.33, 133.45, 129.88, 126.26, 122.65,
65.14, 64.03, 61.44, 60.39, 32.72, 25.71, 25.47. MS (m/z): [M+Na]$^+$ expected for C$_{33}$H$_{41}$NO$_7$Si,
646.24; found, 646.2027 for [M+Na]$^+$.

**Synthesis of compound 1g**

Compound 1f (1.54 g, 4.00 mmol) was dissolved in 40 mL of THF. Phosphorus tribromide (540
mg, 2.00 mmol) was added to the stirring reaction mixture under ice bath. The reaction mixture
was stirred at room temperature and completion of reaction was monitored by TLC after an hour.
After evaporation of THF, the reaction mixture was dissolved in DCM and transferred to separatory funnel. Then 2 M sodium bicarbonate was added to the mixture and the aqueous layer was extracted by DCM three times. The organic layers were collected, dried over anhydrous sodium sulfate, and subjected to purification by flash chromatography on silica gel with gradient of DCM and methanol. Yield: 970 mg, 54%. $^1$H NMR (400 MHz, CDCl$_3$) $\delta$ 8.18-8.15 (td, 2H), 8.01-7.98 (td, 2H), 6.32 (s, 1H), 5.94-5.93 (d, 1H), 5.01 (s, 2H), 4.38-4.35 (t, 2H), 4.24-4.21 (t, 2H), 4.17 (s, 2H), 1.85-1.79 (tt, 2H), 1.78-1.71 (tt, 2H), 1.62-1.48 (p, 4H). $^{13}$C NMR (100 MHz, CDCl$_3$) $\delta$ 165.31, 165.11, 145.87, 137.68, 134.40, 130.52, 129.30, 126.68, 65.82, 65.34, 29.57, 28.69, 28.61, 25.82, 25.79. MS (m/z): [M+Na]$^+$ expected for C$_{17}$H$_{22}$BrNO$_6$S, 470.04; found, 471.9670 for [M+Na]$^+$.

**Synthesis of compounds 1, 5 and 6**

Compound 1g (100 mg, 0.22 mmol) was dissolved in 2 mL of THF. To the stirring reaction mixture three equivalents of pyridine, 4-methoxypyridine, or 4-pyridinecarbonitrile dissolved in THF was added dropwise. After 3 hours precipitate was collected, dissolved in methanol, and precipitated in cold diethyl ether. Precipitation was repeated two more times to get the pure product. Yield ~75 mg, 64%. **Compound 1** $^1$H NMR (400 MHz, CD$_3$OD) $\delta$ 9.06-9.05 (d, 2H), 8.66-8.62 (t, 1H), 8.18-8.13 (m, 4H), 8.03-8.01 (dd, 2H), 6.66 (s, 1H), 6.33 (s, 1H), 5.52 (s, 2H), 4.38-4.35 (t, 2H), 4.20-4.17 (t, 2H), 1.84-1.76 (tt, 2H), 1.73-1.66 (p, 2H), 1.53-1.39 (m, 4H). $^{13}$C NMR (100 MHz, CD$_3$OD) $\delta$ 165.32, 164.44, 147.77, 146.19, 145.09, 133.64, 133.30, 132.9, 129.6, 127.96, 125.98, 65.31, 65.12, 61.58, 28.12, 27.99, 25.25, 25.15. MS (m/z): [M]$^+$ expected for C$_{22}$H$_{27}$N$_2$O$_6$S$^+$, 447.16; found, 448.0984 for [M]$^+$. 

55
Scheme 2.3: Synthesis of compound 2.

**Compound 5** $^1$H NMR (400 MHz, CD$_3$OD) δ 8.76-8.73 (td, 2H), 8.17-8.13 (td, 2H), 8.02-7.98 (td, 2H), 7.55-7.52 (td, 2H), 6.58 (s, 1H), 6.21 (s, 1H), 5.30 (s, 2H), 4.37-4.34 (t, 2H), 4.19-4.16 (t, 2H), 4.15 (s, 3H), 1.83-1.76 (p, 2H), 1.73-1.66 (p, 2H), 1.53-1.39 (m, 4H). $^{13}$C NMR (100 MHz, CD$_3$OD) δ 171.94, 165.33, 164.54, 147.77, 146.17, 134.07, 133.30, 132.34, 129.67, 125.98, 113.10, 65.23, 65.13, 59.57, 57.42, 28.13, 28.02, 25.27, 25.18. MS (m/z): [M]$^+$ expected for C$_{23}$H$_{29}$N$_2$O$_7$S$^+$, 477.17; found, 477.1290 for [M]$^+$.

**Compound 6** $^1$H NMR (400 MHz, DMSO-D$_6$) δ 9.37-9.35 (t, 2H), 8.74-8.72 (dd, 2H), 8.13-8.11 (dd, 2H), 7.97-7.95 (dd, 2H), 7.57 (s, 2H), 6.54 (s, 1H), 6.25 (s, 1H), 5.59 (s, 2H), 4.32-4.29 (t, 2H), 4.11-4.08 (t, 2H), 1.75-1.68 (p, 2H), 1.64-1.57 (p, 2H), 1.45-1.30 (m, 4H). $^{13}$C NMR (100 MHz, DMSO-D$_6$) δ 165.28, 164.56, 148.50, 147.19, 134.18, 133.65, 133.01, 131.49, 130.30, 128.19, 126.55, 115.24, 65.63, 65.57, 61.97, 28.44, 28.24, 25.49, 25.43. MS (m/z): [M]$^+$ expected for C$_{23}$H$_{26}$N$_3$O$_6$S$^+$, 472.15; found, 472.1752 for [M]$^+$.

### 2.6.5 Synthesis of compound 2

**Synthesis of compound 2a**

4-Pyridylacetic acid hydrochloride (500 mg, 2.88 mmol), N,N-Diisopropylethylamine (DIPEA) (1.12 g, 8.6 mmol) and O-(7-Azabenzotriazol-1-yl)-N,N,N',N'-tetramethyluronium hexafluorophosphate (HATU) (1.31 g, 3.46 mmol) was dissolved in 20 mL of DMF. After 30 minutes 4-(2-Aminoethyl) benzenesulfonylamide (577 mg, 2.88 mmol) in DMF was added dropwise to the stirring reaction mixture, and the reaction was stirred overnight at room temperature. Then,
the reaction mixture was transferred to double neck round bottom flask and DMF was evaporated overnight by passing stream of air through the stirring reaction mixture. The reaction mixture was dissolved in DCM and washed by brine twice in a separatory funnel. The organic layer was collected, dried over anhydrous sodium sulfate, and subjected to purification by flash chromatography on silica gel with gradient of DCM and methanol. Yield: 235 mg, 26%. \(^1\)H NMR (400 MHz, CD\(_3\)OD) \(\delta\) 8.46-8.45 (dd, 2H), 7.82-7.79 (td, 2H), 7.36-7.33 (td, 2H), 7.29-7.27 (dd, 2H), 3.52-3.51 (d, 2H), 3.50-3.48 (d, 2H), 2.93-2.88 (dd, 2H). \(^{13}\)C NMR (100 MHz, CD\(_3\)OD) \(\delta\) 172.15, 150.17, 147.53, 145.27, 143.28, 130.56, 130.37, 127.39, 126.87, 126.15, 41.57, 40.93, 36.14.

**Synthesis of compound 2**

Compound 2a (201 mg, 0.63 mmol) was dissolved in 1.5 mL of DMF. To the stirring reaction mixture compound 2b (77 mg, 4.32 mmol) dissolved in DMF were added dropwise. After 3 hours, the reaction mixture was precipitated in cold ether and the precipitate was collected. The collected precipitate was dissolved in methanol and precipitated in cold diethyl ether. Precipitation was repeated two more times to get the pure product. Yield 127 mg, 34%. \(^1\)H NMR (400 MHz, MeOD) \(\delta\) 8.90-8.87 (td, 2H), 7.82-7.80 (dd, 2H), 7.78-7.75 (td, 2H), 7.39-7.36 (td, 2H), 6.68 (s, 1H), 6.37-
6.36 (t, 1H), 5.48-5.47 (d, 2H), 3.81 (s, 2H), 3.76 (s, 3H), 3.59-3.56 (t, 2H), 2.95-2.92 (t, 2H). $^{13}$C NMR (100 MHz, MeOD) $\delta$ 168.09, 164.94, 156.79, 144.30, 143.72, 141.70, 133.54, 133.44, 129.30, 128.21, 125.77, 61.04, 51.65, 41.58, 39.79, 34.64. MS (m/z): [M]$^{+}$ expected for C$_{20}$H$_{24}$N$_{3}$O$_{5}$S$^{+}$, 418.14; found, 418.1065 for [M]$^{+}$.

2.6.6 Synthesis of compound 3

**Synthesis of compound 3a**

4-Pyridylacetic acid hydrochloride (3 g, 17.28 mmol), DIPEA (6.70 g, 51.8 mmol) and HATU (7.88 g, 20.74 mmol) was dissolved in 40 mL of DMF in a round bottom flask and stirred at room temperature. After 30 minutes N-Boc-1,6-hexanediamine (3.74 g, 17.28 mmol) in DMF was added dropwise to the stirring reaction mixture, and the reaction was stirred overnight at room temperature. Then, the reaction mixture was transferred to double neck round bottom flask and DMF was evaporated overnight by passing stream of air through the stirring reaction mixture. The reaction mixture was dissolved in DCM and washed by brine twice in a separatory funnel. The organic layer was collected, dried over anhydrous sodium sulfate, and subjected to purification by flash chromatography on silica gel with gradient of DCM and methanol. Yield: 4.2 g, 72%.

$^{1}$H NMR (400 MHz, CDCl$_3$) $\delta$ 8.59-8.58 (dd, 2H), 7.27-7.25 (dd, 2H), 5.84 (s, 1H), 4.57 (s, 1H), 3.56 (s, 2H), 3.27-3.22 (q, 2H), 3.14-3.09 (q, 2H), 1.77-1.47 (m, 4H), 1.45 (s, 9H), 1.35-1.27 (m, 4H).

$^{13}$C NMR (100 MHz, CDCl$_3$) $\delta$ 168.97, 156.19, 149.94, 144.41, 124.56, 55.22, 43.02, 39.93, 39.31, 29.98, 29.12, 28.43, 25.82, 25.67.

**Synthesis of compound 3b**

Compound 3a (4.2 g, 12.52 mmol) was dissolved in 50 mL of 1:1 Trifluoroacetic acid (TFA)/Methanol and stirred overnight at room temperature. Solvents were evaporated under reduced pressure. TFA was removed by co-evaporation with methanol twice. Then, the residue
was left under a high vacuum overnight. In a separate round bottom flask 4-Sulfamoylbenzoic acid (2.52 g, 12.52 mmol), DIPEA (4.85 g, 37.56 mmol) and HATU (5.71 g, 15.02 mmol) was dissolved in 40 mL of DMF in a round bottom flask and stirred at room temperature. After 30 minutes deprotected residue of compound 3a dissolved in DMF was added dropwise to the stirring reaction mixture, and the reaction was stirred overnight at room temperature. Then, the reaction mixture was transferred to double neck round bottom flask and DMF was evaporated overnight by passing stream of air through the stirring reaction mixture. The reaction mixture was dissolved in DCM and washed by brine twice in a separatory funnel. The organic layer was collected, dried over anhydrous sodium sulfate, and subjected to purification by flash chromatography on silica gel with gradient of DCM and methanol. Yield: 3.1 g, 59%. \( ^1\)H NMR (400 MHz, MeOD) \( \delta \) 8.48-8.46 (dd, 2H), 8.00-7.95 (m, 4H), 7.40-7.38 (dd, 2H), 3.58 (s, 2H), 3.41-3.37 (t, 2H), 3.24-3.20 (t, 2H), 1.67-1.60 (p, 2H), 1.59-1.52 (p, 2H), 1.46-1.38 (m, 4H). \( ^{13}\)C NMR (100 MHz, MeOD) \( \delta \) 170.51, 167.36, 148.55, 146.50, 146.24, 137.80, 127.52, 125.92, 124.75, 41.62, 39.55, 39.04, 28.90, 28.82, 26.16, 26.08. MS (m/z): \([M+Na]^+ \) expected for C\(_{20}\)H\(_{26}\)NO\(_4\)S, 441.17; found, 442.2706 for \([M+Na]^+ \).

**Synthesis of compound 3**

Compound 3b (50 mg, 0.12 mmol) was dissolved in 1 mL of DMF. To the stirring reaction mixture two equivalents of compound 2b (43 mg, 0.24 mmol) dissolved in DMF were added dropwise. After 3 hours, the reaction mixture was precipitated in cold ether and the precipitate was collected. The collected precipitate was dissolved in methanol and precipitated in cold diethyl ether. Precipitation was repeated two more times to get the pure product. Yield ~35 mg, 49%. \( ^1\)H NMR (400 MHz, MeOD) \( \delta \) 8.95-8.92 (td, 2H), 8.64-8.62 (t, 1H), 8.38-8.35 (t, H), 8.04-8.02 (dd, 2H), 7.99-7.94 (m, 4H), 6.64 (s, 1H), 6.34-6.33 (t, 1H), 5.47 (s, 2H), 3.92 (s, 2H), 3.74 (s, 3H), 3.42-
3.37 (q, 2H), 3.25-3.20 (q, 2H), 1.68-1.61 (p, 2H), 1.59-1.53 (p, 2H), 1.45-1.40 (m, 4H). \(^{13}\)C NMR (100 MHz, MeOD) \(\delta\) 166.59, 165.83, 163.36, 155.54, 144.74, 142.82, 136.27, 132.04, 131.67, 126.98, 126.01, 124.39, 59.43, 50.10, 39.84, 37.96, 37.67, 27.33, 27.22, 24.58, 24.53. MS (m/z): [M]\(^+\) expected for C\(_{25}\)H\(_{33}\)N\(_4\)O\(_6\)S\(^+\), 517.21; found, 517.1534 for [M]\(^+\).

2.6.7 Synthesis of compound 4

*Synthesis of compound 4a*

\[
\begin{align*}
\text{HO-C} & \quad \text{Cl} \quad \text{Cl} \quad \text{HO-
NH}_2 \quad \text{TEA, THF} \\
\text{DMF, DCM} & \quad \text{DMF, DCM} & \quad \text{DMF, DCM} & \quad \text{DMF, DCM}
\end{align*}
\]

Scheme 2.5: Synthesis of compound 4.

4-Pentynoic acid (3 g, 30.58 mmol) was dissolved in 40 mL of DCM in a round bottom flask. Then oxalyl chloride (7.76 g, 61.16 mmol) and 3-5 drops of catalytic DMF were added to the stirring reaction mixture sequentially and dropwise. After 2 hours reaction completion was monitored, and solvent was removed under reduced pressure to get 4-Pentynoic acyl chloride. In a separate round
bottom flask 1-Amino-2-methyl-2-propanol (4.09 g, 45.87 mmol) and triethylamine (6.19 g, 61.16 mmol) were dissolved in 200 mL of THF. To the stirring reaction mixture in ice bath, 4-Pentynoic acyl chloride in THF was added dropwise at 0 °C. The reaction mixture was stirred overnight at room temperature. After evaporation of THF, the reaction mixture was dissolved in ethyl acetate and washed by saturated sodium bicarbonate thrice in a separatory funnel. The organic layer was collected, dried over anhydrous sodium sulfate, and left under high vacuum overnight to get the pure product. Yield: 2.20 g, 43%. 1H NMR (400 MHz, CDCl3) δ 6.20 (s, 1H), 3.31-3.30 (d, 2H), 2.59-2.55 (m, 2H), 2.49-2.44 (m, 2H), 2.04-2.03 (t, 1H), 1.25 (s, 6H). 13C NMR (100 MHz, CDCl3) δ 172.00, 83.00, 70.86, 69.50, 50.30, 35.43, 27.27, 15.03. MS (m/z): [M+H]+ expected for C9H13NO2, 169.11; found, 170.1323 for [M+H]+.

Synthesis of compound 4b

Compound 1d (3.00 g, 8.81 mmol) was dissolved in 40 mL of DCM in a round bottom flask. Then oxalyl chloride (2.24 g, 17.62 mmol) and 3-5 drops of catalytic DMF were added to the stirring reaction mixture dropwise and sequentially. After 2 hours reaction completion was monitored, and solvent was removed under reduced pressure to get acyl chloride of compound 1d. In a separate round bottom flask compound 4a (2.20 g, 13.00 mmol) and triethylamine (3.95 g, 39.00 mmol) were dissolved in 30 mL of THF. To the stirring reaction mixture in ice bath, acyl chloride of compound 1c was added dropwise at 0 °C. The reaction mixture was stirred overnight at room temperature. After evaporation of THF, the reaction mixture was dissolved in ethyl acetate and washed by water and brine sequentially in a separatory funnel. The organic layer was collected, dried over anhydrous sodium sulfate, and subjected to purification by flash chromatography on silica gel with gradient of hexanes and ethyl acetate. Yield: 2.6 g, 60%. 1H NMR (400 MHz, CDCl3) δ 7.68-7.65 (m, 4H), 7.46-7.37 (m, 6H), 6.25-6.24 (q, 1H), 6.07-6.05 (q, 1H), 4.37-4.36 (t,
1H), 3.54-3.52 (d, 2H), 2.51-2.47 (dt, 2H), 2.37-2.34 (dt, 2H), 1.93-1.91 (t, 1H), 1.43 (s, 6H), 1.08 (s, 9H). $^{13}$C NMR (100 MHz, CDCl$_3$) δ 171.05, 165.31, 140.32, 135.46, 133.15, 129.85, 127.79, 124.15, 82.90, 82.64, 69.42, 62.26, 47.95, 35.47, 26.81, 24.06, 19.29, 14.94. MS (m/z): [M+Na]$^+$ expected for C$_{29}$H$_{37}$NO$_4$Si, 514.25; found, 514.1833 for [M+Na]$^+$.

**Synthesis of compound 4c**

Compound 4b (2.60 g, 5.29 mmol) was dissolved in 20 mL of THF in a round bottom flask, and 1 M TBAF in THF (6.35 mL, 6.35 mmol) was added to the stirring reaction mixture dropwise. The reaction mixture was stirred overnight at room temperature. After evaporation of THF, the reaction mixture was dissolved in ethyl acetate and washed by water and brine sequentially in a separatory funnel. The organic layer was collected, dried over anhydrous sodium sulfate, and subjected to purification by flash chromatography on silica gel with gradient of hexanes and ethyl acetate. Yield: 0.54 g, 40%. $^1$H NMR (400 MHz, CDCl$_3$) δ 6.47 (s, 1H), 6.18-6.18 (d, 1H), 5.78-5.77 (q, 1H), 4.29-4.28 (d, 2H), 3.58-3.56 (d, 2H), 2.55-2.51 (m, 2H), 2.45-2.41 (t, 2H), 1.99-1.98 (t, 1H), 1.50 (s, 6H). $^{13}$C NMR (100 MHz, CDCl$_3$) δ 171.29, 165.71, 140.50, 126.25, 83.02, 82.89, 69.47, 62.78, 48.04, 35.53, 23.97, 14.99. MS (m/z): [M+Na]$^+$ expected for C$_{13}$H$_{19}$NO$_4$, 276.13; found, 276.1157 for [M+Na]$^+$.

**Synthesis of compound 4d**

Compound 4c (0.54 g, 2.13 mmol) was dissolved in 20 mL of THF. Phosphorus tribromide (0.29 g, 1.06 mmol) was added to the stirring reaction mixture under ice bath. The reaction mixture was stirred at room temperature and completion of reaction was monitored by TLC after an hour. After evaporation of THF, the reaction mixture was dissolved in DCM and transferred to separatory funnel. Then 2 M sodium bicarbonate was added to the mixture and the aqueous layer was extracted by DCM three times. The organic layers were collected, dried over anhydrous sodium
sulfate, and subjected to purification by flash chromatography on silica gel with gradient of DCM and methanol. Yield: 0.36 g, 53%. $^1$H NMR (500 MHz, CDCl$_3$) δ 6.31-6.31 (d, 1H), 5.91-5.91 (d, 1H), 4.19-4.19 (d, 2H), 3.63-3.62 (d, 2H), 2.59-2.55 (m, 2H), 2.48-2.45 (dt, 2H), 2.01-2.00 (t, 1H), 1.55 (s, 6H). $^{13}$C NMR (100 MHz, CDCl$_3$) δ 171.12, 163.96, 138.20, 129.40, 83.51, 82.97, 69.37, 48.02, 35.52, 30.20, 23.76, 14.95. MS (m/z): [M+H]$^+$ expected for C$_{13}$H$_{18}$BrNO$_3$, 316.05; found, 316.0489 for [M+H]$^+$. 

**Synthesis of compound 4**

Compound 3b (30 mg, 0.072 mmol) was dissolved in 1 mL of DMF. To the stirring reaction mixture three equivalents of compound 4d (45.37 mg, 0.143 mmol) dissolved in DMF were added dropwise. After 3 hours, the reaction mixture was precipitated in cold ether and the precipitate was collected. The collected precipitate was dissolved in methanol and precipitated in cold diethyl ether. Precipitation was repeated two more times to get the pure product. Yield 24 mg, 46%. $^1$H NMR (400 MHz, MeOD) δ 8.97-8.96 (dd, 2H), 8.05-8.03 (dd, 2H), 8.01-7.95 (m, 4H), 6.62 (s, 1H), 6.24 (s, 1H), 5.42 (s, 2H), 3.92 (s, 2H), 3.54-3.48 (m, 3H), 3.43-3.39 (t, 2H), 3.26-3.23 (t, 2H), 2.49-2.44 (m, 2H), 2.42-2.38 (m, 2H), 2.29-2.28 (t, 1H), 1.70-1.62 (p, 2H), 1.60-1.54 (p, 2H), 1.44 (s, 6H), 1.44-1.41 (m, 4H), 1.22-1.18 (t, 2H). $^{13}$C NMR (100 MHz, MeOD) δ 172.92, 168.37, 167.37, 163.65, 160.48, 146.23, 145.00, 137.80, 134.54, 133.37, 127.57, 125.94, 125.43, 96.36, 83.50, 82.43, 69.09, 61.28, 46.54, 39.52, 39.10, 34.54, 28.85, 28.78, 26.10, 26.03, 22.72, 14.36. MS (m/z): [M]$^+$ expected for C$_{33}$H$_{44}$N$_5$O$_7$S$^+$, 654.80; found, 654.2219 for [M]$^+$. 

**Compound 1d, compound 2a** and the control TMAc (Figure 2.7a) were synthesized using the procedure provided in the previous literature.$^{27}$

**2.6.8 Supplementary figures**
Figure 2.7: Structure of LD-TMAc reagents and characterization of in-vitro bCAll labeling. (A) Molecular structure of LD-TMAc reagents 1-3 and control LD-TMAc. ESI-MS (24+ charge state) of bCAll (5 μM) incubated with (B) 1 (10 μM), 10 mM pH 7.4 MOPS buffer, 37 °C, 1 h, (C) 2, 3, or control TMAc (10 μM), 10 mM pH 7.4 MOPS buffer, 37 °C, 1 h. Native bCAll has molecular weight of ~29 KDa. Reagent 1 modification has an adduct of ~368 Da. Reagents 2, 3, and control TMAc has modification of ~99 Da. Number of modifications by the LD-TMAc reagents are indicated above each new protein peak.
Figure 2.8: MALDI-MS of bCAII (5 μM) incubated with 1 – 6 (10 μM), 100 μM EZA, PBS buffer, pH 7.4, 37 °C, 2 h. No protein modification is observed for 2 – 4. Only 11%, 7% and 28% of bCAII was modified by 1, 5, and 6 respectively.
Figure 2.9: MS/MS of LD-TMAc labeled S171-N185 peptide. LD-TMAc labeling on this peptide leads to unambiguous assignment of modification site.

Figure 2.10: Crystal structure of bCAll (PDB: 1V9E) with LD-TMAc reagent 3 structure (magenta) superimposed on the zinc coordinated sulfonamide binding site.
### Table 2.11: Solvent accessible surface area (SASA) of Serine, Threonine, Histidine and Lysine on bCAII crystal structure (PDB: 1V9E) calculated using GetArea\textsuperscript{33} with a probe radius of 1.4 Å.

<table>
<thead>
<tr>
<th>Residue</th>
<th>Sequence number</th>
<th>SASA %</th>
</tr>
</thead>
<tbody>
<tr>
<td>SER</td>
<td>1</td>
<td>81.1</td>
</tr>
<tr>
<td>SER</td>
<td>28</td>
<td>0</td>
</tr>
<tr>
<td>SER</td>
<td>55</td>
<td>0</td>
</tr>
<tr>
<td>SER</td>
<td>64</td>
<td>0.2</td>
</tr>
<tr>
<td>SER</td>
<td>72</td>
<td>74.8</td>
</tr>
<tr>
<td>SER</td>
<td>98</td>
<td>48.5</td>
</tr>
<tr>
<td>SER</td>
<td>99</td>
<td>59.9</td>
</tr>
<tr>
<td>SER</td>
<td>104</td>
<td>0</td>
</tr>
<tr>
<td>SER</td>
<td>164</td>
<td>43.4</td>
</tr>
<tr>
<td>SER</td>
<td>171</td>
<td>30.3</td>
</tr>
<tr>
<td>SER</td>
<td>181</td>
<td>30</td>
</tr>
<tr>
<td>SER</td>
<td>195</td>
<td>0</td>
</tr>
<tr>
<td>SER</td>
<td>204</td>
<td>2.1</td>
</tr>
<tr>
<td>SER</td>
<td>215</td>
<td>43.2</td>
</tr>
<tr>
<td>SER</td>
<td>217</td>
<td>9.8</td>
</tr>
<tr>
<td>SER</td>
<td>218</td>
<td>54.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Residue</th>
<th>Sequence number</th>
<th>SASA %</th>
</tr>
</thead>
<tbody>
<tr>
<td>THR</td>
<td>34</td>
<td>40.7</td>
</tr>
<tr>
<td>THR</td>
<td>54</td>
<td>36.2</td>
</tr>
<tr>
<td>THR</td>
<td>84</td>
<td>100</td>
</tr>
<tr>
<td>THR</td>
<td>86</td>
<td>26.7</td>
</tr>
<tr>
<td>THR</td>
<td>107</td>
<td>4.4</td>
</tr>
<tr>
<td>THR</td>
<td>124</td>
<td>23.7</td>
</tr>
<tr>
<td>THR</td>
<td>131</td>
<td>56.1</td>
</tr>
<tr>
<td>THR</td>
<td>167</td>
<td>21.2</td>
</tr>
<tr>
<td>THR</td>
<td>172</td>
<td>32.7</td>
</tr>
<tr>
<td>THR</td>
<td>191</td>
<td>0.4</td>
</tr>
<tr>
<td>THR</td>
<td>197</td>
<td>2.2</td>
</tr>
<tr>
<td>THR</td>
<td>198</td>
<td>18.1</td>
</tr>
<tr>
<td>THR</td>
<td>206</td>
<td>23.1</td>
</tr>
<tr>
<td>THR</td>
<td>226</td>
<td>34.5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Residue</th>
<th>Sequence number</th>
<th>SASA %</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIS</td>
<td>2</td>
<td>11.3</td>
</tr>
<tr>
<td>HIS</td>
<td>3</td>
<td>81</td>
</tr>
<tr>
<td>HIS</td>
<td>9</td>
<td>84.7</td>
</tr>
<tr>
<td>HIS</td>
<td>14</td>
<td>42.6</td>
</tr>
<tr>
<td>HIS</td>
<td>16</td>
<td>38</td>
</tr>
<tr>
<td>HIS</td>
<td>63</td>
<td>17.3</td>
</tr>
<tr>
<td>HIS</td>
<td>93</td>
<td>12.9</td>
</tr>
<tr>
<td>HIS</td>
<td>95</td>
<td>0.1</td>
</tr>
<tr>
<td>HIS</td>
<td>106</td>
<td>0</td>
</tr>
<tr>
<td>HIS</td>
<td>118</td>
<td>1.4</td>
</tr>
<tr>
<td>HIS</td>
<td>121</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Residue</th>
<th>Sequence number</th>
<th>SASA %</th>
</tr>
</thead>
<tbody>
<tr>
<td>LYS</td>
<td>8</td>
<td>89.9</td>
</tr>
<tr>
<td>LYS</td>
<td>17</td>
<td>66.1</td>
</tr>
<tr>
<td>LYS</td>
<td>35</td>
<td>100</td>
</tr>
<tr>
<td>LYS</td>
<td>44</td>
<td>66.3</td>
</tr>
<tr>
<td>LYS</td>
<td>75</td>
<td>32.3</td>
</tr>
<tr>
<td>LYS</td>
<td>79</td>
<td>48.4</td>
</tr>
<tr>
<td>LYS</td>
<td>111</td>
<td>53.6</td>
</tr>
<tr>
<td>LYS</td>
<td>112</td>
<td>44.2</td>
</tr>
<tr>
<td>LYS</td>
<td>125</td>
<td>55.5</td>
</tr>
<tr>
<td>LYS</td>
<td>147</td>
<td>59</td>
</tr>
<tr>
<td>LYS</td>
<td>157</td>
<td>42.7</td>
</tr>
<tr>
<td>LYS</td>
<td>166</td>
<td>29.3</td>
</tr>
<tr>
<td>LYS</td>
<td>168</td>
<td>27</td>
</tr>
<tr>
<td>LYS</td>
<td>170</td>
<td>53.7</td>
</tr>
<tr>
<td>LYS</td>
<td>211</td>
<td>23</td>
</tr>
<tr>
<td>LYS</td>
<td>223</td>
<td>32.4</td>
</tr>
<tr>
<td>LYS</td>
<td>250</td>
<td>29.2</td>
</tr>
<tr>
<td>LYS</td>
<td>259</td>
<td>92.7</td>
</tr>
</tbody>
</table>

**Figure 2.11:** Solvent accessible surface area (SASA) of Serine, Threonine, Histidine and Lysine on bCAII crystal structure (PDB: 1V9E) calculated using GetArea\textsuperscript{33} with a probe radius of 1.4 Å.

**Figure 2.12:** Table of Control TMAc targeted residues on bCAII.
2.6.9 Tandem Mass Spectra of the peptide modifications

![Tandem Mass Spectra of the peptide modifications](image)

**Figure 2.13:** H3 modification by LD-TMAc 1 (+367 Da)

![Tandem Mass Spectra of the peptide modifications](image)

**Figure 2.14:** H63 modification by LD-TMAc 1 (+367 Da)
Figure 2.15: H3 modification by LD-TMAc 2 (+98 Da)

Figure 2.16: H63 modification by LD-TMAc 2 (+98 Da)
Figure 2.17: K166 modification by LD-TMAc 2 (+98 Da)

Figure 2.18: S182 modification by LD-TMAc 2 (+98 Da)
Figure 2.19: T172 modification by LD-TMAc 2 (+98 Da)
Figure 2.20: S64 modification by LD-TMAc 3 (+98 Da)

Figure 2.21: S72 modification by LD-TMAc 3 (+98 Da)
Figure 2.22: S28 modification by LD-TMAc 3 (+98 Da)

Figure 2.23: S171 modification by LD-TMAc 3 (+98 Da)
Figure 2.24: K44 modification by LD-TMAc 3 (+98 Da)

Figure 2.25: K250 modification by LD-TMAc 3 (+98 Da)
**Figure 2.26:** H3 modification by LD-TMAc 3 (+98 Da)

**Figure 2.27:** H2 and H3 modification by LD-TMAc 3 (+98 Da + 98 Da, double modification)
2.6.10 Nuclear Magnetic Resonance (NMR) of the final compounds

Figure 2.28: HNMR of compound 1.

Figure 2.29: $^{13}$C NMR of compound 1.
Figure 2.30: HNMR of compound 2.

Figure 2.31: $^{13}$C NMR of compound 2.
**Figure 2.32:** HNMR of compound 3.

**Figure 2.33:** $^{13}$C NMR of compound 3.
Figure 2.34: HNMR of compound 4.

Figure 2.35: $^{13}$C NMR of compound 4.
Figure 2.36: HNMR of compound 5.

Figure 2.37: $^{13}$C NMR of compound 5.
Figure 2.38: HNMR of compound 6.

Figure 2.39: $^{13}$C NMR of compound 6.
2.7 References


(42) Tu, C.; Silverman, D. N.; Forsman, C.; Jonsson, B. H.; Lindskog, S. Role of Histidine 64 in the Catalytic Mechanism of Human Carbonic Anhydrase II Studied with a Site-Specific Mutant.


3.1 Introduction

There are ~20000 to 25000 proteins in the human body with various functions in structural building blocks, defense mechanisms, molecular transport, signal transduction, and catalytic activities. Proteostasis is the balance of protein concentration, structural integrity, and subcellular localization. When the delicate balance of proteostasis is perturbed, resulting in abnormally high protein activity, it can lead to substantial systemic complications and diseases. Traditional
therapies counteracted this imbalance with small therapeutic molecules that bind and render these biomarker proteins inactive. However, only ~ 7% of the proteome is estimated to be druggable.\textsuperscript{2}

In recent years, targeted protein degradation (TPD) of proteins has emerged as an alternative therapy paradigm. Compared to the traditional drugs that only bind and inactivate a single protein, targeted protein degraders bind the target protein and send them down their natural degradation pathway.\textsuperscript{3} Additionally, targeted degraders are often catalytic, meaning they are not consumed and can perform repeated rounds of protein degradation. Since targeted protein degraders completely obliterate their targets, they can be effective even against proteins long considered undruggable.\textsuperscript{4,5}

\textbf{Figure 3.2:} HyT LD-TMAs’ mechanism of action and possible pathways of protein degradation following hydrophobic tagging.
PROTACs are among the most popular targeted protein degraders; there were at least 20 of them in clinical trials at the end of 2022 (Figure 3.1 A).\textsuperscript{5,6} They are heterobifunctional small molecules that bring together the target protein and E3 liga, resulting in ubiquitylation and proteasomal digestion of the target protein. Although they have been demonstrated effective for various cytosolic proteins, they are ineffective against extracellular and membrane proteins.\textsuperscript{3} Furthermore, out of ~600 E3 ligases in the human body, only ~2 % can be utilized by PROTACs.\textsuperscript{7} To address these limitations, strategies such as LYTACs,\textsuperscript{8} AUTACs, ATTECs, and AbTACs have been developed for TPD of membrane and extracellular proteins by lysosomal pathway.\textsuperscript{3} Hydrophobic tagging (HyT) molecules are another class of degraders that act by binding and tagging their target proteins by hydrophobic moieties such as adamantane,\textsuperscript{9} carboranes,\textsuperscript{10} Boc\textsubscript{3}Arg\textsuperscript{11} and norbornene (Figure 3.1 B).\textsuperscript{12} Consequently, the tagged proteins are denatured or recognized by chaperones and sent down the proteasomal degradation pathway.\textsuperscript{13,14} Although the HyT strategy has been demonstrated effective against various targets such as mutant Huntington protein,\textsuperscript{15} tau protein\textsuperscript{16} and androgen receptors,\textsuperscript{17} sometimes it can be ineffective.\textsuperscript{18} Additionally, the hydrophobicity of the HyT degraders can cause them nonspecifically associate with off-target proteins.\textsuperscript{19} To address this limitation in the HyT TPD strategy, we introduce HyT Ligand-Directed Triggerable Michael Acceptors (LD-TMACs) for targeted degradation of membrane proteins.

In the previous chapter, we demonstrated fast, selective, and multisite protein labeling by LD-TMAcs. Herein, by installing hydrophobic adamantane to the ester group of the α, β unsaturated carbonyl Michael Acceptor, we have functionalized LD-TMAcs as HyT targeted protein degraders (Figure 3.2). Compared to the other HyT degraders in the literature that tag their target proteins with a single hydrophobic moiety, HyT LD-TMAcs can covalently install multiple hydrophobic moieties on a single target protein. We hypothesize that multiple hydrophobic tags concentrated
around the ligand binding site will be more effective in misfolding and degrading the target proteins, especially ones with large molecular weight. The positive charge on the LD-TMAc structure gives them better water solubility, potentially resulting in less nonspecific association and only the hydrophobic part of the molecule is left on the target protein after the protein labeling reaction. In previous studies, hydrophobically tagged cytosolic proteins or membrane proteins with hydrophobic tags on their cytosolic domains were shown to be degraded by the proteasomal pathway. However, to the best of our knowledge, extracellular hydrophobic tagging of membrane proteins and their fate afterward remains unexplored. Cells employ multiple proteostasis pathways to regulate the quality and quantity of the membrane proteins, including secretion via exosomes, lysosomal digestion, ectodomain shedding, and endoplasmic reticulum-associated degradation (ERAD) that depends on the proteasome. This chapter also investigates our hypothesized protein degradation mechanisms following HyT by LD-TMAc (Figure 3.2).

3.2 Results and Discussion

3.2.1 LD-TMAc Hydrophobic tagging and degradation of CAXII

To test their capability, we chose carbonic anhydrase (CA) as a model membrane protein. We functionalized the ester group of LD-TMAc 3 from the previous chapter with hydrophobic adamantane to get CA-targeted HyT LD-TMAc (Figure 3.3 a). Carbonic anhydrase catalyzes the conversion of CO₂ into carbonic acid, and CA inhibitors are clinically used in treating glaucoma, tumor, and obesity and as antimicrobials or antifungals. To characterize the protein labeling by HyT LD-TMAcs, bCAII (5 μM) was incubated with 10 μM HyT LD-TMAcs for 2 hours in PBS buffer at 37 °C. The extent of protein labeling was evaluated using matrix-assisted laser desorption ionization time of flight mass spectrometry (MALDI-TOF-MS) (Figure 3.3 b). After the protein labeling reaction, three new peaks with a separation of ~ 218 Da corresponding to adamantane
Figure 3.3: a) Structure of HyT LD-TMAc; b) MALDI-MS spectrum of control bCAII and hydrophobically tagged bCAII; c) Western blotting analysis of CAXII and Vinculin in MCF7 cells incubated with increasing concentrations of HyT LD-TMAc for c) 3 hours and d) 6 hours; e) Bar graph of relative CAXII abundance from figures c and d. * Represents the number of hydrophobic adamantane tags by HyT LD-TMAc.
adduct appear on the labeled bCAII spectrum, demonstrating up to three tags per protein.

To evaluate HyT-induced degradation by LD-TMAcs, MCF7 cells expressing CAXII on their cell membrane were incubated with increasing concentrations of HyT LD-TMAcs from 0.5 µM to 10 µM for 3 and 6 hours. Then, cell lysates were extracted, and the relative abundance of CAXII was evaluated by Western blotting (Figure 3.3 c-e). When the relative abundance of CAXII from Figure 3.3 c and d was plotted, a trend of decreasing CAXII abundance is observed as the concentration of the HyT probe is increased (Figure 3.3 e). On the other hand, the abundance of the control protein Vinculin is unaffected by increasing concentrations of the HyT probe. CAXII degradation of 76% and 65% is observed for 3 and 6 hours, respectively, with DC50 of 1 µM HyT
The % decrease of degradation at 6 hours relative to 3 hours can be explained by fast reaction and consumption of HyT probes. Although HyT LD-TMAcs have fast reaction kinetics, they are not catalytic as some other targeted protein degraders.

Targeted degradation was further evaluated at incubation of shorter 0.5 and 1 hours and longer time point of 12 hours (Figure 3.4). No trend is observed for CAXII abundance as HyT probe incubation time is increased. Still, the relative abundance of CAXII remained as low as ~ 44 % at 12 hours.
the 10 µM HyT probe, demonstrating sustained degradation. Additionally, our probes show rapid
degradation of ~ 75 % CAXII within 30 minutes at 10 µM HyT probe. To see the time-dependent
change in CAXII abundance relative to HyT probe incubation time, more western blotting
experiments with short increments in probe incubation time will need to be done.

3.2.2 Influence of LD-TMAc labeling on CAXII degradation and IgG recognition

While hydrophobic tagging mediated degradation of proteins has been well established, we
questioned if simple LD-TMAc labeling, regardless of the payload hydrophobicity, would be
enough for protein degradation. It has been demonstrated that nanoparticle-conjugated antibodies
or simply antibodies themselves can cause protein degradation to a certain extent.\textsuperscript{23} To see the
effect of LD-TMAc labeling on CAXII degradation, MCF7 cells were incubated with methyl and
propargyl LD-TMAcs for three hours, followed by Western blotting analysis of relative CAXII
abundance (Figure 3.5). Compared to the 76% CAXII degradation of the 10 µM HyT probe, the
methyl probe causes just about 30% CAXII degradation with the same concentration (Figure 3.5
b, c). Even lesser CAXII degradation of 15% is observed with µM propargyl probe (Figure 3.5 d,
e). One of the explanations for HyT-induced degradation is direct destabilization in the targeted
protein structure. Therefore, it is easy to explain degradation caused by bulky and hydrophobic
adamantane. Although methyl LD-TMAc carries a simple methyl group on its ester functionality,
it can label a single carbonic anhydrase protein up to six times compared to the maximum
modification of 3 adamantane units of HyT LD-TMAc (Figure 2.2 b, Figure 3.3 b). Therefore,
six units of methyl LD-TMAc modifications concentrated around the ligand binding site might be
enough to cause CAXII degradation of 30%. Similarly, although propargyl LD-TMAc carries a
bulkier propargyl unit on its ester functionality, it can label only up to three modifications per CA
(Figure 2.5 b). Since propargyl LD-TMAc is neither hydrophobic nor can label the target protein
up to six times, it manages only half the degradation extent of 15% compared to 30% degradation by methyl LD-TMAc.

Then, we questioned if the decrease in CAXII abundance indicated the Western blotting experiments were indeed due to protein degradation or the inability of IgG to recognize the CAXII post-LD-TMAc labeling. To confirm the antibody recognition post-LD-TMAc labeling, cell lysates from MCF7 cells were extracted, and labeled with HyT LD-TMAc for an hour, followed by SDS-PAGE and western blotting analysis (Figure 3.6 a). CAXII labeling in cell lysates cannot cause any reduction in any protein abundance, as there is no active proteolytic machinery in cell lysates. Therefore, the only plausible explanation for the decrease in protein abundance indicated by Western blotting would be the loss of IgG’s ability to recognize the target protein. Compared
to a 76% reduction of CAXII band intensity following live cell hydrophobic tagging with 10 µM HyT LD-TMAc, cell lysate labeling with the same probe concentration does not significantly reduce band intensity (Figure 3.6 b). Similarly, methyl and propargyl LD-TMAc labeling in cell lysates show no decrease in CAXII abundance with all the probe concentrations tested (Figure 3.6 c, d). Altogether, these results demonstrate that LD-TMAc labeling, whether it is methyl, propargyl, or adamantane, does not cause any loss in antibodies protein recognition capability; and the reduction of CAXII abundance following live cell labeling with HyT LD-TMAc is indeed due to protein degradation.

3.2.3 Investigation of HyT CAXII degradation pathway

Previous studies with A431 and MCF7 cell lines have shown that CAXII gets naturally internalized and digested in lysosomes with a half-life of ~9 hours. Another natural degradation pathway for membrane CA is partial proteolysis by ectodomain shedding caused by cleavage of CAXII or CAIX ectodomain by TAM/ADAM17 metalloprotease. It has been shown that apoptotic conditions upregulate CAIX ectodomain shedding, suggesting that an increased level of circulating CAIX ectodomain can indicate an effective cancer treatment. With the combined knowledge of hydrophobically tagged proteins getting degraded by the proteasome, we hypothesized that hydrophobically tagged CAXII is degraded in one of three ways: ectodomain shedding, proteasomal degradation, or lysosomal digestion (Figure 3.2).

To investigate the possibility of the lysosomal digestion of hydrophobically tagged CAXII, MCF7 cells were incubated with 0.1 µM Bafilomycin for an hour. Bafilomycin prevents acidification by inhibiting the lysosomal proton pump V-ATPase and, therefore, degradation of internalized proteins. After one hour of Bafilomycin incubation, MCF7 cells were treated with increasing concentrations of HyT LD-TMAc probe for three hours, followed by Western blotting.
When the protein degradation extent is compared to hydrophobic tagging with no Bafilomycin preincubation, it is clear that lysosomal inhibition does not significantly change CAXII degradation extent. Having ruled out the lysosomal digestion of HyT CAXII, we investigated the proteasomal digestion pathway using covalent proteasome inhibitor MG132. In the same manner, MCF7 cells were preincubated with 10 µM MG132 for an hour, followed by treatment of increasing concentrations of HyT LD-TMAc for three hours (Figure 3.7 c, d). Gratifyingly, ~30% of CAXII degradation is inhibited by MG132. Compared to 76% CAXII degradation at 10 µM LD-TMAc, only 44% degradation is observed when proteasomal
degradation is inhibited. To test if this 30% decrease in degradation can be pushed to total inhibition, western blotting experiments were repeated with increasing concentrations of MG132 while keeping 5 µM HyT LD-TMAc concentration constant (Figure 3.8). However, although CAXII degradation decreased with increasing concentrations of MG132 inhibitor, more degradation was observed compared to 5 µM LD-TMAc alone. These results suggest that other proteolytic mechanisms might be upregulated when the proteasomal pathway is fully inhibited.

To investigate the possibility of ectodomain shedding and exosomal secretion of HyT CAXII, we synthesized coumarin-conjugated LD-TMAc (Figure 3.9 a). To demonstrate the coumarin labeling, 5 µM bCAII was incubated with 10 µM coumarin LD-TMAc for three hours and analyzed with MALDI-MS (Figure 3.9 b). Although 30% of bCAII remained unlabeled, new bCAII peaks with up to three coumarin labels per protein are observed. Since the coumarin moiety
itself is hydrophobic, we hypothesized that fluorescently labeled CAXII ectodomains would be found in the cell culture media after either coumarin LD-TMAc or a combination of coumarin and HyT LD-TMAc labeling of cells. To test this theory, MCF cells were incubated with coumarin and HyT LD-TMAcs. Then cell lysates and media were collected, subjected to SDS-PAGE, and analyzed for in-gel fluorescence (Figure 3.10 b, c). Cell media was collected directly (method 1) or subjected to lysis buffer before loading to the gel to collect the HyT CAXII possibly secreted in exosomes (method 2). Details of the media collection protocol are elaborated in the methods section. When 5 µM bCAII was labeled with 10 µM coumarin probe for an hour and loaded to the first lane of the gels as a control, a bright fluorescent band corresponding to the bCAII and traces of CA with higher molecular weight was observed. When the cells were incubated with coumarin for one or four hours, two highly fluorescent bands were observed in the media (Figure 3.10 b, c; lane 1, 5); and no significant fluorescence was observed in the cell lysate (Figure 3.10 b, c). The presence of fluorescent bands in the media and the absence of fluorescence in the cell lysates indicates that CAXII ectodomains get shed as soon as they get labeled by the coumarin probe.

**Figure 3.9:** Structure of coumarin LD-TMAc b) MALDI-MS spectrum of control bCAII and coumarin LD-TMAc labeled bCAII.
Figure 3.10: In-gel analysis of MCF7 cell media and lysate post coumarin LD-TMAc and HyT LD-TMAc labeling. a) Gel contents of figures b-e and LD-TMAc labeling reaction conditions of cells which the samples were collected from. Cell media of figure b and c were collected using method 1 and method 2, respectively. Gels b and c are otherwise identical. b, c) In-gel fluorescence; d, e) SYPRO Ruby total protein staining of gels in figure b and c respectively; f, g) Western blotting of CAII and CAXII respectively. Cell media samples for figures f and g were collected using method 1. Arrow indicates expected position of CAXII based on its molecular weight. Arrow indicates CAII. Arrow indicates new protein band that does not appear in control media loaded into lane 3.
When MCF7 was sequentially incubated with the coumarin probe for an hour and the HyT probe for three hours, no fluorescence in the media or the cell lysate was observed due to the media change after coumarin incubation (Figure 3.10 b, c; lane 4, 5). No fluorescence from the cell lysate and media was observed from control cells incubated without any probe (Figure 3.10 b, c). No difference in cell media fluorescence of samples collected by methods 1 and 2 is observed, indicating HyT CAXII is not secreted via exosomes.

When the same gels were stained with SYPRO Ruby for total protein detection, numerous bands were observed from all cell lysate samples. Although the cell labeling reactions were carried out in serum-free media, cells were seeded and incubated in 1% FBS media before the experiments. A single protein band possibly corresponding to traces of BSA was observed in the control cell media. The media from cells incubated with one and three hours of coumarin contained an additional protein band possibly corresponding to the CAXII ectodomain (Figure 3.10 d, e; arrows). No protein was observed in the media cells sequentially incubated with coumarin and HyT probes due to the media exchange following coumarin probe incubation.

To detect and immunologically identify the hCAII and hCAXII in cell lysates and media, gels identical to those in Figure 3.10 c were subjected to Western blotting (Figure 3.10 f, g). Unfortunately, dry transfer times were not optimal for CAII, and only thin protein bands were observed in cell media. Another critical factor that might have contributed to the poor detection of CAII is its inherent low-level expression in MCF7 cells. This possibility is supported by good detection of abundant bCAII in the first lane by the hCAII-targeted primary antibodies. No CAXII was observed in the cell media (Figure 3.10 g). It is likely that our primary antibodies did not recognize the CAXII ectodomain after it was shed. As expected, increasing levels of CAXII degradation was observed in cells incubated with the coumarin probe for one hour, four hours, and
sequential incubation with coumarin and HyT probes (Figure 3.10 g; lanes 8, 4, and 6, respectively).

After in-gel fluorescence analysis of cell lysates and media suggested ectodomain shedding, we investigated the effects of ectodomain shedding inhibitors and activators on CAXII degradation. A Disintegrin and Metalloprotease (ADAM) gene family is known for their role in ectodomain shedding.\(^{21}\) While Batimastat (Bat) and Retinoic acid (RA) are inhibitors, phorbol-12-myristate-13-acetate (PMA) is an activator of ADAM17.\(^{21}\) To evaluate their effect on CAXII degradation rate, Bat and PMA were incubated with or without HyT LD-TMAcs for three hours and Western blotting experiments were carried out to quantify CAXII (Figure 3.11 a, b). As
expected, PMA incubation alone showed 19% degradation, and Bat incubation alone did not significantly reduce CAXII abundance. When 1 µM and 5 µM of HyT probes were incubated with 10 µM of PMA, the relative abundance of CAXII decreased from 50% and 45% to 44% and 35%, respectively. While the PMA activator showed some increase in CAXII degradation, the Bat inhibitor did not cause a significant change in CAXII degradation.

Then, to evaluate the inhibition of ectodomain shedding, lysosomal and proteasomal degradations side by side, 1 µM and 5 µM of HyT probes were incubated with 20 µM of RA, Baf, or MG132 (Figure 3.11 c, d). While incubation of all three inhibitors alone did not cause any changes, CAXII abundance slightly increased when Baf was co-incubated with 1 µM and 5 µM HyT probes from 51% and 38% to 65% and 51%, respectively, in accordance with our previous results with Baf. Unlike Bat inhibitor, co-incubation with RA effectively increased CAXII abundance to 72% and 63% with 1 µM and 5 µM HyT probes, respectively. Finally, co-incubation with MG132 resulted in the most effective inhibition of degradation resulting in 74% and 69%
CAXII abundance with 1 µM and 5 µM HyT, respectively.

3.2.4 Confocal microscopy tracking of HyT CAXII

After western blotting experiments suggested HyT CAXII is degraded either via ectodomain shedding or proteasomal pathway, we wanted to track HyT CAXII by confocal microscopy experiments (Figure 3.12 a). First, confluent MCF7 cells were incubated with 10 µM HyT LD-TMAc for 30 minutes (Figure 3.13). Then the cells were added anti-CAXII FITC IgG, stained with Hoechst, and imaged after 10, 90, and 120 minutes. After 10 minutes of anti-CAXII FITC IgG incubation, CAXII was clearly visible on the cell membrane with green FITC fluorescence. After 90 minutes of FITC IgG incubation, spots of FITC fluorescence are observed outside the cells (Figure 3.13, arrows). After 120 minutes, FITC fluorescence on the cell membrane visibly decreased, and we continued to observe fluorescent spots outside the cells. FITC fluorescent spots were also observed on the cell membrane (Figure 3.13, arrows). After 120 minutes of imaging, cells were washed and imaged to see all the fluorescent spots outside the cell membrane disappear. Some fluorescent spots remained inside the cells (Figure 3.13, arrow). It is possible that the HyT CAXII was trafficked inside the cells to be degraded by the endoplasmic reticulum-associated degradation (ERAD) pathway.

As a control experiment, confluent MCF cells were directly incubated with anti-CAXII FITC IgG (Figure 3.12 b). After 10 minutes of incubation, CAXII on the cell membrane is clearly visible by the green fluorescence of the FITC IgG (Figure 3.14). Over time, FITC fluorescence on the cell membrane got clarified, but there were no aggregations or spots of fluorescence inside or outside the cells even after 120 minutes. Together, these results show the clear and visible effects of LD-TMAc HyT on CAXII.

Since FITC IgG labeling of CAXII is a noncovalent interaction, we questioned if similar
106

results can be obtained by tracking the HyT CAXII after covalent labeling with coumarin LD-TMACs. Confluent MCF7 cells were incubated with coumarin LD-TMAc for 30 minutes, and

Figure 3.13: Confocal microscopy tracking of HyT CAXII. Confluent MCF7 cells were incubated with 10 µM HyT LD-TMAc for 30 minutes. Then anti-CAXII FITC IgG was added to cell media to track the HyT CAXII. Hoechst staining, FITC IgG fluorescence, brightfield image and their merged images were obtained following 10, 90 and 120 minutes of FITC IgG incubation. Cells were imaged one more time after 120 minutes’ imaging and washing. Red, pink, and yellow arrows show aggregations of FITC IgG fluorescence outside cells, on cell membrane and inside cells respectively.
coumarin fluorescence, Hoechst staining, brightfield, and merged images were obtained (Figure 3.15 a). After that, cells were added 10 µM HyT LD-TMAc and images were obtained similarly after 90 and 120 minutes. Coumarin and HyT LD-TMAcs bind to the same ligand binding site, label the same residues, and HyT probes are much faster than the coumarin probes. Therefore, this sequential labeling results in partial labeling of CAXII by coumarin, then more labeling with hydrophobic adamantane.

After 30-minute incubation of MCF7 cells with coumarin LD-TMAc, we saw coumarin fluorescence not only from the membrane but also from the nuclear membrane and arrays of perinuclear vesicles (Figure 3.15 a, arrowheads, and arrows). We hypothesize the cell membrane permeability of the coumarin LD-TMAc allows us to visualize not only CAXII but also cytoplasmic CAII. While the cell membrane fluorescence is due to the CAXII labeling, it has been

**Figure 3.14:** Confocal microscopy imaging of CAXII. Anti-CAXII FITC IgG was added to a confluent MCF7 media and cells were stained with Hoechst. Then FITC fluorescence, Hoechst staining and their merged images were obtained after 10 and 120 minutes.
shown that cytoplasmic CAII can be loosely associated with the nuclear membrane and secretory

Figure 3.15: Confocal microscopy tracking of coumarin LD-TMAc and HyT LD-TMAc labeled CAXII. Coumarin fluorescence, Hoesch staining, brightfield and merged images of MCF7 a) post 30 minutes incubation of 1 µM coumarin LD-TMAc; and b) post 90 and 120 min incubation with 10 µM HyT LD-TMAc following 30 min coumarin labeling. Arrowheads show nuclear membrane staining. Arrows show perinuclear vesicular staining. Arrows show aggregation of coumarin fluorescence near cell membrane. Arrows show vanished coumarin fluorescence on cell membrane.

shown that cytoplasmic CAII can be loosely associated with the nuclear membrane and secretary
granules in the perinuclear region.\textsuperscript{29}

After 90-minute incubation of MCF7 cells with HyT LD-TMAc post coumarin labeling, coumarin fluorescence from both cell membrane and cytoplasmic membranes started to dissipate and spread all over the cytoplasm. Like FITC IgG aggregation, we also saw some aggregations of coumarin fluorescence close to the cell membrane (Figure 3.15 b, arrows). After 120 minutes, fluorescence on the cell membrane completely disappears (Figure 3.15 b, arrows).

3.2.5 PDL-1 targeted HyT LD-TMAc

Having established the degradation of CAXII by HyT LD-TMAcs, we sought more membrane protein targets. PDL-1 is a checkpoint membrane protein that binds to the PD-1 on killer T cells to avoid T cell-induced apoptosis.\textsuperscript{30} While this essential checkpoint keeps unhealthy cells going rogue most of the time, cancer cells slip away from this apoptotic checkpoint by overexpression of PDL-1.\textsuperscript{31} To enhance immunotherapy by targeted degradation of PDL-1, we synthesized LD-TMAc guided by PDL-1 ligand BMS-1166 (Figure 3.16 a).\textsuperscript{32} To evaluate the degradation by PDL-1 HyT probes, MDA-MB-231 cells expressing PDL-1 on their cell membrane were incubated with increasing concentrations of HyT BMS-1166 probes from 5 nM to 5 µM for 6 hours, and PDL-1
abundance was analyzed by Western blotting (Figure 3.16 b). However, the presence of HyT probes did not cause any change in the PDL-1 band intensity. Since BMS-1166 acts by dimerizing two PDL-1 proteins to form a ternary complex, conjugation of bulky HyT LD-TMAc unit might have reduced its efficiency in forming the ternary complex.

3.3 Conclusion

In conclusion, using our LD-TMAc target selective protein labeling platform, we have developed new Hydrophobic Tagging (HyT) membrane protein degraders. By functionalizing the LD-TMAcs with hydrophobic adamantane and benzenesulfonamide ligand on their ester and leaving groups, we have synthesized CAXII-targeted HyT probes. MALDI-MS experiments show that our HyT probes can label bCAII to almost completion in three hours with up to 3 adamantane units per bCAII protein molecule. HyT LD-TMAcs have DC_{50} of 1 µM, and they can cause up to 75% CAXII degradation in 30 minutes at 10 µM. While inhibition of lysosomal acidification does not cause a significant change in HyT LD-TMAc-induced CAXII degradation, proteasomal inhibition results in 30% less CAXII degradation. Inhibition of ectodomain shedding metalloproteases by RA also resulted in 21-25% less CAXII degradation. In-gel fluorescence and total protein staining analysis of cell lysates and media after hydrophobic coumarin labeling of CAXII revealed fluorescent protein bands in the cell media. Confocal microscopy tracking of HyT CAXII by FITC IgG and coumarin labeling suggests that the HyT CAXII proteins end up outside or inside the cells. These results suggest that multiple pathways, including ectodomain shedding and proteasomal pathways, are responsible for the degradation of HyT CAXII. While our PDL-1 targeted HyT LD-TMAc did not cause any protein degradation, we are yet to explore the degradation of more membrane protein targets.

3.4 Acknowledgements
Western blotting experiments were conducted by Jithu Krishna. Confocal microscopy experiments were conducted by Jithu Krishna and Myrat Kurbanov.

3.5 Experimental details

3.5.1 Materials and instrumentation

All chemicals and materials were purchased from commercial sources unless stated otherwise. CAXII primary & secondary antibodies and PDL-1 primary & secondary antibodies were purchased from Abcam. Vinculin primary antibody, Clarity Western ECL Substrate, and Precision Plus Protein Dual Color Standards (protein ladder) were purchased from Bio-Rad Laboratories. Vinculin secondary antibody and Hoechst were purchased from Invitrogen. RIPA buffer, Halt™ Protease Inhibitor Cocktail (100X), BCA kit, and SYPRO™ Ruby Protein Gel Stain were purchased from ThermoFisher Scientific. WGA was purchased from Biotium. 1H NMR and 13C NMR spectra were recorded on either Bruker 400 or 500 MHz NMR spectrometer. MALDI-MS studies were carried out using Bruker UltrafleXtreme MALDI-TOF/TOF. All optical images were captured using a CrestV2-2xTIRF confocal microscope.

3.5.2 MALDI-MS experiments

bCAII (5 μM) was incubated with LD-TMAc probes (10 μM) in PBS buffer, pH 7.4, at 37 °C while shaking for the stated duration. 25 mg/mL Sinapinic acid in 0.2 % TFA, 1:1 Acetonitrile/Water solution was prepared as MALDI matrix solution. After the protein labeling reaction, 5 uL of the reaction mixture and 5 uL of matrix solution were mixed in a new Eppendorf tube by repeated dispensing and aspirations of the pipette used. Then, 1-1.5 µL of the sample mixed matrix solution was loaded onto a ground steel MALDI-MS target plate and left for drying in a fume hood. Finally, the samples loaded on the MALDI target plate were analyzed using Bruker
3.5.3 Western blotting experiments.

Cells were plated in a 24-well plate (0.5 × 10^6 cell/well) in high glucose media. After overnight incubation, cells were washed with PBS and serum free media was added before the treatment. Then, each well was treated with different probe/inhibitor/activator concentrations and incubated for corresponding time points. After the treatment, cells were washed with ice-cold PBS and 150 μL of lysis buffer (RIPA buffer + protease-phosphatase inhibitor cocktail) was added in each well and kept for shaking in ice for 30 minutes. Then, the plates were sonicated for 2 minutes, and each well was collected into a separate Eppendorf tube. The tubes were centrifuged at 14000 rpm for 15 minutes and the supernatant was collected into a new Eppendorf tube. A BCA assay was carried out to find the protein concentrations in each collected tube. From the determined protein concentration, calculated volume of protein was diluted with PBS and 10 μL of 3X loading dye was added to make final volume and concentrations of 40 μL and 10 μg/40 μL respectively. Samples were loaded into 4-20% precast polyacrylamide gel, 8.6 × 6.7 cm (W × L), and the gel was run at 130 V for 1 h. After that, the gel was separated from the gasket and put in a tray with distilled water. The gel was blotted onto PVDF membrane by conducting dry transfer using P0 method (20 V for 1 minute, 23 V for 4 minutes, 25 V for 3 minutes) of iBlot™ 2 Gel Transfer Device for a total duration of 8 minutes. The PVDF membrane was washed with distilled water and blocked by incubation in 5% nonfat milk solution for one hour. The membrane was cut into two pieces corresponding to CAXII and Vinculin. The membrane pieces were separately probed with their respective 1/2000 diluted primary antibody in 5% nonfat milk in 1X Tris-Buffered Saline, 0.1% Tween® 20 Detergent (TBST) for 12 hours at 4 degrees while shaking. Then, they were washed twice with 1X TBST for 10 minutes and probed with 1/5000 diluted secondary
antibody for one hour. After that, membranes were washed twice with 1X TBST for 10 minutes and imaged by incubating in Clarity Western ECL Substrate for 8 minutes in Bio-Rad chemi doc imaging system. Finally, the blots were analyzed and processed using Image Lab software.

3.5.4 Confocal microscopy experiments.

Cells were plated cells in a 35 mm$^2$ glass bottom plate (0.5 × 10$^6$ cells/well) and incubated overnight. After washing the cells with PBS, the media was changed to serum-free media before the treatment with different probe/inhibitor/activator concentrations and incubated for corresponding time points at 37 °C. Then, the cells were incubated with Hoechst and/or WGA for 5 and 15 minutes respectively before imaging. Finally, the cells were imaged under CrestV2 (CrestV2-2xTIRF instrument) for specified duration of time.

3.5.5 Comparative analysis of cell lysate and media proteins for ectodomain shedding.

Cells were plated in a 24-well plate (0.5 × 10$^6$ cell/well) in high glucose media. After overnight incubation, cells were washed with PBS and serum free media was added before treatment. Then, each well was treated with different probe/inhibitor/activator concentrations and incubated for corresponding time points. After the treatment, media from each well was collected and proteins were concentrated in a 3 kDa cutoff filtration tube by centrifugation. Right after the removal of media, the cells were washed with ice-cold PBS and cell lysate was collected. The rest of the media collection procedure was carried out using two different methods. Method 1: After the media collection, a BCA assay was carried out to get the final concentration of the proteins in the media. Then, a calculated amount of protein was diluted with PBS, added 10 μL of 3X loading dye to make 40 μL of 10 μg/40 μL sample and directly loaded onto SDS-PAGE well. Method 2: The rest of the media from method 1 was diluted half with RIPA buffer + protease cocktail and incubated
for 30 minutes in ice while shaking. Then, keeping the half dilution in mind, a calculated amount of protein was further diluted with PBS, added 10 μL 3X loading dye to make 40 μL of 10 μg/40 μL sample, and directly loaded onto SDS-PAGE well.

3.6 Synthetic procedures

3.6.1 Synthesis of CA targeted HyT LD-TMAc, compound 7

Synthetic procedures and characterizations for compounds 1d and 3b are provided in chapter 2. Compounds 7a, 7b, 7c and 7 were synthesized following the synthetic procedures for compounds 4b, 4c, 4d and 4 respectively from Chapter 2.

Characterizations of compound 7a: 1H NMR (400 MHz, CDCl3) δ 7.68-7.65 (m, 4H), 7.45-7.36 (m, 6H), 6.23-6.21 (q, 1H), 6.02-6.00 (q, 1H), 4.39-4.38 (t, 2H), 2.17-2.09 (m, 9H), 1.66-1.64 (t, 6H), 1.08-1.07 (t, 9H). 13C NMR (101 MHz, CDCl3) δ 165.01, 141.20, 135.64, 133.57, 129.88, 127.92, 127.90, 122.98, 81.03, 62.52, 41.45, 36.36, 30.99, 27.00, 19.49. Yield: 60%.
Characterizations of compound 7b: $^1$H NMR (400 MHz, CDCl$_3$) δ 6.16-6.15 (p, 1H), 5.73-5.72 (q 1H), 4.28 (s, 2H), 2.20-2.15 (m, 9H), 1.71-1.64 (m, 6H). $^{13}$C NMR (101 MHz, CDCl$_3$) δ 165.56, 141.02, 125.04, 81.68, 63.03, 41.51, 36.31, 31.01. Yield: 60%.

Characterizations of compound 7c: $^1$H NMR (400 MHz, CDCl$_3$) δ 6.23-6.23 (d, 1H), 5.85-5.84 (q, 1H), 4.14-4.14 (d, 2H), 2.20-2.17 (m, 9H), 1.72-1.64 (m, 6H). $^{13}$C NMR (101 MHz, CDCl$_3$) δ 163.81, 139.11, 128.16, 81.94, 41.41, 36.33, 31.03, 30.06. Yield: 43%.

Characterizations of compound 7: $^1$H NMR (400 MHz, MeOD) δ 8.92-8.90 (d, 2H), 8.04-8.02 (d, 2H), 7.99-7.93 (m, 4H), 6.56 (s, 1H), 6.21 (s, 1H), 5.41-5.41 (d, 2H), 3.91 (s, 2H), 3.41-3.37 (t, 2H), 3.24-3.21 (t, 2H), 2.16-2.05 (m, 9H), 1.71-1.39 (m, 14H). $^{13}$C NMR (101 MHz, MeOD) δ 169.60, 168.82, 164.56, 158.58, 147.75, 145.75, 139.29, 136.48, 134.16, 130.00, 129.06, 127.42, 83.90, 67.01, 62.56, 42.31, 37.14, 32.42, 30.36, 30.26, 27.63, 27.59, 15.54. Yield: 37%.

3.6.2 Synthesis of CA targeted coumarin LD-TMAc, compound 8.

Compounds 8a, 8b, 8c, 8d and 8 were synthesized following the synthetic procedures for compounds 3a, 4b, 4c, 4d, 4 respectively from Chapter 2.

Characterizations of compound 8a: $^1$H NMR (500 MHz, CDCl$_3$) δ 9.22, 9.21, 9.20, 8.59, 7.00, 3.46, 3.45, 3.35, 3.34, 3.33, 3.32, 2.90, 2.88, 2.87, 2.78, 2.77, 2.76, 2.00, 1.99, 1.98, 1.97, 1.96, 1.95, 1.26, 0.07. $^{13}$C NMR (126 MHz, CDCl$_3$) δ 165.67, 163.24, 152.94, 148.55, 148.53, 127.29, 119.91, 108.50, 108.40, 105.84, 71.55, 51.48, 50.44, 50.02, 27.64, 27.54, 21.29, 20.35, 20.26. Yield: 66%.

Characterizations of compound 8b: $^1$H NMR (400 MHz, CDCl$_3$) δ 9.22-9.19 (t, 1H), 8.59 (s, 1H), 7.68-7.65 (m, 4H), 7.42-7.34 (m, 6H), 7.00 (s, 1H), 6.37-6.36 (q, 1H), 6.12-6.11 (q, 1H), 4.44-4.43 (q, 2H), 3.68-3.67 (d, 2H), 3.35-3.31 (m, 4H), 2.91-2.88 (t, 2H), 2.79-2.76 (t, 2H), 2.02-1.95 (m, 4H), 1.53 (s, 6H), 1.07 (s, 9H). $^{13}$C NMR (126 MHz, CDCl$_3$) δ 165.06, 163.82, 163.03, 152.70,
Scheme 3.2: Synthesis of CA targeted coumarin LD-TMAc, compound 8.


Characterizations of compound 8c: $^1$H NMR (400 MHz, CDCl$_3$) δ 9.26-9.23 (t, 1H), 8.62 (s, 1H), 7.01 (s, 1H), 6.27-6.27 (d, 1H), 5.76-5.76 (d, 1H), 4.29-4.28 (d, 2H), 3.71-3.69 (d, 2H), 3.36-3.31 (q, 4H), 2.95-2.92 (t, 1H), 2.90-2.86 (t, 2H), 2.79-2.76 (t, 2H), 2.01-1.94 (p, 4H), 1.59 (s, 6H). $^{13}$C NMR (126 MHz, CDCl$_3$) δ 165.87, 164.33, 163.45, 152.91, 148.63, 148.51, 140.93, 127.31, 126.57, 119.95, 108.69, 108.48, 105.84, 82.68, 62.81, 50.44, 50.02, 48.90, 27.65, 23.68, 21.30, 20.36, 20.27. Yield: 60%.

Characterizations of compound 8d: $^1$H NMR (400 MHz, CDCl$_3$) δ 9.23-9.20 (t, 1H), 8.61 (s, 1H), 7.00 (s, 1H), 6.36-6.36 (d, 1H), 5.92-5.92 (s, 1H), 4.19-4.19 (d, 2H), 3.74-3.72 (d, 2H), 3.35-3.31
Scheme 3.3: Synthesis of PDL-1 targeted BMS-1166 guided HyT LD-TMAc, compound 9.

(q, 4H), 2.89-2.86 (t, 2H), 2.78-2.75 (t, 2H), 2.01-1.94 (m, 4H), 1.60 (s, 6H). $^{13}$C NMR (126 MHz, CDCl$_3$) δ 164.13, 164.03, 163.35, 152.88, 148.51, 148.40, 138.54, 129.07, 127.24, 119.86, 108.89, 108.42, 105.83, 83.29, 50.42, 49.99, 48.66, 30.06, 27.64, 23.63, 21.30, 20.36, 20.28. Yield: 54%.

Characterizations of compound 8: $^1$H NMR (500 MHz, MeOD) δ 9.22-9.19 (t, 1H), 8.95-8.94 (d, 2H), 8.46 (s, 1H), 8.03-8.02 (d, 2H), 7.97-7.92 (q, H), 7.14 (s, 1H), 6.66 (s, 1H), 6.26 (s, 1H), 5.45 (s, 2H), 3.87 (s, 2H), 3.69-3.68 (d, 2H), 3.40-3.35 (m, 6H), 3.19-3.17 (t, 2H), 2.85-2.82 (t, 2H), 2.79-2.77 (t, 2H), 1.99-1.94 (m, 4H), 1.61-1.36 (m, 14H). $^{13}$C NMR (126 MHz, MeOD) δ 168.04, 167.26, 164.53, 163.63, 163.26, 152.67, 148.96, 148.05, 146.25, 145.03, 144.32, 137.75, 134.55, 133.24, 128.49, 127.55, 127.19, 125.92, 125.48, 120.52, 108.02, 106.55, 105.12, 83.64, 61.28, 60.93, 49.90, 49.39, 39.50, 39.16, 28.85, 28.75, 27.04, 26.10, 26.05, 25.99, 22.78, 20.74, 19.78, 19.62. Yield: 55%.

3.6.3 Synthesis of PDL-1 targeted BMS-1166 guided HyT LD-TMAc, compound 9.

BMS-1166 was purchased from Arctom. Compound 9a, and 9 were synthesized following the synthetic procedures for compound 3a, and 4 respectively from Chapter 2.
Characterizations of compound 9a: $^1$H NMR (400 MHz, CDCl$_3$) $\delta$ 8.50-8.47 (d, 2H), 8.17-8.15 (d, 1H), 7.74 (s, 1H), 7.66-7.58 (m, 2H), 7.49-7.45 (t, 1H), 7.37-7.33 (m, 3H), 7.22-7.19 (m, 3H), 6.90-6.88 (d, 1H), 6.79-6.78 (d, 1H), 6.57 (s, 1H), 5.11-5.07 (d, 4H), 4.40 (s, 1H), 4.29 (s, 4H), 4.02-3.99 (d, 1H), 3.85-3.82 (d, 1H), 3.73-3.66 (m, 2H), 3.56 (s, 2H), 3.46 (s, 2H), 3.27-3.25 (d, 1H), 2.92-2.80 (m, 2H), 2.24 (s, 3H), 1.27-1.14 (m, 8H). $^{13}$C NMR (101 MHz, CDCl$_3$) $\delta$ 169.22, 155.96, 155.05, 149.40, 148.23, 143.23, 142.85, 142.57, 137.93, 135.16, 134.51, 134.27, 132.50, 132.07, 131.85, 130.94, 130.46, 129.87, 128.21, 127.47, 125.74, 125.38, 122.67, 120.03, 118.69, 118.35, 117.08, 115.49, 112.99, 100.09, 70.63, 70.09, 69.79, 66.15, 64.60, 64.57, 62.00, 53.69, 53.00, 50.87, 42.98, 42.07, 40.19, 39.09, 38.66, 29.85, 29.19, 28.84, 25.57, 25.53, 18.78, 17.64, 16.40, 12.04. Yield: 44%.

Characterizations of compound 9: $^1$H NMR (400 MHz, MeOD) $\delta$ 8.91-8.89 (d, 2Hz), 8.03-8.02 (d, 3H), 7.89-7.87 (d, 1H), 7.75-7.73 (d, 1H), 7.62-7.58 (t, 1H), 7.48 (s, 1H), 7.41-7.39 (d, 1H), 7.22-7.15 (m, 2H), 6.98 (s, 1H), 6.88-6.86 (d, 1H), 6.74-6.71 (m, 2H), 6.56 (s, 1H), 6.21 (s, 1H), 5.41-5.34 (m, 4H), 5.25 (s, 2H), 4.28 (s, 4H), 3.74-3.71 (m, 2H), 3.15-2.99 (m, 4H), 2.26 (s, 3H), 2.15-2.02 (m, 12 H), 1.89-1.86 (m, 2H), 1.70-1.66 (m, 8H), 1.51-1.22 (m, 8H). $^{13}$C NMR (101 MHz, MeOD) $\delta$ 169.51, 164.58, 158.48, 146.14, 145.74, 144.76, 144.35, 143.98, 136.48, 136.34, 136.11, 135.62, 134.15, 133.85, 133.30, 132.90, 131.34, 131.09, 130.00, 128.73, 128.57, 126.67, 123.52, 119.68, 119.25, 118.02, 115.97, 114.04, 101.12, 83.90, 82.50, 71.51, 71.32, 68.96, 67.01, 66.77, 65.83, 65.80, 63.25, 62.67, 62.56, 54.49, 42.46, 42.34, 42.32, 40.75, 37.29, 37.14, 32.42, 30.26, 27.58, 27.50, 27.45, 26.60, 16.62, 16.60, 15.54. Yield: 55%.
Figure 3.17: $^1$H NMR spectrum of compound 7.

Figure 3.18: $^{13}$C NMR spectrum of compound 7.
Figure 3.19: $^1$H NMR spectrum of compound 8.

Figure 3.20: $^{13}$C NMR spectrum of compound 8.
3.6 References


(8) Banik, S. M.; Pedram, K.; Wisnovsky, S.; Ahn, G.; Riley, N. M.; Bertozzi, C. R.


CHAPTER 4

PHOTO-RESPONSIVE HYDROGEL DRUG DEPOTS FOR IMPLANTABLE MACROSCALE DRUG DELIVERY (MDD) DEVICE

Figure 4.1: Schematic representation and comparison of MDD with parental drug administration. (Reproduced with permission. Copyright 2013, Springer Nature).

4.1 Introduction

Macroscale drug delivery (MDD) devices are medical devices carrying depots of drugs that can be implanted to the targeted disease site (Figure 4.1). They are designed to address the limitations of the traditional systemic administration of drugs, including large, repeated dosing, toxic side effects, and exposure of the labile therapeutics to harsh environments. On the other hand, MDD devices protect the labile therapeutics in their safe depot medium and localize the drug's effect to the target site in which the device is implanted. Most importantly, MDD devices
release the drug in a controlled and sustained manner keeping the dosage within the therapeutic window but not exceeding the dosage that will result in detrimental outcomes.

Among others, hydrogels are one of the most widely used drug depot materials in MDD devices. Hydrogels are gel-like materials composed chiefly of water supported by polymer networks that confer physical, chemical, and biological properties to the hydrogel. Various polymers, from natural, synthetic, to modified natural polymers, have been employed in the formulation of hydrogels. Natural polymers like gelatin, collagen, hyaluronic acid, alginates, and chitosan have been preferred for their non-toxic, non-immunogenic, biodegradable, and cell adhesion traits. On the other hand, synthetic polymers such as polyethylene glycol (PEG), poly(N-isopropylacrylamide) (PNIPAAm), poly(acrylic acid) (PAA), poly(vinyl alcohol) (PVA), and poly(propylene fumarate) (PPF) are preferred for their tunable properties in chemical
structure, mechanical strength, degradation rate, and consistency.\(^7,8\)

To achieve a controlled release of the encapsulated drugs, hydrogel mesh size can be manipulated by swelling or degradation of the polymer network.\(^5\) Depending on the crosslinker concentration, the mesh size ranges from 5-100 nm, and it can be manipulated by swelling caused by internal or external factors such as pH, temperature, or light. Alternatively, mesh size can be increased by degradation of the polymer network by temperature, pH, light or enzymatic hydrolysis causing the drug to leak out.\(^5\) However, controlling the mesh size can be challenging due to the network's heterogeneity and the polymer's polydispersity. Additionally, drug release cannot be controlled after the mesh sizes reach a point of gaping, limiting the extent of controlled release from hours to a day.

Another way of controlling the drug release from hydrogels can be achieved by engineering drug-polymer interactions. This strategy can be necessary for encapsulating drugs
too small to contain by adjusting the mesh size. While sulfonate groups and cyclodextrin can be installed on the polymer chain to increase the electrostatic and hydrophobic interactions with charged and hydrophobic drugs, respectively, the drugs' structure cannot be altered easily to make them charged or hydrophobic. Therapeutic molecules have been directly conjugated to the polymer structure with stable covalent bonds by employing N-hydroxysuccinimide (NHS), thiol-ene, or strain-promoted azide-alkyne cycloaddition (SPAAC) click reactions. Although this type of hydrogel system has been demonstrated in cell culture and tissue engineering applications, the stable covalent polymer-drug bond prohibits the release of the conjugated drugs. Although their drug release cannot be controlled externally; hydrolytic ester, protease cleavable peptides, and engineered bonds cleavable by β elimination have also been used to encapsulate drugs in hydrogels.

In this study, we have developed photo-responsive polymers that can be covalently conjugated with surface lysine residues of native proteins of interest (POI) or amino/hydroxyl groups of small molecule drugs (Figure 4.2 and 4.3). After conjugation, the polymer-drug complex can be crosslinked to form a hydrogel that can be drop-cast onto an MDD device as a drug depot. With our design, covalent conjugation of the drugs, whether they are proteins or small molecules, ensures there will not be any premature drug release regardless of the hydrogel mesh size. Photo-cleavable polymer-drug bonds allow external spatiotemporal control of the drug release. Moreover, the self-immolative nature of the polymer-drug linker allows the traceless release of the conjugated therapeutics.

Growth factors released from hydrogels have been widely studied for tissue engineering and wound healing applications, but release control remains an issue. With enabling features of our hydrogel system, our first objective is to demonstrate the controlled release of proteins
such as growth factors. Our second objective is to illustrate the application of our hydrogel system for delivering small-molecule drugs. Due to its off-target effects and fast clearance rate, Dexamethasone (Dex) release systems have long been researched as an alternative to its systemic use for ophthalmology, anti-inflammatory, cancer, postoperative, and neurological treatment applications. We envision the covalent encapsulation of Dex in our hydrogel systems will enable strict spatiotemporal control over its release and upgrade the current Dex release systems in their applications.

4.2 Results and Discussion

4.2.1 Design and synthesis of the photo-responsive polymers

The design of our polymers consists of two repeating units: ~ 500 Da poly(ethylene glycol) (PEG) methacrylate and light responsive 2-(4-nitro-3-
(((perfluorophenoxy)carbonyl)oxy)methyl)phenoxy)ethyl methacrylate (Figure 4.3a). While the PEG unit ensures water solubility of the polymer in low to high hydrogel concentrations, the nitrobenzyl group is installed for the photo-responsive release. Nitrobenzyl units have been widely used in the formulation of photo-responsive hydrogels. Here, we took the design of those systems one step further by making not just the polymer network but also the polymer-drug covalent bond photo-responsive. Monomers with nitrobenzyl groups are functionalized with pentafluorophenyl (PFP) carbonate groups that can react with amine or hydroxyl groups to generate photo-responsive nitrobenzyl carbamate or carbonate groups, respectively.

In addition to the degradable hydrogel, we have designed a non-degradable hydrogel system in which only the polymer-drug bonds are photo-responsive. While the degradation of the polymer network might facilitate the release of large therapeutics proteins, this might not be necessary to release small molecules, which can quickly diffuse out of large hydrogel meshes. Moreover, keeping the network intact might make the estimation of drug release easier, as the polymer mesh size would not change, resulting in constant retardation of drug release over time.

In addition to the PEG and nitrobenzyl repeating units, the non-degradable hydrogel polymers feature azidoethyl methacrylate units (Figure 4.3). In the first step of non-degradable hydrogel formulation, small molecule drugs with amino or hydroxyl moieties is conjugated to the polymer.

Figure 4.5: a) Synthesis of the photo-responsive polymer P2; b) GPC chromatogram of P2. \( M_n = 26 \text{ kDa}, \text{ PDI} = 1.22 \).
backbone by reacting with PFP carbonate groups. After purification of the drug-conjugated polymer, the solution of drug-polymer conjugate is cross-linked by the SPAAC reaction of the azidoethyl methacrylate unit in the polymer backbone and four arms PEG dibenzocyclooctyne (DBCO). Since this resulting bond is not photo-responsive, the hydrogels is not degraded even while it is photo-responsively releasing the encapsulated drugs.

To optimize the degradable hydrogel polymer P1, various reversible addition–fragmentation chain-transfer (RAFT) reaction conditions with different solvent and initiator combinations has been explored (Figure 4.4 a,b). These solvents include dimethyl
formamide (DMF), toluene, dioxane, dimethyl sulfoxide (DMSO), tetrahydrofuran (THF), and trifluoroethanol (TFE). To conduct the RAFT polymerizations in different temperatures, initiators azobisisobutyronitrile (AIBN) and V-40 with 10-hour half-life temperatures of 65°C and 88°C, respectively, were used. Among these different reaction conditions, DMF and AIBN resulted in the best monomer-to-polymer conversion (Figure 4.4 b). Gel permeation chromatography (GPC) of the polymer showed a polydispersity index (PDI) of 1.18 and a molecular weight of 31 kDa, close to the targeted molecular weight of 30 kDa (Figure 4.4 c). The same reaction conditions were followed to synthesize the non-degradable polymer P2 hydrogel with 26 kDa molecular weight and PDI of 1.22 (Figure 4.5).

To confirm the photoresponsive nature of the polymer, 500 uL of pH 7.4 phosphate-buffered saline (PBS) solution of 2 mg/mL P1 was subjected to 365 nm UV light for 2 hours. UV absorption of the solution from 8 different time points from the UV exposure revealed a decrease in the signature absorption peak of 308 nm for the nitrobenzyl carbonate group resulting from its photocleavage (Figure 4.6b).32

4.2.2 Formulation of protein-encapsulated hydrogels

To evaluate the protein conjugation efficiency by the P1 polymers, various P1:protein ratios of bovine serum albumin (BSA) and equine cytochrome c was incubated overnight in pH 8 PBS buffer while shaking. Then samples from each reaction were run on SDS PAGE and visualized after coomassie staining (Figure 4.7 a,b). As the protein polymer bioconjugation results in large protein-polymer complexes, the resulting complex runs slower on the SDS-PAGE and does not align with the original protein band. Therefore, compared to the thick protein band from the control BSA lane incubated without any P1, increasing polymer concentrations diminished the BSA band from right to left. Encapsulation efficiency was calculated by the ratios of band
intensities from each P1:BSA lane and the control BSA lane with no P1 (Figure 4.7 c). The resulting graphs showed near complete protein conjugation with P1:BSA of 10000 and above.

The same experiment with cytochrome c showed complete protein conjugation with all P1:cytochrome c ratios tested (Figure 4.7 b). The high number of surface-exposed lysine residues and their amine side chains allow fast and efficient bioconjugation of cytochrome c. While BSA has a molecular weight of 66.34 kDa and 32 surface-exposed lysines, cytochrome c has a
molecular weight of 12.38 kDa and 15 surface-exposed lysines (Figure 7 d,e). Solvent accessible surface area (SASA) of the lysine residues was calculated using GetArea.\textsuperscript{34}

Having demonstrated the protein encapsulation capability of P1 polymers, we have formulated fluorescein isothiocyanate (FITC) BSA encapsulated hydrogels and confirmed photoresponsive release (Figure 4.8). Using the ideal polymer/protein bioconjugation ratio 10000, P1 and FITC BSA were incubated for 8 hours in pH 8.5 PBS buffer at room temperature while shaking. Then, a calculated amount of 10 kDa 4 arm PEG amine was added to the reaction mixture to crosslink the protein-polymer bioconjugate overnight and formulate 500 uL 5 weight (wt) % hydrogels in a 1.6 mL Eppendorf tube. Then 200 uL of PBS solution was added on top of the solid hydrogel, and

**Figure 4.8:** Photo-responsive protein release of P1 hydrogels. a) Cumulative FITC BSA release over 2 hours; b) transformation of solid P1 hydrogel to solution; comparison of hydrogel released c) FITC BSA and d) cytochrome c CD spectra with their control counterparts.
the Eppendorf tube was subjected to 2 hours of 365 nm UV exposure. To calculate the FITC BSA release %, the fluorescence of 100 uL solution from the top of the tube was measured (Figure 4.8 a). These results show we get 100 % protein release in ~100 mins of UV exposure. Since the P1 crosslinker bonds are also nitrobenzyl carbonate and photoresponsive, the solid hydrogels slowly turn into solution during the 2 hours of UV exposure (Figure 4.8 b). Circular Dichroism (CD) spectra of the hydrogel-released FITC BSA and cytochrome c align with the spectra of their control counterparts, demonstrating the structural integrity of the proteins after photoreponsive release (Figure 4.8 c,d).

4.2.3 Formulation of small molecule encapsulated hydrogels and their photoresponsive release

To demonstrate the ultimate test of controlled release, we proceeded with encapsulation and photo-responsive release of small molecules. While the release of large therapeutics like proteins can be controlled simply by the hydrogel mesh size, hydrophilic/hydrophobic or covalent interactions must be incorporated into the hydrogel system for the encapsulation and controlled release of small molecules. The PFP carbonate in the P2 polymer backbone allows covalent conjugation of hydroxyl or amine groups containing small molecules, and the resulting nitrobenzyl carbonate or carbamate groups can be degraded photoresponsively to release the drugs.

For ease of release quantification, hydroxyl conjugation was first demonstrated by 7-(Diethylamino)-4-(hydroxymethyl)coumarin conjugation with P2 polymers (Figure 4.9). A conjugation efficiency of 71 % was calculated by the ratio of intensities of the “d” proton on coumarin and “a” proton on the nitrobenzyl group (Figure 4.9 a,b). After coumarin conjugation, $M_n= 26 \text{ kDa}$ and PDI = 1.22 of P2 polymer shifted to $M_n= 27 \text{ kDa}$ and PDI = 1.3 (Figure 4.9).
The color of the transparent P2 solution changed to green coumarin color (Figure 4.9 e).

To demonstrate the photo-responsive release, a 10 mg/mL solution of coumarin conjugated P2 was subjected to 4 hours of UV irradiation, and coumarin release was monitored by HNMR spectroscopy (Figure 4.10). The ratio of proton peak intensities “b” and “a” on the
released and conjugated coumarin revealed 82% coumarin release in 4 hours.

Coumarin fluorescence standard curve was plotted to quantify the release (Figure 4.11 a). To formulate 500 uL, 5 weight% nondegradable hydrogel, a calculated amount of 10 kDa four-arm PEG DBCO crosslinker was added to the solution of coumarin conjugated P2 polymers in PBS. After a couple of pipette resuspensions, the solid hydrogel was formed in a 1.6 mL Eppendorf tube. Before subjecting the Eppendorf tube to UV irradiation, 200 uL of PBS was added on top of the hydrogel to collect the released coumarin. Then, the Eppendorf tube was

Figure 4.10: a) Photo-responsive coumarin release of P2 polymer; b) quantification of photo-responsive coumarin release by HNMR spectroscopy.
subjected to 5, 15, 30, 60, and 120 minutes of pulsative UV irradiations with 1 hour between each pulse to allow diffusion of the released coumarin (Figure 4.11 c). As the amount of coumarin encapsulated in the hydrogel is constant, coumarin release for a duration of UV irradiation decreased over time, with ~75% release for the total UV irradiation of 2 hours and 50 minutes. When the hydrogel was incubated at room temperature and was not subjected to UV light, coumarin release was negligible, amounting to only ~2.5% in 7 days (Figure 4.11 d). Since the coumarin-conjugated P2 polymers were crosslinked by a non-photo responsive DBCO-azide
Figure 4.12: a) Conjugation of dexamethasone with P2 polymer; b) HNMR spectrum of the dexamethasone conjugated P2; GPC chromatogram of P2 polymer c) before (Mn = 26 kDa, PDI = 1.22) and d) after dexamethasone conjugation (Mn = 29 kDa, PDI = 1.28).

bond, the hydrogel stayed in solid form for the entirety of the UV exposure with slight color
change (Figure 4.11 b).

Following the demonstration of coumarin conjugation and release, P2 polymers were conjugated with dexamethasone following the same procedure for coumarin (Figure 4.13). A conjugation efficiency of 51 % was calculated by the intensity ratio of “d” protons on Dex and “b” protons on the nitrobenzyl group (Figure 4.9 a,b). After Dex conjugation, $M_n = 26$ kDa and PDI = 1.22 of P2 polymer shifted to $M_n = 29$ kDa and PDI = 1.28 (Figure 4.9 c,d). Finally, the LC-MS method from the literature was followed to generate a standard curve for the quantitation of dexamethasone release (Figure 4.13).\textsuperscript{35}

4.3 Conclusion

In summary, we have designed and synthesized photo-responsive polymers for covalent encapsulation and release of proteins and small molecules. While photo-degradable P1 hydrogels were formulated for proteins, non-degradable P2 hydrogels were formulated for small molecule
encapsulation and release. The efficiency of protein encapsulation depends on the protein, and it can be 74 nM for proteins like BSA or 300 nM for proteins with a high number of surface-exposed lysine residues such as cytochrome c. Complete protein release of FITC BSA encapsulated hydrogels was achieved in ~100 minutes of UV stimulation, and the structural integrity of the released proteins was confirmed by CD spectroscopy. Using covalently conjugated drug-P2 polymers, drug concentrations of ~10 µM hydrogel and ~75% release were accomplished with a total UV illumination time of 3 hours. Finally, we have optimized the LC-MS quantitation method for Dexamethasone.

4.4 Experimental details

4.4.1 Materials and instrumentation

All chemicals and materials were purchased from commercial sources unless stated otherwise. BSA and cytochrome c from the equine heart were purchased from Sigma-Aldrich. Amine and DBCO 4-arm PEG crosslinkers were purchased from CreativePEGWorks. UV-Vis spectra were recorded on PerkinElmer Lambda 35 UV/Vis spectrometer. Plate reader experiments were conducted on SpectraMax iD. The Dexamethasone standard curve was generated using Waters Xevo TQD Triple Quadrupole Mass Spectrometer. 1H-NMR and 13C-NMR spectra were recorded on either Bruker 400 or 500 MHz NMR spectrometer. Black-Ray UV Bench Lamp |365nm| 115V~60Hz| 0.68 Amps were used in all UV experiments.

4.4.2 Formulation of 5 wt% unloaded hydrogels

To make 5 weight % unloaded hydrogel (no protein or small molecules encapsulated) hydrogel, solution of both polymer and crosslinker components was prepared separately. Although crosslinker solution can be made in bulk and stored in the fridge, the polymer solutions must be
made fresh each time, as the polymer contains reactive and labile carbonate groups. To make ~1 mL solution of the 5 weight % crosslinker solution, 50 mg of the crosslinker is dissolved in 950 uL of PBS solution. With the assumption of PBS having the same 1 mg/uL density as water, this solution will have mass of 50 mg crosslinker plus 950 mg of PBS totaling 1000 mg. 5 % of the total weight of this solution is contributed by the crosslinker, therefore 5 weight % solution. In the same way, 5 weight % polymer solution was prepared.

To calculate how much of each solution is needed to make the final hydrogel, concentrations of reactive PFP carbonate [PFP] and amine groups [NH2] in the polymer and crosslinker solutions respectively were calculated. The molecular weights of P1 and amine crosslinker C\ NH2 are ~30000 g/mol and 10000 g/mol respectively. Molar concentrations of 1 mL, 50 mg/mL solution of polymer [P1] and amine crosslinker [C\ NH2] were calculated following the equations below.

\[
[P1] = \frac{0.05\,g}{30 \times 10^3\,g/mol} \times \frac{1}{10^{-3}L} = 1.67 \times 10^{-3} \, \frac{mol}{L} = 1.67 \times 10^{-3} \, M
\]

\[
[C_{NH2}] = \frac{0.05\,g}{10 \times 10^3\,g/mol} \times \frac{1}{10^{-3}L} = 5 \times 10^{-3} \, \frac{mol}{L} = 5 \times 10^{-3} \, M
\]

The ratio of PEG and PFP carbonate monomers in the P1 polymer is 3:1 and their molecular weights are 500 g/mol and 491.32 g/mol. To calculate the concentrations of PFP reactive groups [PFP], the number of PFP groups in each polymer (N\ PFP) was calculated following the equation below.

\[
N_{PFP} \times 491.32 \, \frac{g}{mol} + N_{PFP} \times 3 \times 500 \, \frac{g}{mol} = 30000 \, \frac{g}{mol}
\]
\[ N_{PFP} \approx 15 \]

Then, using \([P1]\) and \(N_{PFP}\), the concentration of reactive PFP carbonate groups \([PFP]\) in the polymer solution was calculated.

\[
[PFP] = 15 \times [P1] = 15 \times 1.67 \times 10^{-3} \text{ M} \approx 0.025 \text{ M}
\]

Each crosslinker has 4 arms of reactive amines. Therefore, the concentration of amine groups \([NH2]\) is calculated using the formula below.

\[
[NH2] = 4 \times [C_{NH2}] = 4 \times 5 \times 10^{-3} \text{ M} = 0.02 \text{ M}
\]

Finally, the ratio \([PFP]\) and \([NH2]\) is calculated below.

\[
\frac{[PFP]}{[NH2]} = \frac{0.025}{0.02} = 1.25
\]

Therefore, to make a hydrogel of 2.25 mL, 1.25 mL of 4 arm PEG amine solution needs to be added to a solution of 1 mL P1 polymer.

### 4.4.3 Formulation of protein-encapsulated hydrogels

To prepare a protein-encapsulated hydrogel, polymer/protein ratio for near 100% protein conjugation was found following the SPS PAGE experiments (Figure 4.7). For BSA and cytochrome c, this ratio is 10000 and 2500 respectively. Using the procedure provided for the unloaded hydrogels, 5 wt% solution of 4 arm PEG amine crosslinker was prepared in pH 8.5 PBS buffer. Following the procedure for unloaded hydrogels, \([P1]\) in 5 wt% solution was calculated. Then, \([P1]/10000\) and \([P1]/2500\) concentrations of BSA and cytochrome c solutions...
were prepared in pH 8.5 PBS. An appropriate amount of P1 was weighed for the planned experiment and the protein solutions were used directly to prepare 5 wt% polymer solution using the procedure provided for the unloaded hydrogel. Then, protein + polymer solution was incubated in room temperature while shaking for 8 hours. Finally, to the volume of this protein + polymer solution, 1.25 times volume of 5 wt% 4 arm PEG crosslinker solution in pH 8.5 PBS is added and incubated overnight while shaking at room temperature (Considering that some of the PFP carbonates on P1 have reacted with the pre-incubated protein, this volume of 4 arm PEG is excess). Therefore, with 1/2.25 dilution, concentration of [P1] after hydrogel formation [P1H] is $7.42 \times 10^{-4}$ M. [BSA] is $[P1_H]/10000 = 7.42 \times 10^{-8}$ M, and [cyt c] is $[P1_H]/2500 = 2.99 \times 10^{-7}$ M. Using their molecular weight of 66.43 kDa and 12 kDa respectively, weight concentrations of [BSA] = 4.93 µg/mL and [cyt c] = 3.59 µg/mL were calculated.

**4.4.4 Formulation of small molecule encapsulated hydrogels**

After the conjugation of the small molecules to the P2 polymers, similar procedures were followed to make the nondegradable hydrogel. First, 5 wt% of small molecule conjugated P2 and DBCO crosslinker was prepared separately in pH 7.4 PBS. To make the final hydrogel, concentrations of reactive azide [N3] and DBCO [DBOC] in the polymer and crosslinker solutions respectively were calculated. The molecular weights of P2 and DBCO CDBCO are ~30000 g/mol and 10000 g/mol respectively. Since P2 has the same molecular weight as P1, $[P2] = [P1]$. Just like the amine crosslinker, the DBCO crosslinker has 4 reactive arms and molecular weight of 10000 g/mol. Therefore, $[NH2] = [DBCO]$.

The ratio of PFP/PEG/Azide monomers in the P2 polymer is 25:55:20 and their molecular weights are 491.32, 500, and 139.16 g/mol respectively. To calculate the concentrations of N3 reactive groups $[N3]$, the number of azide groups in each polymer ($N_{N3}$) was calculated following
the equation below.

\[
N_{N3} \times 139.16 \frac{g}{mol} + 1.25 \times N_{N3} \times 491.32 \frac{g}{mol} + 2.75 \times N_{N3} \times 500 \frac{g}{mol} = 30000 \frac{g}{mol}
\]

\[N_{N3} \approx 14.1\]

Then, using \([P2]\) and \(N_{N3}\), concentration of reactive azide groups \([N_3]\) in the polymer solution was calculated.

\[[N_3] = 14.1 \times [P2] = 14.1 \times 1.67 \times 10^{-3} M \approx 0.024 M\]

Each crosslinker has 4 arms of reactive amines. Therefore, the concentration of amine groups \([\text{NH2}]\) is calculated using the formula below.

\[[\text{DBCO}] = 4 \times [C_{DBCO}] = 4 \times 5 \times 10^{-3} M = 0.02 M\]

Finally, the ratio \([\text{PFP}]\) and \([\text{NH2}]\) is calculated below.

\[
\frac{[N_3]}{[\text{DBCO}]} = \frac{0.024}{0.02} = 1.2
\]

Therefore, to make hydrogel of 2.2 mL, 1.2 mL of 4 arm PEG DBCO solution needs to be added to a solution of 1 mL \(P2\) polymer. The hydrogel forms in a matter of seconds.

To calculate the small molecule encapsulated in the hydrogel, we need to first calculate \([\text{PFP}]\), and concentration of PFP in the hydrogel \([\text{PFP}_H]\). Since we know that we have 1.25 times PFP as \(N_3\) in the polymer backbone, \([\text{PFP}] = [N_3] \times 1.25 = 0.024 M \times 1.25 = 0.03 M\). Then, from the ratio of polymer and crosslinker solution, we can calculate the concentration of PFP in the hydrogel as \([\text{PFP}_H] = [\text{PFP}]/2.2 = 0.014 M\). The HNMR spectrum reveals that 71% and 51% of the PFP is conjugated with coumarin and Dexamethasone respectively. Therefore,
concentrations of coumarin and Dexamethasone in the hydrogels are $9.94 \times 10^{-3}$ M and $7.14 \times 10^{-3}$ M respectively.

4.4.5 Synthesis of the PFP monomer, P1 and P2 polymers

**Synthesis of compound PFP a**

(Hydroxyethyl)methacrylate (10 g, 76.84 mmol) and triethylamine (11.66 g, 115.26 mmol) were dissolved in 350 mL dichloromethane (DCM). To the stirring reaction mixture in ice bath, solution of 4-Toluenesulfonyl chloride (18.31 g, 96.05 mmol) in DCM was added dropwise. After overnight reaction, the reaction mixture was washed by water and brine sequentially in a separatory funnel. The organic layer was collected, dried over anhydrous sodium sulfate, and subjected to purification by flash chromatography on silica gel with gradient of hexanes and ethyl acetate. Yield: 14.3 g, 33%. $^1$H NMR (400 MHz, CDCl$_3$) δ 7.80-7.78 (d, 2H), 7.35-7.33 (d, 2H), 6.06 (s, 1H), 5.57 (s, 1H), 4.32-4.30 (t, 2H), 4.27-4.25 (t, 2H), 2.45 (s, 3H), 1.89 (s, 3H). $^{13}$C NMR (101 MHz, CDCl$_3$) δ 167.02, 145.22, 135.72, 132.98, 130.12, 128.12, 126.66, 67.81, 62.01, 21.84, 18.34.
Synthesis of compound PFP b

PFP a (12.61 g, 284.33 mmol), 4-(Hydroxymethyl)-3-nitrophenol (5 g, 29.56 mmol), K2CO3 (8.17 g, 59.12 mmol), and 18-crown-6 (3.91 g, 14.78 mmol) were dissolved in 200 mL DMF and stirred overnight at room temperature. After overnight reaction, the reaction mixture was dissolved in ethyl acetate and washed by brine three times. The organic layer was collected, dried over anhydrous sodium sulfate, and subjected to purification by flash chromatography on silica gel with gradient of hexanes and ethyl acetate. Yield: 2.6 g, 31%. 1H NMR (400 MHz, CDCl3) δ 8.19-8.17 (d, 1H), 7.27-7.27 (d, 1H), 6.93-6.90 (dd, 1H), 6.14-6.14 (t, 1H), 5.61-5.60 (t, 1H), 5.01-4.99 (d, 2H), 4.55-4.52 (t, 2H), 4.35-4.33 (t, 2H), 1.95-1.95 (t, 3H). 13C NMR (101 MHz, CDCl3) δ 167.38, 163.27, 162.75, 140.71, 135.96, 128.14, 126.55, 114.69, 113.82, 66.73, 62.96, 62.68, 18.45.

Synthesis of compound PFP monomer

PFP b (1 g, 3.55 mmol) and proton sponge (0.84 g, .913 mmol) were dissolved in 100 mL of DCM. To the stirring solution in ice bath, solution of Bis(pentafluorophenyl) carbonate (1.75 g, 4.44 mmol) in DCM was added dropwise. The reaction was left stirring at room temperature overnight. The reaction mixture was washed by water and brine sequentially in a separatory funnel. The reaction mixture was added more DCM and washed by water and brine sequentially in a separatory funnel. The organic layer was collected, dried over anhydrous sodium sulfate, and subjected to purification by flash chromatography on silica gel with a gradient of hexanes and ethyl acetate. Yield: 1.4 g, 80%. 1H NMR (500 MHz, CDCl3) δ 8.29-8.27 (d, 1H), 7.19-7.18 (d, 1H), 7.01-6.99 (dd, 1H), 6.15-6.15 (t, 1H), 5.62-5.62 (t, 1H), 4.57-4.55 (t, 2H), 4.37-4.35 (t, 2H), 1.96-1.96 (t, 3H). 13C NMR (101 MHz, CDCl3) δ 167.35, 163.27, 162.75, 140.71, 135.96, 128.14, 126.55, 114.69, 113.82, 66.73, 62.96, 62.68, 18.45.
**Synthesis of P1 and P2 polymer**

**PFP monomer** (150 mg, 0.34 mmol), Polyethylene glycol monomethacrylate (PEGMA, 500 g/mol) (516.7 g, 1.03 mmol), and 44-Cyano-4-[(dodecylsulfanylthiocarbonyl)sulfanyl]pentanoic acid (8.97 mg, 0.02 mmol) were dissolved in ~ 2 mL of DMF in a 7 mL vial. To this stirring reaction mixture, AIBN (0.73 mmol, 4.44 × 10⁻³ mmol) in DMF was added and it was stirred for a couple of minutes until everything mixed well. The vial was purged with argon and subjected to three freeze, pump, and thaw cycles. After 15 hours of stirring at 69 °C, the reaction was quenched in liquid nitrogen and the polymer was precipitated in cold diethyl ether. The pure polymer was obtained after three cycles of dissolution in THF and precipitation in diethyl ether. Yield: 96% total monomer conversion by $^1$H NMR. GPC (DMF) $M_n = 31$ kDa, PDI = 1.18. $^1$H NMR (400 MHz, CDCl₃) δ 8.25, 7.19, 7.12, 5.78, 4.33, 4.06, 3.64, 3.37, 1.79, 1.00, 0.88.

Polymer P2 was synthesized following the same procedure above. Yield: 85% total monomer conversion by $^1$H NMR. GPC (TFE) $M_n = 26$ kDa, PDI = 1.22. $^1$H NMR (400 MHz, CDCl₃) δ 8.26, 7.18, 7.09, 5.78, 4.33, 4.08, 3.64, 3.37, 1.66, 1.03, 0.88.

**Conjugation of small molecules with polymer P2**

Polymer P2 (500 mg, 0.235 mmol N₃), Dexamethasone (461 mg, 1.18 mmol) and 4-Dimethylaminopyridine (28.7 mg, 0.235 mmol) were dissolved in 2 mL DMF, and the reaction mixture was left stirring overnight. Then, the reaction mixture was transferred into a 3 kDa cutoff dialysis membrane, and the Dexamethasone conjugated polymer was dialyzed in 1 L of 1:1 DCM, methanol for two days. The dialysis solvent was changed 3 times over the course of 2 days. Yield: 51% conjugation calculated by $^1$H NMR. The same procedure above was followed to synthesize coumarin-conjugated polymer. Yield: 71% conjugation calculated by $^1$H NMR.
Figure 4.14: HNMR of PFP monomer.

Figure 4.15: $^{13}$C NMR of PFP monomer.
4.5 References


Schoenmakers, R. G.; Van De Wetering, P.; Elbert, D. L.; Hubbell, J. A. The Effect of the


CHAPTER 5

SUMMARY AND FUTURE DIRECTIONS

5.1 Summary

In summary, we have briefly discussed reactive chemistries for modifying proteins and their applications in chemical biology, proteomics, and therapeutics. The focus of this thesis was the development of chemistries for the targeted modification of endogenous proteins and their applications in targeted protein degradation (TPD). Additionally, reactive polymers were developed to covalently conjugate therapeutics in hydrogels and tracelessly release them on demand.

In chapter 2, LD-TMAcs for target-selective modification of endogenous proteins was developed. The modular structure of LD-TMAcs allowed us to functionalize them into either covalent inhibitors or multi-site labeling reagents. We have demonstrated that labeling by LD-TMAc covalent inhibitors results in complete loss of activity, and multi-site labeling reagents can label several residues around the binding pocket that include: Histidine, Lysine, Serine, and Threonine. After reacting the model protein carbonic anhydrase with LD-TMAcs, we identified the modified residues around the binding pocket and showed a correlation between their modification extent, proximity to the ligand binding site, and surface accessibility. LD-TMAcs are target selective, and we have demonstrated the utility of this method by selectively labeling membrane-bound carbonic anhydrase XII in live cells.

In chapter 3, the modularity of TMAc allowed us to develop multi-site labeling LD-TMAcs into Hydrophobic Tagging (HyT) targeted protein degraders. HyT LD-TMAcs are functionalized with pendant adamantane units on their ester group, and we have demonstrated that CA-targeted HyT LD-TMAcs can install up to 3 adamantane units per protein molecule. Our western blotting
studies showed HyT LD-TMAcs degraded ~75 CAXII on the MCF7 membrane in 3 hours with DC$_{50}$ of 1 µM. Inhibition of lysosomal and proteasomal degradation and ectodomain shedding did not fully stop HyT-induced protein degradation by LD-TMAcs, indicating HyT CAXII is degraded by multiple pathways. Additionally, tracking of hydrophobically tagged CAXII under confocal microscopy revealed that some CAXII ends up outside the cells while others go inside the cells post-hydrophobic tagging. Finally, PDL-1-targeted HyT LD-TMAcs did not show any protein degradation.

In Chapter 4, we have developed polymers with PEG and reactive nitrobenzyl carbonate groups. While the PEG repeating units confer water solubility, nitrobenzyl carbonate groups can covalently conjugate therapeutics and tracelessly release the conjugates on-demand upon light stimulus. Following covalent protein conjugation, hydrogels were formulated by crosslinking the leftover reactive carbonates on the polymer backbone with 4-arm PEG amine. ~100% protein release was demonstrated in ~100 minutes of UV stimulation, and the structural integrity of the released proteins was confirmed by CD spectroscopy. Additionally, these polymers incorporated azide repeating units to crosslink with PEG DBCO after the covalent conjugation of small molecules on the reactive carbonates. Owing to the non-photo-responsive nature of DBCO-azide bonds, these hydrogels can photo-responsively release the encapsulated small molecules without degradation or converting them into a solution. ~75% photo-responsive release in 3 hours was demonstrated in coumarin-conjugated non-degradable hydrogels.

**NOTE**

5.2 Future Directions

5.2.1 LD-TMAcs to investigate opioid off-targets and dependency

In Chapter 2, we developed highly reactive LD-TMAcs with fast target selective protein labeling kinetics and demonstrated their utility by selectively labeling membrane-bound carbonic
anhydrase XII in live cells. Compared to the other LD approaches with stringent amino acid selectivity, LD-TMAcs have a broad amino acid scope, resulting in unprecedented multisite labeling of the target protein around the ligand binding site. This broad amino acid scope could enable LD-TMAcs to target many proteins where other approaches fall short due to the lack of compatible amino acids around the binding pocket. We envision unique features of our LD-TMAc platform will find use in target/off-target identification and investigation of binding/allostERIC sites.

Opioids are the most potent class of analgesics, and their abuse led to ~90000 overdoses in the United States in 2020. Some of the cardiovascular, respiratory, neurocognitive, and dependency complications of opioids are associated with opioid and their metabolites’ off-target effects. Considering their target selectivity and broad amino acid scope, LD-TMAcs should be able to label opioid targets and off-target proteins when opioids are used as ligands. In this approach, opioid-guided LD-TMAcs could be used to label a crude proteome, followed by an analysis of the labeled proteins with SDS/PAGE and Mass Spectrometry (MS) to identify the hit proteins as targets or off-targets. Identifying the off-targets of opioids and their metabolite fragments will inform what structural features of these molecules results in off target-activities and complications.

Recent studies have shown that opioids with a certain degree of selectivity towards µ
Opioid Receptor - δ Opioid Receptor (MOR-DOR) heterodimers show less severe side effects, dependence, and withdrawal.\(^5\)-\(^7\) However, these opioids do not have enough selectivity and are far from optimal. To optimize this selectivity, structure-activity knowledge of opioid-MOR-DOR binding is essential. It is speculated that the heterodimer selective opioids bind somewhere in the heterodimeric sites, but the X-ray crystal or cryo-EM structure of these heterodimers is unknown to pinpoint the exact binding site.\(^5\) Considering LD-TMAcs’ ability to map the ligand binding site with their multi-site labeling feature, opioid-guided MOR-DOR labeling with LD-TMAcs could shed light on the binding site of the MOR-DOR selective opioids.

5.2.2 Targeted Degradation of GPCRs and Ion Channels

In Chapters 1 and 3, we discussed the use of Halo-Tags in the so-called Hydrophobic Tagging (HyT) approach to target-selectively label proteins with hydrophobic moieties to induce their degradation.\(^8\) However, the application of this approach is strictly limited to chemical biology as Halo-Tag systems are genetically engineered. To the best of our knowledge, endogenous target-selective protein labeling approaches were not employed for HyT-induced targeted protein degradation (TPD) up to this point. Additionally, hydrophobic tagging of extracellular membrane proteins for TPD was not explored.

In chapter 2, by target-selectively degrading membrane CAXII on MCF7 cells, we have shown HyT-induced TPD with an endogenous protein labeling approach for the first time. This is also the first instance of extracellular HyT-induced TPD membrane proteins. In addition to carbonic anhydrase, there are many other disease targets on cell membranes, such as G Protein Coupled Receptors (GPCRs) and ion channels with extracellular domains that can be targeted with small molecules.\(^9\),\(^10\) It will be interesting to see the use of agonists, not just antagonists, for HyT-induced TPD, converting them into antagonists by the outcome. Our next goal is
demonstrating the HyT-induced TPD of GPCRs and ion channels. By synthesizing cetirizine-conjugated LD-TMAc, we aim for the TPD of Histamine Receptor H1 (HRH1) (GPCR) (Figure 5.1). Additionally, we aim to target nicotinic acetylcholine receptors (nAchR) (ion channels) and muscarinic acetylcholine receptors (nAchR) (GPCRs) with carbachol conjugated LD-TMAcs (Figure 5.1).

5.2.3 Circadian synchronized cell culture with stimuli-responsive Dexamethasone-releasing hydrogels.

Every day of our lives is governed by awake and sleep cycles with different physiological processes. This phenomenon is called the circadian rhythm, and it is controlled by the body’s internal clock with fluctuating levels of melatonin and cortisol. Cortisol levels peak in the morning to stimulate wakefulness and alertness and drop towards dusk. Then melatonin levels start to rise in the evening to promote sleep and fade as the day sees light in the morning. In
addition to the systemic circadian rhythm, every cell in our body has its own circadian rhythm and internal clock, synchronized by systemic cortisol. BMAL1 and PER2 proteins are among the key players of cellular circadian rhythm, and their expression levels oscillate daily in opposite phases. In the liver, there are about 3000 genes under BMAL control, and 56 of 100 top-selling drug targets are under the control of cellular circadian rhythm. Internal clock of every cell in our body is synchronized by cortisol; therefore, a drug targeting a particular protein will have different effects during daytime or nighttime as the levels of the targeted protein could be highest or lowest during the time of treatment (Figure 5.2). However, cells in a typical cell culture have non-synchronized clock resulting in the uniform expression level of proteins in the whole cell culture medium. Hence, the same drug in cell culture will result in the same effect no matter what time of the day cells are treated. This result, however, does not reflect what would happen systemically when all the cells have synchronized clocks. Therefore, advanced cell culture

Figure 5.3: PER2 and BMAL1 expression levels in circadian synchronized and non-synchronized cells.
systems with synchronized circadian rhythms are necessary to better reflect the cellular response to drugs.

Dexamethasone (Dex) is about 30 times more potent than cortisol, and it is used to synchronize the circadian rhythm of cell culture by ~ 2-hour Dex treatment right after seeding.\textsuperscript{13,16} However, oscillations in levels of circadian proteins BMAL1 and PER2 dissipate after only 2.5 days (Figure 5.3).\textsuperscript{13} In Chapter 4, we developed hydrogels that can release covalently encapsulated molecules with photo-responsive control. Compared to other hydrogels in the literature, covalent encapsulation of drugs in our system prohibits premature release. Additionally, photo-responsive drug-polymer bonds enable active drug molecules' controlled and traceless release. Toward our future goal, we have synthesized Dex-conjugated polymers, which can be used to make hydrogels that photo-responsively release Dex. Moving forward, our goal in this project is to make hydrogels for advanced cell culture setting that can controllably release Dex to synchronize circadian rhythm in the long term.

5.3 References


(4) Green, J. M.; Sundman, M. H.; Chou, Y. hui. Opioid-Induced Microglia Reactivity


(10) Bagal, S. K.; Brown, A. D.; Cox, P. J.; Omoto, K.; Owen, R. M.; Pryde, D. C.; Sidders, B.; Skerratt, S. E.; Stevens, E. B.; Storer, R. I.; et al. Ion Channels as Therapeutic Targets: A


BIBLIOGRAPHY


Hoelzel, C. A.; Zhang, X. Visualizing and Manipulating Biological Processes by Using HaloTag


Jeon, J.; Shen, B.; Xiong, L.; Miao, Z.; Lee, K. H.; Rao, J.; Chin, F. T. Efficient Method for Site-


Grim, J. C.; Marozas, I. A.; Anseth, K. S. Thiol-Ene and Photo-Cleavage Chemistry for Controlled


Zhang, Y.; Zhang, C.; Jiang, H.; Yang, P.; Lu, H. Fishing the PTM Proteome with Chemical


Krantz, M. J.; Palmer, R. B.; Haigney, M. C. P. Cardiovascular Complications of Opioid Use:


