Exploration and Applications of Boron Mediated Bioconjugation Chemistries

Kaicheng Li

A dissertation

submitted to the faculty of

the department of Chemistry

in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Boston College

Morrissey College of Arts and Sciences

Graduate School

May 2020

Exploration and Applications of Boron Mediated Bioconjugation Chemistries

Kaicheng Li

Advisor: professor Jianmin Gao, Ph.D.

Besides their broad applications in organic chemistry, boronic acids are also increasingly seen in a variety of biological applications. For instance, they have been used in therapeutic drugs for chemotherapy or probes for the detection of saccharides. One unique feature of boronic acids is that they are capable of forming reversible complexes with sugars or even amino acids. Importantly, they are known to have low inherent toxicity to human. In this work, we have focused on two important reactions involving boronic acids: boronate ester formation, in which boronic acids react with diols and iminoboronate formation in which boronic acids form dative bonds with neighboring amino groups. We have demonstrated the potential of these reactions in bacteria targeting or protein modification.

We envisioned that boronic acids could be used as a great warhead to be included in the development of novel antibiotics to counter antibiotic resistance of bacteria, which has emerged to be a serious clinical problem. Different from conventional antibiotics, we decided to utilize reversible covalent chemistries in the design of bacterium binding probes. Inspired by the diol-rich environment on bacterial surface, the first strategy took advantage of the reaction between boronic acid and diols to form boronate esters. We have rationally designed a linear peptide containing five copies of the 'Wulff-type' boronic acids or bicyclic amphiphilic peptides with two copies of boronic acids. We examined their capability of binding to *E. coli* cells or their bactericidal activity against *S. aureus*. Furthermore, we established a synthetic peptide library (OBTC) incorporating the 2-acetylphenyl

boronic acid (2-APBA) warhead to form iminoboronate with the target-lipid II, a precursor for the biosynthesis of peptidoglycan. Although this library failed to generate any potent peptide hits, it provided useful information regarding the development of a synthetic library as well as the screening process.

As an extension of the iminoboronate chemistry, thiazolidinoboronate (TzB) attracted our attention for its unique reaction mechanism, superior kinetics and excellent selectivity towards N-terminal cysteine residues. In this work, we have proposed an additional acyl transfer following TzB formation to turn this reaction into an irreversible conjugation. The new reaction inherits the fast kinetics and outstanding selectivity from the TzB chemistry. Excitingly, the product of TzB mediated acyl transfer survived complex conditions such as SDS-PAGE and LC/MS. This reaction was also applied to modify two recombinant proteins with N-terminal cysteine residues or to create a C5C phage library with two distinct modifications. We have further extended the substrate from cysteine to diaminopropionic acid (Dap), serine and tris base. We were delighted to observe imidazolidino boronate (IzB) formation and oxazolidino boronate (OzB) formation, which led to the design of cysteine-responsive probes or peptides that can be spontaneously cyclized in neutral aqueous conditions.

Table of Contents

List of tables

List of figures

List of schemes

Table of abbreviations

2-APBA	2-acetylphenyl boronic acid
2-FPBA	2-formylphenyl boronic acid
2-MEA	2-methoxyethylamine
4-MPAA	4-mercaptophenylacetic acid
ABC transporter	ATP-binding cassette transporter
Ala, A	Alanine
Alloc	Allyloxycarbonyl
AR	Antibiotic resistance
AzoR	Azo-reductase
B_2Pin_2	Bis(pinacolato)diboron
BIA	Biotin-iodoacetamide
BNCT	Boron neutron capture therapy
Boc	Tert-butoxycarbonyl
BSA	Bovine serum albumin
CBT	Cyanobenzothiazole
cfu	colony forming units
CuAAC	Copper catalyzed alkyne-azide cycloaddition
Cys, C	Cysteine
Dab	Diaminobutyric acid
Dabcyl-Osu	4-((4-(Dimethylamino)phenyl)azo)benzoic acid, succinimidyl ester
Dap	Diaminopropionic acid
DCM	Dichloromethane

DIPEA	N,N-Diisopropylethylamine
DI water	Deionized water
DMF	Dimethylformamide
DMSO	Dimethyl sulfoxide
DNA	Deoxyribose nucleic acid
DTT	Dithiothreitol
EC_{50}	Half maximal effective concentration
E. coli	*Eschericia coli*
ELISA	Enzyme-linked immunosorbent assay
EtOAc	Ethyl Acetate
FAM-Osu	Carboxyfluorescein-succinimydyl ester
FBS	Fetal bovine serum
FDA	Food and drug administration
Fmoc	Fluorenylmethyloxycarbonyl
GlcNAc	N-acetylglucosamine
Gly,G	Glycine
GSH	Glutathione
GSSG	Glutathione disulfide
HBS	Human blood serum
HBTU	(2-(1H-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate
HOBt	Hydroxybenzotriazole
HPLC	High performance liquid chromatography
HRMS	High resolution mass spectrometry
IA	Iodoacetamide
Ile, I	Isoleucine

IPTG	Isopropyl β-D-1-thiogalactopyranoside
iTCEP	immobilized tris(2-carboxyethyl)phosphine
IzB	Imidazolidino boronate
K_d	Equilibrium dissociation constant
LB	Lysogeny broth
LC/MS	High performance liquid chromatograph/ mass spectrometry
LDM	Linchpin directed modification
Leu	Leucine
LPS	Lipopolysaccharide
Lys, K	Lysine
MALDI	Matrix-assisted laser desorption/ionization
Mcl-1	Induced myeloid leukemia cell differentiation protein
MIC	Minimum inhibitory concentration
mRNA	Messenger ribonucleic acid
MRSA	Methicillin-resistant Staphylococcus aureus
MRSE	multidrug-resistant staphylococcus epidermidis
MruNAc	N-acetylmuramic acid
MWCO	Molecular weight cut-off
NCL	Native chemical ligation
NMR	Nuclear magnetic resonance spectroscopy
OBOC	one-bead one-compound
OBTC	one-bead two-compound
OD_{600}	optical density measured at 600 nm
OzB	Oxazolidino boronate
PBS	Phosphate-buffered saline

$Pd(dppf)Cl_2$	(1,1-Bis(diphenylphosphino)ferrocene)dichloropalladium(II)
PEG	polyethylene glycol
PG	Phosphatidylglycero
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPG	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoglycerol
PTMs	Post-translational modifications
RP-HPLC	Reversed phase high performance liquid chromatography
SAR	Structure activity relationship
S. aureus	*Staphylococcus aureus*
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
Ser, S	Serine
SPPS	Solid phase peptide synthesis
SrtA	Sortase A
TCEP	tris(2-carboxyethyl)phosphine
TEA	Triethylamine
TEV	*Tobacco etch virus*
TFA	Trifluoroacetic acid
TIPS	Triisopropylsilane
TOF	Time of flight
tRNA	Transfer ribonucleic acid
Trt	Trityl
Trx	Thioredoxin
TzB	Thiazolidino boronate
UAA	unnatural amino acid
UV-vis	Ultraviolet-visible spectroscopy

Acknowledgement

Firstly, I would like to thank my supervisor Dr. Jianmin Gao. I really appreciate the support I received from him during these years. Although he would often encourage me to think and trouble shoot independently, whenever I was lost on how to proceed on a project, he always provided a different way of thinking and gave me lots of useful advice. He taught me how to think broad and critically, and allowed me to feel confident that I could be an independent scientist after almost 6 years of training under his supervision. I would also like to thank Professors Abhishek Chatterjee, Eranthie Weerapana and Jia Niu for their kind help during my PhD studies. I am grateful for their suggestions on my research, qualifying exam, proposal and thesis defense.

Next, I would like to thank my co-workers, the ones that I spent most of my time with during my PhD period. Firstly, my mentor Dr. Yue Zhao not only mentored me with my experimental skills, but also provided help in my life outside of the lab. I am really thankful that I can still turn to her even after she graduated. Next, I would like to thank Dr. Breanna Zerfas. We collaborated on some projects and she inspired me by her way of thinking and knowledge of science. She also provided me with great help with my oral and written English. I would like to thank Michael Kelly, Samantha Cambray and Kelly McCarthy. They are really friendly and taught me a lot about American culture. They gave useful opinions on my research and helped me with my presentations, resume and dissertation. Next, Wenjian Wang and Chelsea Burke were good listeners to my complaints about research and personal life. They have also been reliable whenever I turn to them for help. I also want to say thank you

to Aaron Maurais in professor Weerapana's lab. He helped us with LC/MS/MS sequencing which played a vitally important role in one of my projects. Natalia Chmielewska and Tuo Ji gave me the opportunity to become a mentor and these clever mentees helped me a lot with my experiments. I would also say thank you to Dr. Anupam Bandyopadhyay, Dr. Azade Hosseini, Lauren Blair and all other colleagues in the ChemBio groups. It has been a great experience to work in the department because of these nice people. I would like to thank all the faculties in the chemistry department who made my research a lot easier.

In graduate school, the valuable friendship made my life enjoyable at BC. These wonderful friends include Pengcheng Ye, Qi Dong, Chaofan Xu and Julia Liu, to name a few. They cheered me up when I felt down and helped me get through the rough days.

Lastly, I would like to say thank you to my family members. Although we were far away from each other, they provide me with the strongest emotional support. They would respect every decision I made and provide any possible help they could offer. My parents and my sister are always saying they are proud of me, but I feel lucky to have them as my family. Without them, I do not think I can accomplish anything.

Financial support of this work is partially provided by the National Science Foundation and the National Institutes of Health through grants CHEM1904874 and GM102735, respectively.

Chapter 1 Introduction

1.1 Boronic acid

The 5th atom, boron has always been special for chemists. The 2010 Nobel prize for chemistry which was awarded to Akira Suzuki and co-workers, is a good example highlighting its importance in chemistry. The famous reaction, Suzuki coupling, has found broad applications in industry as well as academic labs.[1,2] Although boron can form covalent bonds with various atoms such as N and F, it is used in the form of boronic acid in most biological applications due to its biocompatibility, low toxicity and stability in aqueous solution. When boron exists as neutral boronic acid or boronate ester, its empty p orbital provides great accommodation for electron-rich atoms such as nitrogen and oxygen.

Boronic acid containing compounds have gained a lot of attention in the pharmaceutical industry. As a matter of fact, Bortezomib was initially approved by the FDA in 2003 for the treatment of multiple myeloma. Later in 2006, it was approved for the treatment of mantle cell lymphoma. The boronic acid in Bortezomib plays a critical role in binding to the catalytic site of the 26S proteasome.[3] Its binding mode was later revealed by a co-crystal structure of yeast 20s proteasome in which the boronic acid contributes to the binding through both hydrogen bonds as well as a reversible covalent bond (**Figure 1.1.1**).[4] The approval of Bortezomib has opened the door for researchers to develop boronic acid containing compounds for therapeutic purposes, mostly as enzyme inhibitors.[4–6] Recently, tavaborole[7] and crisaborole[8] were approved by the FDA for the treatment of fungi infection and eczema respectively. Boronic acid can also serve as boron neutron capture therapy (BNCT) agents.[9] When irradiated with neutrons, boron-10 transforms to lithium and emits α-particles which can then kill cancer cells. Although related compounds have yet to be approved as

therapeutic medicines, this interesting warhead certainly occupies an important position in pharmaceutical industry.

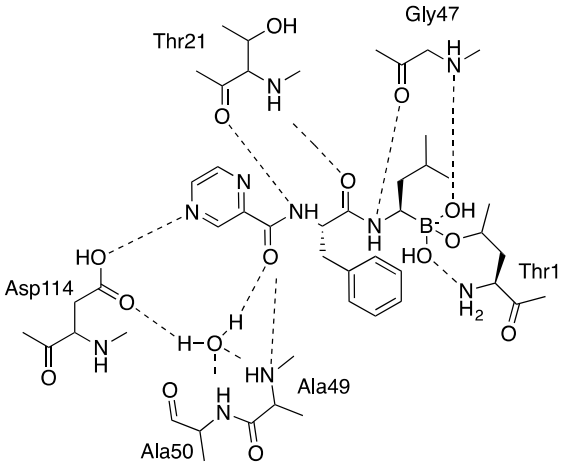


Figure 1.1. Bortezomib binding to yeast 20s proteasome

1.2 Boronate ester formation

Boronic acid is known to react with diol to form boronate esters. The reaction of boronic acid with diol in aqueous solution is reversible and less favored than non-aqueous solution. In fact, boronate esters in aqueous solution exist in quick equilibrium with boronic acids. This unique feature makes it an attractive functional group to be incorporated in biological applications. For instance, boronic acid has been employed to develop different sensors for various saccharides.[10–13] One representative example is the design of a fluorescent molecular tweezer by the Tony James' group.[12] In their design, when a saccharide molecule forms a 1:1 complex with the sensor, it forces the separation between the pyrene molecules, which leads to a quenching of the excimer emission. Moreover, glycoproteins represent more than 50% of the total proteins in mammalian cells, and various Nano particles or polymers highlighting boronic acid headgroups have been developed to detect or enrich these important proteins.[14,15] Similarly, Ronald Raines and co-workers have utilized the

interaction of boronic acids with cell surface glycans to develop cell-penetrating molecules.[16,17] This interesting discovery was latter applied in drug delivery.[18] Boronic acid has also been utilized in designing hydrogels that have unique properties such as pH sensitivity, glucose responsiveness or self-healing capabilities.[19,20]

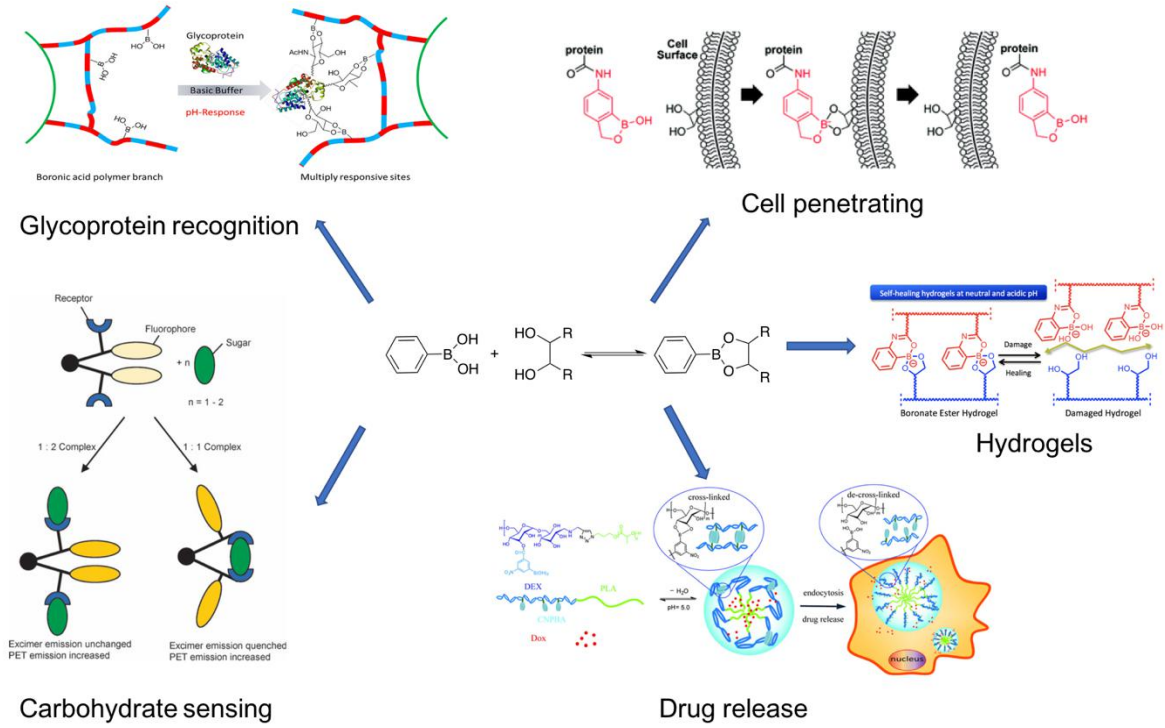


Figure 1.2. Examples of the applications based on boronic acid-diol reaction. Images were taken from reference 12, 14, 16, 18 and 19.

The reaction between diols and boronic acid and related derivatives has been studied quite extensively. It has been shown that this reaction is related to the pKa of boronic acid and diol, steric hinderance, as well as stabilization of the product.[21] It is believed that the association process is more favored in basic conditions presumably due to the tetrahedral structure of boronic acid. Compared to its planar structure in the neutral form, sp3 hybridized boronic acid releases the angular strain of the boronate ester, which drives the equilibrium towards product formation. When a boronic acid

derivative has a lower pKa, a larger amount of it exists as a boronate anion under the same pH, which allows it to bind better to diols in general.

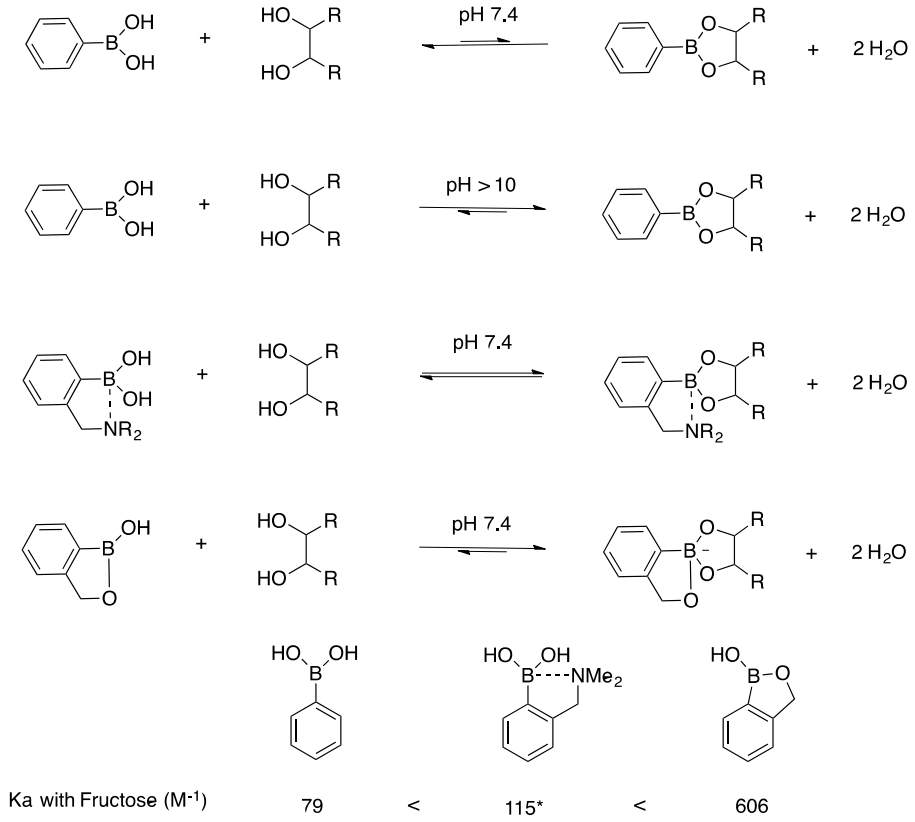


Figure 1.3. Boronic acid analogues binding to fructose determined by NMR in D_2O. Benzylboroxoles and 'Wul-ff-type' aryboronic acids have higher binding affinity with diols. * means 'Wulff-type' was measured in a mixture of CD_3OD/D_2O co-solvent to increase its solubility

'Wulff-type' arylboronic acids are unique boronic acid derivatives. They have a higher affinity to diols compared to phenylboronic acid by enabling an N-B coordination at a lower pH. The neighboring amino group pre-positions the boronic acid in a tetrahedral structure to promote binding to diols. However, these arylboronic acids generally do not have good solubility in water due to their zwitterionic property.

This issue can be easily mitigated by adding hydrophilic functional groups when designing peptides or small molecule derivatives.

Recently, the Hall group has extensively studied benzoboroxoles as superior carbohydrate binders.[22,23] When dissolved in aqueous solution, benzoboroxoles were shown to exist in equilibrium with ortho-hydroxymethyl phenylboronic acid which is analogous to the 'Wulff-type' derivatives. The boronic acid in benzoboroxoles is believed to have a much lower pKa than phenylboronic acid. Better solubility in water and a higher association constant with diols make it a good warhead for targeting diol-containing structures. It is worth mentioning that both tavaborole[7] and crisaborole[8] are benzoboroxole derivatives, although their targets are not necessarily diols.

1.3 Iminoboronate formation

Aldehydes or ketones are known to react with amines to form imines. However, this reaction finds more use in organic solvents rather than aqueous condition due to the high dissociation constant (K_d). When a neighboring boronic acid is available to stabilize the imine product with an N-B coordination, the K_d of the conjugation can be lowered to the millimolar range. This reaction is named as iminoboronate. Since its initial report by James group, iminoboroante has been a star in many biological applications.[24] For example, Mikail Barboiu's lab has been utilizing iminoboronate to make functional G-Quartets;[25] the Gois group has demonstrated reversible protein modification on lysine residues with 2-Formylphenyl boronic acid (2-FPBA);[26] our lab has employed the 2-acetylphenylboronic acid (2-APBA) warhead for bacteria targeting and imaging.[27,28] Recently, 2-APBA has also

been used to increase the binding affinity of a Mcl-1 inhibitor, demonstrating its therapeutic potential.[29]

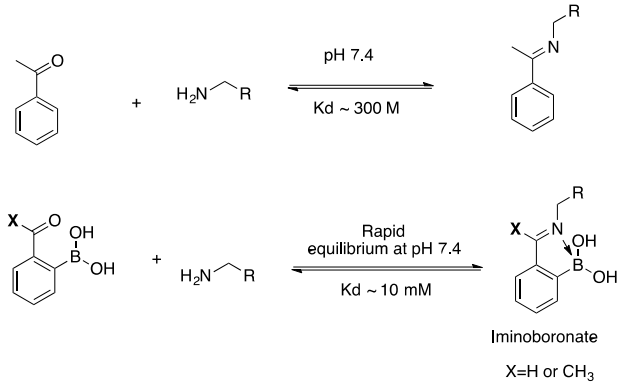


Figure 1.4. Comparison of iminoboronate with ordinary imine formation in neutral aqueous buffer

Dr. Anupam Bandyopadhyay in our lab has extensively studied this reaction. In fact, he expanded the scope of iminoboronate and discovered a reaction named thiazolidinoboronate (TzB). Our lab and the Gois' lab have recently reported, almost at the time, about this fast and chemo-selective reaction between 2-FPBA and cysteine.[30,31] Although N-terminal cysteines, or aminothiol structures is known to react with aldehydes to form thiazolidines, this reaction is generally slow and requires slight acidic condition. TzB formation on the other hand, has a rate constant of 5000 $M^{-1}s^{-1}$, which is one of the fastest bioconjugation chemistries that happens in neutral aqueous conditions.[32,33] The chemo-selectivity was demonstrated by carrying out this reaction in the presence of a range of molecules that are commonly found in biology such as fructose, serine, lysine and glutathione. Importantly, TzB formation is highly selective towards N-terminal cysteines with no observed side reactions with internal cysteines. This is expected as the formation of TzB is initiated by iminoboronate formation between the aldehyde group from 2-FPBA and the α-amine group from N-terminal cysteine. The thiol group from cysteine side chain then attacks the imine

structure to give a five membered thiazolidine ring. In native chemical ligation (NCL) or CBT condensation, however, the reactions start with thiol-exchange or thioimidate formation, resulting in potential side reactions with internal cysteines.[34,35]

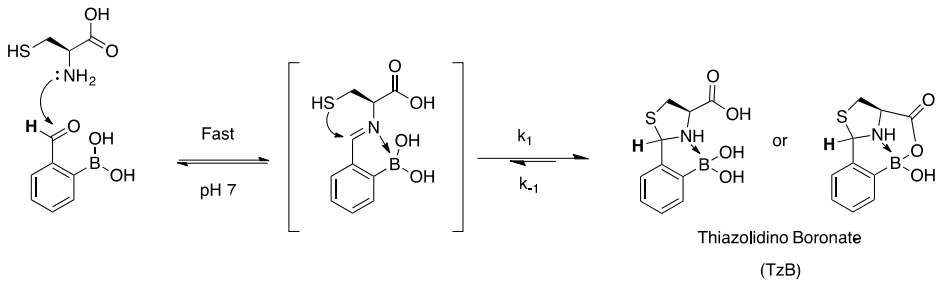


Figure 1.5. Mechanism of thiazolidino boronate formation

In this contribution, we describe the studies that we have carried out about boronic acids. Specifically, we explored the biological applications of boronic acids in: 1) Bacteria targeting through boronate ester formation or iminoboronate formation (Chapter 2); 2) Site-specific modifications on proteins through thiazolidinoboronate mediated acyl transfer (Chapter 3).

1.4 Reference

(1) Miyaura, N.; Suzuki, A. Stereoselective Synthesis of Arylated (. *J. Chem. Soc. Chem. Commun.* **1979**, No. 866, 866–867.

(2) Miyaura, N.; Yamada, K.; Suzuki, A. Our Continuous Discovered. *Tetrahedron Lett.* **1979**, *20* (36), 3437–3440.

(3) Bonvini, P.; Zorzi, E.; Basso, G.; Rosolen, A. Bortezomib-Mediated 26S Proteasome Inhibition Causes Cell-Cycle Arrest and Induces Apoptosis in CD-30+ Anaplastic Large Cell Lymphoma [16]. *Leukemia* **2007**, *21* (4), 838–842. https://doi.org/10.1038/sj.leu.2404528.

(4) Groll, M.; Berkers, C. R.; Ploegh, H. L.; Ovaa, H. Crystal Structure of the Boronic Acid-Based Proteasome Inhibitor Bortezomib in Complex with the Yeast 20S Proteasome. *Structure* **2006**, *14* (3), 451–456. https://doi.org/10.1016/j.str.2005.11.019.

(5) Yang, W.; Gao, X.; Wang, B. Boronic Acid Compounds as Potential Pharmaceutical Agents. *Med. Res. Rev.* **2003**, *23* (3), 346–368. https://doi.org/10.1002/med.10043.

(6) Touchet, S.; Carreaux, F.; Carboni, B.; Bouillon, A.; Boucher, J. L. Aminoboronic Acids and Esters: From Synthetic Challenges to the Discovery of Unique Classes of Enzyme Inhibitors. *Chem. Soc. Rev.* **2011**, *40* (7), 3895–3914. https://doi.org/10.1039/c0cs00154f.

(7) Baker, S. J.; Zhang, Y. K.; Akama, T.; Lau, A.; Zhou, H.; Hernandez, V.; Mao, W.; Alley, M. R. K.; Sanders, V.; Plattner, J. J. Discovery of a New Boron-Containing Antifungal Agent, 5-Fluoro-1,3-Dihydro- 1-Hydroxy-2,1-Benzoxaborole (AN2690), for the Potential Treatment of Onychomycosis. *J.*

Med. Chem. **2006**, *49* (15), 4447–4450. https://doi.org/10.1021/jm0603724.

(8) Akama, T.; Baker, S. J.; Zhang, Y. K.; Hernandez, V.; Zhou, H.; Sanders, V.; Freund, Y.; Kimura, R.; Maples, K. R.; Plattner, J. J. Discovery and Structure-Activity Study of a Novel Benzoxaborole Anti-Inflammatory Agent (AN2728) for the Potential Topical Treatment of Psoriasis and Atopic Dermatitis. *Bioorganic Med. Chem. Lett.* **2009**, *19* (8), 2129–2132. https://doi.org/10.1016/j.bmcl.2009.03.007.

(9) Soloway, A. H.; Tjarks, W.; Barnum, B. A.; Rong, F. G.; Barth, R. F.; Codogni, I. M.; Wilson, J. G. The Chemistry of Neutron Capture Therapy. *Chem. Rev.* **1998**, *98* (4), 1515–1562. https://doi.org/10.1021/cr941195u.

(10) Sandanayake, K. R. A. S.; Shinkai, S. Novel Molecular Sensors for Saccharides Based on the Interaction of Boronic Acid and Amines: Saccharide Sensing in Neutral Water. *J. Chem. Soc. Chem. Commun.* **1994**, No. 9, 1083–1084. https://doi.org/10.1039/C39940001083.

(11) Wu, X.; Li, Z.; Chen, X. X.; Fossey, J. S.; James, T. D.; Jiang, Y. B. Selective Sensing of Saccharides Using Simple Boronic Acids and Their Aggregates. *Chem. Soc. Rev.* **2013**, *42* (20), 8032–8048. https://doi.org/10.1039/c3cs60148j.

(12) Phillips, M. D.; Fyles, T. M.; Barwell, N. P.; James, T. D. Carbohydrate Sensing Using a Fluorescent Molecular Tweezer. *Chem. Commun.* **2009**, No. 43, 6557–6559. https://doi.org/10.1039/b909230g.

(13) Sun, X.; James, T. D. Glucose Sensing in Supramolecular Chemistry. *Chem. Rev.* **2015**, *115* (15), 8001–8037. https://doi.org/10.1021/cr500562m.

(14) Liu, J.; Yang, K.; Shao, W.; Qu, Y.; Li, S.; Wu, Q.; Zhang, L.; Zhang, Y. Boronic Acid-Functionalized Particles with Flexible Three-Dimensional Polymer Branch for Highly Specific Recognition of Glycoproteins. *ACS Appl.*

Mater. Interfaces **2016**, *8* (15), 9552–9556. https://doi.org/10.1021/acsami.6b01829.

(15) Wang, H.; Bie, Z.; Lü, C.; Liu, Z. Magnetic Nanoparticles with Dendrimer-Assisted Boronate Avidity for the Selective Enrichment of Trace Glycoproteins. *Chem. Sci.* **2013**, *4* (11), 4298–4303. https://doi.org/10.1039/c3sc51623g.

(16) Ellis, G. A.; Palte, M. J.; Raines, R. T. Boronate-Mediated Biologic Delivery. *J. Am. Chem. Soc.* **2012**, *134* (8), 3631–3634. https://doi.org/10.1021/ja210719s.

(17) Andersen, K. A.; Smith, T. P.; Lomax, J. E.; Raines, R. T. Boronic Acid for the Traceless Delivery of Proteins into Cells. *ACS Chem. Biol.* **2016**, *11* (2), 319–323. https://doi.org/10.1021/acschembio.5b00966.

(18) Zhao, Z.; Yao, X.; Zhang, Z.; Chen, L.; He, C.; Chen, X. Boronic Acid Shell-Crosslinked Dextran-b-PLA Micelles for Acid-Responsive Drug Delivery. *Macromol. Biosci.* **2014**, *14* (11), 1609–1618. https://doi.org/10.1002/mabi.201400251.

(19) Deng, C. C.; Brooks, W. L. A.; Abboud, K. A.; Sumerlin, B. S. Boronic Acid-Based Hydrogels Undergo Self-Healing at Neutral and Acidic PH. *ACS Macro Lett.* **2015**, *4* (2), 220–224. https://doi.org/10.1021/acsmacrolett.5b00018.

(20) Dong, Y.; Wang, W.; Veiseh, O.; Appel, E. A.; Xue, K.; Webber, M. J.; Tang, B. C.; Yang, X. W.; Weir, G. C.; Langer, R.; et al. Injectable and Glucose-Responsive Hydrogels Based on Boronic Acid-Glucose Complexation. *Langmuir* **2016**, *32* (34), 8743–8747. https://doi.org/10.1021/acs.langmuir.5b04755.

(21) Brooks, W. L. A.; Deng, C. C.; Sumerlin, B. S. Structure-Reactivity Relationships in Boronic Acid-Diol Complexation. *ACS Omega* **2018**, *3* (12), 17863–17870. https://doi.org/10.1021/acsomega.8b02999.

(22) Bérubé, M.; Dowlut, M.; Hall, D. G. Benzoboroxoles as Efficient Glycopyranoside-Binding Agents in Physiological Conditions: Structure and Selectivity of Complex Formation. *J. Org. Chem.* **2008**, *73* (17), 6471–6479. https://doi.org/10.1021/jo800788s.

(23) Dowlut, M.; Hall, D. G. An Improved Class of Sugar-Binding Boronic Acids, Soluble and Capable of Complexing Glycosides in Neutral Water. *J. Am. Chem. Soc.* **2006**, *128* (13), 4226–4227. https://doi.org/10.1021/ja057798c.

(24) Pérez-Fuertes, Y.; Kelly, A. M.; Johnson, A. L.; Arimori, S.; Bull, S. D.; James, T. D. Simple Protocol for NMR Analysis of the Enantiomeric Purity of Primary Amines. *Org. Lett.* **2006**, *8* (4), 609–612. https://doi.org/10.1021/ol052776g.

(25) Arnal-Hérault, C.; Pasc, A.; Michau, M.; Cot, D.; Petit, E.; Barboiu, M. Functional G-Quartet Macroscopic Membrane Films. *Angew. Chemie - Int. Ed.* **2007**, *46* (44), 8409–8413. https://doi.org/10.1002/anie.200702605.

(26) Cal, P. M. S. D.; Vicente, J. B.; Pires, E.; Coelho, A. V.; Veiros, L. F.; Cordeiro, C.; Gois, P. M. P. Iminoboronates: A New Strategy for Reversible Protein Modification. *J. Am. Chem. Soc.* **2012**, *134* (24), 10299–10305. https://doi.org/10.1021/ja303436y.

(27) Bandyopadhyay, A.; McCarthy, K. A.; Kelly, M. A.; Gao, J. Targeting Bacteria via Iminoboronate Chemistry of Amine-Presenting Lipids. *Nat. Commun.* **2015**, *6*, 1–9. https://doi.org/10.1038/ncomms7561.

(28) McCarthy, K. A.; Kelly, M. A.; Li, K.; Cambray, S.; Hosseini, A. S.; Van Opijnen, T.; Gao, J. Phage Display of Dynamic Covalent Binding Motifs Enables Facile Development of Targeted Antibiotics. *J. Am. Chem. Soc.* **2018**, *140* (19), 6137–6145. https://doi.org/10.1021/jacs.8b02461.

(29) Akçay, G.; Belmonte, M. A.; Aquila, B.; Chuaqui, C.; Hird, A. W.; Lamb, M. L.; Rawlins, P. B.; Su, N.; Tentarelli, S.; Grimster, N. P.; et al. Inhibition of Mcl-1 through Covalent Modification of a Noncatalytic Lysine Side Chain. *Nat. Chem. Biol.* **2016**, *12* (11), 931–936. https://doi.org/10.1038/nchembio.2174.

(30) Bandyopadhyay, A.; Cambray, S.; Gao, J. Fast and Selective Labeling of N-Terminal Cysteines at Neutral PH via Thiazolidino Boronate Formation. *Chem. Sci.* **2016**, *7* (7), 4589–4593. https://doi.org/10.1039/C6SC00172F.

(31) Faustino, H.; Silva, M. J. S. A.; Veiros, L. F.; Bernardes, G. J. L.; Gois, P. M. P. Iminoboronates Are Efficient Intermediates for Selective, Rapid and Reversible N-Terminal Cysteine Functionalisation. *Chem. Sci.* **2016**, *7* (8), 5052–5058. https://doi.org/10.1039/c6sc01520d.

(32) Yu, Z.; Pan, Y.; Wang, Z.; Wang, J.; Lin, Q. Genetically Encoded Cyclopropene Directs Rapid, Photoclick-Chemistry- Mediated Protein Labeling in Mammalian Cells. *Angew. Chemie - Int. Ed.* **2012**, *51* (42), 10600–10604. https://doi.org/10.1002/anie.201205352.

(33) Devaraj, N. K.; Weissleder, R.; Hilderbrand, S. A. Tetrazine-Based Cycloadditions: Application to Pretargeted Live Cell Imaging. *Bioconjug. Chem.* **2008**, *19* (12), 2297–2299. https://doi.org/10.1021/bc8004446.

(34) Wang, W.; Gao, J. N, S-Double Labeling of N-Terminal Cysteines via an Alternative Conjugation Pathway with 2-Cyanobenzothiazole. *J. Org. Chem.* **2020**. https://doi.org/10.1021/acs.joc.9b02959.

(35) Ramil, C. P.; An, P.; Yu, Z.; Lin, Q. Sequence-Specific 2-Cyanobenzothiazole Ligation. *J. Am. Chem. Soc.* **2016**, *138* (17), 5499–5502. https://doi.org/10.1021/jacs.6b00982.

Chapter 2 Targeting bacteria with boronic acid containing peptides through reversible covalent interactions

2.1 Introduction

Since the discovery of penicillin, antibiotics have been a standard answer to treat bacterial infections. Many other antibiotics were discovered or rationally designed post-penicillin discovery which greatly reduced the death rates caused by bacterial infection. Unfortunately, antibiotic resistance (AR) has become an increasing clinical concern due to its natural occurrence coupled with the abuse of antimicrobial agents.[1,2] Limiting the use of antibiotics helps to slow down the emergence of antibiotic resistance.[3] However, the development of new antibiotics is always in urgent need to counter the existing and prepare for the emergence of new AR strains.

Although the cell wall and membrane serve as shelters to protect the bacteria from its surrounding environment, they are also the most straight forward and accessible targets for antibiotics. Our lab has been trying to understand the roles of lipids and lipid modification in bacterial growth. Research in this area will not only help us to understand the functions of these important components on bacteria cell surface, but also provide guidance for us to design chemical probes or novel antibiotics for AR bacteria. In this chapter, I will talk about the effort that we have devoted to develop novel antibiotics that target lipid components of bacteria through two reversible reactions: boronate ester formation and iminoboronate formation.

References

(1) Neu, H. C. The Crisis in Antibiotic Resistance. *Science (80-.).* **1992**, *257* (5073), 1064 LP – 1073. https://doi.org/10.1126/science.257.5073.1064.

(2) Davies, J. Origins and Evolution of Antibiotic Resistance. *Microbiologia* **1996**, *12* (1), 9–16. https://doi.org/10.1128/mmbr.00016-10.

(3) Gould, I. M. A Review of the Role of Antibiotic Policies in the Control of Antibiotic Resistance. *J. Antimicrob. Chemother.* **1999**, *43* (4), 459–465. https://doi.org/10.1093/jac/43.4.459.

2.2 Targeting the diol structures on bacteria surface through boronate ester formation

2.2.1 Introduction

2.2.1.1 Diol structures on bacteria

The surface of bacteria presents abundant components containing the diol structure. For example, lipopolysaccharides (LPS) or endotoxins, are the major components of the outer membrane of gram-negative bacteria. They are of great importance to the structural integrity and provide protection for some antibiotics that are lethal to gram-positive bacteria. LPS is composed of three parts: Lipid A, Core oligosaccharide and O-antigen. The structure of O-antigen varies from strain to strain, but it is composed of repeating subunits of oligosaccharide that contains abundant diol structures.[1]

Additional diol-presenting structures from the bacterial surface include lipid II and phosphatidylglycerol. Lipid II is a precursor for peptidoglycan in bacteria and disruption for the biosynthesis of peptidoglycan will lead to cell death.[2–4] It is constructed on the inner side of cytoplasmic membrane and transported through the membrane by a flippase. The N-acetylglucosamine (GlcNAc) in lipid II before polymerization also presents a diol structure that can potentially be targeted, although it is not easily accessible for gram-negative bacteria due to the presence of LPS. Moreover, Phosphatidylglycerol (PG) is a key component in bacteria membrane. It exists in both gram-negative and gram-positive bacteria and contains a diol structure originated from glycerol. These diol-containing structures can be considered as possible targets when designing novel probes or antibiotics.

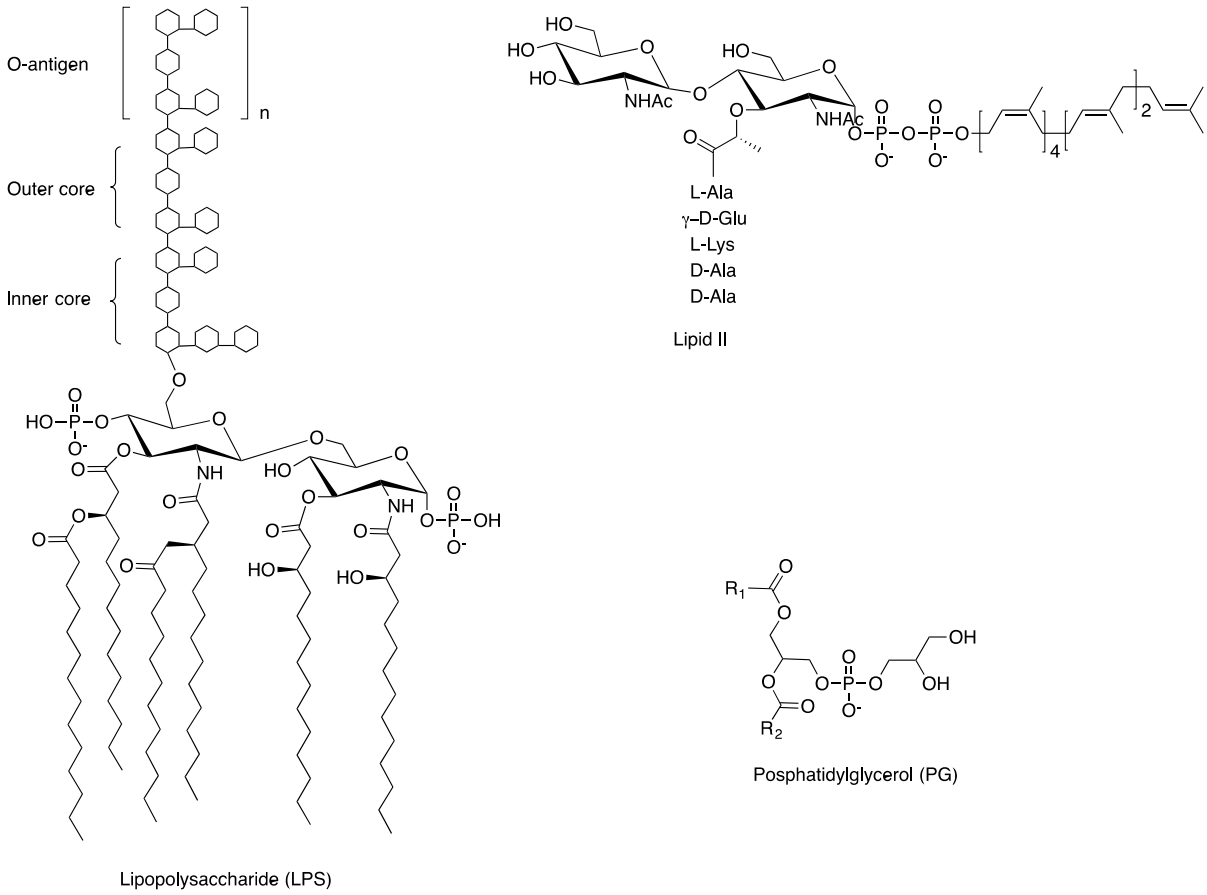


Figure 2.2.1. Structures of diol-containing components on bacteria surface including LPS, PG and lipid II

2.2.1.2 Targeting bacteria with boronic acid containing peptides

The majority of antibiotics bind to their targets through non-covalent driving forces, including hydrogen bonding or electrostatic interaction. To develop a series of new antibiotics, we hypothesized that we could take advantage of reversible covalent interactions. The diol structures mentioned previously are highly conservative to bacteria, thus antibiotics that specifically bind to the diols may be potent drug candidates due to the difficulty of resistance development from these pathogens. As is discussed earlier, boronic acids reacts reversibly with diols, therefore indicating that it could be a good warhead to be included in the design of new antibiotics. Although

there are multiple antibiotics that target the membrane of bacteria, the concept of utilizing the diol structures through covalent interaction is without precedent.

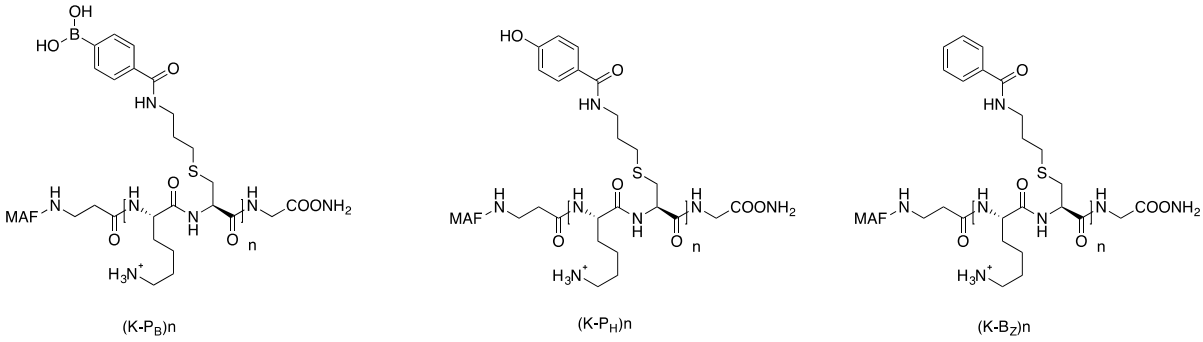
Peptide antibiotics, such as colistin and daptomycin, are quite common in nature. The structure of a peptide can be easily manipulated to obtain optimal properties such as better solubility or higher binding affinity to its target. Importantly, we can incorporate multiple functional groups which may not exist naturally into peptides through standard solid phase peptide synthesis (SPPS). To this end, we sought to design novel antibiotics using peptides as the scaffold and boronic acids as the warhead.

2.2.2 Boronic acid containing peptide design

Boronic acid alone only has mediocre binding affinity towards diols, which means a peptide with only one copy of boronic acid may not provide strong enough binding. Thus, we decided to incorporate multiple copies of this warheads to achieve better binding affinity. Lysine was also included to increase the solubility and binding potency by interacting with the negative charges on bacteria surface through electrostatic interactions or hydrogen bonding. β-alanine was added as a spacer to separate the peptide sequence from the fluorophore. The designed peptide has the following sequence: (K-X)n = *FAM*-βA-(K-X)n-G. X stands for amino acid that presents the boronic acid warhead.

There are two possible ways to incorporate boronic acid into peptides. Boronic acid derivatives can be coupled on resin or in solution after the backbone of the peptide is assembled through SPPS; An alternative way is to synthesize an unnatural amino acid (UAA) bearing the boronic acid warhead and directly incorporate it into the sequence during peptide assembly. The first strategy turned out be difficult, as

was explored by Dr. Yue Zhao in our lab. When multiple copies of boronic acid were conjugated on resin, incomplete reaction resulted in a mixture of peptides which were difficult to separate out and purify. We then decided to incorporate the warhead in the form of an UAA, which turned out to be a more efficient synthetic strategy. We synthesized the desired peptide named (K-P_B)5 as well as some control peptides including (K-P_H)5 and (K-B_Z)5.



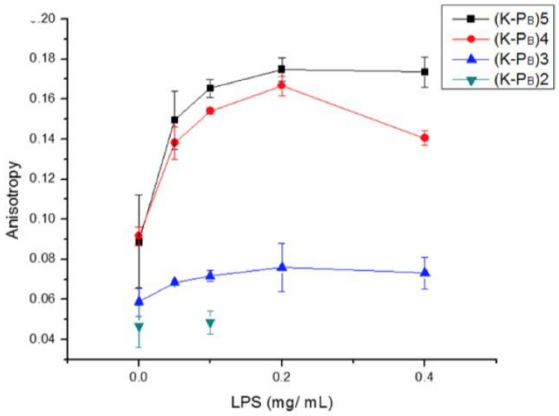


Figure 2.2.2. Anisotropy results for (K-P_B)5 and its control (K-P_B)4, (K-P_B)3 and (K-P_B)2. 0.5μM peptide was incubated with LPS micelle in PBS for 90 min before reading. Excitation wavelength: 495 nm; emission wavelength: 532 nm. The experiment was carried out in triplicates (data collected by Dr. Yue Zhao)

We conducted anisotropy experiments on (K-P_B)n peptides and some of the results were included in **Figure 2.2.2**. (K-P_B)5 and (K-P_B)4 show significant

anisotropy change as the concentration of LPS micelle increases. (K-P_B)5 in this experiment has a slight advantage over (K-P_B)4. This result agrees with our hypothesis that peptides exhibiting multivalence will have an advantage binding to the desired target. However, (K-P_B)5 in the following experiments failed to show stronger binding affinity than (K-P_H)5 or (K-B_Z)5 against LPS micelles, POPC/POPG large unilamellar vesicles or *E. coli* cells (data not shown). These results therefor call into question the function of the boronic acid warhead in (K-P_B)5.

The low affinity of phenylboronic acid binding to diols may have contributed to the failure of (K-P_B)5. As is discussed earlier, the association constant of boronic acid is closely dictated by its structure. In order to achieve a higher binding affinity, we decided to incorporate a better boronic acid warhead into the peptides. The 'Wulff-type' arylboronic acid was finally chosen considering the ease of synthesis. We then proposed a new structure for our peptide named (K-T_B)5 and its control (K-T_Z)5.

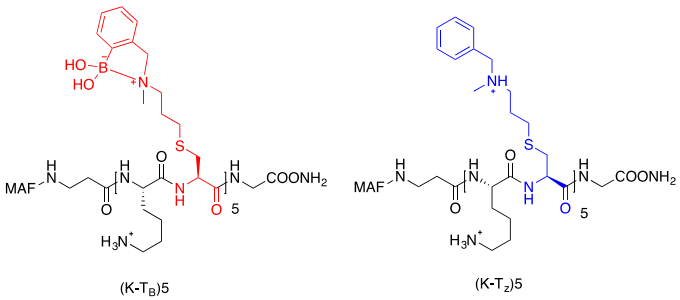
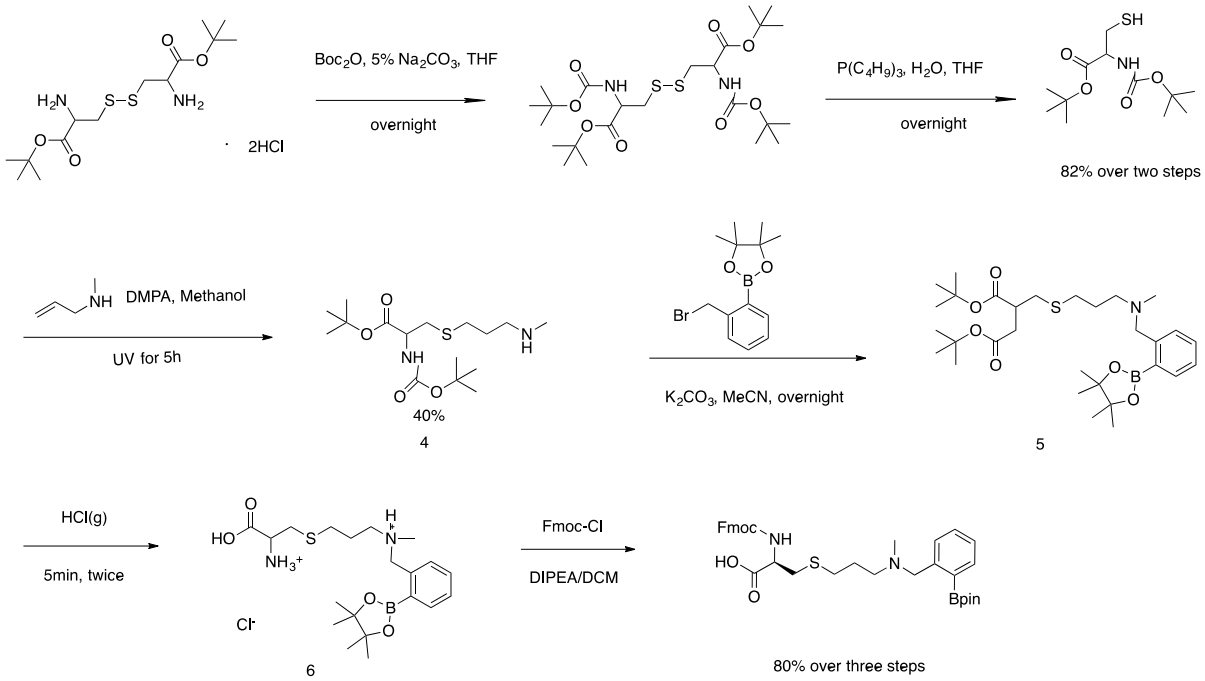


Figure 2.2.3. Structure of (K-T_B)5 containing 'Wulff-type' boronic acid warheads and its control (K-T_Z)5

2.2.3 Synthesis and optimization of the unnatural amino acid



Scheme 2.2.1. Original synthetic route for T_B, designed by Dr. Yue Zhao

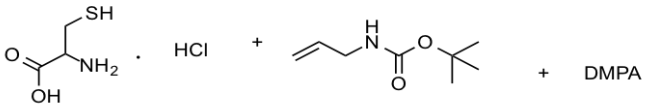
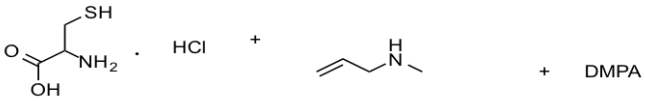
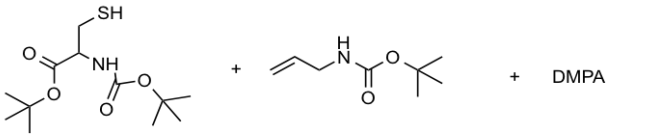
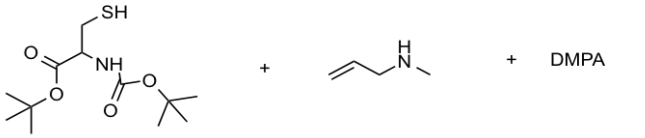
Scheme 2.2.1 showed the original synthetic route for T_B. The reaction starts with Boc protection of L-cystine, followed by reduction of the disulfide bond. The thiol group was then alkylated by a thiol-ene click reaction introducing a secondary amine structure to its side chain. The product was then reacted with 2-Bromobenzyl boronate ester through a SN2 reaction. The synthesis was finalized by removal of the protecting groups and the installation of Fmoc on α-amine. This synthetic route was proven to be a feasible strategy. However, the thio-ene step suffered from low yield and the highest possible yield obtained was around 40%. Importantly, the product was too hydrophilic and the presence of too many side products added to the difficulty of purification. To optimize the thiol-ene click reaction, several comparative reactions was carried out.

Table 2.2.1. Troubleshoot the thiol-ene reaction by tuning the stoichiometry, with yield determined by crude ^{1}H NMR

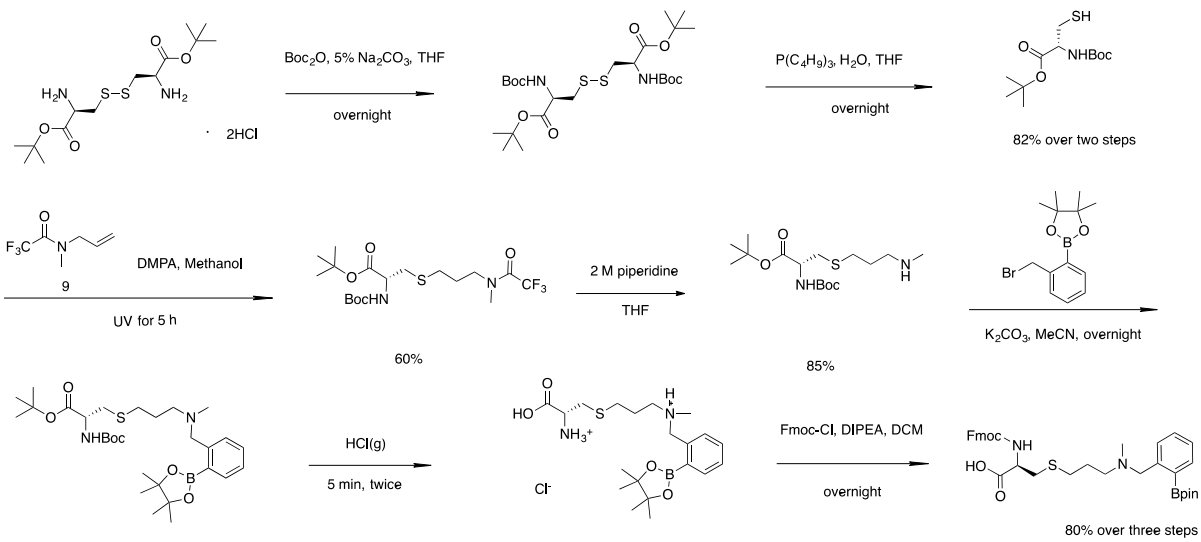
Compound 3:		N-allylmethylamine:		DMPA	Yield
1	:	1	:	0.1	41%
1	:	2	:	0.1	45%
1	:	1	:	0.2	40%
1	:	2	:	0.2	45%

The results in **Table 2.2.1** indicated that a change in stoichiometry was not helpful. Given that the main byproduct was cystine, we hypothesized that oxidation of cysteine happens faster than the click reaction, preventing the thiol from reacting with alkene. We then carried out the same reaction on different substrates.

Table 2.2.2. Troubleshoot the thiol-ene reaction with different substrates

Ingredients	ratio	conversion
	1 eq: 1eq: 0.2 eq	100%
	1 eq: 1eq: 0.2 eq	precipitate
	1 eq: 1eq: 0.2 eq	80%
	1 eq: 1eq: 0.2 eq	40%

A different substrate, Boc protected Allylamine showed significantly higher conversion compared to the substrate with a free secondary amine. Unfortunately, it is not feasible to protect the amine with Boc because the deprotection requires acidic condition, which also removes the tert-butyl group on carboxylic acid and Boc protection on the α-amine. Alloc is also not a good choice due to the presence of an alkene structure. We predicted that a trifluoroacetyl group would be an optimal protecting group due to its orthogonality and compatibility with the click reaction. To our delight, this substrate gave us an improved yield (60%). Importantly, the product became less hydrophilic and more easily purified.



Scheme 2.2.2. New synthetic route for T_B with improved yield in the thiol-ene click reaction

2.2.4 (K-T_B)5-*E. coli* binding experiment

2.2.4.1 Titration in PBS

With the unnatural amino acid at hand, we were able to synthesize the desired peptide and carry out binding experiments towards *E. coli* with flow cytometry. Experimental details are included at the end of this chapter. Consistent with our previous data (collected by Dr. Yue, data not shown), (K-T_B)5 showed significantly

higher binding potency compared to the control peptide (K-P_H)5. The estimated EC_{50} was found to be sub-micromolar. However, it did not show any superior binding potency in comparison to (K-T_Z)5 or (K-T_B)2-(K-T_Z)3. This result implies that boronic acid in (K-T_B)5 may not be the main driving force for bacteria binding. It is also possible that the unnatural amino acid T_Z enables binding through a different mechanism.

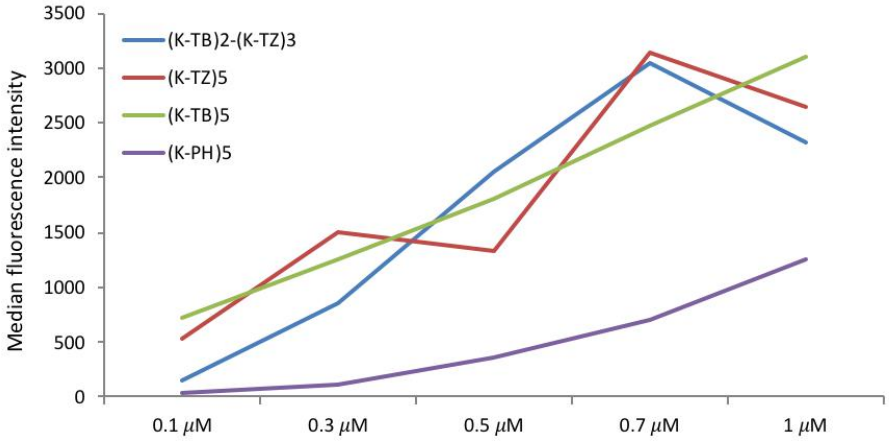


Figure 2.2.4. Flow cytometry titration comparison for (K-T_B)5 and it control peptides against *E. coli*. The peptides were dissolved in DMF as 100 μM stock. Cell density: 2 million cfu/mL

2.2.4.2 Fructose inhibition

In our original design, boronic acids in (K-T_B)5 were intended to serve as warheads to provide stronger binding affinity through covalent interaction with the diol structures on the bacterial surface. However, the results in **Figure 2.2.4** was contradictory to our prediction. In order to confirm the function of boronic acid in (K-T_B)5, we conducted fructose inhibition experiments.

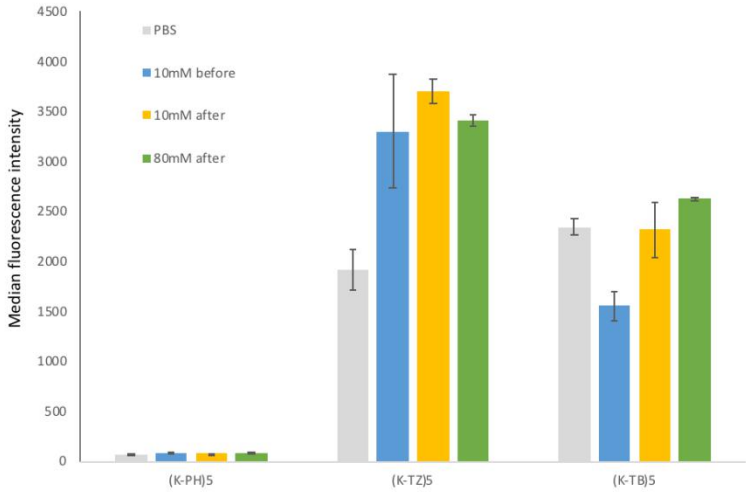


Figure 2.2.5. Fructose inhibition experiment comparing (K-T_B)5, (K-T_Z)5 and (K-P_H)5. Peptide concentration: 0.5 µM; cell density: 2 million cfu/mL. 10 mM before: cells were incubated with fructose for 30 min before the addition of peptides; 10 mM/80 mM after: 10 mM/80 mM fructose was added to the mixture after the cells were incubated with peptides for 30 min. The samples were incubated for another 30 min before readings were taken. All the experiments were carried out with triplicates.

We selected fructose in this experiment because it is known to have a higher binding affinity to boronic acids compared to other sugars such as glucose. The results in **Figure 2.2.5** suggest that fructose indeed inhibited the binding of (K-T_B)5 at 10 mM, if it was added before the peptide was allowed to incubated with *E. coli*. No significant inhibition effect was observed when it was added after the incubation period even with 80 mM concentration, suggesting a slow dissociation kinetics of the bound peptide.

2.2.4.3 Comparison of binding in different media

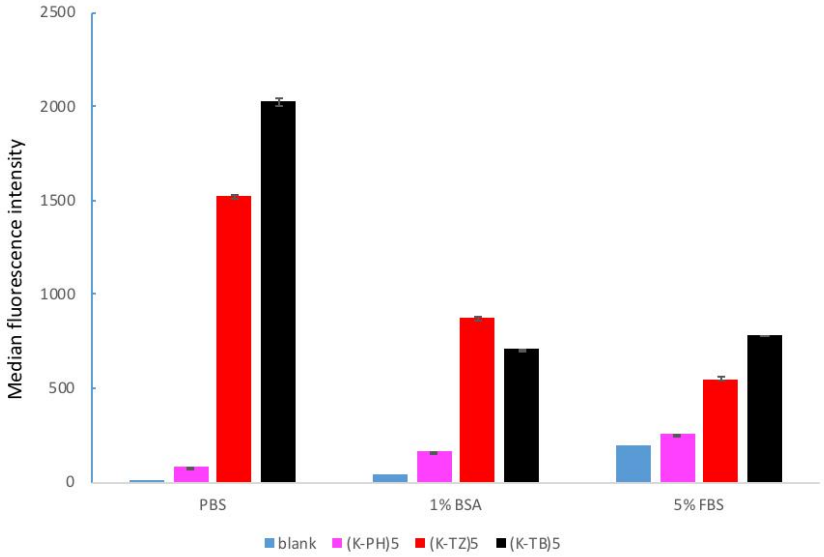


Figure 2.2.6. Flow cytometry experiments in different media. *E.coli* pellets were suspended into PBS, PBS with 1% (w/v) or 5% FBS (v/v). Final cell density: 2 million cfu/mL; peptide concentration: 0.5 μM. All the experiments were carried out in triplicates

Although the result in PBS is discouraging, we examined the binding capacity of these peptides in complex media to find out if the boronic acid would help (K-T_B)5 escape from the hijack of proteins. Flow cytometry experiments were then carried out in PBS, PBS with 1% BSA and PBS with 5% fetal bovine serum (FBS). Unfortunately, the median fluorescent intensity for both (K-T_B)5 and (K-T_Z)5 dropped significantly in 1% BSA or 5% FBS. We can conclude that neither of these peptides can be useful in real application if they cannot overcome the inhibition from serum.

2.2.5 Targeting bacteria with cyclic peptides

Cyclic peptides have attracted the attention of researchers as potential therapeutics. They have a relatively longer half-life in blood serum compared to linear peptides and offer better binding affinity due to their lowered entropy. In fact, there are a good number of antibiotics that are composed of cyclic peptides in nature, such

as nisin, polymyxin and daptomycin. Inspired by these naturally occurring cyclic peptide antibiotics, we decided to design a new set of peptides based on a bicyclic scaffold.

2.2.5.1 Peptide cyclization by iminoboronate

Dr. Anupam Bandyopadhyay in our lab has discovered that a short peptide can go through spontaneous cyclization in neutral aqueous solution when 2-APBA and lysine are both present in the sequence.[5] He designed an unnatural amino acid named AB3 that contains the 2-APBA headgroup and incorporated it into peptides through SPPS. This intramolecular iminoboronate induced cyclization is rapid, reversible and sensitive to pH and small molecules. Importantly, the imine can be further reduced by $NaCNBH_3$ to give a permanent cyclic structure with an N-B coordinated boronic acid.

2.2.5.2 Increasing the cyclization efficiency of iminoboronate mediated peptide cyclization

Although intramolecular iminoboronate formation is independent of peptide concentration, its cyclization efficiency maximizes at 85% in short peptides (<10 residues) and becomes even less efficient when the peptide gets longer. This can be explained by the mediocre binding affinity between 2-APBA and lysine. In an effort to improve the cyclization efficiency, we examined the K_d between 2-APBA and lysine analogues including ornithine, diaminobutyric acid (Dab) and diaminopropionic acid (Dap). When 2-APBA forms iminoboronate with amines, the maximum absorption shifts from 254 nm to 280 nm, enabling assessment of binding affinity through UV-vis based experiments (**Figure 2.2.7a-d**).

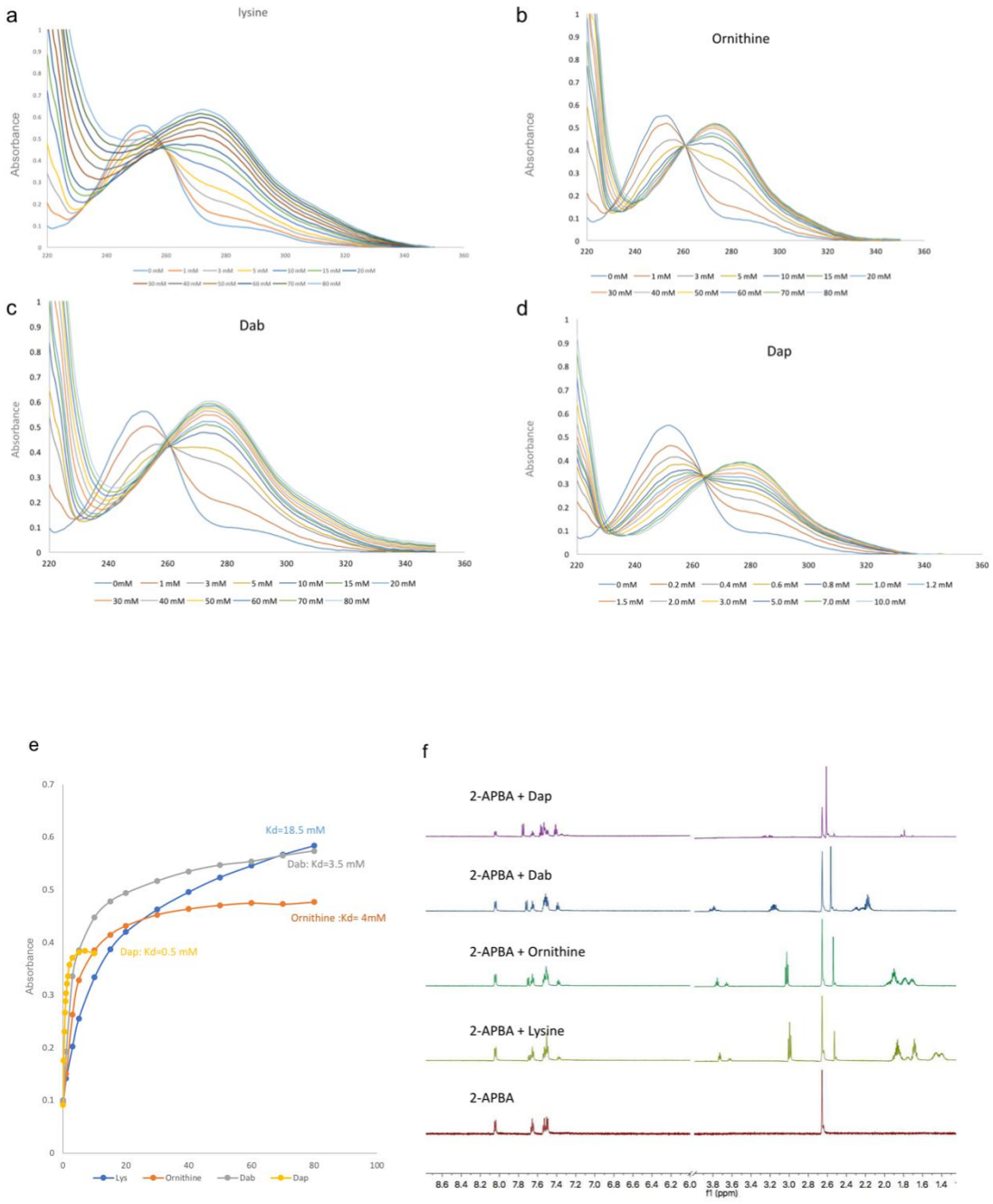


Figure 2.2.7. a)-d) Titration results for lysine, ornithine, Dab and Dap. 2-APBA was dissolved in 1 mL PBS, pH 7.4 in a quartz cuvette (1 cm pass length) at 50 µM. Lysine or its derivatives dissolved in PBS as 1 M stock was gradually added into the solution. UV-vis absorbance from 220 nm to 350 nm was recorded; e) Absorbance at 280 nm was plotted against the concentration of lysine derivatives; f) ^{1}H-NMR showing 2-APBA reacting with lysine derivatives.

Interestingly, it was discovered that the K_d value is positively correlated with side chain length. The best candidate, Dap, has a K_d of 0.5 mM, which is 39 folds lowers than that of lysine. NMR experiments also suggested a better conversion to the iminoboronate product when 2-APBA and Dap were mixed at 5 mM each. Taken together, we concluded that Dap could be a better amino acid than lysine to initiate peptide cyclization.

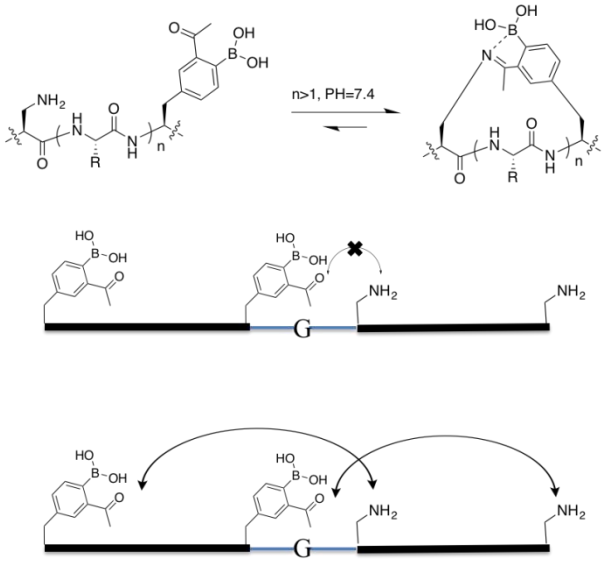


Figure 2.2.8. Iminoboronate between AB3 and Dap mediated peptide cyclization. AB3 does not cyclize with the lysine residue that is too close (i.e. only one amino acid in between)

We synthesized a number of peptides on rink-amide resin to compare the cyclization efficiency between Dap and lysine. Synthesis of Fmoc-AB3 was carried out according to reported procedure.[5] As is shown in **Figure 2.2.9**, KL1 showed complete cyclization at pH 7.4 whereas KL1-Lys only showed 80%. The results for other peptides are summarized in **Table 2.2.3**. A longer peptide KL4 showed more than 90% completion (data not shown) and the reported cyclization efficiency for its lysine mutant was ~50%.

Table 2.2.3. Mass and cyclization efficiency of model peptides

peptide	sequence	cyclic%	mass(m/z)
KL1	Ac-Dap-A-A-G-AB_3-$CONH_2$	100	$[M+H]^+$ cal.578.2740; obs.578.2729
KL1-Lys	Ac-Lys-A-A-G-AB_3-$CONH_2$	80	$[M+H]^+$ cal.620.3210; obs.620.3251
KL2	Ac-Dap-A-A-A-A-G-AB_3-$CONH_2$	100	$[M+H]^+$ cal.720.3483; obs.720.3445
KL3	G-A-A-Dap-R-G-D-AB_3-$CONH_2$	100	$[M+H]^+$ cal.935.4501; obs.935.4478
KL4	Ac-Dap-A-A-A-D-A-A-A-D-G-AB_3-$CONH_2$	>90%	$[M+2H]^{2+}$ cal.537.7366; obs. 537.7353
KL5	Ac-Dap-G-Lys-A-A-G-AB_3-$CONH_2$	100*	$[M+H]^+$ cal. 763.3905; obs.763.3916

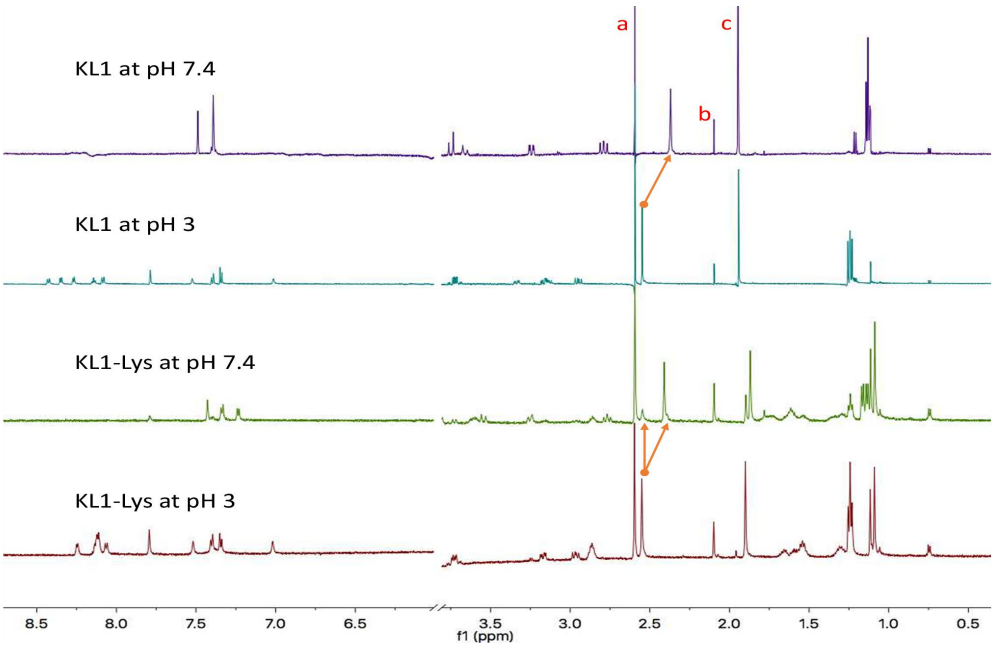


Figure 2.2.9. ^{1}H-NMR comparison between KL1 and KL1-Lys at pH 3 and pH 7.4. The orange arrows highlight the acetyl peak shift of AB3 upon iminoboronate formation. Peak a: DMSO as an internal standard; b: residual acetone; c: acetyl group of the N-terminus. The peptide samples were prepared at 0.5 mM in PBS (with 10% D_2O) and tuned to the desired pH with 0.5 M HCl or NaOH.

It is important to note that a mixed peptide KL5, which includes both Dap and Lys, preferentially cyclizes towards Dap (~87%, **Figure 2.2.10a**). As is mentioned, the imine bond can be quickly reduced by $NaCNBH_3$ to yield a permanently cyclized product. Interestingly, reduction of cyclized KL5 gave a single product, presumably the AB3-Dap cyclic isomer with a free lysine residue (**Figure 2.2.10b**).

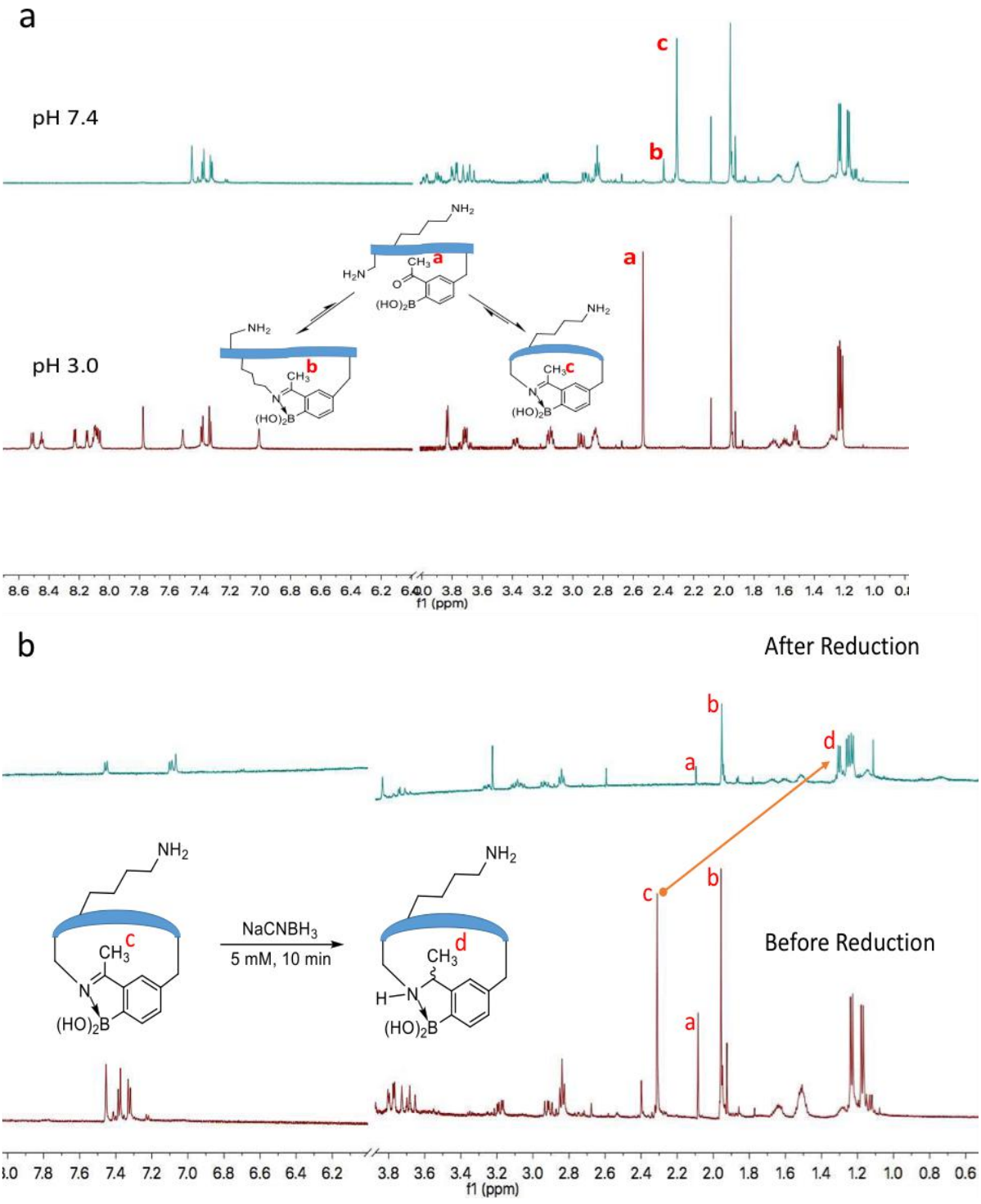


Figure 2.2.10. Cyclization of KL5. a) 1H-NMR of KL5 at pH 3 and pH 7.4. The AB3 acetyl peaks (a, b, c) are labeled in the NMR spectra and accordingly in the peptide structures; b) KL5 before and after reduction

2.2.5.3 Targeting bacteria with an amphiphilic bicyclic peptide

After Dap was determined to be the optimal amino acid for peptide cyclization, we synthesized a small group of peptides that are cyclized by reduced iminoboronate and evaluated their bactericidal capability against *S. aureus*, a gram- positive bacteria.

Table 2.2.4. Peptide sequence in the group

Name	*Sequence*
KL6	WWW-RRR
KL7	RWR-RWR
KL8	RWR-WRW
KL9	WRR-RWW
KL10	RRW-RWW
KL11	WRR-WWR
KL12	RRW-WWR
KL13	RWW-RWW
KL14	WRW-WRW
KL15	WWR-WWR
KL16	WRR-WWW
KL17	RWW-WWW
KL18	WWR-WWW
KL19	RWW-WWW

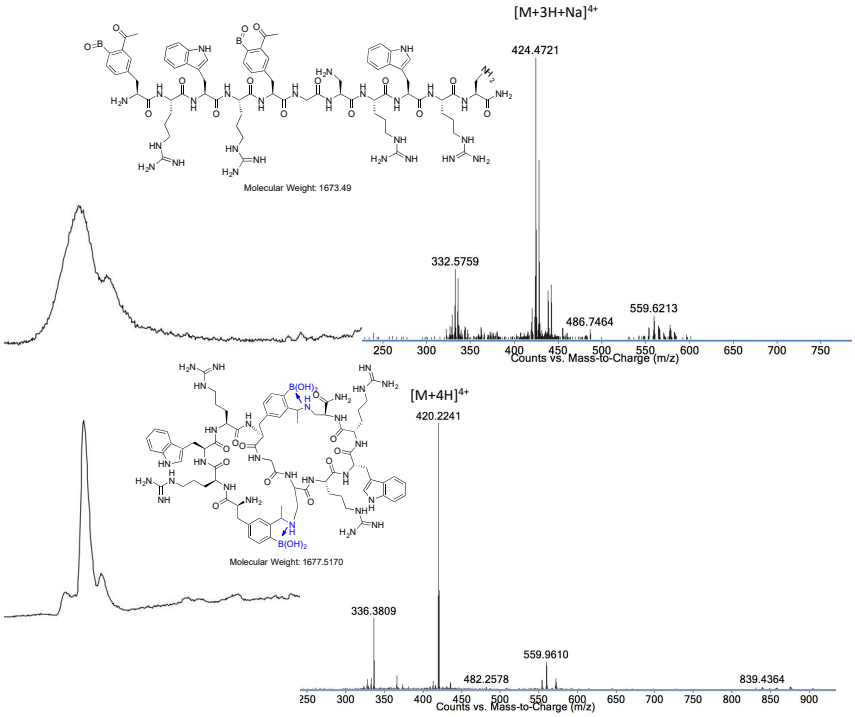


Figure 2.2.11. Example of reduced bicyclic peptide KL7, including the LC trace and mass specs of KL7 before or after reduction

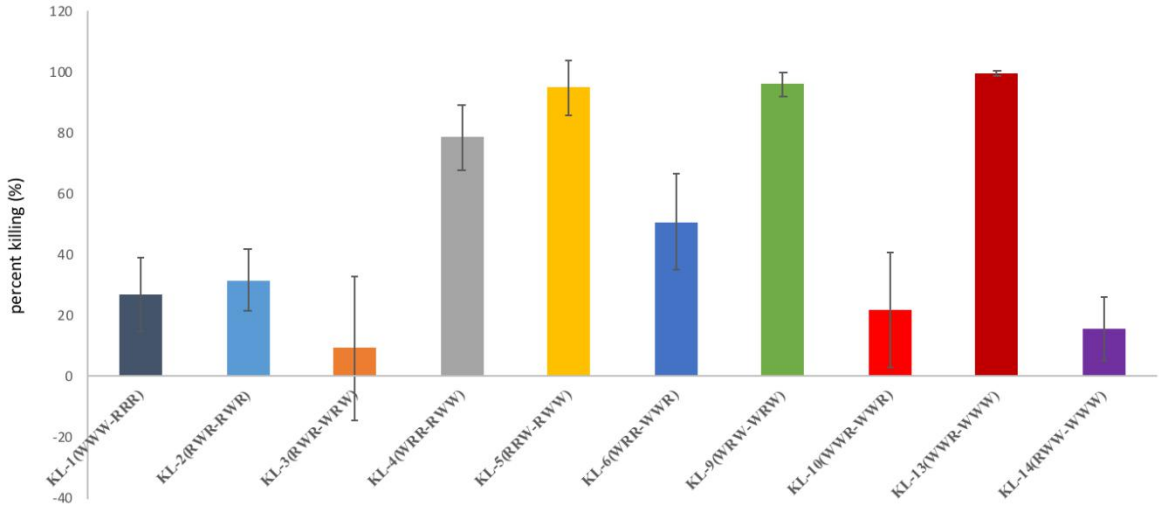


Figure 2.2.12. killing capability of KL1-6, 9, 10, 13 and 14 against *S. aureus* at 1 μM in Tris acetate buffer. All the experiments were carried out in triplicates

We selected some peptides from **Table 2.2.4** and conducted cell killing experiments. Detailed procedures are included at the end of this chapter. As is summarized in **Figure 2.2.12**, some peptides indeed showed cell killing capability at 1 μM. Importantly, peptide KL9, KL10, KL14 and KL18 showed more than 80% killing and they were selected for further investigation. We then examined their bactericidal activity in a more complex media. The killing capability of KL14 and KL18 was evaluated in Tris acetate buffer with 10% human blood serum (HBS). Unfortunately, none of these peptides showed any significant killing in this condition (**Figure 2.2.13a**). This result indicates that the protein molecules in serum were likely to hijack the peptides and demolish their killing capability.

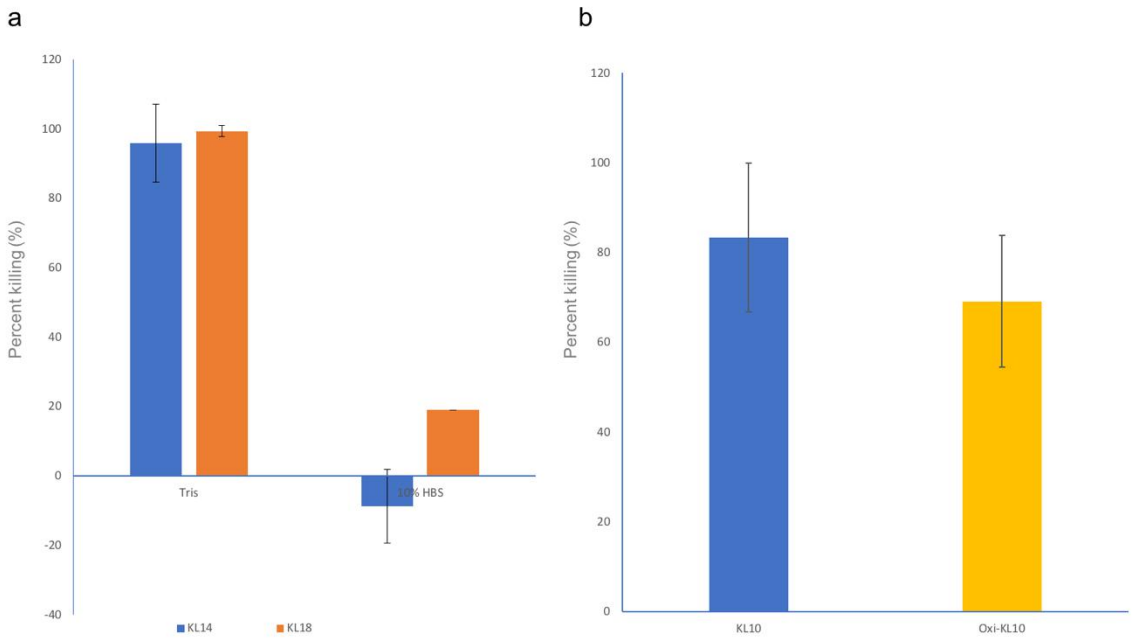


Figure 2.2.13. a) *S. aureus* killing assay in Tris acetate buffer and Tris acetate buffer with 10% human blood serum. Peptide concentration: 1 μM. The addition of HBS significantly demolished the killing capability of KL14 and KL18 against *S. aureus*; b) Comparison of KL10 and oxidized KL10 in killing *S. aureus* in Tris acetate buffer

We also examined the contribution of the boronic acid in KL10 by oxidizing the boronic acid into a hydroxyl group with H_2O_2. The new peptide, Oxi-KL10 was compared to KL10 with the same killing assay in Tris acetate buffer without HBS. Surprisingly, Oxi-KL10 showed comparable killing capability to KL10 (**Figure 2.2.13b**). The killing capability of this series of peptides was most likely to be induced by their amphiphilic nature instead of boronic acid.

2.2.6 Summary

In summary, we have explored the possibility of utilizing boronate ester formation to target the diol structures on bacteria. Although the overall result is not ideal, we could still draw some conclusions from these experiments, which could be useful for future projects. We know from the anisotropy experiment with LPS

micelles that multivalence does have an impact on the binding affinity. However, the fructose inhibition experiment indicates that boronic acid only contributes partially to target binding. This explains why (K-T_Z)5 and (K-T_B)5 did not show significant difference binding to *E. coli*. The results tell us that the function of boronic acid in these peptides is quite limited and boronate ester formation is not the main driving force for binding. Although the EC_{50} of (K-T_B)5 is in the low micromolar range, results indicate that it likely binds through electrostatic or hydrophobic effect. More importantly, these peptides can be easily hijacked by proteins in serum.

We revised our strategy and designed a panel of bicyclic peptide that are permanently cyclized by reduction of an intramolecular iminoboronate, aiming to increase binding potency and proteolytic stability. We found that the replacement of lysine to Dap can increase the cyclization efficiency. Although some bicyclic peptides demonstrated impressive killing capability in Tris acetate buffer, they are unlikely to be applied in true biological conditions due to poor activity in blood serum. The comparison between KL10 and Oxi-KL10 also tells us that boronic acid may not be vitally important in these peptides.

A common problem for the current boronic acid containing peptides, both linear and bicyclic, is that they are unable to overcome the inhibition of serum. This is an important factor to be considered when designing peptides to target bacteria. Although we can consider other boronic acid warheads that bind better to diols, we cannot predict if the new molecule will be able to function in complex media that contains albumin. Rational design, or trial and error based strategies can be time consuming. We decided to develop a screening platform to find peptide probes or antibiotics that can target bacteria in presence of serum.

2.2.7 General information

2.2.7.1 Flow cytometry with *E .coli*

E. coli cells from sub-culture was grown to OD_{600} ~0.6 and spun down at 3000g for 10 min at 4 °C. The supernatant was discarded and the cells were re-suspended into PBS, pH 7.4 and diluted to 2 million cfu/mL for flow cytometry experiment. Peptides from stock solutions were added to make a final concentration of 0.1, 0.3, 0.5, 0.7, 1 μM respectively. The cells were incubated for half an hour before flow cytometry.

For fructose inhibition experiment, the mixture was incubated for an extra 30 min after the addition of fructose.

2.2.7.2 Cell killing experiments with *S. aureus*

A colony of *S. aureus* was selected on an agar plate and pipetted into LB broth to grow overnight at 37 °C on a shaker. An aliquot of the broth was diluted 100 times into fresh broth and sub-cultured until OD600 ~0.6. The cells were centrifuged at 400 rpm for 5 min and the supernatant was discarded. The cell pellet was washed by 10 mM Tris acetate buffer twice and resuspended in fresh buffer. The cells were pipetted into 96 well plate (200 μL in each well) and the peptides in stock solutions were added until a final concentration of 1 μM (including three blanks where only DMSO was added). Each peptide was tested on triplicates. The plate was incubated for 10 min at 37 °C. The sample in each well was diluted 100 times and an aliquot of 100 μL was spread on agar plates. The plates were incubated at 37 °C overnight and colonies on each plate were recorded. The average colonies of each peptide treated samples divided by the average colonies of blanks stands for percent survival of *S. aureus*, which was subtracted by 100% to give the percent killing value.

2.2.8 References

(1) Raetz, C. R. H.; Whitfield, C. Lipopolysaccharide Endotoxins. *Annu. Rev. Biochem.* **2002**, *71* (1), 635–700. https://doi.org/10.1146/annurev.biochem.71.110601.135414.

(2) Lee, K.; Campbell, J.; Swoboda, J. G.; Cuny, G. D.; Walker, S. Development of Improved Inhibitors of Wall Teichoic Acid Biosynthesis with Potent Activity against Staphylococcus Aureus. *Bioorganic Med. Chem. Lett.* **2010**, *20* (5), 1767–1770. https://doi.org/10.1016/j.bmcl.2010.01.036.

(3) Lee, W.; Schaefer, K.; Qiao, Y.; Srisuknimit, V.; Steinmetz, H.; Müller, R.; Kahne, D.; Walker, S. The Mechanism of Action of Lysobactin. *J. Am. Chem. Soc.* **2016**, *138* (1), 100–103. https://doi.org/10.1021/jacs.5b11807.

(4) Swoboda, J. G.; Meredith, T. C.; Campbell, J.; Brown, S.; Suzuki, T.; Bollenbach, T.; Malhowski, A. J.; Kishony, R.; Gilmore, M. S.; Walker, S. Discovery of a Small Molecule That Blocks Wall Teichoic Acid Biosynthesis in Staphylococcus Aureus. *ACS Chem. Biol.* **2009**, *4* (10), 875–883. https://doi.org/10.1021/cb900151k.

(5) Bandyopadhyay, A.; Gao, J. Iminoboronate-Based Peptide Cyclization That Responds to PH, Oxidation, and Small Molecule Modulators. *J. Am. Chem. Soc.* **2016**, *138* (7), 2098–2101. https://doi.org/10.1021/jacs.5b12301.

2.3 Targeting lipid II peptide in gram positive bacteria with cyclic peptides through iminoboronate formation

(The work in this chapter was jointly performed with Dr. Breanna L. Zerfas)

2.3.1 Introduction

Since the discovery of penicillin, antibiotics have been a standard answer to treat bacterial infections. Many other antibiotics were discovered or rationally designed post-penicillin discovery which greatly reduced the death rates caused by bacterial infection. Unfortunately, antibiotic resistance (AR) has become an increasing clinical concern due to its natural occurrence coupled with the abuse of antimicrobial agents.[1,2] Limiting the use of antibiotics helps to slow down the emergence of antibiotic resistance.[3] However, the development of new antibiotics is always in urgent need to counter the existing and prepare for the emergence of new AR strains.

Although the cell wall and membrane serve as shelters to protect the bacteria from its surrounding environment, they are also the most straight forward and accessible targets for antibiotics. Our lab has been trying to understand the roles of lipids and lipid modification in bacterial growth. Research in this area will not only help us to understand the functions of these important components on bacteria cell surface, but also provide guidance for us to design chemical probes or novel antibiotics for AR bacteria. In this chapter, I will talk about the effort that we have devoted to developed novel antibiotics that target lipid II in gram-positive bacteria.

2.3.1.1 Peptidoglycan and antibiotics

As is discussed earlier, peptidoglycan is crucially important for both gram-positive and gram-negative bacteria. It is mainly composed of sugars (N-acetylglucosamine, GlcNAc and N-acetylmuramic acid, MurNAc) and amino acids (stem peptide). Sugars are connected by a β–(1,4)-glycosidic bond and stem peptides on MurNAc are cross linked by transpeptidase. The resulting product is a mesh-like

structure outside of bacteria membrane that is strong and rigid to provide structural strength and counter the osmotic pressure of the cytoplasm.[4]

The synthetic pathway of peptidoglycan has always been a popular target for antibiotics. **Figure 2.3.1.** illustrates the targets of some well-known antibiotics that interrupts the synthetic pathway of peptidoglycan.[5] For example, CDFI binds to an peptidoglycan flippase and stops it from transporting lipid II to the surface of bacteria, disrupting cell wall synthesis;[6] Ramoplanin binds to lipid II to prevent the transglycosylation step;[7] Moenomycin also inhibits the transglycosylation step by mimicking the structure of lipid II and binding to transglycosylases;[8] Bacitracin inhibits the dephosphorylation of C_{55}-isoprenyl pyrophosphate to stop the regeneration of this lipid carrier.[9] A recently discovered antibiotic, Tragocil inhibits the biosynthetic pathway of wall teichoic acid by binding to ABC transporter that exports the precursor to cell surface.[10,11]

There are some antibiotics that also affect the biosynthesis of peptidoglycan but are not presented here. For instance, β-lactams such as penicillin bind to a crucial enzyme (penicillin-binding protein, or transpeptidase) to stop the bond formation between the stem peptides.[12] The complex biosynthesis of peptidoglycan requires a number of enzymes and substrates, leaving the bacteria vulnerable to different types of antibiotics. It is reasonable to believe that more antibiotics will be developed or discovered that target the biosynthetic pathway for peptidoglycan.

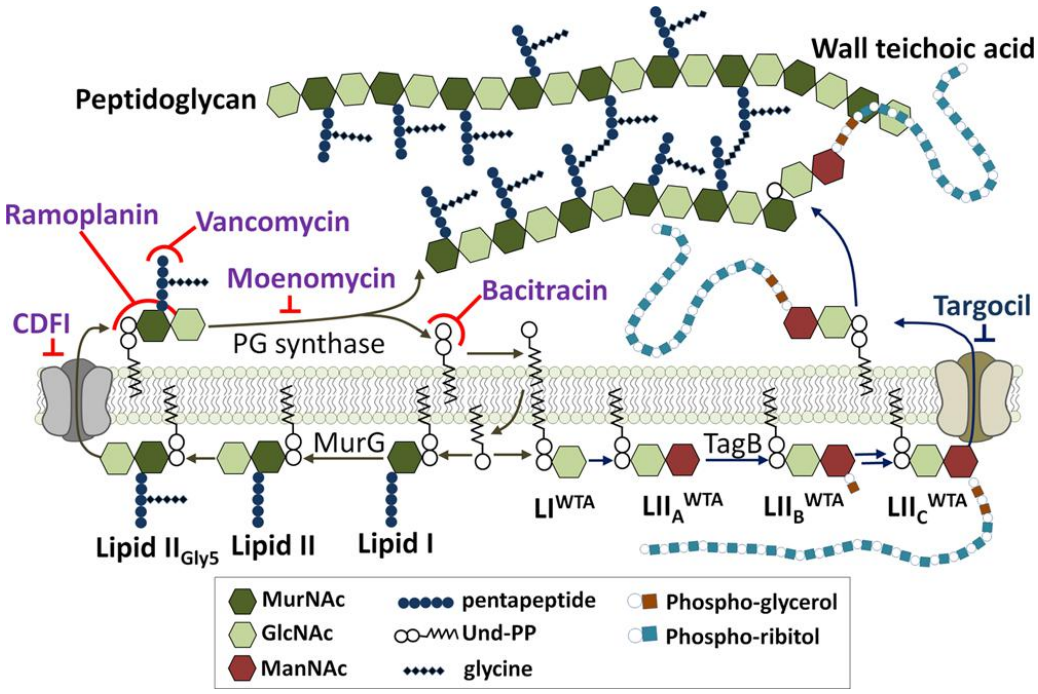


Figure 2.3.1. Biosynthetic pathways of peptidoglycan and the targets of some representative antibiotics. Image was taken from reference 5.

2.3.1.2 Lipid II and vancomycin

Vancomycin is a potent glycopeptide antibiotic that is being used to treat serious gram-positive bacterial infections including methicillin-resistant *S. aureus* (MRSA) and multidrug-resistant *S. epidermidis* (MRSE). It binds to the terminal D-Ala-D-Ala moieties of the stem peptide from lipid II through a five-point hydrogen bonding. This binding event prevents the polymer growth during the process of peptidoglycan synthesis. One common resistance mechanism developed against vancomycin is the mutation from D-Ala-Ala to D-Ala-D-lactase, which causes the loss of a key hydrogen bonding (**Figure 2.3.2c**).[13,14] The fight between human and bacteria will never stop as bacteria continue to develop resistance to newly discovered antibiotics. However, resistance to vancomycin was not observed until about 30 years after its discovery, which tells us that it is not easy for bacteria to develop resistance against antibiotics of this kind. We hypothesized that we could design a potent

antibiotic by mimicking the structure and binding mode of vancomycin. The sequence of stem peptide includes a lysine residue, which is quite intriguing to us given our recent focus on iminoboronate. Importantly, lysine is a highly conserved residue because it is responsible for the linkage formation between stem peptides.[15] By incorporating a 2-APBA warhead in our peptide design, we hypothesized that we might be able to design an antibiotic that has a lower chance for bacteria to develop resistance.

The interaction between 2-APBA and lysine alone is not sufficient to elicit potent binding. We decided to build a multicyclic peptide library to mimic the structure of vancomycin and incorporate the 2-APBA warhead to achieve higher binding affinity

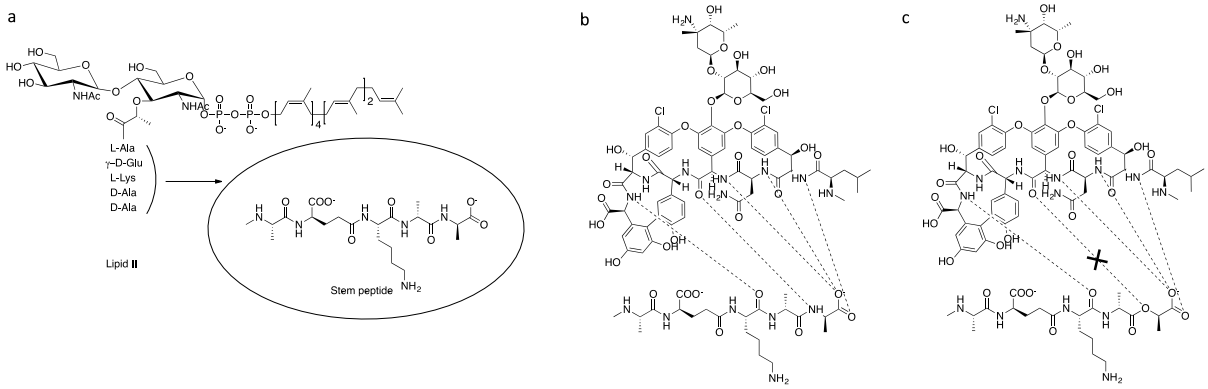


Figure 2.3.2. Lipid II and vancomycin. a) Structure of stem peptide from lipid II; b) Binding mode of vancomycin; c) Mechanism of resistance against vancomycin

2.3.1.3 Screening techniques for peptide discovery

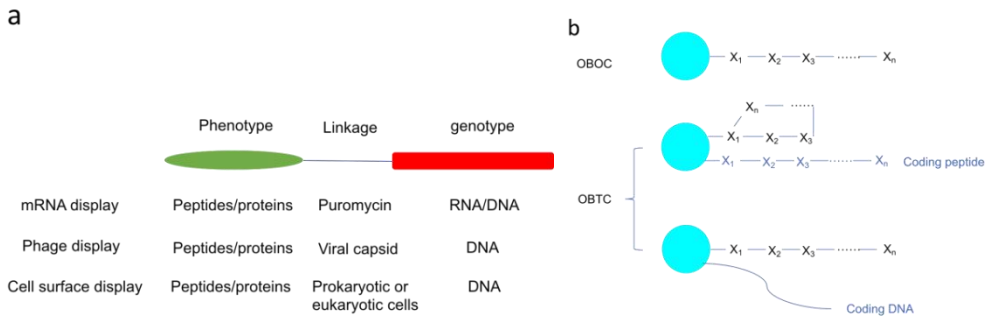


Figure 2.3.3. Biological libraries and synthetic libraries. a) biological library including mRNA display, phage display and cell surface display; b) synthetic libraries including OBOC and OBTC

Pharmaceutical companies usually employ screening techniques as an efficient method to discover drug candidate for a specific target. Screening-based ligand discovery techniques have many advantages over rational design. First of all, a large number of compounds can be evaluated at the same time in early stages instead of tediously conducting the trial and error experiments. Preliminary results from screening provides useful information for the following structure-activity relationship study (SAR). Moreover, by adding BSA or blood serum in the buffer during the screening process, the selected compounds have a good chance to overcome the hijack of small molecules or proteins in real application. Although a common problem with library-based screening techniques is the occurrence of false positives, it can be easily solved by conducting multiple rounds of screening to narrow down the potential hits. Also, hit validation is a crucial step to distinguish false positives from desired compound(s).

Peptide-based libraries can be categorized into biological libraries and synthetic libraries.[16] Biological libraries, including phage display and bacterial display usually consist of two parts, genotypes (DNA sequence that encodes the peptide

sequences) and phenotypes (expressed peptides). Since the work from George Smith,[17] phage display has evolved to become one of the most popular screening platforms for novel peptide discovery. For example, our lab has been utilizing an M13 phage displaying five copies of the same peptides on the N-terminus of the pIII protein to screen against bacteria and protein targets.[18] pVIII protein can also be used to express peptides. However, this coat protein exists in many copies and usually only short peptides are expressed on a fraction of the protein.[16] One advantage of phage display is the large size of the library. The diversity of phage display library goes up as the length of the variant peptides grows. In real application, a phage display library usually contains 10^8 to 10^9 different peptide sequences. Another highlight of phage display includes the ease to obtain, amplify and store, enabling both in vitro as well as in vivo applications. Importantly, more than one round of screening allows for better convergence. However, the disadvantages of phage display are also concerning. For example, the pIII protein is necessary for the invasion of the phage into bacteria, and significant changes to the protein can potentially impair its ability to infect bacteria and amplify. Also, the sequence after each round of amplification could be biased due to the difference in growth rates. Most phase display platforms are also still limited to natural amino acids. Although significant efforts has been devoted to this area, incorporating an unnatural amino acid into the library is still challenging.[19–21] A more popular choice is to modify the phage through site-specific reactions on cysteine or serine to add on new warheads or change the structure of the peptide.

Another biological library is cell surface display, including bacterial and yeast display. Compared to the limited size of polypeptide that can be displayed by phage, cell surface display provides a better platform for large peptides or even proteins. Developed in 1986 by Charbit and Freudl et al,[22,23] bacterial display is becoming more

popular in the field of creating novel vaccines[24] , exploring protein-protein interactions,[25] antibody engineering[26] and epitope mapping of an antibody[27] et al. Emerging later than bacterial display, yeast display is mostly used for protein engineering.[28] Based on eukaryotic cells, yeast display can provide complex post-translational modifications to proteins, which are not available in other display technologies. Similar to phage display, cell surface displays are mostly limited to natural amino acids.

Peptide selection via mRNA display is also attracting more attention. Peptide libraries in mRNA display can be regarded as a fusion of a biological and synthetic library. This platform utilizes a puromycin linkage to fuse the translated peptides to their mRNA progenitors. Sequences of the peptides can be decoded by reverse transcription, PCR and DNA sequencing. The size of the mRNA display (10^{12}-10^{14} members) can be larger than that of phage display whose diversity is compromised due to transformation efficiency.[16] Recently, Suga et al has been applying the so called 'flexizymes' to incorporate unnatural amino acids into mRNA display, making this technique highly competitive compared to other screening platforms.[29,30] However, the library construction is quite complex and costly.

Synthetic libraries compared to biological libraries are generally easier to construct.[31,32] Because the library is built on resin, it can tolerate harsh conditions for a wider range chemistry that may not be suitable on phage or bacteria. Importantly, incorporating unnatural amino acids into peptide sequence can be more straightforward during the construction of synthetic libraries. One major drawback for synthetic libraries is the decoding process. Unlike biological libraries, the sequence of the peptides cannot be easily obtained through DNA sequencing. However, the

development of partial Edman degradation[33,34] and LC/MS/MS[35] provides a good solution to this problem. Although these methods are limited to linear peptides, the emergence of one-bead two-compound library (OBTC) is attracting more attention.[36,37] In an OBTC library, a linear peptide is usually conserved as the coding sequence while the rest of the peptides on the same bead present the designed structure and warhead(s) for screening purposes. A variant of synthetic OBTC library is encoded by DNA.[38] This method provides an easier decoding process. However, Dr. Breanna Zerfas in our lab has explored this strategy without too much success due to cost and time limit. Considering the availability of screening techniques, we decided to construct a synthetic library on resin to incorporate boronic acid containing amino acids. The library only takes a few days to construct and Professor Eranthie Weerapana's lab in Boston College provided help with the decoding process.

2.3.2 Construction of One-bead Two-compound (OBTC) cyclic peptide library

2.3.2.1 Bicyclic peptide library

To mimic the function and structure of vancomycin, we decided to develop a constrained bicyclic peptide library on resin to target bacteria. As is mentioned, one of the biggest challenges in peptide screening on resin is the sequencing method. For One-bead One-compound based strategy, the peptide itself can be directly sequenced by partial Edman degradation and MALDI TOF mass spec or LC/MS/MS. However, our previous efforts to sequence the peptide by Edman degradation was not successful presumably due to a low concentration of peptide on one bead and results from DNA encoded library was not encouraging either. We then decided to collaborate with Professor Weerapana's lab and sequence our peptide through LC/MS/MS. Besides AB3, we chose 7 other amino acid as variants. These amino acids are selected to cover a better diversity of the side chain.

As is discussed in **Chapter 2.2**, we developed a novel strategy to synthesize bicyclic peptide, which can be utilized to construct peptide libraries on resin. However, boronic acid containing amino acids can be difficult to analyze on mass spec due to its propensity to lose a water molecule. To this end, the boronic acids in the test peptide were oxidized into hydroxyl groups by 1mM H_2O_2 solution. Five samples were prepared as described at the end of this chapter and sent out for sequencing. Unfortunately, none of these samples resulted in the right sequence, although the mass in MS1 was correct. A possible explanation is that the acetyl group in AB3 caused some problems because an electrophile is not commonly seen in natural amino acids. This acetyl group was thus reduced into a hydroxyl group by $NaBH_4$ in methanol. Both the oxidation and reduction were proven to be efficient on our own LC/MS, but the peptide still could not be successfully sequenced. We then decided to adopt an OBTC library with a portion of peptide excluding AB3 so it stays linear for the purpose of sequencing

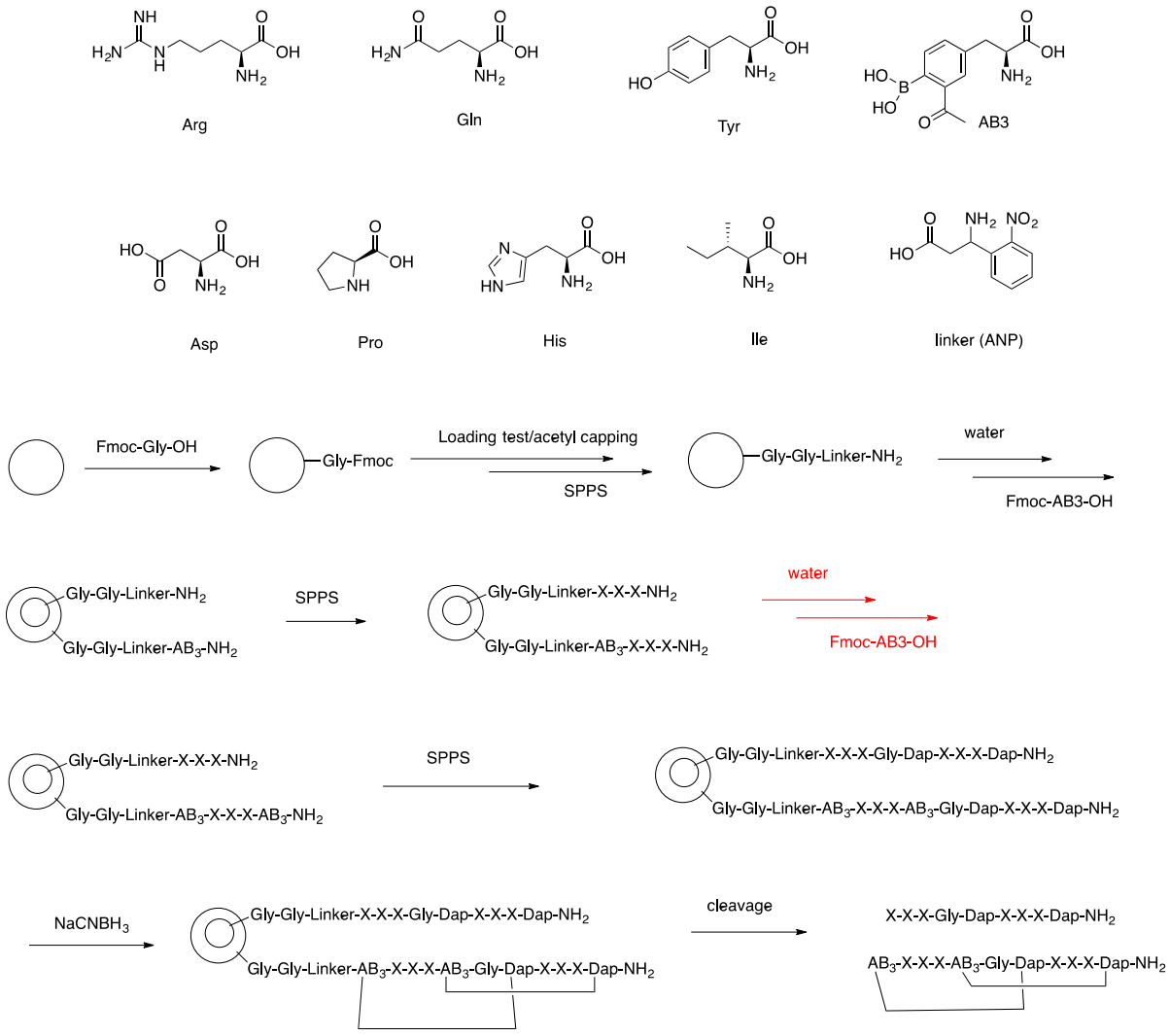
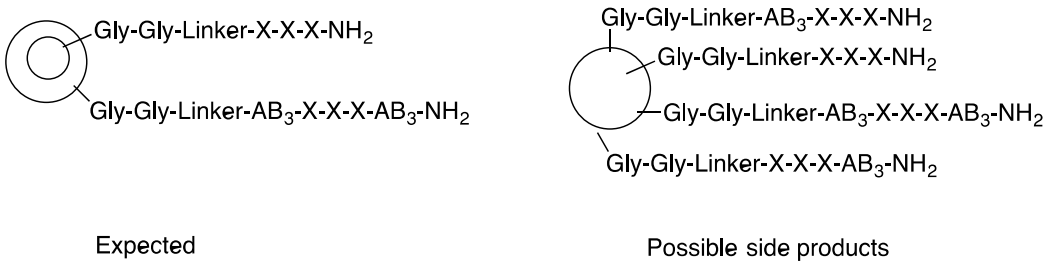


Figure 2.3.4. Original design for bicyclic library cyclized by iminoboronate. X in the sequence stands for randomized amino acid, which can be any of the seven amino acids.

Before synthesizing the library, we carried out a test synthesis of a model peptide on a tentagel resin that could be swelled in both organic and aqueous solution. The peptide was loaded onto the resin through two glycine residues and a photolabile linker. Glycine residues was included to increase the distance between the peptide library and the resin. The photolabile linker is resistant to both acidic and basic condition, which allowed the peptide to remain attached to the resin during TFA treatment to remove the protecting groups on side chains. Importantly, we could release the peptide from resin under UV light, providing an efficient and clean way

for peptide cleavage.

During the synthesis of the OBTC library, we adopted the 'core and shell' strategy. The protocol was adopted according to a reported procedure.[36] Briefly, the first few residues were coupled to the resin though standard SPPS. To couple the AB3 residue, the resin was soaked in DMF for 15 min, followed by 3:1 DMF/H_2O, 1:1 DMF/H_2O, 1:3 DMF/H_2O, and finally in pure water. Water was drained after overnight incubation and the resin was quickly suspended in a solution of activated Fmoc-AB3. It is believed that a portion of the peptides on resin were disguised by a layer of water so that they were not accessible for reactants dissolved in DCM or ether. Peptides without AB3 could then be used as a coding sequence. However, a major flaw in our library design was the need for a second 'core and shell' step. When the resin went through the same procedure, not all the peptides originally in the core were sheltered again by a layer a water. Also, peptides that were exposed for coupling reaction previously were not completely exposed in step. This phenomenon resulted in a complex mixture on the same bead. As a matter of fact, we did not find any precedent where multiple 'core and shell' steps were conducted on the same peptide library.



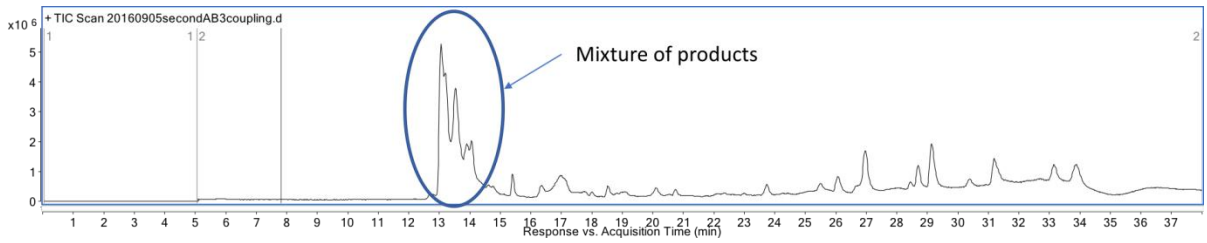


Figure 2.3.5. Expected and unexpected products after the second 'core and shell' step. LC/MS spectrum shows a complex mixture

2.3.2.2 Mono cyclic peptide library

To compromise to the fact that a bicyclic structure could not be easily realized, we decided to build a monocyclic library as a feasible alternative. There are different ways to make cyclic peptides on resin and we chose the most straightforward disulfide bond formation. This disulfide bond can be easily reduced by TCEP to yield a linear peptide for sequencing. Importantly, the 2-APBA warhead in AB3 can be reserved in this design for binding to lysine residues on stem peptides through iminoboronate. The new library would be composed of CXXX(AB3)XXXC, in which X stands for any of the seven amino acids and two cysteine residues would be oxidized to form disulfide bond. This library is much easier to construct due to the fact that it only requires one step of 'core and shell' when coupling the AB3 residue.

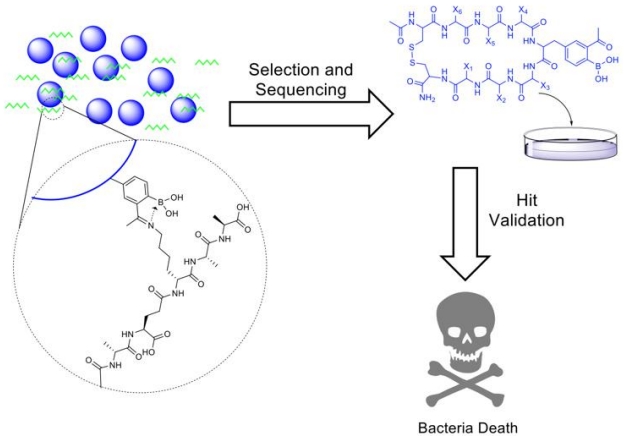
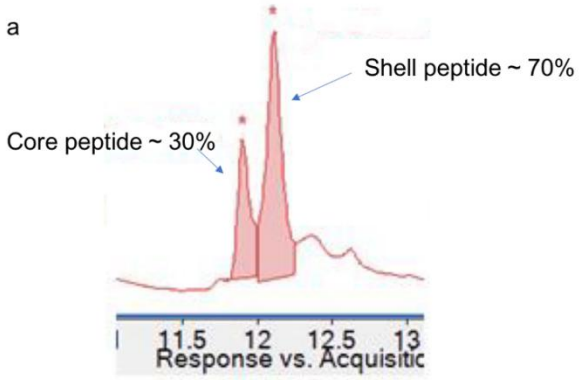


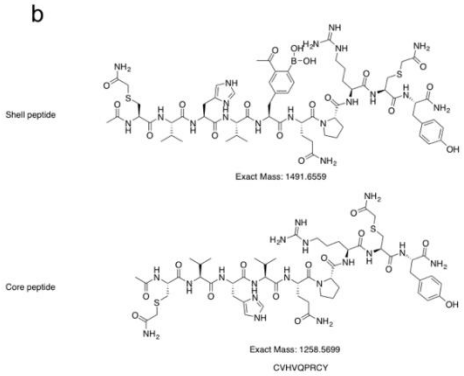
Figure 2.3.6. New design of monocyclic peptide library containing the 2-APBA warhead targeting lipid II in gram- positive bacteria

A test peptide was synthesized with a known sequence by the 'core and shell' strategy. LC trace showed that 30% of the peptide did not contain the AB3 residue. Three beads were picked up under microscope and the peptides on them were cleaved off resin. The N-terminus of the peptide was capped by acetic anhydride and cysteine residues were alkylated by iodoacetamide. Each sample was submitted for sequencing individually to Aaron Maurais in Weerapana's lab. The sequence highlighted by green color in **Figure 2.3.7c** matches the expected peptide sequence (G in the sequencing result stands for acetyl group at the N-terminus). This result confirms that peptide sequence on one bead can be successfully determined by LC/MS/MS.

a



b



c

ID	sequence	avg mass	monoisotopicmass	mass difference (obs mass - correct mass)	10uL_inj_1	10uL_inj_2	10uL_inj_3	Average
KL_29322	GCVHVQPRCY	1260.42	1259.58	0	471	457	554	**494.00**
KL_10114	GCYDVQPRCY	1302.41	1301.54	41.96	324	210	286	**273.33**
KL_29346	GCVHVQRPCY	1260.42	1259.58	0	246	130	212	**196.00**
KL_29315	GCVHVQVRCY	1262.43	1261.59	2.01	111	107	149	**122.33**
KL_67738	GCDYVQPRCY	1302.41	1301.54	41.96	41	42	68	**50.33**
KL_113381	GCRHVQRPCY	1317.47	1316.61	57.03	20	40	75	**45.00**
KL_113357	GCRHVQPRCY	1317.47	1316.61	57.03	0	29	62	**30.33**
KL_46129	GCPHVQPRCY	1258.4	1257.56	-2.02	44	6	31	**27.00**

Figure 2.3.7. Sequencing the test peptide. a) LC trace for the 'core and shell' peptides. Core peptide (without AB3) takes up 30% of the total peptide on the same resin; b) Structure and mass of the 'core and shell' peptides; c) Sequencing result for the test peptide. The correct sequence was highlighted in green color

2.3.2.3. Library construction and quality control

We went forward and built the peptide library by split and pool as is shown in **Figure 2.3.8.** Detailed synthesis is included at the end of this chapter. It is noteworthy that cyclization of the peptide library on resin with DMSO at the last step was not successful. We then applied a method reported by the Fernando group for on-resin disulfide formation using N-chlorosuccinimide.[39] This method allows for complete cyclization within 15 min.

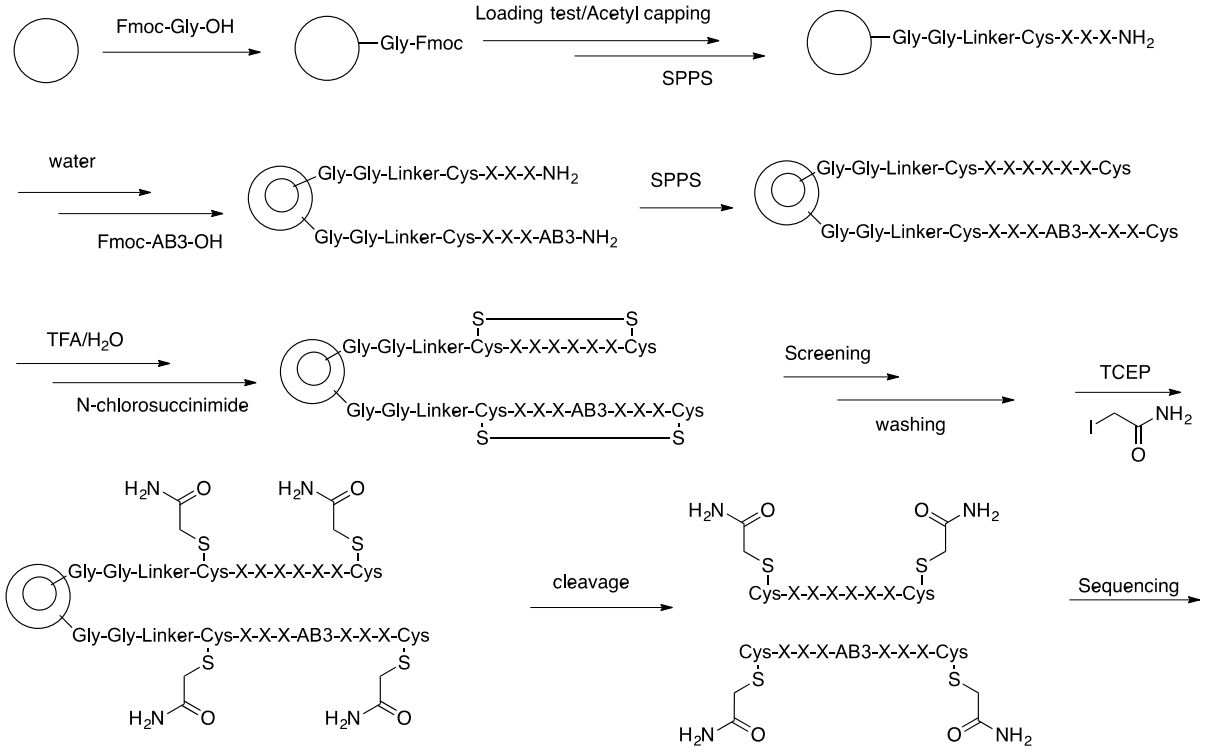


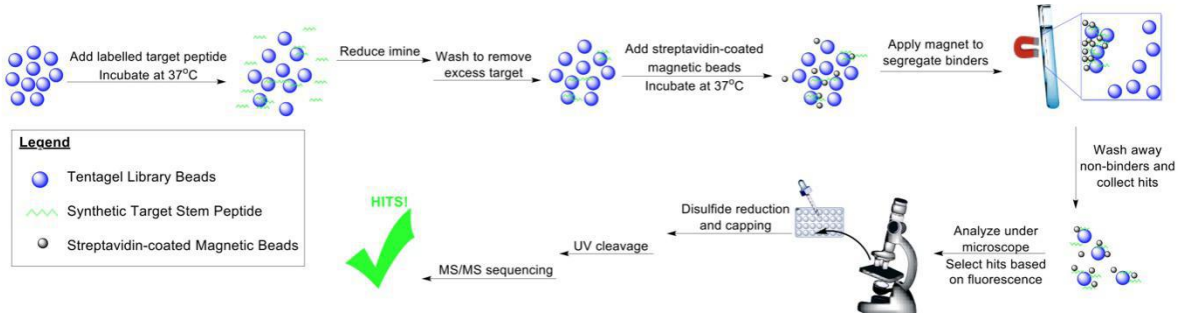
Figure 2.3.8. New design for a monocyclic peptide library cyclized by a disulfide bond

After synthesis of the library, we picked up 20 beads and cleaved the peptides off the resin for sequencing. The result is summarized in **Table 2.3.1**. As is shown by the table, we were able to sequence over 50% of the peptide. Considering we included multiple copies of the same sequence, we believed this was a satisfactory library to move forward for screening.

Table 2.3.1. Quality control for OBTC library

Sample	sequence	sample	sequence
1	GCYPPDYQC	11	NA
2	GCRDYDPDC	12	NA
3	GCHPHIRQC	13	GCIRDHYQC
4	NA	14	NA
5	GCIDYDYRC	15	GCYRHDYHC
6	NA	16	NA
7	GCPDQDHQC	17	GCHPHYPIC
8	NA	18	NA
9	GCIQHHIIC	19	GCYYHIHDC
10	GCIQHHIIC	20	NA

2.3.3 Screening process



Scheme 2.3.1. Screening scheme against lipid II pentapeptide

Screening was carried out according the scheme above. We included two steps of screening to achieve better convergence. A stem peptide derivative was synthesized containing two modifications: biotin and fluorescein. Biotin allows the peptide to bind to streptavidin beads and fluorescein provides a fluorescent readout. When the peptide library was mixed with the stem peptide in PBS (with BSA), some peptide was expected to be bound to the beads, presumably through iminoboronate formation. The imine intermediate was then reduced by sodium boron hydride to yield a permanent bond. Excess stem peptide was then washed away by PBS. The resin was resuspended in PBS with 0.1% BSA. Streptavidin-coated magnetic beads were added

into the mixture and attracted to the peptide beads by biotin. Next, a magnet outside of the falcon tube was applied to attract the magnetic beads. We then built a chain that is comprised of peptide beads-stem peptide-magnetic beads-magnet. Peptide beads that did not have magnetic beads surrounded would eventually be washed away. This first step of screening eliminated peptides that did not recruit any stem peptide. Interestingly, a few thousands of beads survived after this step. It is unrealistic to analyze all of them on LC/MS/MS due to time limit. These beads then went through a second step of screening under microscope. Only the brightest beads were considered to carry potential hits. Detailed screening procedure is described at the end of this chapter.

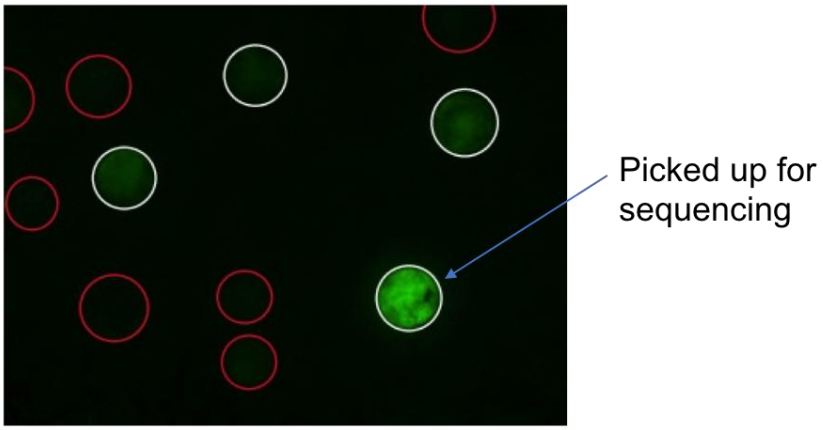


Figure 2.3.9. Representative example of picked up beads under microscope. Only the brightest bead on the same slide was picked up for sequencing

2.3.4 Sequencing result

Peptides from the selected beads were cleaved off resin under UV light in methanol. The sequencing result is summarized in **Table 2.3.2.** To our delight, the majority of the peptides were successfully sequenced. The success rate was 76% (22 out of 29), which was significantly higher than that from quality control (55%, 11 out of 20). We believe the screening process enriched peptides that contain the full

sequence and filtered out the blank beads.

Table 2.3.2. Sequencing result complied

	Sequence	1	2	3	4	5	6
Hit 25	HIDPRR	H	I	D	P	R	R
Hit 18	HQYYYR	H	Q	Y	Y	Y	R
Hit 11	HRHDYR	H	R	H	D	Y	R
Hit 22	HYQRRY	H	Y	Q	R	R	Y
Hit 1	RHYRYQ	R	H	Y	R	Y	Q
Hit 9	RIIHPI	R	I	I	H	P	I
Hit 13	RQHYPQ	R	Q	H	Y	P	Q
Hit 12	RRIYRQ	R	R	I	Y	R	Q
Hit 5	RRRYHR	R	R	R	Y	H	R
Hit 24	IHHRQY	I	H	H	R	Q	Y
Hit 26	IHQYPH	I	H	Q	Y	P	H
Hit 21	IRYHYP	I	R	Y	H	Y	P
Hit 20	YDDDDP	Y	D	D	D	D	P
Hit 15	YIDHQH	Y	I	D	H	Q	H
Hit 16	YRYPPR	Y	R	Y	P	P	R
Hit 29	YIDIQP	Y	I	D	I	Q	P
Hit 14	YHIDQP	Y	H	I	D	D	P
Hit 17	YRIRDI	Y	R	I	R	D	I
Hit 3	PHPRQQ	P	H	P	R	Q	Q
Hit 23	PHRRQI	P	H	R	R	Q	I
Hit 4	PRHYHH	P	R	H	Y	H	H
Hit 28	PIPHRQ	P	I	P	H	R	Q

2.3.5 Hit validation

Unfortunately, we did not observe any repeating sequence partially due to the drawback of this screening platform. It is currently not easy to carry out multiple rounds of screening to achieve better convergence. We decided to pick up some representative sequences and re-synthesize the peptides for hit validation.

2.3.5.1 MIC assay

Our goal for this project is to develop novel antibiotics that target the lipid II stem peptide. One of the most attractive perspectives of the hits is their ability to kill or inhibit the growth of bacteria. We then measured the MIC value of two peptides, Hit 1 and Hit 21, on *S. aureus*. Unfortunately, we did not observe any inhibitory effect from these peptides even at 10 μM. In comparison, Vancomycin efficiently inhibited cell growth at ~ 0.3 μM.

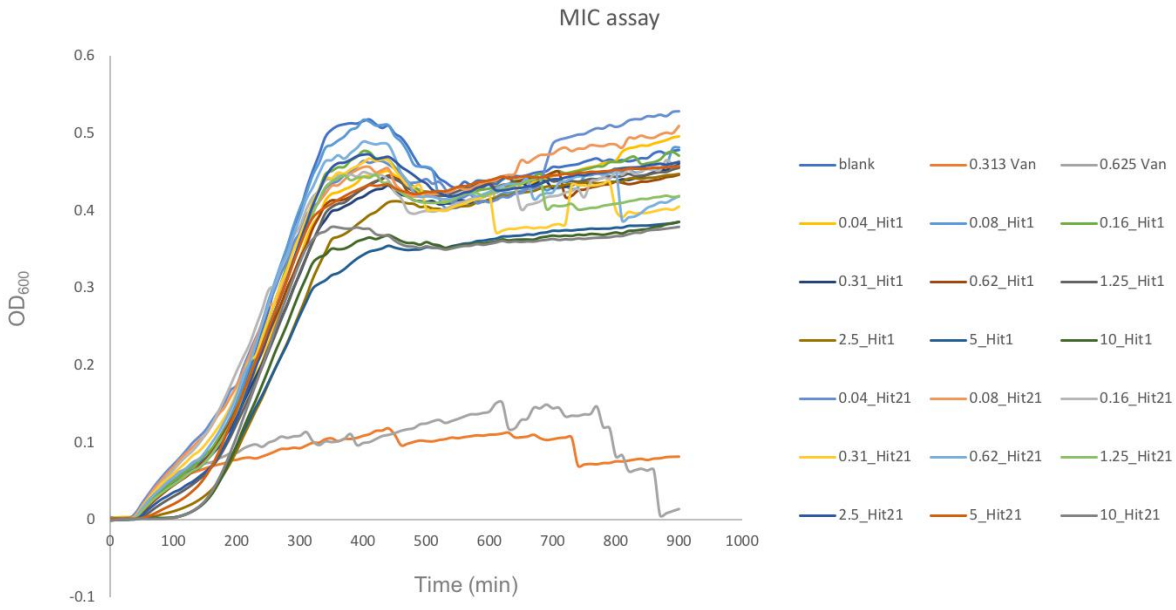


Figure 2.3.10. MIC plots of Vancomycin, Hit 1 and Hit 21 against *S. aureus*. The cells were diluted to 5 $\times 10^5$ cfu/mL in LB media before the addition of peptides. Peptide concentration ranges from 0.04 μM to 10 μM. Vancomycin was used at 0.313 μM and 0.625 μM as a positive control

2.3.5.2 Titration against lipid II

Because cell survival was not part of our criteria during our screening process, it is possible that some peptide hits might actually bind to the stem peptide but did not show any killing capability. However, measuring the dissociation constant between two peptides can be challenging. Our previous design was to install the stem peptide onto a streptavidin coated beads and determine the K_d through anisotropy. A major flaw for this strategy is the interference from streptavidin protein. Lysine residues from this protein can provide extra interaction with the hit peptide through iminoboronate therefore biasing the results. As is discussed earlier, the interaction of 2-APBA (warhead of AB3) and amine to form iminoboronate can be reflected by the change in UV absorbance. Although this assay is only suitable to measure K_d in millimolar range, it can be used as a preliminary method to decide if there is any iminoboronate formation.

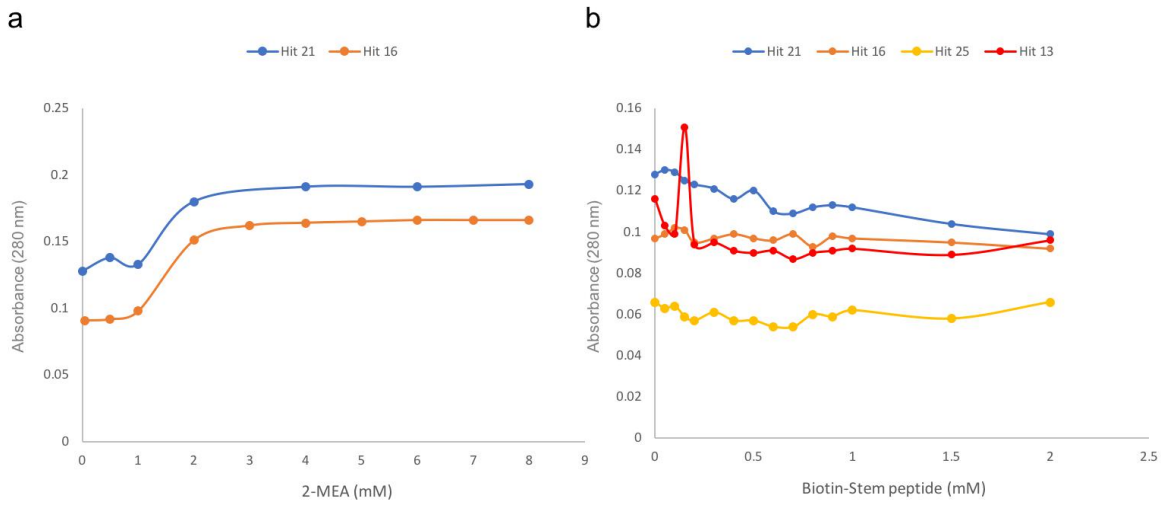


Figure 2.3.11. UV-vis titration of hit peptides against a) 2-MEA and b) biotin-stem peptide. The titration was carried out in the same way as was described previously. The absorbance of the solution at 280 nm was plotted against the concentration of 2-MEA or stem peptide

To validate the feasibility of this assay, we first carried out a UV-vis based titration with a simple compound containing an amino group. We dissolved two potential hits, Hit 16 and Hit 21, separately in PBS buffer at 50 μM, pH 7.4. 2-methoxyethylamine (2-MEA) dissolved in stock was titrated into the solution. The increase of absorbance at 280 nm was recorded and plotted against the concentration of 2-MEA. From **Figure 2.3.11a** we can conclude that AB3 in both peptides can interact with 2-MEA to form iminoboronate. The estimated K_d (~ 2 mM) was slightly lower than simple 2-APBA (~10 mM) probably due to the additional amino acid residues from these peptides. Nevertheless, this experiment agrees with our expectation that the affinity between peptides and amine-containing compounds can be determined by a UV-vis based experiment. We then carried out the same titration experiment with four hits against biotin-stem peptide. Unfortunately, we did not observe a significant increase of absorbance at 280 nm for any of the hits (**Figure 2.3.11b**). The K_d between these four peptides and the stem peptide is likely to be

greater than the highest concentration we have titrated (2 mM). Although these peptides could interact with the stem peptides through other driving forces such as hydrogen binding, we decided not to design new assays to validate this possibility. Our original purpose is to develop a platform to find novel peptides that target lipid II stem peptide through covalent interaction and these preliminary results are sufficient to prove that they are false positives.

2.3.6 Summary

I described in this chapter about our effort to develop a screening technique to discover peptides that targets bacteria through reversible covalent interaction. We have successfully constructed a cyclic peptide library on resin that includes an unnatural amino acid. This library can go through two rounds of selection (affinity based and fluorescent based) to maximize our chance to find the best candidate. Unfortunately, none of the peptides in the hit validation step showed strong binding affinity to our target. Our current failure can be attributed to a few factors. Firstly, the stem peptide is structurally simple so that it does not provide enough functional groups to participate in binding. Also, the peptide library was relatively small compared to other screening techniques such a phage display and mRNA display. A more diverse peptide library can be a solution to this problem. Other pitfalls for our current design include inefficient beads selection under microscope. The criteria for the brightest beads on the same slide is too subtle and the beads were selected based on subjective decisions. A flow cytometer that can sort these fluorescent beads will be a more reasonable option to determine the ‘real hits’. Overall, although the results from hit validation were discouraging, we can still take some useful information out of this experience and we certainly got improved in our experimental skills and troubleshooting abilities.

2.3.7 General information

2.3.7.1 Library synthesis

Tentagel resin (800 mg) preloaded with glycine was swelled in 8 mL DMF for 2 hours. Glycine, photolabile inker and cysteine residues were coupled onto the resin through SPPS. The resin was resuspended in 7 mL of DMF and the first step of 'split and pool' synthesis was to equally split the resin into seven portions. Each portion was individually coupled with one of the seven amino acids shown in **Figure 2.3.4.** for 1 h. The resin was washed 6 times by 2 mL DMF before pooled together. This 'split and pool' procedure was repeated for another two residues. 'Core and shell' coupling was carried out as described previously and three more residues were coupled afterwards by 'split and pool'. After the resin was pooled together, a last cysteine was coupled by standard SPPS. The N-terminus of the library was capped by an acetyl group (1 mL DMF, 0.6 mL DIPEA and 0.5 mL acetic anhydride) before the protecting groups on side chains were globally removed by TFA with 10% water. Peptides on resin were cyclized by 15 mM N-chlorosuccinimide in 5 mL DMF for 15 min. The resin was washed by DMF (3× 8 mL) and DCM (3× 8 mL) and dried for 5 min before storage.

2.3.7.2 Library screening

This screening process was carried out with 350 mg of resin. Based on beads density (~2 million/g, 100 μm) and library diversity (~10^5), we should be able to cover 6-7 repeats of the same sequence. The dry resin was soaked in 20 mL PBS buffer overnight before the screening process. The library was then washed with 20 mL PBS buffer for 3 times and resuspended into 20 mL PBS buffer with 0.1 mg/mL BSA. Dual labelled stem peptide was added to the mixture at a final concentration of 10 μM. The solution was incubated at 37 °C for 1 h while rotating. 38 mg sodium

boron hydride solid was added into the solution to reduce the imine for 15 min. Excess amount of the stem peptide was then washed by PBS (6 × 20 mL). The resin was resuspended in 20 mL PBS with 0.1% BSA. Strepavidin-coated magnetic beads (200 μL) were added into the mixture and incubated for 2 h at 37 °C while rotating. Next, a magnet was applied outside of the falcon tube to attract the magnet beads and the supernatant was discarded. The resin was resuspended in 20 mL PBS (with 0.1 % BSA) and this magnetic pull down process was repeated two more times. The beads that got pulled down by magnet was diluted into 10 mL PBS without BSA. ~10 μL of solution was pipetted on a glass slide so that less than 100 beads will be analyzed at a time (diluted if it was too dense). Only the brightest beads on the same slide will be picked up and considered for sequencing.

2.3.7.3 Sample preparation for sequencing

The beads, either from quality test or screening were resuspended in PBS buffer. Disulfide bonds was reduced on resin by 50 mM TCEP for 30 min and the reduced cysteine residues were capped by 20 mM iodoacetamide in PBS for 1 h. After 6 washes, the beads were diluted into fresh PBS buffer. The amount of buffer was adjusted according to the density of the beads (usually less than 50 beads/10 μL). 10 μ L of PBS was pipetted onto a glass slide each time and analyzed under microscope. The desired beads will be picked into a microcentrifuge tube containing 20 μL of HPLC grade methanol to be cleaved under UV light for 1 h. The solution was then diluted by 100 μL of buffer A from HPLC and filtered through a 0.45 μm PTFE filter. The filter was washed by another 100 μ L buffer A to maximize the recovery of the peptides. Buffer A was then combined and concentrated to ~20 μL on speed-vac. The samples were then transferred into LC/MS vial and submitted to

Aaron Maurais for sequencing.

2.3.7.4 LC/MS/MS sequencing

Sequencing was conducted with the help of Aaron Maurais. MS2 profiles of the peptides were matched to a self-made database that includes all possible sequences. Acetyl group on the N-terminus was specified for glycine; static modifications (+ 57.0214 m/z) were made for cysteine residues and C-terminus (-0.984 m/z) accounting for the IA alkylation and C-terminal amide.

2.3.7.5 MIC assay

MIC values of the hits against *S. aureus* (ATC 6538) were determined by the standard microdilution method. Briefly, Sub-cultured *S. aureus* was let grown until OD_{600} ~ 0.6 and diluted to ~ 5×10^5 cfu/mL in LB. Each well in a sterile 96 well plate was added 200 μL of the cell suspension. Peptides or Vancomycin dissolved in DMSO were added into each well to the indicated concentration. Each concentration was tested on triplicates. The plate was then placed on a plate reader and OD_{600} was monitored overnight with 10 min interval and 15 seconds of shaking before reading.

2.3.8 References

(1) Neu, H. C. The Crisis in Antibiotic Resistance. *Science (80-.).* **1992**, *257* (5073), 1064 LP – 1073. https://doi.org/10.1126/science.257.5073.1064.

(2) Davies, J. Origins and Evolution of Antibiotic Resistance. *Microbiologia* **1996**, *12* (1), 9–16. https://doi.org/10.1128/mmbr.00016-10.

(3) Gould, I. M. A Review of the Role of Antibiotic Policies in the Control of Antibiotic Resistance. *J. Antimicrob. Chemother.* **1999**, *43* (4), 459–465. https://doi.org/10.1093/jac/43.4.459.

(4) Vollmer, W.; Blanot, D.; De Pedro, M. A. Peptidoglycan Structure and Architecture. *FEMS Microbiol. Rev.* **2008**, *32* (2), 149–167. https://doi.org/10.1111/j.1574-6976.2007.00094.x.

(5) Lee, W.; Schaefer, K.; Qiao, Y.; Srisuknimit, V.; Steinmetz, H.; Müller, R.; Kahne, D.; Walker, S. The Mechanism of Action of Lysobactin. *J. Am. Chem. Soc.* **2016**, *138* (1), 100–103. https://doi.org/10.1021/jacs.5b11807.

(6) Huber, J.; Donald, R. G. K.; Lee, S. H.; Jarantow, L. W.; Salvatore, M. J.; Meng, X.; Painter, R.; Onishi, R. H.; Occi, J.; Dorso, K.; et al. Chemical Genetic Identification of Peptidoglycan Inhibitors Potentiating Carbapenem Activity against Methicillin-Resistant Staphylococcus Aureus. *Chem. Biol.* **2009**, *16* (8), 837–848. https://doi.org/10.1016/j.chembiol.2009.05.012.

(7) Fang, X.; Tiyanont, K.; Zhang, Y.; Wanner, J.; Boger, D.; Walker, S. The Mechanism of Action of Ramoplanin and Enduracidin. *Mol. Biosyst.* **2006**, *2* (1), 69–76. https://doi.org/10.1039/b515328j.

(8) Walsh, C. T.; Wencewicz, T. A. Prospects for New Antibiotics: A Molecule-

Centered Perspective. *J. Antibiot. (Tokyo).* **2014**, *67* (1), 7–22. https://doi.org/10.1038/ja.2013.49.

(9) Stone, K. J.; Strominger, J. L. Mechanism of Action of Bacitracin: Complexation with Metal Ion and C₅₅-Isoprenyl Pyrophosphate. *Proc. Natl. Acad. Sci.* **1971**, *68* (12), 3223 LP – 3227. https://doi.org/10.1073/pnas.68.12.3223.

(10) Swoboda, J. G.; Meredith, T. C.; Campbell, J.; Brown, S.; Suzuki, T.; Bollenbach, T.; Malhowski, A. J.; Kishony, R.; Gilmore, M. S.; Walker, S. Discovery of a Small Molecule That Blocks Wall Teichoic Acid Biosynthesis in Staphylococcus Aureus. *ACS Chem. Biol.* **2009**, *4* (10), 875–883. https://doi.org/10.1021/cb900151k.

(11) Lee, K.; Campbell, J.; Swoboda, J. G.; Cuny, G. D.; Walker, S. Development of Improved Inhibitors of Wall Teichoic Acid Biosynthesis with Potent Activity against Staphylococcus Aureus. *Bioorganic Med. Chem. Lett.* **2010**, *20* (5), 1767–1770. https://doi.org/10.1016/j.bmcl.2010.01.036.

(12) Fisher, J. F.; Meroueh, S. O.; Mobashery, S. Bacterial Resistance to β-Lactam Antibiotics: Compelling Opportunism, Compelling Opportunity. *Chem. Rev.* **2005**, *105* (2), 395–424. https://doi.org/10.1021/cr030102i.

(13) McComas, C. C.; Crowley, B. M.; Boger, D. L. Partitioning the Loss in Vancomycin Binding Affinity for D-Ala-D-Lac into Lost H-Bond and Repulsive Lone Pair Contributions. *J. Am. Chem. Soc.* **2003**, *125* (31), 9314–9315. https://doi.org/10.1021/ja035901x.

(14) Okano, A.; Nakayama, A.; Schammel, A. W.; Boger, D. L. Total Synthesis of [Ψ[C(=NH)NH]Tpg4]Vancomycin and Its (4-Chlorobiphenyl)Methyl

Derivative: Impact of Peripheral Modifications on Vancomycin Analogues Redesigned for Dual d-Ala-d-Ala and d-Ala-d-Lac Binding. *J. Am. Chem. Soc.* **2014**, *136* (39), 13522–13525. https://doi.org/10.1021/ja507009a.

(15) Münch, D.; Sahl, H. G. Structural Variations of the Cell Wall Precursor Lipid II in Gram-Positive Bacteria - Impact on Binding and Efficacy of Antimicrobial Peptides. *Biochim. Biophys. Acta - Biomembr.* **2015**, *1848* (11), 3062–3071. https://doi.org/10.1016/j.bbamem.2015.04.014.

(16) Gray, B. P.; Brown, K. C. Combinatorial Peptide Libraries: Mining for Cell-Binding Peptides. *Chem. Rev.* **2014**, *114* (2), 1020–1081. https://doi.org/10.1021/cr400166n.

(17) Smith, G. P. Filamentous Fusion Phage: Novel Expression Vectors That Display Cloned Antigens on the Virion Surface. *Science (80-.).* **1985**, *228* (4705), 1315–1317. https://doi.org/10.1126/science.4001944.

(18) McCarthy, K. A.; Kelly, M. A.; Li, K.; Cambray, S.; Hosseini, A. S.; Van Opijnen, T.; Gao, J. Phage Display of Dynamic Covalent Binding Motifs Enables Facile Development of Targeted Antibiotics. *J. Am. Chem. Soc.* **2018**, *140* (19), 6137–6145. https://doi.org/10.1021/jacs.8b02461.

(19) Owens, A. E.; Iannuzzelli, J. A.; Gu, Y.; Fasan, R. MOrPH-PhD: An Integrated Phage Display Platform for the Discovery of Functional Genetically Encoded Peptide Macrocycles. *ACS Cent. Sci.* **2020**. https://doi.org/10.1021/acscentsci.9b00927.

(20) Oller-Salvia, B.; Chin, J. W. Efficient Phage Display with Multiple Distinct Non-Canonical Amino Acids Using Orthogonal Ribosome-Mediated Genetic Code Expansion. *Angew. Chemie - Int. Ed.* **2019**, *58* (32), 10844–10848.

https://doi.org/10.1002/anie.201902658.

(21) Tian, F.; Tsao, M. L.; Schultz, P. G. A Phage Display System with Unnatural Amino Acids. *J. Am. Chem. Soc.* **2004**, *126* (49), 15962–15963. https://doi.org/10.1021/ja045673m.

(22) Charbit, A.; Boulain, J. C.; Ryter, A.; Hofnung, M. Probing the Topology of a Bacterial Membrane Protein by Genetic Insertion of a Foreign Epitope; Expression at the Cell Surface. *EMBO J.* **1986**, *5* (11), 3029–3037. https://doi.org/10.1002/j.1460-2075.1986.tb04602.x.

(23) Freudl, R.; MacIntyre, S.; Degen, M.; Henning, U. Cell Surface Exposure of the Outer Membrane Protein OmpA of Escherichia Coli K-12. *J. Mol. Biol.* **1986**, *188* (3), 491–494. https://doi.org/10.1016/0022-2836(86)90171-3.

(24) Lee, J. S.; Shin, K. S.; Pan, J. G.; Kim, C. J. Surface-Displayed Viral Antigens on Salmonella Carrier Vaccine. *Nat. Biotechnol.* **2000**, *18* (6), 645–648. https://doi.org/10.1038/76494.

(25) Lu, Z.; Murray, K. S.; Cleave, V. Van; LaVallie, E. R.; Stahl, M. L.; McCoy, J. M. Expression of Thioredoxin Random Peptide Libraries on the Escherichia Coli Cell Surface as Functional Fusions to Flagellin: A System Designed for Exploring Protein-Protein Interactions. *Bio/Technology* **1995**, *13* (4), 366–372. https://doi.org/10.1038/nbt0495-366.

(26) Francisco, J. A.; Campbell, R.; Iverson, B. L.; Georgiou, G. Production and Fluorescence-Activated Cell Sorting of Escherichia Coli Expressing a Functional Antibody Fragment on the External Surface. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90* (22), 10444–10448. https://doi.org/10.1073/pnas.90.22.10444.

(27) Rockberg, J.; Löfblom, J.; Hjelm, B.; Uhlén, M.; Ståhl, S. Epitope Mapping of

Antibodies Using Bacterial Surface Display. *Nat. Methods* **2008**, *5* (12), 1039–1045. https://doi.org/10.1038/nmeth.1272.

(28) Boder, E. T.; Wittrup, K. D. Yeast Surface Display for Screening Combinatorial Polypeptide Libraries. *Nat. Biotechnol.* **1997**, *15* (6), 553–557. https://doi.org/10.1038/nbt0697-553.

(29) Ohuchi, M.; Murakami, H.; Suga, H. The Flexizyme System: A Highly Flexible TRNA Aminoacylation Tool for the Translation Apparatus. *Curr. Opin. Chem. Biol.* **2007**, *11* (5), 537–542. https://doi.org/10.1016/j.cbpa.2007.08.011.

(30) Murakami, H.; Ohta, A.; Ashigai, H.; Suga, H. A Highly Flexible TRNA Acylation Method for Non-Natural Polypeptide Synthesis. *Nat. Methods* **2006**, *3* (5), 357–359. https://doi.org/10.1038/nmeth877.

(31) Lam, K. S.; Lebl, M.; Krchňák, V. The "One-Bead-One-Compound" Combinatorial Library Method. *Chem. Rev.* **1997**, *97* (2), 411–448. https://doi.org/10.1021/cr9600114.

(32) Lam, K. S.; Salmon, S. E.; Hersh, E. M.; Hruby, V. J.; Kazmierski, W. M.; Knapp, R. J. A New Type of Synthetic Peptide Library for Identifying Ligand-Binding Activity. *Nature* **1991**, *354* (6348), 82–84. https://doi.org/10.1038/354082a0.

(33) Youngquist, R. S.; Fuentes, G. R.; Lacey, M. P.; Keough, T. Generation and Screening of Combinatorial Peptide Libraries Designed for Rapid Sequencing by Mass Spectrometry. *J. Am. Chem. Soc.* **1995**, *117* (14), 3900–3906. https://doi.org/10.1021/ja00119a002.

(34) Wang, P.; Arabaci, G.; Pei, D. Rapid Sequencing of Library-Derived Peptides

by Partial Edman Degradation and Mass Spectrometry. *J. Comb. Chem.* **2001**, *3* (3), 251–254. https://doi.org/10.1021/cc000102l.

(35) Webb-Robertson, B. J. M.; Cannon, W. R. Current Trends in Computational Inference from Mass Spectrometry-Based Proteomics. *Brief. Bioinform.* **2007**, *8* (5), 304–317. https://doi.org/10.1093/bib/bbm023.

(36) Lian, W.; Upadhyaya, P.; Rhodes, C. A.; Liu, Y.; Pei, D. Screening Bicyclic Peptide Libraries for Protein-Protein Interaction Inhibitors: Discovery of a Tumor Necrosis Factor-α Antagonist. *J. Am. Chem. Soc.* **2013**, *135* (32), 11990–11995. https://doi.org/10.1021/ja405106u.

(37) Wang, X.; Peng, L.; Liu, R.; Xu, B.; Lam, K. S. Applications of Topologically Segregated Bilayer Beads in "one-Bead One-Compound" Combinatorial Libraries. *J. Pept. Res.* **2005**, *65* (1), 130–138. https://doi.org/10.1111/j.1399-3011.2005.00192.x.

(38) Needels, M. C.; Jones, D. G.; Tate, E. H.; Heinkel, G. L.; Kochersperger, L. M.; Dower, W. J.; Barrett, R. W.; Gallop, M. A. Generation and Screening of an Oligonucleotide-Encoded Synthetic Peptide Library. *Proc. Natl. Acad. Sci. U. S. A.* **1993**, *90* (22), 10700–10704. https://doi.org/10.1073/pnas.90.22.10700.

(39) Postma, T. M.; Albericio, F. N-Chlorosuccinimide, an Efficient Reagent for on-Resin Disulfide Formation in Solid-Phase Peptide Synthesis. *Org. Lett.* **2013**, *15* (3), 616–619. https://doi.org/10.1021/ol303428d.

Chapter 3 Site-specific modifications on proteins

3.1 Introduction

3.1.1 Protein modification

Post-translational modifications (PTMs) on proteins is a widespread process in nature. These modifications not only increase their structural diversity, but also control their activity as well as metabolism.[1] For instance, phosphorylation, which is the most abundant PTM is a common mechanism of regulating enzyme acticities;[2] Ubiquitination, another common PTM, involves a ubiquitin protein being attached to another protein, leading to protein degradation.[3] **Figure 3.1.1** lists some of the common naturally occurring modifications reported in the literature.

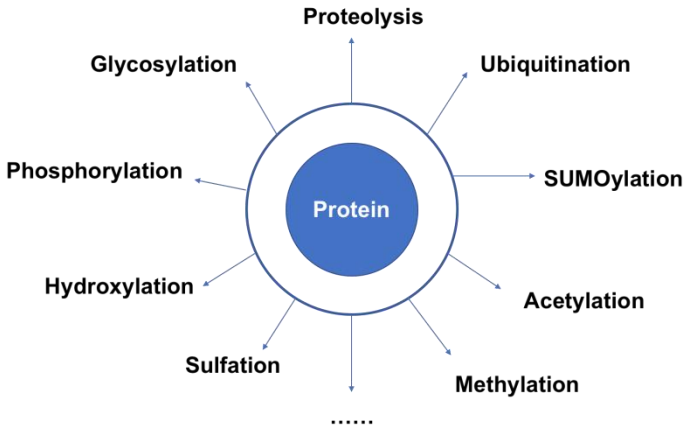
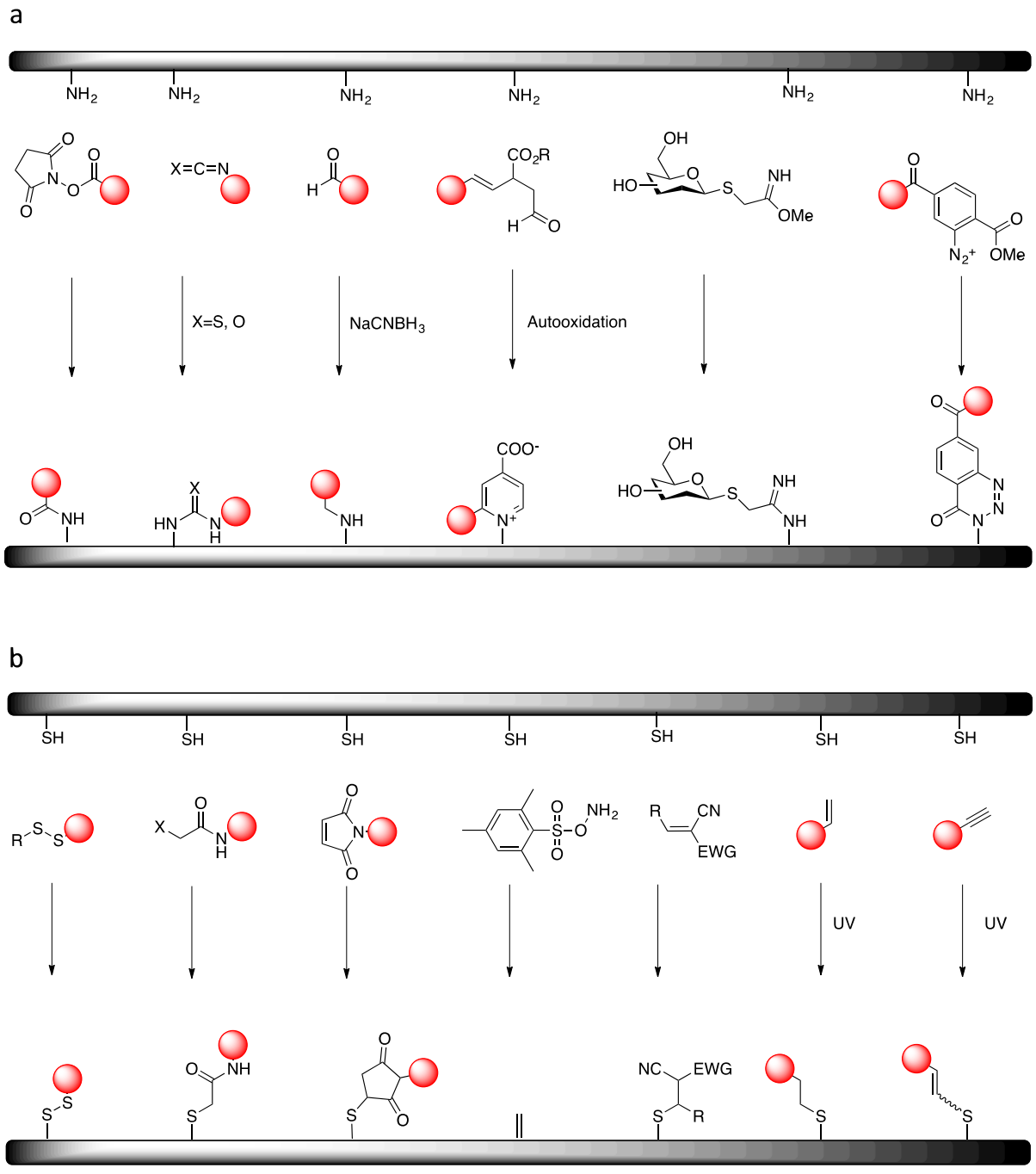


Figure 3.1.1. illustration of some naturally occurring modifications on proteins

In some cases, we would like to install functions that are not commonly found in nature. For example, biotinylation is often used for protein immobilization or pull down;[4] fluorescein succinimidyl ester (FAM-Osu) and FAM-maleimide are often used to label lysine or cysteine residues in a protein;[5] PEGlytion can reduce the immunogenicity of a protein and increase its solubility;[6] Protein-protein conjugate is frequently used in ELISA essays.[7] Ideally, these reactions should have efficient conversions in neutral aqueous conditions to avoid disturbance of the structure and

activity of the protein. Also, ideal reactions should be relatively fast to avoid long incubation time that might lead to the loss of protein function. Researchers have always been interested in developing chemistries that can fulfill these requirements and some representative work will be discussed in the following chapters.

3.1.2 Modifications on natural amino acids



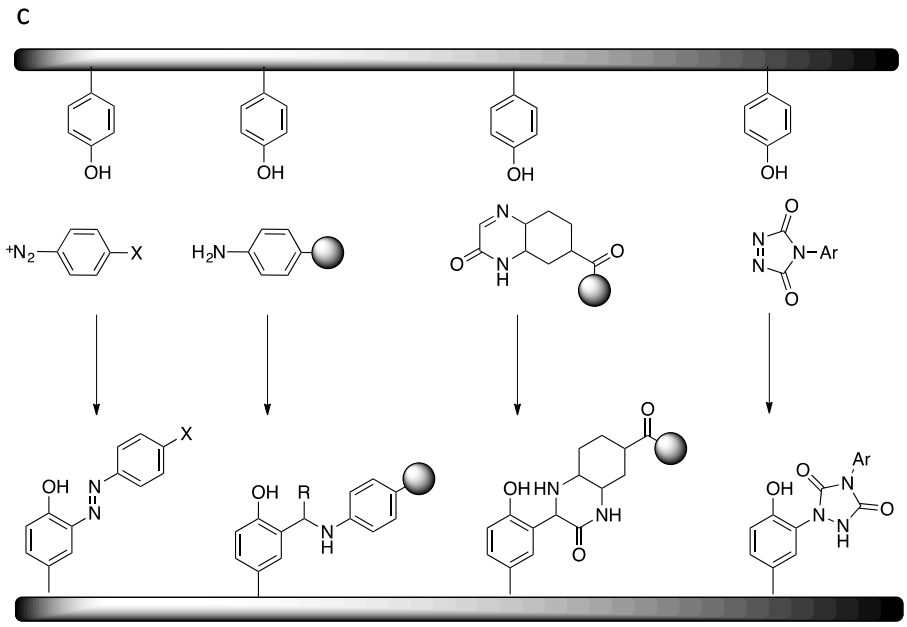


Figure 3.1.2. Selected examples of popular chemistries for protein or peptide on a) lysine; b) cysteine and C) tyrosine

There are 20 standard natural amino acids (not including newly discovered amino acid such as selenocysteine), each with a unique functional group on the side chain, serving as different handles for modification purposes. Some of the popular choices including lysine, cysteine and tyrosine will be discussed in detail as representative examples.

Conventional methods for lysine modification include acylation by succinimidyl ester, isothiocyanate/isocyanate and reductive amination. Recently, Katsumura's group reported the irreversible modification of lysine residues though aza-electrocyclization.[8] Other novel methods includes 2-imno-2-methoxyethyl reagents to form amide bond[9] or diazonium terephthalates to form a stable triazin-4(3H)-one ring.[10] However, the position and the number of lysine residues being modified are not easily predictable or controlled in these reactions and some of them suffer from incomplete conversion.

Cysteine has always been a popular choice due to its low abundance and higher nucleophilicity compared to other residues such as lysine and histidine,[11] making it easier to generate homogenous conjugates. Cysteine can easily undergo disulfide exchange or get alkylated by electrophiles such as α-halocabonyls[12] and Michael acceptors[13]. Thiol-ene[14] or thiol-yne[15] click reactions with alkenes or alkynes are also popular reactions. Sometimes, cysteine is converted to dehydroalanine by oxidative or bis-alkylation elimination and the newly formed amino acid can be used as a handle for later modification such as Michael addition by thiol reagents.[16] Some of these modifications are reversible including Michael addition and arguably disulfide bond formation while others are usually considered irreversible. When a protein has a free and exposed cysteine that is not critical to its function, these reactions can be the best choices to install chemical modifications without the need for protein engineering.

Francis's[17] and Barbas'[18] lab are doing great work in developing chemoselective reactions on tyrosine residues. They modified the ortho position of tyrosine using diazonium salts with various electron withdrawing groups at the para position. Mannich-type reactions on tyrosine with aldehydes and anilines are also explored and applied for protein modification by Francis's lab,[19] although the reactions are somewhat sluggish in aqueous solution. Fujie Tanaka and coworkers extended this work utilizing preformed imines.[20] Recently, Barbas and coworkers developed a novel ene-type click-like reaction for tyrosine using azomaleimides. [21]

Besides the three popular choices mentioned above, other amino acids such as aspartic acid and histidine also attracted some attention, though they will not be discussed further here. The difficulty for chemical modification on natural amino

acids lies in the specificity of the reactions. Although the chemistries mentioned above can be 'amino acid specific', the presence of multiple copies of the same amino acid usually leads to undesired reactions. One way to get around this problem is to incorporate unnatural amino acids with biorthogonal handles into a protein of interest.

3.1.3 Modifications on unnatural amino acids

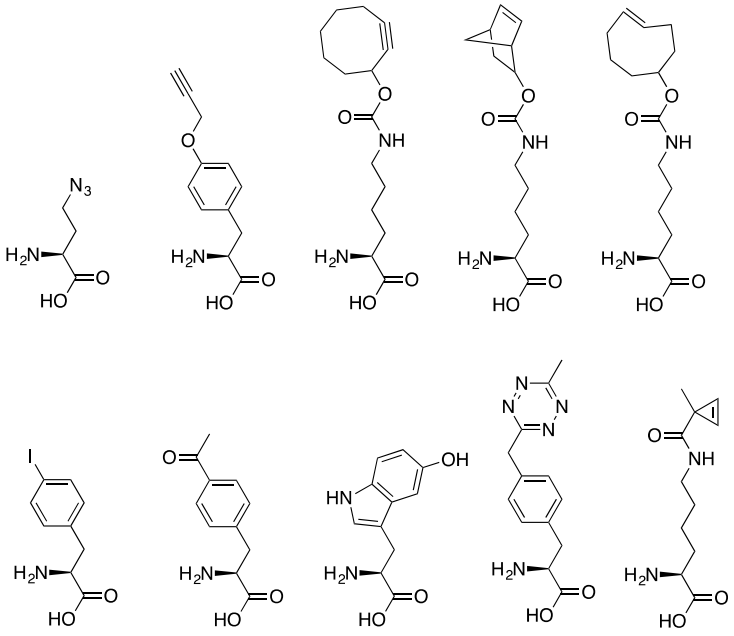


Figure 3.1.3. Examples of different types of unnatural amino acids that have been incorporated into proteins

Unnatural amino acids can carry more diverse functional groups and provide unique handles for chemical modification. A lot of work has already been done to incorporate those amino acids into proteins through the reassignment of codons such as amber codon suppression. The amino acids shown in **Figure 3.1.3** are some representative examples.[22–30] Although this field is rapidly growing, the limitation remains. Incorporation of unnatural amino acid requires access to the plasmids and the required tRNA/aminoacyl-tRNA synthetase (aaRS) pair may not be applicable to

all protein systems. For example, the tRNA/aaRS developed from *E. coli* might not be applicable in mammalian cells. However, with the current effort from researchers, I believe this technology will eventually become readily available and easy to handle.

3.1.4 Site specific modifications on natural amino acids

The reactions mentioned above are only chemoselective to a specific functional group. However, when there are multiple copies of the same amino acid in the protein, site-specificity is difficult to achieve and instead a heterogenous mixture is obtained. An ideal reaction should happen on a single amino acid and proceed to near complete conversion to create a homogenous conjugate. This can be achieved by changing the local environment of the amino acid that needs to be modified. There are multiple examples of site-specific modifications on a single site in a rationally designed sequence or local environment. Some representative examples are discussed below.

3.1.4.1 Site-specific modification on cysteine

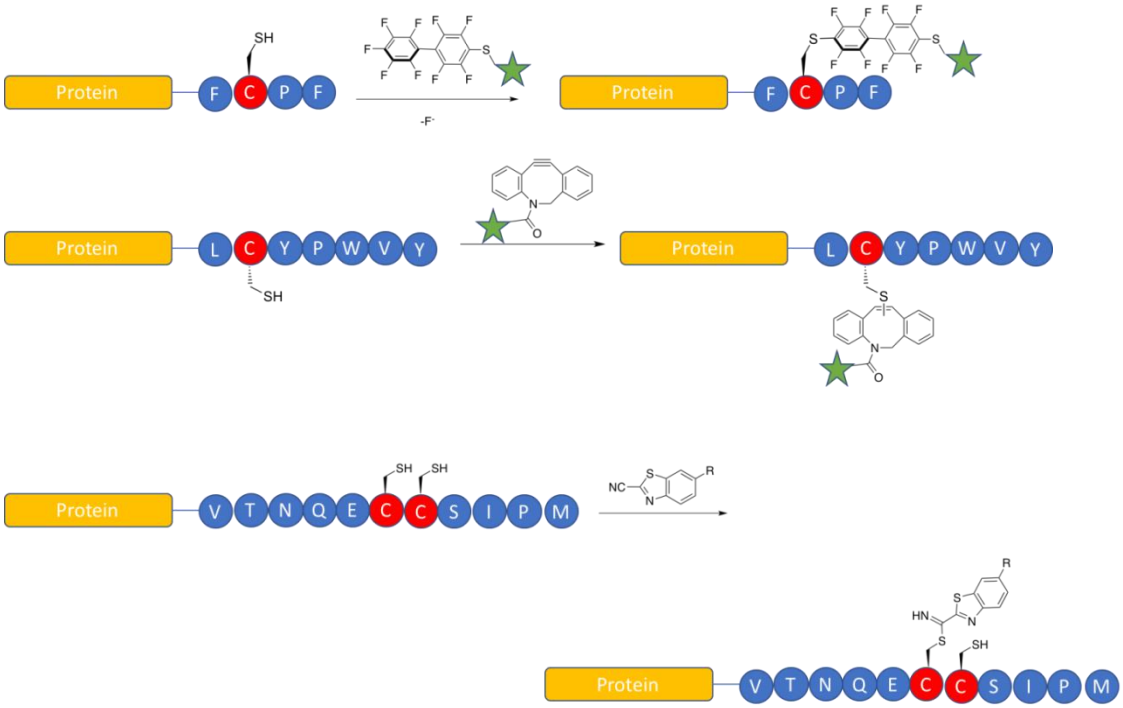


Figure 3.1.4. Site specific modification on cysteine residues through π-clamp[31] or thiol-yne click chemistry[32] or thioimidate formation[33]

In these examples, cysteine gains selectivity through accelerated reaction kinetics. The reactions between perfluoroaromatic reagents or alkynes are generally sluggish in aqueous solution. In the first example, that cysteine residue exists in a 'π-clamp', which positions the thiol group in an optimal position for reaction and lowers the activation energy by ~ 3 kcal/mol. In the second example, thiol-yne reaction rate was increased by over 200 folds, probably due to a lower pKa. Finally, the third example, reported by Qing's group utilizes thioimidate formation. Before the side reaction was discovered,[34] 2-cyanobenzothiazole (CBT) was known to react selectively with N-terminal cysteine residues. The first step of reaction between cysteine and CBT is the attack of thiol to the C-N triple bond to form a thioimidate, which is known to be reversible. However, Qing's group found a cysteine residue in the sequence 'VTNQECCSIPM' that formed a stable thioimidate product without the need for 1,2-aminothiol functionality.[33] It is suspected that the surrounding residues stabilize the product through non-covalent interactions.

3.1.4.2 Site-specific modification on lysine

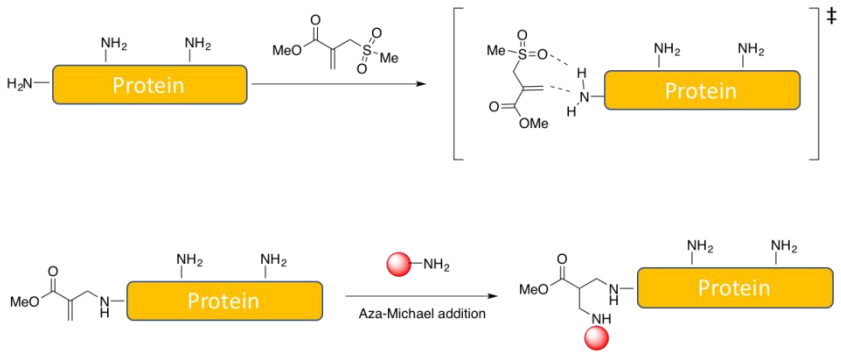


Figure 3.1.5. Lysine residue with the lowest pKa value is site specifically modified by sulfonyl acrylate

Although there are abundant chemistries on lysine residues, only few cases where site-specific modifications are realized. In one example, Goncalo's group designed a sulfonyl acrylate reagent that can regioselectively react with the lysine residue with the lowest pKa value in a protein.[35] This reaction can proceed to completion under mild aqueous condition. However, the regioselectivity is achieved solely based on the difference of pKa values from lysine residues. One can assume that if sulfonyl acrylate is added in excess amount, or multiple lysine residues have a similar pKa value, site-specificity will not be easily achieved by this reaction.

3.1.4.3 Site-specific modification on histidine

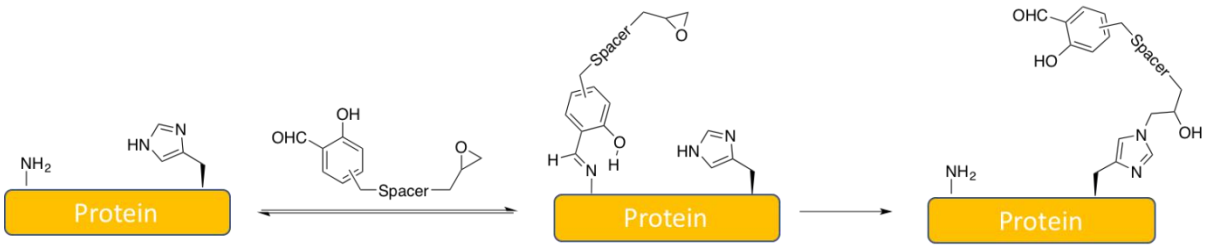


Figure 3.1.6. Linchpin directed modification modifies histidine residues

Rai's group designed a bifunctional group to target histidine site specifically through 'linchpin directed modification (LDM)'.[36] As is illustrated above, the salicylaldehyde group on one side of the molecule reacts rapidly and reversibly with an amino group on a lysine residue, bringing the epoxide group to a nearby histidine. The reaction between epoxide and a free histidine is generally sluggish in aqueous solution. However, LDM would increase the local concentration of epoxide that is in close proximity to a specific histidine so that the reaction rate will be greatly accelerated. The different kinetics of LDM compared to undirected reaction would reduce background reactions and make site-specific modification feasible. Unfortunately, this method may not be generally applicable to most proteins. It

requires a lysine and histidine residue to be in the best location to each other in order to proceed. Furthermore, even for proteins that showed reaction, the reported yield was only ~30%.

3.1.5 Site specific modification on N-terminal residues

The N-terminal residue occupies a unique position in a protein and provides an attractive option for site-specific modification. The α-amine from an amino acid generally has a lower pKa than the side chains from lysine, arginine and tryptophan because of the inductive effect of the nearby carbonyl group, making it possible to be selectively modified.[37] Modification on the N-terminus also causes less disturbance to the structure and the activity of a protein.

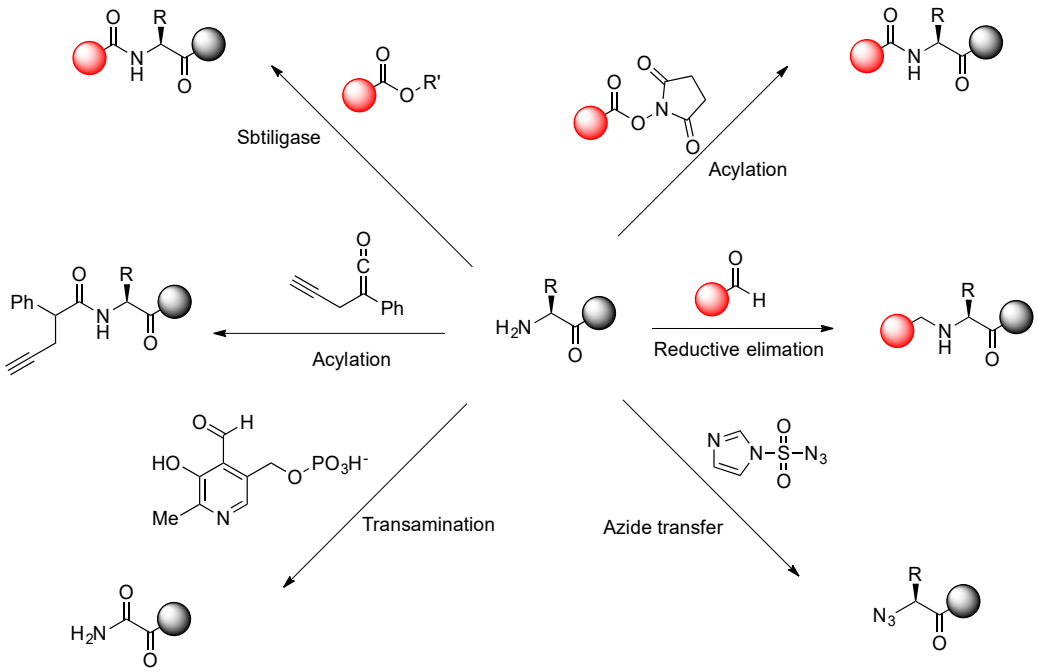


Figure 3.1.7. Chemistries for N-terminal amine modification

Different chemistries have been developed to site specifically modify the N-terminus of a protein. The most common strategy utilizes an activated ester[38] or ketene[39] to acylate the N-terminal amine; aldehyde can react with amines through

reductive amination.[40] It is also possible to convert the N-terminal amine group into an azide through azide transfer[41] followed by an biorthogonal click chemistry. However, utilizing the lower pKa value of the terminal amino group has its limitation. Some lysine residues might have a lower pKa than usual, leading to undesired side reactions.[35] Although there is an extreme example where all the lysine residues were deleted to achieve site specificity on N-terminus,[42] it is not applicable to other proteins because lysine residues are essential for protein structure and function most of the time.

The N-terminal amino group can be converted into a ketone through transamination.[43] Pyridoxal-5′-phosphate reacts with α-amines to form imines followed by hydrolysis to yield an aldehyde or ketone group. This functional group can then react with oxime chemo selectively. However, transamination and oxime ligation both require a long reaction time (~18 hrs).

There are also some examples where N-terminus modification is realized by an enzyme. For example, subtiligase can react with a glycolate ester substrate to form a thioester. This intermediate will react with the N-terminus through acyl transfer to generate a peptide bond. [44] Another popular choice is using a LPXTG tag which can be recognized by SrtA to modify the N-terminus through transpeptidation.[45]

It is worth noting that Francis's lab has developed a novel method using 2-pyridinecarboxyaldehydes to label the N-terminus of protein or peptides.[46] The reaction starts with imine formation followed by a nucleophilic attack of the amide nitrogen from the peptide backbone. However, this reaction requires millimolar concentration of the labelling reagent and 8-16 hrs of incubation time.

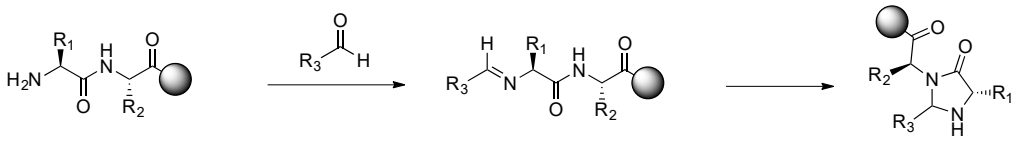


Figure 3.1.8. N-terminal modification utilizing the first and second residue with 2-pyridinecarboxyaldehydes

3.1.6 Site specific modification on N-terminal cysteine residue

Utilizing the side chain functional group together with the N-terminal amine of a peptide/protein is always a good option. This idea has been employed in multiple applications where N-terminal serine/threonine[47–51] or tryptophan[52] are site specifically modified, as is illustrated below. However, the reaction with serine/threonine requires use of a strong oxidant-$NaIO_4$ and the Pictet-Spengler reaction is generally conducted in acetic acid solution. These requirements limit the applications of the chemistries mentioned above.

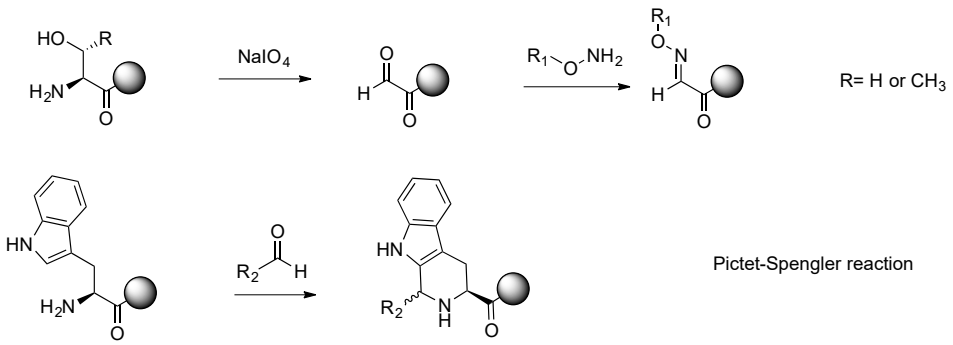


Figure 3.1.9. N-terminal serine/threonine and tryptophan modification

Among all the N-terminal choices, cysteine is the most popularly investigated, as well as utilized amino acid because of its unique aminothiol structure. The most well-known and utilized reactions are native chemical ligation (NCL), thiazolidine formation and CBT condensation.

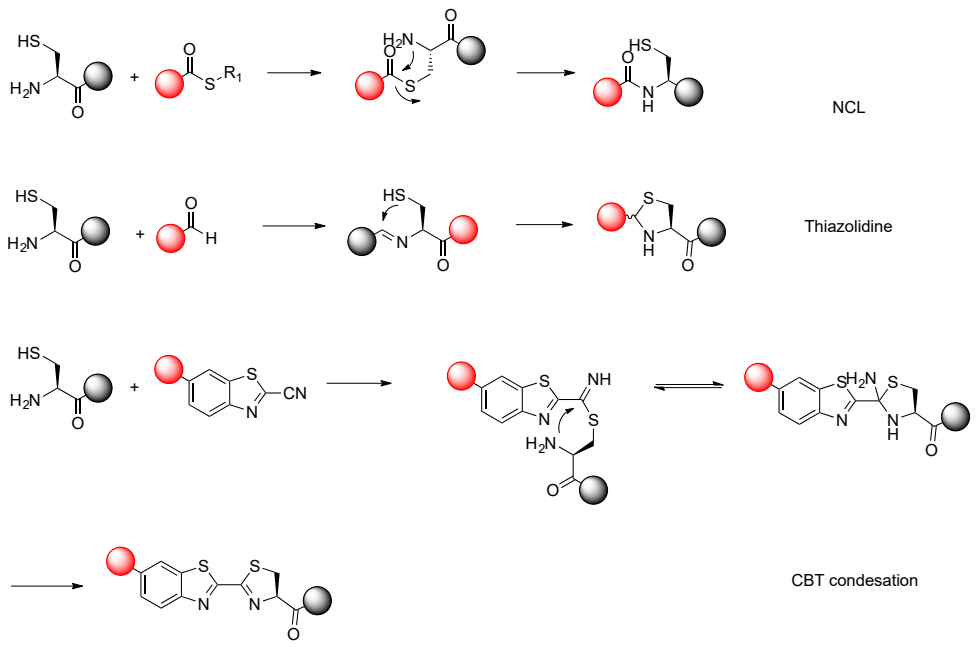


Figure 3.1.10. N-terminal cysteine modification

NCL is being widely used for chemical synthesis or semi-synthesis of proteins,[53,54] as well as protein labeling.[55] The reaction starts with the attack from the thiol group of an N-terminal cysteine residue to the C-terminal thioester. This thiol-exchange step is reversible and newly formed thioester intermediate rearranges by an intramolecular S→N acyl transfer to form a peptide bond. The typical condition for native chemical ligation requires high concentration of 4-mercaptophenylacetic acid to accelerate the thiol-exchange step and at the same time, reduce non-specific reactions. This reaction usually takes a few hours or overnight to finish depending on the neighboring amino acid. One side reaction that our lab has discovered is the non-specific thiol-exchange between the thioester and internal cysteine, although the newly formed thioester can be eventually hydrolyzed.

Aldehydes are known to react with aminothiol to form thiazolidine.[56] The reaction starts with an imine formation between the N-terminal amine and the aldehyde. The thiol group from the cysteine then attacks the imine structure to give a

five membered ring called thiazolidine. This reaction is catalyzed in a slightly acidic aqueous condition and generally takes a few hours to complete. The product is labile in acidic conditions so it is not a reliable method to generate stable conjugates.

The condensation between a N-terminal cysteine and CBT was initially found to be the final step of D-luciferin synthesis[57] and it has already been used to generate protein conjugates.[58–60] The reaction begins by the attack of thiol from cysteine to the cyano group to generate a thioimidate intermediate. N-terminal amine attacks the thioimidate and gives a five membered ring (thiazolidine); leaving of the amine group yields the final luciferin product. This reaction proceeds to completion in neutral aqueous conditions and the product is stable, making it practical for many bioconjugate applications. However, our lab has recently discovered a side reaction with CBT where it conjugates with internal cysteine residues. Although the thioimidate product can potentially hydrolyze, Qing's work suggested that this intermediate can be quite stable in some cases. Importantly, our lab has recently reported a N,S-double labeling of N-terminal cysteine with CBT through an alternative pathway.[34] These observations raise questions about its site-specificity.

The reactions mentioned in this introduction expanded our 'toolkit' to introduce chemical modifications on proteins. Although some of these chemistries discussed here are unique and widely used, there are also limitations associated with them. There is always a need for additional site-specific reactions with improved selectivity and kinetics. This chapter of my thesis will include the effort that our lab has contributed in this area.

3.1.7 References

(1) Spicer, C. D.; Davis, B. G. Selective Chemical Protein Modification. *Nat. Commun.* **2014**, *5*, 1–14. https://doi.org/10.1038/ncomms5740.

(2) Khoury, G. A.; Baliban, R. C.; Floudas, C. A. Proteome-Wide Post-Translational Modification Statistics: Frequency Analysis and Curation of the Swiss-Prot Database. *Sci. Rep.* **2011**, *1*, 1–5. https://doi.org/10.1038/srep00090.

(3) Glickman, M. H.; Ciechanover, A. The Ubiquitin-Proteasome Proteolytic Pathway: Destruction for the Sake of Construction. *Physiol. Rev.* **2002**, *82* (2), 373–428. https://doi.org/10.1152/physrev.00027.2001.

(4) Agard, N. J.; Prescher, J. A.; Bertozzi, C. R. A Strain-Promoted [3 + 2] Azide-Alkyne Cycloaddition for Covalent Modification of Biomolecules in Living Systems. *J. Am. Chem. Soc.* **2004**, *126* (46), 15046–15047. https://doi.org/10.1021/ja044996f.

(5) Prescher, J. A.; Dube, D. H.; Bertozzi, C. R. Chemical Remodelling of Cell Surfaces in Living Animals. *Nature* **2004**, *430* (7002), 873–877. https://doi.org/10.1038/nature02791.

(6) Deiters, A.; Cropp, T. A.; Summerer, D.; Mukherji, M.; Schultz, P. G. Site-Specific PEGylation of Proteins Containing Unnatural Amino Acids. *Bioorganic Med. Chem. Lett.* **2004**, *14* (23), 5743–5745. https://doi.org/10.1016/j.bmcl.2004.09.059.

(7) Kim, C. H.; Axup, J. Y.; Dubrovska, A.; Kazane, S. A.; Hutchins, B. A.; Wold, E. D.; Smider, V. V.; Schultz, P. G. Synthesis of Bispecific Antibodies Using Genetically Encoded Unnatural Amino Acids. *J. Am. Chem. Soc.* **2012**, *134* (24), 9918–9921. https://doi.org/10.1021/ja303904e.

(8) Tanaka, K.; Fukase, K.; Katsumura, S. Exploring a Unique Reactivity of 6-Azaelectrocyclization to Enzyme Inhibition, Natural Products Synthesis, and Molecular Imaging: An Approach to Chemical Biology by Synthetic Chemists. *Synlett* **2011**, No. 15, 2115–2139. https://doi.org/10.1055/s-0030-1261192.

(9) Lee, Y. C. U.; Stowell, C. P.; Krantz, M. J. 2-Imino-2-Methoxyethyl 1-Thioglycosides: New Reagents for Attaching Sugars to Proteins. *Biochemistry* **1976**, *15* (18), 3956–3963. https://doi.org/10.1021/bi00663a008.

(10) Diethelm, S.; Schafroth, M. A.; Carreira, E. M. Amine-Selective Bioconjugation Using Arene Diazonium Salts. *Org. Lett.* **2014**, *16* (15), 3908–3911. https://doi.org/10.1021/ol5016509.

(11) Crankshaw, M. W.; Grant, G. A. Modification of Cysteine. *Curr. Protoc. Protein Sci.* **1996**, *3* (1), 15.1.1-15.1.18. https://doi.org/10.1002/0471140864.ps1501s03.

(12) Lundell, N.; Schreitmüller, T. Sample Preparation for Peptide Mapping - A Pharmaceutical Quality- Control Perspective. *Anal. Biochem.* **1999**, *266* (1), 31–47. https://doi.org/10.1006/abio.1998.2919.

(13) Smith, M. E. B.; Schumacher, F. F.; Ryan, C. P.; Tedaldi, L. M.; Papaioannou, D.; Waksman, G.; Caddick, S.; Baker, J. R. Protein Modification, Bioconjugation, and Disulfide Bridging Using Bromomaleimides. *J. Am. Chem. Soc.* **2010**, *132* (6), 1960–1965. https://doi.org/10.1021/ja908610s.

(14) Li, F.; Allahverdi, A.; Yang, R.; Lua, G. B. J.; Zhang, X.; Cao, Y.; Korolev, N.; Nordenskiöld, L.; Liu, C. F. A Direct Method for Site-Specific Protein Acetylation. *Angew. Chemie - Int. Ed.* **2011**, *50* (41), 9611–9614. https://doi.org/10.1002/anie.201103754.

(15) Conte, M. Lo; Staderini, S.; Marra, A.; Sanchez-Navarro, M.; Davis, B. G.;

Dondoni, A. Multi-Molecule Reaction of Serum Albumin Can Occur through Thiol-Yne Coupling. *Chem. Commun.* **2011**, *47* (39), 11086–11088. https://doi.org/10.1039/c1cc14402b.

(16) Chalker, J. M.; Gunnoo, S. B.; Boutureira, O.; Gerstberger, S. C.; Fernández-González, M.; Bernardes, G. J. L.; Griffin, L.; Hailu, H.; Schofield, C. J.; Davis, B. G. Methods for Converting Cysteine to Dehydroalanine on Peptides and Proteins. *Chem. Sci.* **2011**, *2* (9), 1666–1676. https://doi.org/10.1039/c1sc00185j.

(17) Hooker, J. M.; Kovacs, E. W.; Francis, M. B. Interior Surface Modification of Bacteriophage MS2. *J. Am. Chem. Soc.* **2004**, *126* (12), 3718–3719. https://doi.org/10.1021/ja031790q.

(18) Gavrilyuk, J.; Ban, H.; Nagano, M.; Hakamata, W.; Barbas, C. F. Formylbenzene Diazonium Hexafluorophosphate Reagent for Tyrosine-Selective Modification of Proteins and the Introduction of a Bioorthogonal Aldehyde. *Bioconjug. Chem.* **2012**, *23* (12), 2321–2328. https://doi.org/10.1021/bc300410p.

(19) Joshi, N. S.; Whitaker, L. R.; Francis, M. B. A Three-Component Mannich-Type Reaction for Selective Tyrosine Bioconjugation. *J. Am. Chem. Soc.* **2004**, *126* (49), 15942–15943. https://doi.org/10.1021/ja0439017.

(20) Guo, H. M.; Minakawa, M.; Ueno, L.; Tanaka, F. Synthesis and Evaluation of a Cyclic Imine Derivative Conjugated to a Fluorescent Molecule for Labeling of Proteins. *Bioorganic Med. Chem. Lett.* **2009**, *19* (4), 1210–1213. https://doi.org/10.1016/j.bmcl.2008.12.071.

(21) Ban, H.; Gavrilyuk, J.; Barbas, C. F. Tyrosine Bioconjugation through Aqueous Ene-Type Reactions: A Click-like Reaction for Tyrosine. *J. Am.*

Chem. Soc. **2010**, *132* (5), 1523–1525. https://doi.org/10.1021/ja909062q.

(22) Yu, Z.; Pan, Y.; Wang, Z.; Wang, J.; Lin, Q. Genetically Encoded Cyclopropene Directs Rapid, Photoclick-Chemistry- Mediated Protein Labeling in Mammalian Cells. *Angew. Chemie - Int. Ed.* **2012**, *51* (42), 10600–10604. https://doi.org/10.1002/anie.201205352.

(23) Lang, K.; Davis, L.; Torres-Kolbus, J.; Chou, C.; Deiters, A.; Chin, J. W. Genetically Encoded Norbornene Directs Site-Specific Cellular Protein Labelling via a Rapid Bioorthogonal Reaction. *Nat. Chem.* **2012**, *4* (4), 298–304. https://doi.org/10.1038/nchem.1250.

(24) Santoro, S. W.; Wang, L.; Herberich, B.; King, D. S.; Schultz, P. G. An Efficient System for the Evolution of Aminoacyl-TRNA Synthetase Specificity. *Nat. Biotechnol.* **2002**, *20* (10), 1044–1048. https://doi.org/10.1038/nbt742.

(25) Kodama, K.; Fukuzawa, S.; Nakayama, H.; Kigawa, T.; Sakamoto, K.; Yabuki, T.; Matsuda, N.; Shirouzu, M.; Takio, K.; Tachibana, K.; et al. Regioselective Carbon-Carbon Bond Formation in Proteins with Palladium Catalysis; New Protein Chemistry by Organometallic Chemistry. *ChemBioChem* **2006**, *7* (1), 134–139. https://doi.org/10.1002/cbic.200500290.

(26) Addy, P. S.; Erickson, S. B.; Italia, J. S.; Chatterjee, A. A Chemoselective Rapid Azo-Coupling Reaction (CRACR) for Unclickable Bioconjugation. *J. Am. Chem. Soc.* **2017**, *139* (34), 11670–11673. https://doi.org/10.1021/jacs.7b05125.

(27) Seitchik, J. L.; Peeler, J. C.; Taylor, M. T.; Blackman, M. L.; Rhoads, T. W.; Cooley, R. B.; Refakis, C.; Fox, J. M.; Mehl, R. A. Genetically Encoded Tetrazine Amino Acid Directs Rapid Site-Specific in Vivo Bioorthogonal Ligation with Trans-Cyclooctenes. *J. Am. Chem. Soc.* **2012**, *134* (6), 2898–

2901. https://doi.org/10.1021/ja2109745.

(28) Deiters, A.; Cropp, T. A.; Mukherji, M.; Chin, J. W.; Anderson, J. C.; Schultz, P. G. Adding Amino Acids with Novel Reactivity to the Genetic Code of Saccharomyces Cerevisiae. *J. Am. Chem. Soc.* **2003**, *125* (39), 11782–11783. https://doi.org/10.1021/ja0370037.

(29) Lang, K.; Davis, L.; Wallace, S.; Mahesh, M.; Cox, D. J.; Blackman, M. L.; Fox, J. M.; Chin, J. W. Genetic Encoding of Bicyclononynes and Trans-Cyclooctenes for Site-Specific Protein Labeling in Vitro and in Live Mammalian Cells via Rapid Fluorogenic Diels-Alder Reactions. *J. Am. Chem. Soc.* **2012**, *134* (25), 10317–10320. https://doi.org/10.1021/ja302832g.

(30) Wangt, L.; Zhang, Z.; Brock, A.; Schultz, P. G. Addition of the Keto Functional Group to the Genetic Code of Escherichia Coli. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (1), 56–61. https://doi.org/10.1073/pnas.0234824100.

(31) Zhang, C.; Welborn, M.; Zhu, T.; Yang, N. J.; Santos, M. S.; Van Voorhis, T.; Pentelute, B. L. π-Clamp-Mediated Cysteine Conjugation. *Nat. Chem.* **2016**, *8* (2), 120–128. https://doi.org/10.1038/nchem.2413.

(32) Zhang, C.; Dai, P.; Vinogradov, A. A.; Gates, Z. P.; Pentelute, B. L. Site-Selective Cysteine–Cyclooctyne Conjugation. *Angew. Chemie* **2018**, *130* (22), 6569–6573. https://doi.org/10.1002/ange.201800860.

(33) Ramil, C. P.; An, P.; Yu, Z.; Lin, Q. Sequence-Specific 2-Cyanobenzothiazole Ligation. *J. Am. Chem. Soc.* **2016**, *138* (17), 5499–5502. https://doi.org/10.1021/jacs.6b00982.

(34) Wang, W.; Gao, J. N, S-Double Labeling of N-Terminal Cysteines via an Alternative Conjugation Pathway with 2-Cyanobenzothiazole. *J. Org. Chem.* **2020**. https://doi.org/10.1021/acs.joc.9b02959.

(35) Matos, M. J.; Oliveira, B. L.; Martínez-Sáez, N.; Guerreiro, A.; Cal, P. M. S. D.; Bertoldo, J.; Maneiro, M.; Perkins, E.; Howard, J.; Deery, M. J.; et al. Chemo- and Regioselective Lysine Modification on Native Proteins. *J. Am. Chem. Soc.* **2018**, *140* (11), 4004–4017. https://doi.org/10.1021/jacs.7b12874.

(36) Adusumalli, S. R.; Rawale, D. G.; Singh, U.; Tripathi, P.; Paul, R.; Kalra, N.; Mishra, R. K.; Shukla, S.; Rai, V. Single-Site Labeling of Native Proteins Enabled by a Chemoselective and Site-Selective Chemical Technology. *J. Am. Chem. Soc.* **2018**, *140* (44), 15114–15123. https://doi.org/10.1021/jacs.8b10490.

(37) Sereda, T. J.; Mant, C. T.; Quinn, A. M.; Hodges, R. S. Effect of the α-Amino Group on Peptide Retention Behaviour in Reversed-Phase Chromatography Determination of the PKa Values of the α-Amino Group of 19 Different N-Terminal Amino Acid Residues. *J. Chromatogr. A* **1993**, *646* (1), 17–30. https://doi.org/10.1016/S0021-9673(99)87003-4.

(38) Dozier, J. K.; Distefano, M. D. Site-Specific Pegylation of Therapeutic Proteins. *Int. J. Mol. Sci.* **2015**, *16* (10), 25831–25864. https://doi.org/10.3390/ijms161025831.

(39) Chan, A. O. Y.; Ho, C. M.; Chong, H. C.; Leung, Y. C.; Huang, J. S.; Wong, M. K.; Che, C. M. Modification of N-Terminal α-Amino Groups of Peptides and Proteins Using Ketenes. *J. Am. Chem. Soc.* **2012**, *134* (5), 2589–2598. https://doi.org/10.1021/ja208009r.

(40) Chen, D.; Disotuar, M. M.; Xiong, X.; Wang, Y.; Chou, D. H.-C. Selective N-Terminal Functionalization of Native Peptides and Proteins. *Chem. Sci.* **2017**, *8* (4), 2717–2722. https://doi.org/10.1039/C6SC04744K.

(41) Schoffelen, S.; van Eldijk, M. B.; Rooijakkers, B.; Raijmakers, R.; Heck, A. J.

R.; van Hest, J. C. M. Metal-Free and PH-Controlled Introduction of Azides in Proteins. *Chem. Sci.* **2011**, *2* (4), 701–705. https://doi.org/10.1039/c0sc00562b.

(42) Yamamoto, Y.; Tsutsumi, Y.; Yoshioka, Y.; Nishibata, T.; Kobayashi, K.; Okamoto, T.; Mukai, Y.; Shimizu, T.; Nakagawa, S.; Nagata, S.; et al. Site-Specific Pegylation of a Lysine-Deficient TNF-α with Full Bioactivity. *Nat. Biotechnol.* **2003**, *21* (5), 546–552. https://doi.org/10.1038/nbt812.

(43) Dixon, H. B. F.; Fields, R. B. T.-M. in E. [33] Specific Modification of NH2-Terminal Residues by Transamination. In *Enzyme Structure, Part B*; Academic Press, 1972; Vol. 25, pp 409–419. https://doi.org/https://doi.org/10.1016/S0076-6879(72)25036-4.

(44) Abrahms, L.; Tom, J.; Burnier, J.; Butcher, K. A.; Kossiakoff, A.; Wells, J. A. Engineering Subtilisin and Its Substrates for Efficient Ligation of Peptide Bonds in Aqueous Solution. *Biochemistry* **1991**, *30* (17), 4151–4159. https://doi.org/10.1021/bi00231a007.

(45) Antos, J. M.; Chew, G. L.; Guimaraes, C. P.; Yoder, N. C.; Grotenbreg, G. M.; Popp, M. W. L.; Ploegh, H. L. Site-Specific N- and C-Terminal Labeling of a Single Polypeptide Using Sortases of Different Specificity. *J. Am. Chem. Soc.* **2009**, *131* (31), 10800–10801. https://doi.org/10.1021/ja902681k.

(46) Macdonald, J. I.; Munch, H. K.; Moore, T.; Francis, M. B. One-Step Site-Specific Modification of Native Proteins with 2-Pyridinecarboxyaldehydes. *Nat. Chem. Biol.* **2015**, *11* (5), 326–331. https://doi.org/10.1038/nchembio.1792.

(47) Ng, S.; Jafari, M. R.; Matochko, W. L.; Derda, R. Quantitative Synthesis of Genetically Encoded Glycopeptide Libraries Displayed on M13 Phage. *ACS Chem. Biol.* **2012**, *7* (9), 1482–1487. https://doi.org/10.1021/cb300187t.

(48) Geoghegan, K. F.; Stroh, J. G. Site-Directed Conjugation of Nonpeptide Groups to Peptides and Proteins Via Periodate Oxidation of a 2-Amino Alcohol. Application to Modification at N-Terminal Serine. *Bioconjug. Chem.* **1992**, *3* (2), 138–146. https://doi.org/10.1021/bc00014a008.

(49) Chen, J. K.; Tanaka, A.; Schreiber, S. L.; Lane, W. S.; Braucr, A. W. Biased Combinatorial Libraries; Novel Ligands for the SH3 Domain of Phosphatidylinositol 3-Kinase. *J. Am. Chem. Soc.* **1993**, *115* (26), 12591–12592. https://doi.org/10.1021/ja00079a051.

(50) Chou, Y.; Kitova, E. N.; Joe, M.; Brunton, R.; Lowary, T. L.; Klassen, J. S.; Derda, R. Genetically-Encoded Fragment-Based Discovery (GE-FBD) of Glycopeptide Ligands with Differential Selectivity for Antibodies Related to Mycobacterial Infections. *Org. Biomol. Chem.* **2018**, *16* (2), 223–227. https://doi.org/10.1039/c7ob02783d.

(51) Ng, S.; Bennett, N. J.; Schulze, J.; Gao, N.; Rademacher, C.; Derda, R. Genetically-Encoded Fragment-Based Discovery of Glycopeptide Ligands for DC-SIGN. *Bioorganic Med. Chem.* **2018**, *26* (19), 5368–5377. https://doi.org/10.1016/j.bmc.2018.08.036.

(52) Li, X.; Zhang, L.; Hall, S. E.; Tam, J. P. A New Ligation Method for N-Terminal Tryptophan-Containing Peptides Using the Pictet-Spengler Reaction. *Tetrahedron Lett.* **2000**, *41* (21), 4069–4073. https://doi.org/10.1016/S0040-4039(00)00592-X.

(53) Aimoto, S. Synthesis of Proteins by Native Chemical Ligation. *Tanpakushitsu Kakusan Koso.* **2007**, *52* (13 Suppl), 1804–1805.

(54) Kent, S. B. H. Total Chemical Synthesis of Proteins. *Chem. Soc. Rev.* **2009**, *38* (2), 338–351. https://doi.org/10.1039/b700141j.

(55) Wissner, R. F.; Batjargal, S.; Fadzen, C. M.; Petersson, E. J. Labeling Proteins with Fluorophore/Thioamide Förster Resonant Energy Transfer Pairs by Combining Unnatural Amino Acid Mutagenesis and Native Chemical Ligation. *J. Am. Chem. Soc.* **2013**, *135* (17), 6529–6540. https://doi.org/10.1021/ja4005943.

(56) Zhang, L.; Tam, J. P. Thiazolidine Formation as a General and Site-Specific Conjugation Method for Synthetic Peptides and Proteins. *Anal. Biochem.* **1996**, *233* (1), 87–93. https://doi.org/10.1006/abio.1996.0011.

(57) White, E. H.; McCapra, F.; Field, G. F. The Structure and Synthesis of Firefly Luciferin. *J. Am. Chem. Soc.* **1963**, *85* (3), 337–343. https://doi.org/10.1021/ja00886a019.

(58) Ren, H.; Xiao, F.; Zhan, K.; Kim, Y.-P.; Xie, H.; Xia, Z.; Rao, J. A Biocompatible Condensation Reaction for the Labeling of Terminal Cysteine Residues on Proteins. *Angew. Chemie Int. Ed.* **2009**, *48* (51), 9658–9662. https://doi.org/10.1002/anie.200903627.

(59) Nguyen, D. P.; Elliott, T.; Holt, M.; Muir, T. W.; Chin, J. W. Genetically Encoded 1,2-Aminothiols Facilitate Rapid and Site-Specific Protein Labeling via a Bio-Orthogonal Cyanobenzothiazole Condensation. *J. Am. Chem. Soc.* **2011**, *133* (30), 11418–11421. https://doi.org/10.1021/ja203111c.

(60) Cheng, Y.; Peng, H.; Chen, W.; Ni, N.; Ke, B.; Dai, C.; Wang, B. Rapid and Specific Post-Synthesis Modification of DNA through a Biocompatible Condensation of 1,2-Aminothiols with 2-Cyanobenzothiazole. *Chem. - A Eur. J.* **2013**, *19* (12), 4036–4042. https://doi.org/10.1002/chem.201201677.

3.2 Thiazolidino boronate mediated acyl transfer

3.2.1 Introduction

As is discussed in **Chapter 1**, 2-FPBA conjugates rapidly with a cysteine in aqueous solution to form TzB[1]. The reaction starts with the iminoboronate formation between 2-FPBA and α-amine followed by the nucleophilic attack from cysteine sidechain to the imine intermediate. As a fast and reversible reaction, the application of TzB is limited when a stable conjugate is desired. In fact, we have found that the TzB complex can still be oxidized by oxidative species and dissociates upon dilution or acidification within an hour.[2,3] To overcome this problem so that the TzB chemistry can be tolerated in a broader range of conditions, we examined the possibility of adding an acyl transfer step to make this conjugation irreversible. This idea was inspired by the work from Xucheng Li and James Tam.[4–6] In their contribution, aldehyde mediated O→N acyl transfer was utilized as a ligation method on serine or threonine. We noticed that the intermediate, which is an acetylated oxazolidine ring structure, actually survived HPLC conditions, indicating it is more stable compared to a TzB complex. We hypothesized that acetylation on the thiazolidine ring would stop the regeneration of a C=N bond in the intermediate step making it an irreversible reaction.

In this chapter, thiazolidino boronate mediated acyl transfer will be discussed in detail including its mechanism, kinetics and thermodynamics. Moreover, we will demonstrate that this chemistry can be applied in constructing a phage display library that has two distinct modifications. We believe this methodology can expand the diversity and utility of phage display.

3.2.2 Acetylation of TzB complex with acetyl chloride

Although the N-B coordination is almost as strong as a covalent bond, it is still in quick equilibrium exposing the secondary amine for potential further reaction. To test this hypothesis, an acetylation experiment was carried out on a preformed TzB complex: to 1 mL of 100 mM TzB (cloudy) in PBS was added 1 eq. of acetyl chloride dissolved in 1 mL of acetone. 1 eq. K_2CO_3 solid was added immediately and the mixture was stirred vigorously for 1 hr. The reaction mixture was diluted and subjected for LC/MS analysis.

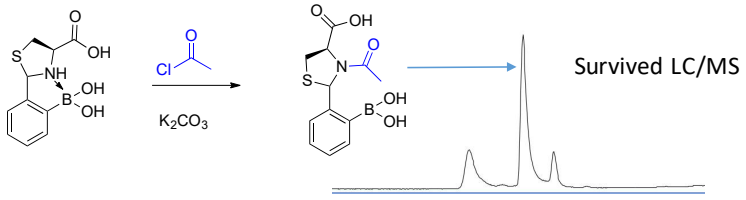


Figure 3.2.1. LC/MS analysis of acetylation reaction on TzB complex. In contrary to non-acetylated TzB complex, the product survived the acidic condition in LC/MS analysis

Contrary to the TzB complex, this conjugate survived the LC/MS conditions (**Figure 3.2.1**). Despite this encouraging result, the addition of acetyl chloride into a protein solution is not ideal because it is likely to react with other amino acids, including lysine and arginine. We proposed a milder way to acylate the secondary amine with an intramolecular acyl transfer reaction, as is shown in **Figure 3.2.2**. Acyl transfer reactions, including O→N, S→N and Se→N transfers, are all widely used in peptide ligations. In our case, when the N-B coordination breaks up, an acyl transfer reaction can potentially occur between the adjacent ester and the revealed secondary amine. Synthesis of the ester derivative of 2-FPBA, named KL42, is attached at the end of this chapter.

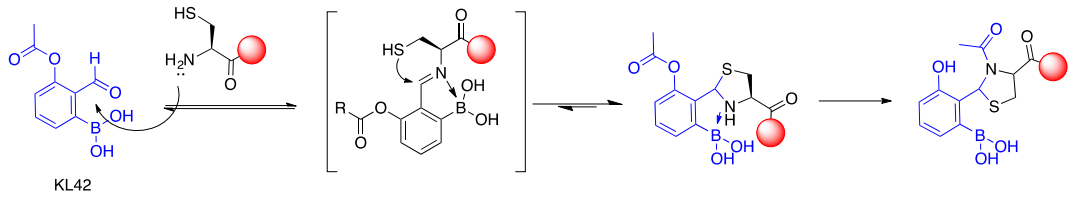


Figure 3.2.2. Proposed mechanism for TzB mediated acyl transfer. Intramolecular acyl transfer results in a stable, acetylated TzB complex

It is well known that salicylaldehyde ester can go through rapid hydrolysis in aqueous solution due to the catalytic effect from the neighboring aldehyde group.[7] As a matter of fact, we have found that the half-life of acetyl salicylaldehyde ester is less than 3 min in PBS buffer, pH 7.4. To test the hydrolytic property of KL42, we monitored its hydrolysis with ^{1}H NMR in PBS containing 15% D_2O. Curve fitting according to a first-order reaction mechanism yielded the half-life ($t_{1/2}$) values. To our delight, the hydrolysis of KL42 is much slower compared to salicylaldehyde ester, probably due to the formation of boroxozole (**Figure 3.2.3**). The slower hydrolytic rate provides enough time for the association between KL42 and N-terminal cysteine residue, as well as the following acyl transfer step.

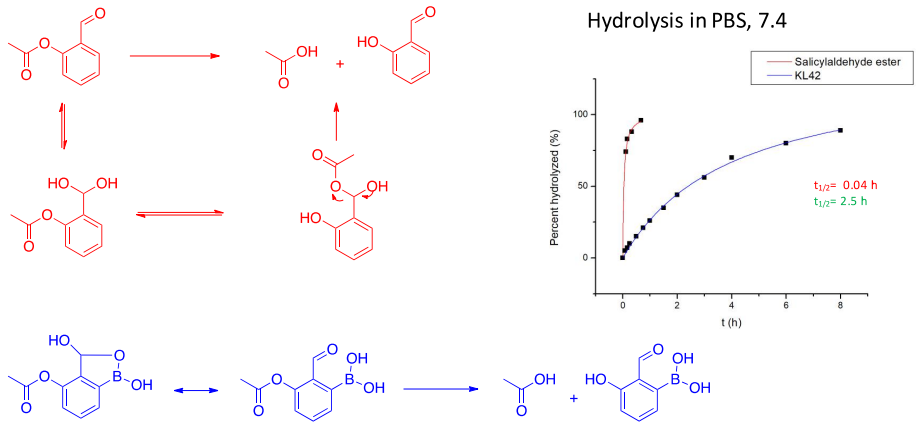


Figure 3.2.3. Mechanism for accelerated hydrolysis of a typical salicylaldehyde ester. The neighboring aldehyde catalyzes the hydrolysis of the ester. The boroxozole formation of KL42 slows down its hydrolysis in PBS

3.2.3 Kinetics of KL42 with CLA peptide

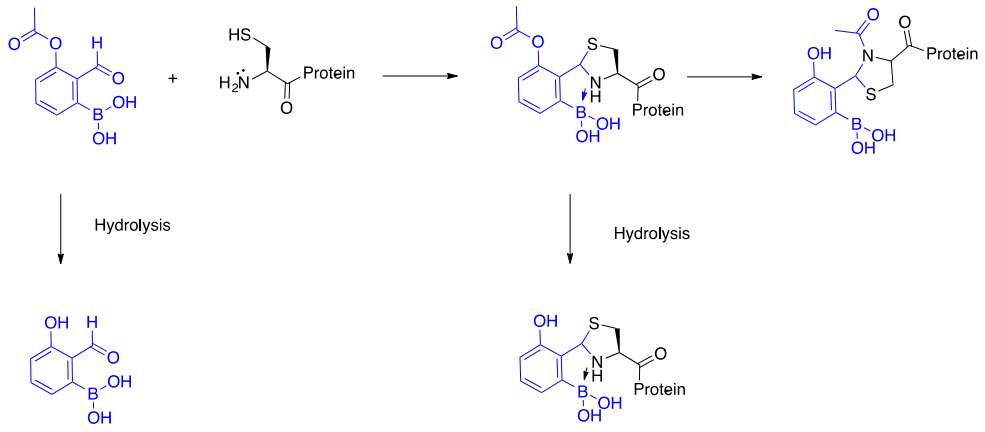


Figure 3.2.4. Possible products and side products between the reaction of KL42 and N-terminal cysteine containing peptides/proteins

Kinetics of KL42 reacting with a model peptide P1 (sequence: Cys-Leu-Ala) at 1 mM was measured by ^{1}H NMR. As is shown by NMR spectrum in **Figure 3.2.5a**, acyl transfer happens during the period of 20 hrs. Upon mixing, a peak at 5.95 ppm showed up (blue star), corresponding to thiazolidino boronate. After 20 hrs of incubation at r.t., the new product, named KL44^{P1} showed peaks at 6.42 and 6.36 ppm (red stars). When analyzed by LC/MS, the mixture only showed one single peak with a mass corresponding to the conjugated product. This result indicates that the two sets of peaks in NMR are isomers. The effort to crystalize KL44^{P1} was not successful. However, we were able to obtain the crystal structure for the conjugates of KL42 with cysteamine (product named as KL44^{P2}) and Cysteine-Aniline (product named as KL44^{P3}). The crystal structures both display an N-acylated thiazolidine core. Importantly, KL44^{P2} presents a cis configuration, contrary to KL44^{P3}. These results indicate that the product of KL42 and aminothiol could exist as isomers, which explains the two species shown in NMR.

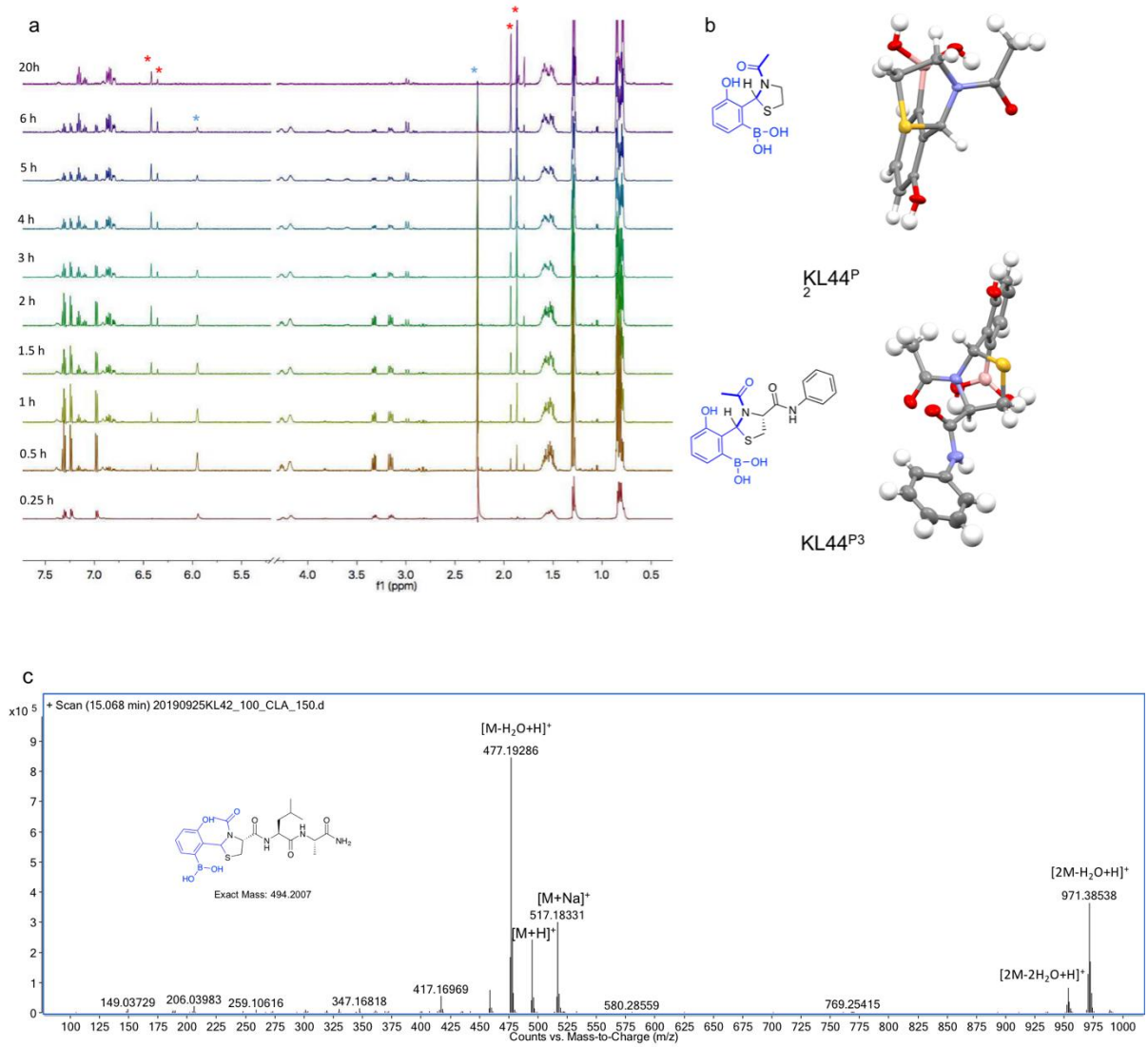


Figure 3.2.5. KL42 reacting with CLA peptide. a) ^{1}H-NMR stacked spectrums monitoring the acyl transfer process. The signature peak of TzB at ~6.0 ppm (blue star)gradually shifted to ~6.4 ppm (red stars); b) Crystal structures of KL44^{P2} and KL44^{P3}; c) Mass spec of KL44^{P1} corresponding to acetylated TzB complex

3.2.4 Stability of the product

The stability of acylated product was examined by LC/MS and compared to a simple TzB complex (non-acylated). To simplify the quantification process, we synthesized a peptide P1* with the sequence CLA(Dap-FAM). KL44^{P1*} showed no dissociation during LC/MS analysis whereas the non-acylated product (KL45^{P1*})

showed 90% degradation. When both conjugates at 100 µM were mixed with free cysteine (1 mM, 10 eq.), KL45^{P1*} completely disappeared from LC trace due to quick exchange with free cysteine and KL44^{P1*} showed no change. Furthermore, little degradation was observed over 6 days when KL44^{P1*} is dissolved in PBS buffer at pH 7.4 or 5.0, corresponding to physiological pH or slight acidic condition in lysosomes respectively.

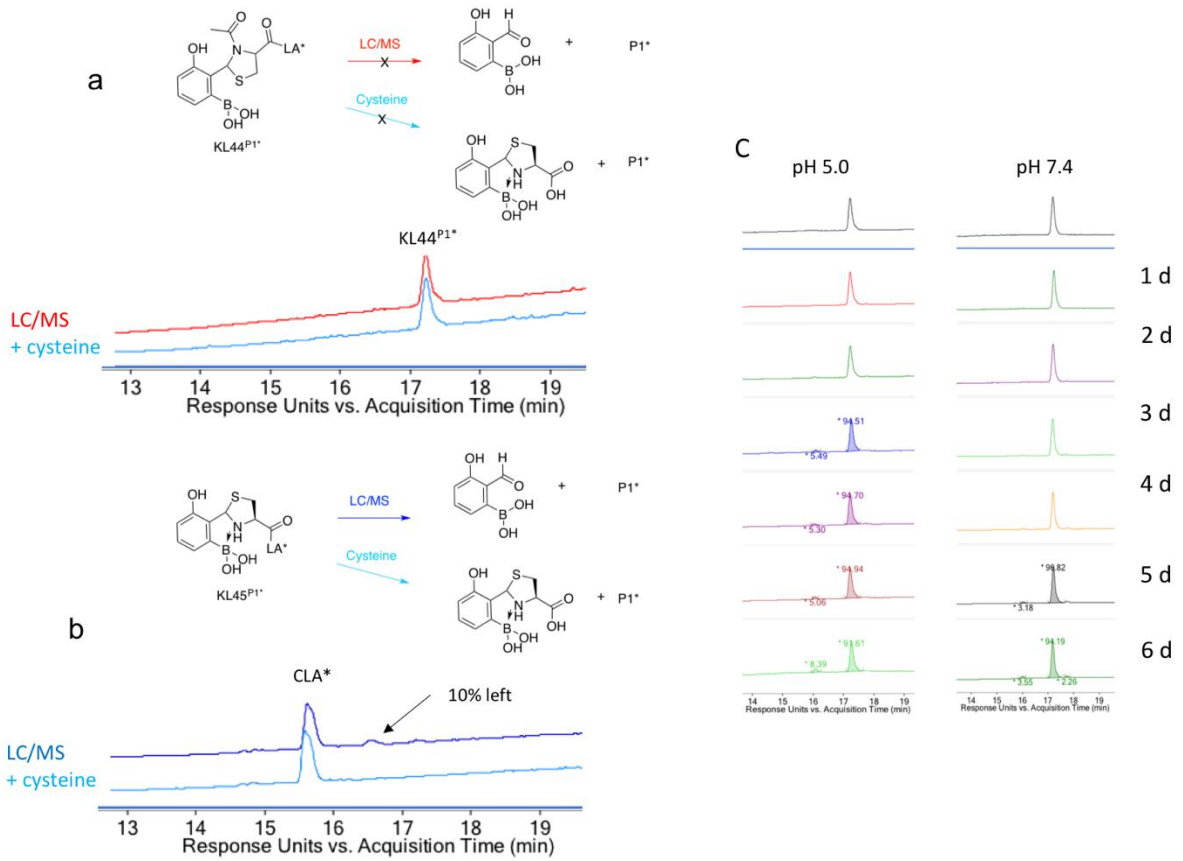


Figure 3.2.6. Stability of acylated TzB complex. a) KL44^{P1*} shows no degradation under LC/MS conditions with and without the addition of cysteine; b) KL45^{P1*} shows 90% dissociation under LC/MS conditions and completely dissociates with addition of cysteine; c) KL44^{P1*} displays negligible degradation, at pH 5.0 or pH 7.4, over 6 days according to LC-MS analysis

3.2.5 pH dependence and concentration dependence

The $t_{1/2}$ of this reaction at pH 7.4 was determined to be 2.9 hrs, which is rather sluggish compared to TzB formation. In order to accelerate the reaction, we investigated the effect of pH on the acyl transfer step. Excitingly, the same experiment was conducted at pH 6.0 and a new $t_{1/2}$ was determined to be 0.3 hrs, which is 10 times shorter than that in pH 7.4. This can be explained by faster dynamics of the N-B coordination under slightly acidic condition, as is illustrated in **Figure 3.2.7b**. Higher proton concentration helps to break the N-B coordination and reveal the secondary amino group for acylation.

The mechanism in **Figure 3.2.4** indicates that the conjugation of KL42 with N-terminal cysteine might be concentration independent because TzB formation is a rapid process. We tested out this hypothesis by monitoring the reaction at 1 mM and 0.2 mM. Plotting percent conversion versus time gave the profile for the reaction kinetics. The $t_{1/2}$ was determined to be 2.9 hrs for 1 mM reaction and 5.9 hrs for 0.2 mM reaction. This less than two folds of difference supports our hypothesis that acyl transfer is the rate limiting step.

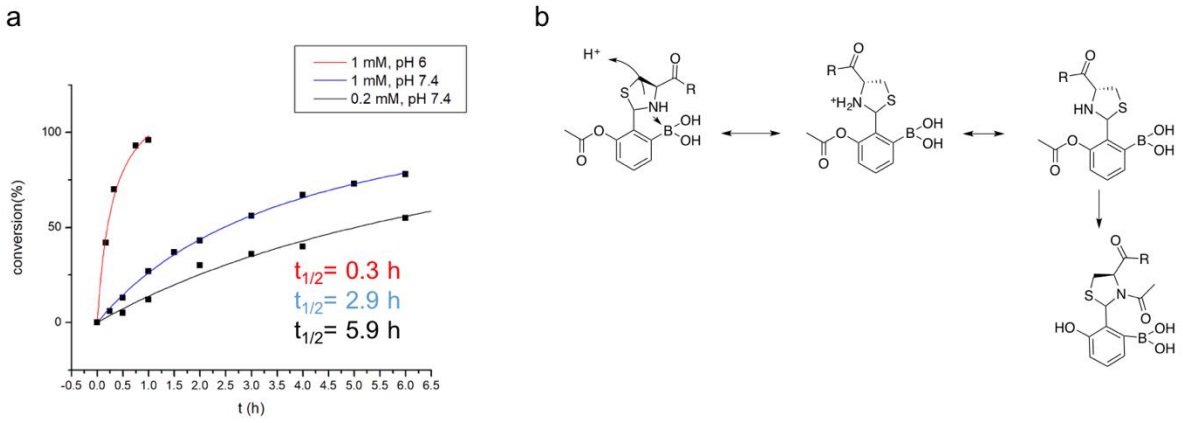


Figure 3.2.7. a) Kinetics of KL42 reacting with P1 at 1 mM, pH 6.0 or 1 mM, pH 7.4 or 0.2 mM, pH 7.4. Reactions were monitored by ^{1}H-NMR. Acyl transfer happens much faster at pH 6.0; b) possible

explanation for faster kinetics at pH 6.0. Higher concentration of proton helps to break the N-B coordination to allow acyl transfer

3.2.6 Reaction at single digit micromolar concentration

Inspired by the fact that the reaction is largely concentration independent, we were curious about the results at a single digit micromolar concentration. We firstly compared the reaction of peptide P1 reacting with KL42 or CBT in their optimal condition.

Monitoring reactions at such a low concentration can be challenging. LC/MS only gives weak and noisy signals for small molecules both on mass and UV trace at low micromolar. To overcome this problem, we concentrated our analytes after the reactions were completed. Briefly, 2 μM KL42 was mixed with 2 uL CLA peptide in 0.02 × PBS, pH 6. The mixture was incubated for 1 h and quenched by 20 μM cysteine and lyophilized. The residue solid was re-dissolved in 40 μL of nano pure water. The sample was spun down and an aliquot of supernatant was was taken for LC/MS analysis. The same condition was used for CBT, except the pH was tuned to 7.4. Not surprisingly, KL42 still gave about 80% yield at such a low concentration, whereas CBT only yielded ~5% product.

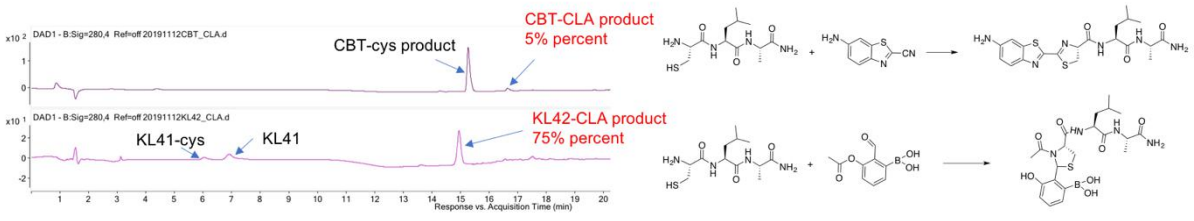


Figure 3.2.8. Comparative study between KL42 and CBT reacting with P1 at 2 μM for 1 h. KL42 gives 75% yield whereas CBT only gives 5% yield

The same comparison was carried out between KL42 and the widely used NCL. NCL is not considered a fast reaction, and it usually requires 4-mercaptophenylacetic acid (4-MPAA) as a catalyst to accelerate the thiol-exchange step. To compare the efficiency of our chemistry and eliminate the need for any additional catalyst, we synthesized an acetyl thioester derived from 4-MPAA. Also, we replaced P1 with P1* to better quantify the conversion because the product of P1 and the thioester contains no chromophore. This comparison follows the same procedure as described above. Not surprisingly, KL42 again showed ~80% conversion, which is significantly higher than that of thioester (< 10%). These results combined suggest that our chemistry performs much better at low concentration and it can potentially be a better choice for intracellular protein modification.

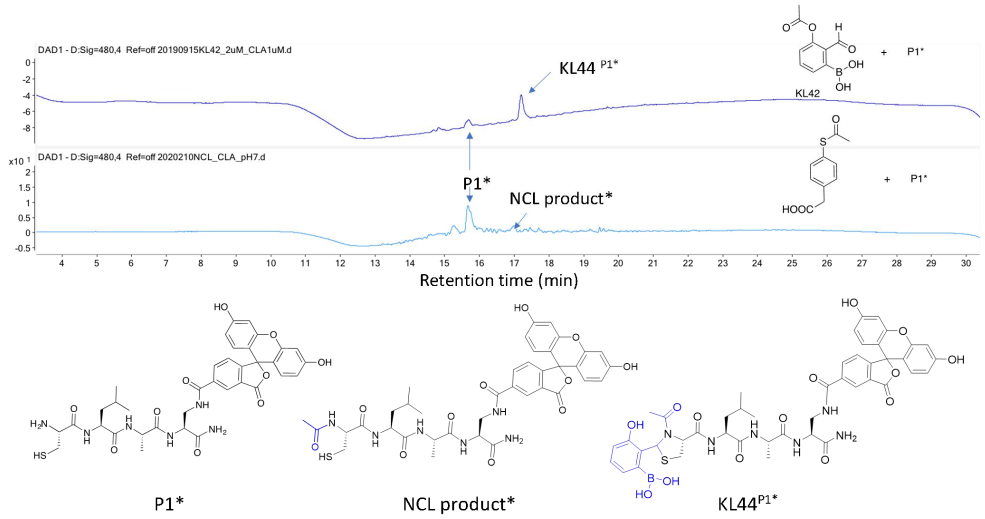


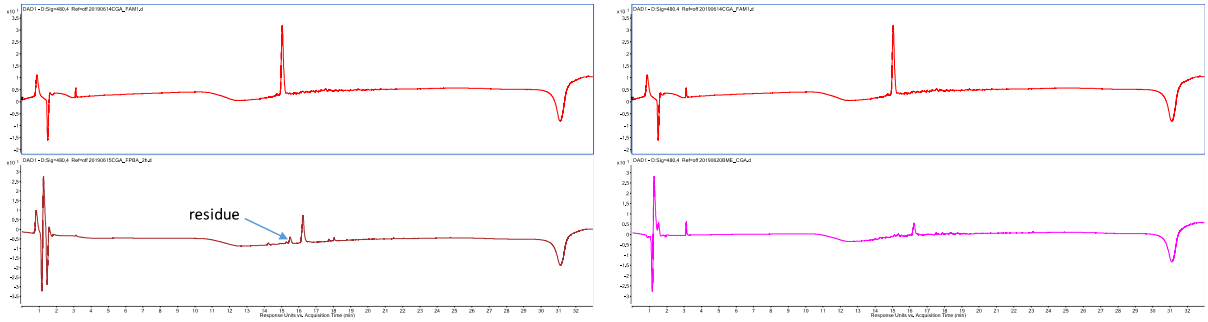
Figure 3.2.9. Comparative study between KL42 and thioester reacting with P1* at 2 μM. KL42 showed ~80% yield whereas thioester showed less than 10% yield

3.2.7 Effect of the neighboring amino acid

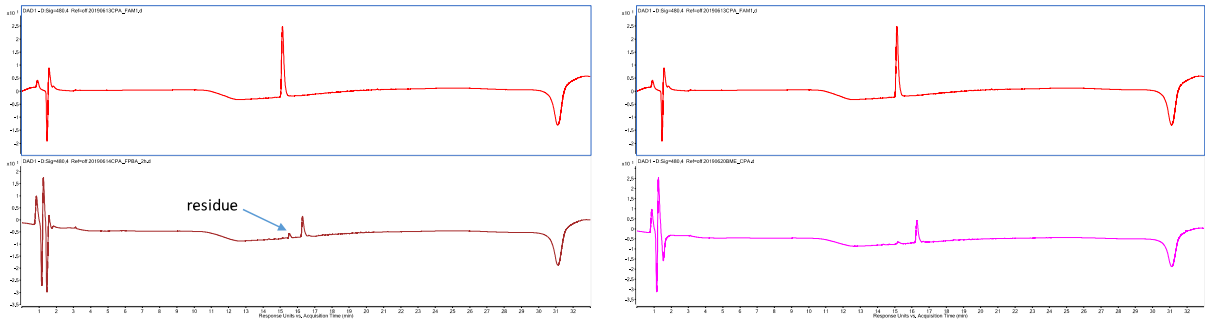
Model peptide CXA-FAM (X= G, P, W) were synthesized to test the structural dependence of the peptides that contains N-terminal cysteine. The peptides

were labelled at 50 μM with 1.5 eq. of KL42 at pH 6.0 for 2 hrs. As is shown in **Figure 3.2.10**, ~90% peptides were converted to products with about 10% oxidation observed consistently. 0.5 mM of 2-mercaptoethanol was added to each reaction to prevent oxidation of the peptides and complete conversion was achieved.

a



b



c

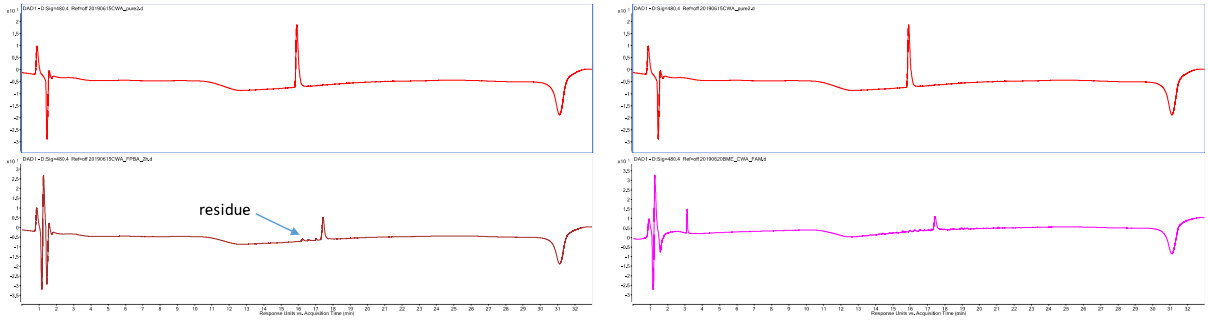


Figure 3.2.10. KL42 labels CXA* at sub-millimolar concentration. a) CGA*, b) CPA* and c) CWA*. Left: without the addition of 2-mercaptoethanol; Right: with the addition of 2-mercaptoethanol. The addition of 2-mercaptoethanol prevented the oxidation of these peptides

3.2.8 Peptide labelling in complex media

To demonstrate the bio-compatibility of this reaction, we carried out this reaction in presence of blood serum. Peptide P1* (100 µM) was mixed with 150 µM KL42 from 10 mM stock in the presence of 10% FBS in PBS pH 6.0 buffer. The sample was incubated for 2 hr and then subjected to LC-MS analysis. The results show efficient conversion of the P1* to KL44^{P1*} even in the presence of 10% FBS. Consistent with previous results, a small degree of oxidized P1* was observed.

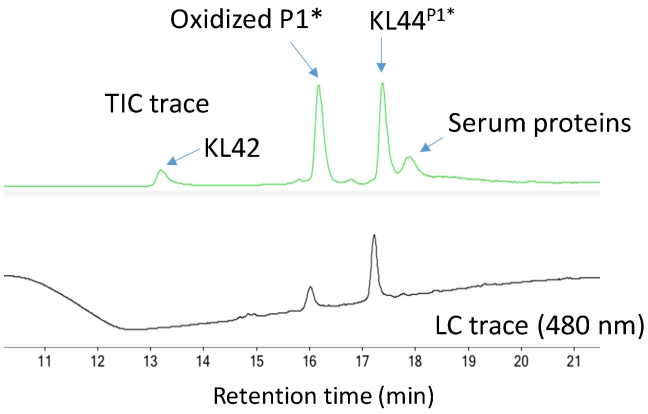


Figure 3.2.11. KL42 efficiently labels P1* in presence of 10% FBS at 100 µM

3.2.9 Selectivity test: peptide dual label

To examine the site selectivity of this reaction, we synthesized a model peptide P7 with the sequence CQRDKYC(Dap-FAM). When 50 µM P7 was treated with 250 µM (5 eq.) of KL42, a single product was observed on LC/MS, corresponding to N-terminal cysteine modified conjugate. Further treating this sample with 0.5 mM of iodoacetamide (IA) yielded a clean product with dual modifications. KL42 in this experiment showed high specificity towards N-terminal cysteine.

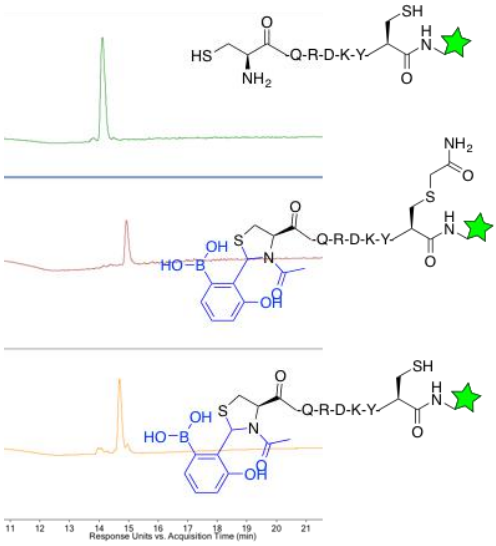


Figure 3.2.12. P7 was sequentially labeled by KL42 and IA. No side reaction was observed even with the addition of 5 eq. of KL42

NCL and CBT condensation were both believed to be site-selective. However, we observed side reactions in which both chemistries elicit internal cysteine modification. When 5.eq. of CBT was mixed with P7 at pH 7.4 for 1 hr, a side product was detected corresponding to the mass of P7 plus two CBT modifications (-18 for luciferin formation, **Figure 3.2.13a**). Together with the work from Qing Ling and others[8,9], we have enough evidence to show that CBT also reacts with internal cysteine residues. In fact, the thioimidate product is stable enough to prevent the following alkylation of the internal cysteine (**Figure 3.2.13b**). The same experiment was conducted with thioester. Not surprisingly, the mass of the major product turned out to be P7 with three additional acetyl groups. This result suggests that after NCL, both internal thiol groups reacted with the excess thioester.

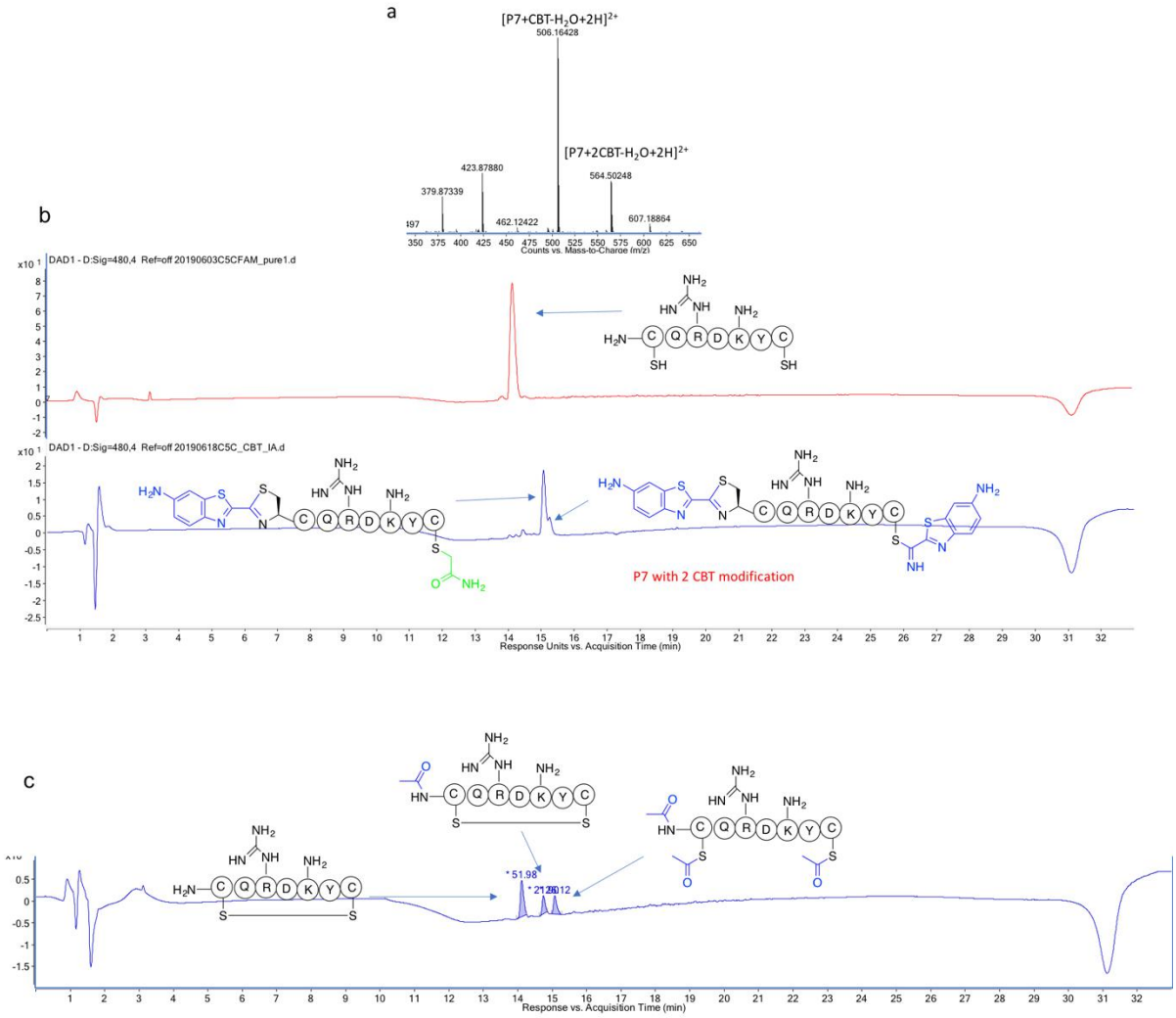


Figure 3.2.13. CBT condensation and NCL showed side reactions. a) Mass spec showing P7 being dual labelled by CBT; b) Internal cysteine failed to be alkylated due to thioimidate formation; c) Internal cysteine residues went through thiol-exchange with thioester, resulting in a mixture of products

3.2.10 Site specific modification on proteins

After we demonstrated the efficient and site-specific modification of peptides with KL42, we next applied this chemistry to proteins. We expressed two recombinant proteins, azo-reductase (AzoR) and thioredoxin (Trx), whose N-terminus was mutated to carry the sequence ENLYGQC (cleaved by TEV protease) or IEGRC (cleaved by factor Xa) respectively. Cleavage of these peptide sequences exposes the N-terminal cysteine for latter modification.

3.2.10.1 Expression of Trx

Trx is a class of small proteins that exist in all organisms. It plays an important role in redox signaling. Trx is a good example to demonstrate the site specificity of our chemistry because it contains five copies of cysteine residues, all of which present a free thiol group when the protein is in its reduced form.[10] A pET16b vector encoding wild type thioredoxin (WT-Trx) in *E.coli* (BL21) was obtained from professor Eranthie Weerapana and the protein was expressed by a standard protocol attached at the end of this section.

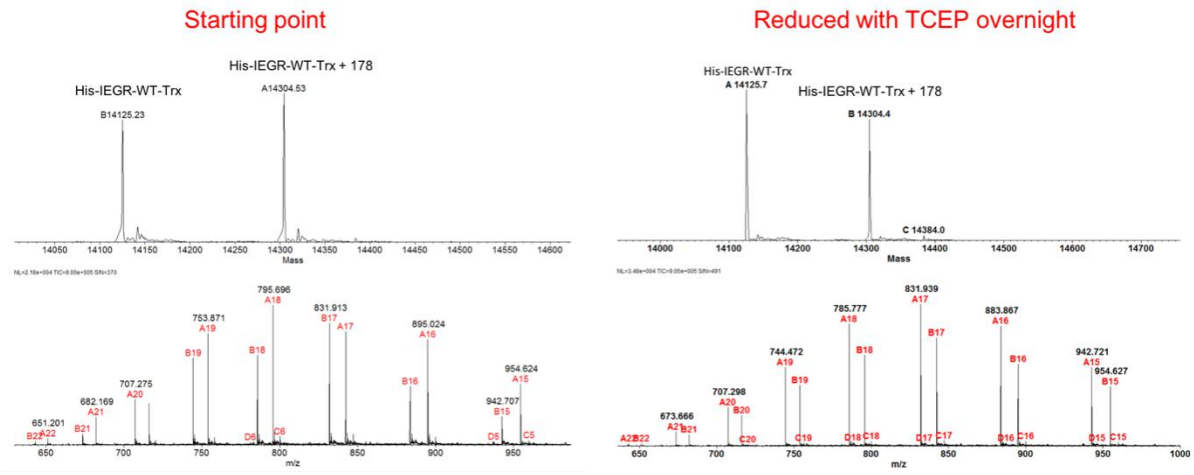


Figure 3.2.14. WT-Trx contains two species after purification, which decreases over time with the addition of TCEP. This unexpected mass corresponds to Trx being modified by a cysteine-glycine peptide through disulfide bond formation

LC/MS data suggests that the purified protein was a 'mixture' with the desired mass and a mass that is 178 greater, corresponding to a cysteine-glycine adduct. This modification presumably happens on one of the cysteine residues through disulfide bond because the intensity of the adduct on mass spec decreases with the addition of TCEP. However, this process is sluggish and the adduct peak did not completely disappear after a week.

Protein sequence:

GHHHHHHHHHHSSGHIEGRHMVKQIESKTAFQEALDAAGDKLVVVD
FSATWCGPCKMIKPFFHSLSEKYSNVIFLEVDVDDCQDVASECEVKCMPTFQF
FKKGQKVGEFSGANKEKLEATINELV

3.2.10.2 Cleavage of Trx

The protein after purification was desalted through a NAP-5 column from GE healthcare into a new buffer: 20 mM Tris, 100 mM NaCl, 2 mM $CaCl_2$, pH 8.0 for factor Xa cleavage. However, initial attempts to cleave Trx turned out be unsuccessful, different conditions and their outcomes are listed in **Table 3.2.1** (- means no cleavage).

Table 3.2.1. Conditions tested for thioredoxin cleavage by changing the incubation time, pH, concentration or temperature. Neither of these changes enabled cleavage of the IEGR sequence

Standard	Incubation time				desalt again	Dilution		pH	Temperature	pH + Tmeperature
	2 h	6 h	overnight	2 days		2 ×	4 ×	6.5	37	6.5/37
-	-	-	-	-	-	-	-	-	-	-

It was hypothesized that Trx could reduce the disulfide bond in Factor Xa to inhibit its activity. To test out this possibility, the protein (0.2 mg/mL) was incubated with a peptide (HA-IEGR-C7C: YDYDVPDYAAIEGRCQVINKNSC at 50 μM) and factor Xa (2 μg) in 50 μL. The peptide got cleaved in less than 3 hrs, while the protein was still intact. This result suggested that the enzyme was still active and the IEGR sequence may not be easily accessible for factor Xa cleavage when Trx was in reduced form. The protein was then oxidized by 0.5 mM H_2O_2 for 2 hrs according to a previous report[11] and the oxidized Trx was cleaved overnight by factor Xa. LC/MS showed obvious peak for cleaved products although the spectrum was messy (data not

shown). We can conclude that oxidation of Trx with H_2O_2 for 2 hrs will generate dimer and His-tag from the dimer can be cleaved by factor Xa. However, this oxidation condition is too harsh which might destroy its activity. We need to find a milder condition to oxidize Trx.

3.2.10.3 Oxidation of Trx by GSSG

Glutathione disulfide (GSSG) can be used as a mild oxidative species for cysteine. In fact, most of Trx was oxidized by 1 mM GSSG, based on deconvoluted mass. Two disulfide bonds were observed with an extra mass of GSH, indicating only one intramolecular disulfide bond was formed besides another disulfide bond with GSH. Cleavage of the oxidized protein was mostly completed after overnight incubation. The mass for cleaved protein [M] was detected after de-convolution as well as [M + GSH]. After TCEP reduction, LC/MS results showed incomplete cleavage of the IEGR sequence. This mixture was purified by Ni-NTA agarose to yield a clean protein.

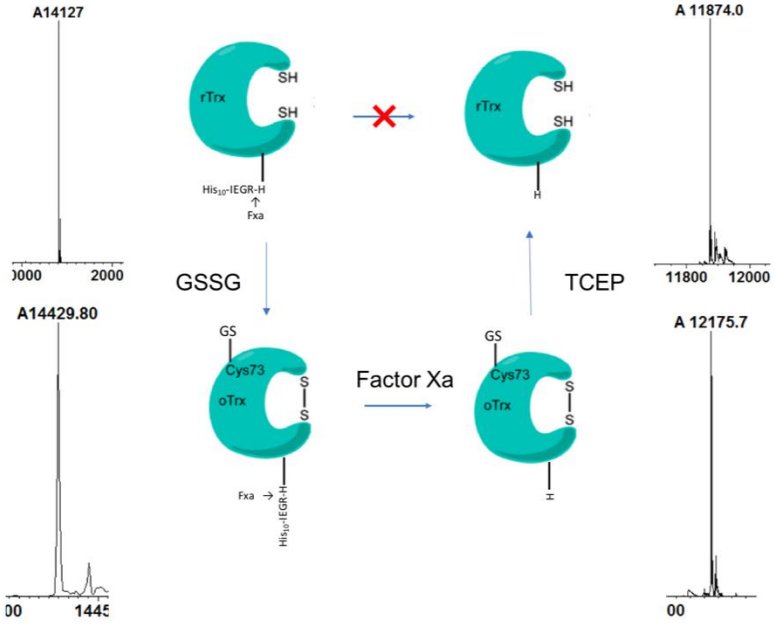


Figure 3.2.15. Scheme of WT-Trx cleavge by factor Xa. The protein has to be oxidized to reveal the N-terminus for enzyme cleavage

3.2.10.4 Expression of Trx with N-terminus Cysteine (Cys-Trx)

To install an N-terminal cysteine in Trx, the histidine after IEGR was mutated to cysteine through a Q5 site-directed mutagenesis kit. Detailed procedures are included at the end of this section.

Protein sequence after mutation:

GHHHHHHHHHHSSGHIEGRCMVKQIESKTAFQEALDAAGDKLVVVD
FSATWCGPCKMIKPFFHSLSEKYSNVIFLEVDVDDCQDVASECEVKCMPTFQF
FKKGQKVGEFSGANKEKLEATINELV

3.2.10.5 Labelling of Cys-Trx

Cleavage of His-tag and the IEGR sequenced follows the protocol described above. The concentration of the protein was diluted to 10 μM and labelled by KL42 at 50 μM for 2 hrs at pH 6.0. The mixture was diluted and subjected to LC/MS analysis, which showed complete labelling of Cys-Trx. Importantly, no side reactions were observed meaning that none of the other cysteine residues were modified by KL42. This observation further supports the specificity of our chemistry.

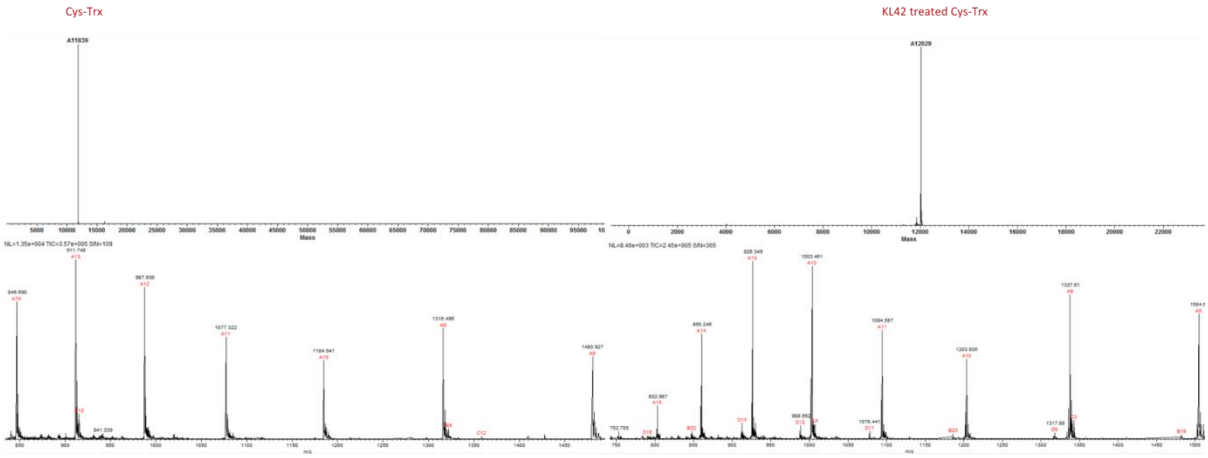


Figure 3.2.16. Cys-Trx was completed labelled by KL42

3.2.10.6 Effect of N-terminal modification of Cys-Trx

Ideally, the modification of Trx should minimally disturb the structure and function of the original protein. To examine the effect of KL42 modification on Trx, we adopted a reported protocol to compare the catalytic activity of Trx before and after modification.[12] Briefly, the catalytic activity of Trx is measured by the reduction of insulin in presence of DTT. When the disulfide bond from insulin gets reduced, the protein will aggregate and bind to the detection dye. The activity of Trx can be determined by the fluorescent intensity after a certain period of time (half an hour in this case). We were pleased to see no significant change in the catalytic activity of Trx after KL42 modification.

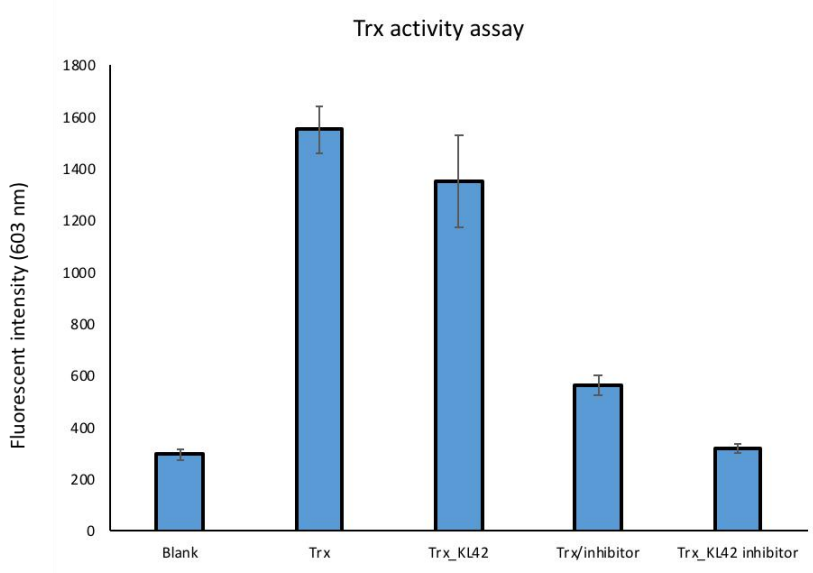


Figure 3.2.17. Modification of KL42 showed no significant effect on the activity of Trx

3.2.10.7 Expression of Azo-reductase protein with N-terminus Cysteine

The labelling efficiency of KL42 was demonstrated on a second protein: Azo reductase. pET28a-TEV-Cys-AzoR was obtained from Professor Hua Lu, Peking University. The protein has a sequence GMENLYFQC at its N-terminus for Tev protease cleavage. In order to accelerate the cleavage process, the sequence after N-terminal cysteine residue was mutated to Cys-Ile-Ser. The mutated protein was expressed in a similar manner described above.

Primer 1: ATTTCTGGTATGAGCAAGGTATTAG

Primer 2: ACATTGGAAGTACAGGTTC

Original sequence:

GMENLYFQCGMSKVLVLKSSILAGYSQSNQLSDYFVEQWREKHSADE
ITVRDLAANPIPVLDGELVGALRPSDAPLTPRQQEALALSDELIAELKAHDVIV
IAAPMYNFNISTQLKNYFDLVARAGVTFRYTENGPEGLVTGKKAIVITSRGGI

HKDGPTDLVTPYLSTFLGFIGITDVKFVFAEGIAYGPEMAAKAQSDAKAAIDSI VSAHHHHHH

Mutated sequence:

GMENLYFQCISGMSKVLVLKSSILAGYSQSNQLSDYFVEQWREKHSA DEITVRDLAANPIPVLDGELVGALRPSDAPLTPRQQEALALSDELIAELKAHD VIVIAAPMYNFNISTQLKNYFDLVARAGVTFRYTENGPEGLVTGKKAIVITSR GGIHKDGPTDLVTPYLSTFLGFIGITDVKFVFAEGIAYGPEMAAKAQSDAKAA IDSIVSAHHHHHH

3.2.10.8 Labelling of Azo-reductase protein with KL42 and KL62

The short peptide before the cysteine residue was cleaved by TEV protease. The protein after cleavage was diluted to 10 μM and labelled not only by KL42, but also by a biotin derivative KL62 (both at 50 μ M). LC/MS analysis suggests a complete conversion of Cys-AzoR to its corresponding products.

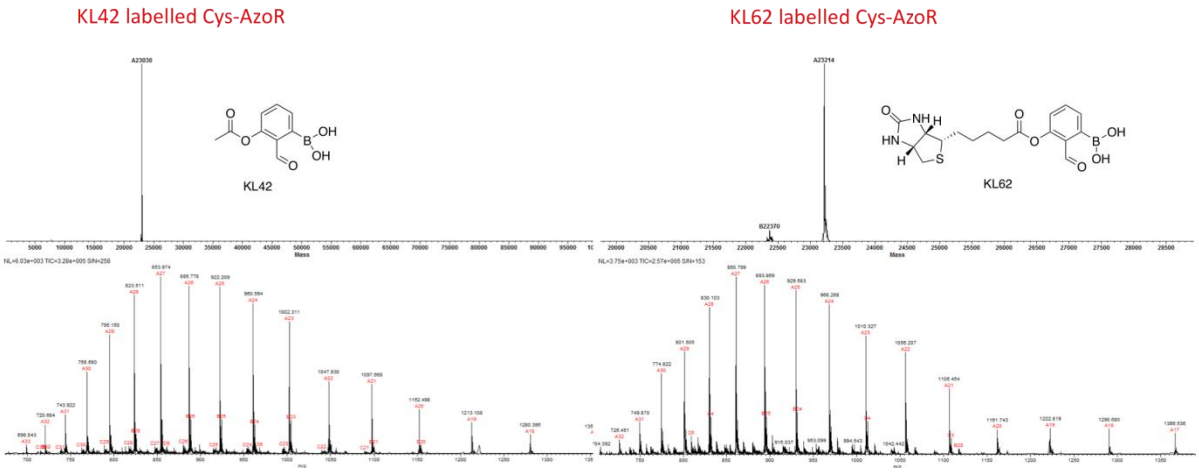


Figure 3.2.18. Cys-AzoR was efficiently labeled by KL42 and KL62

3.2.10.9 Gel experiment with KL72 labelling with Azo reductase

One application of protein labelling is to fix a biotin tag to pull down the protein of interest with streptavidin. We then carried out a gel shift analysis to confirm the biotinylation of Cys-AzoR after treatment with KL62. In a SDS-PAGE gel, the KL62 modified Cys-AzoR should form a complex with streptavidin. This complex have a significantly higher molecular weight than streptavidin or Cys-AzoR alone. The result is shown in **Figure 3.2.19a**, Gel 1. Unfortunately, there was no significant shift of protein bands in lane 6, meaning that there was no biotin-streptavidin complex formation. One possibility is that Cys-AzoR shields the biotin structure on its N-terminus, making it difficult to bind to streptavidin. A solution to this problem is to increase the distance of biotin and Cys-AzoR so it is exposed for binding. To this end, we synthesized another derivative named KL72 by adding a PEG linker between the biotin structure and 2-FPBA. This new molecule can also completely modify Cys-AzoR at 50 μM with no side product (**Figure 3.2.19b**). Excitingly, the KL72-Cys-AzoR conjugate traveled in the gel as streptavidin complexes, as is shown in **Figure 3.2.19a**, Gel 2.

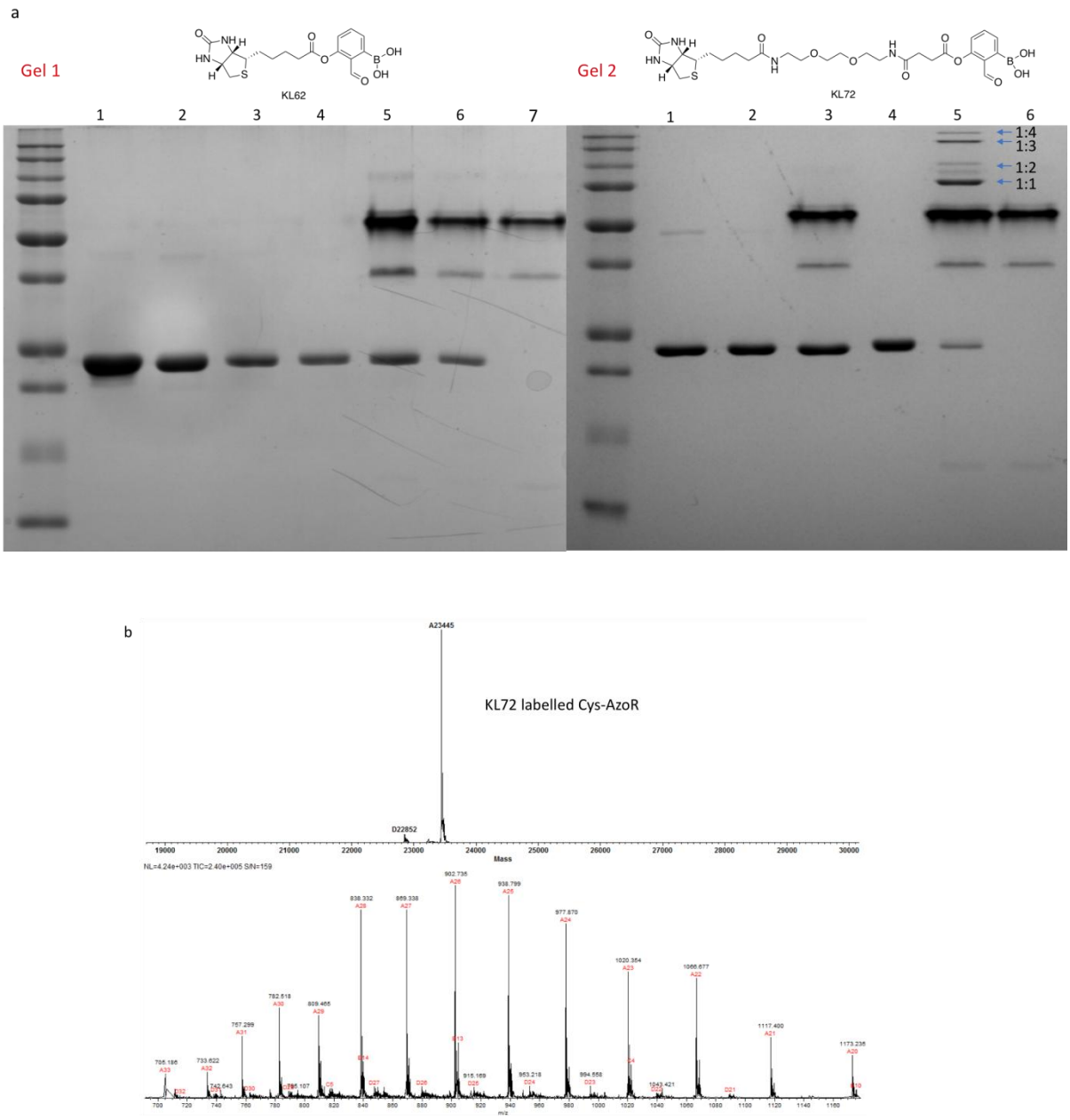


Figure 3.2.19. a) Gel analysis for Cys-AzoR biotinylation; b) Cys-AzoR modified by KL72. Gel 1 assignment: Lane 1: uncleaved Cys-AzoR; Lane 2: cleaved but unmodified Cys-AzoR; Lane 3: KL42 labelled Cys-AzoR; Lane 4: KL62 labelled Cys-AzoR; Lane 5: KL42 labelled Cys-AzoR mixed with streptavidin; Lane 6: KL62 labelled Cys-AzoR mixed with streptavidin; Lane 7: streptavidin alone. Gel 2 assignment: Lane 1: unmodified Cys-AzoR; Lane 2: KL42 labelled Cys-AzoR; Lane 3: KL42 labelled Cys-AzoR mixed with streptavidin; Lane 4: KL72 labelled Cys_Azo; Lane 5: KL72 labelled Cys_Azo mixed with streptavidin; Lane 6: streptavidin alone.

3.2.10.10 Comparison of KL42 and thioester for Cys-AzoR modification

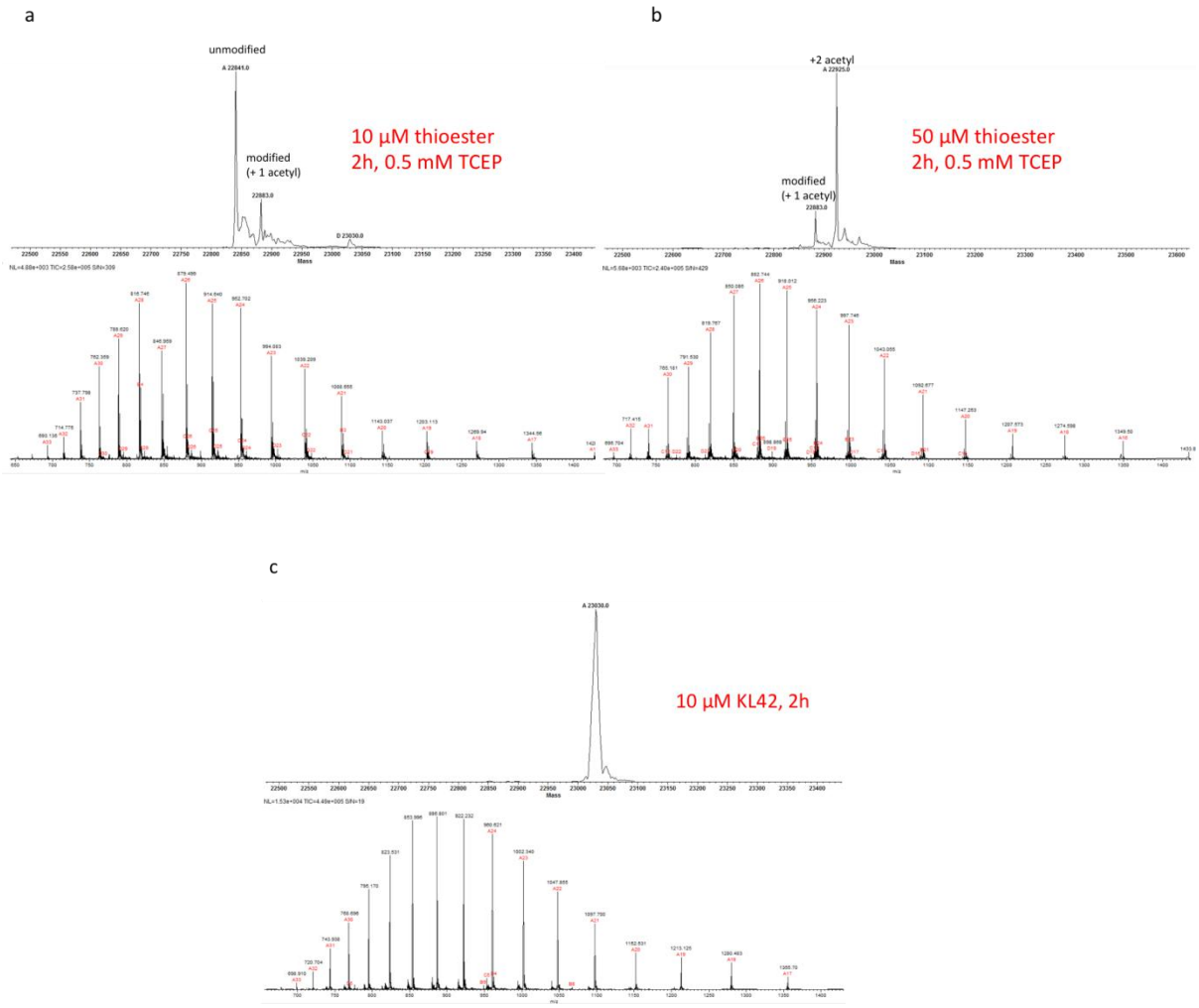


Figure 3.2.20. Cys-AzoR was treated with a) 10 µM thioester; b) 50 µM thioester; c) 10 µM KL42. The protein could not be completely labeled by thioester at 10 µM and increasing the concentration of thioester to 50 µM resulted in non-specific labeling on internal cysteine residues. KL42 efficiently labeled the protein at 10 µM

A comparative study between KL42 and a thioester was carried out on Cys-AzoR. Not surprisingly, KL42 completely labels Cys-AzoR at 1 eq. (10 µM, **Figure 3.2.20c**). On the contrary, there was a significantly amount of unmodified protein if only 1 eq. of thioester was used (**Figure 3.2.20a**). Increasing the concentration of thioester completely consumed all the protein. However, we also observed a side

product with an additional acetyl group compared to the desired product. This is consistent with the previous experiment in which the thioester also modifies internal cysteine residues. Taken together, we can conclude that our chemistry performs better at all the concentrations that have been tested. It provides a more complete conversion at single digit micromolar concentration and shows no side reaction when used at excess amounts. It is overall a superior choice over NCL or even CBT for N-terminal cysteine modification.

3.2.11 Modification on phage

(This assay was carried out by Wenjian Wang)

Phage display is a powerful technique for discovering peptide probes and inhibitors. However, it has been largely limited to natural amino acids. Encoding unnatural amino acid into phage is still under development. To this end, modifications on a phage library to increase the diversity has become a hot topic. We hypothesized that if we can prepare a phage library that contains an N-terminal cysteine and an internal cysteine, we would be able to build a library that has two distinct modifications utilizing our chemistry.

To confirm our hypothesis, we prepared a phage library that contains a factor Xa cleavage site at the N-terminus of the PIII protein, followed by a $C(X)_7C$ sequence (X can be any of the 20 amino acids). After factor Xa cleavage, the disulfide bond was reduced by iTCEP to reveal the cysteine residues. The resulting phage were then treated with KL72 followed by an ELISA assay. To our delight, the results show that KL72 completely modified the N-terminus of the PIII protein, compared to the control (BIA: biotin-iodoacetamide). Importantly, if the phage was pre-treated with KL42, which did not contain a biotin group, the ELISA signal was then greatly

demolished. Capping the N-terminus with an acetyl group also eliminated the ELISA signal for KL72 treated sample. These results further support the fact that our chemistry is highly site-specific.

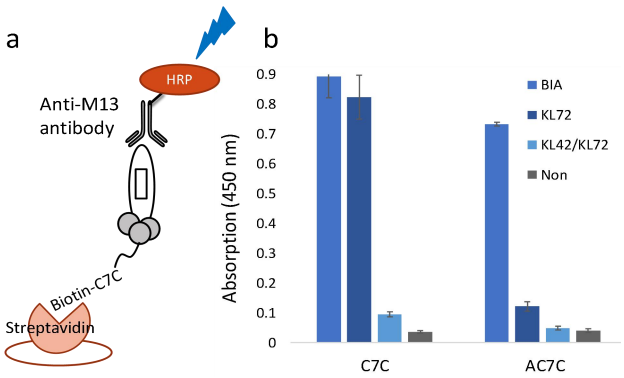


Figure 3.2.21. a) Illustration the ELISA experimental design; b) ELISA results showing KL72 modifies NCys-bearing phage with high efficiency and specificity.

3.2.12 Conclusions

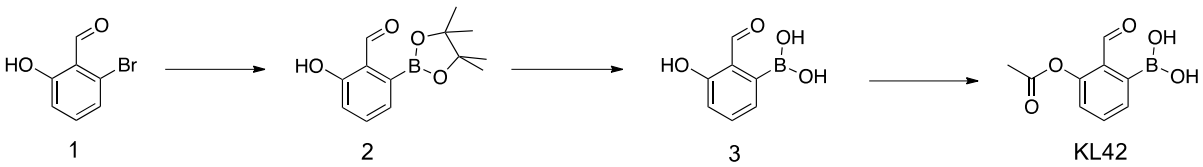
To summarize, we presented here a fast and site-specific reaction that can be applied to protein modification. This new conjugation starts with a rapid formation of TzB intermediate and goes through intramolecular acyl transfer to give a stable product. In fact, the conjugate stayed intact in SDS-PAGE or in HPLC conditions. In all cases, this TzB mediated acyl transfer showed superior selectivity and kinetics to N-terminal cysteine compared to NCL and CBT condensation. Furthermore, we applied this chemistry to phage on the PIII protein and created a library with two distinct chemical modifications. We believe this chemistry will find its use in other biological applications as well.

We also want to expand the scope of this chemistry to cysteine analogous such as N-terminal serine or threonine. These two amino acids are structurally similar to cysteine, presenting a hydroxyl group instead of a thiol group as the sidechain. They

are presumably less reactive towards modifications but more stable in physiological conditions. New reactions on these amino acids will provide more options to site-specifically modify proteins.

3.2.13 General information

3.2.13.1 Synthesis of KL42



Scheme 3.2.1. synthetic route for KL42

2-Hydroxy-6-(4,4,5,5-tetramethyl-1,3,2-dioxaborolan-2-yl)benzaldehyde (**2**)

1 (1.2 g, 6 mmol), B_2Pin_2 (3.82 g, 15 mmol), $Pd(dppf)Cl_2$ (438 mg, 0.6 mmol) and KOAc(1.77 g, 18 mmol) were suspended in 10 mL dioxane. The solution was bubbled with argon for 15 min and then heated at 80 °C. The reaction was quenched after 1 hr with 50 mL of water and the resulting mixture was extracted with EtOAc (50 mL × 3). The organic layers were collected, washed by brine (50 mL) and dried over sodium sulfate. Then the solvent was evaporated and the crude product was purified using silica chromatography (3% EtOAc/Hexanes) to yield 942 mg of product as a white solid (yield: 64%).

^{1}H-NMR (500 MHz, $CDCl_3$) δ 12.00 (s, 1H), 10.64 (s, 1H), 7.54 – 7.38 (m, 2H), 7.03 (dd, J = 6.6, 3.0 Hz, 1H), 1.35 (s, 12H).

^{13}C-NMR (126 MHz, $cdcl_3$) δ 199.89, 162.58, 135.72, 128.03, 124.02, 120.75, 120.47, 84.45, 24.80.

HRMS (TOF-ESI$^+$) m/z calculated for $C_{13}H_{18}BO_4$ $[M+H]^+$ 249.1293, found 249.1251.

2-Formyl-3-hydroxy-phenylboronic acid (**3**)

2 (250 mg, 1 mmol) was dissolved in 1 mL THF and 0.4 mL water, and then $NaIO_4$ (635 mg, 3 mmol) was added. The resulting mixture was stirred for 15 min before the addition of 1 mL HCl (1 N). The resulting solution was left stirring for 4 hr and extracted using EtOAc (30 mL × 3). The organic layers were combined, washed by brine (30 mL) and dried over sodium sulfate. Solvent was evaporated to ~2 mL and then 20 mL hexane was added to precipitate the product. A white solid was obtained (150 mg, 65% yield) and used without further purification.

^{1}H-NMR (600 MHz, acetone-d6) δ 11.72 (s, 1H), 10.48 (d, *J* = 0.7 Hz, 1H), 7.75 (s, 1H), 7.53 (dd, *J* = 8.3, 7.2 Hz, 1H), 7.30 (dd, *J* = 7.2, 1.1 Hz, 1H), 6.97 (dt, *J* = 8.4, 0.8 Hz, 1H).

^{13}C-NMR (151 MHz, acetone-d6) δ 201.77, 164.91, 138.48, 128.16, 125.94, 121.07.

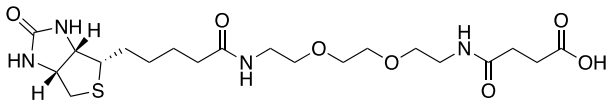
HRMS (TOF-ESI$^+$) m/z calculated for $C_7H_6BO_3$ $[M-H_2O+H]^+$ 149.0405, found 149.0368.

3-Acetoxy-2-formyl-phenylboronic acid (KL42)

3 (33 mg, 0.2 mmol) was mixed with 0.1 mL of acetic anhydride and 0.1 mL of DIPEA. The mixture was stirred overnight and purified by RP-HPLC to yield a yellowish product after lyophilization (38 mg, 93% yield). According to ^{1}H-NMR (Figure S1a), KL42 exists as a dynamic mixture of aldehydes and boroxoles. Hence, ^{13}C-NMR was not recorded and the purity and integrity of the compound is confirmed by LC-MS (Figure S1b).

HRMS (TOF-ESI$^+$) m/z calculated for $C_9H_8BO_4$ $[M-H_2O+H]^+$ 191.0510, found 191.0468

3.2.13.2 Synthesis of KL72

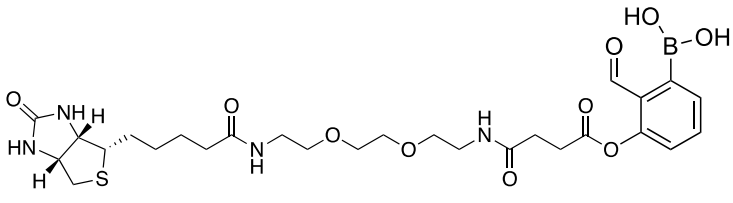

(Biotin-PEG-COOH)

Biotin-OSu (17 mg, 0.05 mmol) was added to a solution of Amino-PEG2-amine (8 mg , 0.054 mmol in 200 µM DMF). To the resulting solution was added 43 µL of DIPEA. The reaction mixture was vigorously shaken on a shaker and sonicated every hour to break the gel that formed. After three hours, 7.5 mg of succinic anhydride was added to the solution and the mixture was shaken for another three hours. The product was purified on RP-HPLC to yield 12 mg of product (50%) as a white solid after lyophilization.

^{1}H-NMR (500 MHz, CD_3OD) δ 4.50 (ddd, *J* = 7.8, 5.0, 0.9 Hz, 1H), 4.32 (dd, *J* = 7.9, 4.5 Hz, 1H), 3.63 (s, 4H), 3.56 (td, *J* = 5.5, 3.1 Hz, 4H), 3.37 (td, *J* = 5.6, 1.9 Hz, 4H), 3.22 (ddd, *J* = 8.9, 5.9, 4.5 Hz, 1H), 2.94 (dd, *J* = 12.8, 5.0 Hz, 1H), 2.71 (d, *J* = 12.7 Hz, 1H), 2.60 (td, *J* = 7.0, 0.9 Hz, 2H), 2.54 – 2.46 (m, 2H), 2.23 (t, *J* = 7.4 Hz, 2H), 1.80 – 1.55 (m, 4H), 1.45 (p, *J* = 7.8 Hz, 2H).

^{13}C-NMR (126 MHz, CD_3OD) δ 174.77, 173.21, 164.69, 69.92, 69.87, 69.19, 61.97, 60.25, 55.56, 39.60, 38.97, 38.88, 35.32, 30.10, 28.86, 28.32, 28.07, 25.42.

HRMS (TOF-ESI$^+$) m/z calculated for $C_{20}H_{35}N_4O_7S$ $[M+H]^+$ 475.2221, found 475.2181.

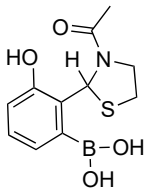


(KL72)

Biotin-PEG-COOH (12 mg, 0.025 mmol) was mixed with a DMF solution (100 µL) of EDC hydrochloride (4.8mg, 0.025 mmol) and HOBt (3.4 mg, 0.025 mmol). 18 µL DIPEA and 6.2 mg of Compound **3** (Figure S1) was added to the mixture and the reaction was stirred for 7 h. The product was purified by RP-HPLC to yield 5 mg of product (32% yield) as a white solid. Due to the dynamic formation of boroxoles similar o KL42, the purity and integrity of KL72 was confirmed by LC-MS (Figure S1c).

HRMS (TOF-ESI$^+$) m/z calculated for $C_{27}H_{38}BN_4O_9S$ $[M-H_2O+H]^+$ 605.2447, found 605.2449.

3.2.13.3 Synthesis of KL44 variants for crystallization

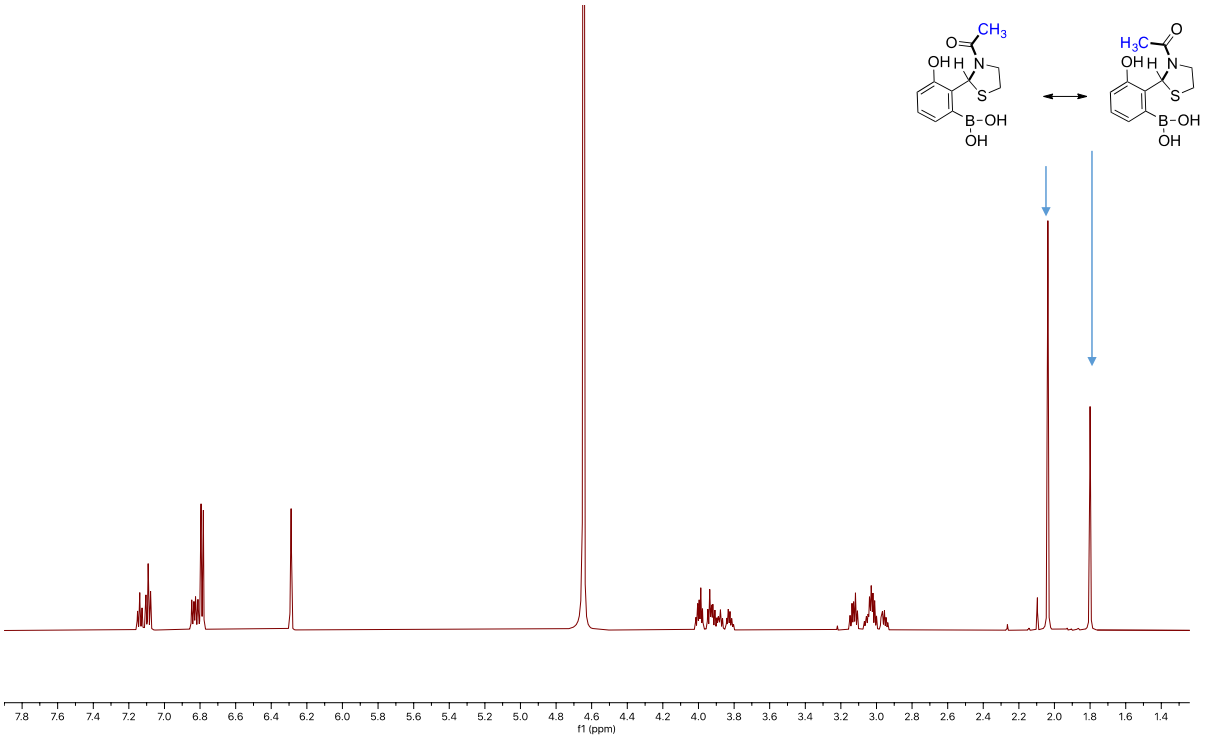


(KL44^{P2})

KL42 (12 mg, 0.058 mmol) was mixed with P2 (cysteamine) (6.6mg, 0.087 mmol) in 50 mL PBS (pH 6.0). The solution was concentrated and the mixture was purified by RP-HPLC to generate 9 mg of product as a white solid (58% yield).

HRMS (TOF-ESI$^+$) m/z calculated for $C_{11}H_{15}BNO_4S$ $[M+H]^+$ 268.0809, found 268.0819.

^{1}H-NMR spectrum (Figure S5) indicates the product consists of *cis* and *trans* isomers of the amide bond. For crystallization, KL44^{P2} (5 mg) was dissolved in 1:1 acetonitrile/water mixture (50 µL) in a 12×35 mm glass vial with a loose cap. Crystals formed after overnight. The solved structure (Figure 2c, Table S1-2) revealed the isomer with cis-amide bond only, which crystallized as a racemic mixture.



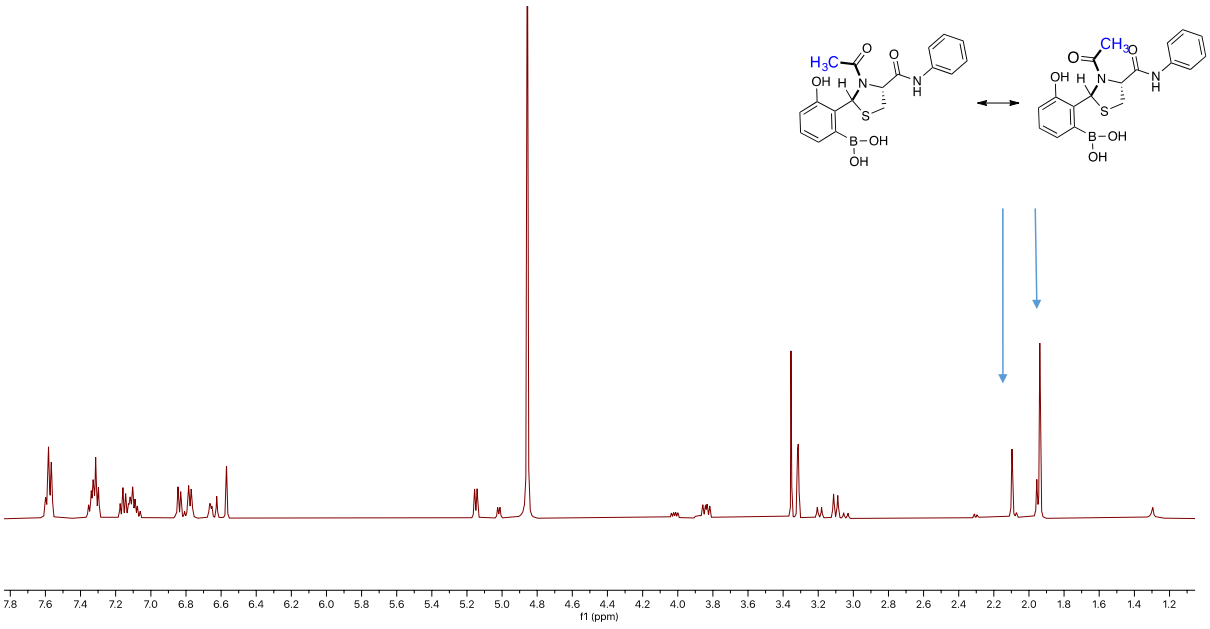
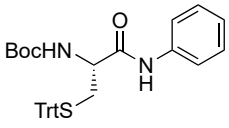


Figure 3.2.22. ^{1}H-NMR spectra of KL44^{P2} and KL44^{P3}. The highlighted peaks (acetyl protons) indicate that these conjugates exist as a pair of isomers in solution.



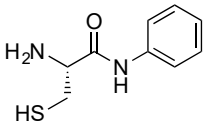
(Boc-Cys(Trt)-aniline)

Boc-Cys(Trt)-OH (232 mg, 0.5 mmol) and N,N-Disuccinimidyl carbonate (128 mg, 0.5 mmol) was dissolved in 2 mL CH_3CN, to which was added 80 µL pyridine. The mixture was stirred for 3 hr and the solution became clear. Aniline (69 mg, 1 mmol) was added and the reaction was stirred at room temperature overnight. After removal of CH_3CN, the residue was treated with 0.1 N HCl and extracted immediately with EtOAc (30 mL × 3). The organic layers were combined, washed by brine (30 mL) and dried over sodium sulfate. After solvent removal, the product was purified by silica column (20% EtOAc/Hexanes) to yield 210 mg product as a white powder (78% yield).

^{1}H-NMR (600 MHz, $CDCl_3$) δ 8.22 (s, 1H), 7.46 (dd, *J* = 7.5, 1.9 Hz, 8H), 7.39 – 7.18 (m, 11H), 7.08 (q, *J* = 8.0, 7.4 Hz, 1H), 5.02 (d, *J* = 7.6 Hz, 1H), 4.01 (s, 1H), 2.73 (ddd, *J* = 63.4, 13.2, 6.4 Hz, 2H), 1.44 (s, 9H).

^{13}C-NMR (151 MHz, $CDCl_3$) δ 171.52, 158.54, 147.07, 140.19, 132.25, 131.55, 130.75, 130.73, 130.63, 129.57, 127.03, 122.61, 122.54, 83.28, 70.01, 56.98, 36.15, 32.37, 30.96.

HRMS (TOF-ESI^{+}) m/z calculated for $C_{33}H_{34}N_2NaO_3S$ $[M+Na]^+$ 561.2182, found 561.2142.



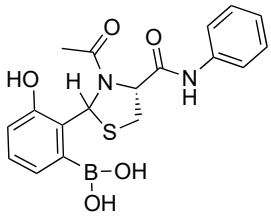
(P3)

Boc-Cys(Trt)-aniline (58 mg, 0.11 mmol) was suspended in 2 mL TFA/DCM/TIPS (5:4:1) cocktail. After stirring for 2 hr, TFA and DCM were removed by vacuum. The residue was dissolved in 2 mL DCM and 13 mL hexane was added to the solution to precipitate the product. The resulting suspension was centrifuged at 7000 g for 15 min and the clear supernatant was discarded. The crude product was purified by RP-HPLC and lyophilization gave the desired product as a colorless oil. (26 mg, 77% yield calculated assuming product exists as a TFA salt).

^{1}H-NMR (500 MHz, CD_3OD) δ 7.65 – 7.55 (m, 2H), 7.40 – 7.26 (m, 2H), 7.20 – 7.09 (m, 1H), 4.16 (dd, J = 7.2, 5.1 Hz, 1H), 3.22 – 2.95 (m, 2H).

^{13}C-NMR (126 MHz, CD_3OD) δ 165.17, 137.48, 128.56, 124.55, 119.91, 55.43, 24.90.

HRMS (TOF-ESI$^+$) m/z calculated for $C_9H_{13}N_2OS$ $[M+H]^+$ 197.0743, found 197.0721.



(KL44^{P3})

KL42 (24 mg, 0.116 mmol) was mixed with P3 (26mg, 0.083 mmol) in 2 mL PBS (pH 6.0) and 2 mL acetonitrile. After 2 hr incubation, the reaction mixture was directly subjected to RP-HPLC purification. The product was obtained as a white powder after lyophilization (18 mg, 56% yield).

HRMS (TOF-ESI$^+$) m/z calculated for $C_{18}H_{20}BN_2O_5S$ $[M+H]^+$ 387.1180, found 387.1161

^{1}H-NMR spectrum (Figure 2.4.22) indicates the product consists of *cis* and *trans* isomers of the amide bond.

3.2.13.4 Crystallographic information

For crystallization, KL44^{P3} (2.5 mg) was dissolved in 4:1 acetonitrile/water mixture (100 μL) in a 12×35 mm glass vial with a loose cap. Crystals were obtained after overnight. The solved structure revealed the isomer with cis-amide bond only, which crystallized as a racemic mixture.

Table 3.2.2. Crystal data and structure refinement for KL44^{P2}.

Identification code	C11H14BNO4S	
Empirical formula	C11 H14 B N O4 S	
Formula weight	267.10	
Temperature	173(2) K	
Wavelength	0.71073 ≈	
Crystal system	Triclinic	
Space group	P-1	
Unit cell dimensions	a = 8.2933(10) ≈	a= 72.920(4)∞.
	b = 9.9741(9) ≈	b= 80.424(4)∞.
	c = 15.6270(18) ≈	g = 79.918(3)∞.
Volume	1207.4(2) ≈^{3}	
Z	4	
Density (calculated)	1.469 Mg/m^3	

Absorption coefficient	0.273 mm^{-1}
F(000)	560
Crystal size	0.600 x 0.210 x 0.140 mm^3
Theta range for data collection	2.155 to 28.433∞.
Index ranges	-11<=h<=11, -13<=k<=13, -20<=l<=20
Reflections collected	29150
Independent reflections	6051 [R(int) = 0.0539]
Completeness to theta = 25.242∞	100.0 %
Absorption correction	Semi-empirical from equivalents
Max. and min. transmission	0.7457 and 0.6869
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	6051 / 6 / 345
Goodness-of-fit on F^2	1.045
Final R indices [I>2sigma(I)]	R1 = 0.0436, wR2 = 0.0940
R indices (all data)	R1 = 0.0636, wR2 = 0.1020
Extinction coefficient	n/a
Largest diff. peak and hole	0.445 and -0.511 e.≈$^{-3}$

Table 3.2.3. Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters (≈^{2}x 10^3) for $KL44^{P2}$. U(eq) is defined as one third of the trace of the orthogonalized U^{ij} tensor.

	x	y	z	U(eq)
S(1)	-967(1)	6765(1)	3826(1)	22(1)

O(1)	2549(2)	3078(1)	3462(1)	22(1)
O(2)	-2406(2)	5357(2)	2669(1)	24(1)
O(3)	4814(2)	5543(2)	1945(1)	26(1)
O(4)	3379(2)	7750(1)	2034(1)	27(1)
N(1)	1943(2)	5289(2)	3634(1)	17(1)
C(1)	4463(2)	3814(2)	4160(1)	27(1)
C(2)	2914(2)	4020(2)	3731(1)	18(1)
C(3)	2162(2)	6383(2)	4055(1)	23(1)
C(4)	821(2)	7608(2)	3772(1)	25(1)
C(5)	325(2)	5428(2)	3332(1)	16(1)
C(6)	362(2)	5767(2)	2324(1)	14(1)
C(7)	-1093(2)	5656(2)	2024(1)	17(1)
C(8)	-1188(2)	5856(2)	1119(1)	22(1)
C(9)	193(2)	6169(2)	504(1)	22(1)
C(10)	1629(2)	6294(2)	788(1)	21(1)
C(11)	1741(2)	6118(2)	1702(1)	16(1)
B(1)	3379(2)	6453(2)	1934(1)	19(1)
S(2)	10034(1)	572(1)	1292(1)	28(1)
O(5)	5231(2)	3136(2)	1577(1)	42(1)
O(6)	10223(2)	2459(1)	2510(1)	25(1)
O(7)	3879(2)	401(2)	3057(1)	25(1)
O(8)	5660(2)	-1575(1)	2912(1)	24(1)
N(2)	6801(2)	1159(2)	1391(1)	19(1)
C(12)	4133(3)	1831(3)	806(2)	46(1)
C(13)	5418(2)	2108(2)	1277(1)	26(1)

C(14)	7268(3)	53(2)	913(1)	29(1)
C(15)	8883(3)	-745(2)	1219(2)	32(1)
C(16)	8157(2)	1521(2)	1743(1)	18(1)
C(17)	8001(2)	1161(2)	2758(1)	17(1)
C(18)	9126(2)	1686(2)	3113(1)	18(1)
C(19)	9101(2)	1419(2)	4036(1)	23(1)
C(20)	7987(2)	582(2)	4613(1)	23(1)
C(21)	6905(2)	33(2)	4272(1)	21(1)
C(22)	6866(2)	332(2)	3341(1)	18(1)
B(2)	5451(3)	-261(2)	3058(1)	19(1)

Table 3.2.4. Crystal data and structure refinement for KL44[P3].

Identification code	C18H19BN2O5S	
Empirical formula	C20 H22 B N3 O5 S	
Formula weight	427.27	
Temperature	173(2) K	
Wavelength	1.54178 ≈	
Crystal system	Monoclinic	
Space group	$P2_1/n$	
Unit cell dimensions	a = 13.6382(3) ≈	a= 90∞.
	b = 6.7051(2) ≈	b= 104.146(2)∞.
	c = 26.3710(7) ≈	g = 90∞.
Volume	2338.38(11) ≈^{3}	

Z	4
Density (calculated)	1.214 Mg/m^3
Absorption coefficient	1.516 mm^{-1}
F(000)	896
Crystal size	0.380 x 0.120 x 0.100 mm^3
Theta range for data collection	3.366 to 66.563∞.
Index ranges	-16<=h<=15, -7<=k<=7, -31<=l<=31
Reflections collected	40489
Independent reflections	4063 [R(int) = 0.0826]
Completeness to theta = 66.563∞	98.7 %
Absorption correction	Semi-empirical from equivalents
Max. and min. transmission	0.7528 and 0.6714
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	4063 / 3 / 284
Goodness-of-fit on F^2	1.044
Final R indices [I>2sigma(I)]	R1 = 0.0487, wR2 = 0.1204
R indices (all data)	R1 = 0.0860, wR2 = 0.1517
Extinction coefficient	n/a
Largest diff. peak and hole	0.372 and -0.307 e.$\approx^{-3}$

Table 3.2.5. Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters ($\approx^2$x 10^3) for $KL44^{P3}$. U(eq) is defined as one third of the trace of the orthogonalized U^{ij} tensor.

	x	y	z	U(eq)

S(1)	6053(1)	3976(1)	7149(1)	46(1)
O(1)	5444(2)	9629(3)	5995(1)	42(1)
O(2)	8467(2)	4280(3)	7415(1)	38(1)
O(3)	6203(2)	6361(3)	5305(1)	44(1)
O(4)	5566(2)	3264(3)	5474(1)	42(1)
O(5)	4652(2)	8599(3)	6996(1)	44(1)
N(1)	5944(2)	6799(3)	6449(1)	33(1)
N(2)	3276(2)	7222(4)	6444(1)	44(1)
B(1)	6313(3)	4675(5)	5592(1)	36(1)
C(1)	7207(2)	9334(5)	6403(1)	44(1)
C(2)	6140(2)	8621(4)	6265(1)	36(1)
C(3)	4903(2)	6099(4)	6370(1)	36(1)
C(4)	4955(2)	4001(5)	6603(1)	42(1)
C(5)	6739(2)	5641(4)	6815(1)	32(1)
C(6)	7472(2)	4615(4)	6556(1)	31(1)
C(7)	8393(2)	4030(4)	6895(1)	33(1)
C(8)	9165(2)	3250(4)	6704(1)	40(1)
C(9)	9034(2)	3004(5)	6170(1)	46(1)
C(10)	8120(2)	3521(5)	5833(1)	42(1)
C(11)	7329(2)	4305(4)	6018(1)	33(1)
C(12)	4285(2)	7462(4)	6634(1)	36(1)
C(13)	2501(2)	8033(5)	6661(1)	42(1)
C(14)	2562(2)	9898(5)	6876(1)	47(1)
C(15)	1789(3)	10569(6)	7089(1)	55(1)
C(16)	977(3)	9382(6)	7092(2)	62(1)

C(17)	911(3)	7531(7)	6870(2)	79(1)
C(18)	1668(3)	6842(6)	6650(2)	65(1)
C(19)	2938(5)	271(8)	5314(2)	106(2)
C(20)	2780(4)	2184(9)	5554(2)	90(2)
N(3)	2690(3)	3671(8)	5743(2)	103(2)

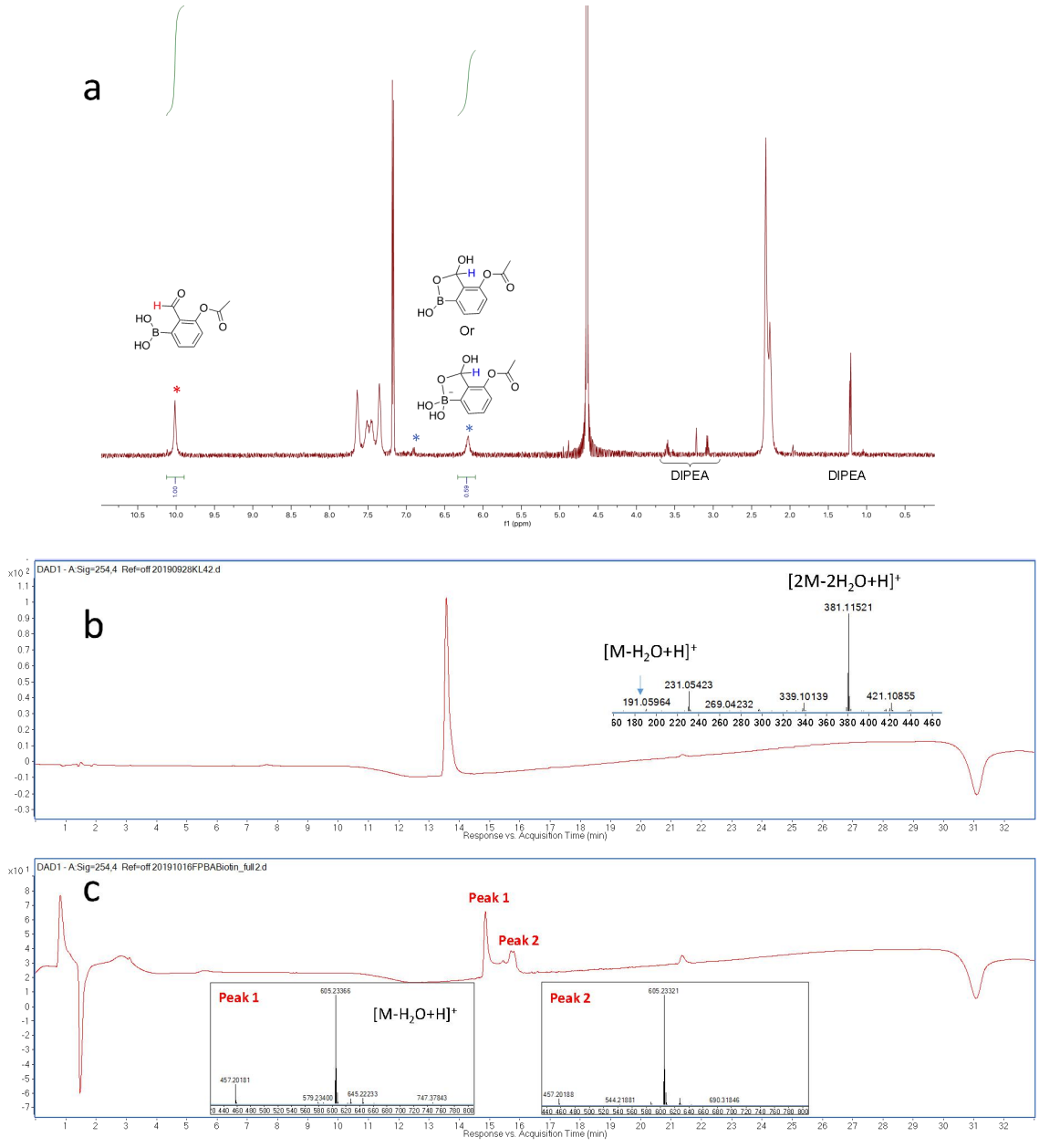


Figure 3.2.23. Characterization data of KL42 and KL72. a) ^{1}H-NMR spectrum of KL42 in water showing KL42 exists as a mixture of isomers. b) LC-MS data confirming the purity and integrity of KL42. c) LC-MS data confirming the purity and integrity of KL72. The LC traces were recorded by monitoring absorbance at 254 nm. Although two peaks were observed in c), the two peaks give identical m/z readout, and presumably represent isomeric forms of KL72.

3.2.13.5 Peptide synthesis

Peptide synthesis was conducted on a peptide synthesizer using Rink amide resin (0.05 mmol) and the crude peptides were purified using RP-HPLC. For fluorophore labelled peptides, namely P1* (CLA*), P4 (CGA*), P5 (CWA*), and P6 (CPA*), Fmoc-Dap(Alloc)-OH residue was used to introduce an Alloc-protected Dap residue on the C-terminus. Before the removal of the last Fmoc, Alloc was orthogonally removed using a mixture of tetrakis triphenylphosphine palladium (50 mg) and phenylsilane (0.3 mL) in DCM (2 mL). Then the deprotected Dap was allowed to couple with 5-carboxyfluorescein (2 eq) with HBTU activation. These peptides were purified by HPLC and stored as a solid to prevent cysteine oxidation. Stock solutions were made fresh before each experiment and the concentration was determined with absorbance at 495 nm.

3.2.13.6 Trx expression

A pET16b vector encoding wild type thioredoxin (WT-Trx) in *E. coli* (BL21) was obtained from professor Eranthie Weerapana. A stab of frozen glycerol stock was injected to 5 mL of LB for overnight growth. To inhibit the growth other pathogens, 0.1 mg/mL ampicillin was added to the LB in advance. The 5 mL overnight growth as poured into 500 mL LB containing 0.1 mg/mL ampicillin and incubate at 37 degree until OD=0.6-0.8. 1 mL was collected for Gel analysis. Protein expression was induced with 0.5 mL of 0.6 M IPTG (final concentration 0.6 mM). The flask was kept

at R.T. overnight. Another 1 mL was collected for Gel analysis. The two 1mL samples was lysed and compared by SDS PAGE. The gel showed a strong band at ~15 KDa from the sample after IPTG induction, indicating overexpression of Trx. The cell culture was centrifuged at 5000 xg for 10 min. The cell pellet was re-suspended in 8 mL of PBS (pH 7.4) and lysed through sonication on ice. The lysate was centrifuged at 5000 xg at 4 °C for 15min and the precipitate was discarded. Protein in the supernatant was purified according to the protocol from ThermoFisher (Qiagen Ni-NTA agarose). Purity of the protein was determined with SDS-PAGE and the mass was confirmed by LC/MS after dilution. The protein was stored at 4 °C in presence of 5 mM TCEP to keep it in its reduced form.

3.2.13.7 Oxidation and cleavage of Trx

1 mM GSSG was added to 1 mg/mL Trx and the solution was incubated for 2 hrs. To the oxidized mixture was added factor Xa according to the manual. The reaction was mostly complete after overnight incubation. The mass for cleaved protein [M] was detected after de-convolution as well as [M + GSH]. 10 mM TCEP was added to the mixture for 2 hrs and the reduction was let go for 3 hrs when complete reduction was observed. LC/MS results showed incomplete cleavage of the IEGR sequence. 0.5 mL protein mixture was suspended to 50 µL Ni-NTA agarose and incubated for 2 hrs on a rotator. The agarose beads were spun down and discarded. LC/MS of the supernatant showed a single mass corresponding to cleaved protein.

3.2.13.8 Expression of Cys-Trx

The primers are designed as the following sequences:

Primer 1:CGAAGGTCGTTGTATGGTGAAACAGATCGAG

Primer 2:ATATGGCCGCTGCTGTGA

The plasmid was obtained by mini-prep and single point mutation was conducted according to the manual. The new plasmid was transformed into the provided chemically-competent *E. coli* cells. Cells were grown on an agar plate (made with 0.1 mg/mL ampicillin) overnight and three colonies were chosen for miniprep and sequencing. The plasmid with the correct sequence was saved and used for transformation to *E. coli* (BL21 DE3) cells by electroporation. The transformed cells were plated on agar plate and let grown overnight. One colony was chosen to grow overnight and then sub-cultured in 500 mL LB media. 1 mL media at OD~0.6 was saved and the cells were re-suspended in 15% glycerol as a stock. To the rest of the media was added IPTG to induce protein expression. The desired protein was purified as described previously. The desired mass and a mass that was 178 greater than expected were both detected on LC/MS, consistent with previous observations.

3.2.13.9 Cleavage of Cys-AzoR

200 μg protein and 4 μL TEV^{plus} were mixed in 100 μL PBS buffer containing 0.5 mM EDTA and 1 mM DTT in a 3000 MWCO dialysis bag. The dialysis bag was stirred in 300 mL of the same buffer excluding the protein and TEV protease. The cleavage was completed within 2 hrs. The resulting protein went through solvent exchange into PBS pH 6.0 for labelling reactions.

3.2.14 References

(1) Bandyopadhyay, A.; Cambray, S.; Gao, J. Fast and Selective Labeling of N-Terminal Cysteines at Neutral PH via Thiazolidino Boronate Formation. *Chem. Sci.* **2016**, *7* (7), 4589–4593. https://doi.org/10.1039/C6SC00172F.

(2) Bandyopadhyay, A.; Cambray, S.; Gao, J. Fast and Selective Labeling of N-Terminal Cysteines at Neutral PH: Via Thiazolidino Boronate Formation. *Chem. Sci.* **2016**, *7* (7), 4589–4593. https://doi.org/10.1039/c6sc00172f.

(3) Li, K.; Kelly, M. A.; Gao, J. Biocompatible Conjugation of Tris Base to 2-Acetyl and 2-Formyl Phenylboronic Acid. *Org. Biomol. Chem.* **2019**, *17* (24), 5908–5912. https://doi.org/10.1039/c9ob00726a.

(4) Liu, H.; Li, X. Serine/Threonine Ligation: Origin, Mechanistic Aspects, and Applications. *Acc. Chem. Res.* **2018**, *51* (7), 1643–1655. https://doi.org/10.1021/acs.accounts.8b00151.

(5) Tam, J. P.; Miao, Z. Stereospecific Pseudoproline Ligation of N-Terminal Serine, Threonine, or Cysteine-Containing Unprotected Peptides. *J. Am. Chem. Soc.* **1999**, *121* (39), 9013–9022. https://doi.org/10.1021/ja991153t.

(6) Li, X.; Lam, H. Y.; Zhang, Y.; Chan, C. K. Salicylaldehyde Ester-Induced Chemoselective Peptide Ligations: Enabling Generation of Natural Peptidic Linkages at the Serine/Threonine Sites. *Org. Lett.* **2010**, *12* (8), 1724–1727. https://doi.org/10.1021/ol1003109.

(7) Walder, J. A.; Johnson, R. S.; Klotz, I. M. Neighboring-Group Participation of Aldehydes and Ketones in Ester Hydrolysis. Mechanism of Hydrolysis of O-Acetylsalicylaldehyde. *J. Am. Chem. Soc.* **1978**, *100* (16), 5156–5159.

https://doi.org/10.1021/ja00484a041.

(8) Wang, W.; Gao, J. N, S-Double Labeling of N-Terminal Cysteines via an Alternative Conjugation Pathway with 2-Cyanobenzothiazole. *J. Org. Chem.* **2020**. https://doi.org/10.1021/acs.joc.9b02959.

(9) Ramil, C. P.; An, P.; Yu, Z.; Lin, Q. Sequence-Specific 2-Cyanobenzothiazole Ligation. *J. Am. Chem. Soc.* **2016**, *138* (17), 5499–5502. https://doi.org/10.1021/jacs.6b00982.

(10) Barglow, K. T.; Knutson, C. G.; Wishnok, J. S.; Tannenbaum, S. R.; Marletta, M. A. Site-Specific and Redox-Controlled S-Nitrosation of Thioredoxin. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108* (35). https://doi.org/10.1073/pnas.1110736108.

(11) Hashemy, S. I.; Holmgren, A. Regulation of the Catalytic Activity and Structure of Human Thioredoxin 1 via Oxidation and S-Nitrosylation of Cysteine Residues. *J. Biol. Chem.* **2008**, *283* (32), 21890–21898. https://doi.org/10.1074/jbc.M801047200.

(12) You, B. R.; Shin, H. R.; Park, W. H. PX-12 Inhibits the Growth of A549 Lung Cancer Cells via G2/M Phase Arrest and ROS-Dependent Apoptosis. *Int. J. Oncol.* **2014**, *44* (1), 301–308. https://doi.org/10.3892/ijo.2013.2152.

3.3 Imidazolidino borornate formation enables peptide cyclization and cysteine detection

3.3.1 Introduction

As is discussed earlier, our lab and the Gois' lab have recently reported a fast, chemoselective reaction named thiazolidino boronate (TzB).[1,2] This reaction was used for site specific modification on a N-terminal cysteine residue in a peptide.

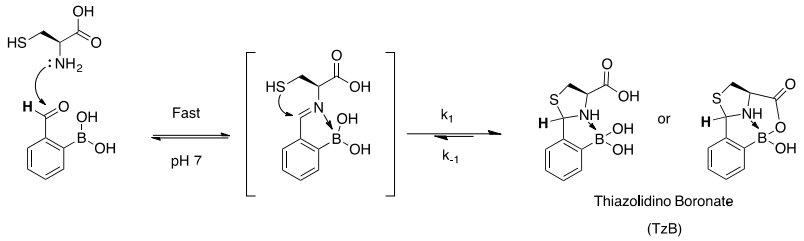


Figure 3.3.1. Mechanism of thiazolidino boronate formation

To have a deeper understanding of the TzB formation, and at the same time to extend the scope of this chemistry, we explored the reactivity of 2-FPBA with cysteine analogues such as serine, diaminopropionic acid (Dap) and tris-base.

3.3.2 The discovery of imidazolidino boronate [3]

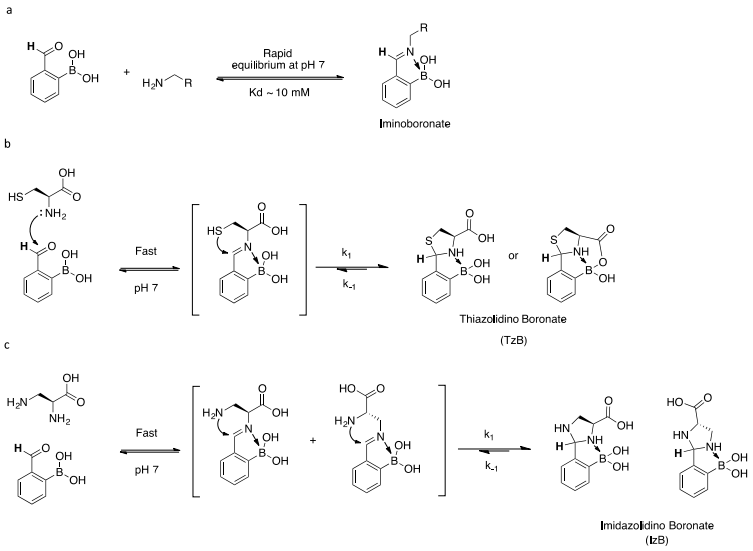


Figure 3.3.2. Mechanism of a) iminoboronate; b) thiazolidinoboronate (TzB); c) imidazolidino boronate formation (IzB)

Dap shares a lot of structure similarity to cysteine: the only difference is the replacement of the thiol group to a less nucleophilic amino group. We hypothesized that it might share a similar mechanism reacting with 2-FPBA.

To explore the interaction of 2-FPBA and Dap, we performed an UV-vis titration. Our lab has been applying UV-vis based experiments to determine the dissociation/ association kinetics as well as thermodynamics for iminoboronate and TzB. When 2-APBA or 2-FPBA forms iminoboronates with amines (including lysine and 2-methoxyethylamine), the maximum absorption wavelength shifts from 254 nm to 280 nm. The K_d can be generated by plotting the absorbance at 280 nm and fitting the data with hyperbola. In the case of cysteine and 2-FPBA, however, no increase of absorption was observed. Instead, the absorption of 2-FPBA at 254 nm decreased with the addition of cysteine due to the loss of conjugation. While studying the K_d between 2-FPBA and Dap, we observed a similar titration profile to cysteine (**Figure 3.3.4**), which agrees with our hypothesis.

We then carried out ^{1}H NMR experiments to explore the details of this reaction. When 2-FPBA and Dap were mixed at 1 mM at pH 7.4, we observed the appearance of peaks at 5.75 ppm and 5.85 ppm, which are signature peaks for the formation of imidazolidine. The masses of the conjugate as well as the starting material were both detected by LC/MS, indicating that some conjugate dissociated under acidic condition (**Figure 3.3.3**).

We have enough evidence to believe that this reaction follows a similar mechanism as the TzB chemistry. However, it is unknown which amine forms the iminoboroante with 2-FPBA in the intermediate step. This reaction is named as imidazolidino boronate (IzB, **Figure 3.3.2c**).

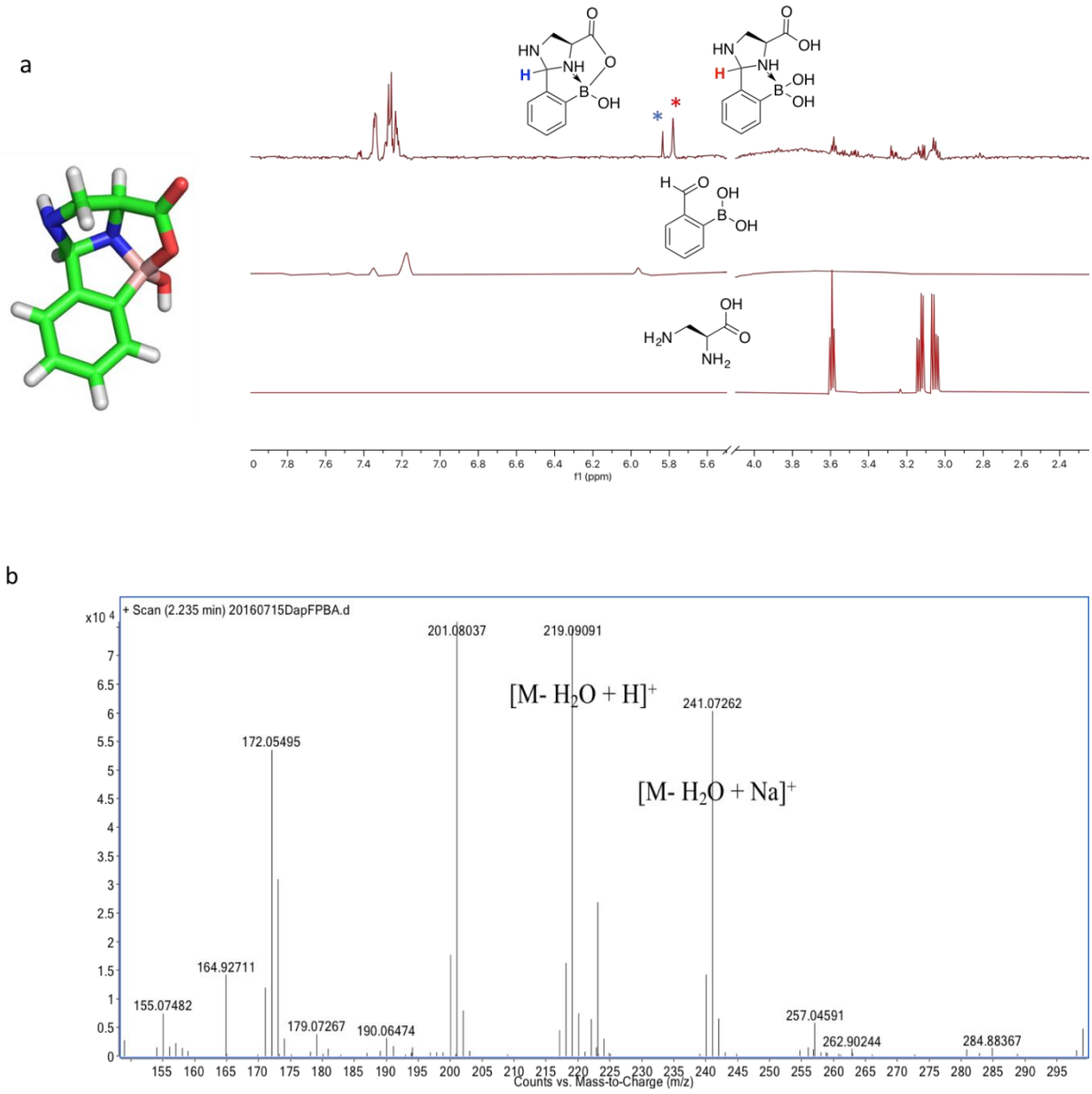


Figure 3.3.3. a) Crystal structure and NMR spectrum illustrating the formation of IzB; b) Mass spec data for IzB complex on LC/MS

3.3.3 Characterization of IzB

3.3.3.1 Thermodynamics of IzB chemistry

We performed an equilibrium titration experiment to determine the stability of the product and examine the reversibility of this chemistry. Detailed procedures are attached at the end of this section. Not surprisingly, IzB conjugation turned out be to reversible reaction and the K_d between 2-FPBA and Dap was determined to be 100

μM (**Figure 3.3.4**). This value is higher compared to that of TzB (~ 1 μM), as the nitrogen from the side chain of Dap is less nucleophilic than sulfur.

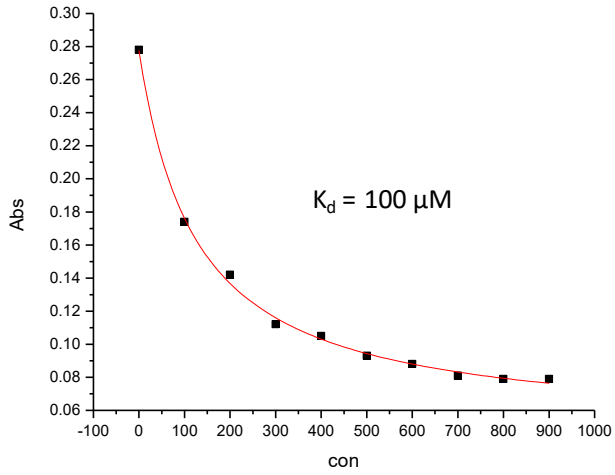


Figure 3.3.4. Titration curve of cysteine with 2-FPBA. Absorbance at 254 nm was plotted against the concentration of Dap. The data was fitted by hyperbola to give the K_d value

3.3.3.2 Kinetics of IzB chemistry

Both the forward and backward reaction followed a typical relaxation kinetics, which was studied by UV-vis based experiments. The relaxation time constant for dissociation and association were determined to be 6.7 ± 0.3 s and 6.2 ± 0.4 s, respectively. Combined with the calculated K_d value, the forward and backward reaction rate constants was calculated to be 820 M^{-1} s^{-1} and 0.082 s^{-1}, respectively. The association kinetics of this boronic acid accelerated conjugation is comparable to that of TzB formation while the dissociation rate is much faster.

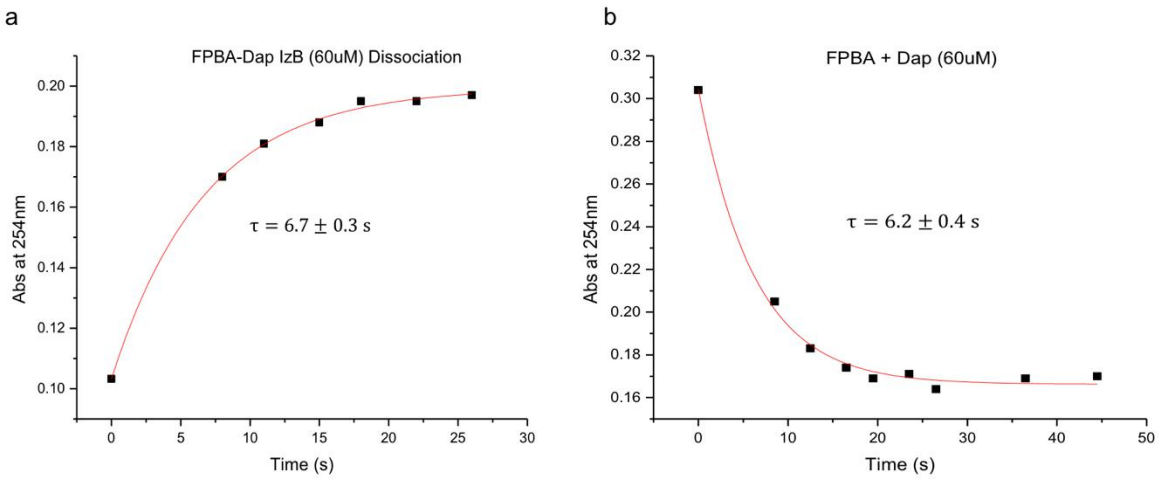


Figure 3.3.5. a) Dissociation and b) association kinetics of IzB. Both curves were fitted by relaxation kinetics

3.3.3.3 Chemoselectivity of IzB chemistry

To demonstrate the chemoselectivity of the IzB reaction, we carried out a small molecule competition experiment where 10 eq. of glucose, serine, lysine, glutathione or 1 eq. of free cysteine were added to IzB complex and the reactions were monitored by NMR. Not surprisingly, the IzB complex showed little interference with these small molecules other than cysteine. This is expected because the K_d of 2-FPBA conjugating with Dap (100 μM) is much lower than a typical iminoboronate (~ 10 mM). It also implies that the K_d between cysteine and 2-FPBA is in the low micromolar range (~ 5 μM). The residual IzB complex after cysteine addition can be attributed to the tricyclic product, which is supposed to be more stable.

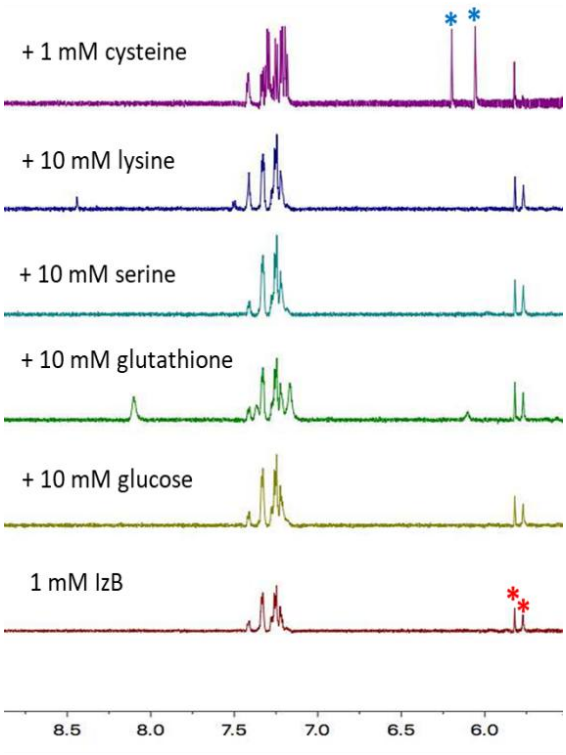


Figure 3.3.6 Stability of IzB in presence of bio-relevant molecules. The IzB complex was stable in presence of 10 eq. of lysine, serine, glutathione or glucose. Cysteine was able to decompose the IzB complex

3.3.4 IzB mediated peptide cyclization

3.3.4.1 Introduction

Cyclic peptides attracted attention from medical chemists because of their superior proteolytic stability compared to linear peptides. The reduced conformational entropy also favors target binding. As is discussed earlier, our lab has reported a novel method for peptide cyclization based on iminoboronate.[4,5]

We envisioned that IzB chemistry can also be utilized in peptide cyclization. When a N-terminal Dap residue and the 2-FPBA headgroup are both incorporated into a peptide, an intramolecular IzB formation can lead to spontaneous peptide cyclization. To test our hypothesis, we synthesized a peptide with the sequence

DapAAAAG(Dap-FPBA) and named it as KL21 as a proof-of-concept example. Synthesis of the peptide is discussed at the end of this section.

3.3.4.2 pH dependence of IzB mediated peptide cyclization

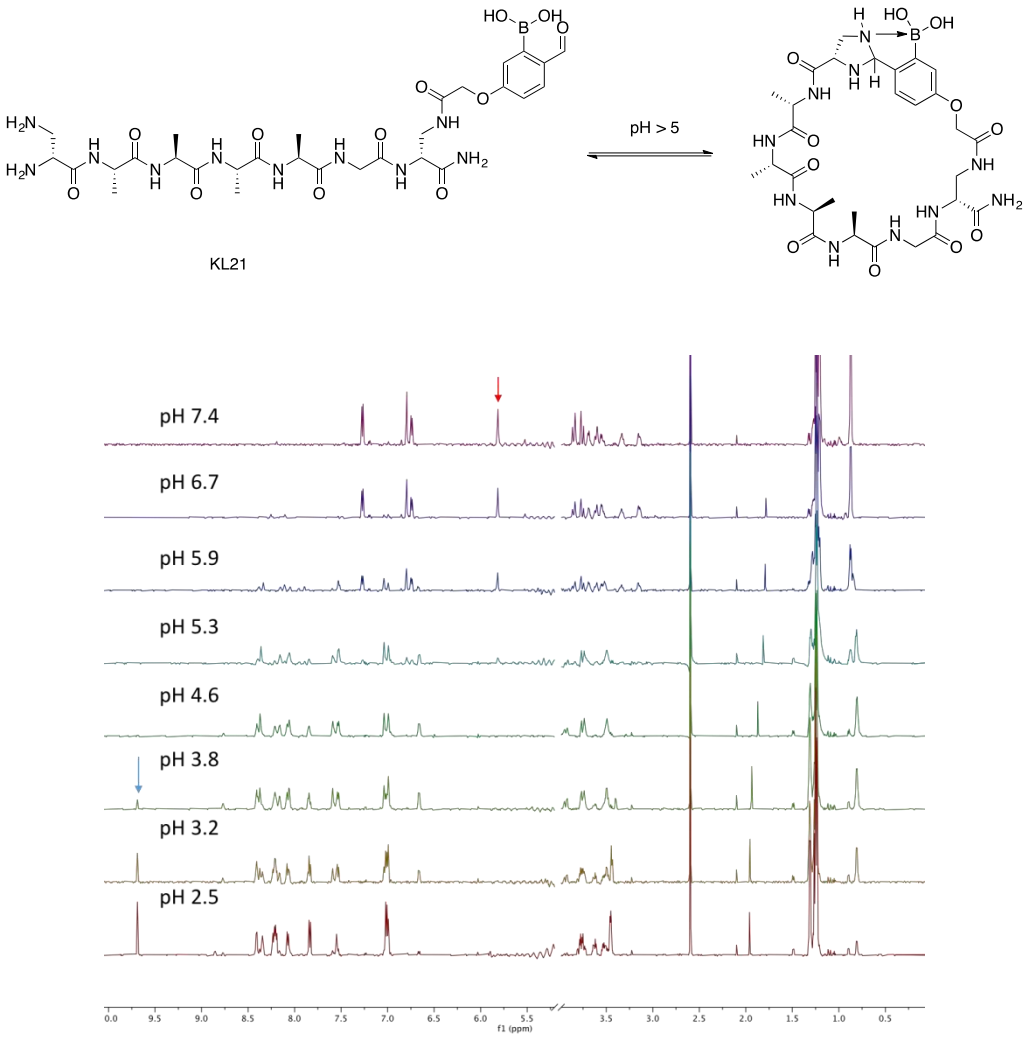


Figure 3.3.7. KL21 cyclizes in a pH dependent manner. Red arrow indicates IzB formation. KL21 was dissolved in phosphate buffer with 20% D_2O in PBS to make a final concentration of 500 μM. The pH was tuned from 2.5 to 7.4 incrementally using either 0.5N HCl or 0.5N NaOH solutions. Cyclization was monitored by the disappearance of the 2-FPBA aldehyde peak at 9.8 ppm and the appearance of the IzB benzylic hydrogen peak around 5.8 ppm.

As is shown in **Figure 3.3.7**, KL21 exists in a linear form under strongly acidic condition. As the pH increases (~ pH 6), the peptide gradually cyclization until

completion, illustrated by the disappearance of the aldehyde peak at 9.8 ppm. In conclusion, we have demonstrated here that the model peptide KL21 can go through spontaneous and complete cyclization in a pH dependent manner, without the requirement for any other reagent. It is worth to note that we had also attempted to apply TzB chemistry in peptide cyclization. However, the peptide is difficult to be synthesized and purified. The initial results suggest that a peptide that contains 2-FPBA headgroup and N-terminal cysteine can possibly aggregate because of intermolecular TzB formation instead of forming a homogeneous cyclized peptide (data not shown).

3.3.4.3 Chemoselectivity of IzB mediated peptide cyclization

We were curious whether the peptide cyclized by the IzB chemistry could go through any ring-opening reaction with the addition of bio-relevant molecules. Not surprisingly, KL21 stays in its cyclic form even with the addition of 10 eq. of serine, glucose, glutathione or lysine, showcasing a good stability.

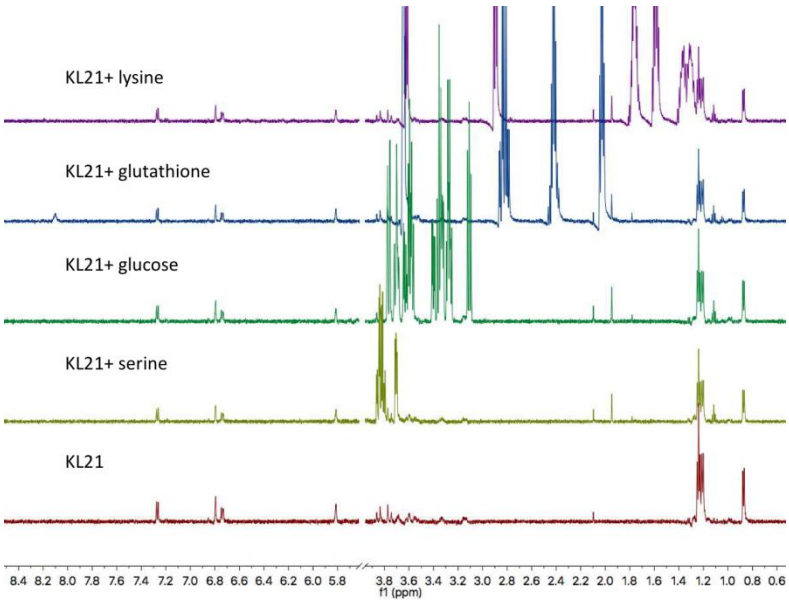


Figure 3.3.8. Cyclic KL21 showed impressive stability towards bio-relevant molecules. KL21 peptide was dissolved in 900 μL phosphate buffer (pH 7.4, 20% D_2O, final KL21 concentration: 200 μM). The solution was mixed with 100 μL stock solutions of lysine (20 mM), serine (20 mM), glutathione (20 mM) and glucose (20 mM) in phosphate buffer. The pH of the mixtures was tuned to 7.4 using 0.5N HCl or 0.5N NaOH. The samples were incubated for 30 minutes before the NMR spectrums were taken.

3.3.5 IzB enables cysteine detection

3.3.5.1 KL21 and KL22 respond to free cysteine in solution

IzB complex was proven to be stable in neutral aqueous condition and it would not easily dissociate even with the addition of competing molecules except free cysteine. Inspired by this observation, we were curious if cyclic KL21 could be linearized by free cysteine. A titration experiment was carried out by adding an increasing amount of cysteine solution to the cyclic peptide. Not surprisingly, KL21 was indeed linearized by the addition of cysteine, as is illustrated by the appearance of TzB signature peaks at 6.0 and 6.2 ppm and the decrease of the IzB peak at 5.8 ppm (**Figure 3.3.9a**). Conversion to the TzB product was quantitatively analyzed by plotting the percentage of integrated peak area of the signature peaks and the EC_{50} value was determined to be 2.0 mM. The same titration was carried out in presence of 10% blood serum (fetal bovine serum, FBS). No significant change in EC_{50} value was observed. This result indicates that the cyclic peptide is quite inert in complex media, further showcasing its chemoselectivity.

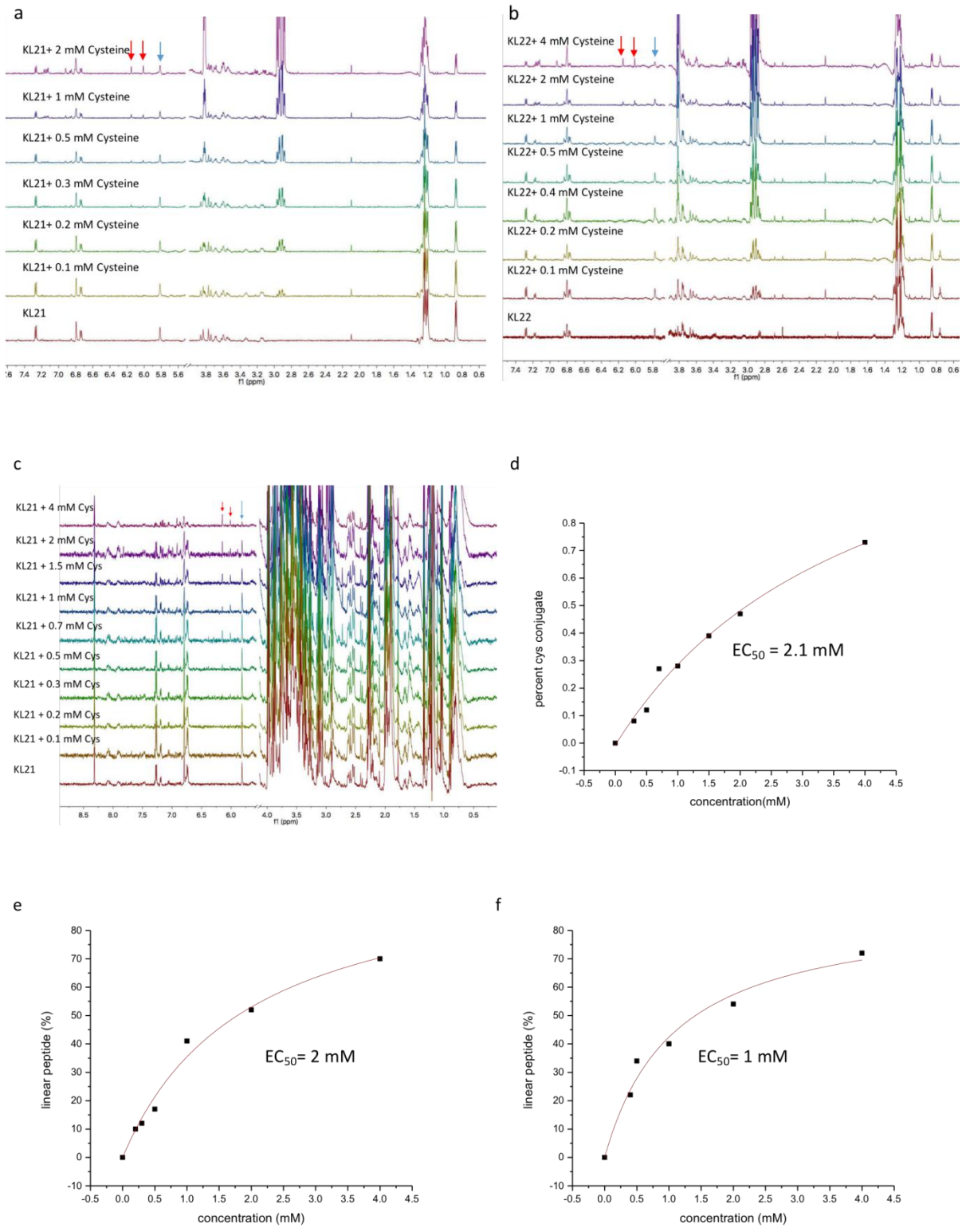


Figure 3.3.9. a), d) Cysteine is titrated into KL21 at PBS, pH 7.4; b), f) Cysteine is titrated into KL22 in PBS pH, 7.4; c), e) Cysteine is titrated into KL21 in presence of 10% FBS. The percentage was calculated by integration of the signature peaks for IzB (5.8 ppm) and TzB (6.0 and 6.2 ppm)

We synthesized a longer peptide KL22 with the sequence DapAAGAAAG(Dap-FPBA) to lower the EC_{50} value to the range of cellular free cysteine concentration (30-200 μM)[6]. This peptide indeed showed increased sensitivity to cysteine (EC_{50} =1 mM). However, it is still higher than desired. We are hesitant to synthesize a longer peptide due to the decreased cyclization efficiency. As a matter of fact, we did not observe peptide cyclization on a maganin derivative, as is shown in the figure below.

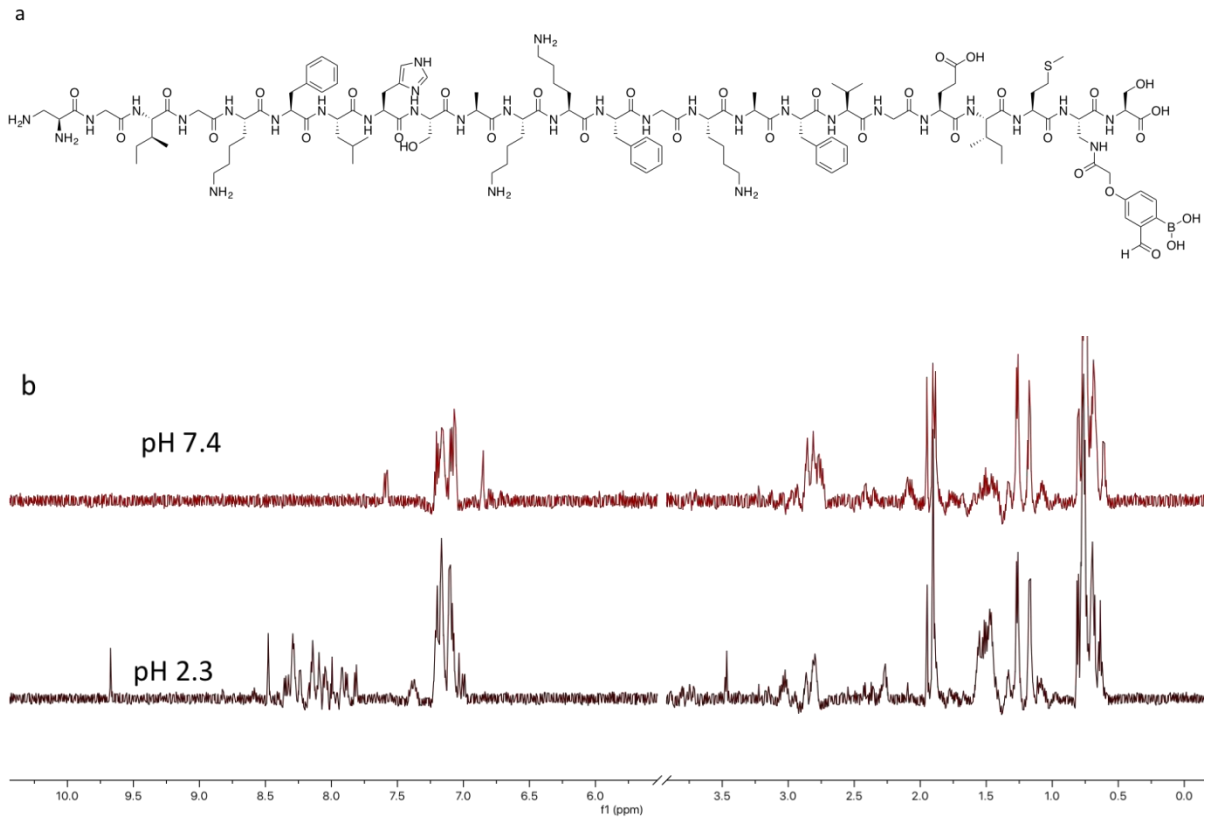


Figure 3.3.10. a) Structure of maganin derivative; b) No cyclization (signature peak at ~ 6.0 ppm) was observed from maganin derivative at pH 7.4

3.3.5.2 Re-cyclization of KL21

Since both TzB and IzB chemistries are reversible, we envisioned that the removal of free cysteine from solution will lead to the re-cyclization of KL21. A

solution of KL21 (0.2 mM) and cysteine (2 mM) was incubated at room temperature. After 60 hrs, we observed almost complete oxidation of free cysteine and the emergence of the IzB signature peak, providing further support for reversible cyclization by IzB.

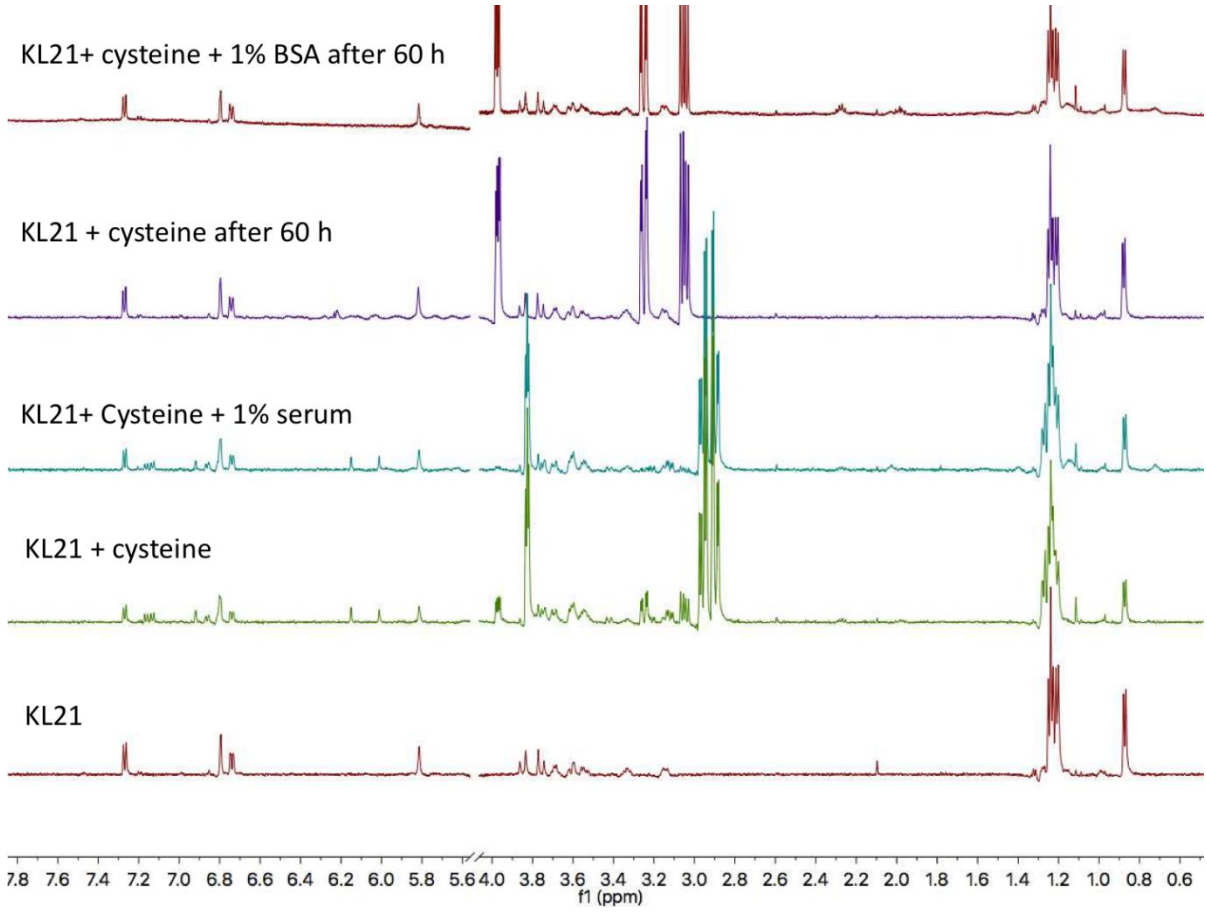


Figure 3.3.11. Oxidation of cysteine leads to re-cyclization of KL42 after 60 hrs

3.3.6 Developing fluorogenic probes for cysteine detection

3.3.6.1 Detect cysteine concentration using cyclic peptides

There are many reports about cysteine detection in solution or in cells with fluorogenic probes. These aldehyde-based probes are generally sluggish and often require acidic conditions.[6–9] To develop a simplified probe for cysteine detection, we designed a peptide named KL23 that contains both a fluorophore and a quencher.

KL23 is not fluorescent in its cyclic conformation. Once the peptide is linearized by cysteine, the fluorophore will be separated from the quencher and emits fluorescence. The quencher, namely Dabcyl, was synthesized into an unnatural amino acid and incorporated into the peptide by SPPS. Fluorescein maleimide was installed through a reserved cysteine near the C-terminus.

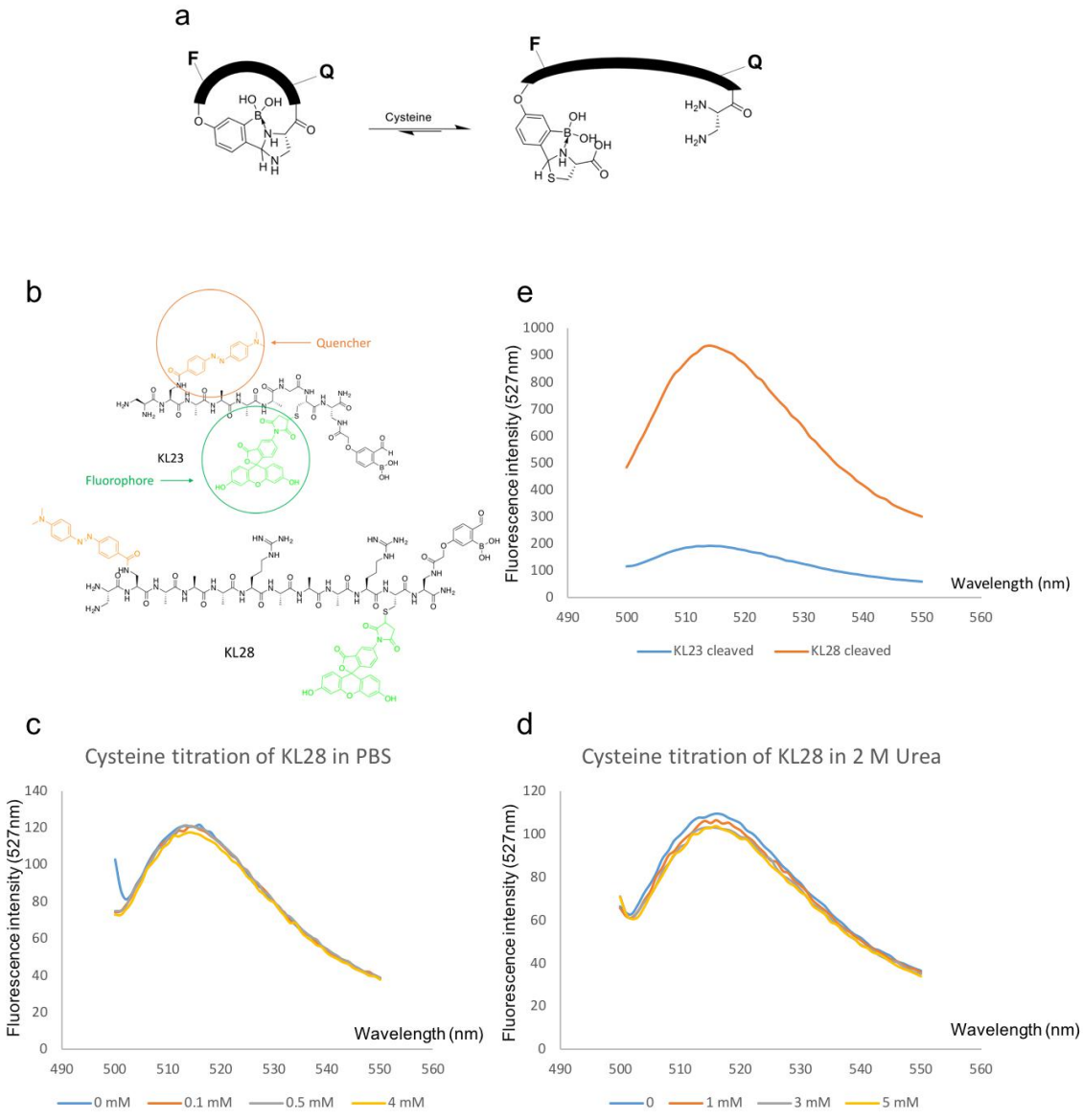


Figure 3.3.12. a) Illustration of peptide design; b) Structure of KL23 and KL28; c) Cysteine titration experiment of KL28; d) Cysteine titration experiment of KL28 in 2M urea; e) Fluorescent profile of

trypsin digested KL23 and KL28. Neither KL23 or KL28 showed fluorescent responses to the addition of free cysteine. Trpsin cleavage induced fluorescence increase for KL28 but not KL23

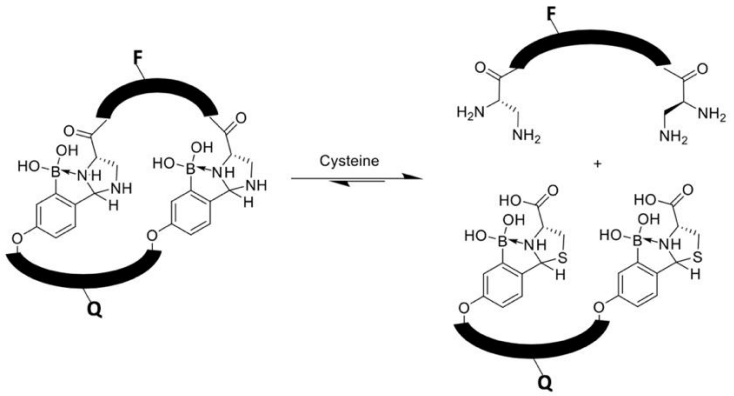
Compared to free fluorescein (FAM) molecule alone, our peptide showed significantly less fluorescence, indicating the quenching effect of Dabcyl. Unfortunately, this peptide fails to give any response to cysteine, presumably due to insufficient fluorophore-quencher separation (data not shown). Based on this failed trail, we hypothesized that KL23 did not provide enough distance to separate the fluorophore and quencher. We then designed a longer peptide that contains 14 residues (KL28). Unfortunately, this peptide again failed to generate any fluorescence with the addition of cysteine (**Figure 3.3.12c**). It is possible that, albeit linearized, KL28 still exists in a curly form so the fluorophore and quencher are still spatially close. Concentrated urea (2 M) was added into the buffer to eliminate possible folding of the peptide. However, the titration profile showed no difference to previous results (**Figure 3.3.12d**).

As a positive control, we digested KL28 with trypsin to completely separate the fluorophore and quencher. KL23 was used as a negative control because of its lack of arginine. As expected, there was an increase in fluorescence from cleaved KL28 compared to KL23 (**Figure 3.3.12e**). We also learned from this experiment that intermolecular Dabcyl was not able to induce fluorescence quenching at such a low concentration. In fact, no significant fluorescent decrease was observed when Dabcyl (up to 5 μM) was titrated into fluorescein (1 μM). This result confirms that Dabcyl indeed quenches the fluorescence of FAM in the same peptide, even in presence of cysteine. We proposed two possible reasons for our failure with KL23 and KL28. The first one is that the peptides we designed were simply not long enough. This hypothesis is difficult to be validated due to decreased cyclization efficiency for a

longer peptide. Another reason could be that the KL23 and KL28 in their linear form were still somewhat coiled due to the hydrophobic interaction between Dabcyl and FAM.

3.3.6.2 Detect cysteine concentration using a peptide pair

We conclude from previous trials that cysteine detection assay based on fluorescence increase cannot be easily realized through the change of intramolecular interactions. We then changed our strategy to bimolecular association. To this end, we designed a peptide pair named KL24 and KL25. KL24 displays two copies of Dap and a fluorescein while KL25 bares two copies of 2-FPBA and one copy of Dabcyl. It is noteworthy that an original attempt to label KL25 precursor that has two copies of 2-FPBA and a free Dap residue with Dabcyl-Osu was not successful. Dabcyl-Osu and KL25 precursor showed no reaction after 6 h due to intramolecular iminoboronate formation between 2-FPBA and Dap. Detailed synthetic routes for both peptides are shown at the end of this chapter.



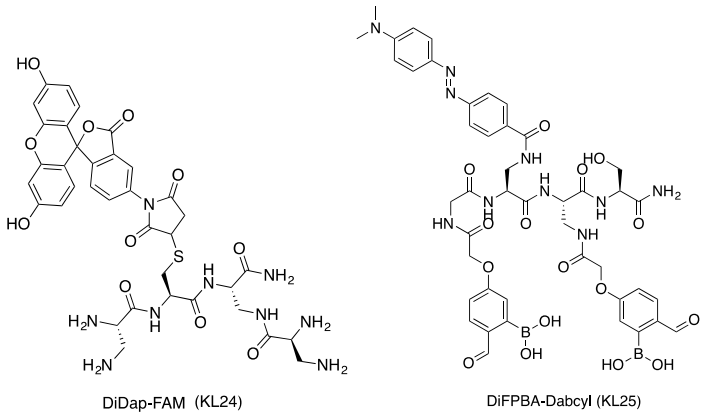


Figure 3.3.13. Design of fluorogenic peptide pair for cysteine detection

The binding affinity of this peptide pair was determined with a fluorescent assay. Titrating KL25 into a solution of KL24 causes a decrease in fluorescence at 517 nm. This experiment was conducted in 96-well plate instead of cuvette to avoid quenching of the fluorophore by repeated excitation. Fluorescent intensity was plotted against the concentration of KL25 and fit by hyperbola equation (**Figure 3.3.14a**). The K_d turned out to be surprisingly low (~0.3 μM), which is almost three orders of magnitudes lower than that of 2-FPBA and Dap monomer (~0.1 mM). This dramatic increase in affinity can be attributed to multivalent effect as well as hydrophobic effect between fluorescein and Dabcyl.

To examine the fluorescent response of this peptide pair towards cysteine, 1 μM of KL25 was mixed with 0.1 μM of KL24 to quench most of the fluorescence before titration. Fluorescent readings were taken while increasing amounts of cysteine from stock solution was gradually titrated into the mixture. To our delight, this peptide pair showed significant responses to cysteine (**Figure 3.3.14b**). Fitting the titration curve yields an EC_{50} value of 0.2 mM, which falls into the physiological concentration range. Moreover, this fluorogenic response is highly specific to cysteine

and it is inert to other bio-relevant molecules including glucose, lysine, glutathione and serine (**Figure 3.2.14b**).

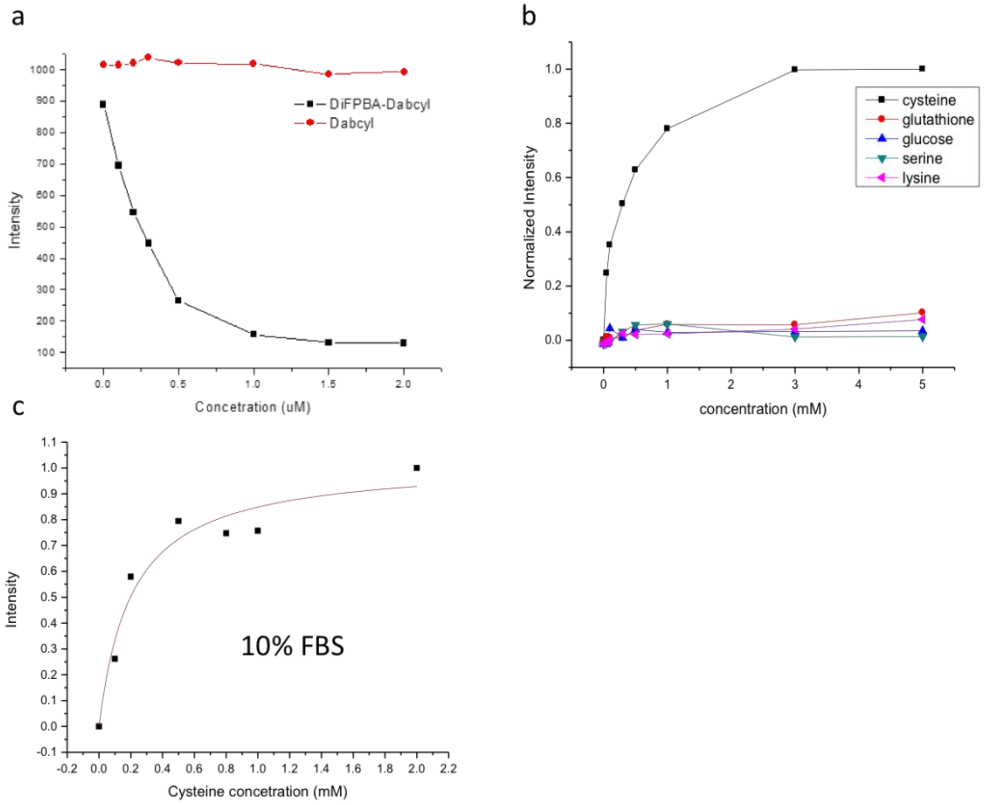


Figure 3.3.14. a) Titration of KL25 into KL24; b) Titration curves of cysteine, glutathione, glucose, serine and lysine into the peptide pair; c) Titration curve of cysteine into the peptide pair in 10% FBS. Peptide pair only showed fluorescent responses to the addition of cysteine

We further explored this reaction in a more complex system (10% FBS). The result was affected to some extent, but the trend agrees with our experiment in PBS. Importantly, the EC_{50} also lies at ~0.2 mM in 10% FBS. The change in fluorescence is possibly caused by lysine residues from the proteins in FBS.

3.3.7 Conclusions

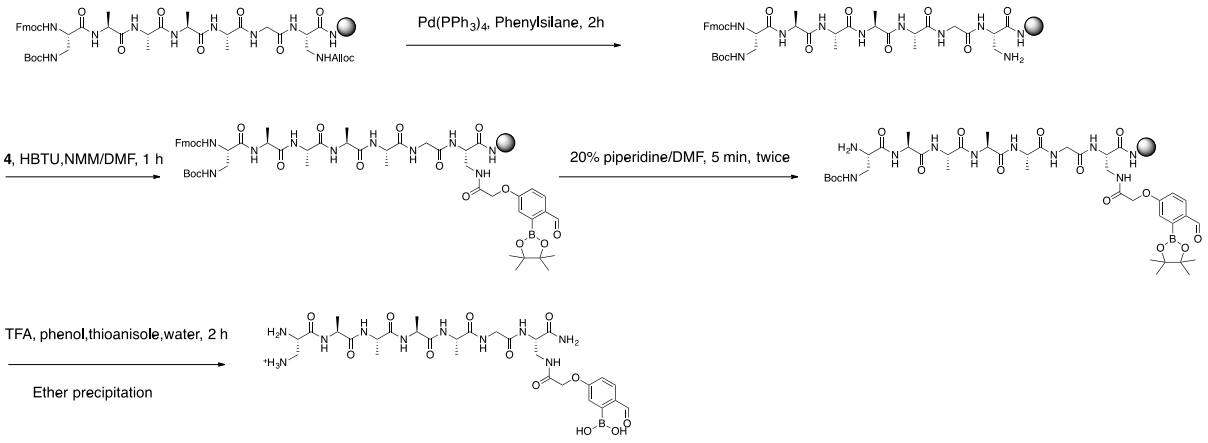
In this chapter, we described a conjugation reaction between 2-FPBA and Dap to form an IzB complex. This reaction shares a similar mechanism to TzB formation

and is rapidly reversible in neutral aqueous solutions with a K_d of ~ 100 μM. Importantly, IzB formation is highly chemo selective, showing little interference with bio-relevant molecules except cysteine. Peptides that includes both Dap and 2-FPBA headgroups can be cyclized spontaneously in neutral aqueous conditions. These cyclic peptides are resistant to small molecules such as lysine and glucose, but they are sensitive to pH and cysteine. We also developed a peptide pair as a fluorogenic reporter for free cysteine detection in solution through the exchange reaction of IzB and TzB. We believe this strategy can be applied not only to free cysteine or proteins with N-terminal cysteine but also molecules that display 1,2-aminothiol structures such as bacillithiol produced in bacteria.[10]

3.3.8 General information

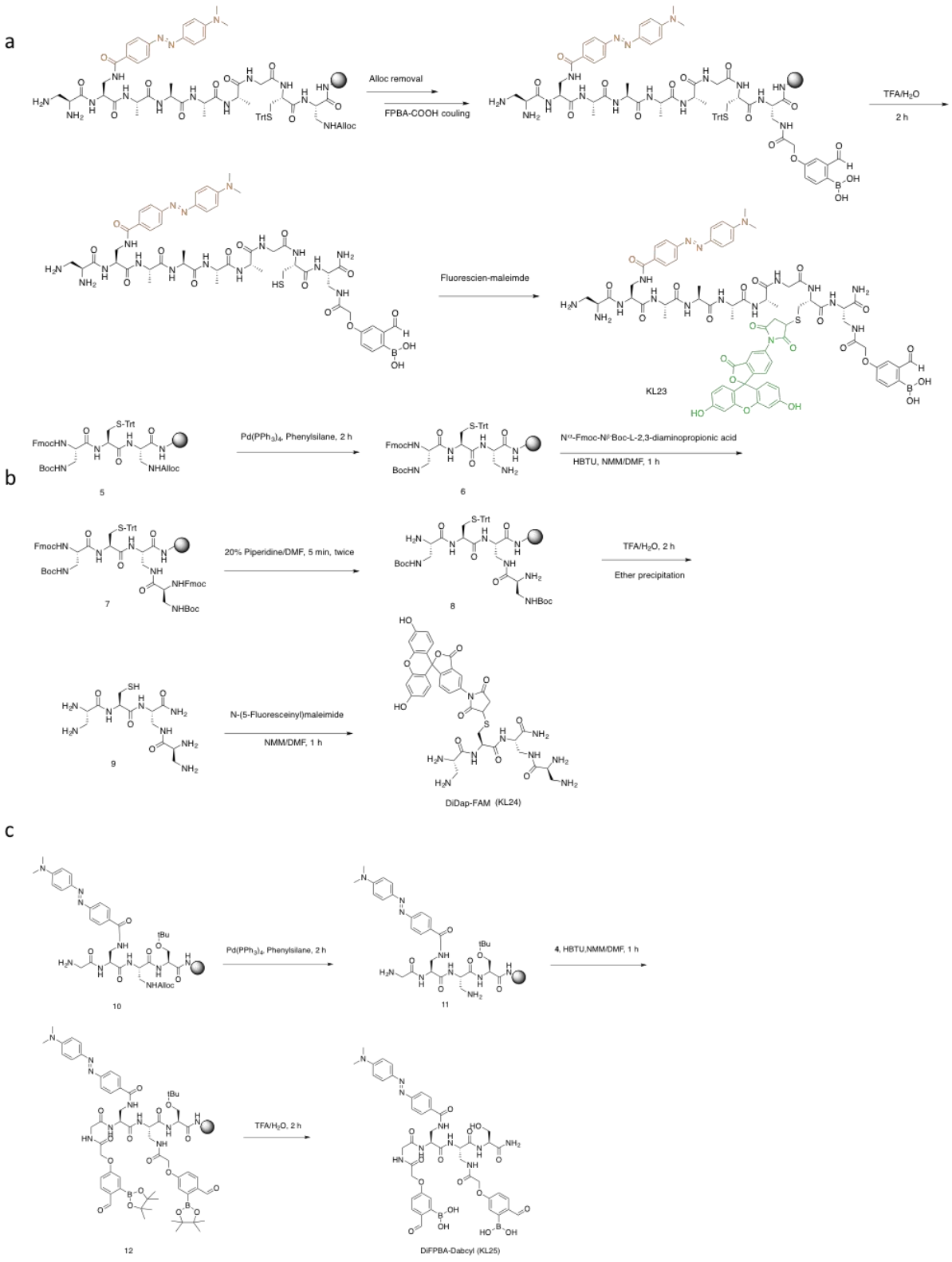
3.3.8.1 Peptide synthesis

For peptide synthesis including KL21 and KL22 and the maganin derivative, we used two kinds of Fmoc protected Dap amino acids that have orthogonal protecting groups on the side chain - Alloc or Boc. Alloc was removed on resin by $P_d(PPh_3)_4$ and phenylsilane in DCM for two hours. The resin was washed by DMF and molecule **4** (whose synthesis will be discussed later) was activated by HBTU and coupled to Dap residue. The Fmoc group on the N-terminal Dap was removed by piperidine and the remaining protecting groups were globally removed by TFA/H_2O (9:1).



Scheme 3.3.1. Synthesis of KL21. KL22 and the maganin derivative were synthesized in a similar manner

For peptides which contain fluorophore and quencher molecules, including KL23 and KL28, an extra cysteine residue was included as a handle for fluorophore labelling.



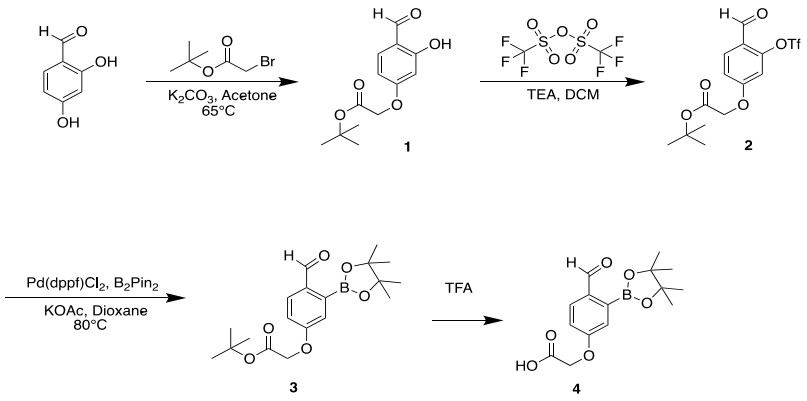
Scheme 3.3.2. Synthetic schemes for a) KL23; b) KL24 and c) KL25

3.3.8.2 Small molecule synthesis

a) Synthesis of Fmoc-Dap-Dabcyl

Dabcyl-Osu (50mg, 0.136 mmol) and Fmoc-Dap-OH (44.5mg, 0.136 mmol) was suspended in 5 mL of Dichloromethane. Triethylamine (70 mg, 0.68 mmol) was added to the mixture and the reaction was stirred overnight. 100 mL HCl (0.02 M) was then added to quench the reaction. The product was extracted by EtOAc (80mL × 3). The organic layers were combined and washed with brine (100 mL) and dried over sodium sulfate. The solvent was evaporated and yielded a dark red solid (70 mg, 88% yield). The crude product was directly used for peptide synthesis without further purification.

b) Synthesis of FPBA-COOH



Scheme 3.3.3. Synthetic scheme for FPBA-COOH

Synthesis of **1**

2,5-dihydroxylbenzaldehyde (3.00 g, 21.74 mmol) was dissolved in 10 mL acetone, to which tert-butyl 2-bromoacetate (4.24 g, 21.43 mmol) and potassium carbonate (5 g, 36.23 mmol) was added. The reaction was stirred at 65 ℃ for 16 h then cooled to room temperature. 150 mL of water was added to the reaction mixture and extracted with ethyl acetate (3 × 150 mL). The organic layers were combined, washed with brine (150 mL) and dried over sodium sulfate. Ethyl acetate was

evaporated and the crude product was purified by silica column using ethyl acetate: hexane (1:10) to yield a white solid (2.90 g, 53%).

^{1}H-NMR (500 MHz, Chloroform-*d*) δ 11.42 (s, 1H), 9.72 (s, 1H), 7.44 (d, *J* = 8.7 Hz, 1H), 6.56 (d, *J* = 11.1 Hz, 1H), 6.36 (d, *J* = 2.4 Hz, 1H), 4.55 (s, 2H), 1.48 (s, 9H).

^{13}C-NMR (126 MHz, Chloroform-*d*) δ 194.47, 166.81, 164.88, 164.25, 135.37, 115.70, 108.45, 101.45, 82.97, 65.48, 28.00.

MS-ESI^{+}: *m/z* calculated for $C_{13}H_{17}O_5$ $[M+H]^+$ 253.1076, observed 253.1055.

Synthesis of **2**

1 (1.51 g, 6.00 mmol) and triethylamine (3.03 g, 30.00 mmol) were dissolved in 15 mL of dichloromethane and stirred at -78 ℃ for 5 min. Trifluoromethanesulfonic anhydride (3.39 g, 12.00 mmol) was added slowly into the mixture over 1 min. The reaction was then allowed to stir at room temperature for 1 h under argon and quenched with 50 mL of saturated sodium bicarbonate. The mixture was stirred for 5 min and then extracted with dichloromethane (3 × 100 mL). The combined organic layer was washed with brine (100 mL) and dried over sodium sulfate. After solvent removal, the crude product was purified by silica column using ethyl acetate: hexane (1:9) to yield a light yellow solid (1.90 g, 83%).

^{1}H-NMR (600 MHz, Chloroform-*d*) δ 10.13 (s, 1H), 7.94 (d, *J* = 8.7 Hz, 1H), 7.00 (dd, *J* = 8.7, 2.3 Hz, 1H), 6.88 (d, *J* = 2.3 Hz, 1H), 4.61 (s, 2H), 1.49 (s, 9H).

^{13}C-NMR (151 MHz, Chloroform-*d*) δ 187.90, 168.90, 166.14, 153.79, 134.82, 125.08, 121.24, 117.31, 111.52, 86.14, 68.56, 30.66.

MS-ESI^{+}: *m/z* calculated for $C_{14}H_{16}F_3O_7S$ $[M+H]^+$ 385.0569, observed 385.0515.

Synthesis of **3**

2 (1.50g, 3.90 mmol), B_2Pin_2 (2.57g, 10.12 mmol), $Pd(dppf)Cl_2$ (300 mg, 0.41 mmol) and potassium acetate (2.00 g, 21.01 mmol) were dissolved in 10 mL of anhydrous dioxane, to which ~100 mg of 3 Å molecular sieves were added. The reaction was flushed with argon for 15 min and allowed to stir for 1 h at 80 °C. The reaction was cooled down to room temperature and water (50 mL) was added to the reaction. The product was extracted with ethyl acetate (3×100 mL). The combined organic layer was washed with brine (100 mL) and dried over sodium sulfate. Ethyl acetate was removed and the product was purified on silica gel column using ethyl acetate: hexane (3:17) to give the desired product a white solid (0.54 g, 38% yield).

^{1}H-NMR (600 MHz, Chloroform-*d*) δ 10.40 (s, 1H), 7.93 (d, *J* = 8.6 Hz, 1H), 7.28 (d, *J* = 2.6 Hz, 1H), 7.03 (dd, *J* = 8.6, 2.7 Hz, 1H), 4.59 (s, 2H), 1.48 (s, 9H), 1.37 (s, 12H).

^{13}C-NMR (151 MHz, Chloroform-*d*) δ 195.67, 169.89, 164.09, 138.03, 132.95, 123.06, 119.73, 87.11, 85.40, 68.22, 30.65, 27.50.

MS-ESI$^+$: *m/z* calculated for $C_{13}H_{16}BO_5$ $[M\text{-}Pin\text{-}H_2O+H]^+$ 263.1091, observed 263.1221.

Synthesis of **4**

3 (200 mg, 0.55 mmol) was dissolved in 2 mL of dichloromethane and stirred at 0 °C, to which 3 mL of trifluoroacetic acid (TFA) was added. The mixture was stirred at room temperature for 1 h before the dichloromethane and TFA were evaporated. The residue was treated with 5 mL of TFA for another hour at room

temperature. Removing TFA completely via evaporation and washing with toluene and dichloromethane yielded an off-white powder as the desired product (150 mg, 89%).

^{1}H NMR (600 MHz, DMSO-d_6) δ 13.12 (s, 1H), 10.13 (s, 1H), 7.87 (d, J = 8.5 Hz, 1H), 7.16 – 7.10 (m, 2H), 4.82 (s, 2H), 1.32 (s, 12H).

^{13}C NMR (151 MHz, DMSO-d_6) δ 195.33, 172.74, 164.62, 137.18, 134.51, 123.14, 119.16, 87.21, 67.67, 27.77.

MS-ESI^{+}: *m/z* calculated for $C_{13}H_{16}BO_5$ $[M\text{-}Pin\text{-}H_2O+H]^{+}$ 207.0465, observed 207.0422.

3.3.8.3 K_d titration

To a 2-FPBA solution (50 μM) was added increasing amounts of L-Dap stock (100 mM) to give final L-Dap concentrations of 100 μM to 900 μM. The absorbance at 254 nm was obtained via NanoDrop reading using a 10 mm path length cuvette. The solution was mixed by pipetting for about 1 min before the measurement. As is mentioned, a decrease in absorbance at 254 nm was observed as 2-FPBA was converted to the IzB complex. Plotting the absorbance at 254 nm against the concentration of Dap gave the binding curve shown below. Fitting the data to a hyperbola equation (y = A + Bx/(C+x)) yielded the apparent K_d value (K_d = C).

3.3.8.4 Kinetic study

For the dissociation reaction, 1 mM 2-FPBA and 1 mM Dap were mixed in PBS buffer at pH 7.4. The mixture was allowed to incubate for 15 minutes at room temperature, then diluted 16.7x to a final concentration of 60 μM and quickly mixed via pipetting. The absorbance profile was recorded approximately every 3 seconds.

The initial absorbance value of the IzB complex was obtained from a separate equimolar mixture of 2-FPBA and Dap (600 μM) pre-incubated for 15 minutes and the absorbance at 254 nm was measured using 1 mm pat/h length, which equals 60 μM concentration of the complex measured by a 10 mm pat/h length cuvette. A rapid increase in absorbance at 254 nm was detected as free 2-FPBA was generated upon dissociation of the IzB complex. This process follows a typical relaxation kinetics. Plotting the absorbance at 254 nm against time followed by curve fitting gave the relaxation time constant of the backward reaction.

For the association reaction, 1 mL of 2-FPBA solution (60 μM) prepared from 10 mM stock solution was added to a cuvette. The absorbance was recorded as the initial time point. 1.2 μL of Dap stock (50 mM) was added to the cuvette to make a final concentration of 60 μM and the solution was quickly mixed via pipetting. The absorbance profile was recorded approximately every 3 seconds. A rapid decrease in absorbance at 254 nm was seen during a period of 1 min as the 2-FPBA was converted to the IzB complex. Plotting absorbance at 254 nm versus time followed by curve fitting gave the forward reaction time constant.

3.3.8.5 IzB complex crystallographic information

300 mM solutions of 2-FPBA (22.5mg in 500 μL) and L-Dap (21.1mg in 500 μL) in water were prepared and combined. 1 mL of MeOH was added to that mixture and the pH was tuned to 7.4 using 1N HCl or 1N NaOH. The solution was filtered through a 0.45 μm PTFE membrane filter (Phenomenex) and allowed to slowly evaporate from a loosely capped 5 mL glass vial at room temperature. After a few days, crystal aggregates were observed and redissolved in pure MeOH. After slow recrystallization at room temperature over a week, single crystals were observed.

Table 3.3.1. Crystal data and structure refinement for the 2-FPBA-L-Dap IzB complex. Structure seen as a tetramer around a central NaCl molecule.

Identification code	C40H44B4ClN8NaO12	
Empirical formula	C40 H44 B4 Cl N8 Na O12	
Formula weight	930.51	
Temperature	100(2) K	
Wavelength	1.54178 Å	
Crystal system	Tetragonal	
Space group	I4	
Unit cell dimensions	a = 17.4670(5) Å	a= 90°.
	b = 17.4670(5) Å	b= 90°.
	c = 6.7244(2) Å	g = 90°.
Volume	2051.59(13) $Å^3$	
Z	2	
Density (calculated)	1.506 Mg/m^3	
Absorption coefficient	1.580 mm^{-1}	
F(000)	968	
Crystal size	0.250 x 0.220 x 0.140 mm^3	
Theta range for data collection	5.064 to 66.529°.	
Index ranges	-20<=h<=11, -20<=k<=19, -7<=l<=7	
Reflections collected	5002	
Independent reflections	1718 [R(int) = 0.0245]	

Completeness to theta = 66.529°	99.3 %
Absorption correction	Semi-empirical from equivalents
Max. and min. transmission	0.7528 and 0.6716
Refinement method	Full-matrix least-squares on F^2
Data / restraints / parameters	1718 / 1 / 163
Goodness-of-fit on F^2	1.039
Final R indices [I>2sigma(I)]	R1 = 0.0227, wR2 = 0.0577
R indices (all data)	R1 = 0.0232, wR2 = 0.0580
Absolute structure parameter	0.032(8)
Extinction coefficient	n/a
Largest diff. peak and hole	0.305 and -0.156 e.$Å^{-3}$

Table 3.3.2. Atomic coordinates (x 10^4) and equivalent isotropic displacement parameters ($Å^2$x 10^3) for the 2-FPBA-L-Dap IzB complex. U(eq) is defined as one third of the trace of the orthogonalized U^{ij} tensor.

	x	y	z	U(eq)
Na(1)	5000	5000	2511(3)	15(1)
Cl(1)	5000	5000	6570(1)	14(1)
O(1)	6341(1)	5315(1)	2576(2)	13(1)
O(2)	7024(1)	5597(1)	5578(2)	13(1)
O(3)	6905(1)	5644(1)	8865(2)	16(1)

N(1)	6699(1)	4269(1)	5014(3)	12(1)
N(2)	7501(1)	3290(1)	6337(3)	17(1)
B(1)	6941(1)	5046(1)	3834(4)	12(1)
C(1)	6914(1)	5280(1)	7325(3)	12(1)
C(2)	6810(1)	4422(1)	7171(3)	14(1)
C(3)	7529(1)	3942(1)	7708(4)	17(1)
C(4)	7248(1)	3617(1)	4474(4)	15(1)
C(5)	7877(1)	4011(1)	3306(3)	13(1)
C(6)	8536(1)	3643(1)	2655(3)	16(1)
C(7)	9065(1)	4057(1)	1550(4)	20(1)
C(8)	8937(1)	4828(1)	1137(3)	18(1)
C(9)	8280(1)	5192(1)	1808(3)	14(1)
C(10)	7735(1)	4783(1)	2899(3)	12(1)

3.3.9 References

(1) Bandyopadhyay, A.; Cambray, S.; Gao, J. Fast and Selective Labeling of N-Terminal Cysteines at Neutral PH via Thiazolidino Boronate Formation. *Chem. Sci.* **2016**, *7* (7), 4589–4593. https://doi.org/10.1039/C6SC00172F.

(2) Faustino, H.; Silva, M. J. S. A.; Veiros, L. F.; Bernardes, G. J. L.; Gois, P. M. P. Iminoboronates Are Efficient Intermediates for Selective, Rapid and Reversible N-Terminal Cysteine Functionalisation. *Chem. Sci.* **2016**, *7* (8), 5052–5058. https://doi.org/10.1039/c6sc01520d.

(3) Li, K.; Weidman, C.; Gao, J. Dynamic Formation of Imidazolidino Boronate Enables Design of Cysteine-Responsive Peptides. *Org. Lett.* **2018**, *20* (1), 20–23. https://doi.org/10.1021/acs.orglett.7b03116.

(4) Bandyopadhyay, A.; Gao, J. Iminoboronate-Based Peptide Cyclization That Responds to PH, Oxidation, and Small Molecule Modulators. *J. Am. Chem. Soc.* **2016**, *138* (7), 2098–2101. https://doi.org/10.1021/jacs.5b12301.

(5) Li, K.; Gao, J. Iminoboronate-Mediated Peptide Cyclization with Lysine Homologues. *Synlett* **2017**, *28* (15), 1913–1916. https://doi.org/10.1055/s-0036-1590795.

(6) Tian, M.; Guo, F.; Sun, Y.; Zhang, W.; Miao, F.; Liu, Y.; Song, G.; Ho, C. L.; Yu, X.; Sun, J. Z.; et al. A Fluorescent Probe for Intracellular Cysteine Overcoming the Interference by Glutathione. *Org. Biomol. Chem.* **2014**, *12* (32), 6128–6133. https://doi.org/10.1039/c4ob00382a.

(7) Yang, Z.; Zhao, N.; Sun, Y.; Miao, F.; Liu, Y.; Liu, X.; Zhang, Y.; Ai, W.; Song, G.; Shen, X.; et al. Highly Selective Red- and Green-Emitting Two-

Photon Fluorescent Probes for Cysteine Detection and Their Bio-Imaging in Living Cells. *Chem. Commun.* **2012**, *48* (28), 3442–3444. https://doi.org/10.1039/c2cc00093h.

(8) Tong, H.; Zhao, J.; Li, X.; Zhang, Y.; Ma, S.; Lou, K.; Wang, W. Orchestration of Dual Cyclization Processes and Dual Quenching Mechanisms for Enhanced Selectivity and Drastic Fluorescence Turn-on Detection of Cysteine. *Chem. Commun.* **2017**, *53* (25), 3583–3586. https://doi.org/10.1039/c6cc09336a.

(9) Wang, W.; Rusin, O.; Xu, X.; Kim, K. K.; Escobedo, J. O.; Fakayode, S. O.; Fletcher, K. A.; Lowry, M.; Schowalter, C. M.; Lawrence, C. M.; et al. Detection of Homocysteine and Cysteine. *J. Am. Chem. Soc.* **2005**, *127* (45), 15949–15958. https://doi.org/10.1021/ja054962n.

(10) Newton, G. L.; Rawat, M.; La Clair, J. J.; Jothivasan, V. K.; Budiarto, T.; Hamilton, C. J.; Claiborne, A.; Helmann, J. D.; Fahey, R. C. Bacillithiol Is an Antioxidant Thiol Produced in Bacilli. *Nat. Chem. Biol.* **2009**, *5* (9), 625–627. https://doi.org/10.1038/nchembio.189.

3.4 Tris-based conjugation with 2-APBA/2-FPBA

3.4.1 Introduction

So far, we have demonstrated that 2-FPBA conjugates rapidly with cysteine and Dap in aqueous solutions to form TzB[1] and IzB[2], respectively. These reactions start with the imine formation between 2-FPBA and α-amine from cysteine or Dap. The nucleophilic group from the side chain of the amino acid, either amine or thiol, further attacks the imine structure and yield a stabilized five-member ring. Serine, as an abundant proteinogenic amino acid, displays a structure that is highly analogous to both cysteine and Dap. To expand the scope of our chemistry and develop more tools for site specific modification on proteins, we are interested to find out if serine would react with 2-FPBA to form oxazolidino boronate (OzB) through a similar mechanism as TzB and IzB formation.

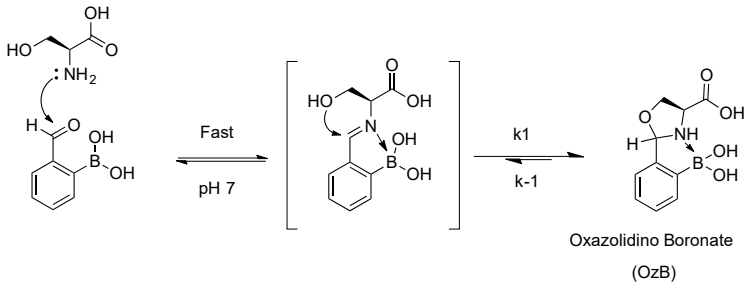


Figure 3.4.1. Proposed mechanism for OzB formation between 2-FPBA and serine

Unfortunately, when 2-FPBA was mixed with serine at low millimolar concentration, little to now reaction was observed. This result is not surprising, as the hydroxyl group from serine is much less nucleophilic than the thiol group from cysteine or the amine group from Dap, which makes it less likely to attack the imine intermediate to from a five-member ring structure (a more stable product). However, when the concentration of serine increases to 100 mM, we observed the emergence of the OzB product (**Figure 3.4.2b**). Although most of the 2-FPBA stays in the form of iminoboronate, there was ~20% product that can be attributed to oxazolidine

formation. No significant change was observed when the reaction was monitored five days later, indicating that this reaction is also reversible and reached equilibrium in a short time (< 30 min). Interestingly, the product is stable enough to be detected by LC/MS (**Figure 3.4.2c**), although we do not have enough data to calculate the percentage that was hydrolyzed during this process. Not surprisingly, the iminoboronate product was completely hydrolyzed in LC/MS thus could not be detected.

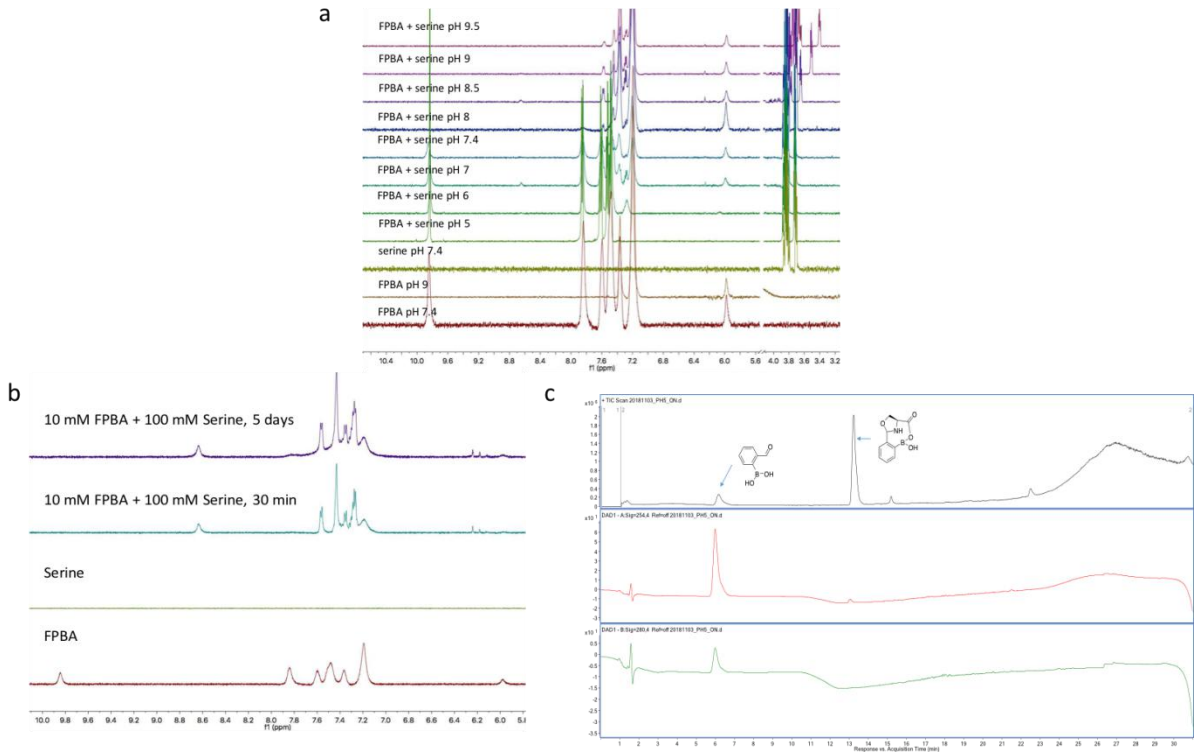


Figure 3.4.2. a) 2-FPBA and serine both at 10 mM under different pH conditions showed little reaction; b) 10 mM 2-FPBA and 100 mM serine showed 20% OzB formation and 80% iminoboronate formation at pH 7.4; c) LC/MS of 2-FPBA-serine conjugate

3.4.2 OzB mediated acyl transfer

Encouraged by the OzB formation, we were wondering if serine could be quantitively converted to a stable product. However, increasing the concentration of

2-FPBA would certainly limit the application of this chemistry. We hypothesized that we could learn from the reaction discussed in **chapter 3.2** and include an extra acyl transfer step following OzB formation. This step could potentially trap the OzB conjugate and drive the reaction to completion.

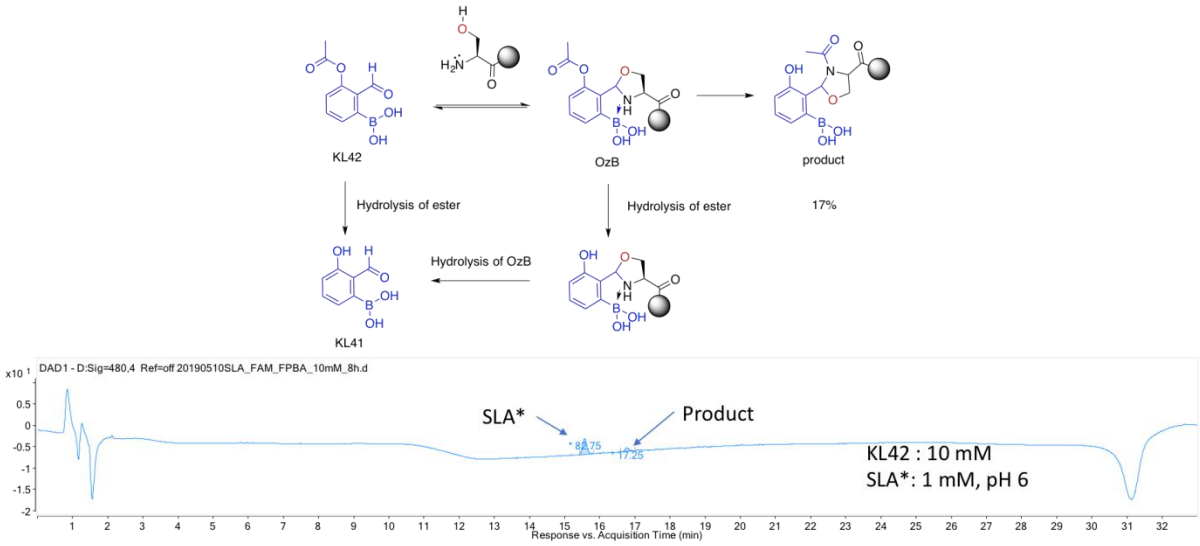


Figure 3.4.3. Reaction of KL42 (10 mM) and SLA* (1 mM) at pH 6.0 after 8 hrs gave 17% conversion

We carried out a reaction between KL42 and a short peptide containing a N-terminal serine residue, SLA* (Ser-Leu-Ala-FAM, in which FAM was coupled through a Dap residue at C-terminus). To our delight, we observed the desired product on LC/MS. Treating 1 mM SLA* peptide with 10 mM KL42 for 8 hrs yielded ~17% conjugate with a mass corresponding to acylated OzB. This is already a great improvement compared to the reaction between 2-FPBA and serine in which we could not observe any reaction at such a low concentration. The low yield can be partially attributed to the difficulty of OzB formation. Based on the result from **Figure 3.4.2a**, when the reaction was carried out at 10 mM concentration, the majority of KL42 stayed unreacted and was likely to be hydrolyzed before it could conjugate with the peptide to form OzB.

3.4.3 Tris with 2-FPBA and 2-APBA[3]

Although we are encouraged by the OzB formation, the requirement for such a high concentration of starting material or long incubation time limits the potential for this reaction on serine. Before we could optimize the structure of KL42, we were also curious if a similar reaction would happen between 2-FPBA and substrates that are structurally similar to serine. Tris base, which is the main ingredient in the often-used Tris buffers, has the necessary functional groups that are required for OzB formation.

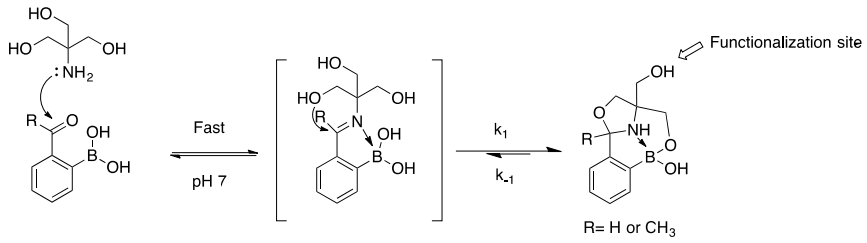


Figure 3.4.4. Proposed mechanism for OzB formation between 2-FPBA and Tris

As a matter of fact, when Tris and 2-FPBA were mixed at 100 mM, both starting materials were completely converted to products. We observed a sharp peak at 6.21 ppm, which is close to the characteristic peak of TzB or IzB complex. Interestingly, we observed the same phenomenon with 2-APBA. For 2-APBA and Tris conjugation, the acetyl peak from 2-APBA shifted from 2.56 ppm to 1.77 ppm, similar to the reported reaction between 2-APBA and cysteine.[4] The spectrums indicate racemic mixtures for both conjugations. The products for these two reactions were later confirmed by the crystallization.

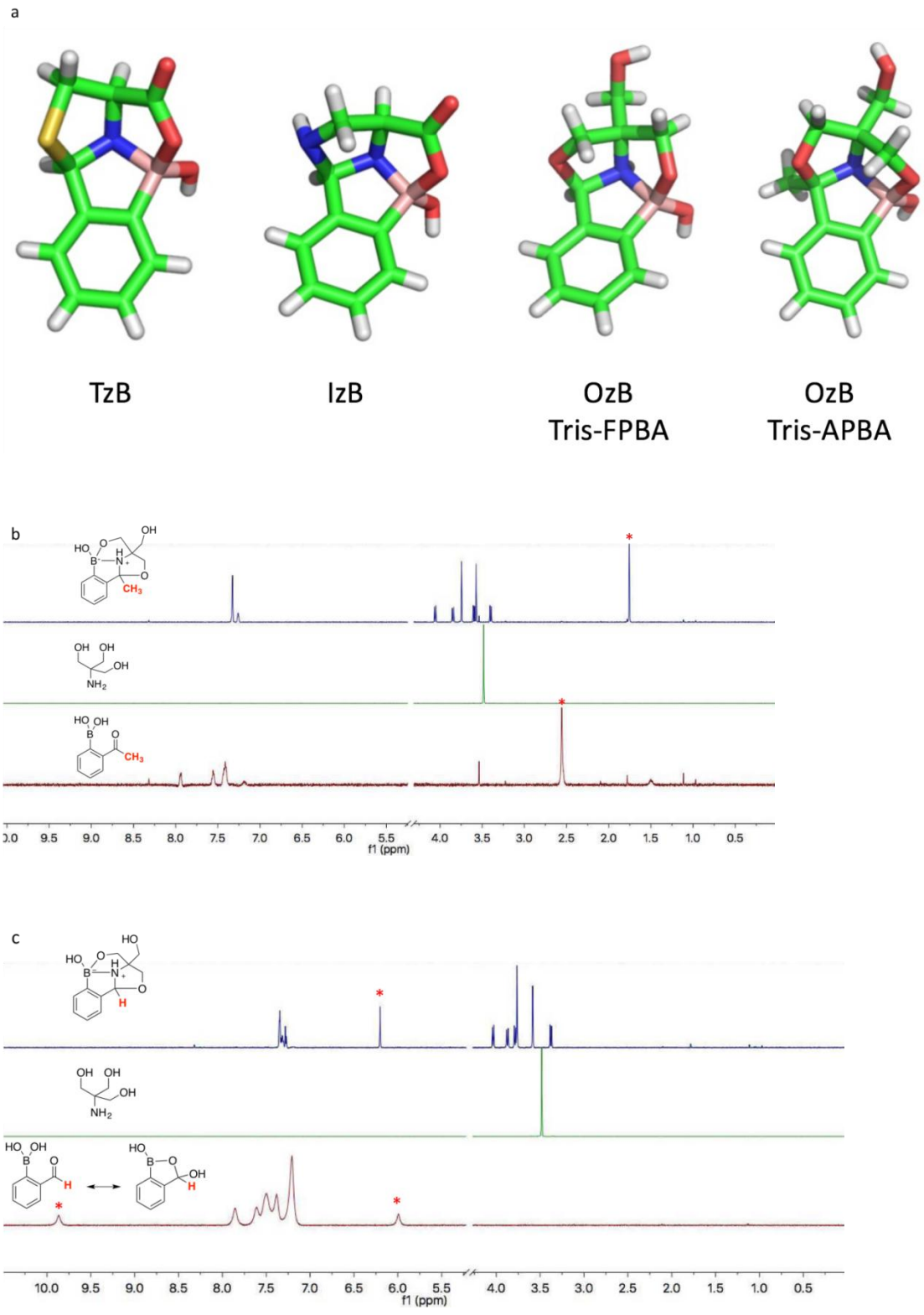


Figure 3.4.5. a) comparison of the crystal structures of TzB, IzB and OzB; b) Tris reacts with 2-APBA; c) Tris reacts with 2-FPBA at 100 mM

Comparing to the reported IzB and TzB chemistry, this reaction has some unique properties despite sharing a similar reaction mechanism. The formation of OzB starts with iminoboronate between the amine and 2-FPBA/2-APBA, followed by nucleophilic attack from one of the hydroxyl groups to form an OzB core; another hydroxyl group cyclizes with boronic acid to give a multicyclic structure. The addition of this boronate ester bond makes it different from TzB and IzB, which will be further discussed in this chapter.

3.4.4 Thermodynamics and kinetics of 2-APBA/2-FPBA-Tris conjugation

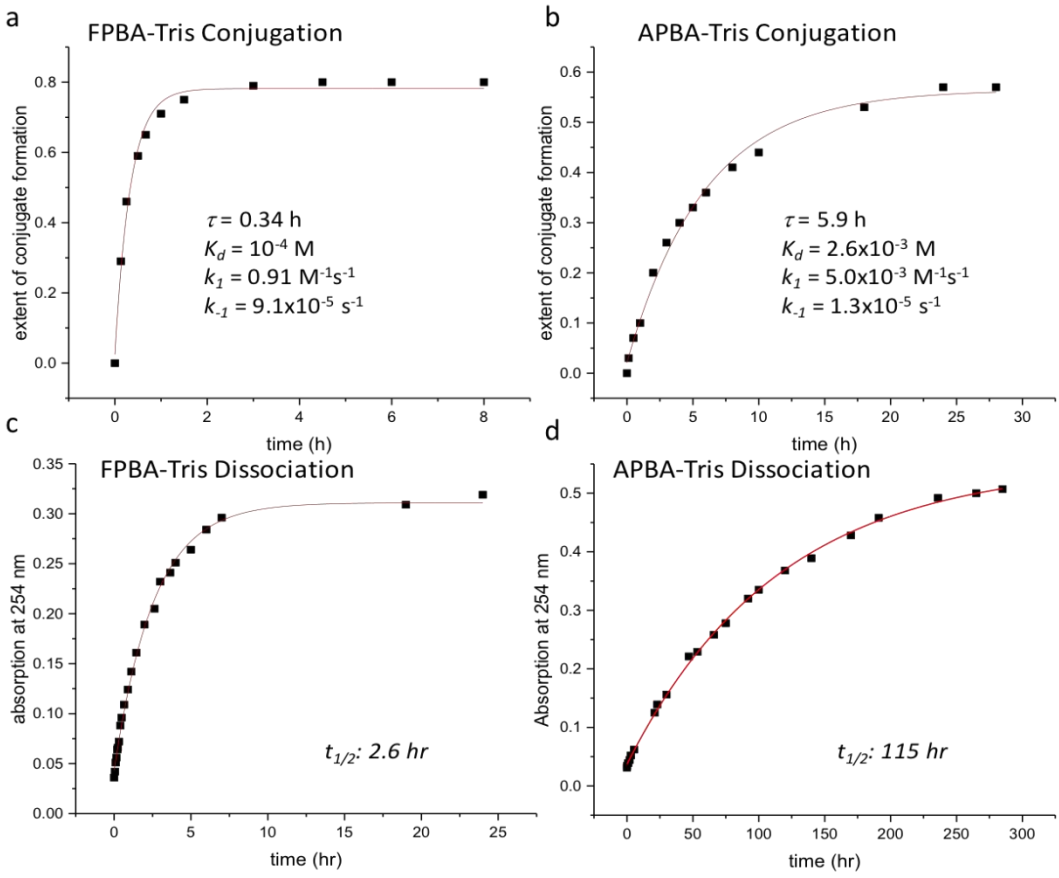


Figure 3.4.6. Association and Dissociation kinetics of Tris conjugation with 2-FPBA or 2-APBA. Association was monitored by ^{1}H-NMR in 20% D_2O at 1 eq.(2 mM for 2-FPBA and 8 mM for 2-APBA). Dissociation process was monitored by UV-vis absorbance at 254 nm in a cuvette, 1 cm path length. The complex was diluted to 100 μM in PBS buffer, pH 7.4

The forward kinetics of this conjugation were determined by mixing equal amount of Tris with 2-FPBA (2 mM) or 2-APBA (8 mM) and monitoring the progression by ^{1}H-NMR. As is illustrated in **Figure 3.4.6a**, 2-FPBA and Tris reached equilibrium after 2 hrs to give 80% conversion. The conjugation of 2-APBA with Tris was significantly slower and it took ~30 hr to reach 57% conversion at equilibrium. These two reactions followed the mechanism of relaxation kinetics. After plotting the data, the forward reaction rate constant (k_1) was determined to be 0.9 and 0.005 $M^{-1}s^{-1}$ for 2-FPBA and 2-APBA respectively. The faster reaction of 2-FPBA can be explained by faster iminoboronate formation due to less steric hinderance compared to 2-APBA.

The rate constants of the backward reactions (k_{-1}) were determined to be 9.1×10^{-5} and 1.3×10^{-5} s^{-1} respectively, which indicate much slower dissociation kinetics in comparison to TzB or IzB complexes. The dissociation kinetics was determined by a dilution experiment and the process is monitored by the increase of absorbance at 254 nm. The half-life for 2-FPBA-Tris and 2-APBA-Tris was calculated to be 2.6 hrs and 115 hrs respectively. For comparison, the TzB complex has a half-life of less than an hour. Based on the distinct difference of dissociation kinetics, we believe Tris conjugation is suitable for various biological applications such as slow drug release.

3.4.5 Bio-compatibility of Tris conjugation

We next examined the bio-compatibility of Tris conjugation in complex media. These reactions at one stoichiometry (0.5 mM for 2-FPBA and 2 mM for 2-APBA) were found to be minimally affected by 10% FBS in PBS or *E. coli* cell lysate (**Figure 3.4.7**).

We also tested the toxicity of 2-APBA and 2-FPBA against *E. coli* cells. The results showed that neither 2-FPBA or 2-APBA showed any toxicity towards bacterial cells at indicated concentrations (**Figure 3.4.7c**). These experiments illustrated the excellent biocompatibility of Tris conjugation.

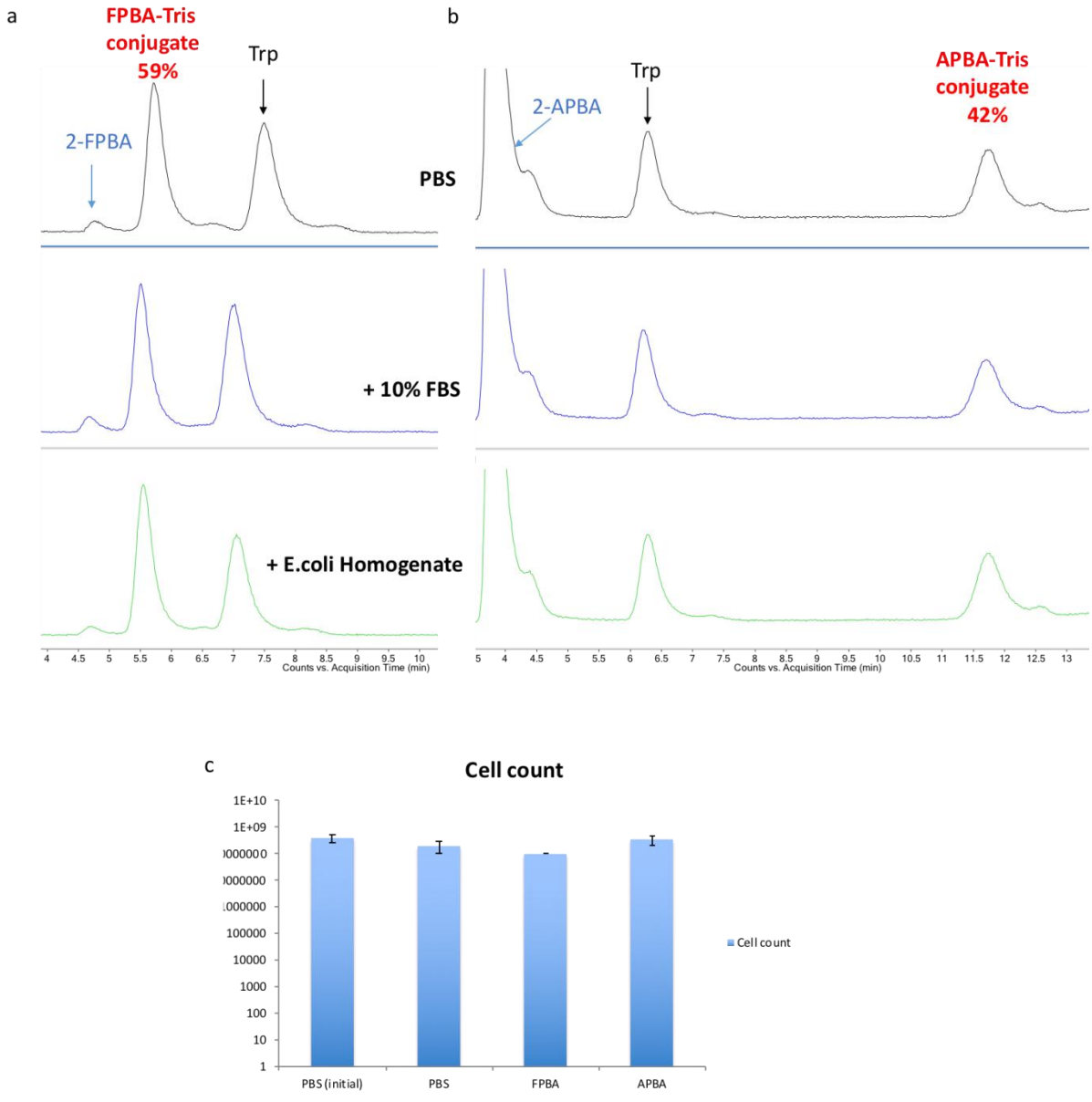


Figure 3.4.7. Tris conjugation with a) 2-FPBA and b) 2-APBA in different media. The results were analyzed by LC/MS.50 μM tryptophan was added as internal standard; c) Toxicity of 2-FPBA (2 mM) and 2-APBA (8 mM) were tested against *E. coli.* This assay was carried out by Michael Kelly.

3.4.6 Conjugation in presence of cysteine

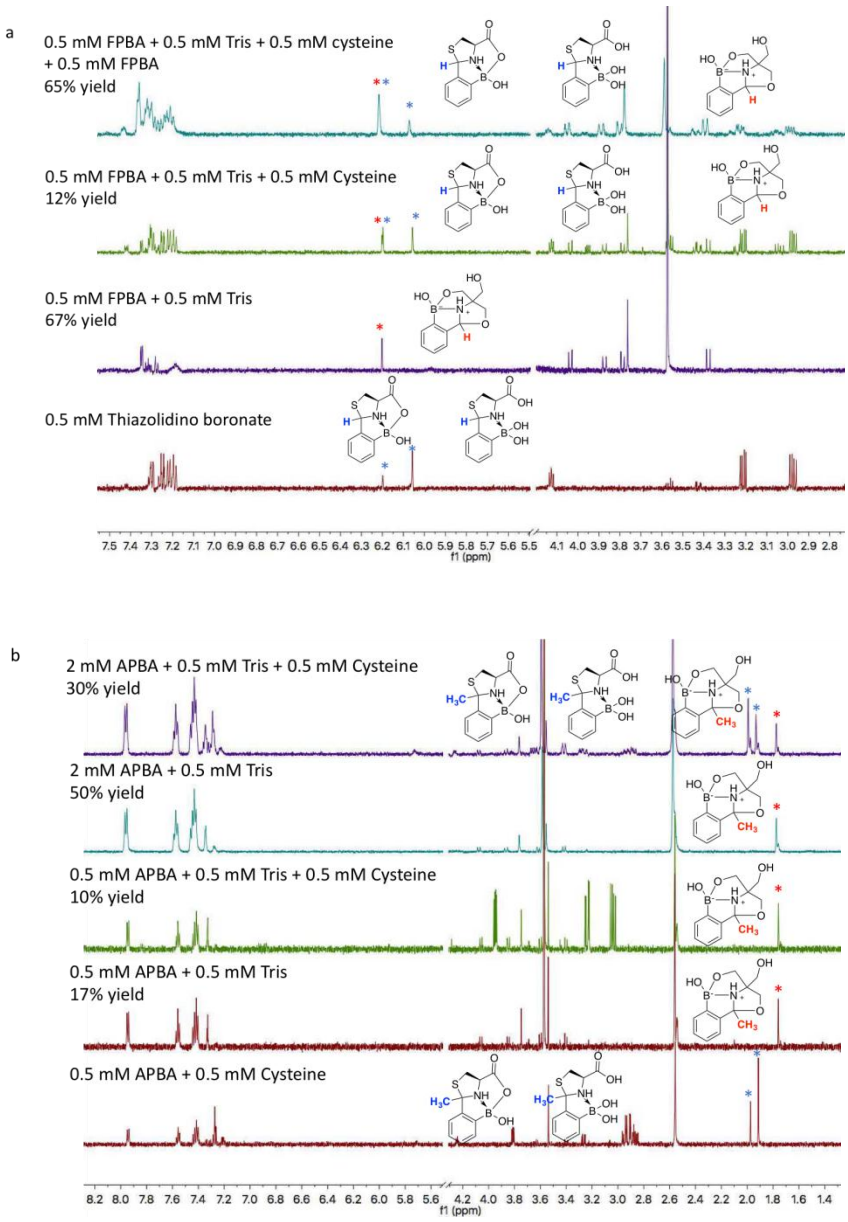


Figure 3.4.8. Tris conjugation with a) 2-FPBA and b) 2-APBA in presence of cysteine

As 2-FPBA and 2-APBA are known to react with free cysteine. We are interested to know how much, if at all, these conjugations will proceed in presence of this competitor. At equimolar concentration (0.5 mM), free cysteine significantly inhibited the 2-FPBA-Tris conjugation (**Figure 3.4.8**) and only 12% Tris conjugate was detected when the reaction reached equilibrium. However, adding another

equivalent of 2-FPBA to the reaction lead to an increase in Tris conversion (65%), comparable to that of cysteine-free conditions. The NMR data only showed the co-existence of the TzB and Tris conjugates with no other side product detected. These results suggest that 2-FPBA can be hijacked by free cysteine resulting in a decreased yield. However, when 2-FPBA was used in excess compared to free cysteine, it can still elicit efficient reaction. A similar phenomenon was observed in the case of 2-APBA.

3.4.7 Resistance towards oxidation

A minor flaw of boron related chemistry is the propensity for boronic acid to get oxidized in presence of oxidative species, even in neutral PBS buffer. Surprisingly, we noticed that APBA-Tris conjugates showed impressive stability towards oxidation. We compared the oxidative stability of these Tris conjugates together with TzB complex against 0.5 mM H_2O_2. After overnight incubation, we observed over 95% oxidation of the TzB complexes, in contrast to 6% and 56% for APBA-Tris and FPBA-Tris conjugates, respectively. It is worth noting that these conjugation reactions can still proceed in the presence of equal molar H_2O_2, although to a lesser extent due to the oxidation of the reactants. This experiment indicates that APBA-Tris or even FPBA-Tris conjugate can serve as a better linkage for cyclic peptides or peptide-drug conjugates in a relatively oxidative environment.

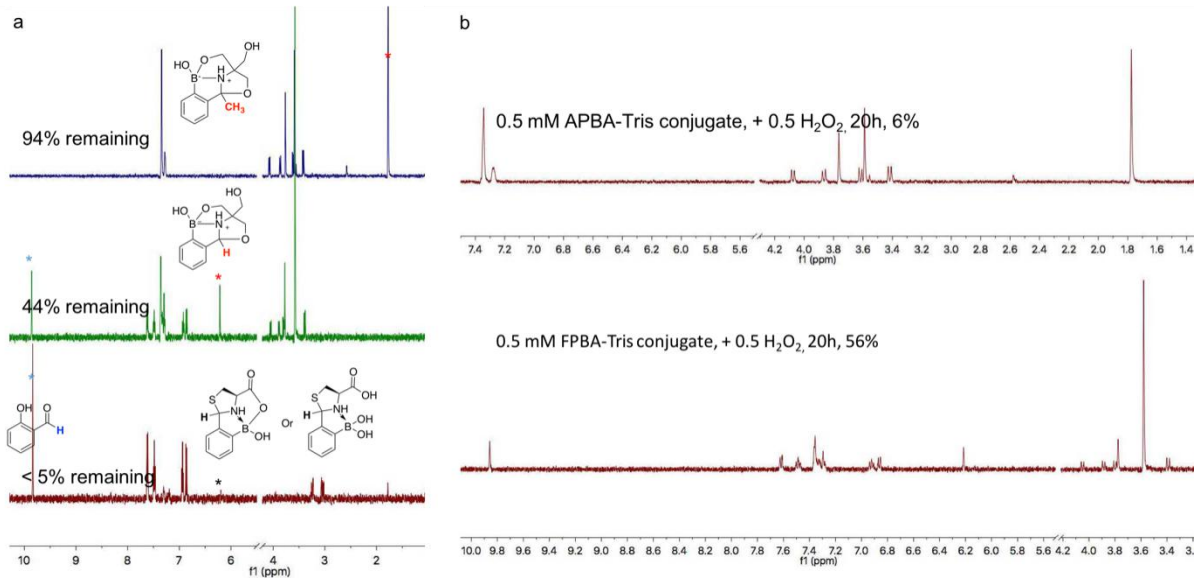
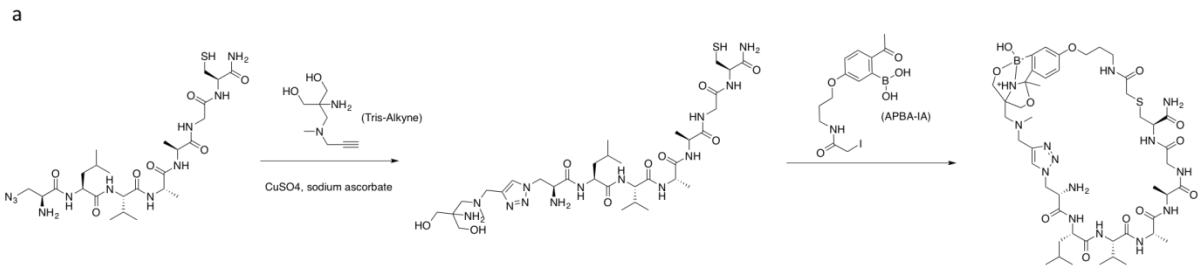


Figure 3.4.9. a) Stability of APBA-Tris, FPBA-Tris and TzB complexes in presence of 0.5 mM H_2O_2. APBA-Tris and FPBA-Tris conjugates showed superior stability against oxidation compared to TzB complex; b) Tris conjugation still proceeds in presence of H_2O_2

3.4.8 Peptide cyclization mediated by Tris conjugation

As a bio-compatible and stable conjugation, we were curious if this chemistry can be applied within a peptide to induce peptide cyclization. We synthesized an alkyne derivative of Tris, which was incorporated into a short peptide KL31 through the following scheme. Briefly, Tris-Alkyne reacted with the azidyl functional group in KL31 through CuAAC chemistry. APBA-IA was then mixed with the peptide in PBS to react with the free cysteine residue.



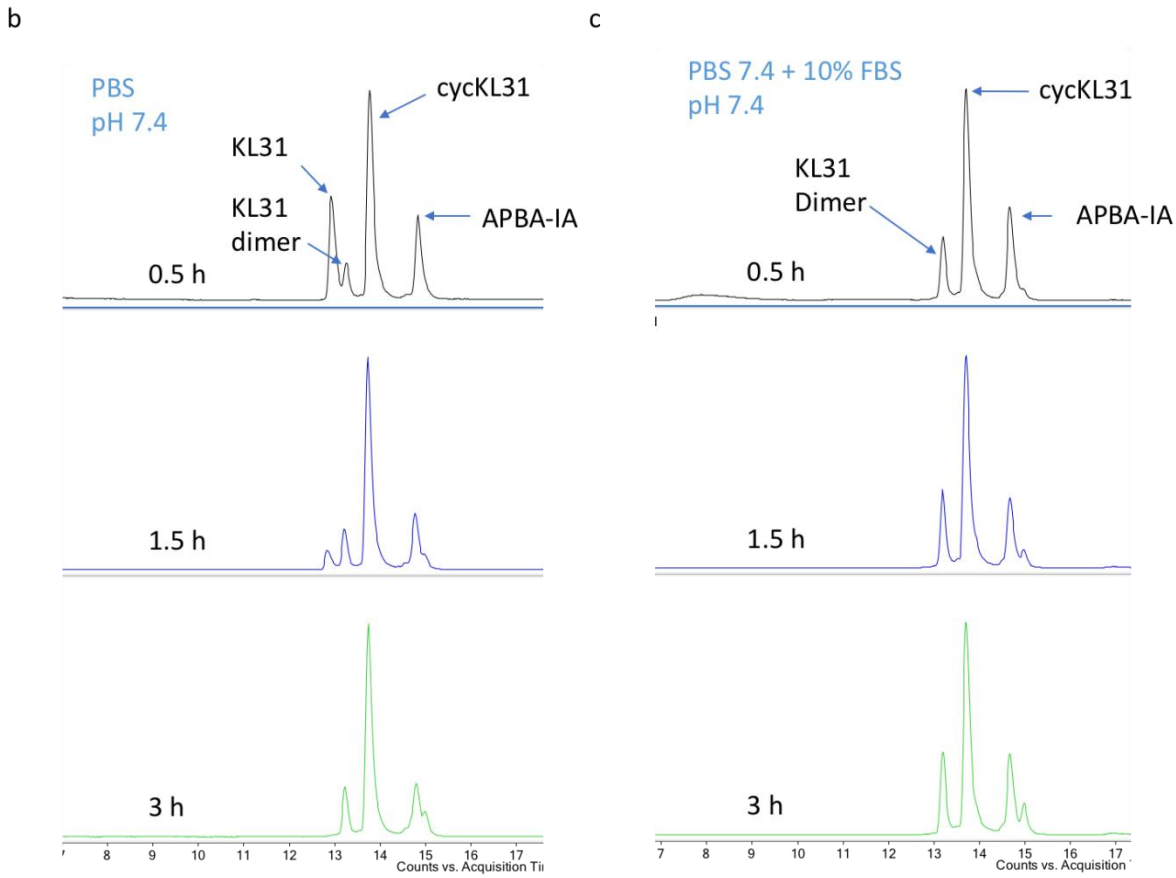


Figure 3.4.10. a) Synthetic scheme for cyclic KL31; b) KL31 cyclization in PBS and c) 10% FBS was monitored by LC/MS

A cyclic peptide was obtained within a couple of hours after the addition of APBA-IA, as is illustrated by the LC/MS data (**Figure 3.4.10b**). Except dimerized peptide and unreacted APBA-IA, no additional species including linear KL31 were detected. This result indicates that the rate limiting step is cysteine alkylation and that intramolecular APBA-Tris conjugation leads to a rapid peptide cyclization. Importantly, this cyclization process was not affected by the addition of 10% FBS. CD spectrums suggested linear and cyclic KL31 both lacked secondary structure.

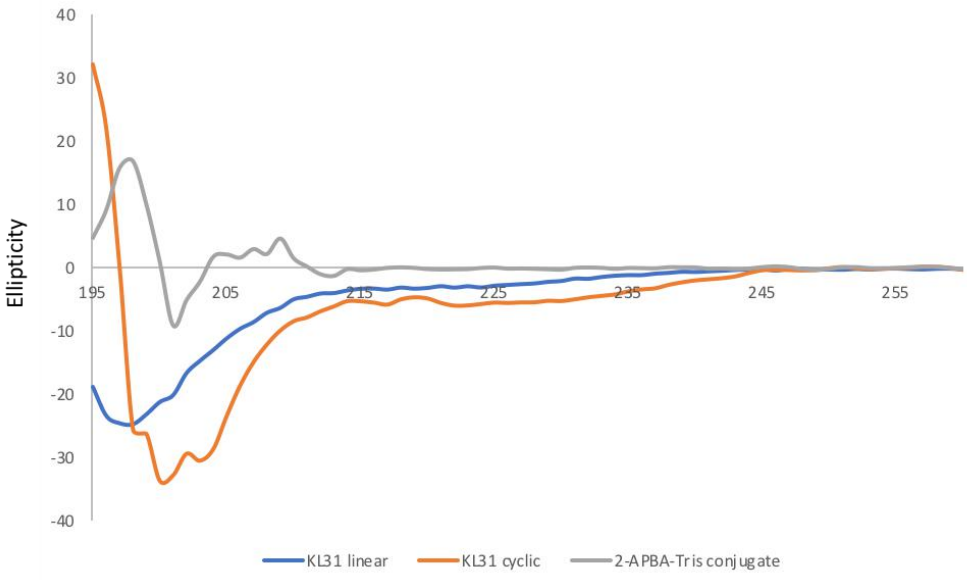


Figure 3.4.11. CD spectrum showing a minimum at 198 nm or 200 nm for linear or cyclic KL31, respectively, indicating that the peptide is structureless before and after cyclization.

3.4.9 Conclusion

In this chapter, we discussed a biocompatible conjugation between Tris and 2-APBA/2-FPBA. Analogous to TzB and IzB formation, the OzB complex shares several similarities (reversibility, initiated by iminoboronate formation, etc.). On the other hand, it presents a significantly different stability towards dilution as well as oxidation. When this chemistry is applied in peptides, it can induce rapid peptide cyclization even in blood serum. Although it requires relatively higher concentration to form the conjugates, the slow dissociation kinetics and excellent biocompatibility can be attractive in various biological applications.

As an extension of the TzB chemistry, we have examined the reaction between 2-FPBA/KL42 and serine. The result suggested that modification on N-terminal serine is much more difficult. Treating SLA peptide with 10 mM KL42 for 8 hrs only gives ~17% yield. The low yield can be attributed to the difficulty of OzB formation.

A possible way to improve the yield would be accelerating the acyl transfer step. Our current plan is to synthesize a thioester derivative of 2-FPBA to enable a faster acyl transfer,[5] which will potentially trap the OzB intermediate and drive the reaction to completion.

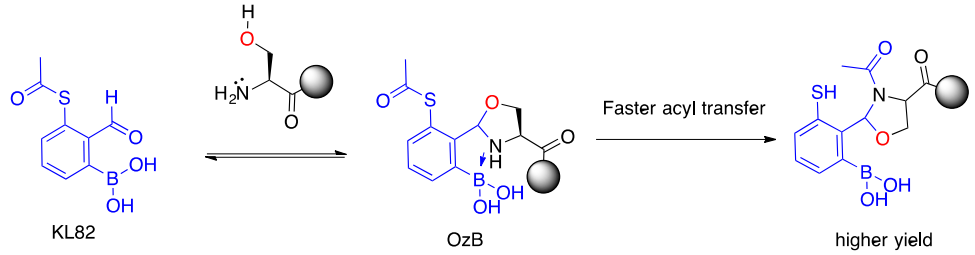
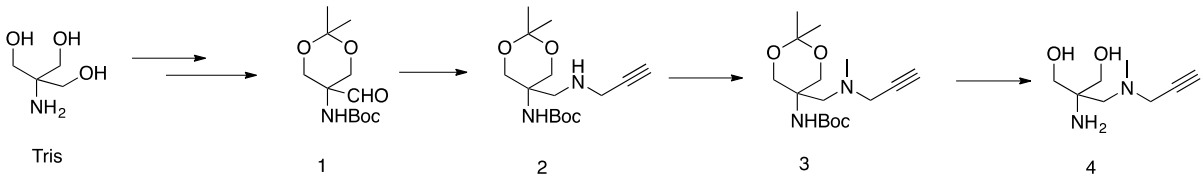


Figure 3.4.12. The faster acyl transfer of a thioester derivative would presumably lead to higher conversion

3.4.10 General information

3.4.10.1 Synthesis of small molecules



Scheme 3.4.1. Synthesis of Tris derivative

Synthesis of **2**

1 was synthesized according to a previously reported procedure.[6] **1** (259 mg, 1 mmol) was dissolved in 5 mL of ethanol, to which was added 55 mg of propargylamine. The mixture was allowed to stir for 2 hours at room temperature before the addition of 75.4 mg of $NaCNBH_3$. The reaction was stirred overnight and

quenched with 30 mL of water. The product was extracted by EtOAc (2 × 50 mL). The organic layer was combined, washed with 50 mL of saturated sodium chloride and dried over sodium sulfate. After solvent removal, the product was purified by silica column (30% EtOAc/Hexane, 1% triethylamine) to give a yellow solid (153 mg, 51% yield). ^{1}H-NMR data suggest the product exists as an ethanol adduct.

^{1}H NMR (500 MHz, Chloroform-*d*) δ 5.01 (s, 1H), 4.06 (d, *J* = 11.7 Hz, 2H), 3.72 – 3.61 (m, 2H), 3.39 (d, *J* = 2.5 Hz, 2H), 2.94 (s, 2H), 2.20 (t, *J* = 2.4 Hz, 1H), 1.47 – 1.31 (m, 15H).

^{13}C NMR (126 MHz, Chloroform-*d*) δ 154.94, 98.60, 98.45, 81.68, 71.66, 64.46, 64.03, 53.18, 51.85, 49.92, 38.77, 28.31, 28.27, 24.30, 22.89.

MS-ESI^{+}: m/z calculated for $C_{15}H_{27}N_2O_4$ $[M+H]^{+}$ 299.1965, found 299.1910.

Synthesis of **3**

2 (90 mg, 0.3 mmol) was dissolved in 1 mL of Acetone and placed on ice bath. K_2CO_3 (210 mg) was added to the solution and the mixture was stirred for 10 min. Iodomethane (42.5 mg) was added and the reaction was allowed to stir overnight at room temperature and then quenched with 50 mL of water. The product was extracted by EtOAc (3 × 50 mL). The organic layer was combined, washed with 50 mL of saturated sodium chloride and dried over sodium sulfate. After solvent removal, the product was purified by silica column (20% EtOAc/Hexane) to give a white solid (60 mg, 64% yield).

^{1}H NMR (500 MHz, Chloroform-*d*) δ 4.88 (s, 1H), 4.05 (d, *J* = 10.4 Hz, 2H), 3.74 (d, *J* = 11.6 Hz, 2H), 3.31 (d, *J* = 2.4 Hz, 2H), 2.85 (s, 2H), 2.40 (s, 3H), 1.42 (t, *J* = 15.4 Hz, 15H).

^{13}C NMR (126 MHz, Chloroform-*d*) δ 154.85, 98.30, 79.56, 72.68, 64.30, 56.39, 52.41, 48.14, 44.59, 24.12, 23.28.

MS-ESI^{+}: m/z calculated for $C_{16}H_{29}N_2O_4$ $[M+H]^{+}$ 313.2122, found 313.2067.

Synthesis of **4**

3 (31 mg, 0.1 mmol) was treated with 2 mL of 95% trifluoroacetic acid and 5% water for 1 h, twice. After removing TFA and water over vacuum, the product was lyophilized and used without purification.

^{1}H NMR (500 MHz, Methanol-d_4) δ 3.72 (d, *J* = 2.3 Hz, 2H), 3.69 (d, *J* = 2.0 Hz, 4H), 3.08 (s, 2H), 2.90 (t, *J* = 2.3 Hz, 1H), 2.70 (s, 3H).

^{13}C NMR (126 MHz, Methanol-d_4) δ 75.78, 75.60, 60.67, 60.34, 55.60, 47.67, 43.21.a

MS-ESI^{+}: m/z calculated for $C_8H_{17}N_2O_2$ $[M+H]^{+}$173.1285, found 173.1217.

3.4.10.2 Peptide Synthesis

KL31 precursor (sequence: AzidoAla-Leu-Val-Ala-Ala-Gly-Cys-NH_2) was synthesized by standard solid phase peptide synthesis on rink amide resin. Fmoc-AzidoAla-OH was synthesized according to literature.[7] The peptide was cleaved off resin using 85% TFA, 5% H2O, 5% phenol and 5% thioanisole and precipitated by cold ether. The crude product was purified via RP-HPLC. **4** was clicked onto the purified peptide with copper catalyst in water. Briefly, 4 mg of peptide was dissolved

in 0.5 mL of DI water, to which was added 3.5 mg of $CuSO_4 \bullet 5H_2O$ and 2.7 mg of sodium L-ascorbate. 7 μL of **4** from 1 M stock in DMF was added to the mixture and the reaction was incubated for 30 min at room temperature. The purified peptides were confirmed by LC/MS to confirm the identity and purity.

3.4.10.3 CD spectrometry

All the samples were filtered before experiments. KL31 was dissolve in PBS to make a 1 mM solution. The spectrums were taken at 25℃, 1 mm path length, scan 190-295 nm with 1 nm step, 10 sec averaging time, 0.33 sec settling time. Spectrums were triplicates with PBS blank subtracted.

3.4.10.4 Toxicity tests for 2-APBA and 2-FPBA

The bacteria (strain ER2738) were grown in LB media to an OD_{600}~ 1.0. Cells were spun down and resuspended in PBS at pH 7.4. 100μL of cells were incubated with 2-APBA (2 mM), 2-FPBA (0.5 mM) or just PBS overnight at 4°C with gentle rocking. Cells were plated before overnight incubation to get an initial cell count and after overnight incubation to monitor cell death. All the experiments were done in triplicates

3.4.11 References

(1) Bandyopadhyay, A.; Cambray, S.; Gao, J. Fast and Selective Labeling of N-Terminal Cysteines at Neutral PH via Thiazolidino Boronate Formation. *Chem. Sci.* **2016**, *7* (7), 4589–4593. https://doi.org/10.1039/C6SC00172F.

(2) Li, K.; Weidman, C.; Gao, J. Dynamic Formation of Imidazolidino Boronate Enables Design of Cysteine-Responsive Peptides. *Org. Lett.* **2018**, *20* (1), 20–23. https://doi.org/10.1021/acs.orglett.7b03116.

(3) Li, K.; Kelly, M. A.; Gao, J. Biocompatible Conjugation of Tris Base to 2-Acetyl and 2-Formyl Phenylboronic Acid. *Org. Biomol. Chem.* **2019**, *17* (24), 5908–5912. https://doi.org/10.1039/c9ob00726a.

(4) Bandyopadhyay, A.; Cambray, S.; Gao, J. Fast Diazaborine Formation of Semicarbazide Enables Facile Labeling of Bacterial Pathogens. *J. Am. Chem. Soc.* **2017**, *139* (2), 871–878. https://doi.org/10.1021/jacs.6b11115.

(5) Raj, M.; Wu, H.; Blosser, S. L.; Vittoria, M. A.; Arora, P. S. Aldehyde Capture Ligation for Synthesis of Native Peptide Bonds. *J. Am. Chem. Soc.* **2015**, *137* (21), 6932–6940. https://doi.org/10.1021/jacs.5b03538.

(6) Ooi, H.; Ishibashi, N.; Iwabuchi, Y.; Ishihara, J.; Hatakeyama, S. A Concise Route to (+)-Lactacystin. *J. Org. Chem.* **2004**, *69* (22), 7765–7768. https://doi.org/10.1021/jo048817o.

(7) Pícha, J.; Buděšínský, M.; Macháčková, K.; Collinsová, M.; Jiráček, J. Optimized Syntheses of Fmoc Azido Amino Acids for the Preparation of Azidopeptides. *J. Pept. Sci.* **2017**, *23* (3), 202–214. https://doi.org/10.1002/psc.2968.

Chapter 4 Conclusions

In this contribution, I described the potential biological applications of boronic acids from two aspects- bacteria targeting and site-specific labeling.

Tremendous effort has been devoted in the area of bacteria targeting due to the emergence of antibiotic resistance. In contrast to conventional antibiotics, we decided to employ reversible covalent interactions in our design. This study is meaningful not only to the field of antibiotic development, but also serves as a proof-of-concept experiment for applying covalent interactions in the design of antibiotics. We have decided to target the lipid components on the bacterial surface due to the conserved structure. Compared to other targets, mutation of lipids requires the involvement of multiple enzymes, meaning that developing resistance to antibiotics of this sort would require mutations in multiple genes, which is an intensive process, even for bacteria.

We have tested several elements that may help to improve the binding affinity of our peptides towards their targets. Firstly, we have designed a linear peptide with five copies of phenylboronic acid warheads and demonstrated that multivalence can significantly enhance target binding. To further increase the potency of our peptide, we changed the warhead to a 'Wulff-type' boronic acid, which indeed showed significant higher binding affinity. However, the function of boronic acids in these peptides was proven to be non-essential for their binding. We also examined the effect of peptide structure in our antibiotic design. Those amphiphilic peptides cyclized by iminoboronate formation demonstrated potent bactericidal activity. Unfortunately, boronic acid once again did not give us any advantage binding to the bacteria. A bigger challenge that we came across was the inhibition effect of serum protein. As a matter of fact, this is problem has also been encountered by other group members.[1]

It is not easy for these rationally designed peptides to get around the problem of serum inhibition, which is the innate drawback of rational design. On the chemical side, we have explored the possibility of utilizing boronate ester or iminoboronate formation in bacterial targeting. Unfortunately, neither of these chemistries was proven to play a critical role in the peptides we have designed. It is not difficult to look for more potent chemistries to enhance binding, but we still cannot predict whether the newly designed molecule can escape the hijack of proteins or other bio-related molecules.

This dilemma led us to select the desired peptides with a screening-based strategy. We attempted to include as much rational design as possible in our library construction. For example, we chose a conserved structure, lipid II as our target and we adopted a cyclic scaffold to mimic the structure of vancomycin. Moreover, we included a 2-APBA headgroup in our sequence, which has been already employed in antibiotic design in our lab.[1,2] Being the first successful synthetic peptide library in the group, the whole process was challenging. Although this library did not give us any positive hits, there are things that we can learn from this experience. Including 2-APBA in peptide design is the novelty of our design. However, it does not ensure success. In fact, we did not detect iminoboronate formation when the concentration of stem peptide went up to the millimolar range. The only explanation is that the rest of the structure did not collaborate with 2-APBA to bind to the target.

One question we have to think about is where we could possibly make a difference if we were to expand upon our current learnings. It is firstly clear that our current design has some flaws due to technical issues. For instance, the beads may have autofluorescence that affects the bead selection step under microscope. We

could use a fluorophore that emits at a longer wavelength to avoid such a problem. Also, a beads-sorting flow cytometer would certainly improve the whole process. Besides these instrumental limitations, we can also improve our design between both the choice of target as well as library design. For instance, targeting stem peptide of lipid II could be a significantly challenging, which has proven to the case after a number of studies in our lab. In fact, we have to date not been able to select potent peptide binder towards this target with different screening platforms including phage display. This is partially due to the relatively small size of the stem peptide. There are not many 'hot spots' that we can go after when we design our peptides. In retrospect, leaving out the oligosaccharide part from lipid II may not be the best choice. The addition of multiple hydroxyl groups for the sugars provides a high chance for extra hydrogen bonding. Chemical synthesis of lipid II was reported in 2001,[3] which could be the guidance for us to synthesize its derivatives for screening purposes. Also, as is discussed earlier, we could potentially build a larger library or even use a different screening platform to maximize our chance to find a potent peptide binder to lipid II.

The other part of this thesis describes the devotion we made to the field of site-specific modification on proteins. These chemistries were developed based on thiazolidino boronate (TzB) formation.[4,5] As an extension of the iminoboronate conjugation, TzB has its advantages over other existing chemistries including NCL and CBT condensation. Besides its supreme kinetics, TzB formation also has significantly better selectivity. Although both NCL and CBT condensation are stated to be site-specific, they do exhibit side reactions with internal cysteine residues in our own hands. Another attractive feature of TzB chemistry is that it is able to proceed to almost completion within an hour period at low micromolar concentrations, which is not possible for the other two chemistries.

We have expanded the TzB chemistry on N-terminal cysteine to N-terminal Dap as well as Tris base and serine/threonine. Due to the fact that cysteine is not the most stable amino acid in the proteome, we sometimes would prefer an inert handle before it gets modified so the protein can be stored for a longer period. Often times, reducing reagents such as TCEP and DTT are required to prevent oxidation or unexpected modification of the cysteine residues. As a matter of fact, N-terminal cysteine containing protein is sometimes modified by pyruvic acid, which is an abundant metabolite in *E.coli*.[6]

2-FPBA reacts with Dap to form imidazolidino boronate (IzB). The IzB and TzB chemistry shares a lot in common. For instance, they are both rapidly reversible and follows similar mechanisms. The dissociation constant of IzB, is much higher than TzB due to the decreased nucleophilicity of an amine side chain. However, this compromised potency helped us to design novel cyclic peptides and peptide pairs that are sensitive to free cysteine. OzB on the other hand, is more challenging and at the same time, attractive. Most ligation chemistries on N-terminal serines are carried out in organic conditions[7,8]. However, some proteins or peptides can be denatured in organic solvents so a chemistry that can proceed in water is still needed. There is only one special case where conjugation was realized on a serine-like structure in aqueous solution.[9] In this reaction, the reagents formed micelles to increase the local concentration and prevent hydrolysis, which indicates that this chemistry is not generally applicable in other circumstances. Conjugation with N-terminal serine is difficult with 2-FPBA alone. We were only able to achieve minimal conjugation at high millimolar concentrations. However, we might be able to find a solution to this problem with the help of an acyl transfer step.

Acyl transfer reactions have been widely used in conjugation chemistries, especially NCL. We have found that our TzB product was stabilized by this step. It follows a unique mechanism in which the acyl transfer is accelerated by the rapid TzB formation. This extra step turns TzB chemistry into an irreversible reaction allowing the TzB chemistry to have a broader range of biological applications. For example, during protein pull down or purification after N-terminal cysteine labelling, it prevents the decomposition of the TzB complex. We have also found that the acyl transfer step is pH sensitive with a much faster kinetics at pH 5-6, which lies in the range of intracellular acidity.[10,11] This indicates that our chemistry can be potentially applied intracellularly in the future.

Acyl transfer also provides a way to label N-terminal serine/threonine. We could potentially make use of OzB formation and an acyl transfer reaction to achieve complete conversion in pure aqueous solution. Our preliminary data suggests that OzB mediated transfer is possible. Incubating 10 mM KL42 with a peptide containing a N-terminal serine residue gave us about 17% yield after 8 hours. Although the conversion is not satisfying, it provides us some guidance on how to improve this reaction. There are two limiting factors that lowers down the yield. Firstly, due to the low affinity of 2-FPBA and serine, there was only a small percentage of OzB intermediate at 10 mM starting material. The majority of KL42 stays free in solution and is hydrolyzed overtime. Secondly, the acyl transfer step is not fast enough to trap the OzB intermediate. Increasing the affinity of 2-FPBA and serine is a non-trivial process because the hydroxyl side chain is not nucleophilic enough to attack the imine intermediate. However, the acyl-transfer step can be potentially accelerated. There are different reports regarding the hydrolytic property and S→N transfer kinetics of a thioester.[12,13] We can conclude that a thioester is equally stable, if not more, than an

ester. Moreover, the S→N transfer kinetics is much faster than O→N transfer (30 times or higher). We are also considering a selenoester derivative of KL42. Se→N transfer in NCL is attracting more attention due to its faster kinetics compared to conventional S→N mediated ligation.[14] A thioester or selenoester derivative of KL42 could be a promising probe for N-terminal serine/threonine modification in aqueous solution.

References

(1) Bandyopadhyay, A.; McCarthy, K. A.; Kelly, M. A.; Gao, J. Targeting Bacteria via Iminoboronate Chemistry of Amine-Presenting Lipids. *Nat. Commun.* **2015**, *6*, 1–9. https://doi.org/10.1038/ncomms7561.

(2) McCarthy, K. A.; Kelly, M. A.; Li, K.; Cambray, S.; Hosseini, A. S.; Van Opijnen, T.; Gao, J. Phage Display of Dynamic Covalent Binding Motifs Enables Facile Development of Targeted Antibiotics. *J. Am. Chem. Soc.* **2018**, *140* (19), 6137–6145. https://doi.org/10.1021/jacs.8b02461.

(3) VanNieuwenhze, M. S.; Mauldin, S. C.; Zia-Ebrahimi, M.; Winger, B. E.; Hornback, W. J.; Saha, S. L.; Aikins, J. A.; Blaszczak, L. C. The First Total Synthesis of Lipid II: The Final Monomeric Intermediate in Bacterial Cell Wall Biosynthesis. *J. Am. Chem. Soc.* **2002**, *124* (14), 3656–3660. https://doi.org/10.1021/ja017386d.

(4) Bandyopadhyay, A.; Cambray, S.; Gao, J. Fast and Selective Labeling of N-Terminal Cysteines at Neutral PH: Via Thiazolidino Boronate Formation. *Chem. Sci.* **2016**, *7* (7), 4589–4593. https://doi.org/10.1039/c6sc00172f.

(5) Faustino, H.; Silva, M. J. S. A.; Veiros, L. F.; Bernardes, G. J. L.; Gois, P. M. P. Iminoboronates Are Efficient Intermediates for Selective, Rapid and Reversible N-Terminal Cysteine Functionalisation. *Chem. Sci.* **2016**, *7* (8), 5052–5058. https://doi.org/10.1039/c6sc01520d.

(6) Gentle, I. E.; De Souza, D. P.; Baca, M. Direct Production of Proteins with N-Terminal Cysteine for Site-Specific Conjugation. *Bioconjug. Chem.* **2004**, *15* (3), 658–663. https://doi.org/10.1021/bc049965o.

(7) Li, X.; Lam, H. Y.; Zhang, Y.; Chan, C. K. Salicylaldehyde Ester-Induced Chemoselective Peptide Ligations: Enabling Generation of Natural Peptidic Linkages at the Serine/Threonine Sites. *Org. Lett.* **2010**, *12* (8), 1724–1727. https://doi.org/10.1021/ol1003109.

(8) Zhang, Y.; Xu, C.; Lam, H. Y.; Lee, C. L.; Li, X. Protein Chemical Synthesis by Serine and Threonine Ligation. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (17), 6657–6662. https://doi.org/10.1073/pnas.1221012110.

(9) Rudd, A. K.; Devaraj, N. K. Traceless Synthesis of Ceramides in Living Cells Reveals Saturation-Dependent Apoptotic Effects. *Proc. Natl. Acad. Sci. U. S. A.* **2018**, *115* (29), 7485–7490. https://doi.org/10.1073/pnas.1804266115.

(10) Han, J.; Burgess, K. Fluorescent Indicators for Intracellular PH. *Chem. Rev.* **2010**, *110* (5), 2709–2728. https://doi.org/10.1021/cr900249z.

(11) Roos, A.; Boron, W. F. Intracellular PH. *Physiol. Rev.* **1981**, *61* (2), 296–434. https://doi.org/10.1152/physrev.1981.61.2.296.

(12) Yang, W.; Drueckhammer, D. G. Understanding the Relative Acyl-Transfer Reactivity of Oxoesters and Thioesters: Computational Analysis of Transition State Delocalization Effects. *J. Am. Chem. Soc.* **2001**, *123* (44), 11004–11009. https://doi.org/10.1021/ja010726a.

(13) Raj, M.; Wu, H.; Blosser, S. L.; Vittoria, M. A.; Arora, P. S. Aldehyde Capture Ligation for Synthesis of Native Peptide Bonds. *J. Am. Chem. Soc.* **2015**, *137* (21), 6932–6940. https://doi.org/10.1021/jacs.5b03538.

(14) Durek, T.; Alewood, P. F. Preformed Selenoesters Enable Rapid Native Chemical Ligation at Intractable Sites. *Angew. Chemie - Int. Ed.* **2011**, *50* (50), 12042–12045. https://doi.org/10.1002/anie.201105512.