Cell-Selective Chemoproteomics for Biological Discovery

Thesis by
Shannon Elizabeth Stone

In Partial Fulfillment of the Requirements for
the degree of
Doctor of Philosophy in Chemistry

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2018
(Defended July 20th, 2017)

Shannon Elizabeth Stone
ORCID: 0000-0002-6617-3874

ACKNOWLEDGEMENTS

The greatest part of my graduate school experience has been the privilege to learn from many brilliant people and be a part of the exceedingly collaborative nature of this scientific community. First, I thank my advisor and mentor in every sense of the word, Prof. Dave Tirrell, for his continued guidance and support. In addition to his immense scientific insight and encouragement throughout the years, I am continuously impressed by his degree of integrity and thoughtfulness, both of which I strive to emulate. Thanks also to the other members of my thesis committee: Prof. Sarkis Mazmanian for many illuminating discussions, both scientific and personal, Prof. Bob Grubbs for advice and guidance, and Prof. Long Cai for astute recommendations on my project.

I have been fortunate to share my time here with many members of the Tirrell lab, whose patience, brilliance, and humor made the time fly by. I will always cherish the days of the Best Office and summer softball with the Tools of Synthetic Biology.

I greatly enjoyed working with Judy Shon, a talented Caltech undergraduate, over the past years. The Proteome Exploration Laboratory is run by the most wonderful and knowledgeable team, who all contributed to most chapters of this thesis. Thank you also to members of both the Newman and Mazmanian labs for sharing knowledge and resources. The staff at the Beckman Imaging Center has also been exceedingly helpful. Thanks to Prof. Linda Hsieh-Wilson for her support, as well as former members of the Hsieh-Wilson lab for their advice and expertise, and especially Dr. Abby Pulsipher, who continues to be a source of scientific and personal inspiration. I was also fortunate to spend a month at the University of Chicago, learning from some of the world's top researchers in *Staphylococcus aureus*, for which I am very grateful.

Members of the Caltech community including the diversity center, health & counseling centers, and graduate office have all immensely enhanced experience: thank

you. I also thank the Chemistry division for its support, especially for help in starting the Women in Chemistry group. I greatly appreciate the enthusiastic lecturers I had during my time at Caltech, especially Justin Bois, who taught me the joys of data analysis.

To my friends who remind me that there is more to the world than science, and who drink wine with me to celebrate or drink wine with me to commiserate, thank you.

I am forever grateful to my family for their love and encouragement from the other side of the country. I would not be where I am today without each of you.

Finally, to Blake, Olive, and Slevin; thank you, I love you, and let the next great adventure begin.

ABSTRACT

Cellular protein synthesis changes rapidly in response to internal and external cues in ways that vary from cell to cell. Global proteomic analyses of microbial communities, tissues, and organisms have provided important insights into the behavior of such systems, but can obscure the diversity of responses characteristic of different cellular subpopulations. Recent advances in cell-specific proteomics—fueled in part by the development of bioorthogonal chemistries, more sensitive mass spectrometers and more advanced mining algorithms—have yielded unprecedented glimpses into how proteins are expressed in space and time. Whereas previous cell-specific proteomic analyses were confined to abundant cells in relatively simple systems, recent advances in chemoproteomics allow researchers to map the protein expression patterns of even rare cells in complex tissues and whole organisms.

Chapter 1 highlights recently developed strategies for cell-selective proteomics, including metabolic labeling strategies such as bioorthogonal noncanonical amino acid tagging (BONCAT). BONCAT is a chemoproteomic technique that enables temporal labeling of proteins using noncanonical amino acids. In the cell-selective version of BONCAT, expressing a mutant aminoacyl-tRNA synthetase under the control of cell-specific genetic elements affords cellular resolution; only cells of interest can selectively incorporate a noncanonical amino acid into proteins for subsequent detection and identification. Chapter 2 details protocols to set up a cell-selective BONCAT system.

While BONCAT had previously been applied to studies of microbial pathogenesis in tissue culture-based models of infection, we sought to further develop the method to identify the proteome of methicillin-resistant *Staphylococcus aureus* (MRSA) within a mouse model of infection, as detailed in Chapter 3. We used this technique to enrich for staphylococcal proteins made within the host and in addition to finding many factors known to be

important for infection, we also found many that had not previously been associated with infection. Screening several of these previously unknown factors *in vivo* led to the discovery of a novel protein important for MRSA infection. This unbiased approach to cell-selectively label pathogenic proteins during infection could be used as a global discovery tool for novel anti-infective strategies.

In Chapter 4, we combine this cell-selective BONCAT strategy with microbial identification after passive clarity technique (MiPACT) to visualize both staphylococcal protein synthesis and ribosomal RNA within whole skin abscesses during infection. In Chapter 5, we continue developing cell-selective BONCAT to study microbial protein synthesis in the context of a living mouse by extending the system to *Bacteroides fragilis*, a common human gut commensal.

Finally, cell-selective BONCAT is wholly dependent on the bioorthogonal nature of the azide and its detection reagents. Fishing out an azide-tagged molecule from the rest of the cellular milieu requires optimization of enrichment-based strategies. In Chapter 6, we describe the development of a peptide to quantitate the gain of our enrichments.

While innovations in mass spectrometry and computational algorithms have facilitated the identification and quantification of thousands of proteins simultaneously from complex samples, this abundance of data does not necessarily lead to biological insight. Cell-specific proteomic techniques will play a key role in the identification of the mechanisms that govern cell specialization and that allow organisms to respond to changing environments. Overall, this work demonstrates the power of cell-selective chemoproteomics to ascertain biological insights in complex systems.

PUBLISHED CONTENT AND CONTRIBUTIONS

(1) Pulsipher, A., Griffin, M. E., **Stone, S. E.**, Brown, J. M., Hsieh-Wilson, L. C. "Directing neuronal signaling through cell-surface glycan engineering." J. Am. Chem. Soc. 2014, 136 (19), 6794-6797. DOI: 10.1021/ja5005174

S.E.S. participated in project conception, optimized the TrkA assay, and contributed to the writing of the manuscript. This article's figures are reproduced in part within Appendix B with permission according to the ACS AuthorChoice terms of use.

(2) Pulsipher, A., Griffin, M. E., **Stone, S. E.**, Hsieh-Wilson, L. C. "Long-lived engineering of glycans to direct stem cell fate." Angew. Chem. Int. Ed. **2015**, 54, 1466-1470. DOI: 10.1002/anie.201409258

S.E.S. participated in project conception, created the plasmids, determined conditions for cell labeling, and contributed to the writing of the manuscript. This article's figures are reproduced in part within Appendix C with permission from the publisher.

(3) Griffin, M. E., Jensen, E. H., Mason, D. E., Jenkins, C. L., **Stone, S. E.**, Peters, E. C., Hsieh-Wilson, L. C. "Comprehensive mapping of *O*-GlcNAc modification sites using a chemically cleavable tag." Mol. Biosys. **2016**, 12, 1756-1759. DOI: 10.1039/c6mb00138f

*S.E.S. synthesized and characterized compound **1** in Figure 1A. This article's figures are reproduced in part within Appendix D with permission according to the Creative Commons Attribution 3.0 Unported License terms of use.*

(4) **Stone, S. E.**, Glenn, W. S., Hamblin, G. D., Tirrell, D. A. "Cell-selective proteomics for biological discovery." Curr. Opin. Chem. Biol. **2017**, 36, 50-57. DOI: 10.1016/j.cbpa.2016.12.026

S.E.S. chose references, devised the outline, designed figures, created the table, and wrote the manuscript. This article including figures is reproduced in Chapter I with permission from the publisher.

(5) Glenn, W. S., **Stone, S. E.**, Ho, S. H., Sweredoski, M. J., Moradian, A., Hess, S., Bailey-Serres, J., Tirrell, D. A. "Bioorthogonal noncanonical amino acid tagging (BONCAT) enables time-resolved analysis of protein synthesis in native plant tissue." Plant Phys. **2017**, 173 (3), 1543-1553. DOI: 10.1104/pp.16.01762

S.E.S. performed data analysis, designed figures, and contributed to the writing of the manuscript. This article's figures are reproduced in part within Appendix E with permission from the publisher.

TABLE OF CONTENTS

LIST OF FIGURES

NOMENCLATURE

aaRS. Aminoacyl-tRNA synthetase
Aha. Azidohomoalanine
Anl. Azidonorleucine
APEX. Ascorbate peroxidase
Azf. 4-azido-L-phenylalanine
BONCAT. Bioorthogonal noncanonical amino acid tagging
CFU. Colony-forming units
Cm. Chloramphenicol
CTAP. Cell-type specific labeling with amino acid precursors
CuAAC. Copper catalyzed alkyne-azide cycloaddition
DAPI. 4,6-Diamidino-2-phenylindole
DBCO. Aza-dibenzocyclooctyne
Dde. N-(1-(4,4-dimethyl-2,6-dioxocyclohexylidene)ethyl)
EF. 4-ethynyl-L-phenylalanine
Erm. Erythromycin
FACS. Fluorescence activated cell sorting
FISH. Fluorescence in situ hybridization
Gent. Gentamicin
GF. Germ-free
GO. Gene ontology
H&E. Hematoxylin and eosin
HCR. Hybridization chain reaction
Hpg. Homopropargylglycine
IP. Intraperitoneal
LB. Luria-Bertani broth
LC. Liquid chromatography
LFQ. Label-free quantitation
LIMMA. Linear models for microarray data
MetRS. Methionyl-tRNA synthetase
MiPACT. Microbial identification after PACT
MRM. Multiple reaction monitoring

MRSA. Methicillin-resistant *Staphylococcus aureus*
MS. Mass spectrometry
ncAA. Noncanonical amino acid
OCT. Optimum cutting temperature
PACT. Passive CLARITY technique
PBS. Phosphate-buffered saline
PCA. Principal coordinate analysis
PCR. Polymerase chain reaction
PEG. Polyethylene glycol
PFA. Paraformaldehyde
Phe. Phenylalanine
PheRS. Phenylalanyl-tRNA synthetase
PMN. Polymorphonuclear leukocyte
PO. Per os (Oral Gavage)
Pra. Propargylglycine
Rcf. Relative centrifugal force
RIMS. Refractive index matching solution
S/C. Sub-cutaneous
SILAC. Stable isotope labeling using amino acids in culture
SORT. Stochastic orthogonal recoding of translation
SPAAC. Strain-promoted azide-alkyne cycloaddition
Spec. Spectinomycin
SPF. Specific-pathogen free
TAMRA. Tetramethylrhodamine
Tet. Tetracycline
TRAP. Translating ribosome affinity purification
TSB. Tryptic soy broth