Incorporation of non-canonical proline residues into proteins expressed in *Escherichia coli*

Thesis by

Stephanie L. Breunig

In Partial Fulfillment of the Requirements for the

Degree of

Doctor of Philosophy in Chemistry

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2023

(Defended May 5, 2023)

Stephanie L. Breunig
ORCID: 0000-0002-8665-6363

ACKNOWLEDGEMENTS

As with any important transition point in life, writing and defending a doctoral dissertation provides a beautiful opportunity to reflect upon those individuals who have shaped one's trajectory in life. I have been incredibly blessed in the relationships, both professional and personal, that have led me to Caltech and helped me through the trials and tribulations that is graduate school.

Of course, one of the most influential individuals at Caltech has been my advisor, David Tirrell. Incredibly, Dave managed to run an active research group while serving as Caltech's provost for the bulk of my time in his lab. Dave has undoubtedly shaped my scientific thinking beyond that, I think, of a typical advisor. His healthy dose of skepticism, true depth of thought, genuine kindness, and ability to ask *just* the right question are unparalleled. I have grown an incredible amount while a member of his lab, and appreciate the intellectual freedom and independence we were granted as his students.

I gratefully acknowledge the other members of my thesis committee: Dennis Dougherty, Steve Mayo, and Brian Stoltz. I cloned my first construct in Dennis's lab shortly after moving to Pasadena, and appreciate the physical organic chemistry perspective he brought to my education. A collaboration between Steve's group and ours was a fun project that, while not explicitly included in this thesis, refined my thinking around proline conformation and insulin dynamics. Brian taught some of the first classes I took at Caltech, and helped to provide a foundation in organic chemistry. His exceptionally friendly nature (and the occasional pickup volleyball appearance) helped me to first realize that yes, even Caltech professors are people, too.

The members of the Tirrell lab deserve my heartfelt gratitude. The diversity of research interests in the lab has significantly widened my knowledge of so many areas of chemical biology, proteomics, and materials science. Much of this thesis was founded upon work spearheaded by Seth Lieblich and Katharine Fang, and later carried on by Alex Chapman. Their efforts truly made this dissertation possible. To those who have played some role in training and mentoring me–Peter Rapp, Bradley Silverman, Alex Chapman, and Josh Baccile–I cannot thank you enough. You gave me the skills to become the experimentalist

that I am today and provided incredible support and encouragement along the way. The Tirrell lab has changed much over the years, from fifteen or so members when I first joined, to the intrepid five who remain. Elliot Mackrell, Sophie Miller, Hanwei Liu, and Grace Wang: it has been a pleasure to journey through grad school with you, and to see each of you become experts in your subfields. I cannot wait to see how you all go on to change the world.

The volleyball community at Caltech and in Pasadena gave me a competitive outlet, especially early on in my graduate career; it's always satisfying to hit a ball when experiments aren't exactly going your way. The Alpine and Triathlon Clubs at Caltech helped keep me sane in the later years of grad school, often by doing some pretty insane things. They made a road runner fall in love with the trails and all things multisport, and have quite literally changed my life. These folks are some of the most genuine, hard-working, and welcoming (not to mention *fast*) athletes you might ever meet, and I have immensely enjoyed training and racing with them.

I cannot say enough about the Catholic communities around Pasadena. They welcomed me with open arms as soon as I arrived in California (Joanie Noonan, Rachel & Nick Bradley, Mary Washburn), and helped in forming a true Catholic community at Caltech (Jorge & Hayley Llop, Elizabeth Lunny, Tim Csernica, Maria Camarca, Katt Redmond, Lauren Conger, Alessandra Flaherty, Nick Friesenhahn). These friends (and *so* many others) have walked with me in various capacities along my spiritual journey, and have shown me exactly why it's impossible to live the Christian life in isolation.

Some friendships go beyond description. Caitlin Lacker–you are a sister in all but blood. I'm not sure I could have made it through COVID, let alone grad school, without your constant love and support. Thank you for being you.

Finally, I would not be here (in all senses of the phrase) without my family. Mom, Dad, Nicole, Tim, and everyone in my extended family–it's impossible for me to fully appreciate all that you have done for me, let alone explain how grateful I am for it all. Knowing that you have my back, no matter what, has given me the freedom to chase my dreams. There is no greater gift than your love.

ABSTRACT

Non-canonical proline residues expand the chemical space about proline, while maintaining some conformational properties of the canonical residue. The translational machinery of *Escherichia coli* can accommodate close structural analogs of proline, which has enabled the production of recombinant proteins that contain non-canonical residues at proline positions. However, proline mutagenesis in *E. coli* is restricted to a relatively small set of proline variants, and protein science and engineering efforts utilizing non-canonical proline residues are limited.

This thesis aims to expand the scope of proline analogs that can be accepted by *E. coli*, and demonstrate the utility of proline mutagenesis in modifying and studying protein behavior. In Chapter II, we describe the incorporation of three aliphatic proline residues into recombinantly-produced insulin, and find that these modest modifications at ProB28 alter the biophysical properties of the therapeutic protein. In particular, the addition of an exocyclic olefin at B28 accelerated insulin fibril formation, while 4-methyl substituents increased the rate of dissociation from the pharmaceutically-formulated insulin hexamer. We expand our proline mutagenesis approach to monomeric insulins in Chapter III. 4-fluorinated proline analogs replaced ProB29 of the fast-acting insulin lispro; 4*S*-fluorination of ProB29 slowed fibril formation. Chapter IV describes the incorporation of the photo-activatable proline analog "photo-proline" into proteins expressed in *E. coli*, and Chapter V discusses our efforts to engineer the *E. coli* prolyl-tRNA synthetase to accommodate more diverse proline substrates. Together, this work expands the proline analogs accessible to recombinant expression in *E. coli*, and demonstrates their use in probing and engineering the biophysical properties of proteins.

PUBLISHED CONTENT AND CONTRIBUTIONS

Breunig, S. L. & Tirrell, D. A. “Incorporation of proline analogs into recombinant proteins expressed in *Escherichia coli*.” *Methods Enzymol.* **2021**, *656*, 545-571. doi: 10.1016/bs.mie.2021.05.008.

SLB collected and prepared all data, and participated in the writing of the manuscript. Excerpts from the introduction, Table 1, and Figures 1 & 2 reprinted with permission from the copyright holder, Elsevier, Inc.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

NOMENCLATURE

2-Me. 2-methylproline (or α-methylproline)

3*R*-OH. 3*R*-hydroxyproline

3*S*-OH. 3*S*-hydroxyproline

44-diF. 4,4-difluoroproline

44-diMe. 4,4-dimethylproline

4ene. 4-methyleneproline

4-keto. 4-oxoproline

4*R*-F. 4*R*-fluoroproline

4*R*-Me. 4*R*-methylproline

4*S*-F. 4*S*-fluoroproline

4*S*-Me. 4*S*-methylproline

4*S*-NH_2. 4*S*-aminoproline

4SSM. 4-site saturation mutagenesis

AA. Amino acid

aaRS. Aminoacyl-tRNA synthetase

Aha. Azidohomoalanine

Anl. Azidonorleucine

ANS. 8-anilino-1-naphthalenesulfonic acid

AUC. Analytical ultracentrifugation

Aze. Azetidine-2-carboxylic acid

β2m. β2 microglobulin

BME. β-mercaptoethanol

BONCAT. Bioorthogonal non-canonical amino acid tagging

CD. Circular dichroism

CFUs. Colony-forming units

CSII. Continuous subcutaneous insulin infusion

dhp. 3,4-dehydroproline

EDT. Ethanedithiol

EGFP. Enhanced green fluorescent protein

ePCR. Error-prone PCR

FACS. Fluorescence-activated cell sorting

FAI. Fast-acting insulin

FlAsH-EDT$_2$. fluorescein arsenical hairpin binder-ethanedithiol

GFP. Green fluorescent protein

HPLC. High performance liquid chromatography

IPTG. Isopropyl β-D-1-thiogalactopyranoside

KP. Insulin lispro

LB. Luria Bertani (medium)

MALDI-TOF MS. Matrix-assisted laser desorption/ionization-time of flight mass spectrometry

MetRS. Methionyl-tRNA synthetase

mRFP1. Monomeric red fluorescent protein 1

ncAA. Non-canonical amino acid

ncPro. Non-canonical proline

PDB. Protein databank

Pip. Piperidine-2-carboxylic acid

Pip-Az. Piperazine-2-carboxylic acid

Pip-OH. (2*S*, 5*S*)-5-hydroxypiperidine-2-carboxylic acid

PPIase: Peptidyl-prolyl isomerase

ProRS. Prolyl-tRNA synthetase

TC. Tetracysteine

TFA. Trifluoroacetic acid

Tfn. Trifluoronorleucine

ThT. Thioflavin T

Thz. 1,3-thiazoline-4-carboxylic acid

Trx. Thioredoxin