

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

The logo for the California Institute of Technology (Caltech), featuring the word "Caltech" in a bold, orange, sans-serif font.

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2023
(Defended May 5, 2023)

© 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; 4S-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

Acknowledgements	iii
Abstract	v
Published Content and Contributions.....	vi
Table of Contents.....	vii
List of Figures.....	xi
List of Tables	xiii
Nomenclature.....	xiv
Chapter I: Non-canonical proline residues in protein science and engineering. 1	
1.1 Contributions	1
1.2 Abstract.....	1
1.3 Main text.....	2
1.3.1 The conformational properties of proline	3
1.3.2 Non-canonical proline residues.....	3
1.3.3 Proline analogs in modifying protein behavior	7
1.3.4. Incorporation of non-canonical proline residues into polypeptide chains	10
1.4 Overview of thesis chapters	13
1.5 References	15
Chapter II: Incorporation of aliphatic proline residues into recombinantly- produced insulin.....	21
2.1 Contributions	21
2.2 Abstract.....	21
2.3 Introduction	22
2.4 Results and discussion.....	25
2.4.1 Aliphatic proline residues are accepted by the <i>E. coli</i> translational machinery	25
2.4.2 Proline analogs do not affect insulin secondary structure or bioactivity	26
2.4.3 4-methylene at position B28 accelerates fibril formation	28
2.4.4 4-methylation of ProB28 speeds hexamer dissociation	30
2.5 Conclusion.....	32
2.6 Materials and methods	33
2.6.1 Chemicals	33
2.6.2 Enzymes.....	34
2.6.3 Strains and plasmids	35
2.6.4 Primers	36
2.6.5 Nucleotide and amino acid sequences	36
2.6.6 Screening ncPro incorporation.....	38
2.6.7 Proinsulin expression	40
2.6.8 Proinsulin refolding.....	41

2.6.9 Insulin maturation and purification.....	42
2.6.10 MALDI-TOF MS	43
2.6.11 Circular dichroism spectroscopy.....	43
2.6.12 Reduction of blood glucose in diabetic mice.....	45
2.6.13 Fibrillation	45
2.6.14 Transmission electron microscopy	46
2.6.15 ANS fluorescence.....	46
2.6.16 Analytical ultracentrifugation	46
2.6.17 Calculations of proline and proline analog conformation.....	47
2.6.18 Models of ins-4R-Me and ins-4S-Me hexamers.....	47
2.7 Supplementary figures and tables	48
2.8 References	56
Chapter III: 4S-fluorination of ProB29 of insulin lispro slows fibril formation.....	61
3.1 Contributions	61
3.2 Abstract.....	61
3.3 Main text.....	62
3.4 Materials and methods	72
3.4.1 Chemicals	72
3.4.2 Enzymes.....	72
3.4.3 Strains and plasmids	72
3.4.4 Primers	73
3.4.5 Nucleotide and amino acid sequences	73
3.4.6 Proinsulin expression	76
3.4.7 Proinsulin refolding.....	77
3.4.8 Lispro maturation and purification	78
3.4.9 Preparation of lispro variants for mouse experiments.....	79
3.4.10 MALDI-TOF MS	79
3.4.11 Reduction of blood glucose in diabetic mice.....	80
3.4.12 Circular dichroism spectroscopy.....	81
3.4.13 Fibrillation	82
3.4.14 Transmission electron microscopy	83
3.4.15 ANS fluorescence.....	83
3.4.16 Analytical ultracentrifugation	83
3.5 Supplementary figures and tables	84
3.6. References	87
Chapter IV: Incorporation of photo-proline into proteins in <i>Escherichia coli</i>	91
4.1 Contributions	91
4.2 Abstract.....	91
4.3 Introduction	92
4.4 Results and discussion.....	94
4.4.1 Incorporation of photo-proline.....	94
4.4.2 <i>proP</i> overexpression leads to ncPro incorporation.....	96
4.4.3 PPIase substrates were not detected with photo-proline	98
4.5 Conclusion.....	100

4.6	Materials and methods	101
4.6.1	Chemicals	101
4.6.2	Enzymes.....	101
4.6.3	Strains and plasmids	101
4.6.4	Primers	101
4.6.5	Nucleotide and amino acid sequences	102
4.6.6	Cloning.....	104
4.6.7	Incorporation of photo-proline into proinsulin under osmotic stress conditions.....	104
4.6.8	Incorporation of photo-proline into proinsulin with <i>proP</i> overexpression	105
4.6.9	Incorporation of photo-pro with <i>proP</i> overexpression for PPIase crosslinking.....	106
4.6.10	Inclusion body isolation	107
4.6.11	LC-ESI-MS.....	107
4.6.12	UV irradiation.....	108
4.6.13	Western blot.....	109
4.7	Supplementary figures and tables	110
4.8	References	115
Chapter V: Toward expanding non-canonical proline incorporation by prolyl-tRNA synthetase engineering.....		118
5.1	Contributions	118
5.2	Abstract.....	118
5.3	Introduction	119
5.4	Results and discussion.....	121
5.4.1	Full-length fluorescent proteins are poor reporters of proline analog incorporation	121
5.4.2	Split-GFP as a reporter for proline analog incorporation.....	123
5.4.3	Efforts to improve split-GFP dynamic range reveals poor GFP1-10 solubility	126
5.4.4	Solubility tags improve GFP1-10 solubility	129
5.4.5	Changing expression protocols does not improve screen performance	133
5.4.6	Alternative GFP11 designs do not improve performance.....	135
5.4.7	Inclusion body analysis by flow cytometry	137
5.4.8	Survival on proline analog-containing medium	138
5.4.9	A small molecule alternative to fluorescent reporter proteins.....	138
5.4.10	Optimization of FIAsh labeling conditions does not improve sorting performance.....	141
5.4.11	Considerations for future ProRS engineering.....	147
5.5	Conclusion	150
5.6	Materials and methods	151
5.6.1	Chemicals	151
5.6.2	Enzymes.....	152

5.6.3	Strains	152
5.6.4	DNA oligos	153
5.6.5	g-Block gene fragments.....	156
5.6.6	Plasmids and cloning.....	158
5.6.7	Library generation	163
5.6.8	Nucleotide and amino acid sequences	165
5.6.9	Non-canonical proline incorporation and protein expression	170
5.6.10	Quantification of ncPro incorporation	171
5.6.11	Western blot analysis of protein solubility	172
5.6.12	FLAsH labeling (optimized conditions).....	173
5.6.13	Flow cytometry and FACS.....	173
5.6.14	Replica plating.....	174
5.6.15	Identification of proline-free carrier proteins	174
5.6.16	Microscopy	175
5.7	Supplementary figures and tables	176
5.8	References	186

LIST OF FIGURES

Figure 1.1. Proline conformational preferences.....	3
Figure 1.2. Non-canonical amino acid mutagenesis.....	11
Figure 2.1. Proline mutagenesis at position B28 of human insulin.....	23
Figure 2.2. Mass spectrometry characterization of insulin variants.....	26
Figure 2.3. Circular dichroism spectroscopy, sedimentation, and bioactivity of insulin variants	27
Figure 2.4. Fibrillation of insulin variants	29
Figure 2.5. Hexamer dissociation kinetics of insulin variants.....	31
Figure 2.6. Sedimentation velocity of insulin and ins-4 <i>S</i> -Me	32
Figure 2.S1. SDS-PAGE after non-canonical proline incorporation	48
Figure 2.S2. Purity of insulin samples assessed by SDS-PAGE.....	49
Figure 2.S3. Circular dichroism controls	50
Figure 2.S4. Models of ins-4 <i>R</i> -Me and ins-4 <i>S</i> -Me hexamers	51
Figure 2.S5. Conformations of proline, 4-methyleneproline, and 3,4-dehydroproline.....	52
Figure 3.1. Proline mutagenesis at position B29 of insulin lispro	63
Figure 3.2. Mass spectrometry characterization of lispro variants.....	67
Figure 3.3. Bioactivity, circular dichroism spectroscopy, and sedimentation of lispro variants.....	68
Figure 3.4. Hexamer dissociation kinetics of lispro variants	69
Figure 3.5. Fibrillation of lispro variants	70
Figure 3.S1. Purity of lispro samples assessed by SDS-PAGE	84
Figure 3.S2. Circular dichroism controls	85
Figure 4.1. Improved incorporation of photo-pro	95
Figure 4.2. <i>proP</i> overexpression leads to proline analog incorporation	97
Figure 4.3. PPIase:client interactions were not detected using photo-proline	99
Figure 4.S1. Thermal degradation products of photo-pro include dhp.....	110
Figure 5.1. mWasabi fluorescence does not correspond to proline analog incorporation efficiencies	122

Figure 5.2. Design of split-GFP to measure proline analog incorporation	124
Figure 5.3. 4S-NH ₂ incorporation does not improve after sorting	125
Figure 5.4. Leaky GFP11 expression and residual proline result in high background fluorescence	127
Figure 5.5. Poor solubility impairs split-GFP performance	127
Figure 5.6. Staggered expression of split-GFP at 25°C.....	130
Figure 5.7. Fluorescence of split supercharged and permuted GFP variants.....	131
Figure 5.8. GFP1-10 solubility tag fusions improve resolution by flow cytometry	132
Figure 5.9. GFP1-10/11 reverse expression.....	134
Figure 5.10. SPI expression of GFP1-10/11	134
Figure 5.11. XTEN-GFP11 fusion proteins to measure proline analog incorporation.....	136
Figure 5.12. Interrogating inclusion body formation by flow cytometry	137
Figure 5.13. A FLAsH-based labeling strategy to measure proline analog incorporation.....	139
Figure 5.14. Optimization of FLAsH labeling and protein expression conditions .	141
Figure 5.15. Sorting after FLAsH labeling reveals poor rescue efficiency and Top7 toxicity.....	142
Figure 5.16. Alternative proline-free target proteins	144
Figure 5.17. Validation of the DHR14 approach.....	145
Figure 5.18. Significant problems persist	147
Figure 5.S1. Proline-containing TC motifs do not accurately measure proline analog incorporation	176
Figure 5.S2. Development of a proline-free TC motif	177
Figure 5.S3. FLAsH labeling after Top7 expression leads to heterogeneous fluorescence	178
Figure 5.S4. FLAsH labeling after DHR14-TC3 expression can measure proline analog incorporation	179
Figure 5.S5. ProRS overexpression increases proline analog incorporation	180
Figure 5.S6. Cheaters overwhelm library samples after sorting	181
Figure 5.S7. Plasmid stability over rounds of transformation and growth	182
Figure 5.S8. Growth depends upon ProRS variant and promoter.....	183

LIST OF TABLES

Table 1.1. Proline analogs used for residue-specific incorporation into recombinant proteins	5
Table 2.S1. Incorporation of non-canonical proline residues.....	53
Table 2.S2. Mass spectrometry characterization of insulin variants.....	54
Table 2.S3. Expression conditions and insulin yields	54
Table 2.S4. Summary of insulin variant characterization	54
Table 2.S5. Sedimentation coefficients of insulin and ins-4S-Me	55
Table 3.S1. Mass spectrometry characterization of lispro variants.....	86
Table 3.S2. Lispro yields.....	86
Table 3.S3. Summary of lispro variant characterization	86
Table 4.1. Photo-pro incorporation efficiencies under osmotic stress conditions	96
Table 4.2. Proline analog incorporation efficiencies with <i>proP</i> overexpression	97
Table 4.S1. 3x-FLAG-tagged PPIases	113
Table 5.S1. Mutation rates for error-prone PCR library generation	184
Table 5.S2. Amino acid sequences of proline-free proteins.....	185

NOMENCLATURE

- 2-Me.** 2-methylproline (or α -methylproline)
3R-OH. 3R-hydroxyproline
3S-OH. 3S-hydroxyproline
44-diF. 4,4-difluoroproline
44-diMe. 4,4-dimethylproline
4ene. 4-methyleneproline
4-keto. 4-oxoproline
4R-F. 4R-fluoroproline
4R-Me. 4R-methylproline
4S-F. 4S-fluoroproline
4S-Me. 4S-methylproline
4S-NH₂. 4S-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
 β 2m. β 2 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-EDT₂. 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