

Engineering the tryptophan
synthase β -subunit for
synthesis of noncanonical
amino acids

Thesis by
Ella Jenná Watkins-Dulaney

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

The Caltech logo, featuring the word "Caltech" in a bold, orange, sans-serif font.

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2022
(Defended June 10, 2021)

© 2021

Ella Jenná Watkins-Dulaney

ORCID: 0000-0002-0585-1598

ACKNOWLEDGEMENTS

I would like to acknowledge the many outstanding people who have mentored and supported me throughout my scientific career. First and foremost, I thank Prof. Frances Arnold for pushing me beyond anything I thought I was ever capable of. You lead an extraordinary life, and I am grateful for the time that you shared that life with me. I will always remember dinners at the Rath Al Fresco, the retreat to the Merv Griffin Estate, the cruise to Mexico, and of course, the Nobel Prize. I would have never thought that the silly Wolfird shirt I designed for a cruise would end up in the Nobel Prize Museum. How strange life can be! It was truly an honor to be trained in your lab, and I am very proud to become an Arnold Lab alumna.

I have had the opportunity to work with fantastic people during my time in the Arnold Lab. Thank you, Dr. Sabine Brinkmann-Chen, for always being someone I was able to confide in. You were the only constant in my time at Caltech, and it was comforting to know that I always had you behind me. The postdocs in the group were integral to my success, and I am thankful for their willingness to lend precious time and advice. I am indebted to Dr. David Romney and Dr. Tina Boville for their mentorship during my time on ‘team TrpB.’ Special thanks to Tina, you are an incredible role model, mentor, and friend. I cannot express how grateful I am that you came into my life. Your charisma and confidence set an example that I needed to see in graduate school (and in life). The world needs more women in science like you. Thank you to Dr. Noah Dunham, Dr. Nicholas Porter, and Dr. David Miller for indulging me in endless discussions about the enolates project (and all other matters of science). I would also like to extend my sincere appreciation to my fellow graduate students Dr. Anders Knight, Nat Goldberg, and Patrick Almhjell. It was such a pleasure to learn and work beside all of you.

I extend my appreciation to Dr. Jimmy D. Gollihar, my mentor in the Ellington Lab at the University of Texas at Austin, who equipped me with many skills that helped me hit the ground running when I arrived at Caltech. He has been there to listen to me and commiserate about the challenges of graduate school during every step of my Ph.D.

Although graduate school was the most difficult thing that I have ever done, I grew to be the happiest and most alive that I have ever been. I owe this happiness in large part to the wonderful group of friends that I made here. Thank you, Tina and Brad Boville for your friendship and support. You two came into my life during the hardest part of my Ph.D. and infused it with a lightheartedness that reminds me not to take life too seriously. Art nights, home-cooked meals, and frequent laughs were just the remedy I needed!

Thank you to Patrick Almhjell for everything. You were the guide and catalyst into a life of outdoor adventures that I had always dreamed of but didn’t know where to start. If I learned anything from you, it would be that if you really want to do something, you just need to try harder. Oh, and to never use the jet color map. You, and the rest of the Pub Club, Kadina Johnston, Nicholas Sarai, and Austin Dulaney filled my memories with skiing and climbing trips that I will cherish forever.

I never thought of myself as someone who would play Dungeons and Dragons, but it became one of the things that brought me the most joy during the COVID-19 pandemic. The Company of 1000 Cloaks distracted me from the chaos of reality and gave me something to look forward to every week. Nicholas Porter, Anders Knight, Silken Jones, Bruce Wittmann, and Austin Dulaney, you all brought adventure into my life in a time when it was desperately needed.

I would like to thank my family for being a constant source of support. Allison and Todd Watkins (mom and dad), you are my role models and I am grateful for everything that you have done for me. It would be impossible to enumerate all of the ways that you set me up for success in life. Thank you for financially supporting me through college. I would not be where I am today without your hard work and generosity.

Finally, I would like to thank my husband, Dr. Austin Dulaney. I hadn't even heard of Caltech until we were applying for graduate school, and I definitely did not believe I could be admitted. Thank you for encouraging me to apply. Thank you for believing in me even when I did not believe in myself. You have been a constant support and I am so grateful for you always being there when I needed you. It was the experience of a lifetime and there is no one I would have rather experienced it all with than you. You bring so much joy and laughter into my life. We did it!

ABSTRACT

The tryptophan synthase β -subunit (TrpB) naturally catalyzes a pyridoxal phosphate cofactor-mediated β -substitution reaction between indole and serine to form L-tryptophan. Almost half a century ago, it was realized that TrpB could accept nucleophiles other than indole to synthesize noncanonical amino acids (ncAAs), which are highly useful small-molecule building blocks that are found in many bioactive molecules. Since then, TrpB has been applied to synthesize a wide range of ncAAs. This thesis details the engineering of TrpB for synthesis of new and useful ncAAs and the application of TrpB as a model to study the principles that govern intra-protein interactions. Chapter I chronicles the history of tryptophan synthase, provides useful information about the enzyme's catalytic cycle, and describes how TrpB has been used to synthesize ncAAs in works preceding this thesis. Chapter II describes the evolution, application, and characterization of TrpB for the synthesis of a blue, fluorescent noncanonical amino acid β -(1-azulenyl)-L-alanine (AzAla). Chapter III details the engineering and mechanistic characterization of TrpB to asymmetrically catalyze C–C bond formation with an entirely new class of nucleophile: ketones. Chapter IV describes the *in vivo* continuous evolution of TrpB which resulted in sequence-diverse TrpB orthologs that have been adapted to function at lower temperatures and display a range of substrate-selectivity profiles. Chapter V describes the development of a deep mutational scanning experiment of combinatorial site-saturation mutagenesis (SSM) libraries for generating a large dataset that maps enzyme sequence to function for the purpose of studying epistasis with machine learning. Overall, the work presented in this thesis expands the repertoire of ncAAs that can be synthesized by TrpB and demonstrates unique applications of TrpB as a model enzyme for continuous *in vivo* directed evolution and for generating a dataset that will be useful to the protein machine learning community.

PUBLISHED CONTENT AND CONTRIBUTIONS

† denotes equal contribution

1. **Watkins-Dulaney, E. J.**, Straathof, S., & Arnold, F. H. Tryptophan Synthase: Biocatalyst Extraordinaire. *ChemBioChem* **22**, 5–16 (2021). doi:10.1002/cbic.202000379

E.J.W.D. prepared the manuscript and all figures.

2. **Watkins, E. J.**,[†] Almhjell, P. J.[†] & Arnold, F. H. Direct Enzymatic Synthesis of a Deep-Blue Fluorescent Noncanonical Amino Acid from Azulene and Serine. *ChemBioChem* **21**, 80–83 (2020). doi:10.1002/cbic.201900497

E.J.W.D. and P.J.A. participated in the conception, design, and execution of the research. E.J.W.D. designed the screen for identifying improved enzyme variants. E.J.W.D. and P.J.A. contributed equally to AzAla purification. P.J.A. performed enzyme kinetics. E.J.W.D. prepared the first manuscript and P.J.A. edited and constructed figures.

3. **Watkins-Dulaney E.J.**, Dunham N.P., Straathof S., Turi S., Arnold F.H., Buller A.R. Asymmetric Alkylation of Ketones Catalyzed by Engineered TrpB. Manuscript in progress.

E.J.W.D. participated in the conception, design, execution of the research, and preparation of the manuscript. N.P.D. participated in the design and analysis of mechanistic experiments and preparation of the manuscript. S.S. participated in execution of round 4 of evolution and found 2-fluoroacetophenone activity. S.T. participated in initial condition optimization. F.H.A. and A.R.B. edited the manuscript.

4. Rix, G. **Watkins-Dulaney, E.J.**, Almhjell, P. J., Boville, C.E., Arnold, F.H., Liu, C.C. Scalable continuous evolution for the generation of diverse enzyme variants encompassing promiscuous activities. *Nat. Commun.* **11**, 1–11 (2020). doi: 10.1038/s41467-020-19539-6

All authors contributed to experimental design and data analysis. G.R. constructed the selection and performed all evolution experiments. E.J.W.D. performed the panel HPLC-MS assay and indole conversion rate measurements on *Tm*TrpBs from variant set 2, and P.J.A. analyzed the results. C.E.B. performed *in vitro* characterizations of *Tm*TrpBs from variant set 1 and performed the thermal shift assay and substrate scope characterizations. G.R. and C.C.L. wrote the manuscript with input and contributions from all authors.

PUBLISHED CONTENT NOT INCLUDED IN THESIS

5. Hammer, S. C. Kubik, G., **Watkins, E. J.**, Huang, S., Minges, H., & Arnold, F. H. Anti-Markovnikov Alkene Oxidation by Metal-Oxo–Mediated Enzyme Catalysis. *Science* **358**, 215–218 (2017). doi: 10.1126/science.aao1482

E.J.W.D. participated in the evolution of the aMOx variant. E.J.W.D. participated in writing and editing the manuscript.

TABLE OF CONTENTS

Acknowledgements	iii
Abstract	v
Published content and contributions	vi
Table of contents.....	viii
List of figures, tables, and schemes.....	xii
Abbreviations.....	xvi
Chapter I: Tryptophan synthase:Biocatalyst extraordinaire	1
Abstract.....	2
1.1 Introduction.....	3
1.2 Properties of tryptophan synthase (TrpS)	4
1.3 Synthesis of noncanonical Trp derivatives	7
1.3.1 Engineering stand-alone TrpB for indole-derived nucleophiles.....	10
1.3.2 β -Branched Trps.....	11
1.4 Engineering stand-alone TrpB for non-indole-derived nucleophiles	13
1.5 Biocatalytic cascades	17
1.5.1 D-amino acids	17
1.5.2 Tryptamine products	19
1.6 Summary and outlook.....	22
Bibliography for Chapter I.....	23
Chapter II: Direct enzymatic synthesis of a deep-blue fluorescent noncanonical amino acid from azulene and serine	34
Abstract.....	35
2.1 Introduction.....	36
2.2 Results	38
Bibliography for Chapter II.....	42
Appendix A: Supplementary information for Chapter II	45
A.1 Figures and tables	45
A.2 Experimental procedures	46
A.2.1 General experimental methods	46
A.2.2 Cloning, expression, and purification of TrpB variants.....	47
A.2.3 Construction of random mutagenesis libraries	48
A.2.4 Library expression and screening	48
A.2.5 Recombination of mutations	49
A.2.6 Small-scale analytical reactions.....	50
A.2.7 Calibration for measuring HPLC yield of AzAla.....	50
A.2.8 Large-scale preparation of AzAla.....	51
A.3 Variant sequences	52
A.4 Structural modeling	54

A.5 Characterization of AzAla	54
A.6 NMR spectra	56
Bibliography for Appendix A	57
Chapter III: Asymmetric alkylation of ketones catalyzed by engineered TrpB	58
Abstract.....	59
3.1 Introduction.....	60
3.2 Results and discussion	62
Bibliography for Chapter III	71
Appendix B: Supplementary information for Chapter III	74
B.1 Supplementary figures	74
B.2 Methods.....	78
B.2.1 General experimental methods	78
B.2.2 Cloning, expression, and purification of TrpB variants	80
B.2.3 Construction of site saturation mutagenesis libraries.....	80
B.2.4 Construction of random mutagenesis libraries	82
B.2.5 Construction of random mutagenesis libraries	83
B.2.6 Library expression and screening	85
B.2.7 Small-scale reactions with heat-treated lysate.....	86
B.2.8 Small-scale reactions with purified protein	86
B.2.9 Sodium cyanoborohydride reduction of 2- fluoroacetophenone product.....	88
B.2.10 Measurement of proton/deuteron exchange rates.....	88
B.3 Enzyme lineage.....	90
B.4 Synthesis, isolation, and characterization of products.....	98
B.4.1 Enzymatic synthesis, isolation, characterization, and standard curve of propiophenone product	98
B.4.2 Enzymatic synthesis, isolation, characterization, and standard curve of 2-cyanoacetophenone product	100
B.4.3 Enzymatic synthesis, isolation, characterization, and standard curve of 2-fluoroacetophenone product.....	102
B.4.4 Synthesis of 2-fluoroacetophenone-2,2- <i>d</i> ₂	104
Bibliography for Appendix B	112
Chapter IV: Scalable, continuous evolution for the generation of diverse enzyme variants encompassing promiscuous activities	113
Abstract.....	114
4.1 Introduction.....	115
4.2 Establishing a selection system for the evolution of <i>Tm</i> TrpB variants	117
4.3 Continuous evolution of <i>Tm</i> TrpB with depth and scale.....	118
4.4 Evolved <i>Tm</i> TrpB variants improve Trp production in vivo and contain cryptic genetic variation	121

4.5 Evolved <i>Tm</i> TrpBs exhibit high primary and promiscuous activity <i>in vitro</i>	123
4.6 A diverse panel of evolved <i>Tm</i> TrpB variants encompasses a variety of useful promiscuous activities with indole analogs.....	125
4.7 Mutations in evolved <i>Tm</i> TrpBs may modulate conformational dynamics and fine tune the active site.....	129
4.8 Discussion.....	130
Bibliography for Chapter IV.....	135
Appendix C: Supplementary information for Chapter IV	138
C.1 Supplementary tables.....	138
C.2 Supplementary figures.....	141
C.3 Methods.....	156
C.3.1 DNA plasmid construction.....	156
C.3.2 Yeast strains and media.....	157
C.3.3 Yeast transformation.....	157
C.3.4 Plating assays.....	158
C.3.5 <i>Tm</i> TrpB evolution.....	159
C.3.6 Growth rate assays.....	159
C.3.7 Enzyme characterization: General experimental methods.....	160
C.3.8 Expression and characterization of set 1 variants.....	161
C.3.8.1 Large scale expression and lysis.....	161
C.3.8.2 Lysate and whole cell small-scale reactions.....	161
C.3.8.3 Thermostability determination.....	162
C.3.8.4 Enzyme kinetics.....	162
C.3.9 Expression and characterization of set 2 variants.....	163
C.3.9.1 Small scale expression and lysis.....	163
C.3.9.2 Indole rate measurements.....	163
C.3.9.3 Substrate scope screen.....	164
C.3.10 Characterization of Tri-100-3-F and Tri-100-1-G.....	165
C.3.10.1 Large-scale expression and purification.....	165
C.3.10.2 Tri-100-3-F PLP-binding assay.....	166
C.3.10.3 Small scale analytical reactions.....	166
Bibliography for Appendix C.....	167
Chapter V: Deep mutational scan of the tryptophan synthase β -subunit.....	169
Abstract.....	170
5.1 Introduction.....	171
5.2 Original design overview.....	172
5.3 Establishing a yeast selection.....	174
5.4 Cloning combinatorial libraries.....	178
5.5 Preparing library amplicons for NGS.....	182
5.6 Proposed <i>E. coli</i> DMS redesign.....	187
5.7 Analysis of NGS data.....	187

Bibliography for Chapter V	189
Appendix D: Supplementary information for Chapter V	192
D.1 Materials.....	192
D.2 Construction of triple-site saturation mutagenesis libraries	192
D.3 NGS amplicon preparation.....	194

LIST OF FIGURES, TABLES, AND SCHEMES

<i>Table Number</i>		<i>Page</i>
A-1	Rate comparisons among native and engineered TrpB enzymes	46
A-2	Gibson assembly primers	50
A-3	TrpB sequencing primers	52
3-1	Ketone substrate scope of TmE8 and deuterium exchange rate constants	70
B-1	Chemicals, reagents, and equipment	79
B-2	Site saturated mutagenesis PCR mix	81
B-3	Site saturated mutagenesis PCR thermal cycler method.....	82
B-4	Gibson assembly primers to insert TrpB into pET22b(+) vector	82
B-5	Random mutagenesis PCR mix	83
B-6	Random mutagenesis PCR thermal cycler method.....	83
B-7	StEP Primary PCR mix	84
B-8	StEP Primary PCR thermal cycler protocol	84
B-9	StEP Secondary PCR mix.....	84
B-10	StEP Secondary PCR thermal cycler protocol	85
B-11	Purified protein reaction conditions.....	87
B-12	Progress curve fit parameters for Figure 3-3D.....	88
B-13	Rate of proton exchange	89
B-14	Summary of directed evolution for enolate activity.....	90
B-15	Sequencing primers	91
C-1	Summary of all cultures passaged for evolution of TmTrpB variants.....	138
C-2	Mutation summary statistics for OrthoRep-evolved TrpB populations	138
C-3	Mutations and identification information for all individual <i>TmTrpB</i> sequences	138
C-4	Kinetic parameters of selected <i>TmTrpB</i> variants at 30 °C.....	140
C-5	Nucleophiles tested in substrate scope screen.....	165
5-1	Sites selected for triple-site combinatorial libraries	178
D-1	Gap/3X SSM Library PCR primers	193
D-2	Gap/3X SSM PCR master mix	194
D-3	Gap/3X SSM PCR thermal cycler protocol	194
D-4	Internal NGS amplicon PCR primers	195
D-5	Internal NGS amplicon PCR master mix	196
D-6	Internal NGS amplicon PCR thermal cycler protocol.....	196
D-7	Nexterra adaptor ('external') primers	196
D-8	External NGS amplicon PCR	197

D-9	External NGS amplicon PCR thermal cycler protocol	197
D-10	Sequencing primer combinations	197

<i>Figure Number</i>		<i>Page</i>
1-1	Products of non-carbon and non-indole nucleophiles produced by TrpS and TrpB	13
2-1	Effect of E104(105)G mutation on Trp and AzAla production	40
2-2	Homology model of <i>Tm9D8*</i> with azulene in the active site.....	41
A-1	Isotryptophan formation catalyzed by TrpB E104(105)G mutants.....	45
A-2	Trace AzAla production by <i>Tm9D8*</i> E105G	45
A-3	Graphs of UV absorbance (A.U.) vs. wavelength for AzAla screen	49
A-4	AzAla calibration curve	51
A-5	HPLC traces of AzAla enantiopurity experiment	55
A-6	¹ H NMR of pure AzAla.....	56
A-7	¹³ C NMR of pure AzAla.....	56
A-8	¹ H NMR of crude AzAla.....	57
3-1	Overview of past and present work	62
3-2	Directed evolution of enolate activity.....	65
3-3	Mechanistic characterization of enolate activity	68
B-1	Location of mutations highlighted in homology model of <i>Tm9D8*</i>	74
B-2	Racemization of enzymatic propiophenone-derived cyclic imine product over time	75
B-3	Sodium cyanoborohydride reduction of 2-fluoroacetophenone product.....	75
B-4	<i>Tm9D8*</i> KIE experiment.	76
B-5	Substrate scope of <i>TmE8</i>	76
B-6	HPLC-MS traces for para-chloro-2-fluoroacetophenone	77
B-7	HPLC-MS traces for 2,2-difluoroacetophenone	77
B-8	HPLC-MS traces for benzoylacetone	78
B-9	Substrate scope reactions	78
B-10	Gibson assembly primers.....	81
B-11	Exchange curves of ketones in deuterated KPi buffer	89
B-12	2-Phenylacetophenone exchange curve in deuterated KPi buffer.....	89
B-13	Propiophenone product standard curve	99
B-14	2-cyanoacetophenone cyclized enamine product dissolved in methanol	101
B-15	2-cyanoacetophenone product standard curve	101
B-16	2-fluoroacetophenone product decomposes to 2-phenylpyrrole upon addition of acid to reaction mixture.....	103

B-17	2-fluoroacetophenone product standard curve	104
B-18	¹ H NMR spectrum of propiophenone product	105
B-19	¹ H NMR spectrum of 2-cyanoacetophenone product	106
B-20	¹³ C NMR spectrum of 2-cyanoacetophenone product	107
B-21	¹ H NMR spectrum of 2-fluoroacetophenone product	108
B-22	¹⁹ F NMR spectrum of 2-fluoroacetophenone product	109
B-23	¹ H NMR spectrum of 2-fluoroacetophenone-2,2- <i>d</i> ₂ product.....	110
B-24	¹³ C NMR spectrum of 2-fluoroacetophenone-2,2- <i>d</i> ₂ product.....	111
4-1	Pipeline for the use of OrthoRep continuous directed evolution.....	118
4-2	Selection trajectories for ten replicate cultures.....	120
4-3	<i>TmTrpB</i> homology model and table depicting consensus mutations of the ten cultures shown in Figure 4-1	121
4-4	In vivo activity and diversity of individual <i>TmTrpB</i> variants from OrthoRep-evolved populations	123
4-5	Promiscuous activities of a panel of evolved <i>TmTrpBs</i>	128
4-6	Conceptual similarities between natural enzyme ortholog evolution and OrthoRep evolution.....	134
C-1	Evaluation of indole-dependent TRP5 complementation of TrpB variants	141
C-2	In vivo Trp production by evolved TrpBs.....	142
C-3	In vitro Trp production by evolved TrpBs with heat treated lysate	143
C-4	In vitro Trp and Trp analog production with purified enzyme	144
C-5	Thermal shift assay on various <i>TmTrpBs</i>	145
C-6	Michaelis-Menten plots for rate of Trp production at saturating serine for evolved <i>TmTrpB</i> variants	146
C-7	Relatedness of TrpB panel sequences generated by OrthoRep evolution	147
C-8	TrpB panel indole activity by initial rate of Trp formation	148
C-9	TrpB panel activity with indole analogs by HPLC yield.....	152
C-10	Substrate activity profiles for large scale purification of variants B5 and G6.....	153
C-11	Commonly observed mutations at the α -subunit interaction interface	154
C-12	Commonly observed mutations to residues near a catalytic α -helix	154
C-13	First- and second-shell active site mutations.....	155
C-14	Sequence divergence for natural and OrthoRep-evolved TrpBs	156
5-1	Deep mutational scan overview	173
5-2	Genomic integration by homologous recombination.....	175
5-3	Pilot selection with TrpB variants integrated into yeast genome.....	176

5-4	Growth curve assay of BY4742 Δ Trp5 expressing <i>Tm9D8*</i> in minimal media.....	177
5-5	Site-saturation mutagenesis libraries highlighted in <i>Pyrococcus furiosus</i> scaffold (PDB: 6AMH).....	179
5-6	Combinatorial library primer design	181
5-7	Triple SSM library cloning scheme	181
5-8	Template recombination and yeast emulsion scheme	183
5-9	General overview of Illumina sequencing.....	184
5-10	Library diversity strategies and amplicon primer design.....	186
5-11	Pseudocode of the data analysis package	188
5-12	Example of poor and good quality forward and reverse read data.....	189
D-1	Snapshot of TrpB gene with annotations of seed regions of internal primers and the amplicons that would be generated.....	198

<i>Scheme Number</i>		<i>Page</i>
1-1	Native transformation catalyzed by TrpS.....	5
1-2	Catalytic cycle of TrpB.....	7
1-3	TrpS- and TrpB-catalyzed synthesis of Trp analogs.....	8
1-4	One-pot C5-arylation of β -methylTrp	12
1-5	Reaction scheme and product scope of engineered TrpB with nitroalkanes as nucleophiles for ncAA synthesis	14
1-6	Reaction scheme and product scope of engineered TrpB with oxindoles as nucleophiles for ncAA synthesis	16
1-7	TrpB-catalyzed synthesis of β -(1-azulenyl)-l-alanine (AzAla)	17
1-8	Biocatalytic cascade for synthesis of D-Trp derivatives	18
1-9	One-pot chemoenzymatic approach for the stereoinversion of β -methylTrp to form D- β -methylTrp.	19
1-10	Enzymatic synthesis of psilocybin.....	20
1-11	One-pot, two-step synthesis of tryptamine derivatives with TrpB and <i>Ruminococcus gnavus</i> Trp decarboxylase (RgnTDC).....	22
2-1	Synthesis of AzAla and Native TrpB activity	37
2-2	Parallels between indole and azulene in the TrpB reaction	38
B-1	TrpB catalyzed propiophenone+serine reaction.....	98
B-2	TrpB catalyzed 2-cyanoacetophenone+serine reaction.....	100
B-3	TrpB catalyzed 2-fluoroacetophenone+serine reaction	102

ABBREVIATIONS

Å	Ångstrom
ACN	acetonitrile
AcOH	acetic acid
AzAla	β -(1-azulenyl)-L-alanine
Ala	alanine
BSA	bovine serum albumin
BCA	bicinchoninic acid assay
carb	carbenicillin
COMM domain	communication domain
DAAT	D-alanine aminotransferase
d.r.	diastereomeric ratio
E(A-A)	enzyme amino acrylate
E(Ain)	enzyme internal aldimine
E(Aex ₁)	enzyme external aldimine intermediate 1
E(Aex ₂)	enzyme external aldimine intermediate 2
E(Q ₁)	enzyme quinonoid intermediate 1
E(Q ₂)	enzyme quinonoid intermediate 2
ee	enantiomeric excess
<i>Ec</i>	<i>Escherichia coli</i>
epPCR	error-prone PCR
GC	gas chromatography
GC-MS	gas chromatography with mass spectrometry
GOI	gene of interest
H	hour(s)
HCl	hydrochloric acid
HPLC	high performance liquid chromatography
HPLC-MS	high performance liquid chromatography with mass spectrometry
IGP	indole-3-glycerol phosphate
IPGT	isopropyl β -D-1-thiogalactopyranoside
<i>k</i>	rate constant
<i>K_m</i>	Michaelis constant
Kpi	potassium phosphate
LAAD	L-amino acid deaminase
LAAO	L-amino acid oxidase
LB	Lysogeny Broth (Luria-Bertani medium)
LC-MS	liquid chromatography with mass spectrometry
MeOH	methanol
min	minute(s)
ncAA	noncanonical amino acid
OD ₆₀₀	optical density at 600 nm

PCR	polymerase chain reaction
PDB	Protein Data Bank
PP	propiophenone
<i>Pf</i>	<i>Pyrococcus furiosus</i>
<i>Pf</i> TrpB	TrpB from <i>Pyrococcus furiosus</i>
Ph	phenyl
pK _a	acid dissociation constant
PLP	pyridoxal 5'-phosphate
<i>Rgn</i>	<i>Ruminococcus gnavus</i>
<i>Rgn</i> TDC	Trp decarboxylase from <i>Ruminococcus gnavus</i>
RT	room temperature or retention time
Ser	serine
StEP PCR	staggered extension process PCR
SOC	Super Optimal broth with Catabolite repression
SSM	site saturation mutagenesis
TB	Terrific Broth
Thr	threonine
TM	melting temperature
<i>Tm</i>	<i>Thermotoga maritima</i>
<i>Tm</i> TrpB	TrpB from <i>Thermotoga maritima</i>
Trp	tryptophan
TrpA	Tryptophan synthase α -subunit
TrpB	Tryptophan synthase β -subunit
TrpS	Tryptophan Synthase $\alpha\beta\beta\alpha$ dimeric complex
TTN	total turnover number
UV-vis	ultraviolet visible
Val	valine
V _{max}	maximum rate
V ₀	initial velocity
WT	wild type
2F	2-fluoroacetophenone
[S]	concentration of substrate, S

Chapter 1

TRYPTOPHAN SYNTHASE: BIOCATALYST EXTRAORDINAIRE

Material from this chapter appears in “**Watkins-Dulaney, E. J.**, Straathof, S., Arnold, F. Tryptophan Synthase: Biocatalyst Extraordinaire. *ChemBiochem*. 2021 Jan; 22(1):5-16. DOI: 10.1002/cbic.202000379.”

ABSTRACT

Tryptophan synthase (TrpS) has emerged as a paragon of noncanonical amino acid (ncAA) synthesis and is an ideal biocatalyst for synthetic and biological applications. TrpS catalyzes an irreversible, C–C bond forming reaction between indole and serine (Ser) to make L-tryptophan (Trp); native TrpS complexes possess fairly broad specificity for indole analogs, but are difficult to engineer to extend substrate scope or to confer other useful properties due to allosteric constraints and their heterodimeric structure. Directed evolution freed the catalytically relevant TrpS β -subunit (TrpB) from allosteric regulation by its TrpA partner and has enabled dramatic expansion of the enzyme's substrate scope. This chapter examines the long and storied career of TrpS from the perspective of its application in ncAA synthesis and biocatalytic cascades.

1.1 Introduction

Noncanonical amino acids (ncAAs) enable researchers to interact with and modify life at the molecular level, and are a vital tool for many modern biological studies. Defined as amino acids that are not genetically encoded, ncAAs bear chemical motifs not found in the 20 canonical amino acids and can alter the characteristics of molecules that incorporate them. Though often referred to as unnatural amino acids, many ncAAs do occur naturally as post-translationally modified peptide residues or as intermediates in biosynthesis of secondary metabolites.¹ Nature thus demonstrates that ncAAs can serve as handles to manipulate biochemical properties. Furthermore, substituting canonical amino acids with ncAAs imbues molecules with different functionalities while minimally perturbing structure.² As such, ncAAs are seeing growing applications in research, where they are useful as biophysical probes,^{2,3} are introduced into polypeptides to create improved or entirely new functions,^{4,5} and are incorporated into bioactive small molecules and peptide therapeutics.⁶⁻⁸

A barrier to realizing the potential of ncAAs is that they are challenging to synthesize owing to laborious protection and deprotection sequences necessary to prevent epimerization of the chiral center or undesired reactivity with the amine and carboxylate groups.⁹ Simpler, more effective, and more direct routes to ncAAs are necessary to better harness their potential applications. A promising approach for ncAA synthesis is to use enzymes, which can perform transformations with exquisite precision in the presence of multiple reactive centers without the need for protecting groups. With carefully tuned active sites, enzymes can overcome regio- and stereoselectivity challenges by directing substrates and reactive intermediates during a catalytic cycle.

Another major advantage of enzymes is that they can be combined in one-pot biocatalytic cascades to access value-added products from simple and inexpensive starting materials.¹⁰ However, enzymes from biosynthetic pathways to naturally occurring ncAAs may not be practical to engineer or scale up if, for example, they catalyze reversible reactions, express poorly in recombinant hosts, are allosterically regulated, or have limited substrate scopes.¹¹ Nevertheless, directed evolution has empowered biocatalysis to be well poised to contribute

to ncAA synthesis, and examples of new, engineered ncAA synthases and enzyme cascades are emerging.¹¹

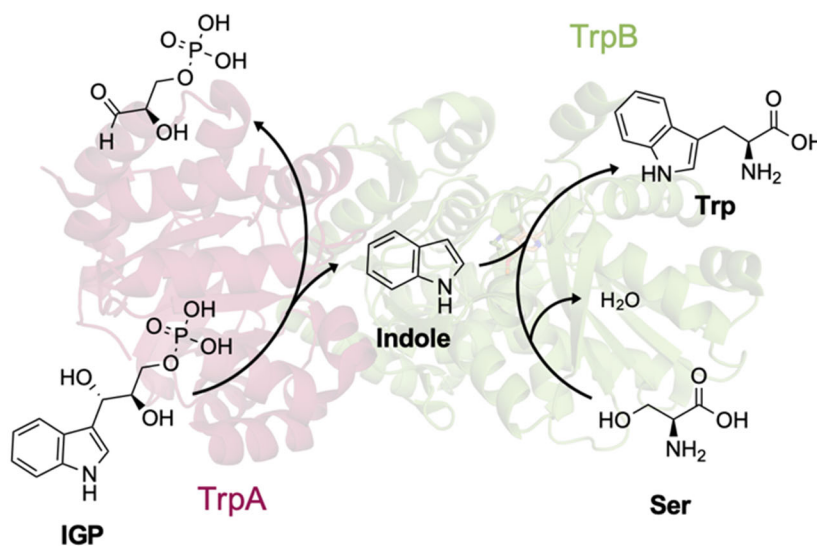
Tryptophan synthase (TrpS) is a premier example of an enzyme that can be used for scalable ncAA synthesis. TrpS possesses covetable qualities for an ncAA synthase: it forms a C–C bond between readily available starting materials to make L-tryptophan (Trp) and closely related derivatives in a single enzymatic step.¹² TrpS exists as a heterodimeric complex comprised of two α - and β -subunits (TrpA and TrpB, respectively) that work together to transform indole glycerol phosphate (IGP) and L-serine (Ser) into Trp (**Scheme 1-1**). The TrpA subunit is responsible for the cleavage of IGP into indole and glyceraldehyde and does not directly participate in the C–C bond forming step. In fact, TrpA can be bypassed entirely by providing indole analogs to the enzyme complex, where they are transformed by the TrpB subunit into Trp analogs. Although TrpS can be used to synthesize a variety of Trp-based ncAAs, directed evolution of the catalytically relevant TrpB subunit to create a stand-alone enzyme dramatically simplified engineering efforts and allowed for a systematic expansion of accessible ncAA products.¹³

This chapter provides an overview of how TrpS and its laboratory-evolved TrpB progeny have been used to produce ncAAs. We also give examples of how this enzyme has been incorporated in biocatalytic cascades to access D-amino acids and tryptamine products.

1.2 Properties of tryptophan synthase (TrpS)

Tryptophan synthase is a pyridoxal 5'-phosphate (PLP)-dependent enzyme that has captured the interest of enzymologists and bioengineers for over half a century. The study of TrpS dates back to when the burgeoning field of molecular biology had barely taken its first steps. Discovered in the 1940s, TrpS has served as a model enzyme for a wide range of investigations, from proving gene-protein collinearity¹⁴ to studying the evolution and nature of allostery,¹⁵ conceptualizing and understanding vectorial catalysis and substrate channeling,¹⁶ and, most relevant to this review, synthesis of ncAAs.

TrpS is found in all domains of life as an $\alpha\beta\beta\alpha$ heterodimeric complex that catalyzes the formation of Trp from IGP and Ser (**Scheme 1-1**). The α -subunit (TrpA) and β -subunit (TrpB) experience mutual allosteric activation, the evolutionary history and nature of which are still an active area of research.^{17,18} The two subunits interact with one another through rigid-body motion of the TrpB communication (COMM) domain and a monovalent cation (MVC) binding site within TrpB. When IGP binds TrpA, it initiates a conformational change activating TrpB to promote formation of the (PLP)-bound amino-acrylate derived from Ser. The TrpB subunit then reciprocally stimulates TrpA to induce retro-aldol cleavage of IGP, releasing indole.¹⁹ Once released, indole diffuses along a 25-Å long tunnel to the β -subunit where it can immediately participate in a PLP-mediated β -addition reaction, releasing water and Trp.



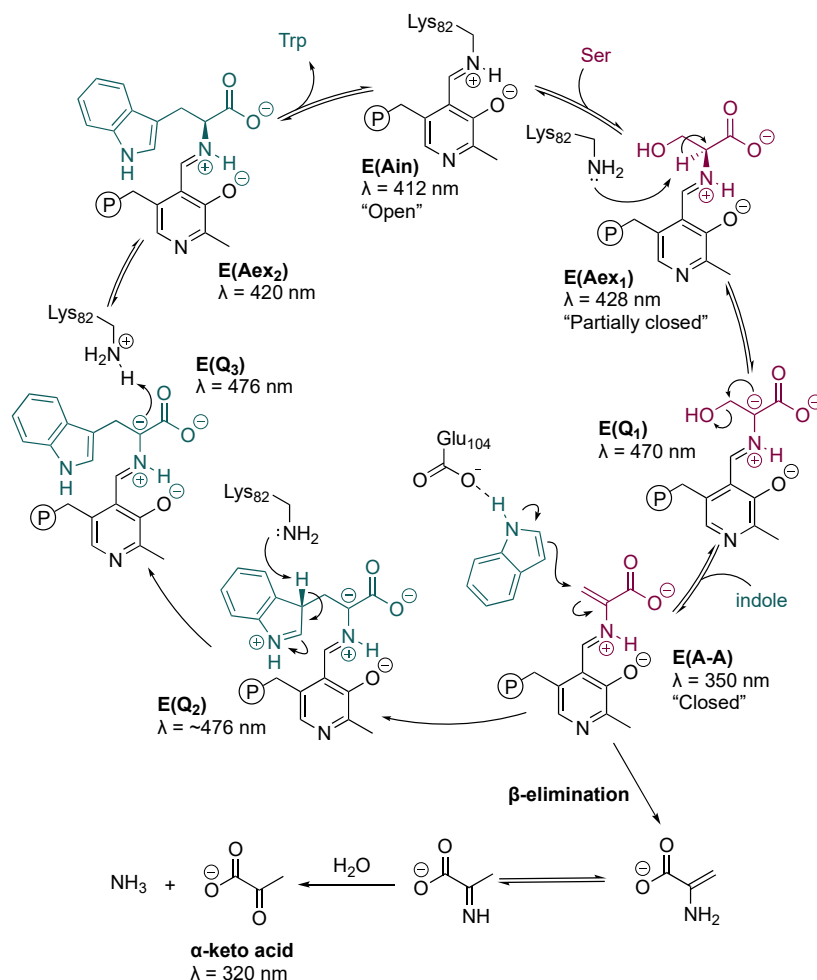
Scheme 1-1. Native transformation catalyzed by TrpS. TrpA (left, pink) performs a retro-aldol cleavage on indole glycerol phosphate (IGP) releasing indole and glyceraldehyde phosphate. Indole diffuses to the TrpB subunit (right, green) which catalyzes a PLP-mediated β -substitution reaction between indole and l-serine (Ser), releasing water and l-tryptophan (Trp). (PDB:5E0K)

The TrpB PLP cofactor absorbs in the UV-vis region, and each reactive intermediate possesses a characteristic spectral trace, allowing observation of the catalytic cycle via UV-vis spectroscopy. In the TrpB resting state, PLP is covalently bound to the ϵ -nitrogen of a lysine residue (K82, *Pyrococcus furiosus* TrpB, *PfTrpB*, numbering) through a protonated

Schiff-base linkage referred to as the internal aldimine, E(Ain) (λ_{\max} = 412 nm).²⁰ In the first stage of the catalytic cycle (**Scheme 1-2**), Ser enters the active site and replaces the lysine via transimination to form an external aldimine intermediate, E(Aex1) (λ_{\max} = 428 nm). This step is concomitant with a rigid-body conformational change in the TrpB COMM domain, with the enzyme adopting a ‘partially closed’ state.

PLP-dependent enzyme specificity is largely dependent on alignment of the bond to be broken with the π molecular orbital system of PLP.²¹⁻²³ TrpB promotes C α deprotonation by using a hydrogen bonding network formed with the Ser carboxylate that locks the C–H bond periplanar to the PLP π system.²³ The free K82 residue deprotonates the C α of Ser, ablating the chiral center and forming a carbanion that is delocalized by the PLP cofactor to form a quinonoid intermediate, E(Q1) (λ_{\max} = 470 nm).²⁴ Subsequent elimination of the hydroxyl group forms the electrophilic amino-acrylate species, E(A-A) (λ_{\max} = 350 nm), which is poised for attack by the indole nucleophile. During this step, the COMM domain assumes a ‘fully closed’ conformation that is stabilized by TrpA.^{17,25} If no indole is present, a kinetically competing transimination reaction with the active site lysine can occur, releasing dehydroalanine that hydrolyzes to form ammonia and pyruvate (β -elimination pathway).²⁶ If indole is present, it arrives in the active site of TrpB and is positioned by the catalytic glutamate (E104, PfTrpB numbering) for nucleophilic attack. The catalytic glutamate is important for controlling the regioselectivity of the reaction; mutagenesis reveals its crucial role to effect C–C bond formation at C3 over a C–N bond at N1.²⁷

The beginning of the second stage of the TrpB catalytic cycle is marked by irreversible nucleophilic attack by indole on the E(A-A) to form a second quinonoid intermediate, E(Q2) (λ_{\max} = ~476 nm). The (S)-indolenene species is quickly deprotonated, restoring aromaticity, to reach a third and final quinonoid intermediate, E(Q3) (λ_{\max} = 476 nm).²⁸ C α is then re-protonated by K82 stereospecifically to form the Trp-bound external aldimine, E(Aex2) (λ_{\max} = 420 nm), re-establishing the chiral center and completing Trp formation.^{29,30} Trp release from the enzyme via transimination by K82 returns PLP to the E(Ain) resting state and completes the catalytic cycle.³⁰



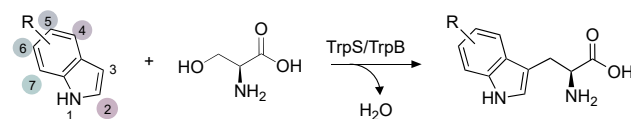
Scheme 1-2. Catalytic cycle of TrpB.

1.3 Synthesis of noncanonical Trp derivatives

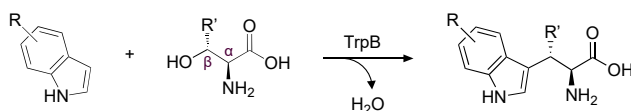
Tryptophan derivatives (**Scheme 1-3**) are a subclass of ncAAs that have been used extensively as probes for chemical biology. Trp itself is a major source of UV absorption and fluorescence in proteins, and its spectral properties, which are highly influenced by the surrounding environment, have been leveraged to study protein dynamics, folding, and ligand binding.³¹ Substitutions on the indole moiety, such as in 4-cyanoTrp and 5-hydroxyTrp, as well as Trp isosteres like azaTrps (**1**, **2**, **3**, **4**) can enhance or alter these spectroscopic properties to exhibit higher quantum yields or shift excitation/emission spectra.^{32–34} Decorations and substitutions on the indole side chain bestow many other useful biochemical properties: fluorinated Trps are used in ¹⁹F NMR studies,³⁵ selenophene and

thienyl functional groups are used for phasing crystallographic structures,³⁶ and halides can allow for site-specific modification through palladium-catalyzed coupling reactions with alkenes and alkynes.³⁷ Like many other ncAAs, Trp derivatives are biosynthetic precursors to compounds that exhibit diverse pharmacological activities, including anticancer, antibiotic, immunosuppressant, and phytotoxic properties.^{38,39}

(a) TrpS/TrpB synthesis of tryptophan analogs



(b) TrpB synthesis of β -branched tryptophan analogs

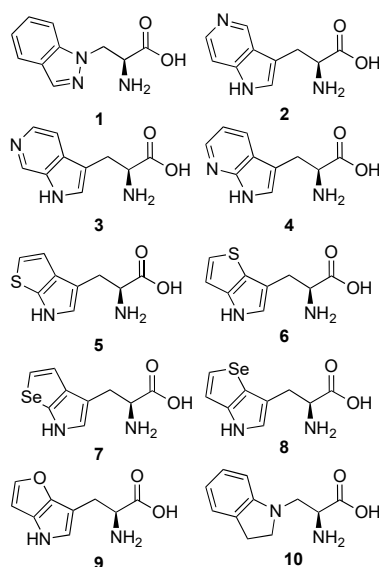


(c) TrpS/B Trp substrate profile

R	2	4	5	6	7
CH3	●	●	●	●	●
OCH3	○	●	●	●	○
OH	○	○	●	●	○
NH2	○	○	○	○	○
N3	○	○	○	○	○
F	●	●	○	○	○
Cl	●	●	○	○	○
Br	●	●	○	○	○
I	○	○	○	○	○
NO2	●	●	○	○	○
CHO	○	○	○	○	○
CN	○	○	○	○	○
CONH2	○	○	○	○	○
B(OH)2	○	○	○	○	○
CF3	○	○	○	○	○
CCH	○	○	○	○	○

Legend: ○ TrpS, ● Both, ● TrpB

(d) Other tryptophan isosteres



Scheme 1-3. TrpS- and TrpB-catalyzed synthesis of Trp analogs. (a) TrpB-catalyzed synthesis of Trp analogs (b) TrpB-catalyzed synthesis of β -branched Trp analogs. (c) TrpS/TrpB substrate profile: modified indoles that any TrpS or TrpB variants have been demonstrated to accept. Position represents carbon where substitution, R, occurs on indole moiety. The profile is not to be interpreted as a TrpS/TrpB selectivity profile, and gaps in activity may be due to lack of testing (not all indole derivatives are readily available). (d) Other tryptophan isosteres: TrpS and TrpB catalyze the synthesis of a number of Trp analogs bearing heteroatom substitutions.

Shortly after the discovery of TrpS, researchers began using substituted methylindoles (2-, 4-, 5-, 6-, 7-CH₃-indole) to gain insights into the enzyme's mechanism and pathway regulation.^{29,40,41} The synthetic utility of TrpS, however, was first realized in 1974 when Wilcox synthesized a series of Trp derivatives (5-F-, 6-F-, 5-OH-, 5-MeO-, 6-MeO-, 2-CH₃-, 5-CH₃-, 7-CH₃-Trp; **1**, **4**) from Ser and indole analogs using TrpS from *Escherichia coli* (*Ec*TrpS).⁴² During the following decades, various wild-type TrpS homologs were shown to have activity on a number of other decorated indoles and indole isosteres: Saito et al. demonstrated the synthesis of azido-substituted Trps (4-, 5-, 6-, 7-N₃-Trp) using a TrpS from *Neurospora crassa*;⁴³ the Phillips group applied TrpS from *Salmonella typhimurium* for the synthesis of chloroTrps (4-, 5-, 6-, 7-Cl-Trp),⁴⁴ sulfur, selenium, and oxygen-containing Trp isosteres (**5**, **6**, **7**, **8**, **9**),⁴⁵⁻⁴⁷ as well as azaTrps (**1**, **2**, **3**, **4**),⁴⁸ Goss and colleagues prepared an exceptionally diverse set of substituted Trp analogs including methyl- (2-, 4-, 5-, 6-, 7-CH₃-Trp),⁴⁹ amino- (4-, 6-, 7-NH₂-Trp),⁵⁰ halo- (4-, 5-, 6-, 7-F; 4-, 5-, 6-, 7-Cl; 5-, 6-, 7-Br; 7-I-Trp),⁴⁹⁻⁵² and nitroTrp (7-NO₂-Trp)⁵² using TrpS from *E. coli* and *Salmonella enterica*. TrpS has also been found to catalyze a C–N bond forming reaction with indoline to form dihydroisotryptophan (**10**).^{27,53,54}

Despite this ability to produce desirable compounds, TrpS still has limitations that restrict practical and widespread use. Even under optimized conditions, ncAA yields with TrpS catalysts are typically under 50%.^{49,52} Insolubility of indole compounds in water also limits the substrate concentrations. Although this was partially remedied by addition of co-solvents, the TrpS homologs used lacked solvent tolerance, limiting the effectiveness of this solution.⁴⁴ Furthermore, extensive engineering of the TrpB subunit for improved activity or expanded substrate scope was impeded by the need for the TrpA subunit, which does not directly participate in the coupling of indole and Ser, but nonetheless increases metabolic load on host cells. Unfortunately, without co-expression and allosteric activation from their corresponding TrpAs, native TrpBs lose most of their activity, rendering them all but useless.^{30,55,56}

1.3.1 Engineering stand-alone TrpB for indole-derived nucleophiles

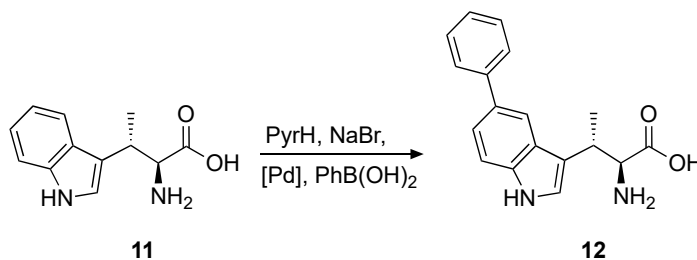
Enzyme homologs are valuable assets that often display divergent activities with non-natural substrates. Javier Murciano-Calles and co-workers investigated homologs of *PfTrpB* for activity on 5-substituted indoles.⁵⁷ They recombined activating mutations discovered by Buller et al. into the TrpB derived from the hyperthermophilic bacterium *Thermotoga maritima* (*TmTrpB*, 64% sequence identity to *PfTrpB*) and found a variant with broadly improved activity toward 5-substituted indoles (5-CH₃-, 5-OCH₃-, 5-Cl-, 5-Br-, 5-NO₂-, 5-CHO-, 5-CN-, 5-B(OH)₂-Trp) compared to previous catalysts.

With a set of stand-alone TrpB enzymes that were straightforward to express and engineer in hand, David Romney and his team aimed to broaden the platform's substrate scope to include challenging indoles on which TrpS had previously shown poor activity.^{56,57} The resultant panel of evolved TrpB enzymes accepted indoles bearing different substitution patterns and functional moieties such as halogen (4-F-; 6-, 7-Cl-; 4-, 6-, 7-Br-; 5-, 7-I-; 5,6-Cl₂-, 5-Br-7-F-, 5-Cl-7-I-Trp), nitro (4-, 5-, 6-, 7-NO₂-Trp), cyano (4-, 5-, 6-, 7-CN-Trp), carboxamide (5-CONH₂-Trp), boronate (5-, 6-B(OH)₂-Trp), and trifluoromethyl groups (5-CF₃-Trp) with most isolated yields ranging from 70–99%.¹³ Christina Boville and colleagues continued engineering one of Romney's variants, *TmTrpB*^{2F3}, to improve activity for 4-cyanoTrp, a useful blue fluorescent ncAA with high quantum yield and lifetime.⁵⁸ This transformation demonstrated a marked improvement over the best synthetic route, which was a palladium-catalyzed cyanation reaction that achieved a maximal yield of 10%.³² Cells from one liter of *E. coli* shake-flask culture expressing the final variant, *TmTrpB*^{9D8*}, could synthesize 4-cyanoTrp on a larger scale (800 mg, 49% yield). Notably, laboratory-evolved *TmTrpB*^{9D8*} was discovered to function better at lower temperatures (such as 37 °C), providing for future possible *in vivo* applications.⁵⁸

1.3.2 β -Branched Trps

Beta-branched amino acids are found in many useful bioactive natural products and pharmaceuticals; however the presence of two adjacent chiral centers makes them particularly challenging to synthesize.^{59,60} Buller's team engineered their stand-alone TrpB to accept L-threonine (Thr) as the electrophile to produce β -methylTrp (**Scheme 1-3b**).⁶¹ Unlike in earlier works, where modified electrophiles ultimately produced the same amino-acrylate as Ser,^{62,63} the use of Thr generated an entirely new β -substituted amino-acrylate-like species (amino-crotonate) that diastereoselectively formed a second chiral center upon C–C bond formation. This is remarkable: Thr is a universal and abundant metabolite that TrpS naturally discriminates against. Buller discovered that native TrpS actually binds Thr efficiently, but binding results in decreased affinity for indole and disrupts the allosteric signaling that regulates the catalytic cycle. These effects translate to a >82,000 fold-preference for Ser over Thr in the native enzyme complex when both substrates are present.⁶⁴ However, in the absence of Ser competition, *Pf*TrpS—and more importantly *Pf*TrpB^{WT}—showed trace activity with indole and Thr, providing the foothold necessary to apply directed evolution. Starting from an evolutionary intermediate from their previous campaign that had better activity with Thr than the wild-type enzyme, *Pf*TrpB^{4D11}, two rounds of evolution accumulating three new mutations resulted in *Pf*TrpB^{2B9}, which exhibited a >6,000-fold boost in activity for β -methylTrp formation relative to wild-type *Pf*TrpB.⁶¹

Shortly after this work appeared, the Micklefield group published an engineered *Sf*TrpS bearing one mutation, L166V, that could also catalyze the formation of β -branched Trps (β -methyl- ; β -methyl-2-, 4-, 6-, 7-CH₃-; β -methyl-4-, 7-F-; β -methyl-7-Cl-; β -methyl-7-OMe-Trp) from Thr and indole analogs.⁶⁵ Similar to *Pf*TrpS, the enzyme struggled with 5-substituted indoles. Instead of applying further evolution to increase the substrate scope, the authors took a different approach, using *Sf*TrpS^{L166V} to synthesize β -methylTrp (**11**), which they then derivatized chemoenzymatically. The flavin-dependent Trp-5-halogenase PyrH was used with MgCl₂ or NaBr to create halogenated 5-substituted Trps (5-Br-, 5-Cl-Trp) which could undergo a palladium-catalyzed cross coupling reaction with phenylboronic acid in the same pot to create 5-phenyl- β -methylTrp (**Scheme 1-4, 12**).



Scheme 1-4. One-pot C5-arylation of β -methylTrp (11). Flavin-dependent Trp-halogenase PyrH brominates C5, which then participates in a palladium-catalyzed Suzuki cross-coupling to install the aryl group (12).⁶⁵

Christina Boville and her team engineered stand-alone *Pf*TrpB^{2B9} to accept β -branched Ser analogs with longer alkyl chains such as β -ethyl- and β -propylSer.⁶⁶ Bulkier alkyl chains at the β -position were thought to hinder nucleophilic attack, allowing the competing β -elimination to unproductively consume the electrophile. Initial activity with *Pf*TrpB^{2B9} on β -ethylSer and indole was too low for high-throughput screening, so the authors mutated an active-site residue presumed to clash sterically with the alkyl β -substitution. Investigating the very same residue mutated by Micklefield et al.,⁶⁵ Boville and colleagues discovered that both valine (Val) and alanine (Ala) improved activity. While Ala was slightly less beneficial than Val, the authors rationalized that Ala may provide more room in the active site for electrophiles with longer β -alkyl substituents. This boosted activity enough to enable Boville's team to use a high-throughput UV-based screen with libraries generated by random mutagenesis for three more rounds of evolution. The final variant, *Pf*TrpB^{7E6}, was assayed against combinations of electrophiles (Thr, β -ethyl-, β -propylSer) and nucleophiles (indole; 2-, 4-, 5-, 6-, 7-CH₃-; 4-, 5-F-; 5-Cl-; 7-aza-indole) to determine its substrate scope and generality. Although *Pf*TrpB^{7E6} was only evolved on β -ethylSer, the mutations substantially improved activity for Thr and β -propylSer as well. Notably, *Pf*TrpB^{7E6} required only one equivalent of Thr to achieve a 3.5-fold greater yield for β -methylTrp than the parent *Pf*TrpB^{2B9} did with ten equivalents of Thr. X-ray crystallography, measurement of the deamination rates, and UV-spectrophotometric evidence supported the hypothesis that evolution stabilized the closed conformation of the enzyme and generated a more persistent amino-acrylate that was less prone to unproductive β -elimination.⁶⁶

1.4 Engineering stand-alone TrpB for non-indole-derived nucleophiles

TrpB proved its mettle as a noncanonical Trp synthase, but it remained to be seen whether the enzyme could become a more generalized ncAA synthase. The observation of activity with thiol- and nitrogen-based nucleophiles provided precedent for the possibility that TrpB can accept molecules that are not explicitly indole-like to create C–N,^{27,54,56,67,68} C–S,⁶⁹ and C–Se⁶⁹ bonds (**Figure 1-1**). In principle, any sufficiently activated nucleophile that fits in the active site could react with the amino-acrylate. Carbon-based nucleophiles would be attractive synthons for TrpB, allowing for enzymatic C–C bond formation to make a broad panel of ncAAs. One major challenge, however, is that strong carbon-based nucleophiles, which are normally accessed via deprotonation of weakly-acidic C–H bonds, are highly disfavored in water because of their high basicity ($pK_a > 7$). Nevertheless, enzymes are known to exert profound effects on substrates to lower activation barriers, making the endeavor at least worth investigating. To our surprise and satisfaction, we discovered that TrpB can react with a number of carbon nucleophiles to form novel ncAAs.

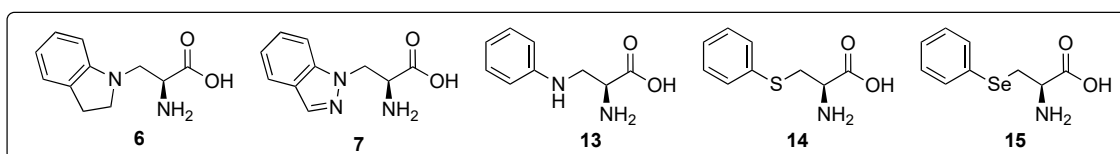
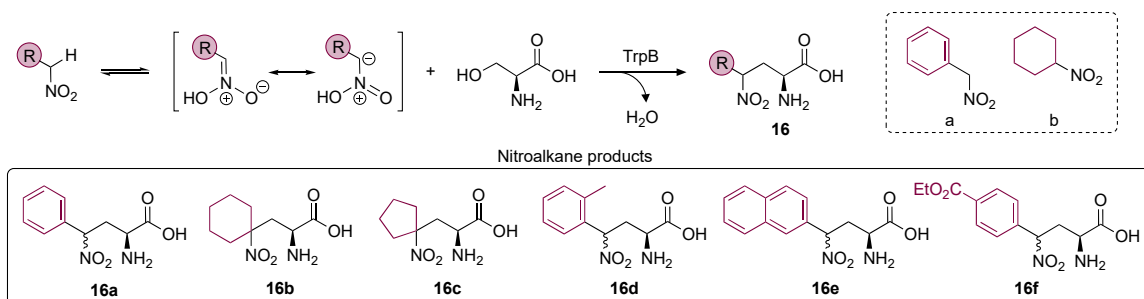


Figure 1-1. Products of non-carbon and non-indole nucleophiles produced by TrpS and TrpB. ^{27,54,56,67-69}

Nitroalkanes readily tautomerize to form a nucleophilic carbon alpha to the nitro group and have been used in the past as substrates for C–C bond formation reactions with electrophile-activating enzymes.⁷⁰⁻⁷² They have also been shown to react with chemically formed amino-acrylates to synthesize a wide range of amino acids, albeit under harsh conditions and with no enantioselectivity.⁷³⁻⁷⁶ This led David Romney to hypothesize that nitroalkanes could act as nucleophiles in the TrpB β -elimination reaction.⁷⁷ This was indeed the case, and many of the pre-existing TrpB variants they tested displayed at least some activity with (nitromethyl)benzene (**Scheme 1-5, a**) and the more sterically unwieldy nitrocyclohexane (**Scheme 1-5, b**). At the standard screening temperature of 75 °C,

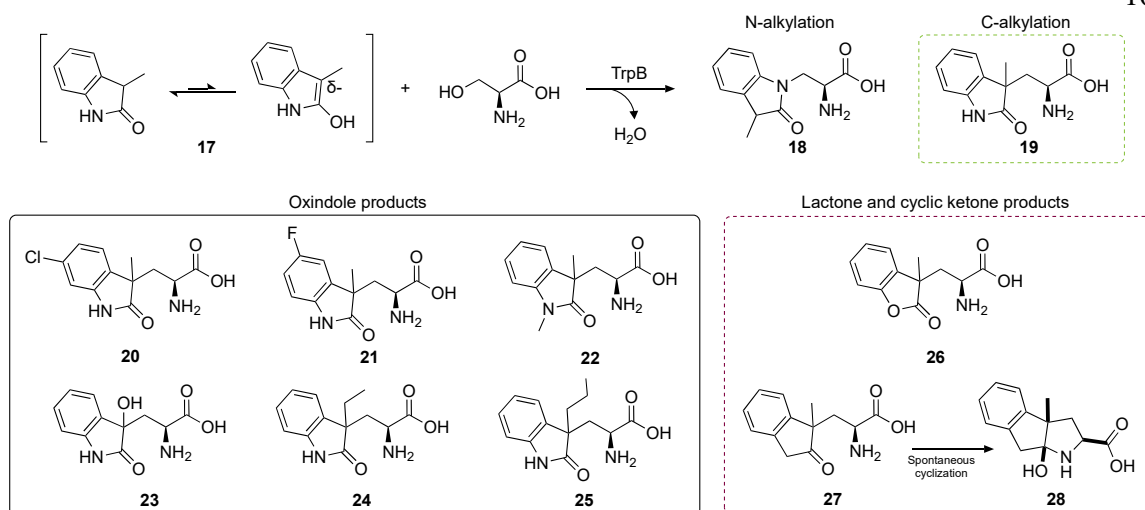
(nitromethyl)benzene was found to decompose, leading the authors to reduce the reaction temperature to 50 °C. The variant with the best activity on both substrates was subjected to several rounds of site-saturation mutagenesis (SSM) targeting the active site to ultimately produce two specialized variants that both exhibited a maximum of 2,700 turnovers with their respective substrates. Because nucleophilicity is dependent on pH, the authors investigated the effect of pH and discovered that higher pH (pH 9.0) did not necessarily improve the initial reaction rate, but did result in higher total turnovers and consequently higher yield. The authors probed the substrate scopes of the two enzymes with a panel of nitrocyclohexane and (nitromethyl)benzene derivatives (**Scheme 1-5, boxed**). Unfortunately, the carbon alpha to the nitro group is still readily deprotonated after product formation, resulting in stereoablation of the products with newly formed chiral centers. Nevertheless, the authors suggested that the platform could be engineered for α -nitro-substituted substrates to synthesize ncAAs with multiple chiral centers.



Scheme 1-5. Reaction scheme and product scope (boxed) of engineered TrpB with nitroalkanes as nucleophiles for ncAA synthesis. Nitroalkanes readily tautomerize in water to form a nucleophilic carbon species that reacts with the amino-acrylate intermediate in TrpB to form a new C–C bond. Model substrates a and b (dashed box) were used for directed evolution of TrpB.⁷⁷

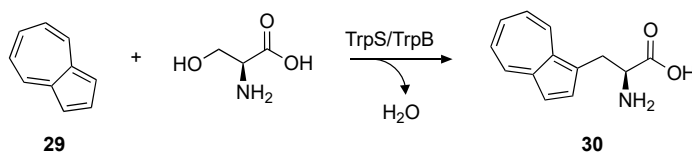
Biologically active compounds are replete with all-carbon quaternary centers whose regio- and stereoselective formation is a challenge for both organic synthesis and biocatalysis. In a recent study, Markus Dick and Nicholas Sarai engineered TrpB for selective quaternary bond formation with 3-substituted oxindoles.⁷⁸ 3,3-Disubstituted oxindoles are a common motif in natural and synthetic bioactive compounds. Similarly to the tautomerization of nitroalkanes to form a nucleophilic carbon, oxindoles exist in an equilibrium between keto and enol

tautomers, the latter of which is nucleophilic. In contrast to the nitroalkanes, whose tautomerization in water could be readily observed by NMR, the equivalent tautomer could not be observed for oxindoles, which suggested that there could be a significant activation barrier for TrpB to overcome in order for nucleophilic attack to occur. 3-Methyloxindole (**17**) was initially tested due to its abundance in synthetic and natural compounds. However, the methyl group appeared to sterically hinder nucleophilic attack from C₃, and the initial TrpB variants that were tested primarily formed the N-alkylation product (**Scheme 1-6, 18**). A few variants, however, also formed the desired C₃-alkylation product, providing a foothold for directed evolution. To prevent N-alkylation, Dick and Sarai chose to use 1,3-dimethyloxindole (1-CH₃-**17**) as the model substrate to begin evolution. Impressively, a single generation was enough to switch the chemoselectivity almost entirely from N- to C₃-alkylation, allowing the authors to continue with their original substrate, 3-methyloxindole, for the remainder of the evolution. After three more rounds of mutagenesis and screening, they obtained variant *Pf*TrpB^{quat} that exhibited >99% chemoselectivity for C₃-alkylation and a 52% yield (122 mg) from 1 mmol of 3-methyloxindole using only 100 mL of *E. coli* cell culture. They determined that the enzyme was *S,S*-stereoselective with 3-methyloxindole, though stereoselectivity decreased with bulkier substitutions at C₃. The enzyme could also tolerate ketone and lactone structures, demonstrating its ability to create quaternary centers with a diverse suite of carbonyl-containing nucleophiles bearing a tertiary carbon (**Scheme 1-6, solid black box and dashed pink box**).



Scheme 1-6. Reaction scheme and product scope (boxed) of oxindoles as nucleophiles for TrpB-catalyzed ncAA synthesis. Initially, TrpB variants catalyzed primarily N-alkylation of **17** to form **18**. With directed evolution, the regio- and chemoselectivity were switched to favor the desired C–C product (**19**, green dashed box). The final evolved variant accepted oxindoles with aryl (**20**, **21**), N-methyl (**22**), and C₃ substitutions (**23**, **24**, **25**, black box). It also catalyzed the formation of lactone (**26**) and cyclic ketone (**27** & **28**) products (pink dashed box), which suggests that evolution could be applied to expand the scope of ncAA synthesis to encompass other nucleophiles.⁷⁸

The ability of TrpB to accept non-indole substrates such as nitroalkanes and oxindoles inspired other members of our lab to explore more molecules with known nucleophilic character such as the cyclic aromatic hydrocarbon azulene (Scheme 7, 29). Until this point, carbon nucleophiles accepted by TrpS and TrpB possessed heteroatoms that could stabilize the accumulation of charge during nucleophilic attack. Although azulene has no heteroatoms, it experiences a permanent dipole readily apparent in its resonance structure, which is a cycloheptatrienyl cation (tropylium) fused to a cyclopentadienyl anion (Cp⁻). We hypothesized that the electron accumulation of the Cp⁻ stabilized by the tropylium system could promote nucleophilic attack by azulene on the amino-acrylate to form the ncAA β-(1-azulenyl)-L-alanine (**Scheme 1-7**, AzAla, 30).⁷⁹ AzAla is a blue Trp isostere whose fluorescent properties have been leveraged for spectroscopic studies,^{80–82} but whose use was limited due to its difficult, time-sensitive, multi-step synthesis.⁸³ The discovery, and engineering of, TrpB AzAla activity is detailed in Chapter II.



Scheme 1-7. TrpB-catalyzed synthesis of β -(1-azulenyl)-l-alanine (AzAla) (30) from azulene (29) and serine.⁷⁹

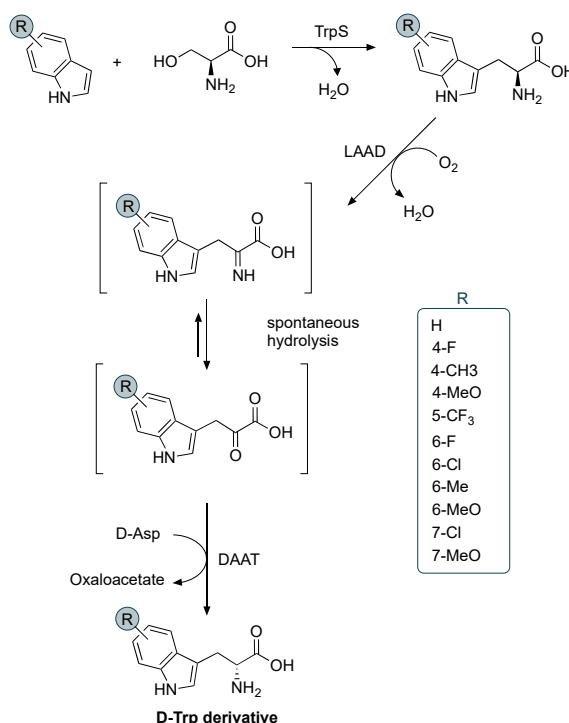
1.5 Biocatalytic cascades

Designing and optimizing enzymatic cascades can be a laborious process, and to date many do not best their synthetic rivals. Nonetheless, their continued development is paramount for realizing the potential of biocatalysis as a sustainable route to many of the world's chemicals. The ever-expanding catalog of engineered biocatalysts and advances in metabolic engineering have transformed the once pipe dream of whole-cell biocatalytic cascades into an attainable reality. In this realm, stereo- and regioselective enzymes that can be engineered easily and expressed heterologously reign supreme. It is thus not surprising that TrpS and TrpB—which are simple to use and boast large scopes of biologically relevant products—have already been used in a number of cascades. Some of the *in vivo* implementations have required only the host's native TrpS, which speaks to the latent potential of this remarkable complex. Others have made use of the simplicity provided by the stand-alone TrpB platform. In the following section, we highlight notable applications of TrpS and TrpB in biocatalytic cascades and discuss their biotechnological relevance.

1.5.1 D-Amino acids

Although L-amino acids comprise an overwhelming majority of amino acids in natural and synthetic compounds, their mirror counterparts are still found in many bioactive molecules and are important targets for enantiopure synthesis. D-Amino acids face similar synthetic challenges to L-amino acids but lack their diverse abundance of synthases, making direct biocatalytic access difficult. Unfortunately, TrpS' strict retention of stereoselectivity for making the L-amino acid hinders its ability to be repurposed as a D-amino acid synthase. Nevertheless, because TrpS still represents a simple way to make Trp derivatives, numerous groups have combined TrpS with downstream enzymes to access various D-Trps.

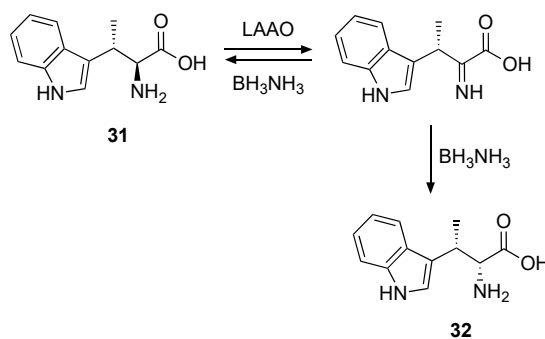
Parmeggiani and colleagues coupled TrpS from *S. enterica* (*Se*TrpS) with an L-amino acid deaminase (LAAD) followed by an engineered D-alanine aminotransferase (DAAT) to synthesize d-Trp derivatives in a one-pot, two-step transformation (**Scheme 1-8**).⁸⁴ LAAD and DAAT were found to possess promiscuous activity for L-Ser, putting Ser in direct competition with L-Trp for stereoinversion and lowering the overall enantiomeric excess (ee) of the product. To circumvent this, the cascade was converted into a one-pot telescopic system whereby the stereoinversion biocatalysts were introduced after the TrpS-mediated synthesis was complete. This proved to be an effective strategy to synthesize numerous D-Trp derivatives at gram scale with high yields and high ee.



Scheme 1-8. Biocatalytic cascade for synthesis of D-Trp derivatives. In the first reaction, a substituted indole is transformed into its respective Trp derivative by TrpS. In the second reaction, an L-amino acid deaminase (LAAD) deaminates the Trp to form an imine intermediate, which then spontaneously hydrolyzes to the α -keto acid. A D-alanine aminotransferase (DAAT) transaminates the α -keto acid with D-Asp, forming oxaloacetate and the D-Trp. Adapted with permission from ref. [85]. Copyright: 2019, ACS Catalysis.

In their preparation of β -branched Trps using TrpS, the Micklefield group used an L-amino acid oxidase (LAAO) and excess ammonia borane to produce the enantiopure d-configured

epimer of β -methylTrp (**Scheme 1-9**, (2*R*,3*S*)- β -methylTrp, **32**).⁶⁵ Ammonia borane nonspecifically reduces the imine product formed by the LAAO to form both the L- and D-amino acids; LAAO re-oxidizes the L-isomer to the imine while the D-isomer accumulates. The synthesis was performed in two steps in separate pots and resulted in an overall yield of 66%. While the authors only demonstrated this proof of concept with β -methylTrp (**31**), the LAAO they used exhibits a broad substrate scope, and it is likely that the method could be used to make other derivatized D- β -branched Trps.⁸⁵



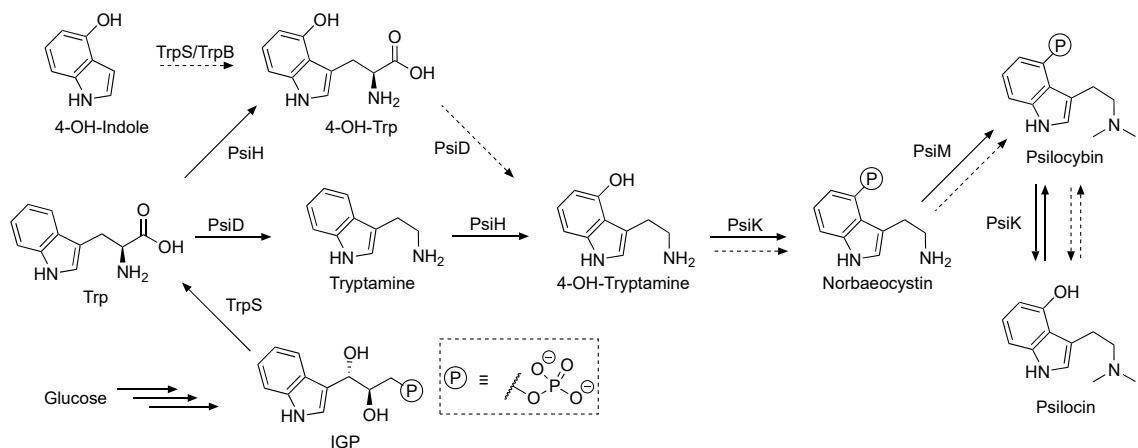
Scheme 1-9. One-pot chemoenzymatic approach for the stereoinversion of β -methylTrp (31**) to form D- β -methylTrp (**32**).**⁶⁵

1.5.2 Tryptamine products

In addition to its essential role as a proteinogenic amino acid, Trp is a precursor to numerous primary and secondary metabolites across all domains of life.⁸⁶ Tryptamines are one such class of Trp-derived molecules that possess a wide range of bioactive properties. Reflecting their importance, significant efforts to develop synthetic approaches for tryptamines have resulted in several effective methodologies.⁸⁷⁻⁹⁰ However, there is still room for biocatalysis to improve upon the cost, sustainability, and level of oversight needed for their synthesis. Recently, there has been interest in the study of psychoactive natural products like psilocybin as treatments for psychological and neurological afflictions. Psilocybin is a hallucinogenic tryptamine that is an effective treatment option for patients with anxiety,⁹¹ substance addiction,^{92,93} and depression,^{94,95} and it is possible that the molecule will be approved as a pharmaceutical drug. Unfortunately, psilocybin is produced only in very small and

inconsistent amounts by the mushroom *Psilocybe cubensis*, making commercial extraction impractical.⁹⁶

Recent elucidation of the natural biosynthetic pathway by the Hoffmeister group revealed that TrpS lies upstream of only four enzymes, PsiH, PsiD, PsiK, and PsiM (which provide monooxygenase, decarboxylase, kinase, and methyltransferase activities, respectively) to reach psilocybin (**Scheme 1-10, black arrows**).⁹⁷ The relative simplicity of the pathway coupled with the product's newfound pharmacological relevance encouraged them to investigate whether the cascade could be expressed in a model host organism to provide a scalable biosynthetic route. Indeed, the proteins expressed in *E. coli* allowed Fricke et al. to validate the putative activities of each enzyme and demonstrate an *in vitro* biosynthetic cascade of psilocybin.⁹⁸ In characterizing the enzymes, they discovered, perhaps unsurprisingly, that *P. cubensis* TrpB (*Pc*TrpB) could accept 4-hydroxyindole to produce 4-hydroxyTrp. Fortuitously, PsiD could accept 4-hydroxyTrp, which obviated the need for PsiH and further simplified the *in vitro* synthesis to a four-enzyme cascade that starts from 4-hydroxyindole (**Scheme 1-10, dashed arrows**).



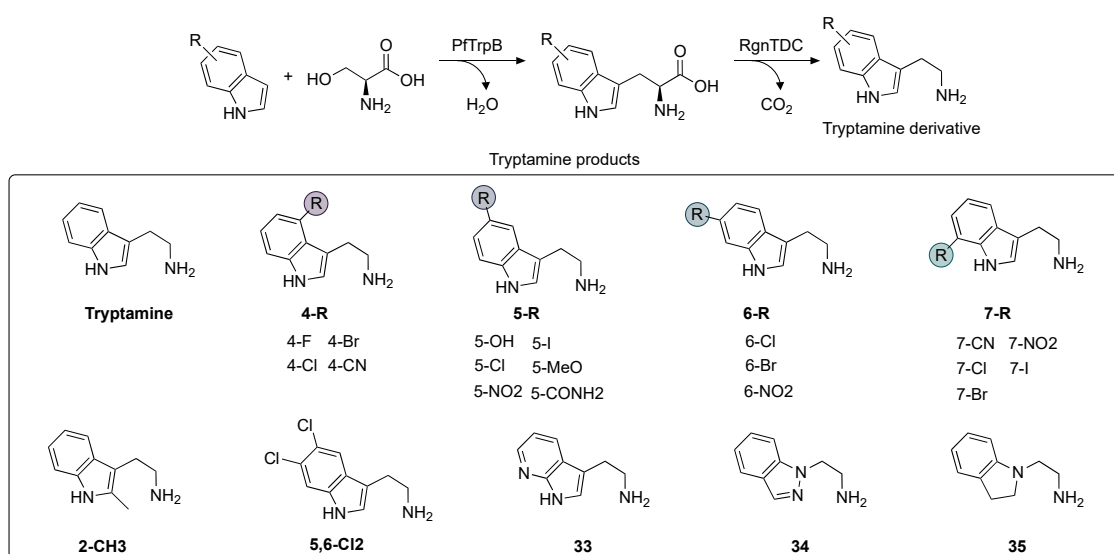
Scheme 1-10. Enzymatic synthesis of psilocybin. Black arrows represent the natural pathway and the pathway implemented by Milne et al. in *S. cerevisiae*.⁹⁹ Dashed arrows represent the synthetic pathway that exploits promiscuity of TrpS to synthesize 4-hydroxyTrp from 4-hydroxyindole.^{98,100}

Since then, there have been several instantiations of the cascade in different hosts. The first *in vivo* attempt ported the four enzymes downstream of TrpS into *Aspergillus nidulans*, accomplishing a modest 110 mg/L titer and more importantly establishing precedence for further *in vivo* applications.¹⁰¹ Adams et al. transferred the cascade without PsiH into *E. coli*, instead exploiting the promiscuity of *EcTrpS* to synthesize 4-hydroxyTrp from 4-hydroxyindole that was provided exogenously.¹⁰⁰ Attempts at *EcTrpS* overexpression did not improve titer, with native levels of expression sufficient for the pathway's flux. Scale up led to 1.16 g/L of psilocybin after 72 hours, a ten-fold enhancement over the previous method. Most recently, Milne et al. transferred the entire pathway into *Saccharomyces cerevisiae* for the *de novo* biosynthetic production of psilocybin.⁹⁹ They chose to use the natural functionality of TrpS to engineer a biosynthetic route to psilocybin rather than exploit the enzyme's promiscuity for relatively costly 4-hydroxyindole. A fed-batch fermentation process yielded 627 mg/L of psilocybin and 580 mg/L of psilocin, the dephosphorylated bioactive form of psilocybin, after 200 hours. However, their metabolically engineered pathway takes about three-fold longer to reach approximately the same titer as the *in vivo* pathway starting from 4-hydroxyindole put forth by Adams et al.¹⁰⁰

Psilocybin is one of the most well-known examples of a tryptamine, perhaps second only to the neurotransmitter serotonin. However, tryptamines are an abundant motif among alkaloid natural products, and substitutions around the aromatic indole ring have profound effects on their bioactive properties. In Nature, substitutions are installed after Trp biosynthesis by specific enzymes, not unlike the natural psilocybin pathway. Harnessing these enzymes for biocatalysis to produce tryptamine derivatives, however, presents the arduous task of identifying and expressing separate tailoring enzymes for different Trp modifications.¹⁰²

The modular and convergent nature of TrpB to combine substituted indoles with Ser offers a simpler and more general method to access Trp analogs. Buller and colleagues hypothesized that coupling a stand-alone TrpB with a promiscuous Trp decarboxylase would create a simple and streamlined route to diverse tryptamines.¹⁰³ No Trp decarboxylases that accept a broad range of substrates had been reported, so they tested a variant from the gut

microbe *Ruminococcus gnavus* (*RgnTDC*), whose active site appeared to be large enough to accommodate substituted indoles. *RgnTDC* possessed relatively high promiscuous activity suitable for immediate biocatalytic application with TrpB. Because the two enzymes operate at dramatically different optimal temperatures (75 vs. 37 °C), *PfTrpB*^{2B9} was combined with *RgnTDC* in a one-pot, two-step reaction to produce a range of tryptamine derivatives with isolated yields ranging from 12–99% (**Scheme 1-11**). The ability to access a large number of products by combining TrpB and *RgnTDC* is a testament to the versatility and power that generalist enzymes bring to biocatalysis.



Scheme 1-11. One-pot, two-step synthesis of tryptamine derivatives with TrpB and *Ruminococcus gnavus* Trp decarboxylase (*RgnTDC*). Top: reaction scheme, bottom boxed: product scope.

1.6 Summary and outlook

TrpS is a remarkable enzyme and ideal ncAA synthase. Its long history as a model enzyme led to substantial knowledge about its allosteric regulation and catalytic mechanism that granted enzyme engineers the ability to re-imagine its function. In this chapter, we have covered major applications of TrpS, from humble beginnings preparing simple Trp analogs to becoming a powerful platform of TrpB enzymes that produce entirely new-to-nature ncAAs, as well as its use in biocatalytic cascades to make new Trp-derived products. The no-frills TrpB platform dramatically simplifies enzyme engineering efforts, enabling the

rapid expansion and exploration of the synthase's substrate scope. Recent success in evolving TrpB to catalyze C–C bonds with new, non-indole nucleophiles bodes well for the platform's further expansion into ncAA space. We envision that the discovery of highly stable and evolvable TrpB variants that function well at moderate temperatures will promote future applications in cascades, replacing multi-step syntheses that currently must accommodate different temperature optima. Mesophilic TrpB variants may also be used for *in vivo* synthesis to improve intracellular delivery of the ncAA (indole passes through cellular membranes more readily than a charged amino acid), reducing the amount of product that must be supplied exogenously, which could improve incorporation rates and reduce costs. *In vivo* synthesis and incorporation of ncAAs into proteins or secondary metabolites might also be used for robust biocontainment.^{100,104,105}

Although many substituted indoles are commercially available, one of the roadblocks for biocatalytic synthesis of Trp derivatives by TrpS and TrpB is the high price and limited availability of these substrates. This issue is highlighted by Milne et al., who used PsiH to install a hydroxyl group on Trp rather than provide expensive 4-hydroxyindole to TrpS directly.⁹⁹ However, as we reach beyond the domain of naturally occurring chemical motifs, it is more difficult to find enzymes like PsiH that can make desired Trp modifications. In this non-natural space the TrpB platform truly shines for ncAA synthesis; it is simpler to engineer TrpB to accept new derivatives than it is to discover and engineer whole new enzymes that modify Trp. Therefore, advances in indole analog synthesis will continue to make TrpB a desirable route to Trp analogs. We expect that further applications of TrpS and TrpB are on the horizon and demonstrations like those from Parmeggiani et al.⁸⁴ and McDonald et al.¹⁰³ will inspire others to use this biocatalyst extraordinaire.

Chapter I Bibliography

1. Blaskovich, M. A. *Handbook on syntheses of amino acids: General routes for the syntheses of amino acids*. (Oxford University Press; American Chemical Society, 2010).
2. Budisa, N. *Engineering the genetic code: Expanding the amino acid repertoire for the*

design of novel proteins. (John Wiley & Sons, 2006).

3. Saleh, A. M., Wilding, K. M., Calve, S., Bundy, B. C., & Kinzer-Ursem, T. L. Non-canonical amino acid labeling in proteomics and biotechnology. *J. Biol. Eng.* **13**, 43 (2019).
4. Liu, C. C. & Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **79**, 413–444 (2010).
5. Beatty, K. E. & Tirrell, D. A. Noncanonical Amino Acids in Protein Science and Engineering. in *Protein Engineering* 127–153 (Springer Berlin Heidelberg, 2009). doi:10.1007/978-3-540-70941-1_5
6. Pommerehne, K., Walisko, J., Ebersbach, A., & Krull, R. The antitumor antibiotic rebeccamycin—challenges and advanced approaches in production processes. *Appl. Microbiol. Biotechnol.* **103**, 3627–3636 (2019).
7. Lam, H. Y. et al. Total Synthesis of Daptomycin by Cyclization via a Chemoselective Serine Ligation. *J. Am. Chem. Soc.* **135**, 6272–6279 (2013).
8. Rezhdo, A., Islam, M., Huang, M., & Van Deventer, J. A. Future prospects for noncanonical amino acids in biological therapeutics. *Curr. Opin. Biotechnol.* **60**, 168–178 (2019).
9. Kieffer, M. E., Repka, L. M., & Reisman, S. E. Enantioselective synthesis of tryptophan derivatives by a tandem Friedel-Crafts conjugate addition/asymmetric protonation reaction. *J. Am. Chem. Soc.* **134**, 5131–5137 (2012).
10. France, S. P., Hepworth, L. J., Turner, N. J., & Flitsch, S. L. Constructing Biocatalytic Cascades: In Vitro and in Vivo Approaches to de Novo Multi-Enzyme Pathways. *ACS Catal.* **7**, 710–724 (2017).
11. Almhjell, P. J., Boville, C. E., & Arnold, F. H. Engineering enzymes for noncanonical amino acid synthesis. *Chem. Soc. Rev.* **47**, 8980–8997 (2018).
12. Phillips, R. S. Synthetic applications of tryptophan synthase. *Tetrahedron: Asymmetry* **15**, 2787–2792 (2004).
13. Romney, D. K., Murciano-Calles, J., Wehrmüller, J. E., & Arnold, F. H. Unlocking Reactivity of TrpB: A General Biocatalytic Platform for Synthesis of Tryptophan Analogues. *J. Am. Chem. Soc.* **139**, 10769–10776 (2017).

14. Yanofsky, C. Using studies on tryptophan metabolism to answer basic biological questions. *J. Biol. Chem.* **278**, 10859–10878 (2003).
15. Barends, T. R., Dunn, M. F., & Schlichting, I. Tryptophan synthase, an allosteric molecular factory. *Curr. Opin. Chem. Biol.* **12**, 593–600 (2008).
16. Huang, X., Holden, H. M., & Raushel, F. M. Channeling of Substrates and Intermediates in Enzyme-Catalyzed Reactions. *Annu. Rev. Biochem.* **70**, 149–180 (2001).
17. Dunn, M. F. Allosteric regulation of substrate channeling and catalysis in the tryptophan synthase holoenzyme complex. *Arch. Biochem. Biophys.* **519**, 154–166 (2012).
18. Schupfner, M., Straub, K., Busch, F., Merkl, R., & Sterner, R. Analysis of allosteric communication in a multienzyme complex by ancestral sequence reconstruction. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 346–354 (2020).
19. Banik, U., Zhu, D. M., Chock, P. B., & Miles, E. W. The Tryptophan Synthase $\alpha\beta_2$ Complex: Kinetic Studies with a Mutant Enzyme (β K87T) To Provide Evidence for Allosteric Activation by an Aminoacrylate Intermediate. *Biochemistry* **34**, 12704–12711 (1995).
20. Caulkins, B. G. et al. Protonation states of the tryptophan synthase internal aldimine active site from solid-state NMR spectroscopy: Direct observation of the protonated schiff base linkage to pyridoxal-5'-phosphate. *J. Am. Chem. Soc.* **136**, 12824–12827 (2014).
21. Dunathan, H. C. Stereochemical aspects of pyridoxal phosphate catalysis. in *Advances in Enzymology and Related Areas of Molecular Biology* (ed. Meister, A.) **35**, 79–134 (John Wiley & Sons, Inc, 1971).
22. Toney, M. D. Controlling reaction specificity in pyridoxal phosphate enzymes. *Biochim. Biophys. Acta (BBA)- Proteins Proteomics* **1814**, 1407–1418 (2011).
23. Blumenstein, L. et al. β Q114N and β T110V mutations reveal a critically important role of the substrate α -carboxylate site in the reaction specificity of tryptophan synthase. *Biochemistry* **46**, 14100–14116 (2007).

24. Caulkins, B. G. et al. NMR Crystallography of a Carbanionic Intermediate in Tryptophan Synthase: Chemical Structure, Tautomerization, and Reaction Specificity. *J. Am. Chem. Soc.* **138**, 15214–15226 (2016).
25. Ngo, H. et al. Allosteric regulation of substrate channeling in tryptophan synthase: modulation of the L-serine reaction in stage I of the β -reaction by α -site ligands. *Biochemistry* **46**, 7740–7753 (2007).
26. Miles, E. W. & McPhie, P. Evidence for a rate determining proton abstraction in the serine deaminase reaction of the β 2 subunit of tryptophan synthetase. *J. Biol. Chem.* **249**, 2852–2857 (1974).
27. Brzovic, P. S., Kayastha, A. M., Miles, E. W., & Dunn, M. F. Substitution of glutamic acid 109 by aspartic acid alters the substrate specificity and catalytic activity of the β -subunit in the tryptophan synthase holoenzyme complex from *Salmonella typhimurium*. *Biochemistry* **31**, 1180–1190 (1992).
28. Phillips, R. S., Wilson Miles, E., & Cohen, L. A. Differential inhibition of tryptophan synthase and of tryptophanase by the two diastereoisomers of 2,3-dihydro-L-tryptophan. *J. Biol. Chem.* **260**, 14665–14670 (1985).
29. Cash, M. T., Miles, E. W., & Phillips, R. S. The reaction of indole with the aminoacrylate intermediate of *Salmonella typhimurium* tryptophan synthase: Observation of a primary kinetic isotope effect with 3-[^2H]indole. *Arch. Biochem. Biophys.* **432**, 233–243 (2004).
30. Lane, A. N. & Kirschner, K. The Catalytic Mechanism of Tryptophan Synthase from *Escherichia coli*. *Eur. J. Biochem.* **129**, 571–582 (1983).
31. Sinkeldam, R. W., Greco, N. J., & Tor, Y. Fluorescent Analogs of Biomolecular Building Blocks: Design, Properties, and Applications. *Chem. Rev.* **110**, 2579–2619 (2010).
32. Hilaire, M. R. et al. Blue fluorescent amino acid for biological spectroscopy and microscopy. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 6005–6009 (2017).
33. Ross, J. B. A. et al. Spectral enhancement of proteins: Biological incorporation and fluorescence characterization of 5-hydroxytryptophan in bacteriophage λ cI repressor. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 12023–12027 (1992).

34. Talukder, P. et al. Tryptophan-based fluorophores for studying protein conformational changes. *Bioorganic Med. Chem.* **22**, 5924–5934 (2014).
35. Seifert, M. H. J. et al. Slow exchange in the chromophore of a green fluorescent protein variant. *J. Am. Chem. Soc.* **124**, 7932–7942 (2002).
36. Bae, J. H. et al. Incorporation of β -selenolo[3,2-b]pyrrolyl-alanine into proteins for phase determination in protein X-ray crystallography. *J. Mol. Biol.* **309**, 925–936 (2001).
37. Boutureira, O. & Bernardes, G. J. L. Advances in chemical protein modification. *Chem. Rev.* **115**, 2174–2195 (2015).
38. Jordan, P. A. & Moore, B. S. Biosynthetic Pathway Connects Cryptic Ribosomally Synthesized Posttranslationally Modified Peptide Genes with Pyrroloquinoline Alkaloids. *Cell Chem. Biol.* **23**, 1504–1514 (2016).
39. Alkhalaf, L. M. & Ryan, K. S. Biosynthetic manipulation of tryptophan in bacteria: Pathways and mechanisms. *Chem. Biol.* **22**, 317–328 (2015).
40. Hall, A. N., Lea, D. J., & Rhydton, H. N. The behaviour of the Bz-methylindoles as substrates and inhibitors for *Neurospora crassa* tryptophan synthase. *Biochem. J.* **84**, 12–16 (1962).
41. Held, W. A. & Smith, O. H. Regulation of the *Escherichia coli* tryptophan operon by early reactions in the aromatic pathway. *J. Bacteriol.* **101**, 202–208 (1970).
42. Wilcox, M. The enzymatic synthesis of l-tryptophan analogues. *Anal. Biochem.* **59**, 436–440 (1974).
43. Saito, A. & Rilling, H. C. Synthesis of Photoreactive Tryptophan Analogs: Enzymatic Conversion of Azidoindoles to Azidotryptophans. *Prep. Biochem.* **11**, 535–546 (1981).
44. Lee, M. & Phillips, R. S. Enzymatic synthesis of chloro-L-tryptophans. *Bioorganic Med. Chem. Lett.* **2**, 1563–1564 (1992).
45. Phillips, R. S., Cohen, L. A., Annby, U., Wensbo, D., & Gronowitz, S. Enzymatic synthesis of Thia-L-tryptophans. *Bioorganic Med. Chem. Lett.* **5**, 1133–1134 (1995).
46. Welch, M. & Phillips, R. S. Enzymatic syntheses of 6-(4H-selenolo[3,2-b]pyrrolyl)-L-alanine, 4-(6H-selenolo[2,3-b]pyrrolyl)-L-alanine, and 6-(4H-furo[3,2-

- b]pyrrolyl)-L-alanine. *Bioorganic Med. Chem. Lett.* **9**, 637–640 (1999).
47. Boles, J. O., Henderson, J., Hatch, D., & Silks, L. A. P. Synthesis and incorporation of [6,7]-selenatryptophan into dihydrofolate reductase. *Biochem. Biophys. Res. Commun.* **298**, 257–261 (2002).
 48. Sloan, M. J. & Phillips, R. S. Enzymatic synthesis of aza-l-tryptophans: The preparation of 5- and 6-Aza-l-tryptophan. *Bioorganic Med. Chem. Lett.* **2**, 1053–1056 (1992).
 49. Goss, R. J. M. & Newill, P. L. A. A convenient enzymatic synthesis of l-halotryptophans. *Chem. Commun.* 4924 (2006). doi:10.1039/b611929h
 50. Winn, M., Roy, A. D., Grüşchow, S., Parameswaran, R. S., & Goss, R. J. M. A convenient one-step synthesis of l-aminotryptophans and improved synthesis of 5-fluorotryptophan. *Bioorganic Med. Chem. Lett.* **18**, 4508–4510 (2008).
 51. Perni, S., Hackett, L., Goss, R. J. M., Simmons, M. J., & Overton, T. W. Optimisation of engineered *Escherichia coli* biofilms for enzymatic biosynthesis of l-halotryptophans. *AMB Express* **3**, 66 (2013).
 52. Smith, D. R. M. et al. The first one-pot synthesis of l-7-iodotryptophan from 7-iodoindole and serine, and an improved synthesis of other l-7-halotryptophans. *Org. Lett.* **16**, 2622–2625 (2014).
 53. Dunn, M. F. et al. The interconversion of *E. coli* tryptophan synthase intermediates is modulated by allosteric interactions. *Indian J. Biochem. Biophys.* **24**, 44–51 (1987).
 54. Roy, M., Keblawi, S., & Dunn, M. F. Stereoelectronic Control of Bond Formation in *Escherichia coli* Tryptophan Synthase: Substrate Specificity and Enzymatic Synthesis of the Novel Amino Acid Dihydroisotryptophan. *Biochemistry* **27**, 6698–6704 (1988).
 55. Kirschner, K., Wiskocil, R. L., Foehn, M., & Rezeau, L. The tryptophan synthase from *Escherichia coli*. *Eur. J. Biochem.* **60**, 513–523 (1975).
 56. Buller, A. R. et al. Directed evolution of the tryptophan synthase β -subunit for stand-alone function recapitulates allosteric activation. *Proc. Natl. Acad. Sci.* **112**, 14599–14604 (2015).
 57. Murciano-Calles, J., Romney, D. K., Brinkmann-Chen, S., Buller, A. R., & Arnold, F. H. A Panel of TrpB Biocatalysts Derived from Tryptophan Synthase through the

- Transfer of Mutations that Mimic Allosteric Activation. *Angew. Chemie - Int. Ed.* **55**, 11577–11581 (2016).
58. Boville, C. E., Romney, D. K., Almhjell, P. J., Sieben, M., & Arnold, F. H. Improved Synthesis of 4-Cyanotryptophan and Other Tryptophan Analogues in Aqueous Solvent Using Variants of TrpB from *Thermotoga maritima*. *J. Org. Chem.* **83**, 7447–7452 (2018).
 59. Zou, Y. et al. Stereospecific biosynthesis of β -methyltryptophan from L-tryptophan features a stereochemical switch. *Angew. Chemie - Int. Ed.* **52**, 12951–12955 (2013).
 60. Sawai, Y., Mizuno, M., Ito, T., Kawakami, J., & Yamano, M. A practical synthesis of enantiopure β -methyltryptophan ethyl ester for a preparation of diabetes drug. *Tetrahedron* **65**, 7122–7128 (2009).
 61. Herger, M. et al. Synthesis of β -Branched Tryptophan Analogues Using an Engineered Subunit of Tryptophan Synthase. *J. Am. Chem. Soc.* **138**, 8388–8391 (2016).
 62. Esaki, N., Tanaka, H., Miles, E. W., & Soda, K. Enzymatic synthesis of S-substituted L-cysteines with tryptophan synthase of *Escherichia coli*. *FEBS Lett.* **161**, 207–209 (1983).
 63. Kumagai, H. & Miles, E. W. The 8 protein of *Escherichia coli* tryptophan synthetase II. New β -elimination and β -replacement reactions. *Biochem. Biophys. Res. Commun.* **44**, 1271–1278 (1971).
 64. Buller, A. R., Van Roye, P., Murciano-Calles, J., & Arnold, F. H. Tryptophan Synthase Uses an Atypical Mechanism To Achieve Substrate Specificity. *Biochemistry* **55**, 7043–7046 (2016).
 65. Francis, D., Winn, M., Latham, J., Greaney, M. F., & Micklefield, J. An Engineered Tryptophan Synthase Opens New Enzymatic Pathways to β -Methyltryptophan and Derivatives. *ChemBioChem* **18**, 382–386 (2017).
 66. Boville, C. E. et al. Engineered biosynthesis of β -alkyl tryptophan analogs. *Angew. Chemie Int. Ed.* **9**, 14764–14768 (2018).
 67. Ferrari, D., Yang, L. H., Miles, E. W., & Dunn, M. F. β D305A mutant of tryptophan synthase shows strongly perturbed allosteric regulation and substrate specificity.

- Biochemistry* **40**, 7421–7432 (2001).
68. Tanaka, H. et al. Production of a novel tryptophan analog, β -l-indazole-L-alanine with tryptophan synthase of *Escherichia coli*. *FEBS Lett.* **196**, 357–360 (1986).
 69. Esaki, N. & Soda, K. Preparation of sulfur and selenium amino acids with microbial pyridoxal phosphate enzymes. *Methods Enzymol.* **143**, 291–297 (1987).
 70. Garrabou, X., Verez, R., & Hilvert, D. Enantiocomplementary synthesis of γ -nitroketones using designed and evolved carboligases. *J. Am. Chem. Soc.* **139**, 103–106 (2017).
 71. Garrabou, X., Macdonald, D. S., & Hilvert, D. Chemoselective Henry Condensations Catalyzed by Artificial Carboligases. *Chem. - A Eur. J.* **23**, 6001–6003 (2017).
 72. Milner, S. E., Moody, T. S., & Maguire, A. R. Biocatalytic approaches to the Henry (nitroaldol) reaction. *European J. Org. Chem.* 3059–3067 (2012). doi:10.1002/ejoc.201101840
 73. Bigge, C. F., Wu, J. P., Drummond, J. T., Coughenour, L. L., & Hanchin, C. M. Excitatory amino acids: 6-phosphonomethyltetrahydro-4-pyrimidinecarboxylic acids and their acyclic analogues are competitive N-methyl-D-aspartic acid receptor antagonists. *Bioorganic Med. Chem. Lett.* **2**, 207–212 (1992).
 74. Crossley, M. & Tansey, C. A Convenient Method for the Synthesis of β -Substituted α Amino Acids. Diastereoselective Conjugate Addition of Nitronates to a Chiral Dehydroalanine Derivative. *Aust. J. Chem.* **45**, 479–481 (1992).
 75. Crossley, M. J., Fung, Y. M., Potter, J. J., & Stamford, A. W. Convenient route to γ -nitro- α -amino acids: Conjugate addition of nitroalkanes to dehydroalanine derivatives. *J. Chem. Soc. - Perkin Trans. I* 1113–1121 (1998). doi:10.1039/a707065i
 76. Ballini, R. et al. Improved Synthesis of γ -Nitro- α -Amino Ester and Acid Derivatives. *Synthesis (Stuttg)*. 296–300 (2005). doi:10.1055/s-2004-834927
 77. Romney, D. K., Sarai, N. S., & Arnold, F. H. Nitroalkanes as Versatile Nucleophiles for Enzymatic Synthesis of Noncanonical Amino Acids. *ACS Catal.* **9**, 8726–8730 (2019).
 78. Dick, M., Sarai, N. S., Martynowycz, M. W., Gonen, T., & Arnold, F. H. Tailoring Tryptophan Synthase TrpB for Selective Quaternary Carbon Bond Formation. *J. Am.*

- Chem. Soc.* **141**, 19817–19822 (2019).
79. Watkins, E. J., Almhjell, P. J., & Arnold, F. H. Direct enzymatic synthesis of a deep-blue fluorescent noncanonical amino acid from azulene and serine. *ChemBioChem* **21**, 80–83 (2020).
 80. Gosavi, P. M., Moroz, Y. S., & Korendovych, I. V. β -(1-Azulenyl)-l-alanine-a functional probe for determination of pKa of histidine residues. *Chem. Commun.* **51**, 5347–5350 (2015).
 81. Moroz, Y. S., Binder, W., Nygren, P., Caputo, G. A., & Korendovych, I. V. Painting proteins blue: β -(1-azulenyl)-l-alanine as a probe for studying protein–protein interactions. *Chem. Commun.* **49**, 490–492 (2013).
 82. Baumann, T. et al. Site-Resolved Observation of Vibrational Energy Transfer Using a Genetically Encoded Ultrafast Heater. *Angew. Chemie - Int. Ed.* **58**, 2899–2903 (2019).
 83. Stempel, E., Kaml, R. F. X., Budisa, N., & Kalesse, M. Painting argyryns blue: Negishi cross-coupling for synthesis of deep-blue tryptophan analogue β -(1-azulenyl)-L-alanine and its incorporation into argyryrin C. *Bioorganic Med. Chem.* **26**, 5259–5269 (2018).
 84. Parmeggiani, F. et al. One-Pot Biocatalytic Synthesis of Substituted d -Tryptophans from Indoles Enabled by an Engineered Aminotransferase. *ACS Catal.* **9**, 3482–3486 (2019).
 85. Roff, G. J., Lloyd, R. C., & Turner, N. J. A Versatile Chemo-Enzymatic Route to Enantiomerically Pure β -Branched α -Amino Acids. *J. Am. Chem. Soc.* **126**, 4098–4099 (2004).
 86. Parthasarathy, A. et al. A Three-Ring circus: Metabolism of the three proteogenic aromatic amino acids and their role in the health of plants and animals. *Front. Mol. Biosci.* **5**, 29 (2018).
 87. Olsen, E. K. et al. Marine AChE inhibitors isolated from *Geodia barretti*: Natural compounds and their synthetic analogs. *Org. Biomol. Chem.* **14**, 1629–1640 (2016).
 88. Muratore, M. E. et al. Enantioselective Brønsted acid-catalyzed N-acyliminium cyclization cascades. *J. Am. Chem. Soc.* **131**, 10796–10797 (2009).

89. Bartolucci, S., Mari, M., Bedini, A., Piersanti, G., & Spadoni, G. Iridium-catalyzed direct synthesis of tryptamine derivatives from indoles: Exploiting N-protected β -amino alcohols as alkylating agents. *J. Org. Chem.* **80**, 3217–3222 (2015).
90. Righi, M. et al. Synthesis of tryptamine derivatives via a direct, one-pot reductive alkylation of indoles. *J. Org. Chem.* **77**, 6351–6357 (2012).
91. Grob, C. S. et al. Pilot study of psilocybin treatment for anxiety in patients with advanced-stage cancer. *Arch. Gen. Psychiatry* **68**, 71–78 (2011).
92. Johnson, M. W., Garcia-Romeu, A., & Griffiths, R. R. Long-term follow-up of psilocybin-facilitated smoking cessation. *Am. J. Drug Alcohol Abuse* **43**, 55–60 (2017).
93. Bogenschutz, M. P. et al. Psilocybin-assisted treatment for alcohol dependence: A proof-of-concept study. *J. Psychopharmacol.* **29**, 289–299 (2015).
94. Carhart-Harris, R. L. et al. Psilocybin for treatment-resistant depression: fMRI-measured brain mechanisms. *Sci. Rep.* **7**, 13187 (2017).
95. Carhart-Harris, R. L. et al. Psilocybin with psychological support for treatment-resistant depression: an open-label feasibility study. *The Lancet Psychiatry* **3**, 619–627 (2016).
96. Bigwood, J. & Beug, M. W. Variation of psilocybin and psilocin levels with repeated flushes (harvests) of mature sporocarps of *Psilocybe cubensis* (earle) singer. *J. Ethnopharmacol.* **5**, 287–291 (1982).
97. Fricke, J., Blei, F., & Hoffmeister, D. Enzymatic Synthesis of Psilocybin. *Angew. Chemie - Int. Ed.* **56**, 12352–12355 (2017).
98. Blei, F., Baldeweg, F., Fricke, J., & Hoffmeister, D. Biocatalytic Production of Psilocybin and Derivatives in Tryptophan Synthase-Enhanced Reactions. *Chem. - A Eur. J.* **24**, 10028–10031 (2018).
99. Milne, N. et al. Metabolic engineering of *Saccharomyces cerevisiae* for the production of psilocybin and related tryptamine derivatives. *Metab. Eng.* **60**, 25–36 (2020).
100. Adams, A. M. et al. In vivo production of psilocybin in *E. coli*. *Metab. Eng.* **56**, 111–119 (2019).
101. Hoefgen, S. et al. Facile assembly and fluorescence-based screening method for

- heterologous expression of biosynthetic pathways in fungi. *Metab. Eng.* **48**, 44–51 (2018).
102. Latham, J., Brandenburger, E., Shepherd, S. A., Menon, B. R. K., & Micklefield, J. Development of Halogenase Enzymes for Use in Synthesis. *Chem. Rev.* **118**, 232–269 (2018).
103. McDonald, A. D., Perkins, L. J., & Buller, A. R. Facile in Vitro Biocatalytic Production of Diverse Tryptamines. *ChemBioChem* **20**, 1939–1944 (2019).
104. Acevedo-Rocha, C. G. & Budisa, N. Xenomicrobiology: a roadmap for genetic code engineering. *Microb. Biotechnol.* **9**, 666–676 (2016).
105. Agostini, F. et al. Biocatalysis with Unnatural Amino Acids: Enzymology Meets Xenobiology. *Angew. Chemie - Int. Ed.* **56**, 9680–9703 (2017).

DIRECT ENZYMATIC SYNTHESIS OF A DEEP-BLUE
FLUORESCENT NONCANONICAL AMINO ACID FROM AZULENE
AND SERINE

Material from this chapter appears in: “**Watkins E. J.**, Almhjell P. J., Arnold F. H., Direct enzymatic synthesis of a deep-blue fluorescent noncanonical amino acid from azulene and serine. *Chembiochem*. 2020 Jan 15;21(1-2):80-83. doi: 10.1002/cbic.201900497.”

E.J.W.D. and P.J.A. participated in the conception, design, and execution of the research. E.J.W.D. designed the screen for identifying improved enzyme variants. E.J.W.D. and P.J.A. contributed equally to AzAla purification. P.J.A. performed enzyme kinetics. E.J.W.D. prepared the first manuscript and P.J.A. edited and constructed figures.

ABSTRACT

We report a simple, one-step enzymatic synthesis of the blue fluorescent noncanonical amino acid β -(1-azulenyl)-L-alanine (AzAla). Using an engineered tryptophan synthase β -subunit (TrpB), stereochemically pure AzAla can be synthesized at scale starting from commercially available azulene and L-serine. Mutation of a universally conserved catalytic glutamate in the active site to glycine has only a modest effect on native activity with indole but abolishes activity on azulene, suggesting that this glutamate activates azulene for nucleophilic attack by stabilization of the aromatic ion.

2.1 Introduction

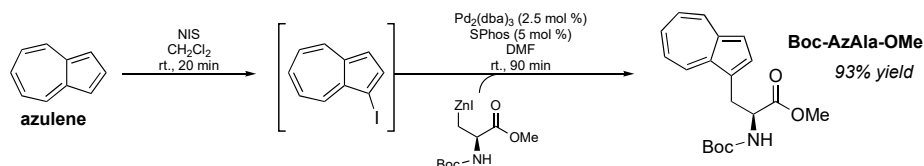
Proteins and peptides can be imbued with new chemical and physical properties via the inclusion of noncanonical amino acids (ncAAs). These molecules resemble the natural building blocks of proteins, but contain distinct structures and functional groups. When incorporated into proteins, ncAAs can serve as handles for chemical reactions or as spectroscopic probes to examine protein function, including reactivity, localization, and interaction with other biomolecules.¹⁻³ Unfortunately, applications of many potentially useful ncAAs are limited owing to their high cost and lack of availability. The paucity of available ncAAs also hinders engineering of aminoacyl tRNA synthetases (aaRS) necessary for site-specific, *in vivo* ncAA incorporation into proteins.

The ncAA β -(1-azulenyl)-l-alanine (AzAla, **Scheme 2-1**) is a tryptophan (Trp) isostere with unique fluorescent properties that make it a useful spectroscopic probe for investigating protein dynamics and protein-protein interactions. It can be incorporated into proteins in place of Trp without significantly disturbing tertiary structure or function.⁴⁻⁸ In contrast to Trp, its spectroscopic properties are insensitive to the environment, making it ideal in contexts where local conditions and quenchers (e.g., methionine, histidine) could complicate analysis of fluorescent signals.⁴ Its qualities have been leveraged for Förster resonance energy transfer (FRET) experiments to elucidate protein-protein interactions as well as vibrational energy transfer (VET) studies to probe anisotropic energy flow within proteins.^{9,10} Recently, a method for synthesis of AzAla via Negishi cross-coupling was described (**Scheme 2-1a**).¹¹ Although proceeding with good yields on gram scale, the multi-step process is highly time-sensitive and uses precious metal catalysts and organic solvents. A simpler route to AzAla would expand its applications in biochemical studies.

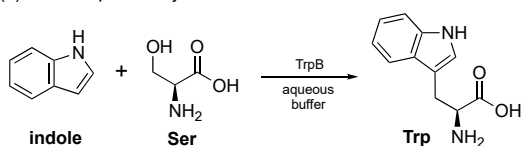
Researchers have begun to look to enzymes as complementary or alternative approaches for the synthesis of enantiomerically pure ncAAs.¹² Enzymes can perform enantio- and regioselective chemistry in the presence of reactive moieties such as primary amines, obviating the need for expensive and intricate chiral catalysts, chiral separations, and protecting groups. To this end, our lab previously reported the directed evolution of the

tryptophan synthase β -subunit (TrpB) as a stand-alone biocatalytic platform for the synthesis of diverse tryptophan analogues (**Scheme 2-1b**).^{13,14} Here we report a simple, efficient route for synthesis of AzAla from stable, commercially available starting materials using an engineered TrpB (**Scheme 2-1c**).

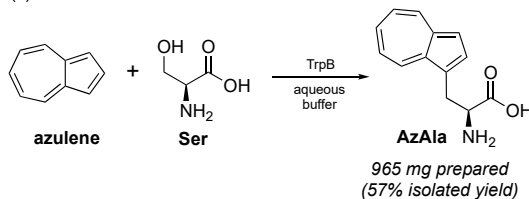
(a) Current synthetic route to AzAla



(b) Native TrpB activity



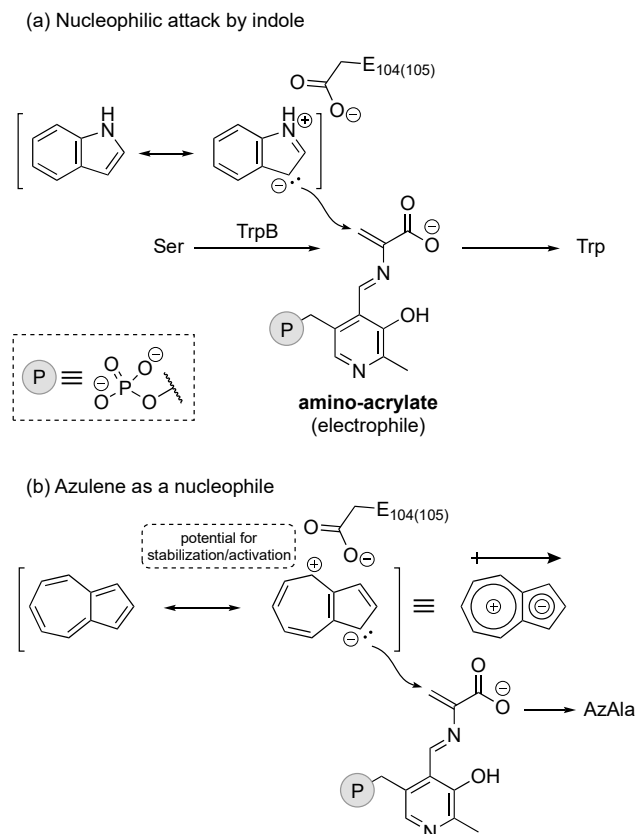
(c) This work



Scheme 2-1. Synthesis of AzAla and Native TrpB activity. (a) Current synthetic route to Boc-protected AzAla. (b) TrpB natively catalyzes the condensation of indole and serine to form tryptophan. (c) Single-step biocatalytic synthesis of AzAla from azulene and serine described in this work.

TrpB is a Type-II pyridoxal phosphate (PLP)-dependent enzyme that natively performs a conjugate addition reaction between indole and L-serine (Ser) to make Trp. During the catalytic cycle, Ser binds the PLP cofactor, and subsequent β -elimination and release of water forms an electrophilic amino-acrylate species (**Scheme 2-2a**). A highly conserved active-site glutamate stabilizes the accumulation of positive charge on the pyrrole ring of indole and helps facilitate the nucleophilic attack on the amino-acrylate to form a new C–C bond that produces Trp. The similarity of AzAla to Trp prompted us to investigate whether TrpB could accept azulene as a nucleophile in the place of indole. Unlike indole, azulene lacks heteroatoms that can help stabilize the accumulation of charge during nucleophilic attack. Despite this, azulene has a permanent dipole exemplified by its resonance structure of a cycloheptatrienyl cation (tropylium) fused to a cyclopentadienyl anion (Cp^-). We hypothesized that, analogous to indole, the buildup of electron density on the Cp^- could promote nucleophilic attack by azulene in the TrpB catalytic cycle, while the tropylium

system could stabilize the resulting positive charge (**Scheme 2-2b**). Aromatic ions are common moieties in synthetic chemistry¹⁵ and as enzyme inhibitors,¹⁶ but there are few reports of enzymes that can interact productively with aromatic ions in their catalytic cycles. We were therefore unsure if the active site of TrpB could accommodate or activate such a substrate, or if the reactivity of azulene would be sufficient for nucleophilic attack.



Scheme 2-2. Parallels between indole and azulene in the TrpB reaction.

2.2 Results

We first examined the conversion of azulene and Ser to AzAla using a small panel of previously engineered TrpB variants from the thermophilic organisms *Pyrococcus furiosus* (PfTrpB) and *Thermotoga maritima* (TmTrpB). The variants were selected to provide an efficient sampling of the engineered TrpB evolutionary lineage beginning from wild-type TrpS and ending with stand-alone TrpB variants evolved for activity with different indole and serine analogues. Nearly every enzyme we tested demonstrated significant activity for

this reaction, the exception being variants in which the highly conserved glutamate mentioned above was mutated to glycine.

The significant effect of the E104(105)G mutation suggested that this conserved catalytic residue may be playing an important role in the non-native azulene reaction. We explored this possibility by examining two engineered variants with and without the E104(105)G mutation: *Pf5G8*, which exhibits optimal activity at 75 °C,¹³ and *Tm9D8**, which exhibits optimal activity at lower temperatures such as 37 °C.¹⁷ Challenging the enzymes with indole demonstrated that this mutation only modestly decreases the rate of Trp formation (**Figure 2-1**), with an additional slight decrease in the chemoselectivity of the reaction that leads to formation of trace amounts of isoTrp (a product of the N-alkylation of indole, shown in Figure S2 and described previously¹³). In contrast, the E104(105)G mutation exerts a profound effect on AzAla production, practically abolishing all activity with azulene. Product was only detected in low amounts when the reaction was performed with high catalyst loadings overnight (0.1 mol %, 16 h; Appendix A, **Figure A-2**). We speculate that the glutamate residue stabilizes the tropylium cation to facilitate nucleophilic attack from Cp⁻. However, further mechanistic studies are required to elucidate the role of this mutation.

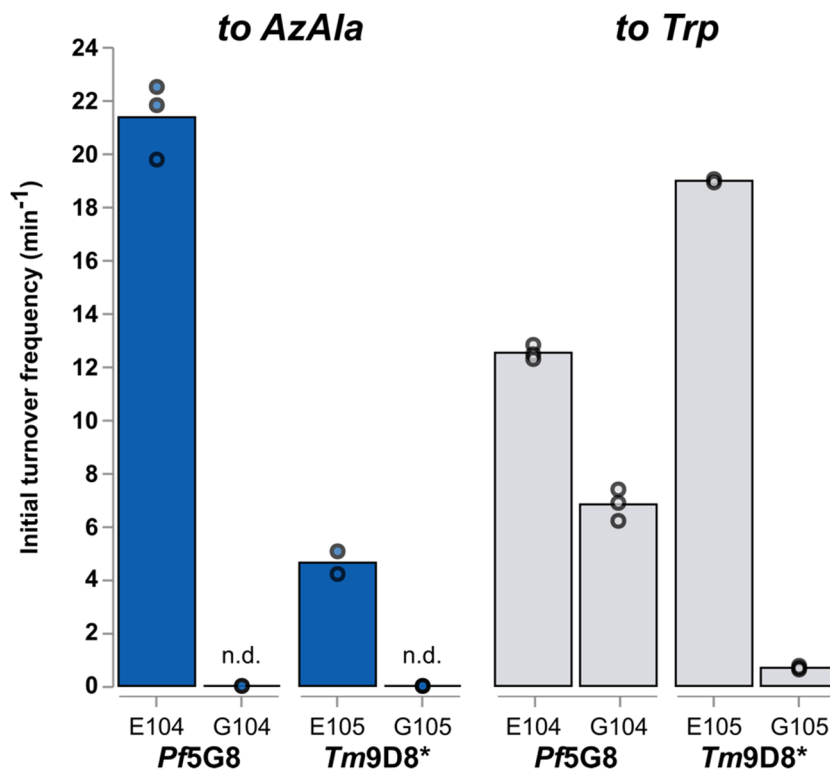


Figure 2-1. Effect of E104(105)G mutation on Trp and AzAla production. Mutation of a conserved glutamate residue affects the rate of AzAla production more significantly than that of native Trp production. Bars represent the average of two to three replicates, with replicates being shown as individual points. Reaction conditions can be found in Table A-1 n.d. = not detected by LC-MS.

Next, we wished to develop a biocatalytic method for the production of AzAla at scale. Although *Pf5G8* catalyzed the reaction roughly 4.5-fold faster at its optimum temperature of 75 °C than *Tm9D8** at 37 °C, azulene readily sublimates at high temperatures, making its containment at 75 °C difficult. We thus opted to use directed evolution to improve the activity of *Tm9D8** at 37 °C to create a stand-alone biocatalyst for the production of AzAla *in vivo* and *in vitro*.

A single round of random mutagenesis and screening identified two variants containing the mutations F184S and W286R. These mutations were combined to yield the final variant *TmAzul*, which had a three-fold improved rate of AzAla formation compared to *Tm9D8** (14.0 turnovers per minute; Appendix A, **Table A-1**). As we have not been successful in obtaining a crystal structure of *TmTrpS* and its variants, we constructed a homology model

of *Tm9D8** to try to understand the locations of each mutation and their possible effects (Figure 2-2). The W286R mutation sits on a flexible loop, where its role in catalysis is difficult to infer. F184S sits directly in the active site and is one of only a handful of residues whose side chains are in close proximity to the azulene substrate. This residue may interact directly with the substrate during catalysis or adjust the active site to be more accommodating for AzAla synthesis.

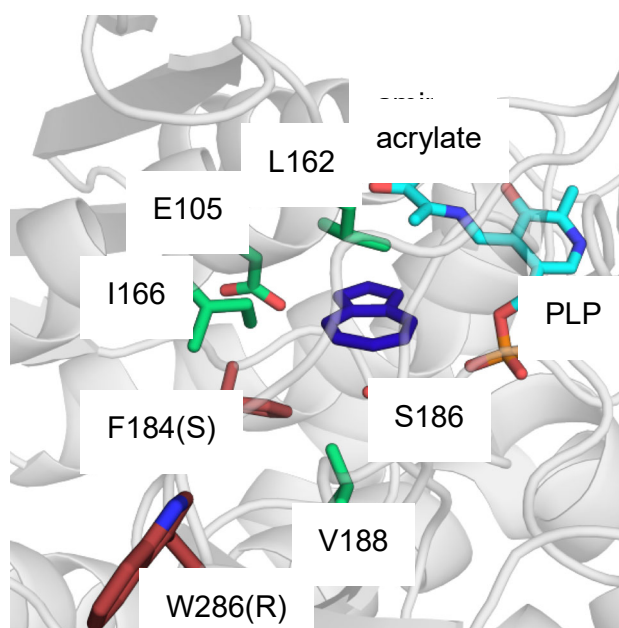


Figure 2-2. Homology model of *Tm9D8 with azulene in the active site.** Azulene (deep blue) is shown in a putative productive binding mode within a model of *Tm9D8** with the amino-acrylate intermediate (cyan) formed in the active site (see Appendix A, Section A.4 for model construction). Active-site residues in the native *TmTrpB* enzyme (green) and the mutations found in this study (red, mutation in parentheses) are shown as sticks.

Because the *Escherichia coli* expression cultures were heat treated at 75 °C for three hours prior to screening, *TmAzul* retains high thermostability, and highly pure enzyme can be obtained simply by heating the *E. coli* expression host at 75 °C for >1 h, pelleting the denatured *E. coli* proteins by centrifugation, and collecting the enzyme-bearing supernatant. To demonstrate scalability of this method, we synthesized AzAla on a gram scale using heat-treated lysate from a 1-L culture expressing the evolved *TmAzul* variant. We found that increasing the concentration of DMSO in the reaction increased the reaction rate at this scale,

presumably by keeping azulene from forming insoluble crystals in the reaction mixture due to the sparing solubility of azulene in aqueous buffer. We thus used 20% DMSO cosolvent, and the reaction progress was monitored by taking small aliquots of the reaction mixture and combining with an equal volume of ethyl acetate to observe the relative ratio of azulene to AzAla in the organic and aqueous layers, respectively. After 48 hours, the product was purified by removing any remaining azulene by extraction with ethyl acetate, removing the aqueous solvent *in vacuo*, and precipitating AzAla from the remaining DMSO cosolvent by the addition of excess ethyl acetate. The crude precipitate was collected by filtration and then further purified by reverse-phased column chromatography to afford 965 mg of pure AzAla (57% isolated yield). Crude azulene can be recovered from ethyl acetate by gently evaporating off the solvent under a constant stream of nitrogen and reused. The enantiopurity of the isolated AzAla product is >99% ee.

In conclusion, we have described a new-to-nature reaction catalyzed by tryptophan synthase and identified a conserved residue that is critical for this non-native reaction. Based on comparisons to the native reaction, we suggest that E104(105) stabilizes the aromatic ions in azulene to facilitate nucleophilic attack from Cp^- . To improve access to this useful ncAA, we have engineered a highly active enzyme catalyst that synthesizes enantiomerically pure AzAla from commercially available starting materials in a single step. Given that an engineered synthetase/tRNA pair has been reported for this ncAA, *TmAzul* also closes the gap for *in vivo* synthesis and incorporation of AzAla.¹⁰

Chapter II Bibliography

1. Liu, C. C. & Schultz, P. G. Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* **79**, 413–444 (2010).
2. Neumann, H. Rewiring translation - Genetic code expansion and its applications. *FEBS Lett.* **586**, 2057–2064 (2012).
3. Agostini, F. et al. Biocatalysis with Unnatural Amino Acids: Enzymology Meets Xenobiology. *Angew. Chemie - Int. Ed.* **56**, 9680–9703 (2017).

4. Gosavi, P. M., Moroz, Y. S., & Korendovych, I. V. β -(1-Azulenyl)-l-alanine-a functional probe for determination of pKa of histidine residues. *Chem. Commun.* **51**, 5347–5350 (2015).
5. Loidl, G. et al. Synthesis of β -(1-azulenyl)-L-alanine as a potential blue-colored fluorescent tryptophan analog and its use in peptide synthesis. *J. Pept. Sci.* **6**, 139–144 (2000).
6. Sartori, E. et al. An oligopeptide doubly labelled with an azulene chromophore and a TEMPO radical. Azulene triplet generation by enhanced ISC from S2. *Chem. Phys. Lett.* **385**, 362–367 (2004).
7. Venanzi, M. et al. Structural properties and photophysical behavior of conformationally constrained hexapeptides functionalized with a new fluorescent analog of tryptophan and a nitroxide radical quencher. *Biopolymers* **75**, 128–139 (2004).
8. Mazzuca, C. et al. Mechanism of membrane activity of the antibiotic trichogin GA IV: A two-state transition controlled by peptide concentration. *Biophys. J.* **88**, 3411–3421 (2005).
9. Moroz, Y. S., Binder, W., Nygren, P., Caputo, G. A., & Korendovych, I. V. Painting proteins blue: β -(1-azulenyl)-l-alanine as a probe for studying protein-protein interactions. *Chem. Commun.* **49**, 490–492 (2013).
10. Baumann, T. et al. Site-Resolved Observation of Vibrational Energy Transfer Using a Genetically Encoded Ultrafast Heater. *Angew. Chemie - Int. Ed.* **58**, 2899–2903 (2019).
11. Stempel, E., Kaml, R. F. X., Budisa, N., & Kalesse, M. Painting argyrins blue: Negishi cross-coupling for synthesis of deep-blue tryptophan analogue β -(1-azulenyl)-L-alanine and its incorporation into argyrin C. *Bioorganic Med. Chem.* **26**, 5259–5269 (2018).
12. Almhjell, P. J., Boville, C. E., & Arnold, F. H. Engineering enzymes for noncanonical amino acid synthesis. *Chem. Soc. Rev.* **47**, 8980–8997 (2018).
13. Romney, D. K., Murciano-Calles, J., Wehrmüller, J., & Arnold, F. H. Unlocking Reactivity of TrpB: A General Biocatalytic Platform for Synthesis of Tryptophan

- Analogues. *J. Am. Chem. Soc.* **139**, 10769–10776 (2017).
14. Buller, A. R. et al. Directed evolution of the tryptophan synthase β -subunit for stand-alone function recapitulates allosteric activation. *Proc. Natl. Acad. Sci.* **112**, 14599–14604 (2015).
 15. Komatsu, K. & Kitagawa, T. Cyclopropenylum cations, cyclopropenones, and heteroanalogues - Recent advances. *Chem. Rev.* **103**, 1371–1427 (2003).
 16. Himmel, D. M. et al. Structure of HIV-1 Reverse Transcriptase with the Inhibitor β -Thujaplicinol Bound at the RNase H Active Site. *Structure* **17**, 1625–1635 (2009).
 17. Boville, C. E., Romney, D. K., Almhjell, P. J., Sieben, M., & Arnold, F. H. Improved Synthesis of 4-Cyanotryptophan and Other Tryptophan Analogues in Aqueous Solvent Using Variants of TrpB from *Thermotoga maritima*. *J. Org. Chem.* **83**, 7447–7452 (2018).

Appendix A

SUPPLEMENTARY INFORMATION FOR CHAPTER II

A.1 Figures and tables

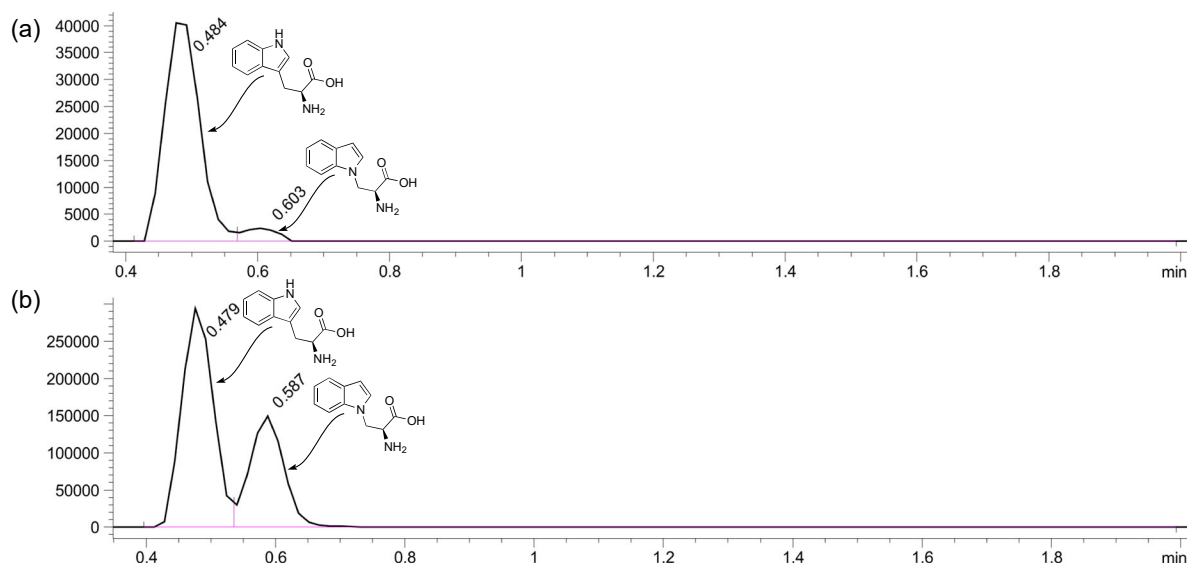


Figure A-1. Isotryptophan formation catalyzed by TrpB E104(105)G mutants. Traces shown as total ion count filtered by $[M+H]^+$ for Trp ($m/z = 205$). (a) A trace amount of isoTrp is formed by Pf5G8 E104G alongside Trp. (b) Tm9D8* E105G also produces isoTrp as the minor product. Reactions prepared according to **Section A.3.7**, with conditions given in Table S1.

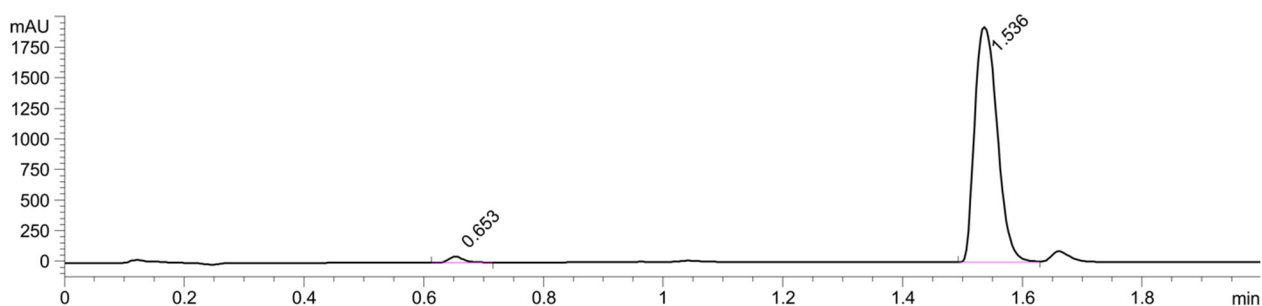


Figure A-2. Trace AzAla production by Tm9D8* E105G. After 16 hours at 37 °C with 0.1 mol % catalyst loading, only a small amount of AzAla (peak at 0.653 min) is formed by Tm9D8* E105G. The reaction was prepared according to **Section A.3.7**, but the reaction mixture was not diluted after centrifugation to increase the concentration of product in the sample. Trace AzAla yield is also seen for Pf5G8 E104G at 75 °C, but the sublimation of azulene interferes with analysis of the reaction at long timescales.

Table A-1. Rate comparisons among native and engineered TrpB enzymes^[a]

catalyst	initial turnover frequency (min ⁻¹)			
	to Trp		to AzAla	
	E104(105)	G104(105)	E104(105)	G104(105)
TmTrpS[b]	174 ± 1	—	12.0 ± 0.4	—
Pf5G8[b]	12.5 ± 0.3	6.8 ± 0.6	21.4 ± 1.4	n.d.
Tm9D8*[c]	19.0 ± 0.1	0.68 ± 0.1[d]	4.6 ± 0.6	n.d.
TmAzul[c]	—	—	14.0 ± 0.4	—

^[a]Reactions performed according to Section 3.7 with 0.02 mol % catalyst loading and allowed to react for 15 minutes (Trp production) or 1 hour (AzAla production). ^[b]Reactions run at 75 °C. ^[c]Reactions run at 37 °C. ^[d]Reaction time and catalyst loading increased to 2 hours and 0.1 mol % catalyst loading to improve quantification. n.d. = not detected by LC-MS.

A.2 Experimental procedures

A.2.1 General experimental methods

Chemicals and reagents were purchased from commercial sources and used without further purification. Proton and carbon NMR were recorded on a Bruker 400 MHz (100 MHz) spectrometer equipped with a cryogenic probe. Proton chemical shifts are reported in ppm (δ) relative to tetramethylsilane and calibrated using the residual solvent resonance (D₂O, δ 4.79 ppm). Data are reported as follows: chemical shift (multiplicity [singlet (s), doublet (d), doublet of doublets (dd), doublet of doublet of doublets (ddd), triplet (t), triplet of doublets (td), multiplet (m)], coupling constants [Hz], integration).

All cultures were grown in Terrific Broth supplemented with 100 μ g/mL carbenicillin (TB_{carb}). Cultures were shaken in the New Brunswick Innova 4000 shaker (shaking diameter 19 mm), with the exception of 96-well plates, which were shaken in the Multitron INFORS HT shaker (shaking diameter 50 mm). Lysis buffer was composed of 50 mM potassium phosphate, pH 8.0 (KPi buffer), supplemented with 100 μ M pyridoxal 5'-phosphate (PLP).

Lysis was performed in 75 °C water bath (Fisherbrand™ Isotemp™ Digital-Control Water Baths: Model 220) for >1 h. Reactions were performed in 50 mM KPi, pH 8.0, unless otherwise stated. High-resolution mass spectrometry (HRMS) was conducted on an Agilent 6200 TOF using electrospray ionization (ESI) to ionize the sample. Liquid chromatography/mass spectrometry (LCMS) was performed on an Agilent 1290 UPLC-LCMS equipped with a C-18 silica column (1.8 μm, 2.1 × 50 mm) using CH₃CN/H₂O (0.1% acetic acid by volume): 5% to 95% CH₃CN over 2 min; 1 mL/min.

A.2.2 Cloning, expression, and purification of TrpB variants

The genes encoding *TmTrpA*^{WT} (Uniprot P50908), *TmTrpB*^{WT} (Uniprot G4FDT2), *Pf5G8*, *Pf5G8 E104G*, *Tm9D8**, and *Tm9D8* E105G* were previously cloned into pET22b(+) with a C-terminal 6x His tag. Protein expression of the variants was carried out in *Escherichia coli* BL21(DE3) *E. coli* Express[®] cells (Lucigen) by inoculating 5 mL TB_{carb} with a single colony and incubating this pre-culture overnight at 37 °C and 230 rpm. For expression, 2.5 mL culture were used to inoculate 500 mL TB_{carb} in a 2-L flask and incubated at 37 °C and 130 rpm for 2.5 hours to reach OD₆₀₀ 0.6–0.8. Cultures were chilled on ice for 20–30 minutes, and protein expression was induced with a final concentration of 1 mM isopropyl β-D-thiogalactopyranoside (IPTG). Expression proceeded at 20 °C and 130 rpm for approximately 24 h. Cells were harvested by centrifugation at 10,000g for 20 minutes at 4 °C and the supernatant was decanted. The pellet was stored at –20 °C until further use.

For protein purification, cells were thawed and were resuspended in 4 mL lysis buffer/g pellet. Cells were heat-treated at 75 °C for >1 h. The supernatant was collected from clarified lysate following centrifugation for 20 min at 14,000g and 4 °C. Purification was performed with a 1-mL Ni-NTA gravity flow column at room temperature. Buffer A: 20 mM imidazole, 25 mM KPi buffer, Buffer B: 500 mM imidazole, 25 mM KPi buffer. The column was equilibrated with 10 column volumes (CV) Buffer A. Subsequently, heat-treated lysate was loaded onto column and washed with 10 CV Buffer A. Next, 10 CV 1:1 Buffer A: Buffer B were added to elute non-target proteins from the column. Protein was eluted with 3 CV

Buffer B. Proteins were dialyzed into 50 mM KPi buffer, flash frozen in liquid nitrogen, and stored at $-80\text{ }^{\circ}\text{C}$.

To obtain the purified *TmTrpS* complex, *TmTrpA*^{WT} was co-purified with *TmTrpB*^{WT}. Heat-treated lysate of both *TmTrpA*^{WT} and *TmTrpB*^{WT} were mixed together in a roughly 5:1 ratio (concentrations determined relative band intensities from SDS-PAGE analysis) and purification was followed as described above.

A.2.3 Construction of random mutagenesis libraries

Random mutagenesis libraries were generated with the *Tm9D8** gene as template by the addition of 200–400 μM MnCl_2 to a *Taq* (New England Biolabs) PCR reaction as previously reported.¹ PCR fragments were treated with *DpnI* for 1 h at $37\text{ }^{\circ}\text{C}$, purified by gel extraction, and then inserted into a pET22b(+) vector via Gibson assembly.² The Gibson assembly product was purified and concentrated using Zymo DNA Clean and Concentrate–5 kit (Catalog #: D4004). BL21(DE3). *E. coli* Express[®] cells were transformed with the Gibson assembly product. Libraries generated with 200, 300, and 400 μM MnCl_2 were tested (one 96-well plate, each) to determine which library gave the optimal balance of high diversity and low rate of inactivation. The chosen library was then tested further (see below).

A.2.4 Library expression and screening

Individual colonies were grown in 300 μL TB_{carb} in deep-well 96-well polypropylene plates and grown overnight at $37\text{ }^{\circ}\text{C}$, 250 rpm, 80% humidity. The following day, 20 μL overnight culture were used to inoculate 630 μL TB_{carb} cultures in deep-well 96-well plates and grown at $37\text{ }^{\circ}\text{C}$, 250 rpm. After 2.5 h, cultures were chilled on ice for 20–30 minutes and protein expression was induced upon addition of 50 μL IPTG (final conc. 1 mM) diluted in TB_{carb} . Cultures were shaken at $20\text{ }^{\circ}\text{C}$, 250 rpm for 20–24 h, after which they were subjected to centrifugation at 4,000g for 10 min. The cell pellets could be frozen at $-20\text{ }^{\circ}\text{C}$ until further use or used immediately.

Pellets were lysed in 300 μL lysis buffer and heat-treated lysate clarified by centrifugation at 4,000g for 10 min. To UV-transparent 96-well assay plates (Coplugs, catalog #:290-8120-

0AF) charged with 10 μL azulene dissolved in DMSO (final conc. 0.625 mM), 30 μL heat-treated lysate was transferred using Microlab NIMBUS96 liquid handler (Hamilton), followed by addition of 70 μL serine (final conc. 10 mM), and 90 μL 50 mM KPi buffer with a 12-channel pipet. Reactions were sealed with Microseal 'B' PCR plate sealing film (BioRad, catalog #: MSB1001) and incubated in a 37 $^{\circ}\text{C}$ water bath. Reaction progress was monitored by measuring absorption at 340 nm over the course of 24 h, in which more active variants retained AzAla in solution while the wells of inactive variants lost azulene due to absorption into the plastic assay plate and/or due to sublimation from solution. This can be seen below (**Figure A-3**), in a case where *Tm9D8** is not given Ser, resulting in no activity, as opposed to the case where the addition of Ser allows for AzAla formation and retention of signal at 340 nm:

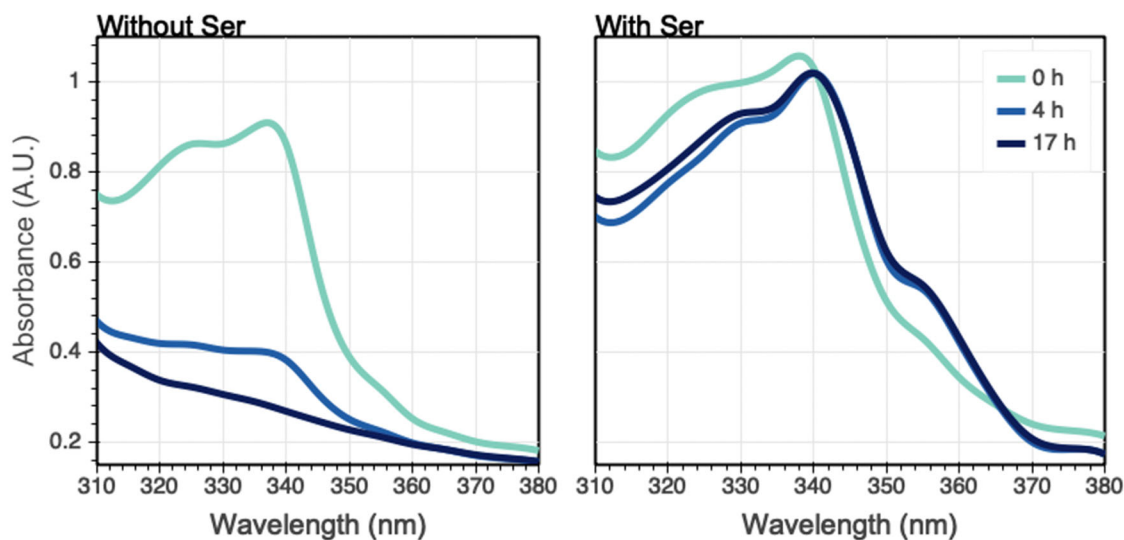


Figure A-3. Graphs of UV absorbance (A.U.) vs. wavelength for AzAla screen. Reactions (200 μL) were run in UV-transparent plates for 17 hours, incubated at 37 $^{\circ}\text{C}$. Lysate (30 μL) and 0.625 mM azulene were added to each well. No-serine control received 50 mM Kpi instead of 10 mM serine in 50 mM Kpi. Averaged data from two replicates of plate. Initially, samples have similar absorbance, but over time serine (-) control wells lose absorbance while serine (+) wells maintain signal.

A.2.5 Recombination of mutations

Site-directed mutagenesis was performed to combine mutations identified through screening to be beneficial for AzAla production. Gibson primers were designed to recombine the

F184S and W286R mutations. Using Tm9D8* W286R as template, the F184S mutation was introduced. Two PCRs were performed using Phusion polymerase (New England Biolabs), each fragment originating from the F184S site and ending in the center of the ampicillin resistance gene. The PCR fragments were treated with *DpnI* for 1 h at 37 °C, purified by gel extraction, and then combined via Gibson assembly.² The Gibson assembly product was purified using Zymo DNA Clean and Concentrate–5 kit (Catalog #: D4004). BL21(DE3) E. coli Express® cells were transformed with the Gibson assembly product.

Table A-2. Gibson assembly primers

Fragment	Size (Kb)	Forward primer (5' to 3')	Reverse primer (5' to 3')
1	4.8	CCTGCAGACCACCTATTACGTGTCGGCTCTGTG GTTGGTCCGCATCCATATCCG	CTGCCATAACCATGAGTGATAAACTGCGGCAAC TTACTTCTGACAACGATCG
2	1.7	CCAACCTACTTCTGACAACGATCGGAGGACCGA AGGAGCTAACCGCTTTTTTTCG	GCGTGACTGGATTACCAACCTGCAGACCACCTATT ACGTG

A.2.6 Small-scale analytical reactions

All analytical reactions were performed in 2-mL glass HPLC vials (Agilent) charged with 10 µL azulene or indole (final conc. 10 mM) dissolved in DMSO (5% v/v), followed by addition of serine (final conc. 10 mM) and purified enzyme diluted in 50 mM KPi buffer to a final volume of 200 µL. Reactions were incubated at 75 °C or 37 °C for 24 h. The reaction was then diluted with 800 µL of 1:1 CH₃CN/1 M aq. HCl and subjected to centrifugation at 20,000g. For indole, this reaction mixture was analyzed directly by UHPLC-MS at 277 nm, representing the isosbestic point between indole and tryptophan and allowing quantification of yield by comparing the substrate and product peak areas.³ For azulene, the reaction mixture was further diluted by adding 10 µL of this mixture into 190 µL of 1:1 CH₃CN/1 M aq. HCl and analyzed by UHPLC-MS. The yield was estimated comparing the integration of the product peak at 254 nm to a calibration curve (**Section A.3.8**).

A.2.7 Calibration for measuring HPLC yield of AzAla

Solutions of azulene and an authentic AzAla standard were made in 1:1 CH₃CN/1 M aq. HCl (total conc. 1 mM) and mixed in different ratios (9:1, 3:1, 1:1, 1:3, 1:9) in duplicate and analyzed by HPLC. The ratios of substrate to product peaks at 254 nm were correlated to the actual ratios by a linear relationship (**Figure A-4**).

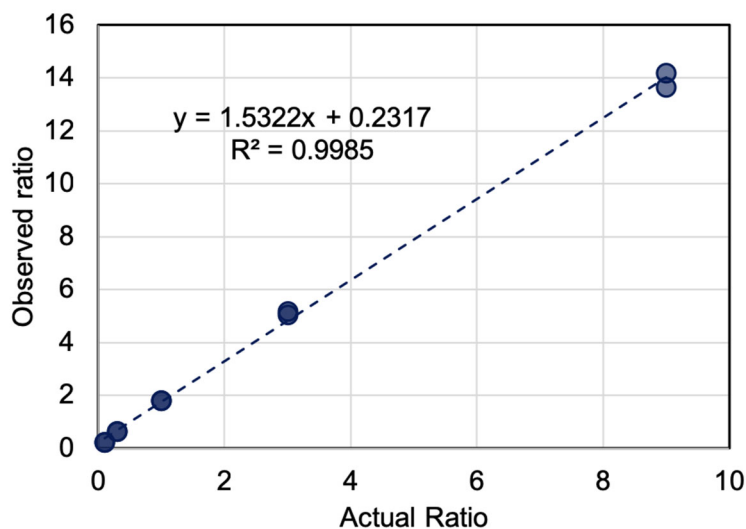


Figure A-4. AzAla calibration curve.

A.2.8 Large-scale preparation of AzAla

Two 500 mL cultures of BL21(DE3) cells expressing *TmAzul* were grown according to expression conditions described in Section 3.2. The expression cultures were centrifuged for 20 min and 14,000g. The pellets (14 g) were resuspended in 50 mL lysis buffer (100 μ M PLP, 50 mM KPi buffer) and lysed at 75 °C for >1 h. Heat-treated lysate was centrifuged 20 min, 14,000g, and the lysate decanted into a fresh container.

To a 500-mL Erlenmeyer flask azulene (7.8 mmol, 1.0 g) and serine (8.6 mmol, 0.90 g), 30 mL DMSO (20% v/v), 70 mL 50 mM KPi buffer, and 50-mL heat-treated lysate were added. The reaction was covered in aluminum foil, placed in 37 °C incubator, and shaken at 180 rpm. After 48 h, the reaction was washed with ethyl acetate to remove any remaining azulene, allowing crude starting material to be recovered from ethyl acetate by gently evaporating off the solvent under a constant stream of nitrogen. The aqueous layer was concentrated *in vacuo* and resulting in a dense paste comprised of DMSO, AzAla, and buffer salts. Ethyl acetate was added to remove DMSO and precipitate the amino acid product. The blue solid was filtered and washed with ethyl acetate. At this stage, AzAla is relatively pure by NMR and can be used for further applications. For highly pure AzAla, the blue solid was solubilized in 1 M HCl and immediately purified on a Biotage Isolera One purification system, using C-18

silica as the stationary phase, CH₃CN (0.1% acetic acid) as the strong solvent, and H₂O (0.1% acetic acid) as the weak solvent. AzAla should not spend long durations of time in strongly acidic conditions (such as 1 M HCl) as the product can decompose over time. The purified product fractions were combined and concentrated *in vacuo* to afford pure AzAla (blue solid, 0.965 g, 57% yield).

A.3 Variant sequences

Variants identified through screening were DNA sequenced using Sanger Sequencing (Laragen) to determine their identities. The DNA sequences of all TrpB genes tested in this project are included here. All variants were cloned into a pET22b(+) vector as described above. T7 and T7-terminator primers were used to sequence all variants.

Table A-3. TrpB sequencing primers

Primer	Direction	Sequence
T7	Forward (5' to 3')	taatacgactcactataggg
T7-term	Reverse (3' to 5')	ctagttattgctcagegggt

T_mTrpA^{WT}:

ATGAAAAGGTTTTATCGCGTACATCCCGGCTGGTTTTCCGGATCTGGAAAACCACCCGTAAAAATTCTGATCGCACTGAACGA
GCTGGTATTACCGGTGTTGAAATTGGTGTCCCGTTCCTCCGACCCGGTTGCGGATGGTCCGGTGATCCAACCTGGCGCATAG
CGTTGCTCTGCGTAACGGTGTGACTATTAATAAAAAATTCTGGAAATGCTGTCCGAGATTTCCGTAGATTACGACCTGTACCT
GATGTCTTACCTGAACCCGATCGTTAATTACCCTGAAGCAAAAGAGAAACTGCTGGACGAACCTGAAGAAGCTGGGCGTTA
AAGGCCTGATTATCCCAGACCTGCCGCTGCGTGAAGTAAAAAACGTTGACATCGCTTACCCGATCGTTCCATTTCGTTGCAC
CGAATACCAAAGACGAAGAGATCGACCTGATCAACTCCGTGCAGGCTCCGTTCTGCTACTATACTCTCGTTACGGTGTGA
ACTGGTGAACGCGAAGACCTGCCGTTTGCAGATCACATCAAACGCGTGAAAGAACGTATCAAACCTGCCACTGTTCCGTCGG
TTTTCGGTATCTCCCGTCACGAACAAGTTAAAAAGTTTTGGGAAATCGCTGATGGTGTATTGTTGGCAGCGCACTGGTCCG
CATCATGGAAGAAAACCCGAAAGATGAGATCCACGTAAGTTGTTGAAAAAGTTAAAGAGCTGCTGGGCAAAAtga

T_mTrpB^{WT}:

ATGAAAGCTACTTCCGGTCCGTACGGTGGCCAGTACGTGCCGAAATCCTGATGCCAGCTCTGGAAGAAGCTGGAAGCTGC
GTACGAAAGAAATCATGAAAGATGAGTCTTTCTGGAAAAGAAATCAATGACCTGCTGCGGATTATGCGGGTCTCCGACTC
CGCTGACTTCCGACGTCGTCTGTCCGAAAAATACGGTGTCTGCATCTATCTGAAACGTGAAGACCTGCTGCATACTGGTG
CGCATAAAATCAATAACGCTATCGGCCAGGTTCTGCTGGCAAAAAAATGGGCAAAACCCGTATCATTGCTGAAACGGGT
GCTGGTCAGCACGGCGTAGCAACTGCTACCGCAGCAGCGCTGTTCCGGTATGGAATGTGTAATCTATATGGGCGAAGAAGA
CACGATCCGCCAGAAAACCGAACGTTGAACGTATGAAACTGTGGGTGCTGAAAGTTGTACCCGTAATAATCCGGTAGCCGTA
CCCTGAAAGACGCAATTAACGAAGCTCTGCGTACTGGATTACCAACCTGCAGACCACCTATTACGTGATCGGCTCTGTG
GTTGGTCCGCATCCATATCCGATTATCGTACGTAACCTCCAAAAGTTATCGGCGAAGAGACCAAAAAACAGATTCTGGA
AAAAGAAGGCCGCTCTGCCGACTACATCGTTGCGTGGTGGTGGTCTAACGCTGCCGATCTTCTATCCGTTTTAT
CGATTCTGGTGTGAAAGCTGATCGGCGTAGAAGCCGGTGGCGAAGGTCTGAAACCCGTAATAATCCGGTAGCCGTA
AAGGTAATAATCGGCTACTGCACGGTTCTAAGACGTTCTGCTGAGGATGACTGGGGTCAAGTTCAGGTGACGCACTCC
GTCTCCGCTGGCCTGGACTACTCCGGTGTCCGGTCCGGAACACGCTATTGGCGTGAGACCGGTAAGTGTGCTGACGATGC
TGTGACCGATGAAGAAGCTCTGGACGCATTATCGAACTGTCTCGCTGGAAGGCATATCCAGCCCTGGAGTCTTCTC
ACGCACTGGCTTATCTGAAGAAGATCAACATCAAGGTAAAGTTGTTGGTGGTTAATCTGCTGGTCTGGTGTGACAAGGAT
CTGGAATCTGTAAGTGAACCCCGTATGTTCCGCAACGCATCCGCCctgagcaccaccatcaccatcactga

Pf5G8:

ATGTGGTTCCGGTGAATTTGGTGGTCAGTACGTGCCAGAAACGCTGGTTGGACCCCTGAAAGAGCTGGAAAAAGCTTACAA
 ACGTTTCAAAGATGACGAAGAATTCATCGTCAGCTGAATTACTACCTGAAAACCTGGGCAGGTCGTCCAACCCCACTGT
 ACTACGCAAAAACGCCCTGACTGAAAAAATCGGTGGTGCTAAAGTCTACCTGAAAACGTGAAGACCTGGTTCACGGTGGTGCA
 CACAAGACCAACAACGCCATCGGTCAGGCACTGCTGGCAAAGCTCATGGGTAAAACCTGCTGATCGCTGAGACCCGGTG
 CTGGTCAGCACGGCGTAGCGACTGCAATGGCTGGTGCCTGCTGGGCATGAAAGTGGACATTTACATGGGTGCTGAGGAC
 GTAGAACGTCAGAAATGAACGTATTCCGTATGAAGCTGCTGGGTGCAAACGTAATTCCAGTTAACTCCGGTTCTCGCAC
 CCTGAAAAGACGCAATCGACGAGGCTCTGCGTGATTGGGTGGTACTTTTGAATACACCCACTACCTAATCGGTTCCGTGGT
 CGGTCCACATCCGTATCCGACCATCGTTCGTGATTTTCAGTCTGTTATCGGTCTGAGGGCTAAAGCGCAGATCCCGGAGGC
 TGAAGGTCAGCTGCCAGATGTAATCGTTGCTTGTGTTGGTGGTGGCTAACCGCATGGGTATCTTTTACCCGTTCTGTGAA
 CGACAAAAAAGTTAAGCTGGTTGGCGTTGAGGCTGGTGGTAAAGGCCTGGAATCTGGTAAGCATTCGGCTAGCCTGAACG
 CAGGTCAGGTTGGTGTGTTCCATGGCATGCTGCTACTTTCTGCAGGACGAAGAAGGTCAGATCAAACCAAGCCACTCC
 ATCGCACCAAGGTTGATTATCCAGGTGTTGGTCCAGAACACGCTTACCTGAAAAAATTCAGCGTGTGAAATACGTGGC
 TGTAACCGATGAAGAAGCACTGAAAGCGTTCATGAACTGAGCCGTACCGAAGGTATCATCCAGCTCTGGAATCTGCGC
 ATGCTGTGGCTTACGCTATGAAACTGGCTAAGGAAAATGTCTCGTGATGAGATCATCATCGTAAACCTGTCTGGTCTGGTG
 ACAAAGACCTGGATATTGTCTGAAAGCGTCTGGCAACGTGCTcgagcaccaccaccaccactga

Pf5G8 E104G

ATGTGGTTCCGGTGAATTTGGTGGTCAGTACGTGCCAGAAACGCTGGTTGGACCCCTGAAAGAGCTGGAAAAAGCTTACAA
 ACGTTTCAAAGATGACGAAGAATTCATCGTCAGCTGAATTACTACCTGAAAACCTGGGCAGGTCGTCCAACCCCACTGT
 ACTACGCAAAAACGCCCTGACTGAAAAAATCGGTGGTGCTAAAGTCTACCTGAAAACGTGAAGACCTGGTTCACGGTGGTGCA
 CACAAGACCAACAACGCCATCGGTCAGGCACTGCTGGCAAAGCTCATGGGTAAAACCTGCTGATCGCTGGGACCCGGTG
 CTGGTCAGCACGGCGTAGCGACTGCAATGGCTGGTGCCTGCTGGGCATGAAAGTGGACATTTACATGGGTGCTGAGGAC
 GTAGAACGTCAGAAATGAACGTATTCCGTATGAAGCTGCTGGGTGCAAACGTAATTCCAGTTAACTCCGGTTCTCGCAC
 CCTGAAAAGACGCAATCGACGAGGCTCTGCGTGATTGGGTGGTACTTTTGAATACACCCACTACCTAATCGGTTCCGTGGT
 CGGTCCACATCCGTATCCGACCATCGTTCGTGATTTTCAGTCTGTTATCGGTCTGAGGGCTAAAGCGCAGATCCCGGAGGC
 TGAAGGTCAGCTGCCAGATGTAATCGTTGCTTGTGTTGGTGGTGGCTAACCGCATGGGTATCTTTTACCCGTTCTGTGAA
 CGACAAAAAAGTTAAGCTGGTTGGCGTTGAGGCTGGTGGTAAAGGCCTGGAATCTGGTAAGCATTCGGCTAGCCTGAACG
 CAGGTCAGGTTGGTGTGTTCCATGGCATGCTGCTACTTTCTGCAGGACGAAGAAGGTCAGATCAAACCAAGCCACTCC
 ATCGCACCAAGGTTGATTATCCAGGTGTTGGTCCAGAACACGCTTACCTGAAAAAATTCAGCGTGTGAAATACGTGGC
 TGTAACCGATGAAGAAGCACTGAAAGCGTTCATGAACTGAGCCGTACCGAAGGTATCATCCAGCTCTGGAATCTGCGC
 ATGCTGTGGCTTACGCTATGAAACTGGCTAAGGAAAATGTCTCGTGATGAGATCATCATCGTAAACCTGTCTGGTCTGGTG
 ACAAAGACCTGGATATTGTCTGAAAGCGTCTGGCAACGTGCTcgagcaccaccaccaccactga

Tm9D8*:

ATGAAAGGCTACTTCCGGTCCGTACGGTGGCCAGTACGTGCCGAAATCCTGATGGGAGCTCTGGAAGAAGCTGGAAGCTGC
 GTACGAAGGAATCATGAAAGATGAGTCTTCTGGAAGAATTCATGACCTGCTGCGCGATTATGCGGGTCTCCGACTC
 CGCTGTACTTCGCACGTCGTCTGTCCGAAAAATACGGTGTCTCGGTATATCTGAAACGTGAAGACCTGCTGCATACTGGTG
 CGCATAAAATCAATAACGCTATCGGCCAGGTTCTGCTGGCAAACTAATGGGCAAAACCCGTATCATTGCTGAAACGGGT
 GCTGGTACGACCGGCTAGCAACTGCTACCGCAGCAGCGCTGTTCCGGTATGGAATGTGTAATCTATATGGGCGAAGAAGA
 CACGATCCGCCAGAACTAAACGTTGAACGTATGAACTGCTGGGTGCTAAAGTTGTACCGGTAATAATCCGGTAGCCGTA
 CCCTGAAAAGACGCAATTGACGAAGCTCTGCGTGACTGGATTACCAACCTGCAGACCACCTATTACGTGTTCCGGTCTGTG
 GTTGGTCCGCATCCATATCCGATTATCGTACGTAACCTCCAAAAGGTTATCGGCGAAGAGACCAAAAAACAGATTCCAGA
 AAAAGAAGGCGCTCTGCCGACTACATCGTTGCGTGCGTGAGCGGTGGTTCTAACGCTGCCGGTATCTCTATCCGTTTAT
 CGATTCTGGTGTGAAGCTGATCGGCGTAGAAGCCGGTGGCGAAGGTTCTGGAACCGGTAACATCGCGCTTCTCTGTCTGA
 AAGGTAATAATCGGCTACCTGCACGGTTCGTAAGACGTTCTGTTCTGCAGGATGACTGGGGTCAAGTTCAAGTGAGCCACTC
 GTCTCCGCTGGCCTGGACTACTCCGGTGTCCGGTCCGGAACACGCTATTGGCGTGAGACCGGTAAGTGTGTACGATGC
 TGTGACCGATGAAGAAGCTCTGGACGATTCATCGAACTGTCTCGCTGGAAGGCATCATCCAGCCCTGGAGTCTTCTC
 ACGCACTGGCTTATCTGAAGAAGATCAACATCAAGGGTAAAGTTGTGGTGGTTAATCTGTCTGGTCTGGTGACAAGGAT
 CTGGAATCTGTACTGAACCACCCGTATGTTCCGCAACGCATCCGCCTcgagcaccaccaccaccactga

Tm9D8* E105G:

ATGAAAGGCTACTTCCGGTCCGTACGGTGGCCAGTACGTGCCGAAATCCTGATGGGAGCTCTGGAAGAAGCTGGAAGCTGC
 GTACGAAGGAATCATGAAAGATGAGTCTTCTGGAAGAATTCATGACCTGCTGCGCGATTATGCGGGTCTCCGACTC
 CGCTGTACTTCGCACGTCGTCTGTCCGAAAAATACGGTGTCTCGGTATATCTGAAACGTGAAGACCTGCTGCATACTGGTG
 CGCATAAAATCAATAACGCTATCGGCCAGGTTCTGCTGGCAAACTAATGGGCAAAACCCGTATCATTGCTGAAACGGGT
 GCTGGTACGACCGGCTAGCAACTGCTACCGCAGCAGCGCTGTTCCGGTATGGAATGTGTAATCTATATGGGCGAAGAAGA
 CACGATCCGCCAGAACTAAACGTTGAACGTATGAACTGCTGGGTGCTAAAGTTGTACCGGTAATAATCCGGTAGCCGTA
 CCCTGAAAAGACGCAATTGACGAAGCTCTGCGTGACTGGATTACCAACCTGCAGACCACCTATTACGTGTTCCGGTCTGTG
 GTTGGTCCGCATCCATATCCGATTATCGTACGTAACCTCCAAAAGGTTATCGGCGAAGAGACCAAAAAACAGATTCCAGA
 AAAAGAAGGCGCTCTGCCGACTACATCGTTGCGTGCGTGAGCGGTGGTTCTAACGCTGCCGGTATCTCTATCCGTTTAT
 CGATTCTGGTGTGAAGCTGATCGGCGTAGAAGCCGGTGGCGAAGGTTCTGGAACCGGTAACATCGCGCTTCTCTGTCTGA
 AAGGTAATAATCGGCTACCTGCACGGTTCGTAAGACGTTCTGTTCTGCAGGATGACTGGGGTCAAGTTCAAGTGAGCCACTC
 GTCTCCGCTGGCCTGGACTACTCCGGTGTCCGGTCCGGAACACGCTATTGGCGTGAGACCGGTAAGTGTGTACGATGC
 TGTGACCGATGAAGAAGCTCTGGACGATTCATCGAACTGTCTCGCTGGAAGGCATCATCCAGCCCTGGAGTCTTCTC
 ACGCACTGGCTTATCTGAAGAAGATCAACATCAAGGGTAAAGTTGTGGTGGTTAATCTGTCTGGTCTGGTGACAAGGAT
 CTGGAATCTGTACTGAACCACCCGTATGTTCCGCAACGCATCCGCCTcgagcaccaccaccaccactga

AAAAGAAGGCCGTCTGCCGACTACATCGTTGCGTGCCTGAGCGGTGGTTCTAACGCTGCCGGTATCTTCTATCCGTTTAT
 CGATTCTGGTGTGAAGCTGATCGGCGTAGAAGCCGGTGGCGAAGGTCTGGAAACCGGTAAACATGCGGCTTCTCTGCTGA
 AAGGTAATAATCGGCTACCTGCACGGTTCTAAGACGTTTCGTTCTGCAGGATGACTGGGGTCAAGTTCAGGTGAGCCACTCC
 GTCTCCGCTGGCCTGGACTACTCCGGTGTCCGGTCCGGAACACGCCTATTGGCGTGAGACCGGTAAAGTGTGTACGATGC
 TGTGACCGATGAAGAAGCTCTGGACGCATTTCGAACTGTCTCGCCTGGAAGGCATCATCCAGCCCTGGAGTCTTCTC
 ACGCACTGGCTTATCTGAAGAAGATCAACATCAAGGTAAAGTTGTGGTGGTTAATCTGTCTGGTCTGGTGGTGAACAAGGAT
 CTGGAATCTGTACTGAACCACCCGTATGTTTCGCGAACGCATCCGCCctgagcaccaccaccaccactga

TmAzul:

ATGAAAAGGCTACTTCGGTCCGTACGGTGGCCAGTACGTGCCGGAATCCTGATGGGAGCTCTGGAAGAACTGGAAGCTGC
 GTACGAAGGAATCATGAAAGATGAGTCTTTCTGGAAAGAATTCAATGACCTGCTGCGGATTATGCGGGTCTGCCGACTC
 CGCTGACTTCCGACGTCGCTGTCCGAAAAATACGGTGTCTCGCTATATCTGAAACGTGAAGACCTGCTGCATACTGGTG
 CGCATAAAAATCAATAACGCTATCGGCCAGGTTCTGCTGGCAAAACTAATGGGCAAAAACCCGTATCATTGCTGAAACGGGT
 GCTGGTCAGCACGGCGTAGCAACTGCTACCGCAGCAGCGCTGTTCCGGTATGGAATGTGTAATCTATATGGGCGAAGAAGA
 CACGATCCGCCAGAAAATAAACGTTGAACGTATGAAACTGCTGGGTGCTAAAGTTGTACCGGTAAAAATCCGGTAGCCGTA
 CCCTGAAAGACGCAATTGACGAAGCTCTGCGTACTGGATTACCAACCTGCAGACCACCTATTACGTGTCGGCTCTGTG
 GTTGGTCCGCATCCATATCCGATTATCGTACGTAACCTCCAAAAGGTTATCGGCGAAGAGACCAAAAAACAGATTCCAGA
 AAAAGAAGCCGCTCTGCCGACTACATCGTTGCGTGCCTGAGCGGTGGTTCTAACGCTGCCGGTATCTTCTATCCGTTTAT
 CGATTCTGGTGTGAAGCTGATCGGCGTAGAAGCCGGTGGCGAAGGTCTGGAAACCGGTAAACATGCGGCTTCTCTGCTGA
 AAGGTAATAATCGGCTACCTGCACGGTTCTAAGACGTTTCGTTCTGCAGGATGACAGGGGTCAAGTTCAGGTGAGCCACTCC
 GTCTCCGCTGGCCTGGACTACTCCGGTGTCCGGTCCGGAACACGCCTATTGGCGTGAGACCGGTAAAGTGTGTACGATGC
 TGTGACCGATGAAGAAGCTCTGGACGCATTTCGAACTGTCTCGCCTGGAAGGCATCATCCAGCCCTGGAGTCTTCTC
 ACGCACTGGCTTATCTGAAGAAGATCAACATCAAGGTAAAGTTGTGGTGGTTAATCTGTCTGGTCTGGTGGTGAACAAGGAT
 CTGGAATCTGTACTGAACCACCCGTATGTTTCGCGAACGCATCCGCCctgagcaccaccaccaccactga

A.4 Structural modeling

A crystal structure of a homologous TrpS complex (*Salmonella typhimurium*, *StTrpS*) has been reported with the amino-acrylate in the β -subunit stabilized in the active site by benzimidazole, a competitive inhibitor (PDB: 4HPX).⁴ This TrpB structure served as the template for construction of a homology model of *Tm9D8** (57% sequence identity) using the SWISS-MODEL program.⁵ The alignment of *Tm9D8** homology model and the *StTrpB* structure allowed benzimidazole and the amino-acrylate to be placed within the *Tm9D8** structure in PyMOL, and the five-membered ring of azulene was aligned directly to the five-membered ring of benzimidazole to simulate a productive binding pose.

A.5 Characterization of AzAla

The enzymatic AzAla product was characterized by NMR, HRMS, and chiral derivatization. The spectra were taken in a mixture of D₂O/5% DCl to increase solubility, resulting in deuterium exchange of one of the protons on the five-membered rings. ¹H NMR (400 MHz, D₂O) δ 7.85 (dd, $J = 10.9, 9.6$ Hz, 2H), 7.29 (s, 1H), 7.16 (t, $J = 9.9$ Hz, 1H), 6.72 (dt, $J = 11.7, 9.8$ Hz, 2H), 3.89 (dd, $J = 7.1, 5.8$ Hz, 1H), 3.23 (dd, $J = 15.3, 5.9$ Hz, 1H), 3.15 (dd, $J = 15.3, 7.1$ Hz, 1H). ¹³C NMR (100 MHz, D₂O) δ 172.35, 140.79, 138.75, 137.59, 137.54,

137.48, 136.66, 133.71, 123.74, 123.09, 120.83, 54.63, 27.64. HRMS (m/z) for $[M+H]^+$ $C_{13}H_{14}NO_2$ requires 216.1019, observed 216.1019.

Enantiopurity was determined by derivatization with enantiopure and racemic FDNP-alanamide. In a 2-mL vial, AzAla (0.5 μ mol) was dissolved in 1 M aq. $NaHCO_3$ (100 μ L), to which 10 μ L of a 33-mM FDNP-alanamide solution in acetone (0.33 μ mol) was added. The vial was shaken for 2 h at 230 rpm, 37 $^{\circ}C$. The reaction was allowed to cool to room temperature, then diluted with 1:1 CH_3CN /1 M aq. HCl (600 μ L). The solution was analyzed via LC-MS (40% to 95% CH_3CN , monitored by using total ion count filtered for the expected mass of 468). Absolute stereochemistry for AzAla was inferred by analogy to L-tryptophan and determined to have >99% enantiomeric excess.

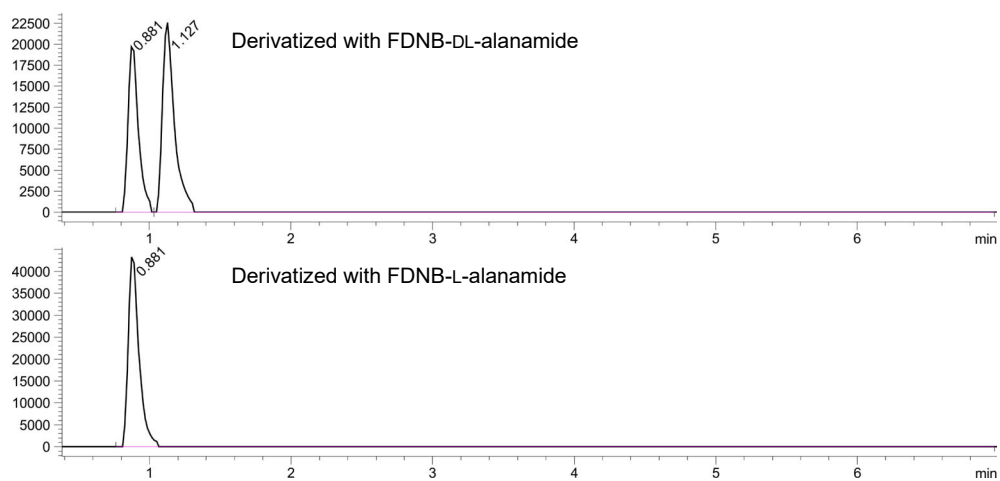
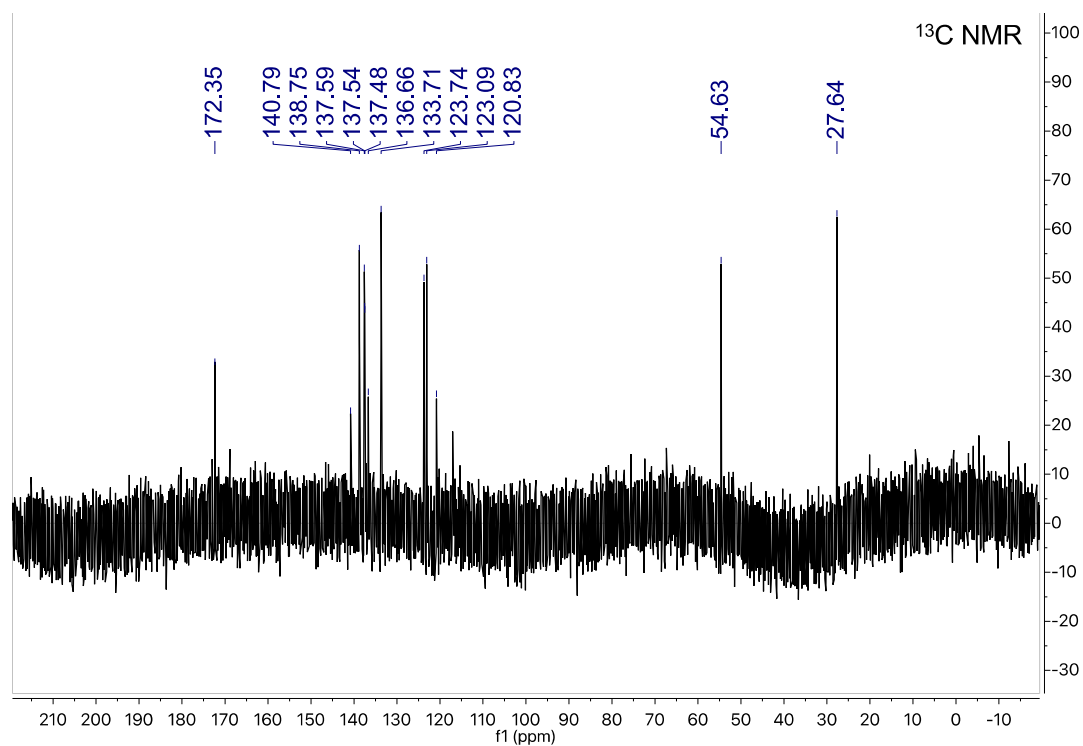
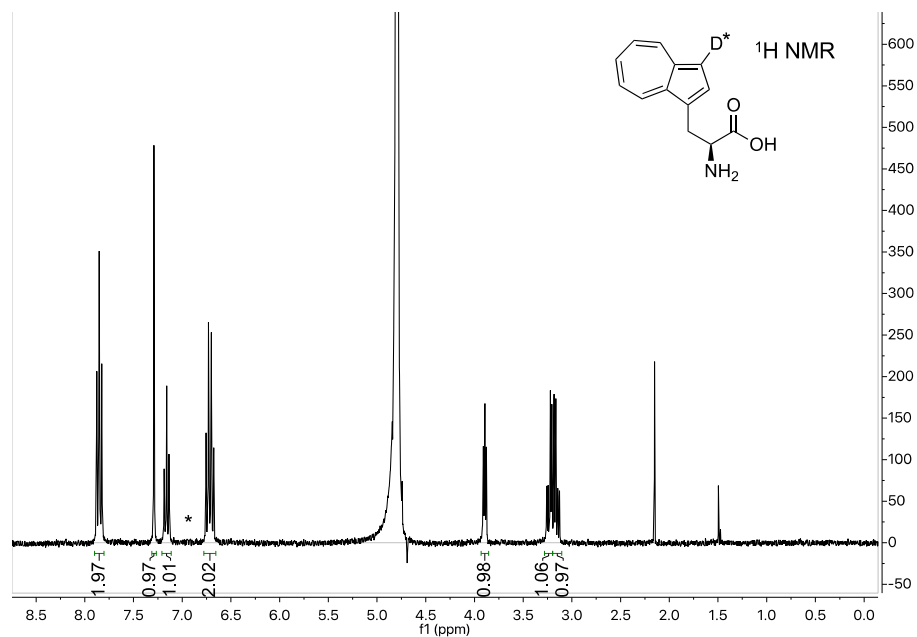


Figure A-5. HPLC traces of AzAla enantiopurity experiment.

A.6 NMR spectra



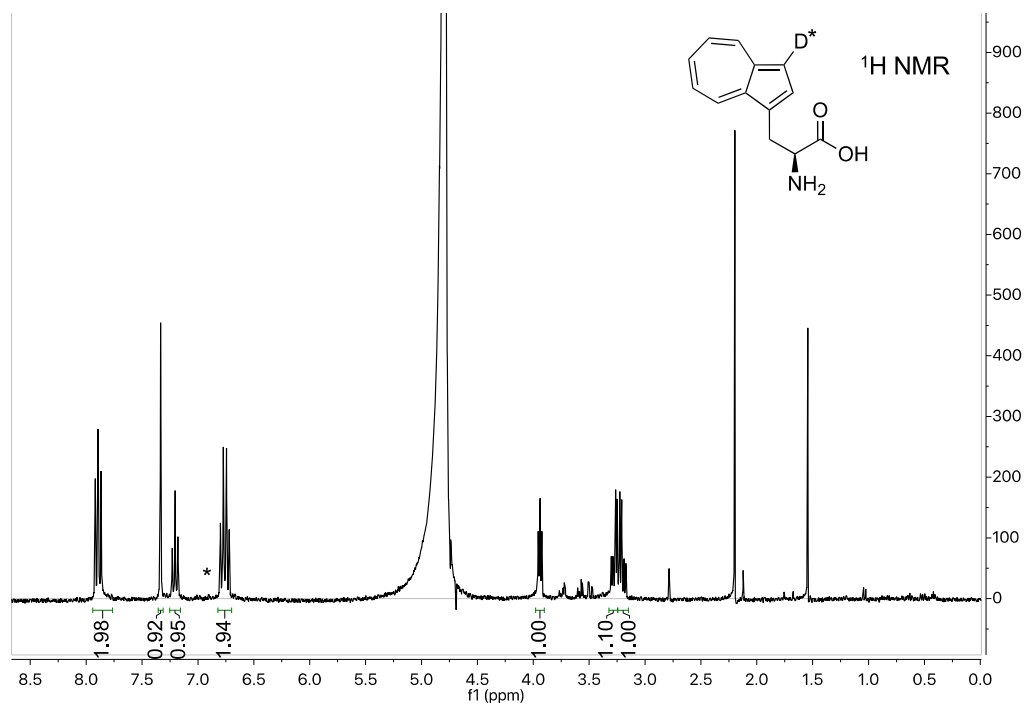


Figure A-8. ^1H NMR of crude AzAla. *Acidic conditions in deuterated solvents result in the exchange of one of the Cp- protons (6.9 ppm) for deuterium.

Bibliography for Appendix A

1. Boville, C. E. et al. Engineered Biosynthesis of β -Alkyl Tryptophan Analogs. *Angew. Chemie Int. Ed.* (2018). doi:10.1002/anie.201807998
2. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
3. Romney, D. K., Murciano-Calles, J., Wehrmüller, J., & Arnold, F. H. Unlocking Reactivity of TrpB: A General Biocatalytic Platform for Synthesis of Tryptophan Analogues. *J. Am. Chem. Soc* **139**, 10769–10776 (2017).
4. Niks, D. et al. Allosterity and substrate channeling in the tryptophan synthase bienzyme complex: Evidence for two subunit conformations and four quaternary states. *Biochemistry* **52**, 6396–6411 (2013).
5. Waterhouse, A. et al. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).

ASYMMETRIC ALKYLATION OF KETONES CATALYZED BY ENGINEERED TRPB

Material from this chapter appears in: **Watkins-Dulaney E. J.**, Dunham N. P., Straathof S., Turi S., Arnold F. H., Buller A. R. Asymmetric Alkylation of Ketones Catalyzed by Engineered TrpB. Manuscript in progress.

E.J.W.D. participated in the conception, design, execution of the research, and preparation of the manuscript. N.P.D. participated in the design and analysis of mechanistic experiments and preparation of the manuscript. S.S. participated in the execution of round 4 of evolution and found 2-fluoroacetophenone activity. S.T. participated in initial condition optimization.

ABSTRACT

The β -subunit of tryptophan synthase (TrpB) catalyzes a PLP-mediated β -substitution reaction between indole and serine to form L-Trp. A succession of TrpB protein engineering campaigns to expand the enzyme's nucleophile substrate range has enabled the biocatalytic production of diverse noncanonical amino acids (ncAAs). Here, we show that ketone-derived enolates can serve as nucleophiles in the TrpB reaction to achieve the asymmetric alkylation of ketones, an outstanding challenge in synthetic chemistry. We engineered TrpB by directed evolution to catalyze the asymmetric alkylation of propiophenone and 2-fluoroacetophenone with a high degree of selectivity. In reactions with propiophenone, preference for the opposite product diastereomer emerges over the course of evolution, demonstrating that full control over the stereochemistry at the new chiral center can be achieved. The addition of this new reaction to the TrpB platform is a crucial first step toward the development of efficient methods to synthesize noncanonical prolines and other chirally dense nitrogen heterocycles.

3.1 Introduction

The formation of carbon–carbon (C–C) bonds in a controllable and selective manner is central to the construction of the diverse organic scaffolds upon which biological systems depend. While the fundamental challenges of C–C bond formation are constant, the ensuing solutions put forth by Nature and the synthetic chemist are distinct yet complementary. Nature has met these challenges with a relatively modest set of C–C bond-forming enzymes that operate under mild conditions and exhibit astonishing control over chemo-, regio-, and stereo-selectivities.^{1,2} By contrast, chemists have accumulated an arsenal of creative methods, many of which boast impressively large scopes of useful substrates. To overcome deficiencies in selectivity, however, chemists often employ precious-metal catalysts and complex ligands, and rely on addition and removal of protecting groups, which can quickly result in inefficient, multistep syntheses. Furthermore, most of these reactions are not compatible with aqueous solutions, and instead require harsh conditions and environmentally hostile solvents. New catalytic methods for C–C bond formation that draw from the vast array of nucleophilic substrates in the synthetic literature while leveraging the renowned selectivity of enzymes would be highly useful.

The tryptophan synthase β -subunit (TrpB) is a pyridoxal phosphate (PLP)-dependent, C–C bond-forming enzyme that catalyzes the last stage in tryptophan biosynthesis. Natively, TrpB is bound to, and allosterically activated by, the tryptophan synthase α -subunit (TrpA), which cleaves indole-3-glycerol phosphate into indole and glyceraldehyde-3-phosphate.³ Indole diffuses along a 25 Å tunnel from TrpA to TrpB, where the PLP cofactor mediates C–C bond formation between C₃ of free indole (in preference to N₁) and an electrophilic serine-derived amino-acrylate species (E(A-A), **Figure 3-1A**). A recent engineering study demonstrated that activity lost by removal of TrpA could be recovered by directed evolution, leaving the stand-alone TrpB subunit as a simpler, more readily programmable platform for the synthesis of noncanonical amino acids (ncAAs).⁴

The electrophilic amino-acrylate intermediate of the native TrpS complex can react with nucleophilic species other than indole: the complex can accept molecules such as indoline,

aniline, benzenethiol, and benzeneselenol to form new C–N,^{4–9} C–S,¹⁰ and C–Se bonds,¹⁰ respectively (**Figure 3-1B**). By engineering the stand-alone TrpB (*vide supra*), Buller, Arnold, and coworkers systematically broadened the scope of ncAAs that are accessible using various nucleophiles, including an assortment of substituted indoles,^{4,8,11,12} azulene,¹³ nitroalkanes,¹⁴ and oxindoles¹⁵ (**Figure 3-1B**). Impressively, the TrpB variant engineered to utilize 3-substituted oxindoles facilitates the formation of quaternary C–C bonds in preference to C–N bond formation in a powerful demonstration of how enzyme selectivity can be leveraged for the construction of C–C bonds that are otherwise challenging to make.¹⁵

In this study, we investigated the capacity of ketone-derived enolates to serve as nucleophiles in TrpB-catalyzed reactions (**Figure 3-1C**). While these compounds are one of the most well-researched carbon-based nucleophiles in the synthetic literature, there are still no general methods for their direct asymmetric α -alkylation.¹⁶ Enzymatic reactions with enolates are relatively rare, and to our knowledge there are no examples engineering enzymes to accept enolates if they are not already the native substrate. Nevertheless, we hypothesized TrpB might be able to accomplish this feat, and that realizing this chemistry may inspire others to investigate further chemoenzymatic reactions with enolates. At the outset, we considered four main obstacles to achieving selective alkylation of ketones: (1) ketone tautomerization can result in multiple enolate stereo- and regioisomers, the latter stemming from asymmetrical ketones with multiple deprotonation sites; (2) the facial approach of the enolate to the electrophile dictates the C α stereochemistry and therefore must be precisely controlled to achieve stereoselectivity; (3) the substituted ketone product may *also* tautomerize and racemize the newly set chiral center; (4) enols are thermodynamically disfavored in aqueous conditions, so the population of reactive nucleophile available in solution would be very small. We reasoned that if TrpB could accept enolates as nucleophiles, the selectivity imparted by the enzyme active site could conquer the first two obstacles.

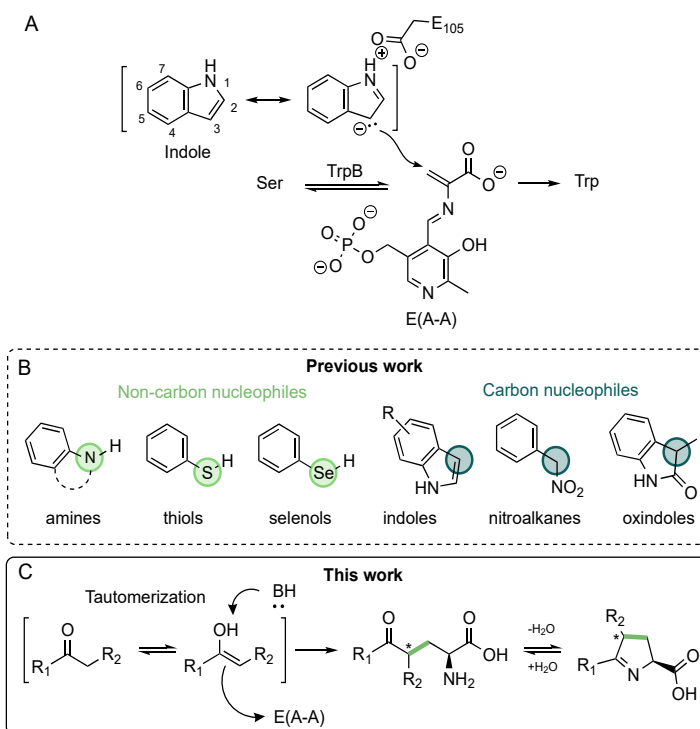


Figure 3-1. Overview of past and present work. (a) The C–C bond-forming step in TrpB-catalyzed reactions: nucleophilic addition of C₃ of indole to the amino acrylate (E(A-A)). (b) Nucleophiles previously shown to be used by TrpS/TrpB. (c) Enolate nucleophiles, supplied to the reaction as their ketone tautomer, can be used by TrpB to form linear ncAAs that undergo spontaneous cyclization.

3.2 Results and discussion

We began by screening a panel of purified TrpB enzymes from previous evolution projects for activity with serine and propiophenone. We selected propiophenone because it absorbs light in the ultraviolet (UV) region, bears a bulky phenyl substituent that resembles other non-indole nucleophiles that react with TrpB, and has been highlighted as a ketone for which no direct asymmetric alkylation methods have been reported.¹⁶ A *Thermotoga maritima* TrpB (*Tm*TrpB) variant previously engineered to synthesize 4-cyanotryptophan¹¹—Tm9D8*—displayed trace activity with propiophenone (30 TTN, **Figure 3-2A**) to form a mixture of cyclized imine isomers as detected by HPLC-MS (*vide infra*).

To improve this novel enzymatic reactivity, we initiated a directed evolution campaign using a combination of site-saturation and random mutagenesis and screened the mutant libraries for increases in product formation, remaining agnostic as to whether beneficial mutations resulted from enhanced enzymatic activity or expression. We screened lysate reactions at 37 °C during the first round of engineering, as this temperature elicited the highest activity of the Tm9D8* variant in previous reactions with 4-cyanotryptophan.¹¹ Optimizing the reaction conditions, we found product formation with the propiophenone substrate reached a maximum at 55 °C, and this temperature was used for subsequent rounds of directed evolution. Three rounds consecutively accrued the mutations F184H (TmE1, 26 TTN), I174T (TmE2, 28 TTN), and F41Y (TmE3, 39 TTN; **Figure 3-2A**).

Constrained to product detection by HPLC-MS, limited analytic screening capacity quickly became an obstacle to discovering improved variants. Thus, we explored screening by Multiple Injections in a Single Run (MISER), a flow-injection analysis method in which each sample in a 96-well plate is injected sequentially, passed through a short guard column, and analyzed by the mass spectrometer.¹⁷ This strategy dramatically improved screening throughput, decreasing the time per 96-well plate from 5 hours to 35 minutes. With MISER, we screened a larger error-prone PCR (epPCR) random mutagenesis library and found an improved variant bearing four additional mutations: K39I, E40G, P54Q, and I195T (TmE4, 53 TTN; **Figure 3-2A**). During this round of engineering, however, the poor signal-to-noise ratio of the MISER screen resulted in a number of false positives that impeded the overall workflow. Therefore, we decided to examine the substrate scope of TmE4 with the prospect of uncovering a new substrate that could be detected by MISER more reliably. To our delight, TmE4 exhibited much higher activity with 2-fluoroacetophenone (90 TTN, **Figure 3-2B**), which is consistent with the expected activating effect of the fluorine on the C α -H bond.

For the remainder of the evolution, we screened for activity with serine and 2-fluoroacetophenone using the MISER analysis method. In the next two rounds of engineering, several variants that yielded improvements were identified (TmE5, TmE6.3, and TmE6.6; **Figure 3-2B** and **Appendix B, Table B-14**). With a considerable number of

mutations now observed throughout the lineage, we generated a recombination library using staggered extension process (StEP) PCR,¹⁸ which allowed each residue of interest to either retain its mutation or revert back to wild type. Of the 600 colonies screened, the best variant displayed a further 2-fold increase in activity over the best round 6 variant, TmE6.3, by recombining mutations I166V and S228G (TmE7, 700 TTN; **Figure 3-2B**).

Because enolates are structurally and electronically dissimilar to the reactive indole enamine-like species, we investigated the roles of active-site residues E105 and H184 by introducing alanine mutations at each site and measuring changes in activity for both indole and 2F-acetophenone. E105 facilitates nucleophilic attack on the amino acrylate by stabilizing the accumulation of positive charge on the pyrrole ring of indole (**Figure 3-1A**); the E105G mutation in *TmTrpB* was shown to be deleterious to activity with indole and several analogs.¹³ Previous studies have shown that mutations to H184, which is spatially adjacent to E105, also strongly affect activity,^{11,14} although the specific role of histidine is less clear. The E105A mutation reduced the indole activity of TmE7 unsurprisingly by 93%, but only decreased the 2-fluoroacetophenone activity by 30% (**Figure 3-2C**). While the H184A mutation had no effect on indole activity, we were pleasantly surprised to find that it improved 2-fluoroacetophenone activity 3.8-fold. TrpB TmE7 with the H184A mutation (TmE8, 3140 TTN, 86% yield by HPLC; **Figure 3-2B**) was used for further studies of this new reactivity.

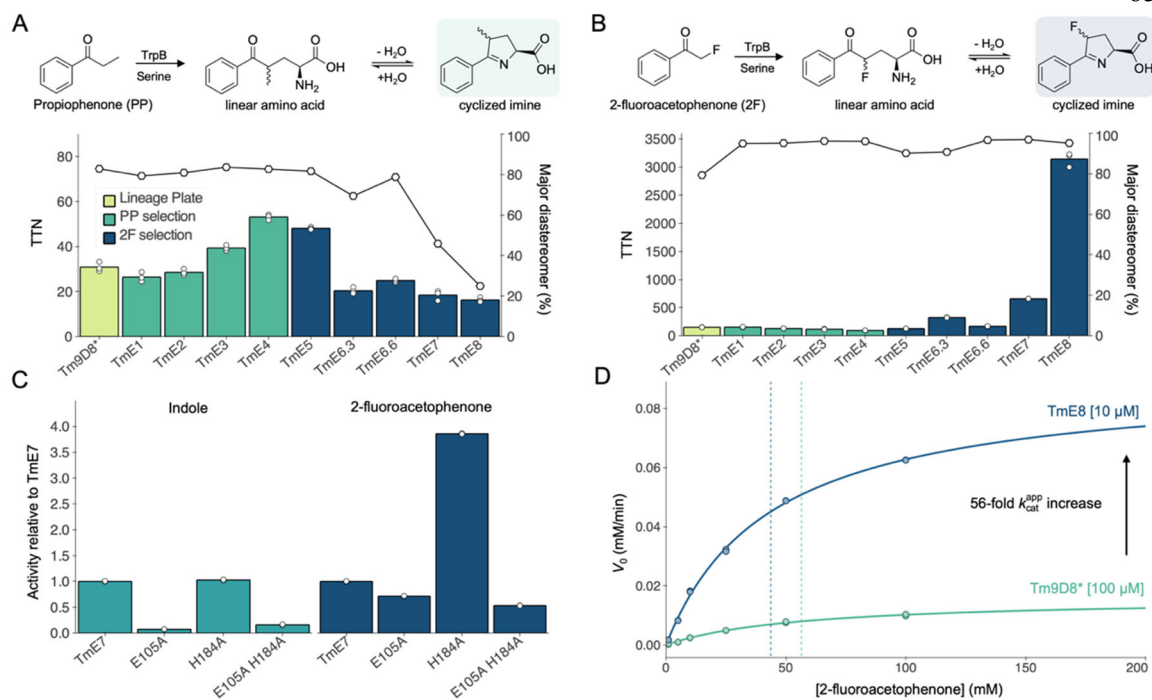


Figure 3-2 Directed evolution of enolate activity. (a) Activity of variants in the TrpB evolutionary lineage with propiophenone (PP) and serine. Bars represent the average of two or three replicates shown as individual points. Colors refer to the substrate used for screening. The diastereoselectivity, shown as individual points, for the cyclized imine product was measured for each variant. (b) Activity of variants in the TrpB evolutionary lineage with 2-fluoroacetophenone (2F) and serine. Bars represent the average of two or three replicates shown as individual points. Colors refer to the substrate used for screening. The diastereoselectivity, shown as individual points, for the cyclized imine product was measured for each variant. (c) Activities of alanine mutants with indole and 2-fluoroacetophenone, relative to TmE7. (d) Michaelis-Menten kinetics of Tm9D8* (teal) and TmE8 (dark blue). Dotted lines represent the K_M^{app} . Reaction conditions for Figure 2 can be found in **Appendix B, Section B.2.8** and **B.2.9**

Using the best variant for each respective substrate, we isolated the products from enzymatic reactions for NMR characterization, as reported syntheses for making the authentic standards were absent from the literature. The enzymatic products of the reactions with propiophenone and 2-fluoroacetophenone were each identified to be a mixture of two imine isomers. Reapplication of the propiophenone-derived imine product mixture as an authentic standard showed that directed evolution had a strong influence on the diastereomeric outcome (**Figure 3-2A**). Although we had not screened specifically for diastereoselectivity, the preference for a single diastereomer was inverted between Tm9D8* and TmE8 (82:18 to 24:76

diastereomeric ratio, *d.r.*, **Figure 3-2A**), demonstrating that stereochemistry at the newly formed chiral center can be tuned with this platform. In reactions with 2-fluoroacetophenone, the enzymes exhibited an even higher degree of selectivity, with the final variant, TmE8, boasting a *d.r.* of 95:5 (**Figure 3-2B**). However, we observed an erosion in selectivity at prolonged reaction times (**Appendix B, Figure B-2**), suggesting the newly formed chiral center racemizes through tautomerization in solution, as discussed above. In principle, subsequent reduction of the product would fix the stereochemistry, and indeed we observed the expected shifts in product mass by HPLC-MS upon treatment with sodium cyanoborohydride (**Appendix B, Figure B-3**), but the resulting mixture of diastereomers rendered rigorous product characterization problematic. Leveraging the selectivity of a suitable reductase in an enzymatic cascade may be the best solution to this problem (*vide infra*).

To gain insight into the mechanism of this new transformation, we first explored the kinetics of product formation in reactions with 2-fluoroacetophenone catalyzed by Tm9D8* and TmE8 (**Figure 3-2D**). TmE8 exhibited a 56-fold increase in apparent turnover number ($k_{\text{cat}}^{\text{app}}$) compared to Tm9D8* (9.0 min⁻¹ and 0.16 min⁻¹, respectively), which is likely the principal factor contributing to the 20-fold increase in TTN observed across the evolutionary lineage. The change in the apparent K_M (K_M^{app}) was modest (57 mM and 44 mM for Tm9D8* and TmE8, respectively). The dramatic improvement in $k_{\text{cat}}^{\text{app}}$ without a significant change in the strikingly high K_M^{app} values prompted us to investigate whether the enzyme was binding the ketone and facilitating its tautomerization in the active site or simply reacting with the small population of the enol tautomer present in solution (**Figure 3-3A**).

We measured the rate of deuterium exchange into 2-fluoroacetophenone from deuterated buffer in the absence of enzyme (**Figure 3-3B**). Under these conditions, the rate of uncatalyzed tautomerization is an order of magnitude faster than the enzymatic reaction (**Figure 3-3C**). We measured the rate of protium exchange into 2,2-*d*₂-2-fluoroacetophenone in unlabeled buffer, from which we calculated a kinetic isotope effect (KIE) of 5.7 for uncatalyzed deprotonation of the starting material (**Figure 3-3B**). We note that the

background rate of enolization is calculated by measuring the formation of isotopically labeled ketone resulting from *two* tautomerization events and is a *conservative* approximation of the in-solution tautomerization rate, as a single-step conversion to the enol would be marginally faster. To learn more about tautomerization in the enzyme active site, we extracted the initial velocity (V_0) of product formation from the progress curves of TmE8 with the unlabeled 2-fluoroacetophenone substrate and 2,2- d_2 -2-fluoroacetophenone and determined a KIE of 5.8 (**Figure 3-3C**). The remarkable similarity of these two KIEs suggests that the KIE measured in the enzymatic reaction may also originate from in-solution tautomerization of the ketone. We therefore considered a second scenario wherein TmE8 binds the enol tautomer directly (**Figure 3-3A**, Enol-binding pathway). Based on known equilibrium constants of acetophenone tautomerization,^{19,20} we estimate that the 2-fluoroacetophenone-derived enol tautomer is present at low μM concentrations for these experiments. Under these conditions, approximately a single turnover of enzyme (present at 10 μM) would deplete the starting material entirely, obliging the system to then wait for tautomerization to deliver more substrate. In such a scenario, the KIE observed in the enzymatic reaction would then exactly match the KIE for the uncatalyzed exchange, which was observed.

A key distinguishing feature of the ketone-binding vs enol-binding models is the kinetic order of the reaction in catalyst. For the ketone-binding model, initial velocities should be directly proportional to catalyst concentration. However, for the enol-binding model, the reaction is fundamentally limited by the speed of the uncatalyzed tautomerization, which would manifest as non-linearity between catalyst concentration and initial velocity. To examine this hypothesis, we extracted V_0 values from progress curves with increasing concentrations of TmE8 (10 μM –150 μM , **Figure 3-3D**). We observed a clear break in the linearity between [TmE8] and V_0 , with the rate of the enzyme-catalyzed reaction asymptotically approaching the 2-fluoroacetophenone exchange rate (**Figure 3-3E**). We therefore conclude that TmE8 does not bind the ketone substrate and catalyze tautomerization, but rather reacts directly with the enol tautomer that forms in solution.

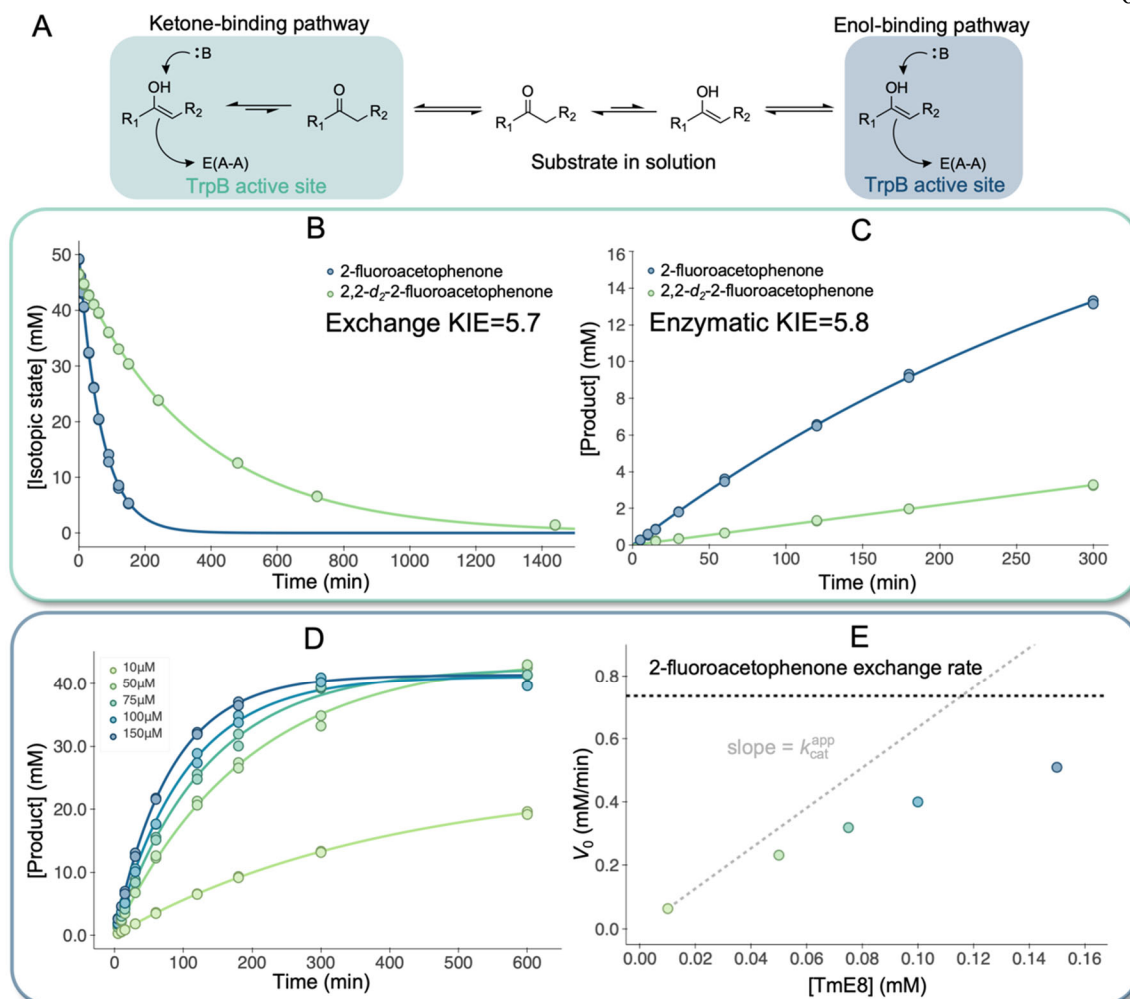




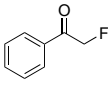
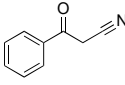
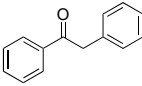
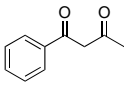
Figure 3-3. Mechanistic characterization of enolate activity. (a) Possible pathways to enolate formation. (b) Exchange curves of 2-fluoroacetophenone and 2,2- d_2 -2-fluoroacetophenone in deuterated buffer and unlabeled buffer, respectively (c) TmE8 (10 μ M) progress curves with 2-fluoroacetophenone and 2,2- d_2 -2-fluoroacetophenone at 50 mM substrate loading. (d) TmE8 progress curves at varying [enzyme] with 50 mM 2-fluoroacetophenone. (e) Initial velocity of product formation extracted from the progress curves in D represented as a function of [TmE8]. Details of nonlinear regression analyses can be found in **Appendix B, Section B.2.8** and **B.2.10**.

The enol-binding mechanism provides clarity to other observations that arose during our studies. The parent enzyme, Tm9D8*, is over 50-fold slower than the final variant and has a much lower KIE of 3.9 (**Appendix B, Figure B-4**), which we hypothesize is due to the larger steady-state concentration of enol tautomer in solution resulting from slower consumption by the enzyme. Furthermore, if the K_M for the substrate were truly 50 mM, then V_0 would

be impacted by changes in both k_{cat} and K_M under the conditions we used to screen for improvements during directed evolution. However, we observed virtually no change in K_M^{app} , despite the significant structural differences between indole and acetophenones. Instead, the enol-binding mechanism suggests the K_M for the true substrate is in the low μM regime, thus providing no selective pressure for this parameter to change. In order to achieve such a low K_M , we speculate that the TrpB active site must have exquisite complementarity for the substrate.

To understand the effect of $C\alpha$ substitutions, we evaluated the substrate scope of the engineered variants and measured the rates of deuterium exchange into those substrates that are compatible with aqueous solutions (**Table 3-1** and **Appendix B, Section B.2.10**). Comparing observed rate constants (k_{exg}), propiophenone tautomerizes an order of magnitude slower than acetophenone ($5.86 \times 10^{-4} \text{ min}^{-1}$ vs. $4.14 \times 10^{-3} \text{ min}^{-1}$, respectively). However, Tm9D8* exhibited a higher level of activity with propiophenone than acetophenone, suggesting that small changes in substitutions at the $C\alpha$ position are distinguished by the enzyme, and positioning of the resulting enol is crucial. In agreement with this hypothesis, TmE8 showed varying degrees of activity with different substituents at $C\alpha$. Notably, TmE8 catalyzed the reaction with 2-cyanoacetophenone, a substrate electronically similar to 2-fluoroacetophenone, with 250 TTN (**Appendix B, Figure B-9**). Intriguingly, the product of this reaction isomerizes to the corresponding enamine, which extends the π conjugation from the arene to the nitrile, producing an intense absorption band in the visible region (**Appendix B, Figure B-14**). TmE8 showed no activity with 2-phenylacetophenone, suggesting that it does not readily accommodate bulkier substituents at $C\alpha$; however, trace activity was observed with benzoylacetone (< 1 TTN). We also found that a variant stemming from *Pyrococcus furiosus* TrpB, PfTrpB^{7E6}, exhibited similar activity with benzoylacetone (See **Appendix B, Figure B-9**). Thus, it appears that multiple TrpB scaffolds are capable of catalyzing this new reaction, broadening the sequence space available to subsequent engineering efforts with other substrates of interest.

Table 1. Ketone substrate scope of TmE8 and deuterium exchange rate constants (k_{exg}).

Entry	Substrate	TTN in 24 h	k_{exg} (min^{-1})
1		3	4.14×10^{-3}
2		<1	5.86×10^{-4}
3		3150	1.47×10^{-2}
4		250	n.d. ^[a]
5		n.d.	5.15×10^{-2}
6		<1	n.d. ^[b]

[a] 2-cyanoacetophenone is insoluble in buffer. [b] Benzoylacetone hydrolyzes at 55 °C.

As it stands currently, the maximum rate of product formation using TrpB is approximately limited by k_{exg} , but this limitation does not preclude its potential utility. For example, assuming that a TrpB variant was limited exclusively by k_{exg} , a reaction with 50 mM 2-fluoroacetophenone would reach completion in less than 4 hours. Under the same conditions, a reaction with acetophenone would near completion in 12 hours. Nevertheless, emulating Nature's machinery may comprise the ultimate solution. Despite the fact that the keto-enol equilibrium strongly favors the ketone in aqueous conditions, enolates are common nucleophiles in biology, and several strategies to overcome these unfavorable thermodynamics have evolved naturally. For example, type I aldolases form an intermediate Schiff base between the substrate carbonyl and an active-site lysine residue to effectively reduce the pKa of the C α proton.²¹ Although TmE8 does not boast such sophistication, there are no obvious factors that would directly preclude engineering the machinery necessary to effect enolization by an analogous mechanism.

In conclusion, we have demonstrated that TrpB can be evolved for the asymmetric alkylation of ketones and have provided a glimpse into the scope of substrates that may be compatible with this new enzymatic transformation. Given that TrpB TmE8 does not influence the keto-enol tautomerization rate, yet exhibits control over the formation of diastereomeric products, we posit that the active site has been engineered to bind a particular configuration of the nucleophilic enol tautomer and direct its approach to the electrophilic amino acrylate. The linear products of the TrpB-catalyzed reactions presented herein spontaneously cyclize to form derivatives of 1-pyrroline-5-carboxylic acid,²² which could be further processed by a suitable imine reductase in a cascade system to furnish chirally dense noncanonical prolines.²³ Furthermore, expanding this platform to include aldehydes as nucleophilic substrates and β -branched amino acids as amino-acrylate precursors would provide access to each carbon substituent on the nitrogen heterocycle. Finally, this study demonstrates that enolates can be intercepted by enzymes, raising the possibility that other classes of enzymes could also be leveraged for asymmetric ketone alkylation in the future.

Chapter III Bibliography

1. Fesko, K. & Gruber-Khadjawi, M. Biocatalytic Methods for C–C Bond Formation. *ChemCatChem* **5**, 1248–1272 (2013).
2. Schmidt, N. G., Eger, E., & Kroutil, W. Building Bridges: Biocatalytic C–C-Bond Formation toward Multifunctional Products. *ACS Catal.* **6**, 4286–4311 (2016).
3. Watkins-Dulaney, E., Straathof, S., & Arnold, F. Tryptophan Synthase: Biocatalyst Extraordinaire. *ChemBioChem* **22**, 5–16 (2021).
4. Buller, A. R. et al. Directed evolution of the tryptophan synthase β -subunit for stand-alone function recapitulates allosteric activation. *Proc. Natl. Acad. Sci.* **112**, 14599–14604 (2015).
5. Roy, M., Keblawi, S., & Dunn, M. F. Stereoelectronic Control of Bond Formation in *Escherichia coli* Tryptophan Synthase: Substrate Specificity and Enzymatic Synthesis of the Novel Amino Acid Dihydroisotryptophan. *Biochemistry* **27**, 6698–6704 (1988).
6. Brzovic, P. S., Kayastha, A. M., Miles, E. W., & Dunn, M. F. Substitution of glutamic acid 109 by aspartic acid alters the substrate specificity and catalytic activity of the β -

- subunit in the tryptophan synthase bienzyme complex from *Salmonella typhimurium*. *Biochemistry* **31**, 1180–1190 (1992).
7. Tanaka, H. et al. Production of a novel tryptophan analog, β -1-indazole-L-alanine with tryptophan synthase of *Escherichia coli*. *FEBS Lett.* **196**, 357–360 (1986).
 8. Murciano-Calles, J., Romney, D. K., Brinkmann-Chen, S., Buller, A. R., & Arnold, F. H. A Panel of TrpB Biocatalysts Derived from Tryptophan Synthase through the Transfer of Mutations that Mimic Allosteric Activation. *Angew. Chemie - Int. Ed.* **55**, 11577–11581 (2016).
 9. Ferrari, D., Yang, L. H., Miles, E. W., & Dunn, M. F. β D305A mutant of tryptophan synthase shows strongly perturbed allosteric regulation and substrate specificity. *Biochemistry* **40**, 7421–7432 (2001).
 10. Esaki, N. & Soda, K. Preparation of sulfur and selenium amino acids with microbial pyridoxal phosphate enzymes. *Methods Enzymol.* **143**, 291–297 (1987).
 11. Boville, C. E., Romney, D. K., Almhjell, P. J., Sieben, M., & Arnold, F. H. Improved Synthesis of 4-Cyanotryptophan and Other Tryptophan Analogues in Aqueous Solvent Using Variants of TrpB from *Thermotoga maritima*. *J. Org. Chem.* **83**, 7447–7452 (2018).
 12. Romney, D. K., Murciano-Calles, J., Wehrmüller, J. E., & Arnold, F. H. Unlocking Reactivity of TrpB: A General Biocatalytic Platform for Synthesis of Tryptophan Analogues. *J. Am. Chem. Soc.* **139**, 10769–10776 (2017).
 13. Watkins, E. J., Almhjell, P. J., & Arnold, F. H. Direct enzymatic synthesis of a deep-blue fluorescent noncanonical amino acid from azulene and serine. *ChemBioChem* **21**, 80–83 (2020).
 14. Romney, D. K., Sarai, N. S., & Arnold, F. H. Nitroalkanes as Versatile Nucleophiles for Enzymatic Synthesis of Noncanonical Amino Acids. *ACS Catal.* **9**, 8726–8730 (2019).
 15. Dick, M., Sarai, N. S., Martynowycz, M. W., Gonen, T., & Arnold, F. H. Tailoring Tryptophan Synthase TrpB for Selective Quaternary Carbon Bond Formation. *J. Am. Chem. Soc.* **141**, 19817–19822 (2019).
 16. Cano, R., Zakarian, A., & McGlacken, G. P. Direct Asymmetric Alkylation of

- Ketones: Still Unconquered. *Angew. Chemie - Int. Ed.* **56**, 9278–9290 (2017).
17. Welch, C. J. et al. MISER chromatography (multiple injections in a single experimental run): The chromatogram is the graph. *Tetrahedron Asymmetry* **21**, 1674–1681 (2010).
 18. Zhao, H. & Zha, W. In vitro ‘sexual’ evolution through the PCR-based staggered extension process (StEP). *Nat. Protoc.* **1**, 1865–1871 (2006).
 19. Keeffe, J. R., Kresge, A. J., & Toullec, J. Acid-catalyzed enolization of acetophenone: catalysis by bisulfate ion in sulfuric acid solutions. *Can. J. Chem.* **64**, 1224–1227 (1986).
 20. Haspra, P., Sutter, A., & Wirz, J. Acidity of Acetophenone Enol in Aqueous Solution. *Angew. Chemie Int. Ed. English* **18**, 617–619 (1979).
 21. Dean, S. M., Greenberg, W. A., & Wong, C. H. Recent advances in aldolase-catalyzed asymmetric synthesis. *Adv. Synth. Catal.* **349**, 1308–1320 (2007).
 22. Bodner, M. J. et al. Definition of the Common and Divergent Steps in Carbapenem β -Lactam Antibiotic Biosynthesis. *ChemBioChem* **12**, 2159–2165 (2011).
 23. Mangas-Sanchez, J. et al. Imine reductases (IREDs). *Curr. Opin. Chem. Biol.* **37**, 19–25 (2017).

SUPPLEMENTARY INFORMATION FOR CHAPTER III

B.1 Supplementary figures

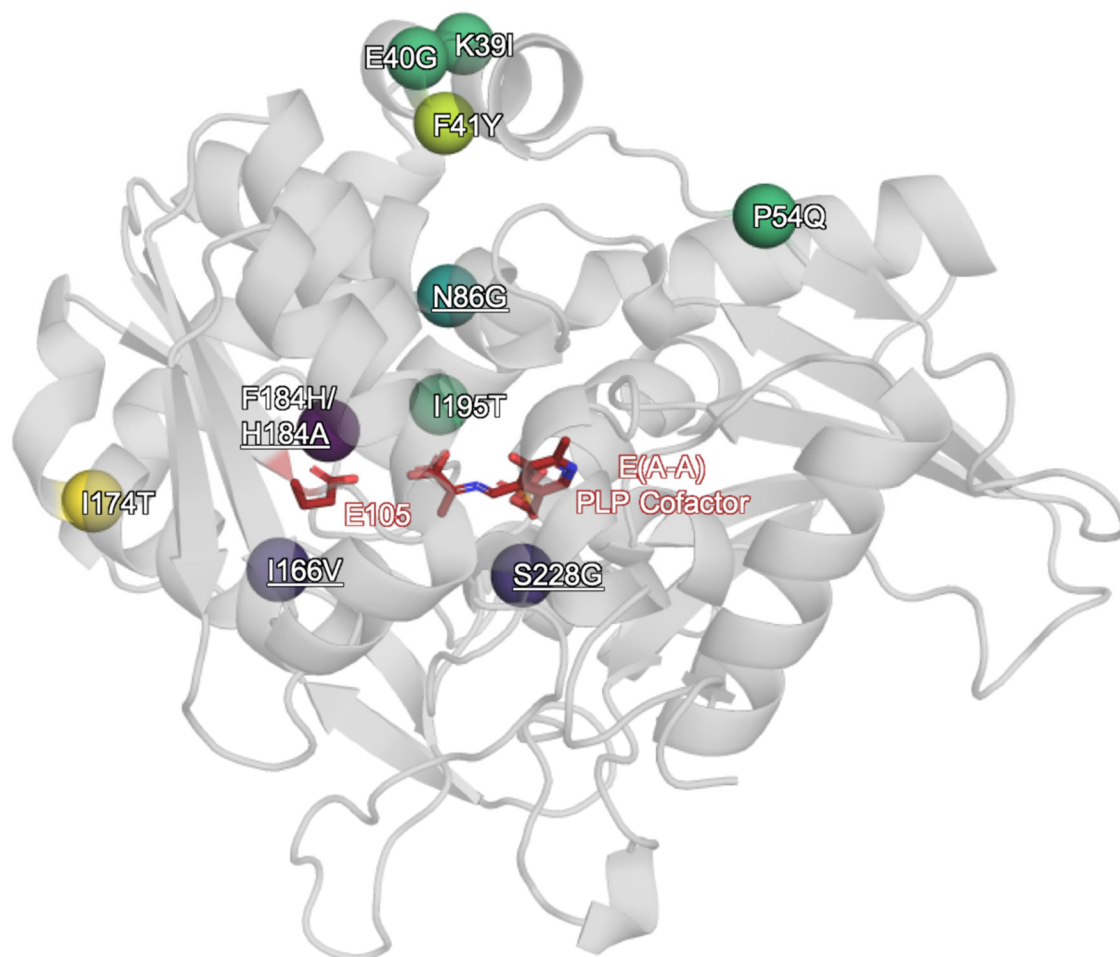


Figure B-1. Location of mutations highlighted in homology model of Tm9D8*¹.



Figure B-2. Racemization of enzymatic propiophenone-derived cyclic imine product over time. Reactions were set up according to Section B.2.8. Reaction conditions: 50 mM propiophenone in DMSO (5% v/v), 50 mM serine, 100 μ M purified TmE2. Reactions were incubated at 37 $^{\circ}$ C and quenched after 20 minutes and 24 hours. HPLC-MS chromatograms are cyclic imine ion mass+1 (204). Cyclic imine retention times are \sim 1.71 and \sim 1.95 min.

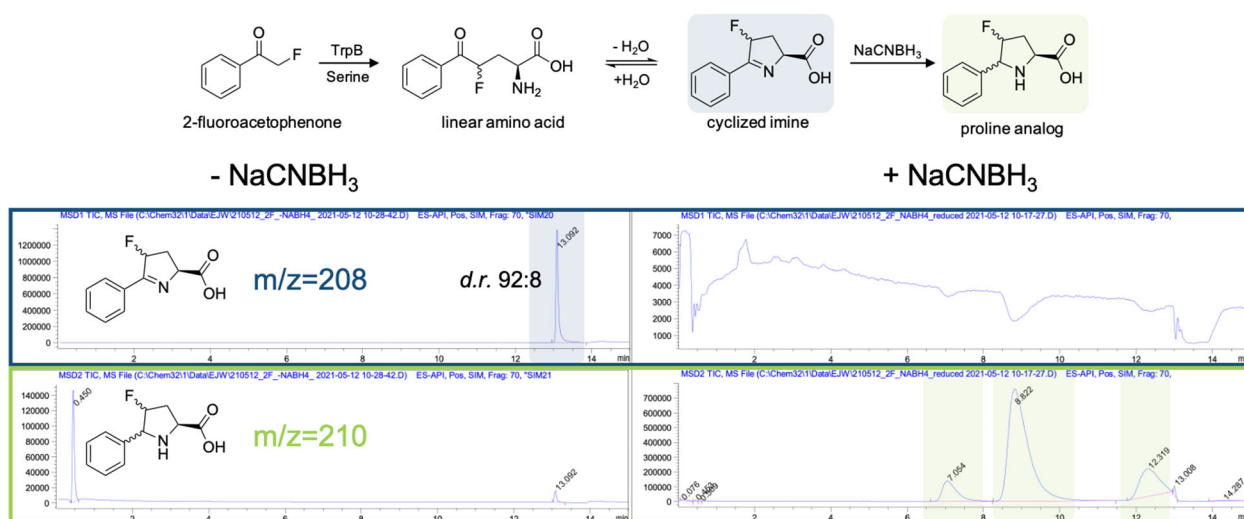


Figure B-3. Sodium cyanoborohydride reduction of 2-fluoroacetophenone product. Reactions were set up according to Section B.2.9.

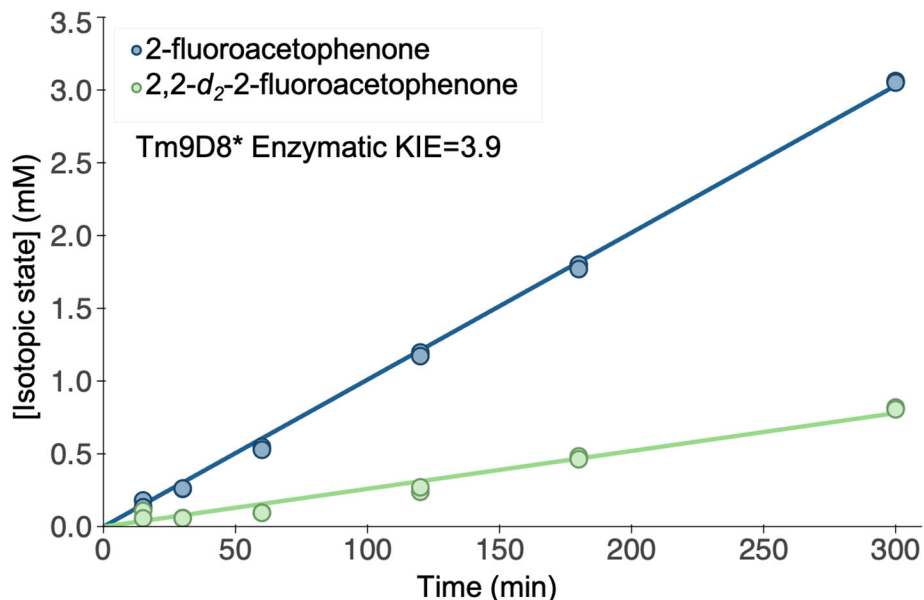


Figure B-4. Tm9D8* KIE experiment. Reactions were set up according to **Section B.2.8**. Reaction conditions were identical to those in **Figure 3-3B** except enzyme loading: 50 mM nucleophile in DMSO (5% v/v), 50 mM serine, 100 μ M purified Tm9D8*, incubation at 55 $^{\circ}$ C.

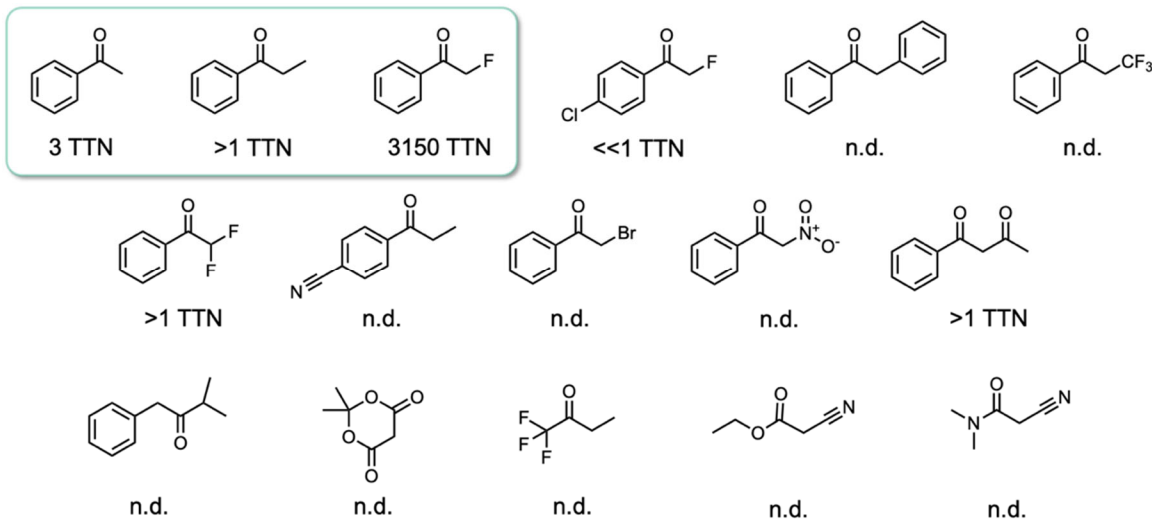


Figure B-5. Substrate scope of TmE8. Boxed substrates reactions were set up according to **Section B.3.7**. Reaction conditions: 50 mM nucleophile in DMSO (5% v/v), 50 mM serine, 5 or 100 μ M TmE8, 24 h incubation at 55 $^{\circ}$ C TTN were determined from standard curve. All other reactions were set up according to **Section B.2.7**. Reaction conditions: 10 mM nucleophile in DMSO (5% v/v), 50 mM serine, 100 μ L TmE8 lysate, 24 h incubation at 55 $^{\circ}$ C.

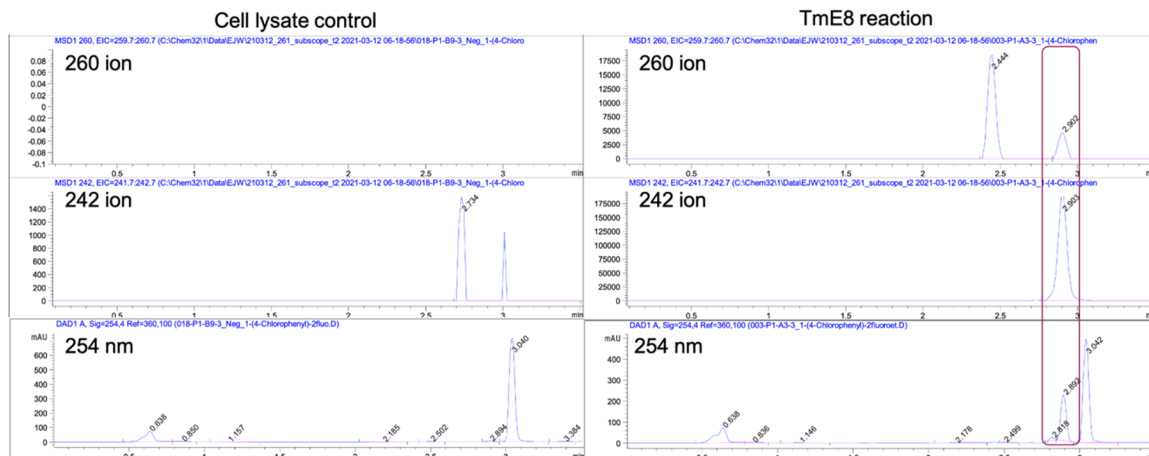
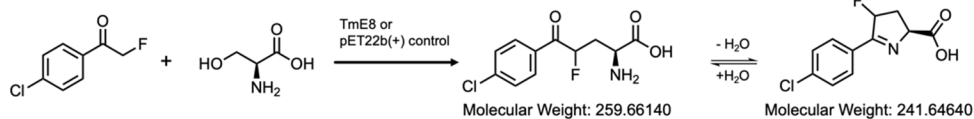


Figure B-6. HPLC-MS traces for para-chloro-2-fluoroacetophenone. Reactions were set up according to Section B.2.7. Reaction conditions: 10 mM para-chloro-2-fluoroacetophenone in DMSO (5% v/v), 50 mM serine, 100 μ L TmE8 lysate or pET22b(+) lysate control, 24 h incubation at 55 $^{\circ}$ C.

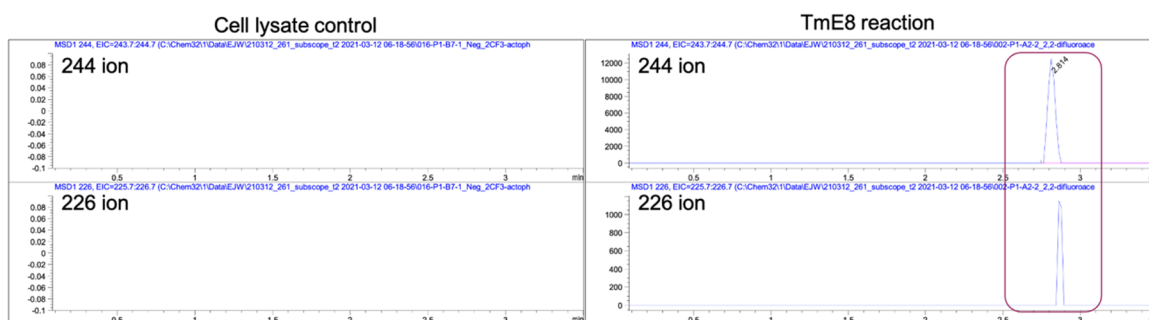
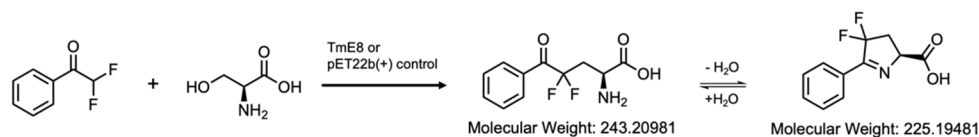


Figure B-7. HPLC-MS traces for 2,2-difluoroacetophenone. Reactions were set up according to Section B.2.7. Reaction conditions: 10 mM 2,2-difluoroacetophenone in DMSO (5% v/v), 50 mM serine, 100 μ L TmE8 lysate or pET22b(+) lysate control, 24 h incubation at 55 $^{\circ}$ C.

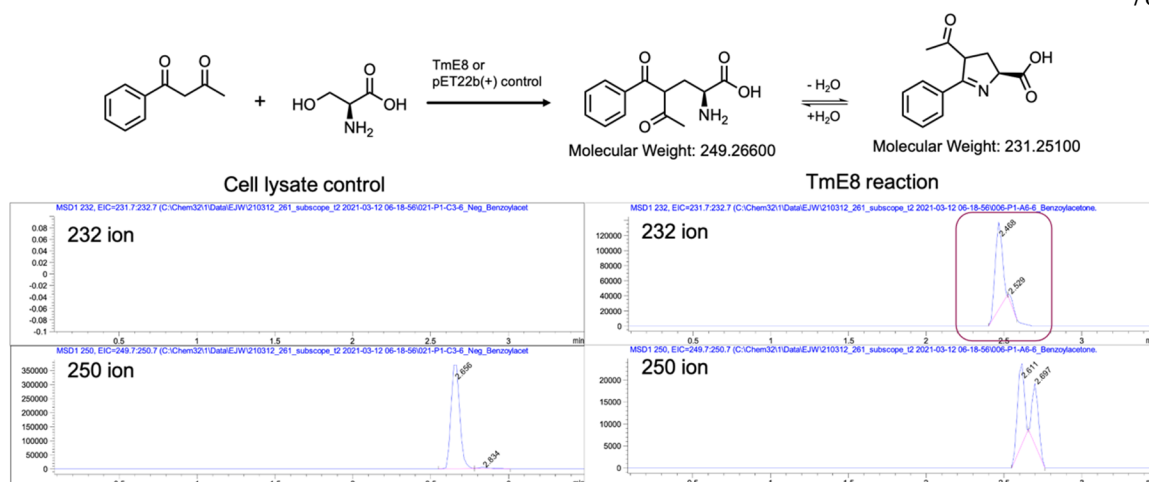


Figure B-8. HPLC-MS traces for benzoylacetone. Reactions were set up according to Section B.2.7. Reaction conditions: 10 mM benzoylacetone in DMSO (5% v/v), 50 mM serine, 100 μ L TmE8 lysate or pET22b(+) lysate control, 24 h incubation at 55 $^{\circ}$ C.

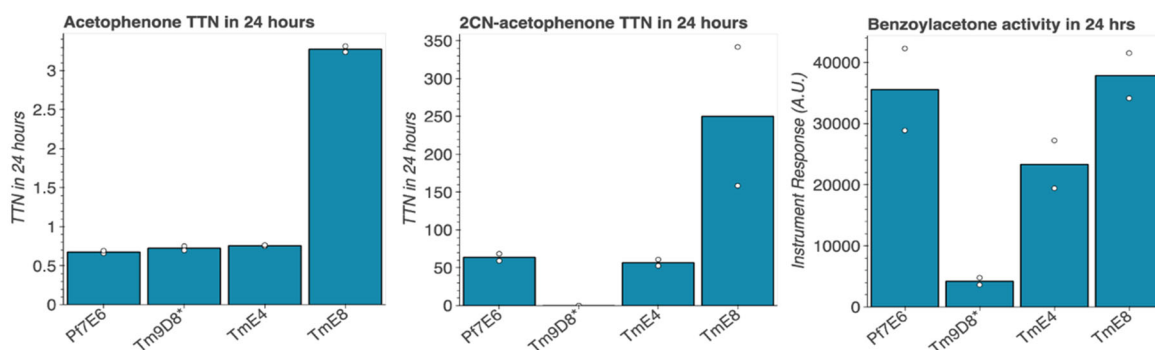


Figure B-9. Substrate scope reactions. Reactions were set up according to Section B.2.8. Reaction conditions: 50 mM nucleophile in DMSO (5% v/v), 50 mM serine, purified protein (Acetophenone: 100 μ M, max TTN 500; 2CN-acetophenone: 5 μ M, max TTN 10,000; Benzoylacetone: 100 μ M, max 500 TTN) 24 h incubation at 55 $^{\circ}$ C.

B.2 Methods

B.2.1 General experimental methods

Chemicals and reagents were purchased from commercial sources and used without further purification. NMR spectra were recorded on a Varian Unity/Inova 500 spectrometer operating at 500 MHz and 125 MHz for 1H and ^{13}C , respectively, a Bruker Avance 400 spectrometer equipped with a cryoprobe operating at 400 MHz and 101 MHz for 1H and ^{13}C , respectively, or a Varian Mercury 300 spectrometer operating at 300 MHz and 282 MHz for

^1H and ^{19}F , respectively. Chemical shifts are reported in ppm (δ) and calibrated using the residual solvent resonance. For ^1H NMR: D_2O , δ 4.79; DMSO-d_6 , δ 2.50; CDCl_3 , δ 7.26. For ^{13}C NMR: DMSO-d_6 , δ 39.52; CDCl_3 , 77.16. Data are reported as follows: chemical shift (multiplicity [singlet (s), doublet (d), doublet of doublets (dd), doublet of doublet of doublets (ddd), triplet (t), triplet of doublets (td), multiplet (m)], coupling constants [Hz], integration).

All cultures were grown in Terrific Broth supplemented with 100 $\mu\text{g/mL}$ carbenicillin (TB_{carb}). Cultures were shaken in New Brunswick Innova 4000 (shaking diameter 19 mm) shakers, with the exception of 96-well plates, which were shaken in a Multitron INFORS HT (shaking diameter 50 mm) shaker. Lysis buffer was composed of 50 mM potassium phosphate, pH 8.0 (KPi buffer), supplemented with 200 μM pyridoxal 5'-phosphate (PLP), unless otherwise specified. Reactions were performed in 50 mM KPi, pH 8, unless otherwise specified. Liquid chromatography/mass spectrometry (LCMS) was performed on an Agilent 1260 HPLC-MS equipped with an S18 silica guard column (Guard: Poroshell 120, UHPLC Guard, ESC18, 4.6 mm; Column: Poroshell 120 ESC18, 4.6x50 mm, 2.7 μm) using $\text{CH}_3\text{CN}/\text{H}_2\text{O}$ (0.1% acetic acid by volume): 5% to 95% CH_3CN over 3 min; 1 mL/min.

Table B-1. Chemicals, reagents, and equipment

Chemical, reagent, or equipment	Supplier	Catalog #
ArctiSeal™ 96 Well Square Silicone/PTFE Coating	Arctic White LLC	AWSM-1003S
AcroPrep wvPTFE 96-well filter plates	VWR	76308-642
Biotage Sfär silica C18-D 12 gram column	Biotage	FSUD-0401-0012
HPLC 2-mL Glass Vials	Agilent	5183-2068
BugBuster 10X Protein Extraction Reagent (Novagen)	Millipore Sigma	70921-4
Bradford 1X dye reagent	Bio-Rad	#5000205
Clean and Concentrate -5 kit	Zymo	D4004
cOmpete™ His-Tag Purification Column	Sigma Aldrich	6781543001
<i>DpnI</i>	New England Biolabs	R0176
E. cloni® EXPRESS BL21(DE3) cells	Lucigen	60300
Gel DNA extraction kit	Zymo	D4008
Glass-lined polypropylene 96-well reaction plates	VWR	10036-272
Hen egg white lysozyme	Sigma Aldrich	
Isotemp™ Digital-Control Water Bath	Fisherbrand™	Model 220
Microlab NIMBUS96 liquid handler	Hamilton	
Taq	New England Biolabs	M0273
Phusion	New England Biolabs	M0530
Pierce™ BCA Protein Assay Kit	ThermoFisher	23225
Quick Start™ Bradford 1x Dye Reagent	BioRad	500-0205
2.0-mL deep 96-well polypropylene Masterblock	USA Scientific	5678-0285

B.2.2 Cloning, expression, and purification of TrpB variants

The gene encoding Tm9D8* was previously cloned into pET22b(+) with a C-terminal 6x His tag. Protein expression of the variants was carried out in Lucigen *E. coli*[®] EXPRESS BL21(DE3) cells by inoculating 5 mL of TB_{carb} with a single colony and incubating this pre-culture overnight at 37 °C and 230 rpm. For expression, 8 mL culture was used to inoculate 800 mL Tb_{carb} (1:100 dilution) in a 1.8-L Fernbach flask and shaken at 200 rpm and 37 °C for 5 hours. Cultures were chilled on ice for 20–30 minutes, and protein expression was induced with a final concentration of 1 mM isopropyl β-D-thiogalactopyranoside (IPTG). Expression proceeded at 200 rpm and 20 °C for approximately 24 h. Cells were harvested by centrifugation at 10,000g for 20 minutes at 4 °C, and the supernatant was decanted. The pellet was stored at –20 °C until further use.

For protein purification, cells were thawed and resuspended in 4 mL lysis buffer per gram of pellet (standard lysis buffer: 1 mg/mL lysozyme, 0.1 mg/mL bovine pancreas DNase I, 1X BugBuster). BugBuster was added after the pellet had been mostly resuspended to prevent foaming. Cells were gently agitated on a rocking platform at room temperature for 30 minutes followed by heat-treatment in a 75 °C water bath for >1 h. The supernatant was collected from clarified lysate following centrifugation for 20 min at 14,000g and 4 °C. Purification was performed with an AKTA purifier FPLC system (GE Healthcare) and a 1-mL Ni-NTA column. Protein was eluted by applying a linear gradient of 100 mM to 500 mM imidazole in 25 mM KPi buffer, pH 8.0 and 100 mM NaCl. Fractions containing purified protein were dialyzed into 50 mM KPi buffer, pH 8.0, flash frozen in liquid nitrogen, and stored at –80 °C. Protein concentrations were determined using the Bio-Rad Quick Start™ Bradford Protein Assay.

B.2.3 Construction of site-saturation mutagenesis libraries

Site-saturation mutagenesis libraries were generated with the Tm9D8* gene (and evolutionary variants) using degenerate ‘NNK’ primers. Forward and reverse primers were designed to possess an overlapping region with a T_m = 55 °C and an independent binding

region with a $T_m = 60\text{ }^\circ\text{C}$ (not including the degenerate site). The entire plasmid was amplified in a single PCR using the Phusion polymerase. PCR fragments were treated with *DpnI* for 1 h at $37\text{ }^\circ\text{C}$, analyzed via agarose gel, purified using Zymo Gel DNA extraction kit, and then inserted into a pET22b(+) vector via Gibson assembly.² The Gibson assembly product was purified and concentrated using the Zymo DNA Clean and Concentrate–5 kit. Lucigen *E. coli*[®] EXPRESS BL21(DE3) cells were transformed with the Gibson assembly product.

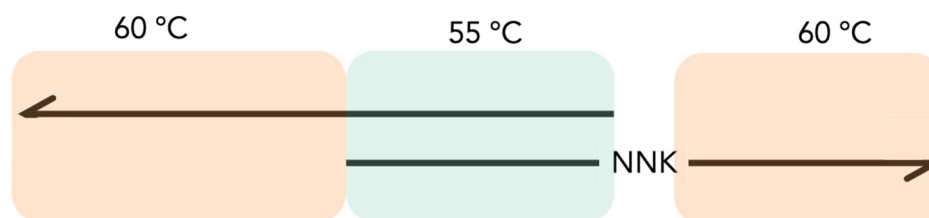


Figure B-10. Gibson assembly primers.

Table B-2. Site saturated mutagenesis PCR mix

Component	Volume (μL)
primer_F (10 μM)	1
primer_R (10 μM)	1
Template (~50-100 ng)	0.5
NTP mix (10 mM)	1
Phusion	0.5
HF buffer (5x)	10
water	33.5
DMSO	0.5

Table B-5. Random mutagenesis PCR mix

Component	Volume (μL)		
	200 μM MnCl_2	300 μM MnCl_2	400 μM MnCl_2
Template DNA (~100 ng/mL prep)	1	1	1
10x Taq buffer	10	10	10
dNTPs (10 mM)	4	4	4
TrpB_F primer (10 μM)	2	2	2
TrpB_R primer (10 μM)	2	2	2
MnCl_2 (1 mM)	20	30	40
Taq polymerase	1.6	1.6	1.6
Water	69.4	59.4	49.4

Table B-6. Random mutagenesis PCR thermal cycler method

	Step	Temp ($^{\circ}\text{C}$)	Time (min)
1	Initial denature	95	2:00
2	Denature	95	0:30
3	Anneal	55	0:30
4	Extend	68	1:25
Repeat 2–4 X30			
5	Final extension	68	10:00
6	Hold	10	∞

B.2.5 Construction of StEP recombination libraries

Recombination libraries were generated using the staggered extension process (StEP) PCR method as previously reported.⁴ All winning variants from rounds 1–6b were combined in a single template master mix such that each StEP PCR reaction received 35 ng/variant. The TrpB gene was amplified using the TrpB_F and TrpB_R primers from Table S5. The extension temperature ranged a gradient from 44.6–61.1 $^{\circ}\text{C}$ across eight PCR samples. A small quantity (2.5 μL) of each PCR was ran on an analytical gel to determine the lowest successful extension temperature. The samples were digested with *DpnI* for 1 hour at 37 $^{\circ}\text{B}$. The PCR with the lowest successful extension temperature(s) were used as template for a secondary PCR to amplify the library. The secondary PCR was analyzed via an analytical gel, purified using the Zymo Gel Extraction kit, and then inserted into a pET22b(+) vector via Gibson assembly.²

Table B-7. StEP Primary PCR mix

Component	Volume (μL)
Template (35 ng/variant)	Variable
10x Standard Taq buffer	2
dNTPs (10 mM)	0.4
TrpB_F primer (1 μM)	1
TrpB_R primer (1 μM)	1
Taq polymerase	0.33
Water	Q.S. 20
Total	20

Table B-8. StEP Primary PCR thermal cycler protocol

	Step	Temp ($^{\circ}\text{C}$)	Time (min)
1	Initial Denature	95	05:00
2	Denature	95	00:10
3	Anneal	55	00:05
4	Extend	Gradient*	00:05
Repeat 2–4 X120			
5	1	10	hold

Table B-9. StEP Secondary PCR mix

Component	Volume / reaction
Template	5
5x GC buffer	20
DMSO	4
dNTPs (10 mM)	2
TrpB_F primer (10 μM)	2
TrpB_R primer (10 μM)	2
Phusion	1
Water	60
Total	100

Table B-10. StEP Secondary PCR thermal cycler protocol

	Step	Temp (°C)	Time (min)
1	Initial denature	98	0:30
2	Denature	98	0:10
3	Anneal	55	0:15
4	Extend	72	0:30
Repeat 2–4 X30			
5	Final extension	68	10:00
6	Hold	10	∞

B.2.6 Library expression and screening

Individual colonies were grown in 300 μ L TB_{carb} in deep-well 96-well polypropylene plates sealed with EasyApp Microporous Film and grown overnight at 37 °C, 250 rpm, and 80% humidity. The following day, 20 μ L overnight culture were used to inoculate 630 μ L TB_{carb} cultures in deep-well 96-well plates and grown at 37 °C and 250 rpm. After 2 h, cultures were chilled on ice for 20–30 minutes and protein expression was induced upon addition of 50 μ L IPTG (final conc. 1 mM) diluted in TB_{carb}. Cultures were shaken at 20 °C and 250 rpm for 20–24 h, after which they were subjected to centrifugation at 5,000g for 5–10 min. The cell pellets could be frozen at –20 °C until further use or used immediately.

Pellets were resuspended in 300 μ L standard lysis buffer, heat treated for >1 h in a 75 °C water bath, and then the heat-treated lysate clarified by centrifugation at 5,000g for 10 minutes. To a glass-lined 96-well deep-well plate charged with 10 μ L of propiophenone or 2-fluoroacetophenone dissolved in DMSO (final conc. 10 mM nucleophile, 5% v/v DMSO), 100 μ L heat-treated lysate were transferred using a Microlab NIMBUS96 liquid handler, followed by the addition of 90 μ L serine (final conc. 10 mM) in 50 mM KPi. Reactions were sealed with silicone ArctiSeal™ and incubated in a 55 °C water bath. After 18–24 h (overnight), reactions were quenched with 600 μ L 2:1 acetonitrile/1M HCl and centrifuged at 5,000g for 10 minutes. For MISER reactions, an acetonitrile/1M HCl mixture was supplemented with 10 mM benzimidazole as internal standard. The supernatant was transferred with the liquid handling robot to a wwPTFE 96-well filter atop a polypropylene HPLC plate and centrifuged at 5,000g for 10 minutes. Each well was analyzed by HPLC-

MS, and improved variants ('hits') were defined as wells that exhibited higher activity relative to activity exhibited by parent expressed on the same plate.

All hits were validated using the procedure described in **Section B.2.7**.

B.2.7 Small-scale reactions with heat-treated lysate

Potential hits were streaked out on agar plates, and single colonies were picked to inoculate 5 mL overnight cultures. The next day, the cultures were used to inoculate expression cultures, miniprepped for sequencing, and prepared as glycerol stocks. In 250-mL Erlenmeyer flasks, 50 mL culture were inoculated with 500 μ L overnight culture, each, and grown for 2.5–5 hours at 250 rpm. The cells were placed on ice for 20–30 minutes and induced with a final concentration of 1 mM IPTG. Expression proceeded at 20 °C and 250 rpm for approximately 24 h. Cells were harvested by centrifugation at 10,000g for 20 minutes at 4 °C, and the supernatant was decanted. The pellet was stored at –20 °C until further use or used immediately. The cell pellet was resuspended in 4 mL/g standard lysis buffer and lysed for >1 h in water bath at 75 °B. The heat-treated lysate was clarified by centrifugation for 20 min at 14,000g and 4 °B. Protein concentration was roughly approximated using Bradford assay.

A 2-mL HPLC vial was charged with 10 μ L propiophenone or 2-fluoroacetophenone dissolved in DMSO (final conc. 10 mM), 100 μ L heat-treated lysate, and 90 μ L serine (final conc. 10 mM) in 50 mM KPi. The reactions were incubated in a water bath at 55 °C. After 18–24 h (overnight), reactions were quenched with 600 μ L 2:1 acetonitrile/1M HCl. Reactions were transferred to 1.7-mL tube and centrifuged 14,000g for 10 minutes. The supernatant was analyzed by HPLC-MS.

B.2.8 Small-scale reactions with purified protein

All analytical reactions were performed in 2-mL glass HPLC vials charged with 10 μ L nucleophile (final concentration variable) dissolved in DMSO (5% v/v), followed by the addition of serine (final concentration 50–100 mM), PLP (final concentration 200 μ M), and purified protein (concentration adjusted depending on desired catalyst loading) diluted in 50

mM KPi buffer to a final volume of 200 μ L. Reactions were incubated at 55 °B. The reaction was then quenched by the addition of either 2:1 CH₃CN/H₂O + 8 mM caffeine or 2:1 CH₃CN/1M HCl. Reactions measuring selectivity (**Figure 3-2 A-B**) were quenched after 20 minutes to prevent racemization. Reactions measuring total turnover number (TTN) were quenched after 24 hours. The 2F-acetophenone product was found to decompose slowly with the addition of acid, therefore the workup mixture was adjusted to be 1:1 acetonitrile/ddiH₂O + 8 mM caffeine (internal standard). This workup is not recommended for regular use as it is not as effective at crashing out protein and can result in aggregation of protein on the HPLC column. Reactions were transferred to 1.7-mL tubes and centrifuged 14,000g for 10 minutes. The supernatant was analyzed by HPLC-MS.

Table B-11. Purified protein reaction conditions

Fig.	[Enzyme] (μ M)	[Serine] (mM)	Nucleophile	[Nucleophile] (mM)	Max TTN
2A	100	50	Propiophenone	50	500
2B	5	50	2-fluoroacetophenone	50	10,000
2C	10	50	Indole	10	1,000
			2-fluoroacetophenone	10	1,000
2D	100 (Tm9D8*), 10 (TmE8)	100	2-fluoroacetophenone	1, 2.5, 5, 10, 25, 50, 100	Varied
3B	10	50	2-fluoroacetophenone	50	10,000
3D	Varied	50	2-fluoroacetophenone	50	Varied
S4	100	50	2-fluoroacetophenone	50	500
			2-fluoroacetophenone-2,2- <i>d</i> ₂	50	500
	100		Acetophenone		500
S9	5	50	2-cyanoacetophenone	50	10,000
	100		Benzoylacetone		500

For Figure 3-3D experiment progress curve fitting:

We can fit the curves to the following equation

$$P(t) = \Delta A(1 - e^{-kt})$$

where P is mM product, t is time in minutes, and k is the observed rate constant. ΔA is a constant that describes the change in amplitude of the exponential function, which will change as a function of [substrate]. To calculate the initial rate of product formation (at 50 mM [substrate]), k is multiplied by ΔA .

Table B-12. Progress curve fit parameters for Figure 3-3D

Enzyme concentration	ΔA	k	V_0 (mM/min)
10 μM	24.7	0.00257	0.0635
50 μM	44.3	0.00523	0.232
75 μM	42.6	0.00750	0.319
100 μM	41.2	0.00969	0.400
150 μM	41.4	0.0123	0.509

B.2.9 Sodium cyanoborohydride reduction of 2-fluoroacetophenone product

Two mL HPLC vials were charged with 10 μL 2-fluoroacetophenone dissolved in DMSO (final concentration 50 mM), followed by the addition of serine (final concentration 50 mM), PLP (final concentration 200 μM), and purified TmE8 (100 μM final concentration) diluted in 50 mM KPi buffer to a final volume of 200 μL . Reactions were incubated at 55 $^{\circ}\text{C}$ for 20 minutes. To one vial, 200 μL of 1M HCl were added, followed by the addition of 10 μL 1M sodium cyanoborohydride in 6.25M NaOH solution. To the control vial, 210 μL of 50 mM KPi buffer was added. The vials were incubated at 55 $^{\circ}\text{C}$ for 2 minutes, then reactions were quenched with 600 μL 2:1 acetonitrile:1 M HCl. Reactions were transferred to 1.7-mL tubes and centrifuged 14,000g for 10 minutes. The supernatant was analyzed by HPLC-MS.

B.2.10 Measurement of proton/deuteron exchange rates

A 2-mL HPLC vial was charged with 10 μL ketone dissolved in DMSO (final concentration 50 mM except 2-Phenylacetophenone, which was run at 1 mM due to solubility issues), 190 μL either protonated or deuterated 50 mM KPi, pH 8.0. Reactions were prepared in duplicate, quenched by the extraction of ketone into 600 μL 1:1 hexanes/ethyl acetate, and measured by GSMS. The amount of isotopically labeled ketone species was estimated by the extracted area of the ion mass and corrected for natural isotopic abundance.

The curves can be fit to an exponential:

$$C_{isotope}(t) = \Delta A(e^{-kt})$$

where $C_{isotope}$ is the concentration of the isotopically labeled substrate (H^2 for the protonated substrate and D^2 for the deuterated substrate), t is time in minutes, and k is the observed rate constant. ΔA is a constant that describes the change in amplitude of the exponential function,

which will change as a function of [substrate]. To calculate your initial rate of exchange (at 50 mM [substrate]), you simply multiply k by ΔA .

Table B-13. Rate of proton exchange

Condition	ΔA	k_{exg} (min^{-1})	V_0 (mM/min)
Propiophenone	50.0	0.000586	0.293
Acetophenone	49.8	0.00414	0.206
2-Fluoroacetophenone	50.0	0.0147	0.735
2-Phenylacetophenone	n.d. ^[a]	0.0515	n.d. ^[a]

^[a] ΔA and V_0 not determined because of ketone solubility issues.

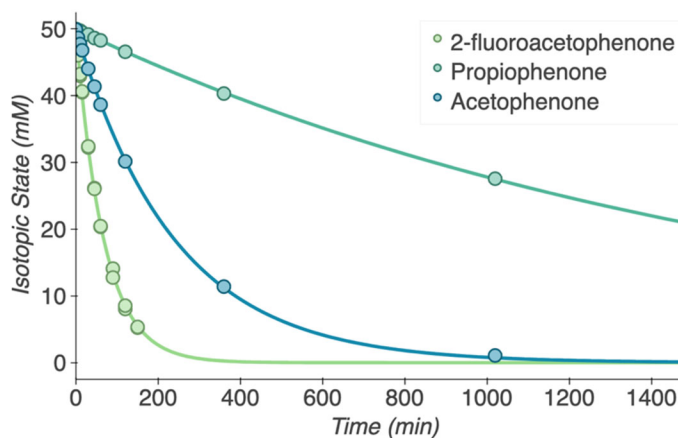


Figure B-11. Exchange curves of ketones in deuterated KPi buffer.

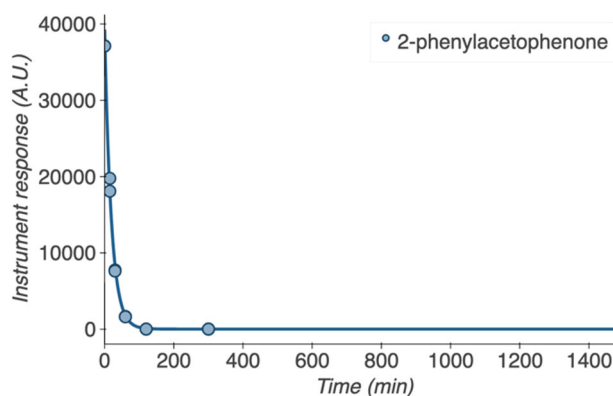


Figure B-12. 2-Phenylacetophenone exchange curve in deuterated KPi buffer.

B.3 Enzyme lineage

Table B-14. Summary of directed evolution for enolate activity

Round	Parent(s)	Library	Screening substrate	Mutations	Evolved variant name	Culture collection #
1	Tm9D8*	Site saturation mutagenesis 105X, 162X, 184X, 228X, 292X		F184H	TmE1	4981
2	TmE1	Random mutagenesis	Propiophenone	I174T	TmE2	4980
3	TmE2	Random mutagenesis		F41Y	TmE3	4981
4	TmE3	Random mutagenesis		K39I, E40G, P54Q, I195T	TmE4	4982
		Site saturation mutagenesis		N86G	TmE5	4983
5	TmE4	85X, 86X, 87X, 117X, 118X, 119X, 198X, 199X, 200X, 201X,		A87D	TmE5.2	
6A	TmE5.1	Site saturation mutagenesis N86X		-	-	
	TmE5.2	Site saturation mutagenesis A87X	2F-acetophenone	N86S	TmE6.1	
				T100P	TmE6.2	
				I166V	TmE6.3	4984
				T179A	TmE6.4	
6B	TmE5.1	Random mutagenesis		I223V, S228G	TmE6.5	
				S228G	TmE6.6	4985
	TmE5.2	Random mutagenesis		-	-	
	TmE6.1	Random mutagenesis		A169T	TmE6.7	
7	TmE5.1	StEP recombination of all previous variants*		I166V, S228G	TmE7	4986
8	TmE7	Alanine scanning		H184A	TmE8	4987

*Variants in grey did not make final lineage but were used in round 7 StEP recombination

Variant sequences

Variants identified through screening were DNA sequenced using Sanger Sequencing (Laragen) to determine their identities. The DNA sequences of all TrpB genes tested in this paper are included here. All variants were cloned into a pET22b(+) vector as described above. T7 and T7-terminator primers were used to sequence all variants.

Table B-15. Sequencing primers

Primer	Direction	Sequence
T7	Forward (5' to 3')	taatacgactcactataggg
T7-term	Reverse (3' to 5')	ctagtattgctcagcgggtg

Pf7E6

Mutations: -

Arnold Lab Culture Collection (ALCC): 4699

```
ATGTGGTTCCGGTGAATTTGGTGGTCCAGTACGTGCCAGAAACGCTGGTAGGACCCTGAAAGAGCTGGAAAAAGCTTACAA
ACGTTTCAAAGATGACGAAGAGTTCAATCGTCAGCTGAATTACTACCTGAAAACCTGGGCAGGTCGTCCAACCCCCTGT
ACTACGCAAAACGCCTGACTGAAAAATCGGTGGTGTAAAATATACCTGAAACGTGAAGACCTGGTTCACGGTGGTGC
ACACAAGACCAACAACGCCATCGGTCAGGCACCGCTGGCAAAGCTCATGGGTAAAACCTGCTGATCGCTGAGACCGGT
GCTGGTCAGCACGGCGTAGCGACTGCAATGGCTGGTGCAGTCTGGGCATGAAAGTGGACATTTACATGGGTGCTGAGGA
CGTAGAACGTCAGAAAAATGAACGTATTCCGTATGAAGCTGCTGGGTGCAAACGTAATTCCAGTTAACTCCGGTTCTCGCA
CCGCGAAAAGACGCAATCAACGAGGCTCTGCGTGATTGGGAAGCTACTTTTGAATACACCCACTACCTAATCGGTTCCGTG
GTCGGTCCACATCCGTATCCGACCATCGTTGCTGATTTCAGTCTGTTATCGGTCGTGAGGCTAAAGCGCAGATCCTGGAG
GCTGAAGGTCAGCTGCCAGATGTAATCGTTGCTTGTGTTGGTGGTGGCTCTAACGCGATGGGTATCTTTACCCGTTCCGTG
AACGACAAAAAGTTAAGCTGGTTGGCGTTGAGGCTGGTGGTAAAGGCTGGAATCTGGTAAGCATTCCGCTAGCCTGAA
CGCAGGTCAGGTTGGTGTGTTGCATGGCATGTGTCCTACTTTCTGCAGGACGAAGAAGGTCAGATCAAACCAAGCCACT
CCATCGCACCAGGTTGGATTATCCAGGTGTGGTCCAGAACACGCTTACTGAAAAAAATTCAGCGTGTGAATACGTG
ACAGTAAACCGATGAAGAAGCACTGAAAGCGTTCATGAACTGAGCCGTACCGAAGGATCATCCCAGCTCTGGAATCTGC
GCATGCTGTGGCTTACGCTATGAAACTGGCTAAGGAAATGTCTCGTGATGAGATCATCATCGTAAACCTGTCTGGTCCGTG
GTGACAAAAGACCTGGATATTGTCTGAAAAGCATCTGGCAACGTGCTcgagcaccaccaccaccactga
```

*Tm9D8**

Mutations: -

Arnold Lab Culture Collection (ALCC): 4907

```
ATGAAAGGCTACTTCCGGTCCGTACGGTGGCCAGTACGTGCCGAAATCCTGATGGGAGCTCTGGAAGAACTGGAAGCTGC
GTACGAAGGAATCATGAAAGATGAGTCTTTCTGGAAAAGAAATCAATGACCTGCTGCGGATTATGCGGGTCTGCCACTC
CGCTGACTTCCGACGTCGCTGTCCGAAAAATACGGTGTCTCGGTATATCTGAAACGTGAAGACCTGCTGCATACTGGTG
CGCATAAAAATCAATAACGCTATCGGCCAGGTTCTGCTGGCAAAACTAATGGGCAAAACCCGTATCATTGCTGAAACGGGT
GCTGGTCAGCACGGCGTAGCAACTGTACCGCAGCAGCGCTGTTCCGGTATGGAATGTGTAATCTATATGGGCGAAGAAGA
CACGATCCGCCAGAAAATAAACGTTGAACGTATGAAACTGCTGGGTGCTAAAAGTTGTACCAGTAAAAATCCGGTAGCCGTA
CCCTGAAAGACGCAATTGACGAAGCTCTGCGTACTGGATTACCAACCTGCAGACCACCTATTACGTGTTCCGGCTCTGTG
GTTGGTCCGCATCCATATCCGATTATCGTACGTAACCTCAAAGGTTATCGGCGAAGAGACCAAAAAACAGATTCCAGA
AAAAGAAGGCCGCTGCCGACTACATCGTTGCGTGGTGGTGAGCGGTGGTTCTAACGCTGCCGTTACTTCTATCCGTTTAT
CGATTCTGGTGTGAAAGCTGATCGGCCTAGAAGCCGGTGGCGAAGGTTGGAACCGGTAAACATGCGGCTTCTCTGTGA
AAGGTAATCCGGCTACCTGCACGGTCTAAGACGTTCTGCTGAGGATGACTGGGGTCAAGTTCAGGTGAGCCACTCC
GTCTCCGCTGGCCTGGACTACTCCGGTGTCCGGTCCGGAACACGCTTATTGGCGTGAGACCGGTAAAAGTGTGTACGATGC
TGTGACCGATGAAGAAGCTCTGGACGCATTATCGAACTGTCTGCGCTGGAAGGCATCATCCCAGCCCTGGAGTCTTCTC
ACGCACTGGCTTATCTGAAGAAGATCAACATCAAGGGTAAAGTTGTGGTGGTTAATCTGTCTGGTGGTGTGACAAGGAT
CTGGAATCTGTACTGAACCCCGTATGTTCCGGAACGCATCCGCCctgagcaccaccaccaccactga
```


TmE1:

Mutations: F184H

Arnold Lab Culture Collection (ALCC): 4979

ATGAAAGGCTACTTCGGTCCGTACGGTGGCCAGTACGTGCCGAAATCCTGATGGGAGCTCTGGAAGAACTGGAAGCTGC
 GTACGAAAGGAATCATGAAAGATGAGTCTTTCTGGAAAGAATTCAATGACCTGCTGCGGATTATGCGGGTCTCCGACTC
 CGCTGACTTCGCACGTCGTCTGTCCGAAAAATACGGTGTCTCGCGTATATCTGAAACGTGAAGACCTGCTGCATACTGGTG
 CGCATAAAATCAATAACGCTATCGGCCAGGTTCTGCTGGCAAACTAATGGGCAAAACCCGTATCATTGCTGAAACGGGT
 GCTGGTCAGCACGGCGTAGCAACTGTACCGCAGCAGCGCTGTTCGGTATGGAATGTGTAATCTATATGGGCGAAGAAGA
 CACGATCCGCCAGAACTAAACGTTGAACGTATGAAACTGCTGGGTGCTAAAGTTGTACCGGTAAAAATCCGGTAGCCGTA
 CCCTGAAAGACGCAATTGACGAAGCTCTGCGTACTGGATTACCAACCTGCAGACCACCTATTACGTG**CAT**GGCTCTGTG
 GTTGGTCCGCATCCATATCCGATTATCGTACGTAACCTCCAAAAGGTTATCGGCGAAGAGACCAAAAAACAGATTCCAGA
 AAAAGAAGGCCGTCTGCCGACTACATCGTTGCGTGCGTGAGCGGTGGTTCTAACGCTGCCGGTATCTTCTATCCGTTTAT
 CGATTCTGGTGTGAAGCTGATCGGCGTAGAAGCCGGTGGCGAAGGTCTGGAACCCGGTAAACATGCGGCTTCTCTGCTGA
 AAGGTAATTCGGCTACTGCACGGTCTAAGACGTTTCGTTCTGCAGGATGACTGGGGTCAAGTTCAGGTGAGCCACTCC
 GTCTCCGCTGGCCTGGACTACTCCGGTGTCCGGTCCGGAACACGCCTATTGGCGTGAGACCCGGTAAAGTGTGTACGATGC
 TGTACCGGATGAAGAAGCTCTGGACGATTTCATCGAAGCTGTCTCCGCTGGAAGGCATCATCCAGCCCTGGAGTCTTCTC
 ACGCACTGGCTTATCTGAAGAAGATCAACATCAAGGGTAAAGTTGTGGTGGTTAATCTGTCTGGTCTGGTGACAAGGAT
 CTGGAATCTGACTGAACCACCCGTATGTTCCGGAACGCATCCGCCctgagcaccaccaccaccactga

TmE2:

Mutations: I174T

Arnold Lab Culture Collection (ALCC): 4980

ATGAAAGGCTACTTCGGTCCGTACGGTGGCCAGTACGTGCCGAAATCCTGATGGGAGCTCTGGAAGAACTGGAAGCTGC
 GTACGAAAGGAATCATGAAAGATGAGTCTTTCTGGAAAGAATTCAATGACCTGCTGCGGATTATGCGGGTCTCCGACTC
 CGCTGACTTCGCACGTCGTCTGTCCGAAAAATACGGTGTCTCGCGTATATCTGAAACGTGAAGACCTGCTGCATACTGGTG
 CGCATAAAATCAATAACGCTATCGGCCAGGTTCTGCTGGCAAACTAATGGGCAAAACCCGTATCATTGCTGAAACGGGT
 GCTGGTCAGCACGGCGTAGCAACTGTACCGCAGCAGCGCTGTTCGGTATGGAATGTGTAATCTATATGGGCGAAGAAGA
 CACGATCCGCCAGAACTAAACGTTGAACGTATGAAACTGCTGGGTGCTAAAGTTGTACCGGTAAAAATCCGGTAGCCGTA
 CCCTGAAAGACGCAATTGACGAAGCTCTGCGTACTGG**ACT**ACCAACCTGCAGACCCTATTACGTGCATGGCTCTGTG
 GTTGGTCCGCATCCATATCCGATTATCGTACGTAACCTCCAAAAGGTTATCGGCGAAGAGACCAAAAAACAGATTCCAGA
 AAAAGAAGGCCGTCTGCCGACTACATCGTTGCGTGCGTGAGCGGTGGTTCTAACGCTGCCGGTATCTTCTATCCGTTTAT
 CGATTCTGGTGTGAAGCTGATCGGCGTAGAAGCCGGTGGCGAAGGTCTGGAACCCGGTAAACATGCGGCTTCTCTGCTGA
 AAGGTAATTCGGCTACTGCACGGTCTAAGACGTTTCGTTCTGCAGGATGACTGGGGTCAAGTTCAGGTGAGCCACTCC
 GTCTCCGCTGGCCTGGACTACTCCGGTGTCCGGTCCGGAACACGCCTATTGGCGTGAGACCCGGTAAAGTGTGTACGATGC
 TGTGACCGATGAAGAAGCTCTGGACGATTTCATCGAAGCTGTCTCGCCTGGAAGGCATCATCCAGCCCTGGAGTCTTCTC
 ACGCACTGGCTTATCTGAAGAAGATCAACATCAAGGGTAAAGTTGTGGTGGTTAATCTGTCTGGTCTGGTGACAAGGAT
 CTGGAATCTGACTGAACCACCCGTATGTTCCGGAACGCATCCGCCctgagcaccaccaccaccactga

TmE3:

Mutations: F41Y

Arnold Lab Culture Collection (ALCC): 4981

ATGAAAGGCTACTTCGGTCCGTACGGTGGCCAGTACGTGCCGGAATCCTGATGGGAGCTCTGGAAGAACTGGAAGCTGC
 GTACGAAAGGAATCATGAAAGATGAGTCTTTCTGGAAAGAA**TACA**ATGACCTGCTGCGCGATTATGCGGGTTCGTCGACTC
 CGCTGACTTCGCACGTCGTCTGTCCGAAAAGTACGGTGCCTCGCGTATATCTGAAACGTGAAGACCTGCTGCATACTGGTG
 CGCATAAAATCAATAACGCTATCGGCCAGGTTCTGCTGGCAAACTAATGGGCAAAACCCGTATCATTGCTGAAACGGGT
 GCTGGTCAGCACGGCGTAGCAACTGCTACCGCAGCAGCGCTGTTCCGGTATGGAATGTGTAATCTATATGGGCGAAGAAGA
 CACGATCCGCCAGAACTAAACGTTGAACGTATGAAACTGCTGGGTGCTAAAGTTGTACCGGTAAAATCCGGTAGCCGTA
 CCCTGAAAGACGCAATTGACGAAGCTCTGCGTGACTGGACTACCAACCTGCAGACCACCTATTACGTGCATGGCTCTGTG
 GTTGGTCCGCATCCATATCCGATTATCGTACGTAACCTCCAAAAGGTTATCGGCGAAGAGACCAAAAAACAGATTCCAGA
 AAAAGAAGGCCGTGCGCGACTACATCGTTGCGTGCGTGAGCGGTGGTTCTAACGCTGCCGGTATCTTCTATCCGTTTAT
 CGATTCTGGTGTGAAGCTGATCGGCGTAGAAGCCGGTGGCGAAGGTCTGGAACCCGGTAAACATGCGGCTTCTCTGCTGA
 AAGGTAAAATCGGCTACTGCACGGTCTAAGACGTTTCGTTCTGCAGGATGACTGGGGTCAAGTTCAGGTGAGCCACTCC
 GTCTCCGCTGGCCTGGACTACTCCGGTGTCCGGTCCGGAACACGCCTATTGGCGTGAGACCCGGTAAAGTGTGTACGATGC
 TGTACCCGATGAAGAAGCTCTGGACGATTATCGAAGCTGCTCGCCTGGAAGGCATCATCCAGCCCTGGAGTCTTCTC
 ACGCACTGGCTTATCTGAAGAAGATCAACATCAAGGGTAAAGTTGTGGTGGTTAATCTGTCTGGTCTGGTGACAAGGAT
 CTGGAATCTGACTGAACCACCCGTATGTTCCGGAACGCATCCGCCctgagcaccaccaccaccactga

TmE4:

Mutations: K39I, E40G, P54Q, I195T

Arnold Lab Culture Collection (ALCC): 4982

ATGAAAGGCTACTTCGGTCCGTACGGTGGCCAGTACGTGCCGGAATCCTGATGGGAGCTCTGGAAGAACTGGAAGCTGC
 GTACGAAAGGAATCATGAAAGATGAGTCTTTCTGG**ATAGGA**TACAATGACCTGCTGCGCGATTATGCGGGTTCGTCGACTC
AGCTGACTTCGCACGTCGTCTGTCCGAAAAGTACGGTGCCTCGCGTATATCTGAAACGTGAAGACCTGCTGCATACTGGT
 GCGCATAAAATCAATAACGCTATCGGCCAGGTTCTGCTGGCAAACTAATGGGCAAAACCCGTATCATTGCTGAAACGGG
 TGCTGGTTCAGCACGGCGTAGCAACTGCTACCGCAGCAGCGCTGTTCCGGTATGGAATGTGTAATCTATATGGGCGAAGAAG
 ACACGATCCGCCAGAACTAAACGTTGAACGTATGAAACTGCTGGGTGCTAAAGTTGTACCGGTAAAATCCGGTAGCCGT
 ACCCTGAAAGACGCAATTGACGAAGCTCTGCGTGACTGGACTACCAACCTGCAGACCACCTATTACGTGCATGGCTCTGT
 GGTGGTCCGCATCCATATCCG**ACT**ATCGTACGTAACCTCCAAAAGGTTATCGGCGAAGAGACCAAAAAACAGATTCCAG
 AAAAGAAGGCCGCTGCGGACTACATCGTTGCGTGCGTGAGCGGTGGTTCTAACGCTGCCGGTATCTTCTATCCGTTA
 TCGATTCTGGTGTGAAGCTGATCGGCGTAGAAGCCGGTGGCGAAGGTCTGGAACCCGGTAAACATGCGGCTTCTCTGCTG
 AAAGGTAAAATCGGCTACTGCACGGTCTAAGACGTTTCGTTCTGCAGGATGACTGGGGTCAAGTTCAGGTGAGCCACTC
 CGTCTCCGCTGGCCTGGACTACTCCGGTGTCCGGTCCGGAACACGCCTATTGGCGTGAGACCCGGTAAAGTGTGTACGATG
 CTGTGACCGATGAAGAAGCTCTGGACGATTATCGAAGCTGCTCGCCTGGAAGGCATCATCCAGCCCTGGAGTCTTCTC
 ACGCACTGGCTTATCTGAAGAAGATCAACATCAAGGGTAAAGTTGTGGTGGTTAATCTGTCTGGTCTGGTGACAAGGAT
 CTGGAATCTGACTGAACCACCCGTATGTTCCGGAACGCATCCGCCctgagcaccaccaccaccactga

TmE5:

Mutations: N86G

Arnold Lab Culture Collection (ALCC): 4983

ATGAAAGGCTACTTCGGTCCGTACGGTGGCCAGTACGTGCCGAAATCCTGATGGGAGCTCTGGAAGAACTGGAAGCTGC
 GTACGAAGGAATCATGAAAGATGAGTCTTTCTGGATAGGATACAATGACCTGCTGCGCGATTATGCGGGTCTGCCGACTC
 AGCTGTACTIONCGCACGTCTGTCCGAAAAGTACGGTGTCTCGCGTATACTGAAACGTGAAGACCTGCTGCATACTGGT
 GCGCATAAAAATCAATGGGCTATCGGCCAGGTTCTGCTGGCAAACCTAATGGGCAAACCCGATCATTGCTGAAACGG
 GTGCTGGTCAGCACGGCGTAGCAACTGCTACCGCAGCAGCGCTGTTCGGTATGGAATGTGTAATCTATATGGGCGAAGAA
 GACACGATCCGCCAGAACTAAACGTTGAACGTATGAACTGCTGGGTGCTAAAGTTGTACCGGTAATAATCCGGTAGCCG
 TACCCTGAAAGACGCAATTGACGAAGCTCTGCGTACTGGACTACCAACCTGCAGACCACCTATTACGTGCATGGCTCTG
 TGGTTGGTCCGCATCCATATCCGACTATCGTACGTAACCTCCAAAAGGTTATCGGCGAAGAGACCAAAAAACAGATTCCA
 GAAAAAGAAGGCCGCTGCCGACTACATCGTTGCGTGCCTGAGCGGTGGTTCTAACGCTGCCGGTATCTTCTATCCGTTT
 ATCGATTCTGGTGTGAAGCTGATCGGCGTAGAAGCCGGTGGCGAAGGTTCTGGAACCCGGTAAACATGCGGCTTCTCTGCT
 GAAAGGTAATAATCCGGTACCTGCACGGTTCTAAGACGTTCTGCTGAGGATGACTGGGGTCAAGTTCAGGTGAGCCACT
 CCGTCTCCGCTGGCCTGGACTACTCCGGTGTCCGGTCCGGAACACGCCTATTGGCGTGAGACCCGGTAAAGTGTGTACGAT
 GCTGTGACCGATGAAGAAGCTCTGGACGCATTATCGAAGCTGCTCGCTGGAAGGCATATCCAGCCCTGAGACTCTTC
 TCACGCACTGGCTTATCTGAAGAAGATCAACATCAAGGGTAAAGTTGTGGTGGTTAATCTGTCTGGTCTGGTGACAAGG
 ATCTGGAATCTGTACTGAACCACCCGATGTTTCGCGAACGCATCCGCCctgagaccaccaccaccactga

TmE5.2:

Mutations: A87D

ATGAAAGGCTACTTCGGTCCGTACGGTGGCCAGTACGTGCCGAAATCCTGATGGGAGCTCTGGAAGAACTGGAAGCTGC
 GTACGAAGGAATCATGAAAGATGAGTCTTTCTGGATAGGATACAATGACCTGCTGCGCGATTATGCGGGTCTGCCGACTC
 AGCTGTACTIONCGCACGTCTGTCCGAAAAGTACGGTGTCTCGCGTATACTGAAACGTGAAGACCTGCTGCATACTGGT
 GCGCATAAAAATCAATAACGATATCGGCCAGGTTCTGCTGGCAAACCTAATGGGCAAACCCGATCATTGCTGAAACGG
 GTGCTGGTCAGCACGGCGTAGCAACTGCTACCGCAGCAGCGCTGTTCGGTATGGAATGTGTAATCTATATGGGCGAAGAA
 GACACGATCCGCCAGAACTAAACGTTGAACGTATGAACTGCTGGGTGCTAAAGTTGTACCGGTAATAATCCGGTAGCCG
 TACCCTGAAAGACGCAATTGACGAAGCTCTGCGTACTGGACTACCAACCTGCAGACCACCTATTACGTGCATGGCTCTG
 TGGTTGGTCCGCATCCATATCCGACTATCGTACGTAACCTCCAAAAGGTTATCGGCGAAGAGACCAAAAAACAGATTCCA
 GAAAAAGAAGGCCGCTGCCGACTACATCGTTGCGTGCCTGAGCGGTGGTTCTAACGCTGCCGGTATCTTCTATCCGTTT
 ATCGATTCTGGTGTGAAGCTGATCGGCGTAGAAGCCGGTGGCGAAGGTTCTGGAACCCGGTAAACATGCGGCTTCTCTGCT
 GAAAGGTAATAATCCGGTACCTGCACGGTTCTAAGACGTTCTGCTGAGGATGACTGGGGTCAAGTTCAGGTGAGCCACT
 CCGTCTCCGCTGGCCTGGACTACTCCGGTGTCCGGTCCGGAACACGCCTATTGGCGTGAGACCCGGTAAAGTGTGTACGAT
 GCTGTGACCGATGAAGAAGCTCTGGACGCATTATCGAAGCTGCTCGCTGGAAGGCATATCCAGCCCTGAGACTCTTC
 TCACGCACTGGCTTATCTGAAGAAGATCAACATCAAGGGTAAAGTTGTGGTGGTTAATCTGTCTGGTCTGGTGACAAGG
 ATCTGGAATCTGTACTGAACCACCCGATGTTTCGCGAACGCATCCGCCctgagaccaccaccaccactga

TmE6.1:

Mutations: N86S

ATGAAAGGCTACTTCGGTCCGTACGGTGGCCAGTACGTGCCGAAATCCTGATGGGAGCTCTGGAAGAACTGGAAGCTGC
 GTACGAAGGAATCATGAAAGATGAGTCTTTCTGGATAGGATACAATGACCTGCTGCGCGATTATGCGGGTCTGCCGACTC
 AGCTGTACTIONCGCACGTCTGTCCGAAAAGTACGGTGTCTCGCGTATACTGAAACGTGAAGACCTGCTGCATACTGGT
 GCGCATAAAAATCAATAGTGATATCGGCCAGGTTCTGCTGGCAAACCTAATGGGCAAACCCGATCATTGCTGAAACGG
 GTGCTGGTCAGCACGGCGTAGCAACTGCTACCGCAGCAGCGCTGTTCGGTATGGAATGTGTAATCTATATGGGCGAAGAA
 GACACGATCCGCCAGAACTAAACGTTGAACGTATGAACTGCTGGGTGCTAAAGTTGTACCGGTAATAATCCGGTAGCCG
 TACCCTGAAAGACGCAATTGACGAAGCTCTGCGTACTGGACTACCAACCTGCAGACCACCTATTACGTGCATGGCTCTG
 TGGTTGGTCCGCATCCATATCCGACTATCGTACGTAACCTCCAAAAGGTTATCGGCGAAGAGACCAAAAAACAGATTCCA
 GAAAAAGAAGGCCGCTGCCGACTACATCGTTGCGTGCCTGAGCGGTGGTTCTAACGCTGCCGGTATCTTCTATCCGTTT
 ATCGATTCTGGTGTGAAGCTGATCGGCGTAGAAGCCGGTGGCGAAGGTTCTGGAACCCGGTAAACATGCGGCTTCTCTGCT
 GAAAGGTAATAATCCGGTACCTGCACGGTTCTAAGACGTTCTGCTGAGGATGACTGGGGTCAAGTTCAGGTGAGCCACT
 CCGTCTCCGCTGGCCTGGACTACTCCGGTGTCCGGTCCGGAACACGCCTATTGGCGTGAGACCCGGTAAAGTGTGTACGAT
 GCTGTGACCGATGAAGAAGCTCTGGACGCATTATCGAAGCTGCTCGCTGGAAGGCATATCCAGCCCTGAGACTCTTC
 TCACGCACTGGCTTATCTGAAGAAGATCAACATCAAGGGTAAAGTTGTGGTGGTTAATCTGTCTGGTCTGGTGACAAGG
 ATCTGGAATCTGTACTGAACCACCCGATGTTTCGCGAACGCATCCGCCctgagaccaccaccaccactga

TmE6.2:

Mutations: T100P

ATGAAAGGCTACTTCGGTCCGTACGGTGGCCAGTACGTGCCGAAATCCTGATGGGAGCTCTGGAAGAACTGGAAGCTGC
 GTACGAAGGAATCATGAAAGATGAGTCTTTCTGGATAGGATACAATGACCTGCTGCGCGATTATGCGGGTCGTCGGACTC
 AGCTGTACTTCGCACGTCGCTGTCCGAAAAGTACGGTGTCTCGGTATAATCTGAAACGTGAAGACCTGCTGCATACTGGT
 GCGCATAAAATCAATGGGGCTATCGGCCAGGTTCTGCTGGCAAACTAATGGGCAAACCCCGTATCATTGCTGAAACGG
 GTGCTGGTCAGCACGGCGTAGCAACTGCTACCGCAGCAGCGCTGTTCCGGTATGGAATGTGTAATCTATATGGGCGAAGAA
 GACACGATCCGCCAGAACTAAACGTTGAACGTATGAAACTGCTGGGTGCTAAAGTTGTACCGGTAAAATCCGGTAGCCG
 TACCCTGAAAGACGCAATTGACGAAGCTCTGCGTGACTGGACTACCAACCTGCAGACCACCTATTACGTGCATGGCTCTG
 TGGTTGGTCCGCATCCATATCCGACTATCGTACGTAACCTCCAAAAGGTTATCGGCGAAGAGACCAAAAAACAGATTCCA
 GAAAAAGAAAGCCCGCTGCCGACTACATCGTTGCGTGCGTGAGCGGTGGTTCTAACGCTGCCGGTATCTTCTATCCGTTT
 ATCGATTCTGGTGTGAAGCTGATCGGCGTAGAAGCCGGTGGCGAAGGTCTGGAACCGGTAAACATCGCGCTTCTCTGCT
 GAAAGGTAAAATCCGGTACCTGCACGGTTCTAAGACGTTCTGCTGAGGATGACTGGGGTCAAGTTCAGGTGAGCCACT
 CCGTCTCCGCTGGCCTGGACTACTCCGGTGTGGTCCGGAACACGCCTATTGGCGTGAGACCGGTAAAGTGTGTACGAT
 GCTGTGACCGATGAAGAAGCTCTGGACGCATTCATCGAACTGTCTCGCCTGGAAGGCATCATCCAGCCCTGGAGTCTTC
 TCACGCACTGGCTTATCTGAAGAAGATCAACATCAAGGGTAAAGTTGTGGTGGTTAATCTGTCTGGTCTGGTGACAAGG
 ATCTGGAATCTGTACTGAACCACCCGTATGTTCCGGAACGCATCCGCCctgagcaccaccaccaccactga

TmE6.3:

Mutations: I166V

Arnold Lab Culture Collection (ALCC): 4984

ATGAAAGGCTACTTCGGTCCGTACGGTGGCCAGTACGTGCCGAAATCCTGATGGGAGCTCTGGAAGAACTGGAAGCTGC
 GTACGAAGGAATCATGAAAGATGAGTCTTTCTGGATAGGATACAATGACCTGCTGCGCGATTATGCGGGTCGTCGGACTC
 AGCTGTACTTCGCACGTCGCTGTCCGAAAAGTACGGTGTCTCGGTATAATCTGAAACGTGAAGACCTGCTGCATACTGGA
 GCGCATAAAATCAATGGGGCTATCGGCCAGGTTCTGCTGGCAAACTAATGGGCAAACCCGTATCATTGCTGAAACGGG
 TGCTGGTCAGCACGGCGTAGCAACTGCTACCGCAGCAGCGCTGTTCCGGTATGGAATGTGTAATCTATATGGGCGAAGAAG
 ACACGATCCGCCAGAACTAAACGTTGAACGTATGAAACTGCTGGGTGCTAAAGTTGTACCGGTAAAATCCGGTAGCCGT
 ACCCTGAAAGACGCA~~GTT~~GACGAAGCTCTGCGTGACTGGACTACCAACCTGCAGACCACCTATTACGTGCATGGCTCTGT
 GGTGGTCCGCATCCATATCCGACTATCGTACGTAACCTCCAAAAGGTTATCGGCGAAGAGACCAAAAAACAGATTCCAG
 AAAAAGAAGGCCCGCTGCCGACTACATCGTTGCGTGCGTGAGCGGTGGTTCTAACGCTGCCGGTATCTTCTATCCGTTTA
 TCGATTCTGGTGTGAAGCTGATCGGCGTAGAAGCCGGTGGCGAAGGTCTGGAACCGGTAAACATGCGGCTTCTCTGCTG
 AAAGGTAAAATCCGGTACCTGCACGGTTCTAAGACGTTCTGCTGAGGATGACTGGGGTCAAGTTCAGGTGAGCCACTC
 CGTCTCCGCTGGCTGGACTACTCCGGTGTGGTCCGGAACACGCCTATTGGCGTGAGACCGGTAAAGTGTGTACGATG
 CTGTGACCGATGAAGAAGCTCTGGACGCATTCATCGAACTGTCTCGCCTGGAAGGCATCATCCAGCCCTGGAGTCTTCTC
 ACGCACTGGCTTATCTGAAGAAGATCAACATCAAGGGTAAAGTTGTGGTGGTTAATCTGTCTGGTCTGGTGACAAGGAT
 CTGGAATCTGTACTGAACCACCCGTATGTTCCGGAACGCATCCGCCctgagcaccaccaccaccactga

TmE6.4:

Mutations: T179A

ATGAAAGGCTACTTCGGTCCGTACGGTGGCCAGTACGTGCCGAAATCCTGATGGGAGCTCTGGAAGAACTGGAAGCTGC
 GTACGAAGGAATCATGAAAGATGAGTCTTTCTGGATAGGATACAATGACCTGCTGCGCGATTATGCGGGTCGTCGGACTC
 AGCTGTACTTCGCACGTCGCTGTCCGAAAAGTACGGTGTCTCGGTATAATCTGAAACGTGAAGACCTGCTGCATACTGGT
 GCGCATAAAATCAATGGGGCTATCGGCCAGGTTCTGCTGGCAAACTAATGGGCAAACCCGTATCATTGCTGAAACGGG
 TGCTGGTCAGCACGGCGTAGCAACTGCTACCGCAGCAGCGCTGTTCCGGTATGGAATGTGTAATCTATATGGGCGAAGAAG
 ACACGATCCGCCAGAACTAAACGTTGAACGTATGAAACTGCTGGGTGCTAAAGTTGTACCGGTAAAATCCGGTAGCCGT
 ACCCTGAAAGACGCAATTGACGAAGCTCTGCGTGACTGGACTACCAACCTGCAG~~GCC~~ACCTATTACGTGCATGGCTCTGT
 GGTGGTCCGCATCCATATCCGACTATCGTACGTAACCTCCAAAAGGTTATCGGCGAAGAGACCAAAAAACAGATTCCAG
 AAAAAGAAGGCCCGCTGCCGACTACATCGTTGCGTGCGTGAGCGGTGGTTCTAACGCTGCCGGTATCTTCTATCCGTTTA
 TCGATTCTGGTGTGAAGCTGATCGGCGTAGAAGCCGGTGGCGAAGGTCTGGAACCGGTAAACATGCGGCTTCTCTGCTG
 AAAGGTAAAATCCGGTACCTGCACGGTTCTAAGACGTTCTGCTGAGGATGACTGGGGTCAAGTTCAGGTGAGCCACTC
 CGTCTCCGCTGGCTGGACTACTCCGGTGTGGTCCGGAACACGCCTATTGGCGTGAGACCGGTAAAGTGTGTACGATG
 CTGTGACCGATGAAGAAGCTCTGGACGCATTCATCGAACTGTCTCGCCTGGAAGGCATCATCCAGCCCTGGAGTCTTCTC
 ACGCACTGGCTTATCTGAAGAAGATCAACATCAAGGGTAAAGTTGTGGTGGTTAATCTGTCTGGTCTGGTGACAAGGAT
 CTGGAATCTGTACTGAACCACCCGTATGTTCCGGAACGCATCCGCCctgagcaccaccaccaccactga

TmE6.5:

Mutations: I223V, S228G

ATGAAAGGCTACTTCGGTCCGTACGGTGGCCAGTACGTGCCGAAATCCTGATGGGAGCTCTGGAAGAACTGGAAGCTGC
 GTACGAAGGAATCATGAAAGATGAGTCTTCTGGATAGGATACAATGACCTGCTGCGCGATTATGCGGGTCTCCGACTC
 AGCTGTACTTCGCACGTCGCTGTCCGAAAAGTACGGTGTCTCGGTATAATCTGAAACGTGAAGACCTGCTGCATACTGGT
 GCGCATAAAAATCAATGGGGCTATCGGCCAGGTTCTGCTGGCAAACTAATGGGCAAAACCCGTATCATTGCTGAAACGGG
 TGCTGGTACGACACGGCGTAGCAACTGCTACCGCAGCAGCGCTGTTCCGGTATGGAATGTGTAATCTATATGGGCGAAGAAG
 ACACGATCCGCCAGAACTAAACGTTGAACGTATGAAGCTGCTGGGTGCTAAAGTTGTACCGGTAATAATCCGGTAGCCGT
 ACCCTGAAAAGACGCAATTGACGAAGCTCTGCGTACTGGACTACCAACCTGCAGACCACCTATTACGTGCATGGCTCTGT
 GGTGGTCCGCATCCATATCCGACTATCGTACGTAACCTCCAAAAGGTTATCGGCGAAGAGACCAAAAAACAGATTCCAG
 AAAAAGAAGGCCCGCTGCCGGACTACGTCGTTGCGTGGCTGGCGGTTCTAACGCTGCCGGTATCTTCTATCCGTTT
 ATCGATTCTGGTGTGAAGCTGATCGGCGTAGAAGCCGGTGGCGAAGGTTCTGGAACCGGTAACATGCGGCTTCTCTGCT
 GAAAGGTAAAATCGGCTACCTGCACGGTTCTAAGACGTTCTGCTGAGGATGACTGGGGTCAAGTTCAGGTGAGCCACT
 CCGTCTCCGCTGGCCTGGACTACTCCGGTGTGGTCCGGAACACGCCTATTGGCGTGAGACCGGTAAGTGTGTACGAT
 GCTGTGACCGATGAAGAAGCTCTGGACGCATTATCGAACTGTCTCGCCTGGAAGGCATCATCCAGCCCTGGAGTCTTC
 TCACGCACTGGCTTATCTGAAGAAGATCAACATCAAGGGTAAAGTTGTGGTGGTTAATCTGTCTGGTCTGGTGACAAGG
 ATCTGGAATCTGTACTGAACCACCCGTATGTTCCGCAACGCATCCGCCctgagcaccaccaccaccactga

TmE6.6:

Mutations: S228G

Arnold Lab Culture Collection (ALCC): 4985

ATGAAAGGCTACTTCGGTCCGTACGGTGGCCAGTACGTGCCGAAATCCTGATGGGAGCTCTGGAAGAACTGGAAGCTGC
 GTACGAAGGAATCATGAAAGATGAGTCTTCTGGATAGGATACAATGACCTGCTGCGCGATTATGCGGGTCTCCGACTC
 AGCTGTACTTCGCACGTCGCTGTCCGAAAAGTACGGTGTCTCGGTATAATCTGAAACGTGAAGACCTGCTGCATACTGGT
 GCGCATAAAAATCAATGGGGCTATCGGCCAGGTTCTGCTGGCAAACTAATGGGCAAAACCCGTATCATTGCTGAAACGGG
 TGCTGGTACGACACGGCGTAGCAACTGCTACCGCAGCAGCGCTGTTCCGGTATGGAATGTGTAATCTATATGGGCGAAGAAG
 ACACGATCCGCCAGAACTAAACGTTGAGCGTATGAACTGCTGGGTGCTAAAGTTGTACCGGTAATAATCCGGTAGCCGT
 ACCCTGAAAAGACGCAATTGACGAAGCTCTGCGTACTGGACTACCAACCTGCAGACCACCTATTACGTGCATGGCTCTGT
 GGTGGTCCGCATCCATATCCGACTATCGTACGTAACCTCCAAAAGGTTATCGGCGAAGAGACCAAAAAACAGATTCCAG
 AAAAAGAAGGCCCGCTGCCGGACTACATCGTTGCGTGGCTGGCGGTTCTAACGCTGCCGGTATCTTCTATCCGTTT
 ATCGATTCTGGTGTGAAGCTGATCGGCGTAGAAGCCGGTGGCGAAGGTTCTGGAACCGGTAACATGCGGCTTCTCTGCT
 GAAAGGTAAAATCGGCTACCTGCACGGTTCTAAGACGTTCTGCTGAGGATGACTGGGGTCAAGTTCAGGTGAGCCACT
 CCGTCTCCGCTGGCCTGGACTACTCCGGTGTGGTCCGGAACACGCCTATTGGCGTGAGACCGGTAAGTGTGTACGAT
 GCTGTGACCGATGAAGAAGCTCTGGACGCATTATCGAACTGTCTCGCCTGGAAGGCATCATCCAGCCCTGGAGTCTTC
 TCACGCACTGGCTTATCTGAAGAAGATCAACATCAAGGGTAAAGTTGTGGTGGTTAATCTGTCTGGTCTGGTGACAAGG
 ATCTGGAATCTGTACTGAACCACCCGTATGTTCCGCAACGCATCCGCCctgagcaccaccaccaccactga

TmE6.7:

Mutations: A169T

ATGAAAGGCTACTTCGGTCCGTACGGTGGCCAGTACGTGCCGAAATCCTGATGGGAGCTCTGGAAGAACTGGAAGCTGC
 GTACGAAGGAATCATGAAAGATGAGTCTTCTGGATAGGATACAATGACCTGCTGCGCGATTATGCGGGTCTCCGACTC
 AGCTGTACTTCGCACGTCGCTGTCCGAAAAGTACGGTGTCTCGGTATAATCTGAAACGTGAAGACCTGCTGCATACTGGT
 GCGCATAAAAATCAATAGTGATATCGGCCAGGTTCTGCTGGCAAACTAATGGGCAAAACCCGTATCATTGCTGAAACGGG
 TGCTGGTACGACACGGCGTAGCAACTGCTACCGCAGCAGCGCTGTTCCGGTATGGAATGTGTAATCTATATGGGCGAAGAAG
 ACACGATCCGCCAGAACTAAACGTTGAACGTATGAACTGCTGGGTGCTAAAGTTGTACCGGTAATAATCCGGTAGCCGT
 ACCCTGAAAAGACGCAATTGACGAGACTCTGCGTACTGGACTACCAACCTGCAGACCACCTATTACGTGCATGGCTCTGT
 GGTGGTCCGCATCCATATCCGACTATCGTACGTAACCTCCAAAAGGTTATCGGCGAAGAGACCAAAAAACAGATTCCAG
 AAAAAGAAGGCCCGCTGCCGGACTACATCGTTGCGTGGCTGGCGAAGGTTCTAACGCTGCCGGTATCTTCTATCCGTTT
 TCGATTCTGGTGTGAAGCTGATCGGCGTAGAAGCCGGTGGCGAAGGTTCTGGAACCGGTAACATGCGGCTTCTCTGCTG
 AAAGGTAAAATCGGCTACCTGCACGGTTCTAAGACGTTCTGCTGAGGATGACTGGGGTCAAGTTCAGGTGAGCCACTC
 CGTCTCCGCTGGCCTGGACTACTCCGGTGTGGTCCGGAACACGCCTATTGGCGTGAGACCGGTAAGTGTGTACGATG
 CTGTGACCGATGAAGAAGCTCTGGACGCATTATCGAACTGTCTCGCCTGGAAGGCATCATCCAGCCCTGGAGTCTTCTC
 ACGCACTGGCTTATCTGAAGAAGATCAACATCAAGGGTAAAGTTGTGGTGGTTAATCTGTCTGGTCTGGTGACAAGGAT
 CTGGAATCTGTACTGAACCACCCGTATGTTCCGCAACGCATCCGCCctgagcaccaccaccaccactga

TmE7:

Mutations: I166V S228G

Arnold Lab Culture Collection (ALCC): 4986

ATGAAAGGCTACTTCGGTCCGTACGGTGGCCAGTACGTGCCGAAATCCTGATGGGAGCTCTGGAAGAACTGGAAGCTGC
 GTACGAAGGAATCATGAAAGATGAGTCTTTCTGGATAGGATACAATGACCTGCTGCGCGATTATGCGGGTCTCCGACTC
 AGCTGTACTTCGCACGTCTGTCCGAAAAGTACGGTGTCTCGGTATATCTGAAACGTGAAGACCTGCTGCATACTGGT
 GCGCATAAAATCAATGGGGCTATCGGCCAGGTTCTGCTGGCAAACTAATGGGCAAAACCCGTATCATTGCTGAAACGGG
 TGCTGGTCAGCACGGCGTAGCAACTGCTACCGCAGCAGCGCTGTTCCGGTATGGAATGTGTAATCTATATGGGCGAAGAAG
 ACACGATCCGCCAGAACTAAACGTTGAACGTATGAACTGCTGGGTGCTAAAGTTGTACCGGTAATAATCCGGTAGCCGT
 ACCCTGAAAGACGCA **GTT**GACGAAGCTCTGCGTGACTGGACTACCAACCTGCAGACCCTATTACGTGCATGGCTCTGT
 GGTGGTCCGCATCCATATCCGACTATCGTACGTAACCTCCAAAAGGTTATCGGCGAAGAGACCAAAAAACAGATTCCAG
 AAAAAGAAGGCCCTGCCGACTACATCGTTGCGTGCGT **GGC**GGTGGTTCTAACGCTGCCGGTATCTTCTATCCGTTT
 ATCGATTCTGGTGTGAAGCTGATCGGCGTAGAAGCCGGTGGCGAAGGTTCTGGAAACCGGTAACATGCGGCTTCTCTGCT
 GAAAGGTAATAATCGGCTACCTGCACGGTTCTAAGACGTTCTGCTGAGGATGACTGGGGTCAAGTTCAGGTGAGCCACT
 CCGTCTCCGCTGGCCTGGACTACTCCGGTGTCCGGTCCGGAACACGCCTATTGGCGTGAGACCGGTAAGTGTGTACGATG
 GCTGTGACCGATGAAGAAGCTCTGGACGCATTTCGAACCTGCTCGCCTGGAAGGCATCATCCAGCCCTGGAGTCTTC
 TCACGCACTGGCTTATCTGAAGAAGATCAACATCAAGGGTAAAGTTGTGGTGGTTAATCTGTCTGGTCTGGTGACAAGG
 ATCTGGAATCTGTACTGAACCACCCGTATGTTCCGGAACGCATCCGCCctgagcaccaccaccaccactga

TmE8:

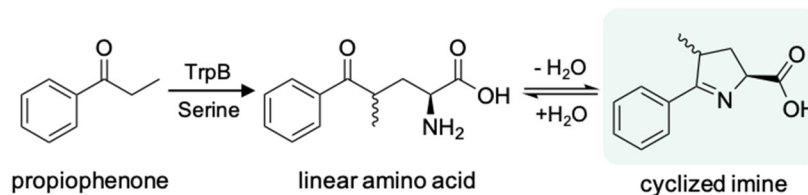
Mutations: H184A

Arnold Lab Culture Collection (ALCC): 4987

ATGAAAGGCTACTTCGGTCCGTACGGTGGCCAGTACGTGCCGAAATCCTGATGGGAGCTCTGGAAGAACTGGAAGCTGC
 GTACGAAGGAATCATGAAAGATGAGTCTTTCTGGATAGGATACAATGACCTGCTGCGCGATTATGCGGGTCTCCGACTC
 AGCTGTACTTCGCACGTCTGTCCGAAAAGTACGGTGTCTCGGTATATCTGAAACGTGAAGACCTGCTGCATACTGGT
 GCGCATAAAATCAATGGGGCTATCGGCCAGGTTCTGCTGGCAAACTAATGGGCAAAACCCGTATCATTGCTGAAACGGG
 TGCTGGTCAGCACGGCGTAGCAACTGCTACCGCAGCAGCGCTGTTCCGGTATGGAATGTGTAATCTATATGGGCGAAGAAG
 ACACGATCCGCCAGAACTAAACGTTGAACGTATGAACTGCTGGGTGCTAAAGTTGTACCGGTAATAATCCGGTAGCCGT
 ACCCTGAAAGACGCAGTTGACGAAGCTCTGCGTGACTGGACTACCAACCTGCAGACCCTATTACGTG **GCG**GGCTCTGT
 GGTGGTCCGCATCCATATCCGACTATCGTACGTAACCTCCAAAAGGTTATCGGCGAAGAGACCAAAAAACAGATTCCAG
 AAAAAGAAGGCCCTGCCGACTACATCGTTGCGTGCGTGGGCGGTGGTTCTAACGCTGCCGGTATCTTCTATCCGTTA
 TCGATTCTGGTGTGAAGCTGATCGGCGTAGAAGCCGGTGGCGAAGGTTCTGGAAACCGGTAACATGCGGCTTCTCTGCTG
 AAAGGTAATAATCGGCTACCTGCACGGTTCTAAGACGTTCTGCTGAGGATGACTGGGGTCAAGTTCAGGTGAGCCACTC
 CGTCTCCGCTGGCCTGGACTACTCCGGTGTCCGGTCCGGAACACGCCTATTGGCGTGAGACCGGTAAGTGTGTACGATG
 CTGTGACCGATGAAGAAGCTCTGGACGCATTTCGAACCTGCTCGCCTGGAAGGCATCATCCAGCCCTGGAGTCTTCTC
 ACGCACTGGCTTATCTGAAGAAGATCAACATCAAGGGTAAAGTTGTGGTGGTTAATCTGTCTGGTCTGGTGACAAGGAT
 CTGGAATCTGTACTGAACCACCCGTATGTTCCGGAACGCATCCGCCctgagcaccaccaccaccactga

B.4 Synthesis, isolation, and characterization of products

B.4.1 Enzymatic synthesis, isolation, characterization, and standard curve of propiophenone product



Scheme B-1. TrpB catalyzed propiophenone+serine reaction.

Two 800 mL cultures of BL21 (DE3) cells expressing TmE4 were grown according to expression conditions described in **Section B.2.2**. The expression cultures were centrifuged for 20 min and 14,000g. The pellets (33 g) were resuspended in 120 mL lysis buffer, rocked on a platform gently at room temperature for 30 minutes followed by heat-treatment in a 75 °C water bath for > 1 h. Heat-treated lysate was centrifuged for 20 min at 14,000g, and the lysate was decanted into a fresh container. To a 500-mL flask, serine (726 mg, 6.90 mmol, 1.02 equiv.) and lysate (120 mL) were added, sparged with argon for 10 minutes to prevent oxidation of the products, followed by the addition of propiophenone (919.8 mg, 6.75 mmol, 1.00 equiv.) and 15 mL KPi buffer to bring the reaction volume to 135 mL. The reaction was incubated in a water bath at 55 °C, swirled every few hours, and quenched with 100 mL acetonitrile after 24 hours. The reaction mixtures were centrifuged to remove precipitated cell mass, and the acetonitrile was removed *in vacuo*. The aqueous reaction mixture was then frozen and lyophilized to concentrate the sample for purification on a Biotage Isolera One purification system, using C₁₈ silica as the stationary phase (Sfär silica C₁₈ 12-gram column), CH₃CN (0.1% acetic acid) as the strong solvent, and H₂O (0.1% acetic acid) as the weak solvent. The product was obtained as a white solid (9.5 mg, 0.7% yield). ¹H NMR (500 MHz, D₂O, 23 °C, δ): [mixture of diastereomers] 8.01–7.95 (m, 2 H major + minor), 7.85–7.79 (m, 1 H major + minor), 7.69–7.63 (m, 2 H major + minor), 5.03 (t, *J* = 8.5 Hz, 1 H minor), 4.93 (dd, *J* = 10.5, 4.7 Hz, 1 H major), 1.37 (d, *J* = 7.4 Hz, 3 H major), 1.34 (d, *J* = 7.5 Hz, 3 H

minor). Due to the complexity of the spectrum, individual signals from the C_{β} and C_{γ} protons could not be identified with certainty.

Solutions of authentic propiophenone product standard were made in 1:1 $\text{CH}_3\text{CN}/1 \text{ M aq. HCl}$ (total conc. 1 mM). Two hundred μL of each standard were placed in a vial and diluted with 600 μL of 1:1 acetonitrile/ H_2O + 8 mM caffeine (final caffeine concentration was 6 mM) and analyzed by HPLC. The ratios of product peaks to area of HPLC absorbance at 254 nm were correlated to the actual ratios by a linear relationship (**Figure B-13**).

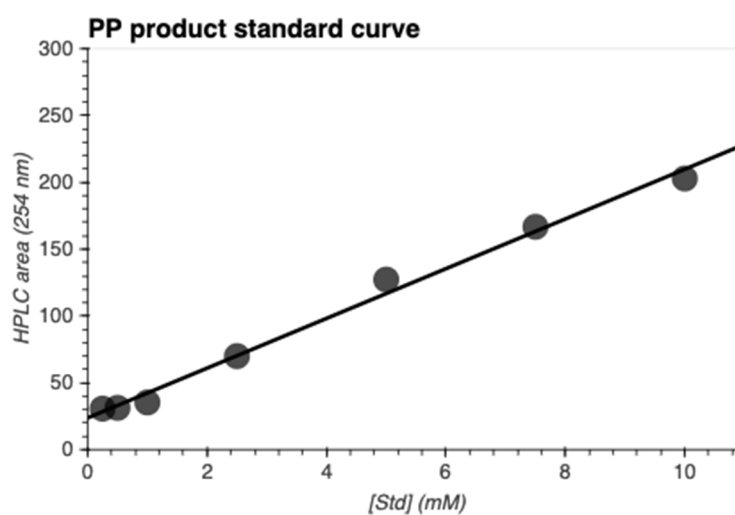
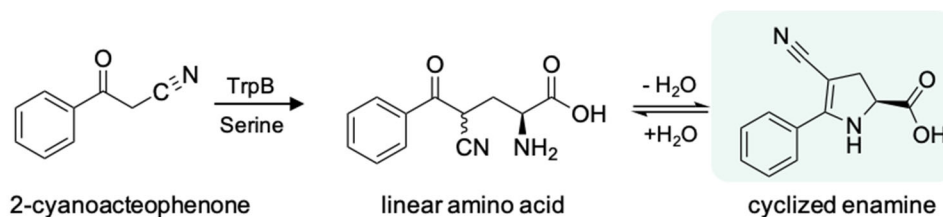


Figure B-13. Propiophenone product standard curve. $y = 0.0534x - 1.24$, $R^2 = 0.9924$.

B.4.2 Enzymatic synthesis, isolation, characterization, and standard curve of 2-cyanoacetophenone product



Scheme B-2. TrpB catalyzed 2-cyanoacetophenone+serine reaction.

An 800 mL culture of BL21 (DE3) cells expressing TmE8 was grown according to expression conditions described in **Section B.2.2**. The expression culture was centrifuged for 20 min at 14,000g. The pellet (15 g) was resuspended in 60 mL lysis buffer, rocked on platform gently at room temperature for 30 minutes, followed by heat-treatment in a 75 °C water bath for > 1 h. To a 250-mL flask, serine (724 mg, 6.89 mmol, 5.00 equiv.), 2-cyanoacetophenone (200 mg, 1.38 mmol, 1.00 equiv.), 50 mL lysate, and 5 mL KPi buffer were added. The reaction was incubated in a water bath at 55 °C, swirled every few hours, and quenched with 50 mL acetonitrile after 24 hours. The reaction mixtures were centrifuged to remove precipitated cell mass, and acetonitrile removed *in vacuo*. The aqueous reaction mixture was then frozen and lyophilized to concentrate the sample for purification on a Biotage Isolera One purification system, using C18 silica as the stationary phase (Sfär silica C18 12-gram column), CH₃CN (0.1% acetic acid) as the strong solvent, and H₂O (0.1% acetic acid) as the weak solvent. The product was obtained as a green solid (191.17 mg, 64.7% yield). ¹H NMR (500 MHz, DMSO-d₆, 23 °C, δ): 12.93 (br s, 1 H), 7.74–7.70 (m, 2 H), 7.53–7.49 (m, 3 H), 7.34 (br s, 1 H), 4.46 (dd, J = 12.4, 6.5 Hz, 1 H), 3.20 (dd, J = 14.7, 12.4 Hz, 1 H), 2.93 (dd, J = 14.7, 6.5 Hz, 1 H). ¹³C NMR (101 MHz, DMSO-d₆, 23 °C, δ): 173.9, 161.3, 130.8, 129.2, 128.7, 126.9, 120.3, 71.0, 58.3, 34.1.

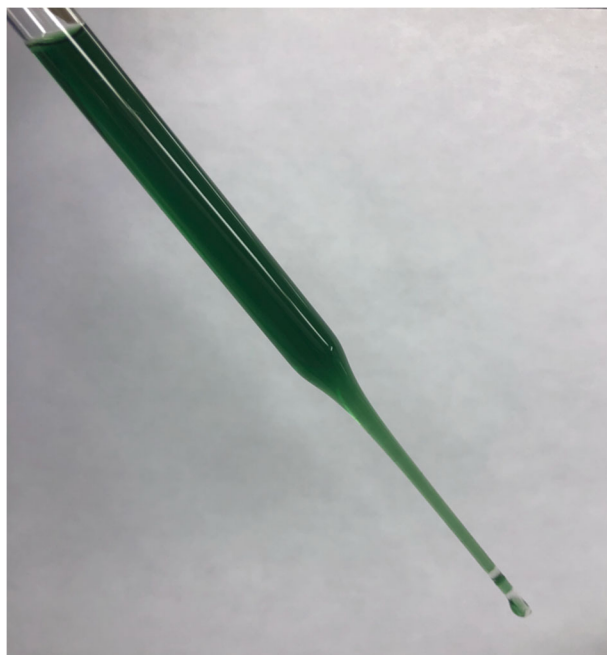


Figure B-14. 2-cyanoacetophenone cyclized enamine product dissolved in methanol.

Solutions of authentic 2-cyanoacetophenone product standard were made in methanol (**Figure B-14**). Two hundred μL of each standard were placed in a vial and diluted with 600 μL of 1:1 acetonitrile/ H_2O + 8 mM caffeine (Final caffeine concentration 6 mM) and analyzed by HPLC. The ratios of product peaks to area of HPLC absorbance at 254 nm were correlated to the actual ratios by a linear relationship (**Figure B-15**).

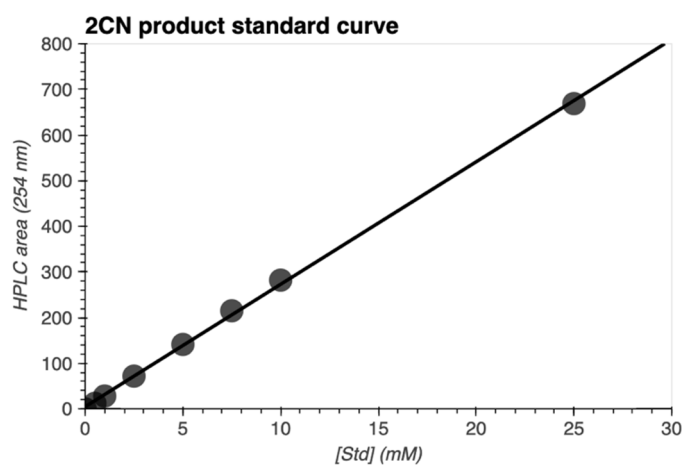
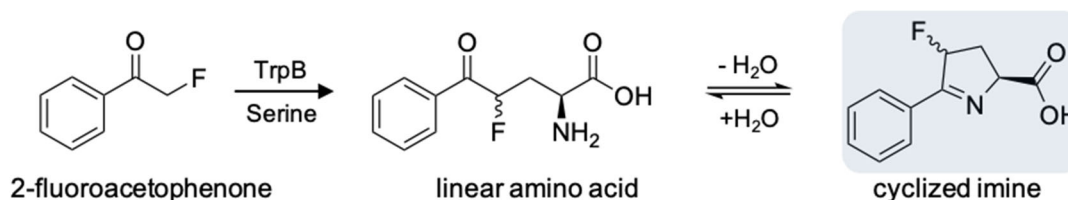


Figure B-15. 2-Cyanoacetophenone product standard curve. $y = 0.0085x$, $R^2 = 0.9986$.

B.4.3 Enzymatic synthesis, isolation, characterization, and standard curve of 2-fluoroacetophenone product



Scheme B-3. TrpB catalyzed 2-fluoroacetophenone+serine reaction.

A 500 mL culture of BL21 (DE3) cells expressing TmE8 was grown according to expression conditions described in **Section B.2.2**. The expression culture was centrifuged for 20 min at 14,000g. The pellet (10 g) was resuspended in 40 mL lysis buffer rocked on platform gently at room temperature for 30 minutes followed by heat-treatment in a 75 °C water bath for > 1 h. To a 250-mL flask, serine (760.76 mg, 7.24 mmol, 5.00 equiv.), 2-fluoroacetophenone (200 mg, 1.45 mmol, 1.00 equiv.), and 40 mL lysate were added. The reaction was incubated in a water bath at 55 °C, swirled every few hours, quenched with 1 M HCl after 48 hours, and extracted with ethyl acetate to remove any excess 2-fluoroacetophenone starting material. Upon addition of acid, the 2-fluoroacetophenone product slowly decomposes to 2-phenylpyrrole (validated by authentic standard on GSMS), which is a bright pink/purple color (**Figure B-16**). However, 2-phenylpyrrole is easily removed from the aqueous layer during the ethyl acetate extraction.

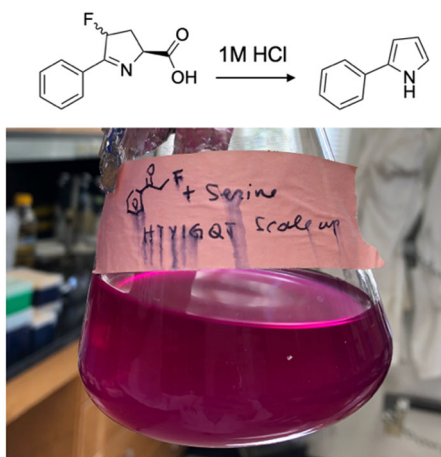


Figure B-16. 2-fluoroacetophenone product decomposes to 2-phenylpyrrole upon addition of acid to reaction mixture.

The aqueous layer was lyophilized to concentrate the sample for purification on a Biotage Isolera One purification system, using C18 silica as the stationary phase (Sfär silica C18 12-gram column), methanol as the strong solvent, and H₂O as the weak solvent. The cyclized 2-fluoroacetophenone product was obtained as a brown solid (16 mg, 5.3% yield). ¹H NMR (500 MHz, D₂O, 23 °C, δ) [mixture of diastereomers]: 7.82–7.76 (m, 2 H major + minor), 7.46–7.41 (m, 1 H major + minor), 7.41–7.35 (m, 2 H major + minor), 6.12 (dddd, J = 54.1, 7.1, 2.8, 1.3 Hz, 1 H major), 6.02 (ddd, J = 53.9, 7.9, 4.2 Hz, 1 H minor), 4.75 (dddd, J = 6.8, 5.6, 5.6, 1.2 Hz, 1 H major), 4.51 (td, J = 8.8, 5.3 Hz, 1 H minor), 2.74 (ddt, J = 22.6, 14.6, 8.1 Hz, 1 H minor), 2.44 (dddd, J = 30.7, 15.0, 7.6, 2.7 Hz, 1 H major), 2.23 (dddd, J = 30.3, 15.1, 7.1, 5.7 Hz, 1 H major), 2.02 (dddd, J = 31.8, 14.6, 5.4, 4.2 Hz, 1 H minor). ¹⁹F NMR (282 MHz, D₂O, 23 °C, δ) [mixture of diastereomers]: –170.5 (dddd, J = 54.0, 31.7, 22.2, 9.5 Hz, 1 F minor), –173.5 (dddd, J = 54.1, 30.7, 30.3, 5.6 Hz, 1 F major).

Solutions of authentic 2-fluoroacetophenone cyclized product standard were made in acetonitrile. Two hundred μL of each standard were placed in a vial and diluted with 600 μL of 1:1 acetonitrile/H₂O + 8 mM caffeine (Final caffeine concentration 6 mM) and analyzed by HPLC. The ratios of caffeine to product peaks at 254 nm were correlated to the actual ratios by a linear relationship (**Figure B-17**).

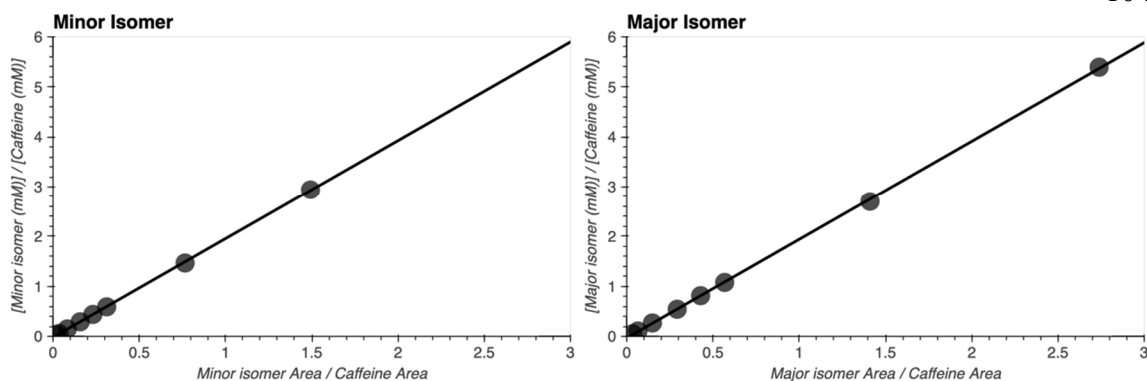


Figure B-17. 2-fluoroacetophenone product standard curve. Minor isomer: $y = 1.97x - 0.017$, $R^2 = 0.9998$. Major isomer: $y = 1.97x - 0.031$, $R^2 = 0.9998$.

To calculate 2-fluoroacetophenone product concentration, the following equation is used:

$$C_P = RF * \frac{A_P}{A_C} * C_C$$

where C_P is the concentration of product (mM), RF is the response factor (1.97093347), A_P is the area of the product, A_C is the area of the caffeine internal standard, and C_C is the concentration of caffeine (6 mM).

B.4.4 Synthesis of 2-fluoroacetophenone-2,2- d_2

2-Fluoroacetophenone-2,2- d_2 was synthesized and purified according to a previously reported method.⁵ ^1H NMR (500 MHz, DMSO- d_6 , 23 °C, δ) 7.89 (d, $J = 8.0$ Hz, 2 H), 7.68 (t, $J = 7.4$ Hz, 1 H), 7.55 (t, $J = 7.8$ Hz, 2 H). ^{13}C NMR (125 MHz, DMSO- d_6 , 23 °C, δ) 193.9 (d, JCF = 14.1 Hz), 134.0, 133.5, 128.9, 127.6 (d, JCF = 1.6 Hz), 83.4 (dt, JCF = 173.5 Hz, JCD = 23.4 Hz). ^{19}F NMR (282 MHz, DMSO- d_6 , 23 °C, δ) -234.2 (p, $J = 6.9$ Hz).

^1H NMR spectrum of propiophenone product
500 MHz, D_2O , 23 °C

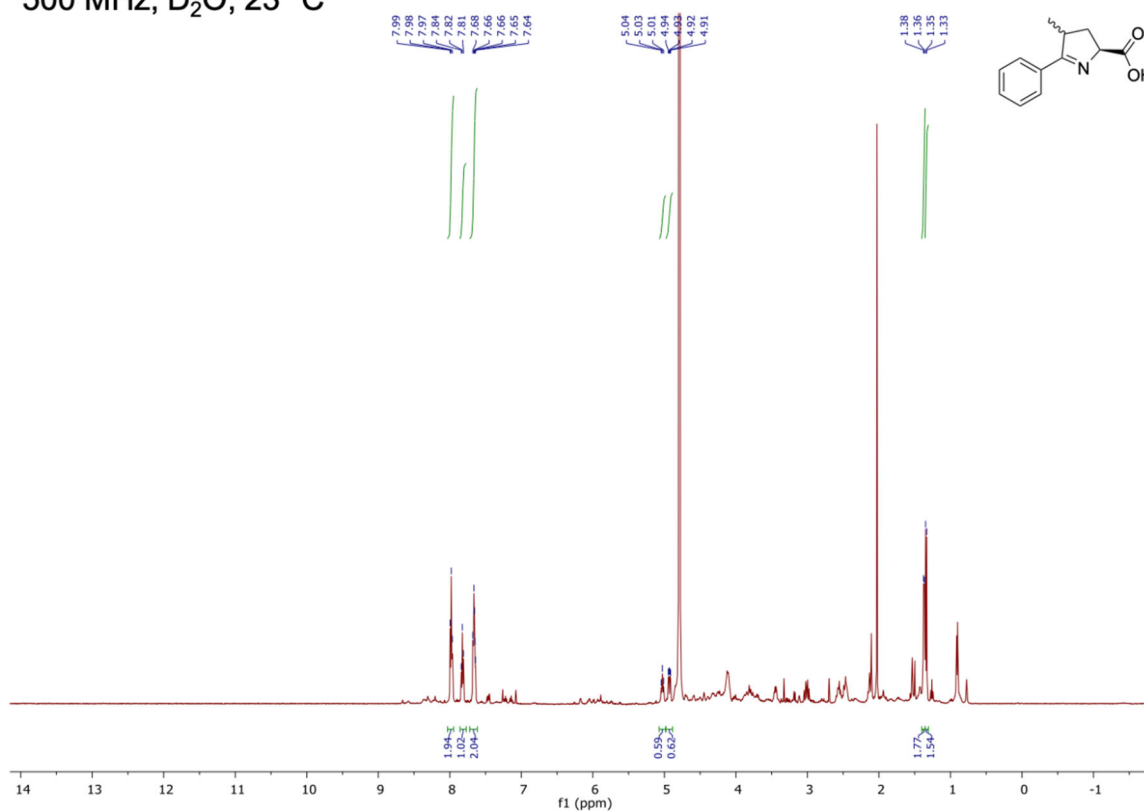


Figure B-18. ^1H NMR spectrum of propiophenone product.

¹H NMR spectrum of 2-cyanoacetophenone product
500 MHz, DMSO-d₆, 23 °C

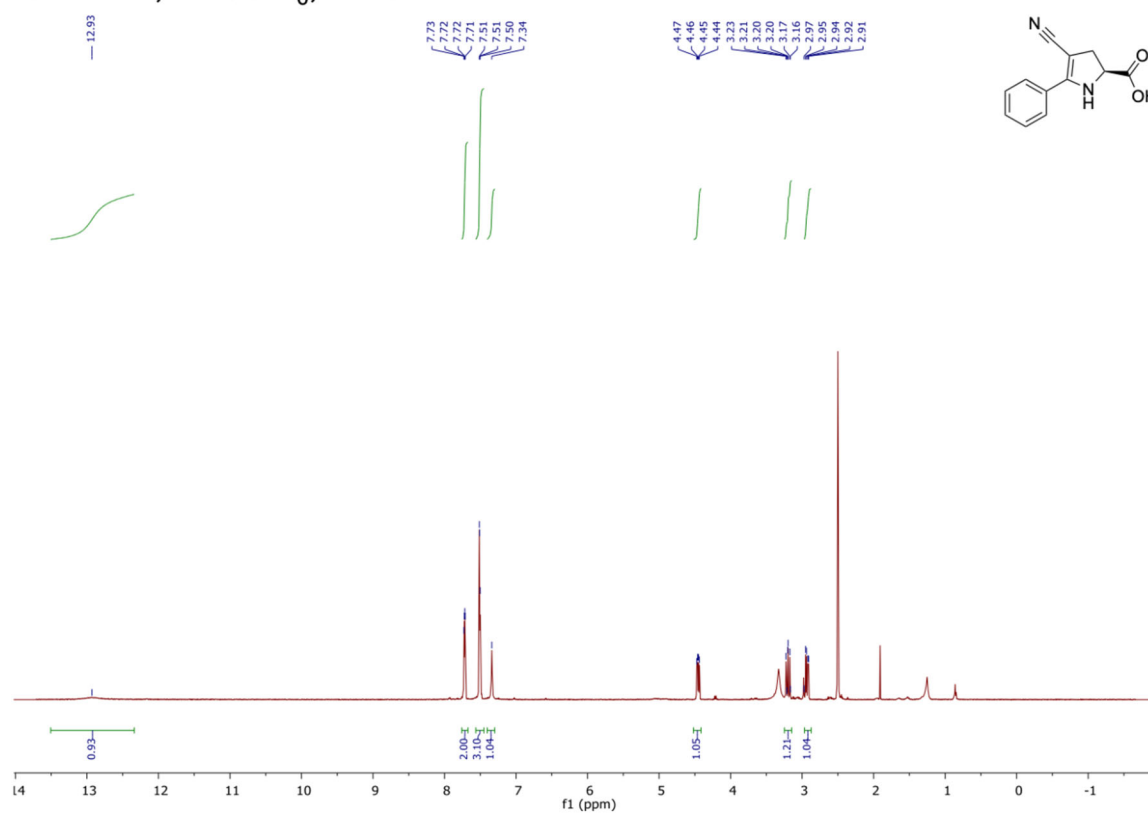


Figure B-19. ¹H NMR spectrum of 2-cyanoacetophenone product.

^{13}C NMR spectrum of 2-cyanoacetophenone product
101 MHz, $\text{DMSO-}d_6$, 23 °C

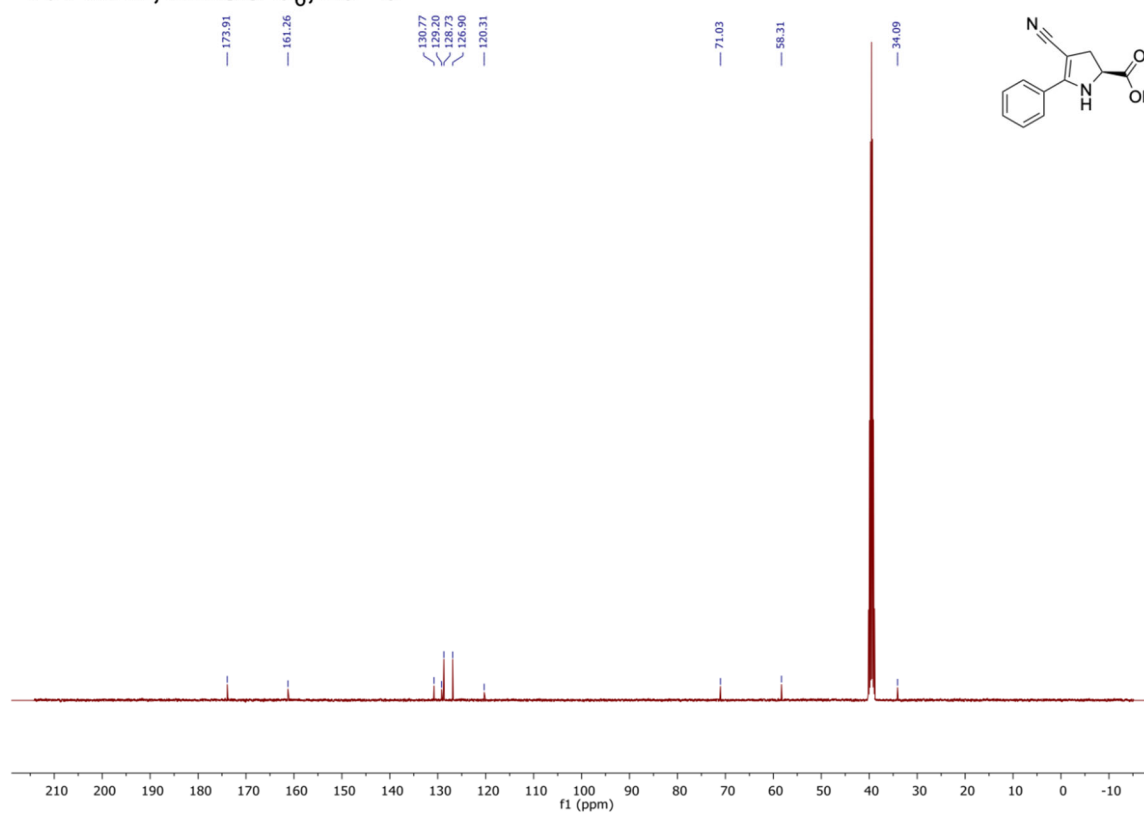


Figure B-20. ^{13}C NMR spectrum of 2-cyanoacetophenone product.

¹H NMR spectrum of 2-fluoroacetophenone product
500 MHz, D₂O, 23 °C

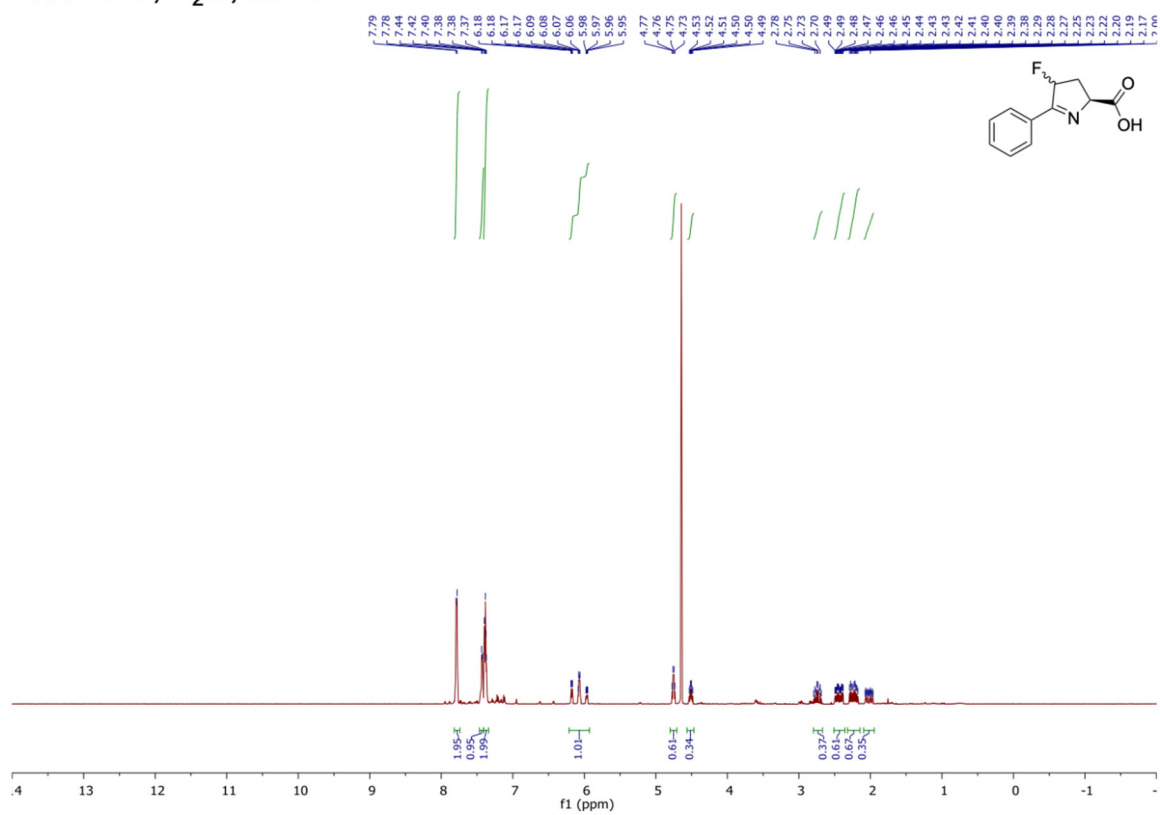


Figure B-21. ¹H NMR spectrum of 2-fluoroacetophenone product.

^{19}F NMR spectrum of 2-fluoroacetophenone product
282 MHz, D_2O , 23 °C

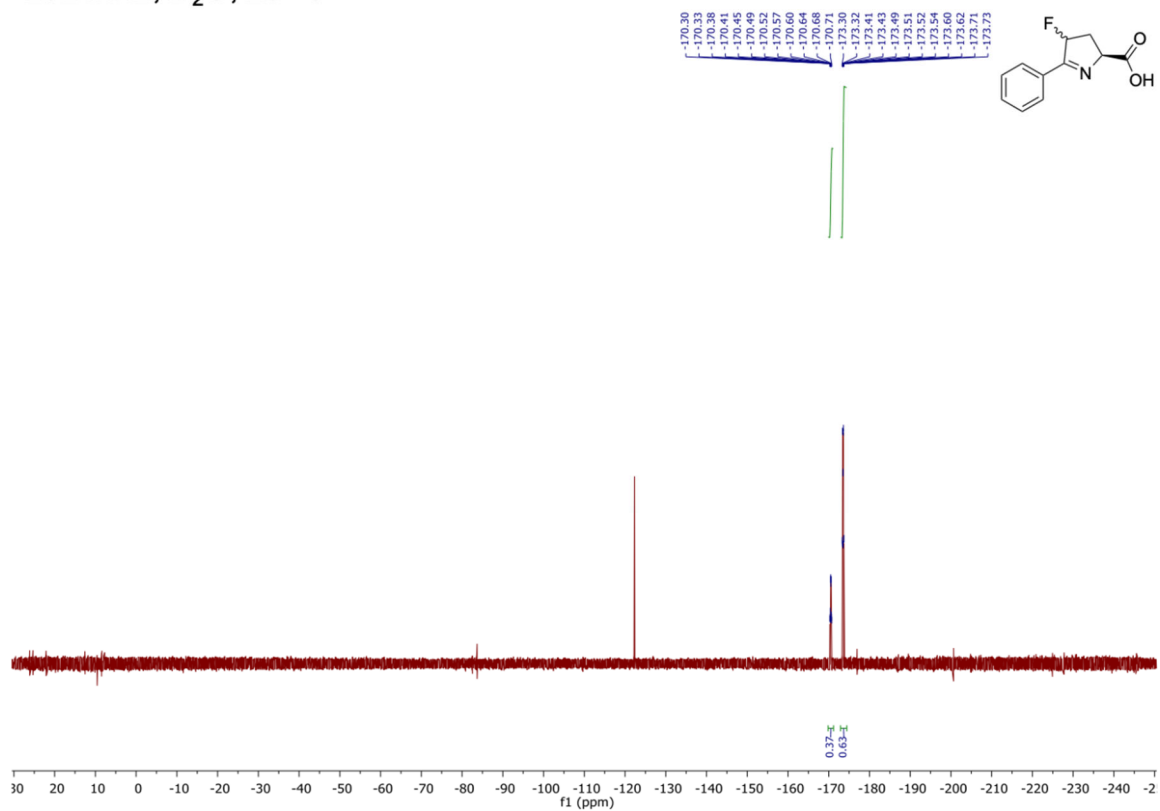


Figure B-22. ^{19}F NMR spectrum of 2-fluoroacetophenone product.

^1H NMR spectrum of 2-fluoroacetophenone-2,2- d_2
500 MHz, $\text{DMSO-}d_6$, 23 °C

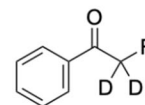
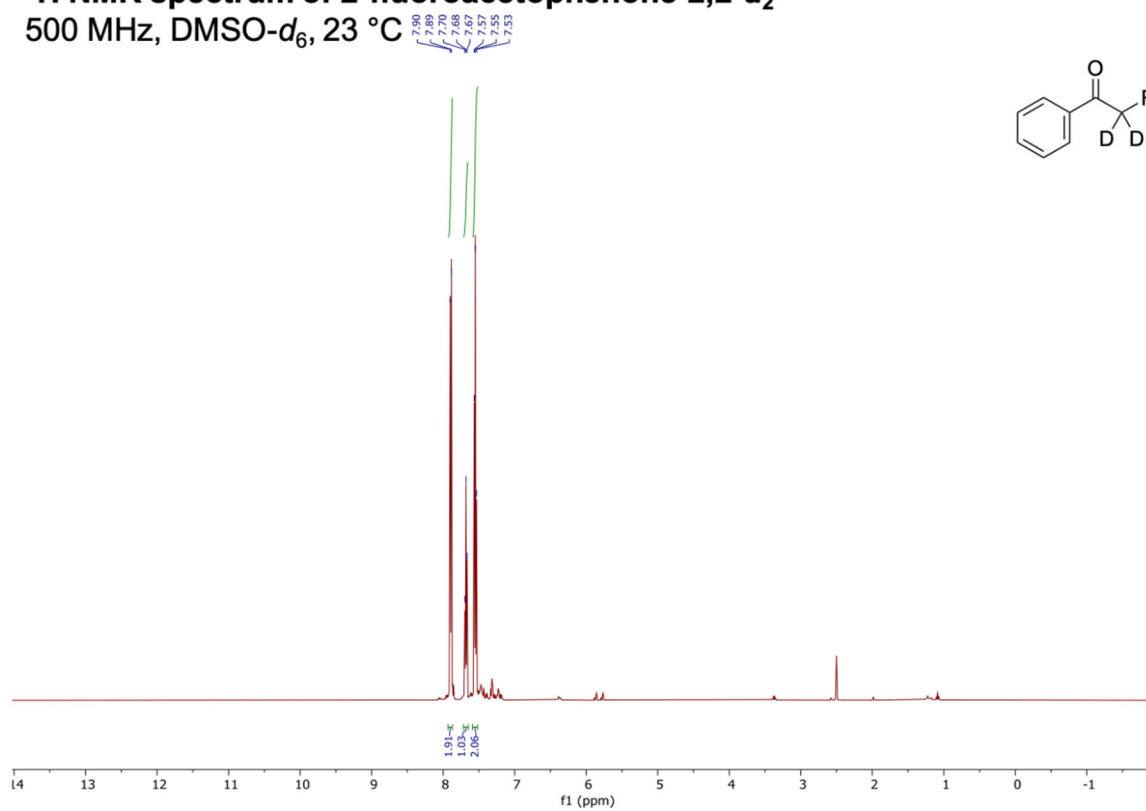


Figure B-23. ^1H NMR spectrum of 2-fluoroacetophenone-2,2- d_2 product.

^{13}C NMR spectrum of 2-fluoroacetophenone-2,2- d_2
125 MHz, $\text{DMSO-}d_6$, 23 °C

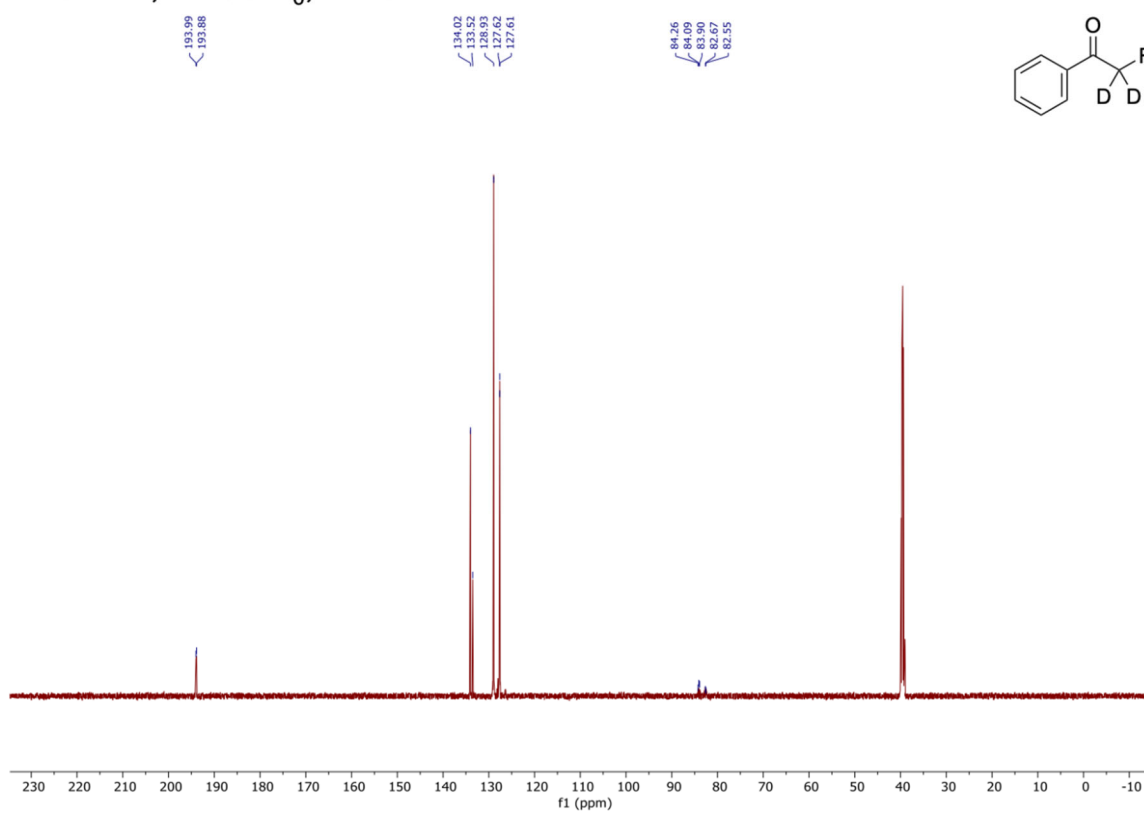


Figure B-24. ^{13}C NMR spectrum of 2-fluoroacetophenone-2,2- d_2 product.

Bibliography for Appendix B

1. Watkins, E. J., Almhjell, P. J., & Arnold, F. H. Direct enzymatic synthesis of a deep-blue fluorescent noncanonical amino acid from azulene and serine. *ChemBioChem* **21**, 80–83 (2020).
2. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
3. Boville, C. E. et al. Engineered Biosynthesis of β -Alkyl Tryptophan Analogs. *Angew. Chemie Int. Ed.* (2018). doi:10.1002/anie.201807998
4. Zhao, H. & Zha, W. In vitro ‘sexual’ evolution through the PCR-based staggered extension process (StEP). *Nat. Protoc.* **1**, 1865–1871 (2006).
5. Catalano, D. et al. Conformation of ω -fluoroacetophenone: An NMR study using a liquid crystalline solvent. *J. Chem. Soc. Perkin Trans. 2* 1823–1826 (1998). doi:10.1039/a800368h

SCALABLE, CONTINUOUS EVOLUTION FOR THE GENERATION
OF DIVERSE ENZYME VARIANTS ENCOMPASSING
PROMISCUOUS ACTIVITIES

Material from this chapter appears in: “Rix, G., **Watkins-Dulaney, E. J.**, Almhjell, P. J., Boville, C. E., Arnold, F. H., & Liu, C. C. (2020). Scalable continuous evolution for the generation of diverse enzyme variants encompassing promiscuous activities. *Nature Communications*, 11(1), 1-11.”

All authors contributed to experimental design and data analysis. G.R. constructed selection and performed all evolution experiments. E.J.W.D. performed the panel HPLC-MS assay and indole conversion rate measurements on *TmTrpBs* from variant set 2, and P.J.A. analyzed the results. C.E.B. performed in vitro characterizations of *TmTrpBs* from variant set 1 and performed the thermal shift assay and substrate scope characterizations. G.R. and C.C.L. wrote the manuscript with input and contributions from all authors.

ABSTRACT

Enzyme orthologs sharing identical primary functions can have different promiscuous activities. While it is possible to mine this natural diversity to obtain useful biocatalysts, generating comparably rich ortholog diversity is difficult, as it is the product of deep evolutionary processes occurring in a multitude of separate species and populations. Here, we take a first step in recapitulating the depth and scale of natural ortholog evolution on laboratory timescales. Using a continuous directed evolution platform called OrthoRep, we rapidly evolved the *Thermotoga maritima* tryptophan synthase β -subunit (*TmTrpB*) through multi-mutation pathways in many independent replicates, selecting only on *TmTrpB*'s primary activity of synthesizing L-tryptophan from indole and L-serine. We find that the resulting sequence-diverse *TmTrpB* variants span a range of substrate profiles useful in industrial biocatalysis and suggest that the depth and scale of evolution that OrthoRep affords will be generally valuable in enzyme engineering and the evolution of new biomolecular functions.

4.1 Introduction

Natural enzymes typically have many orthologs. While the primary activity of orthologous enzymes is largely the same,¹ promiscuous functions not under selective pressure can vary widely.^{2,3} Such variation may be attributed to the deep and distinct evolutionary histories shaping each ortholog, including long periods of neutral drift, recalibration of primary activity, and adaptation to new host environments such as temperature. These rich histories act to produce extensive genetic diversity, which underpins different promiscuity profiles.²

Diversity in promiscuous functions across orthologs is of both fundamental and practical importance. An enzyme's reserve of promiscuous activities dictates what secondary reactions, environmental changes, or niches the enzyme can accommodate.^{4,5} Diversity in promiscuous activities therefore contributes to the basic robustness of life and adaptation. An enzyme's reserve of promiscuous activities can also be mined in the application of enzymes for biocatalysis.^{6,7} Ortholog diversity therefore expands the range of reactions at the disposal of enzyme engineers, supporting the growing role of "green" enzymatic processes in the chemical and pharmaceutical industries.⁸⁻¹⁰

Inspired by the remarkable ability of enzyme orthologs to encompass promiscuous activities, we asked whether we could extend the substrate scope of useful enzymes by evolving multiple versions of an enzyme in the laboratory, selecting only for its primary function. Although this idea has been explored before using classical directed evolution approaches, most notably through the generation of cryptic genetic variation with neutral drift libraries,¹¹⁻¹⁴ we recognized that our recently developed continuous evolution system, OrthoRep, may be considerably better poised for this challenge.^{15,16} Classical directed evolution mimics evolution through an iterative procedure that involves diversifying a gene of interest (GOI) *in vitro* (e.g., through error-prone PCR), transforming the resulting GOI library into cells, and selecting or screening for desired activities, where each cycle of this procedure represents one step in an evolutionary search.¹⁷ However, since each cycle is manually staged, classical directed evolution does not readily admit depth and scale during exploration of functional sequence space—it is difficult to carry out many iterations to mimic lengthy evolutionary

searches (depth), let alone do so in many independent experiments (scale). Yet evolutionary depth and scale are precisely the two characteristics responsible for ortholog diversity in nature. Natural orthologs have diversified from their ancestral parent over great evolutionary timescales, allowing for the traversal of long mutational pathways shaped by complex selection histories (depth). Natural orthologs are also the result of numerous independent evolutionary lineages, since spatially separated species and populations are free to take divergent mutational paths and experience different environments (scale). Systems that better mimic the depth and scale of natural enzyme evolution, but on laboratory timescales, are thus needed for the effective generation of enzyme variants that begin to approach the genetic and promiscuity profile diversity of orthologs.

OrthoRep is such a system. In OrthoRep, an orthogonal error-prone DNA polymerase durably hypermutates an orthogonal plasmid (p1) without raising the mutation rate of the host *Saccharomyces cerevisiae* genome.¹⁶ Thus, GOIs encoded on p1 rapidly evolve when cells are simply passaged under selection. By reducing the manual stages of classical directed evolution down to a continuous process where cycles of diversification and selection occur autonomously *in vivo*, OrthoRep readily accesses depth and scale in evolutionary search.^{16,18} Here, we apply OrthoRep to the evolution of the *Thermotoga maritima* tryptophan synthase β -subunit (*TmTrpB*) in multiple independent continuous evolution experiments, each carried out for at least 100 generations. While we only pressured *TmTrpB* to improve its primary activity of coupling indole and serine to produce tryptophan, the large number of independent evolution experiments we ran (scale) and the high degree of adaptation in each experiment (depth) resulted in a panel of variants encompassing expanded promiscuous activity with indole analogs. In addition to the immediate value of these newly evolved *TmTrpBs* in the synthesis of tryptophan analogs, our study offers a new template for enzyme engineering where evolutionary depth and scale is leveraged on laboratory timescales to generate effective variant collections covering broad substrate scope.

4.2 Establishing a selection system for the evolution of *TmTrpB* variants

To evolve *TmTrpB* variants using OrthoRep, we first needed to develop a selection where yeast would rely on *TmTrpB*'s primary enzymatic activity for growth. *TmTrpB* catalyzes the PLP-dependent coupling of L-serine and indole to generate L-tryptophan (Trp) in the presence of the tryptophan synthase α -subunit, *TmTrpA*.¹⁹ In *T. maritima* and all other organisms that contain a heterodimeric tryptophan synthase complex, TrpA produces the indole substrate that TrpB uses and the absence of TrpA significantly attenuates the activity of TrpB through loss of allosteric activation.^{19,20} TRP5 is the *S. cerevisiae* homolog of this heterodimeric enzyme complex, carrying out both TrpA and TrpB reactions and producing Trp for the cell. We reasoned that by deleting the *TRP5* gene and forcing *S. cerevisiae* to rely on *TmTrpB* instead, cells would be pressured to evolve high stand-alone *TmTrpB* activity in order to produce the essential amino acid Trp in indole-supplemented media (**Figure 4-1**). This selection pressure would also include thermoadaptation, as yeast grow at mesophilic temperatures in contrast to the thermophilic source of *TmTrpB*. Therefore, the selection on *TmTrpB*'s primary activity would be multidimensional—stand-alone function, temperature, and neutral drift implemented when desired—and could result in complex evolutionary pathways that serve our goal of maximizing functional variant diversity across replicate evolution experiments. In addition, the multidimensional selection also serves practical goals as stand-alone activity is useful in biosynthetic applications (enzyme complexes are difficult to express and use *in vitro*) and activity at mesophilic temperatures is more compatible with heat-labile substrates, industrial processes where heating costs can compound, or *in vivo* applications in model mesophilic hosts (e.g. *S. cerevisiae* or *Escherichia coli*).

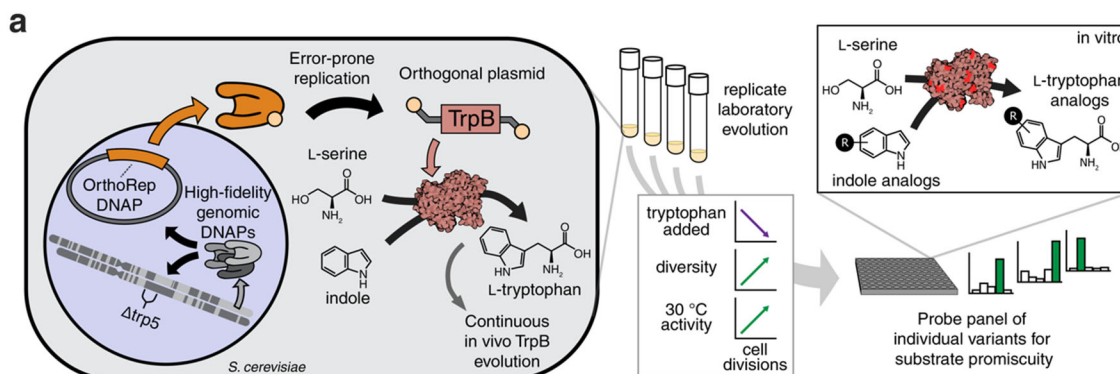


Figure 4-1. Pipeline for the use of OrthoRep continuous directed evolution to generate many diverse, functional *Tm*TrpB sequences. *Tm*TrpB variants are first evolved in replicate for Trp production in yeast. OrthoRep enables replicate evolution through error-prone replication of an orthogonal plasmid by an orthogonal polymerase, maintaining low error rates in genome replication. By encoding a *Tm*TrpB variant on this plasmid in a tryptophan synthase (TRP5) deletion mutant, *Tm*TrpB may be both continuously diversified and selected for through gradual reduction in Trp supplied in the growth medium. Evolved populations containing many diverse, functional individuals may then be randomly sampled and tested for activity with indole analogs.

To test this selection, we turned to a positive control *Tm*TrpB variant called *Tm*Triple. This variant was previously engineered to enable stand-alone activity, free from dependence on allosteric activation by TrpA, through a minimal set of three mutations.⁷ We found that *Tm*Triple rescued TRP5 function in a $\Delta trp5$ strain in an indole-dependent manner, validating our selection (**Appendix C, Figure C-1**). Notably, *Tm*Triple, along with other *TrpB* variants tested, only supported complementation when expressed from a high-strength promoter (**Appendix C, Figure C-1**). This highlighted the opportunity for substantial adaptation and drift even in evolution experiments that start from already engineered *Tm*TrpB variants.

4.3 Continuous evolution of *Tm*TrpB with depth and scale

We encoded wild-type (wt) *Tm*TrpB, *Tm*Triple, as well as a nonsense mutant of *Tm*Triple, *Tm*TripleQ90*, onto OrthoRep's p1 plasmid, which is replicated by a highly error-prone orthogonal DNA polymerase. *Tm*TripleQ90* was included because reversion of the stop codon at position 90 in *Tm*TripleQ90* would act as an early indication that adaptation was occurring, giving us confidence to continue passing our evolution experiments for several weeks to maximize evolutionary search depth. In all three OrthoRep $\Delta trp5$ strains, the initial

TmTrpB sequences enabled only minimal indole-dependent complementation (**Appendix C, Figure C-1**). This was expected for wt *TmTrpB*, which has low stand-alone enzymatic activity and *TmTripleQ90**, which has a premature stop codon, and was unsurprising for *TmTriple*, since *TmTriple* displayed indole-dependent complementation only when artificially overexpressed (**Appendix C, Figure C-1**).

To continuously evolve *TmTrpB*, we passaged cells encoding wt *TmTrpB*, *TmTriple*, or *TmTripleQ90** on OrthoRep in the presence of 100 μ M indole while reducing the amount of Trp in the medium over time. In total, six 100-mL and twenty 3-mL cultures were passaged, each representing a single independent evolutionary trajectory. Passages were carried out as 1:100 dilutions where Trp concentrations were decreased in the (N+1)th passage if cells grew quickly in the Nth passage, until Trp was fully omitted. All six of the 100-mL cultures, and four of the twenty 3-mL cultures fully adapted and were capable of robust growth in indole-supplemented media lacking Trp after 90–130 generations (13–20 passages) (**Figure 4-2; Appendix C, Table C-1**). Populations that did not achieve growth in the absence of Trp still adapted, but stopped improving at \sim 5 μ M supplemented Trp, suggesting a suboptimal local fitness maximum that is easier to escape through the greater sequence diversity represented in larger populations. This could explain the different success rates in reaching full adaptation between the 3-mL and 100-mL populations. Cultures that did adapt fully were passaged for an additional \sim 40 generations without increasing selection stringency to allow for accumulation of further diversity through neutral drift.

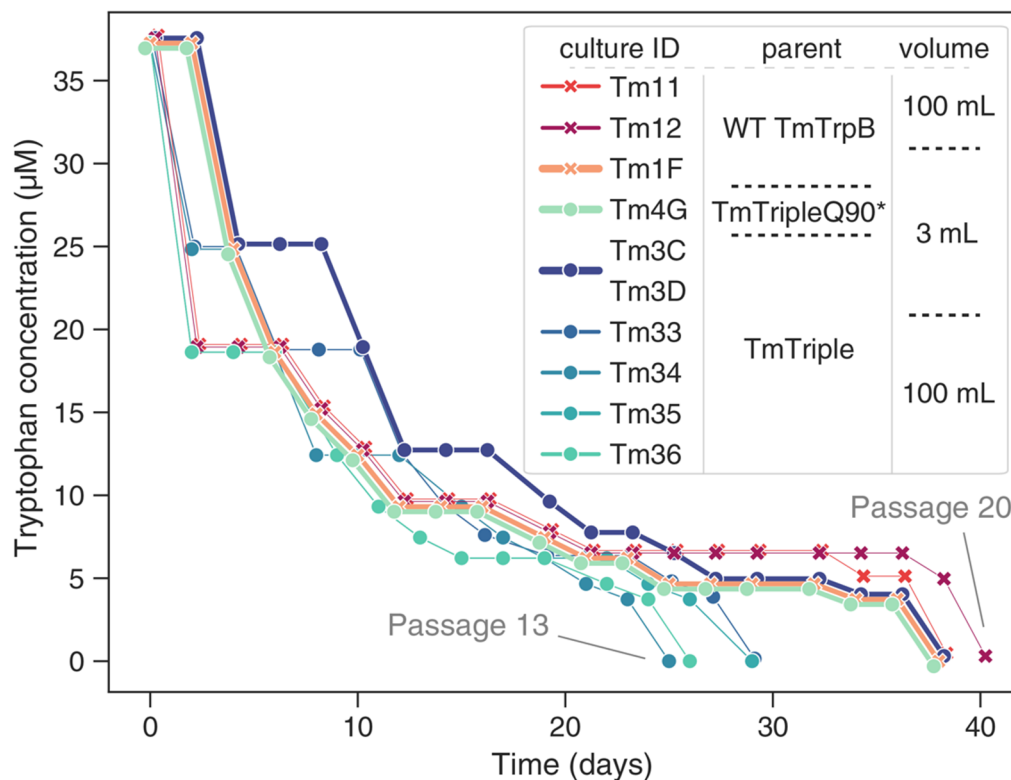


Figure 4-2. Selection trajectories for ten replicate cultures that evolved sufficient *TmTrpB* activity to support cell growth without supplemented Trp. Each point represents a single 1:100 dilution (passage) into fresh indole-supplemented growth medium. Trp concentration of fresh media was reduced when high saturation was achieved in the previous passage. Plots are slightly offset from true values to allow for visibility of all selection trajectories. *Tm3C* and *Tm3D* are plotted as one line as their trajectories were identical.

For each of the ten fully adapted populations, we PCR amplified and bulk sequenced the *TmTrpB* alleles on the p1 plasmid. Mutations relative to the parent *TmTrpB* variant detected at >50% frequency in each population were deemed consensus mutations for that population, with the exception of reversion of the stop codon in populations evolving *TmTripleQ90**. This stop codon reversion occurred at 100% frequency in the relevant populations and was not counted in any subsequent analyses due to its triviality. An average of 5.6 (\pm 2.3 s.d.) and a range of 3–11 consensus amino-acid changes per population were observed (**Figure 4-3; Appendix C, Table C-2**). Some of these mutations occurred at residues previously identified as relevant in conformational dynamics (e.g., N167D and S302P).^{20–22} Most mutations

observed, however, have not been previously identified in laboratory engineering experiments, suggesting that even the consensus of these populations explored new regions of *TmTrpB*'s fitness landscape, doing so with diversity across replicates (**Figure 4-3**) that might translate to diversity in promiscuous activities across evolved variants.

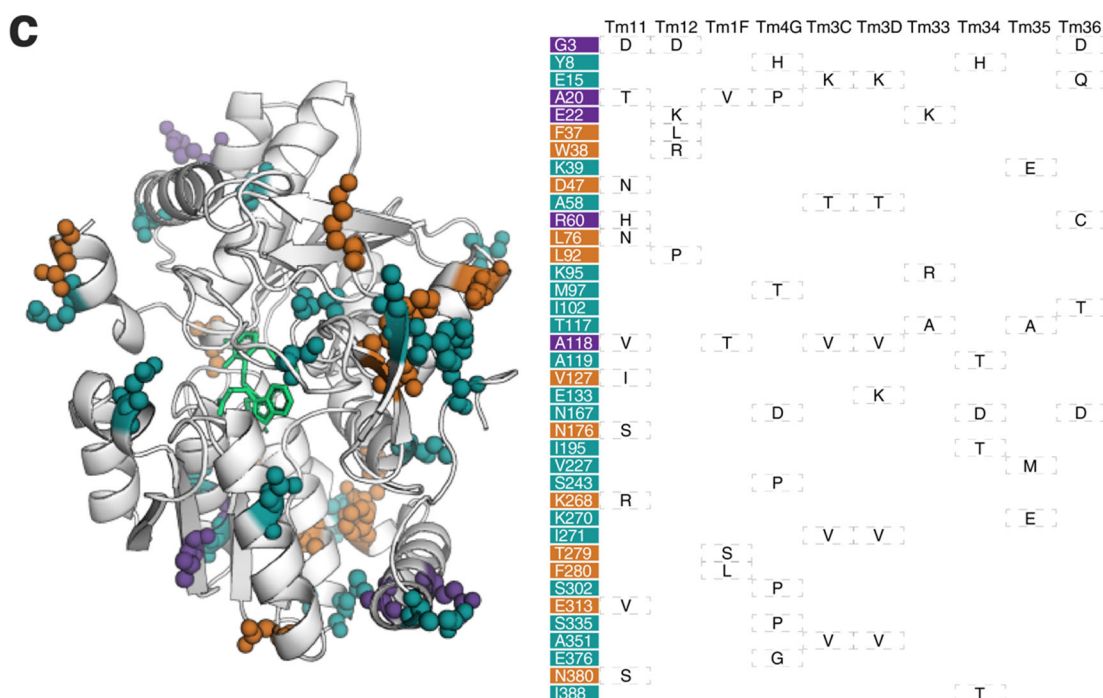


Figure 4-3. *TmTrpB* homology model and table depicting consensus mutations of the ten cultures shown in **Figure 4-1**. Mutations are colored by their appearance in populations evolved from wt *TmTrpB* (orange), *TmTriple* or *TmTripleQ90** (green), or both (purple).

4.4 Evolved *TmTrpB* variants improve Trp production *in vivo* and contain cryptic genetic variation

To ensure that evolved *TmTrpB* variants and not potential host genomic mutations were primarily responsible for each population's adaptation, we cloned individual *TmTrpBs* into a standard low copy yeast nuclear plasmid under a promoter that approximates expression from p1,^{23,24} transformed a fresh $\Delta trp5$ strain with the constructs, and tested for the ability of the variants to support indole-dependent growth in the absence of Trp (**Figures 4-4; Appendix C, Figure C-2**). Sixteen *TmTrpB* mutants were tested, representing one or two

individual variants from each of the ten fully adapted populations. We found that 12 of the 16 TrpB variants complemented growth to a similar degree as TRP5 when supplemented with 400 μ M indole, demonstrating substantial improvement over their wt *Tm*TrpB and *Tm*Triple parents (**Figure 4-4a**).

Unsurprisingly, this set of clonal *Tm*TrpBs contained more sequence diversity than the consensus sequences of the ten populations from which they were taken. Together, the variants tested comprised a total of 85 unique amino-acid substitutions, with an average of 8.7 (\pm 2.1 s.d.) and a range of 5–13 non-synonymous mutations per variant (variant set 1; **Appendix C, Tables C-2 and C-3**). Since the 12 *Tm*TrpBs from this set exhibiting complementation were all similarly active in their primary activity yet mutationally diverse (**Figure 4-4b**), we may conclude that our scaled evolution experiments generated substantial cryptic genetic variation. We note that four of 16 *Tm*TrpB variants exhibited similar or lower Trp productivity compared to their parent (**Appendix C, Figure C-2B**). We suspect that the multicopy nature of p1 in the OrthoRep system allowed for deleterious mutations that appeared toward the end of the experiment to be maintained for a period of time without experiencing purifying selection if they arose in the same cell as functional variants, explaining the presence of these low activity *Tm*TrpBs. Indeed, this multicopy “buffering” may have worked to our advantage by promoting genetic drift under selection, facilitating both greater adaptation and greater diversity of evolutionary pathways across replicates (see **Section 4.8 Discussion**). This may partly account for the high activity and high cryptic genetic variation present in the evolved *Tm*TrpBs.

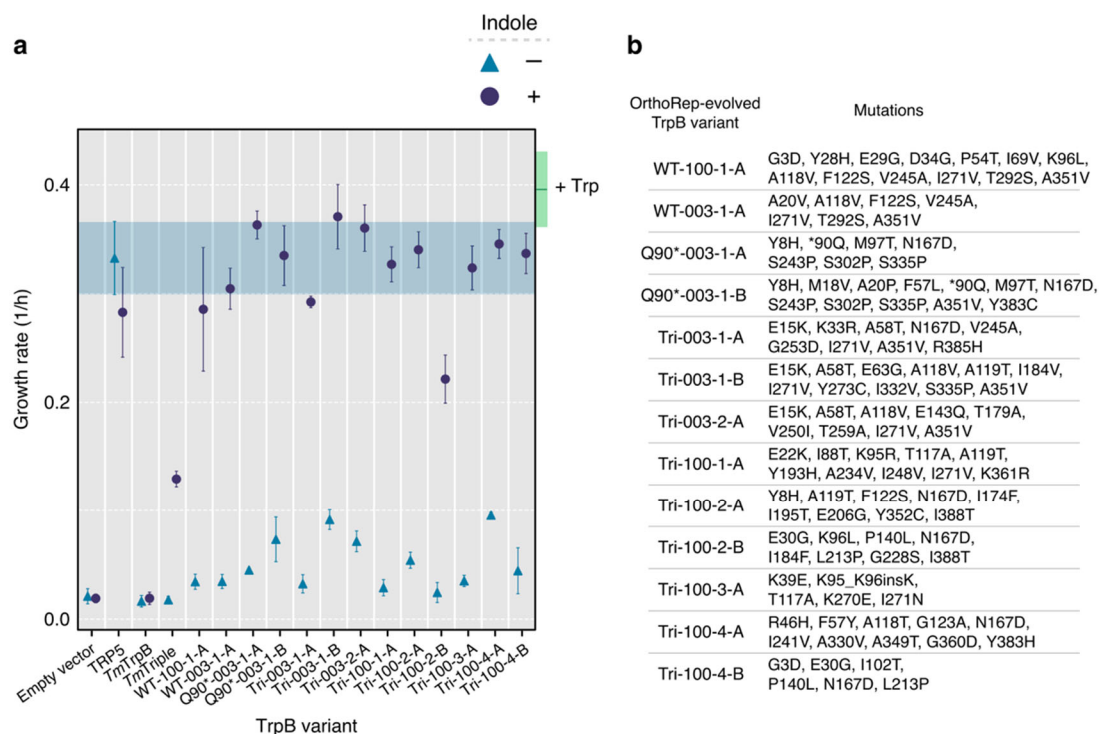


Figure 4-4. *In vivo* activity and diversity of individual *TmTrpB* variants from OrthoRep-evolved populations. (a) Evaluation of TRP5 complementation by evolved variants through a growth rate assay. Maximum growth rates over a 24-h period for $\Delta trp5$ yeast strains transformed with a nuclear plasmid expressing the indicated *TmTrpB* variant, grown in medium with or without 400 μ M indole. Points and error bars represent mean \pm s.d. for four biological replicates, respectively. Shaded area is the mean \pm s.d. growth rate for the TRP5 positive control (i.e. plasmid expressing the endogenous yeast TRP5). Green box indicates the mean \pm s.d. growth rate for all strains shown when Trp is supplemented. Growth rates for individual replicates in all three media conditions are shown in **Appendix C, Figure C-12**. Note that growth rates below ~ 0.15 per hour correspond to cultures that did not enter the exponential phase; in these cases, the reported growth rate is not meaningful and instead can be interpreted as no quantifiable growth. (b) Parent populations from which OrthoRep-evolved variants shown in (b) are derived and all non-synonymous mutations present in each.

4.5 Evolved *TmTrpB*s exhibit high primary and promiscuous activity *in vitro*

We further characterized the evolved *TmTrpB*s *in vitro* to approximate conditions of industrial application, make kinetic measurements, and test whether promiscuous activity could be detected. Nine *TmTrpB* variants were sampled from those that supported robust indole-dependent growth in the $\Delta trp5$ strain, cloned into an *E. coli* expression vector with a

C-terminal polyhistidine tag, and overexpressed. To mimic streamlined purification conditions compatible with biocatalytic application of *TmTrpBs*, we generated heat treated *E. coli* lysates (1 hour incubation at 75 °C) and tested them for their ability to couple indole and serine to produce Trp at 30 °C. Three of the nine OrthoRep-evolved *TrpBs*, WT-003-1-A, Q90*-003-1-A, and Tri-100-2-A, demonstrated improved activity over the benchmark *TmTriple*, as measured by total turnover number (TTN) (**Appendix C, Figure C-3**). Conveniently, each of these variants was evolved from a different starting point, meaning that wt *TmTrpB*, *TmTriple*, and *TmTripleQ90** were all viable starting points for reaching high activity *TmTrpBs*. (See **Appendix C, Table C-3** for an explanation of variant naming conventions where each name designates the source of the variant. For example, WT-003-1-A designates variant A taken from the first replicate of a 3-mL evolution experiment starting from wt *TmTrpB*.)

Since the benchmark *TmTriple* against which we compared the evolved *TmTrpBs* was engineered through classical directed evolution involving screening *E. coli* lysates, whereas our *TmTrpB* variants were evolved in yeast but expressed in *E. coli* for characterization, it is likely that the high-activity evolved *TmTrpBs* would compare even more favorably if normalized by expression. We therefore purified WT-003-1-A, Q90*-003-1-A, and Tri-100-2-A by immobilized metal affinity chromatography (IMAC) and reevaluated their activity for coupling indole with serine to generate Trp. By TTN, all three variants showed a 4- to 5-fold increase in activity over *TmTriple* at 30 °C (**Appendix C, Figure C-4**). At 75 °C, however, WT-003-1-A had only ~2-fold higher activity than *TmTriple*, while the other two variants were less active than *TmTriple*. Since the thermostability of WT-003-1-A, Q90*-003-1-A, and Tri-100-2-A had not been reduced dramatically ($T_{50} > 83.7$ °C; **Appendix C, Figure C-5**), adaptation in these variants occurred at least partially by shifting the activity temperature profile. This is a practically valuable adaptation, since thermostable enzymes that operate at mesophilic temperatures allow for greater versatility in application without sacrificing durability and ease of purification through heat treatment.

Further testing of WT-003-1-A, Q90*-003-1-A, and Tri-100-2-A revealed that all three enzymes had at least a 22-fold higher k_{cat}/K_M for indole than did *TmTriple* at 30 °C (Appendix C, Table C-4 and Figure C-6). Finally, testing for production of Trp analogs revealed that these variants' improved performance with indole transferred to alternate substrates (Appendix C, Figure C-4), validating their utility as versatile biocatalysts and also the hypothesis that continuous evolution of *TmTrpB* variants can uncover promiscuous activities for which they were not selected.

4.6 A diverse panel of evolved *TmTrpB* variants encompasses a variety of useful promiscuous activities with indole analogs

Given the exceptional performance of WT-003-1-A, Q90*-003-1-A, and Tri-100-2-A and their ability to transfer primary activity to new substrates as promiscuous activity, we decided to further sample the variant diversity generated across the multiple *TmTrpB* evolution experiments. We cloned 60 randomly chosen *TmTrpBs* from the ten continuous evolution populations into *E. coli* expression vectors for *in-vitro* characterization. These 60 *TmTrpBs* represent extensive diversity, with an average of 9.3 (\pm 2.8 s.d.) non-synonymous mutations per variant and a total of 194 unique amino acid changes across the set; in addition, each sequence encoded a unique protein (variant set 2; Appendix C, Tables C-2 and C-3). Since each variant had multiple non-synonymous mutations (up to 16) accumulated through >100 generations of adaptation and neutral drift, the depth of OrthoRep-based evolution was indeed leveraged in their evolution. We visualized these sequences, together with the consensus sequences of the populations from which they were derived, as nodes in a force directed graph related by shared mutations (Appendix C, Figure C-7). With only one exception, all individual sequences cluster near the consensus sequence for their population, meaning that interpopulation diversity exceeded intrapopulation diversity. Thus, the scale of OrthoRep-based evolution was also leveraged in these variants—if fewer independent evolution experiments had been run, the reduction in diversity would not be recoverable from sampling more clones.

Preparations of *Tm*TrpBs WT-003-1-A, Q90*-003-1-A, and Tri-100-2-A, the 60 new variants, and four top-performing TrpB benchmark variants from past classical directed evolution campaigns (including *Tm*Triple) were all tested for product formation with indole by UV absorption and nine indole analogs by high performance liquid chromatography-mass spectrometry (HPLC-MS) to detect substrate promiscuity (**Figure 4-5a**). The panel of 63 OrthoRep-evolved *Tm*TrpB variants exhibited an impressive range of activities (**Figure 4-5b**). First, we observed that a number of variants had primary activities with indole that surpass the benchmark *Tm*Triple in lysate, with initial velocities of Trp formation up to 3-fold higher than WT-003-1-A (**Figure 4-5b; Appendix C, Figure C-8**) whose k_{cat}/K_M for indole is $1.37 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$, already 28-fold higher than *Tm*Triple's at saturating serine concentrations (**Appendix C, Table C-4 and Figure C-6**). Second, direct comparison of some of the best panel variants to *Tm*Triple revealed dramatic general activity improvements for multiple indole analogs (**Figure 4-5b**). For example, across the three most versatile variants (Q90*-003-1-C, Tri-003-1-D, and Q90*-003-1-D) the maximum fold-improvement in product yields over *Tm*Triple were 37, 5, 19, and 50 using substrates 5-CN, 7-CN, 5-Br, and 6-Br, respectively (**Figure 4-5c**). Finally, with the exception of 6-Br and azulene, at least one variant from the OrthoRep-evolved panel converted the indole analog substrates as well as or better than benchmark TrpBs Pf2B9, *Tm*Azul, and *Tm*9D8*, which had been deliberately engineered toward new substrate scopes, though at higher temperatures (**Figure 4-5b; Appendix C, Figure C-9**).^{7,21,25,26}

The diverse properties represented in our 63 variants were not just limited to primary activity increases on indole and promiscuous activities for indole analogs. Multiple variants from the panel also exhibited substantial improvements in selectivity for differently substituted indoles, which could be useful when working with substrate mixtures that may be less expensive to use industrially. For example, we observed many *Tm*TrpBs with greater selectivity for 7-Br over 5-Br as compared to all four of the benchmark engineered TrpBs (**Figure 4-5d**). Another variant in the panel, Tri-100-1-G, stood out for having appreciable activity with nearly all substrates tested, including 6-CN and 5-CF₃, which are poorly utilized by most other TrpBs, likely due to electron-withdrawing effects of their respective moieties.

Notably, the ability to accept 5-CF3 as a substrate was unique to Tri-100-1-G: all other variants, as well as the benchmark TrpBs, showed no detectable product formation with this substrate (**Figure 4-5e; Appendix C, Figure C-9**). Repeating the reaction with purified enzyme in replicate confirmed the observed activity (**Appendix C, Figure C-10**). Tri-100-1-G may therefore be a promising starting point for new engineering efforts to access exotic Trp analogs. In short, despite having been selected for native activity with indole, OrthoRep-evolved *Tm*TrpBs have extensive and diverse activities on a range of non-native substrates, demonstrating the value of depth and scale in the evolution of enzyme variants.

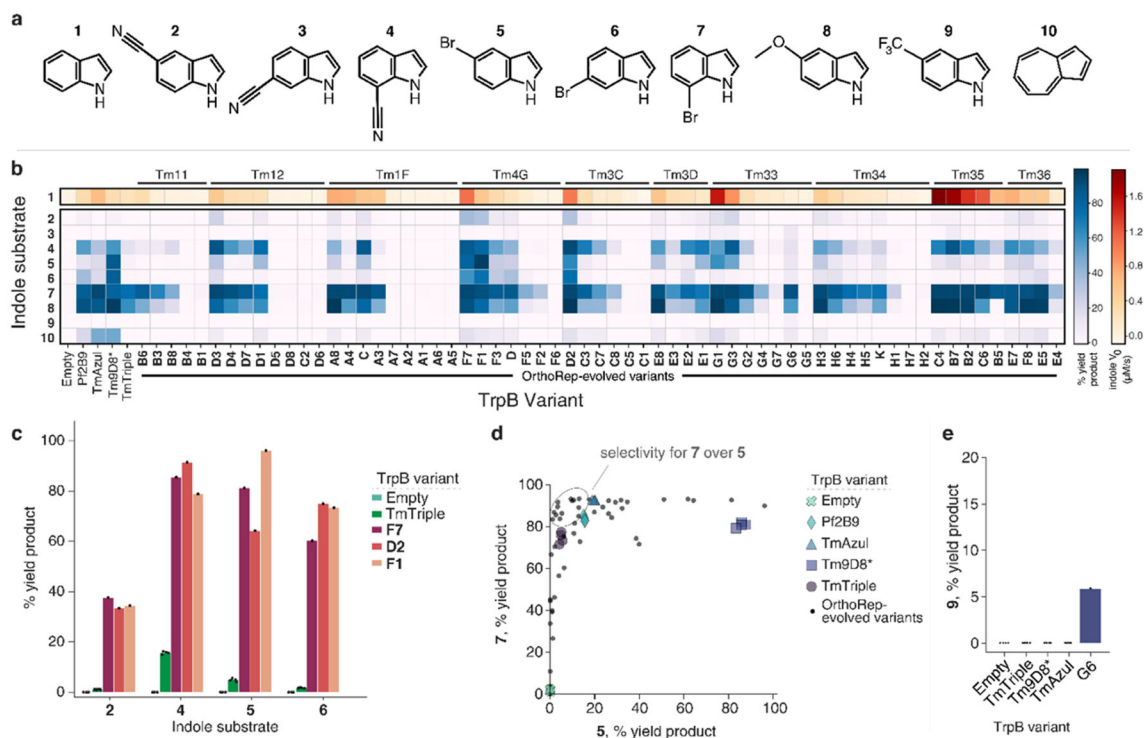


Figure 4-5. Promiscuous activities of a panel of evolved *TmTrpBs*. (a) Indole substrates used to test the substrate scope of a panel of *TmTrpB* variants. 5-CN, 5-cyanoindole; 6-CN, 6-cyanoindole; 7-CN, 7-cyanoindole; 5-Br, 5-bromoindole; 6-Br, 6-bromoindole; 7-Br, 7-bromoindole; 5-MeO, 5-methoxyindole; 5-CF₃, 5-trifluoromethylindole. (b) Heatmap of *TrpB* activities reported as yield of the *Trp* analog produced from indicated substrates where 100% yield corresponds to full conversion of the indole analog to the *Trp* analog. Reactions were carried out using heat treated (1 h at 75 °C) cell lysate, yield was measured by HPLC-MS, and V₀ is the initial rate of *Trp* formation from indole at saturating serine concentrations. Panel *TmTrpB* variants are ordered first by the parental cultures from which they were derived, then by activity with indole. Reactions with OrthoRep-evolved variants other than WT-003-1-A, Q90*-003-1-A, and Tri-100-2-A were performed in one replicate and all other reactions performed in quadruplicate. Empty designates expression vector without any *TrpB* encoded. (c) Bar graph of selected indole analog activities from panel b. Points represent percentage HPLC yield for individual replicates, bars represent mean yield for multiple replicates or yield for a single replicate. (d) Activities of all variants shown in **b** for reactions with substrates 7-Br and 5-Br to show selectivity. Individual replicates are shown for empty vector and benchmark *TrpBs* and only mean values are shown for OrthoRep-evolved variants tested in replicate (i.e. WT-003-1-A, Q90*-003-1-A, and Tri-100-2-A) for clarity. (e) Heat-treated lysate activity with 5-CF₃ for indicated *TrpB* variants.

4.7 Mutations in evolved *Tm*TrpBs may modulate conformational dynamics and fine tune the active site

Of the ~200 unique mutations in the OrthoRep-evolved *Tm*TrpBs that we characterized, there were some mutations whose effects could be rationalized from comparison to previous work. Since the *Tm*TrpBs had to evolve stand-alone activity, it is unsurprising that many of the mutations we observed have been implicated in the loss of allosteric regulation by TrpA. For example, Buller et al. previously examined a series of engineered variants from *Pyrococcus furiosus* TrpB (*Pf*TrpB) and found that evolution for stand-alone activity was facilitated by a progressive shift in the rate-limiting step from the first to the second stage of the catalytic cycle as well as stabilization of the ‘closed’ conformation of the enzyme.²⁷ That work implicated eight residues in this mechanism, seven of which correspond to homologous sites where we observed mutations in the evolved *Tm*TrpB variants (i.e., P14, M18, I69, K96, L274, T292, and T321). Another mutation, N167D, present in three of the ten consensus sequences for evolved populations (**Figure 4-3**), has also been implicated in stabilizing the closed state.²¹ Additional mutations observed but not studied before (e.g., S277F, S302P, and A321T) could also reasonably alter the allosteric network linking *Tm*TrpB activity to its natural *Tm*TrpA partner, based on existing structures and molecular dynamics simulations on the homologous *Pf*TrpA/*Pf*TrpB complex.^{22,27} Taken together, these mutations are likely implicated in converting allosteric activation by *Tm*TrpA into constitutive activity to establish stand-alone function of *Tm*TrpBs.

During the evolution of stand-alone activity, not only must allosteric activation by *Tm*TrpA be recapitulated by mutations in *Tm*TrpB, the surface of *Tm*TrpB that normally interacts with *Tm*TrpA must adjust to a new local environment. Consistent with this adaptation, all consensus sequences for the ten successfully evolved populations from which our *Tm*TrpB variants were sampled contain a mutation to at least one of a set of five residues located on the canonical TrpA interaction interface (**Figure 4-3; Appendix C, Figure C-11**). These mutations might improve solubility by increasing hydrophilicity (e.g. G3D, Y8H, and A20T) or form new intramolecular interactions that compensate for lost interactions with *Tm*TrpA, among other possibilities.

We also detected strong convergent evolution in a region near the catalytic lysine, K83, which directly participates in *TmTrpB*'s catalytic cycle through covalent binding of PLP and multiple proton transfers (**Appendix C, Figure C-12**).¹⁹ For example, A118 was mutated in the consensus sequence of four of the ten fully adapted populations, while adjacent residues T117 or A119 were mutated in an additional three (**Figure 4-3**). Furthermore, the three populations in which these residues were not mutated contained other consensus mutations that are either part of the α -helix to which K83 belongs, or, like residues 117–119, within ~ 8 Å of this helix (**Figure 4-3; Appendix C, Figure C-12**). We hypothesize that the α -helix harboring K83 is a focal point of evolution, whereby mutations in its vicinity may finely adjust the positioning of K83 and the PLP cofactor to improve catalysis, perhaps as compensation for structural changes induced by thermo adaptation. Some OrthoRep-evolved variants also contained mutations to first- and second-shell active site residues (**Appendix C, Figure C-13**), which may directly modulate the activity of *TmTrpBs*, although these mutations were rare. Taken together, we hypothesize that these mutations near the active site residues of TrpB were adaptive or compensatory.

The ~ 20 mutations considered above are rationalized with respect to their impact on *TmTrpB*'s primary catalytic activity. While substrate promiscuity changes may be influenced by these explainable mutations, previous literature suggests that substrate specificity is globally encoded by amino acids distributed across an entire enzyme.²⁸ Indeed, the majority of the ~ 200 mutations found in our panel of *TmTrpBs* were far away from *TmTrpB*'s active site and not rationalizable based on the known structural and kinetic properties of TrpBs. We suspect that the cryptic genetic variation this majority of mutations encompasses contributes to the diversity in substrate scope across our variants.

4.8 Discussion

In this work, we showed how the depth and scale of evolutionary search available in OrthoRep-driven protein evolution experiments could be applied to broaden the secondary promiscuous activities of *TmTrpB* while only selecting on its primary activity. The significance of this finding can be divided into two categories, one concerning the practical

utility of the new *Tm*TrpB variants we obtained and the second concerning how this evolution strategy may apply to future enzyme evolution campaigns and protein engineering in general.

Practically, the new *Tm*TrpBs should find immediate use in the synthesis of Trp analogs. Trp analogs are valuable chiral precursors to pharmaceuticals as well as versatile molecular probes, but their chemical synthesis is challenged by stereoselectivity requirements and functional group incompatibility. This has spurred enzyme engineers to evolve TrpB variants capable of producing Trp analogs,^{20,21,25,26} but the capabilities of available TrpBs are still limited. Compared to existing engineered TrpBs, our new panel of variants has substantially higher activity for the synthesis of Trp and Trp analogs at moderate temperatures from almost all indole analogs tested and also accepts indole analogs, such as 5-CF₃ (**Figure 4-5a**), for which benchmark TrpBs used in this study showed no detectable activity (**Figure 4-5e**). In fact, only one TrpB variant has shown detectable activity for this substrate in previous classical directed evolution campaigns.²¹ In addition, at least one member of the panel accepted each of the nine indole analogs we used to profile promiscuity, suggesting that additional indole analogs and non-indole nucleophiles not assayed here will also be accepted as substrates.^{29,30} Finally, the evolved *Tm*TrpBs are both thermostable and adapted for enzymatic activity at 30 °C. This maximizes their industrial utility, as thermostability predicts a protein's durability and can be exploited for simple heat-based purification processes, while mesophilic activity is compatible with heat-labile substrates, industrial processes where heating costs can compound, or *in-vivo* applications in model mesophilic hosts (e.g. *S. cerevisiae* or *E. coli*).

Of more general significance may be the process through which the *Tm*TrpBs in this study were generated. Previous directed evolution campaigns aiming to expand the substrate scope of TrpB screened directly for activity on indole analogs to guide the evolution process,^{21,26} whereas this study only selected for *Tm*TrpB's primary activity on indole. Yet this study still yielded *Tm*TrpBs whose secondary activities on indole analogs were both appreciable and diverse. Why?

A partial explanation may come from the high primary activities of OrthoRep-evolved *TmTrpB*s, as validated by kinetic measurements showing that variants tested have k_{cat}/K_M values for indole well in the $10^5 \text{ M}^{-1} \text{ s}^{-1}$ range. Since OrthoRep drove the evolution of *TmTrpB* in a continuous format for >100 generations, each resulting *TmTrpB* is the outcome of many rounds of evolutionary improvement and change (evolutionary depth). This contrasts with previous directed evolution campaigns using only a small number of manual rounds of diversification and screening. Continuous OrthoRep evolution, on the other hand, allowed *TmTrpB*s to become quite catalytically efficient with minimal researcher effort. We suggest that the high primary catalytic efficiencies also elevated secondary activities of *TmTrpB*, resulting in the efficient use of indole analogs. However, this explanation is not complete, as evolved *TmTrpB*s with similar primary activity on indole had differences in secondary activities (**Figure 4-5**). In other words, high primary activities did not uniformly raise some intrinsic set of secondary activities in *TmTrpB*, but rather influenced if not augmented the secondary activities of *TmTrpB* in different ways. We attribute this to the fact that we ran our evolution experiments in multiple independent replicates (evolutionary scale). Each replicate could therefore evolve the same primary activity through different mutational paths, the idiosyncrasies of which manifest as distinct secondary activities. A third explanation for the promiscuous profile diversity of these *TmTrpB* variants is that each replicate evolution experiment had, embedded within it, mechanisms to generate cryptic genetic variation without strong selection on primary activity. Many of the clones we sampled from each *TmTrpB* evolution experiment had novel promiscuity profiles but mediocre primary activity with indole (**Figure 4-5**). We believe this is because OrthoRep drove *TmTrpB* evolution in the context of a multicopy plasmid such that non-neutral genetic drift from high-activity sequences could occur within each cell at any given point. Therefore, *TmTrpB* sequences with fitness-lowering mutations could persist for short periods of time, potentially allowing for the crossing of fitness valleys during evolution experiments and, at the end of each evolution experiment, a broadening of the genetic diversity of clones even without explicitly imposed periods of relaxed selection. Since enzyme orthologs are capable of specializing toward different sets of secondary activities when pressured to do so,^{2,3} non-neutral genetic drift from different consensus sequences across independent population

should also access different secondary activities, further explaining the diversity of promiscuous activity profiles across clones selected from replicate evolution experiments. The combination of these mechanisms likely explains the variety of properties encompassed by the panel of *Tm*TrpBs.

Our approach to *Tm*TrpB evolution was inspired by the idea of gene orthologs in nature. Orthologs typically maintain their primary function while diversifying promiscuous activities through long evolutionary histories in different species.^{2,31} We approximated this by evolving *Tm*TrpB through continuous rounds of evolution, mimicking long histories, and in multiple replicates, mimicking the spatial separation and independence of species. Such depth and scale of evolutionary search is likely responsible for the substrate scope diversity of the *Tm*TrpBs we report even as they were selected only on their primary activity. We recognize that the evolved *Tm*TrpBs represent lower diversity than natural orthologs. For example, the median amino acid sequence divergence between orthologous human and mouse proteins is 11%,³² while the median divergence between pairs of variants from our experiment is 4.3% with a maximum of 8% (**Appendix C, Figure C-14**). Still, this level of divergence between functional variants is substantial for a laboratory protein evolution experiment and suggests that it is realistic to model future work on the processes of natural ortholog evolution (**Figure 4-6**). For example, it should be straightforward to scale our experiments further, to hundreds or thousands of independent populations each evolving over longer periods of time. This would better simulate the vastness of natural evolution. It should also be possible to deliberately vary selection schedules by adding competitive *Tm*TrpB inhibitors (such as the very indole analogs for which they have promiscuous activity), changing temperatures, or cycling through periods of weak and strong selection at different rates. Such evolutionary courses would approximate complexity in natural evolutionary histories. These modifications to OrthoRep-driven *Tm*TrpB evolution should yield greater cryptic genetic diversity, which may result in further broadening of promiscuous functions. The generation of cryptic genetic diversity at depth and scale should also be useful in efforts to predict protein folding and the functional effects of mutations via co-evolutionary analysis.^{33,34} Indeed, catalogs of natural orthologs have proven highly effective in fueling

such computational efforts, so our ability to mimic natural ortholog generation on laboratory timescales may be applicable to protein biology at large. Within the scope of enzyme engineering, we envision that the process of continuous replicate evolution, selecting only on primary activities of enzymes, will become a general strategy for expanding promiscuous activity ranges of enzymes as we and others extend it to new targets.

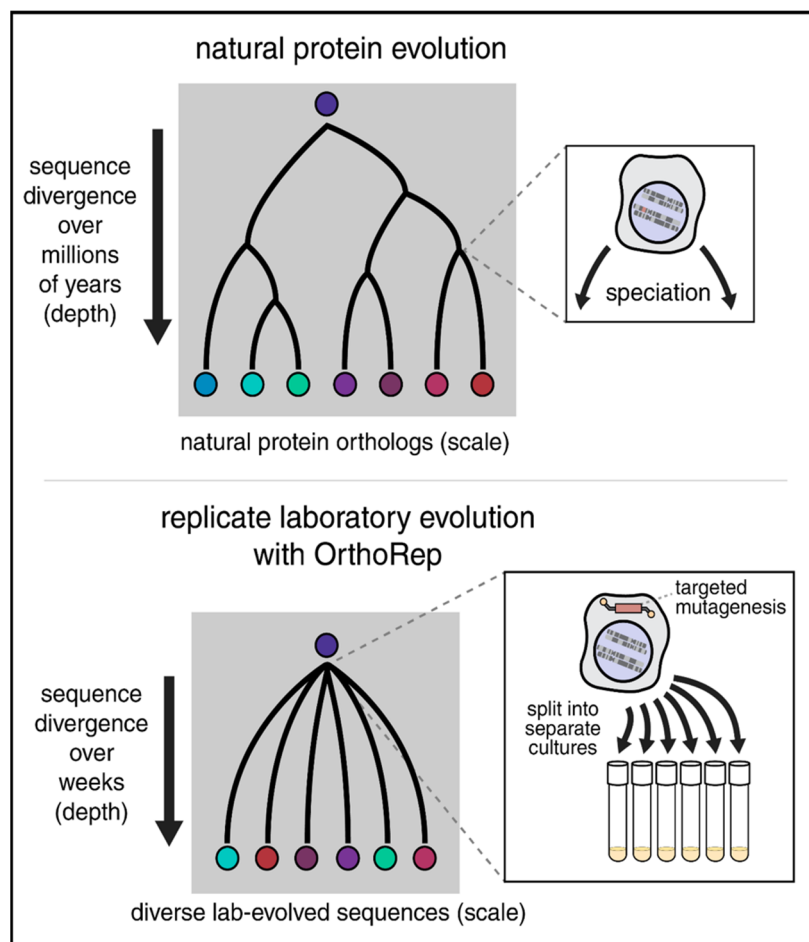


Figure 4-6. Conceptual similarities between natural enzyme ortholog evolution and OrthoRep evolution. Splitting OrthoRep cultures into many replicates can be seen as a form of speciation occurring by spatial separation. Complex selection schedules may emulate varied selection histories of natural orthologs, generating substantial sequence divergence across replicates. Evolved OrthoRep cultures contain diverse populations of sequences akin to quasi species owing to high mutation rates⁴¹.

Chapter IV Bibliography

1. Tatusov, R. L., Koonin, E. V., & Lipman, D. J. A genomic perspective on protein families. *Science* **278**, 631–637 (1997).
2. Khanal, A., McLoughlin, S. Y., Kershner, J. P., & Copley, S. D. Differential Effects of a Mutation on the Normal and Promiscuous Activities of Orthologs: Implications for Natural and Directed Evolution. *Mol. Biol. Evol.* **32**, 100–108 (2015).
3. Baier, F. et al. Cryptic genetic variation shapes the adaptive evolutionary potential of enzymes. *Elife* **8**, 1–20 (2019).
4. Aharoni, A. et al. The ‘evolvability’ of promiscuous protein functions. *Nat. Genet.* **37**, 73–76 (2005).
5. Tawfik, O. K. and D. S. Enzyme Promiscuity: A Mechanistic and Evolutionary Perspective. *Annu. Rev. Biochem.* **79**, 471–505 (2010).
6. Leveson-Gower, R. B., Mayer, C., & Roelfes, G. The importance of catalytic promiscuity for enzyme design and evolution. *Nat. Rev. Chem.* **3**, 687–705 (2019).
7. Murciano-Calles, J., Romney, D. K., Brinkmann-Chen, S., Buller, A. R., & Arnold, F. H. A Panel of TrpB Biocatalysts Derived from Tryptophan Synthase through the Transfer of Mutations that Mimic Allosteric Activation. *Angew. Chemie - Int. Ed.* **55**, 11577–11581 (2016).
8. Devine, P. N. et al. Extending the application of biocatalysis to meet the challenges of drug development. *Nat. Rev. Chem.* **2**, 409–421 (2018).
9. Truppo, M. D. Biocatalysis in the Pharmaceutical Industry: The Need for Speed. *ACS Med. Chem. Lett.* acsmedchemlett.7b00114 (2017). doi:10.1021/acsmedchemlett.7b00114
10. Almhjell, P. J., Boville, C. E., & Arnold, F. H. Engineering enzymes for noncanonical amino acid synthesis. *Chem. Soc. Rev.* **47**, 8980–8997 (2018).
11. Zheng, J., Payne, J. L., & Wagner, A. Cryptic genetic variation accelerates evolution by opening access to diverse adaptive peaks. *Science* **365**, 347–353 (2019).
12. Gupta, R. D. & Tawfik, D. S. Directed enzyme evolution via small and effective neutral drift libraries. *Nat. Methods* **5**, 939–942 (2008).
13. Bloom, J. D., Romero, P. A., Lu, Z., & Arnold, F. H. Neutral genetic drift can alter

- promiscuous protein functions, potentially aiding functional evolution. *Biol. Direct* **2**, 7–10 (2007).
14. Bershtein, S., Goldin, K., & Tawfik, D. S. Intense Neutral Drifts Yield Robust and Evolvable Consensus Proteins. *J. Mol. Biol.* **379**, 1029–1044 (2008).
 15. Ravikumar, A., Arrieta, A., & Liu, C. C. An orthogonal DNA replication system in yeast. *Nat. Chem. Biol.* **10**, 175–177 (2014).
 16. Ravikumar, A., Arzumanyan, G. A., Obadi, M. K. A., Javanpour, A. A., & Liu, C. C. Scalable, Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds. *Cell* **175**, 1946-1957.e13 (2018).
 17. Packer, M. S. & Liu, D. R. Methods for the directed evolution of proteins. *Nat. Rev. Genet.* **16**, 379–394 (2015).
 18. Zhong, Z. et al. Automated Continuous Evolution of Proteins in Vivo . *ACS Synth. Biol.* (2020). doi:10.1021/acssynbio.0c00135
 19. Dunn, M. F. Allosteric regulation of substrate channeling and catalysis in the tryptophan synthase holoenzyme complex. *Arch. Biochem. Biophys.* **519**, 154–166 (2012).
 20. Buller, A. R. et al. Directed evolution of the tryptophan synthase β -subunit for stand-alone function recapitulates allosteric activation. *Proc. Natl. Acad. Sci.* **112**, 14599–14604 (2015).
 21. Romney, D. K., Murciano-Calles, J., Wehrmüller, J. E., & Arnold, F. H. Unlocking Reactivity of TrpB: A General Biocatalytic Platform for Synthesis of Tryptophan Analogues. *J. Am. Chem. Soc.* **139**, 10769–10776 (2017).
 22. Maria-Solano, M. A., Iglesias-Fernández, J., & Osuna, S. Deciphering the Allosterically Driven Conformational Ensemble in Tryptophan Synthase Evolution. *J. Am. Chem. Soc.* **141**, 13049–13056 (2019).
 23. Zhong, Z., Ravikumar, A., & Liu, C. C. Tunable Expression Systems for Orthogonal DNA Replication. *ACS Synth. Biol.* **7**, 2930–2934 (2018).
 24. Lee, M. E., DeLoache, W. C., Cervantes, B., & Dueber, J. E. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth. Biol.* **4**, 975–986 (2015).
 25. Watkins, E. J., Almhjell, P. J., & Arnold, F. H. Direct enzymatic synthesis of a deep-

blue fluorescent noncanonical amino acid from azulene and serine.

ChemBioChem **21**, 80–83 (2020).

26. Boville, C. E., Romney, D. K., Almhjell, P. J., Sieben, M., & Arnold, F. H. Improved Synthesis of 4-Cyanotryptophan and Other Tryptophan Analogues in Aqueous Solvent Using Variants of TrpB from *Thermotoga maritima*. *J. Org. Chem.* **83**, 7447–7452 (2018).
27. Buller, A. R. et al. Directed Evolution Mimics Allosteric Activation by Stepwise Tuning of the Conformational Ensemble. *J. Am. Chem. Soc.* **140**, 7256–7266 (2018).
28. Wrenbeck, E. E., Azouz, L. R., & Whitehead, T. A. Single-mutation fitness landscapes for an enzyme on multiple substrates reveal specificity is globally encoded. *Nat. Commun.* **8**, 1–10 (2017).
29. Dick, M., Sarai, N. S., Martynowycz, M. W., Gonen, T., & Arnold, F. H. Tailoring Tryptophan Synthase TrpB for Selective Quaternary Carbon Bond Formation. *J. Am. Chem. Soc.* **141**, 19817–19822 (2019).
30. Romney, D. K., Sarai, N. S., & Arnold, F. H. Nitroalkanes as Versatile Nucleophiles for Enzymatic Synthesis of Noncanonical Amino Acids. *ACS Catal.* **9**, 8726–8730 (2019).
31. O’Maille, P. E. et al. Quantitative exploration of the catalytic landscape separating divergent plant sesquiterpene synthases. *Nat. Chem. Biol.* **4**, 617–623 (2008).
32. Makałowski, W., & Boguski, M. S. Evolutionary parameters of the transcribed mammalian genome: An analysis of 2,820 orthologous rodent and human sequences. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 9407–9412 (1998).
33. Marks, D. S. et al. Protein 3D structure computed from evolutionary sequence variation. *PLoS One* **6**, (2011).
34. Stiffler, M. A. et al. Protein Structure from Experimental Evolution. *Cell Syst.* **10**, 15-24.e5 (2020).

SUPPLEMENTARY INFORMATION FOR CHAPTER IV

C.1 Supplementary tables

Table C-1. Summary of all cultures passaged for evolution of *TmTrpB* variants.

TrpB variant (strain)	Culture volume (mL)	Total number passaged	Total number successful
wt <i>TmTrpB</i> (GR-Y053)	3	4	1
	100	2	2
<i>TmTriple</i> (GR-Y055)	3	8	2
	100	4	4
<i>TmTripleQ90*</i> (GR-Y057)	3	8	1
	100	0	0

Table C-2. Mutation summary statistics for OrthoRep-evolved *TrpB* populations.

variant set	total number of sequences	non-synonymous mutations			synonymous mutations	
		total unique mutations	mean	standard deviation	mean	standard deviation
consensus	10	43	5.6	2.3	3.0	2.3
1	16	85	8.7	2.1	6.3	4.4
2	60	194	9.3	2.8	6.5	3.0

Table C-3. Mutations and identification information for all individual *TmTrpB* sequences.

variant set	variant ID	variant name	evolution population	number of non-synonymous mutations	number of synonymous mutations	starting sequence
1	A	Tm11g	Tm11	13	3	TmTrpB
1	B	Tm12f	Tm12	9	7	TmTrpB
1	C	Tm1Ff	Tm1F	7	1	TmTrpB
1	D	Tm4Ge	Tm4G	6	11	<i>TmTripleQ90*</i>
1	E	Tm4Gf	Tm4G	11	15	<i>TmTripleQ90*</i>
1	F	Tm3Cc	Tm3C	9	3	<i>TmTriple</i>
1	G	Tm3Cf	Tm3C	9	3	<i>TmTriple</i>

1	H	Tm3Cg	Tm3C	11	4	TmTriple
1	R	Tm3Df	Tm3D	9	6	TmTriple
1	J	Tm33f	Tm33	10	5	TmTriple
1	K	Tm34e	Tm34	9	7	TmTriple
1	L	Tm34f	Tm34	8	6	TmTriple
1	M	Tm35f	Tm35	5	3	TmTriple
1	N	Tm36b	Tm36	7	11	TmTriple
1	O	Tm36e	Tm36	10	1	TmTriple
1	P	Tm36f	Tm36	6	14	TmTriple
2	B1	Tm11-B1	Tm11	16	11	wt TmTrpB
2	B3	Tm11-B3	Tm11	12	7	wt TmTrpB
2	B4	Tm11-B4	Tm11	14	10	wt TmTrpB
2	B6	Tm11-B6	Tm11	14	9	wt TmTrpB
2	B8	Tm11-B8	Tm11	12	9	wt TmTrpB
2	D5	Tm12-D5	Tm12	9	10	wt TmTrpB
2	D4	Tm12-D4	Tm12	7	5	wt TmTrpB
2	D8	Tm12-D8	Tm12	7	6	wt TmTrpB
2	D7	Tm12-D7	Tm12	9	8	wt TmTrpB
2	D6	Tm12-D6	Tm12	10	5	wt TmTrpB
2	D3	Tm12-D3	Tm12	7	9	wt TmTrpB
2	C2	Tm12-C2	Tm12	9	9	wt TmTrpB
2	D1	Tm12-D1	Tm12	8	9	wt TmTrpB
2	A1	Tm1F-A1	Tm1F	9	4	wt TmTrpB
2	A7	Tm1F-A7	Tm1F	10	8	wt TmTrpB
2	A6	Tm1F-A6	Tm1F	8	4	wt TmTrpB
2	A5	Tm1F-A5	Tm1F	8	6	wt TmTrpB
2	A4	Tm1F-A4	Tm1F	7	5	wt TmTrpB
2	A3	Tm1F-A3	Tm1F	7	7	wt TmTrpB
2	A2	Tm1F-A2	Tm1F	8	4	wt TmTrpB
2	A8	Tm1F-A8	Tm1F	6	2	wt TmTrpB
2	F3	Tm4G-F3	Tm4G	14	8	TmTripleQ90*
2	F5	Tm4G-F5	Tm4G	14	8	TmTripleQ90*
2	F6	Tm4G-F6	Tm4G	11	9	TmTripleQ90*
2	F7	Tm4G-F7	Tm4G	9	6	TmTripleQ90*
2	F2	Tm4G-F2	Tm4G	16	6	TmTripleQ90*
2	F1	Tm4G-F1	Tm4G	11	9	TmTripleQ90*
2	C8	Tm3C-C8	Tm3C	10	5	TmTriple
2	C5	Tm3C-C5	Tm3C	15	6	TmTriple
2	D2	Tm3C-D2	Tm3C	9	4	TmTriple

2	C1	Tm3C-C1	Tm3C	6	4	TmTriple
2	C7	Tm3C-C7	Tm3C	12	3	TmTriple
2	C3	Tm3C-C3	Tm3C	7	5	TmTriple
2	E8	Tm3D-E8	Tm3D	8	11	TmTriple
2	E3	Tm3D-E3	Tm3D	13	4	TmTriple
2	E1	Tm3D-E1	Tm3D	10	3	TmTriple
2	E2	Tm3D-E2	Tm3D	12	8	TmTriple
2	G1	Tm33-G1	Tm33	5	5	TmTriple
2	G2	Tm33-G2	Tm33	9	6	TmTriple
2	G3	Tm33-G3	Tm33	6	5	TmTriple
2	G4	Tm33-G4	Tm33	10	8	TmTriple
2	G7	Tm33-G7	Tm33	7	2	TmTriple
2	G5	Tm33-G5	Tm33	8	4	TmTriple
2	G6	Tm33-G6	Tm33	6	5	TmTriple
2	H3	Tm34-H3	Tm34	8	5	TmTriple
2	H6	Tm34-H6	Tm34	10	15	TmTriple
2	H4	Tm34-H4	Tm34	12	10	TmTriple
2	H7	Tm34-H7	Tm34	10	10	TmTriple
2	H2	Tm34-H2	Tm34	13	4	TmTriple
2	H1	Tm34-H1	Tm34	9	15	TmTriple
2	H5	Tm34-H5	Tm34	9	10	TmTriple
2	C6	Tm35-C6	Tm35	6	2	TmTriple
2	B2	Tm35-B2	Tm35	8	4	TmTriple
2	B5	Tm35-B5	Tm35	5	2	TmTriple
2	B7	Tm35-B7	Tm35	6	3	TmTriple
2	C4	Tm35-C4	Tm35	7	3	TmTriple
2	E5	Tm36-E5	Tm36	6	8	TmTriple
2	E7	Tm36-E7	Tm36	8	7	TmTriple
2	E4	Tm36-E4	Tm36	6	6	TmTriple
2	F8	Tm36-F8	Tm36	7	5	TmTriple

Table C-4. Kinetic parameters of selected *Tm*TrpB variants at 30 °C.

variant	k_{cat} [95% credible region] (s^{-1})	K_M [95% credible region] (μM)	k_{cat}/K_M [95% credible region] ($mM^{-1} s^{-1}$)
<i>Tm</i>Triple	0.2 [0.16, 0.31]	41.23 [14.32, 192.66]	4.89 [1.54, 12.12]
C	0.53 [0.49, 0.58]	3.89 [1.85, 7.99]	137.22 [70.29, 276.04]
D	0.77 [0.72, 0.83]	5.79 [3.82, 8.8]	133.38 [91.24, 193.71]
K	0.62 [0.59, 0.66]	5.58 [3.99, 7.91]	111.89 [81.52, 152.25]

C.2 Supplementary figures

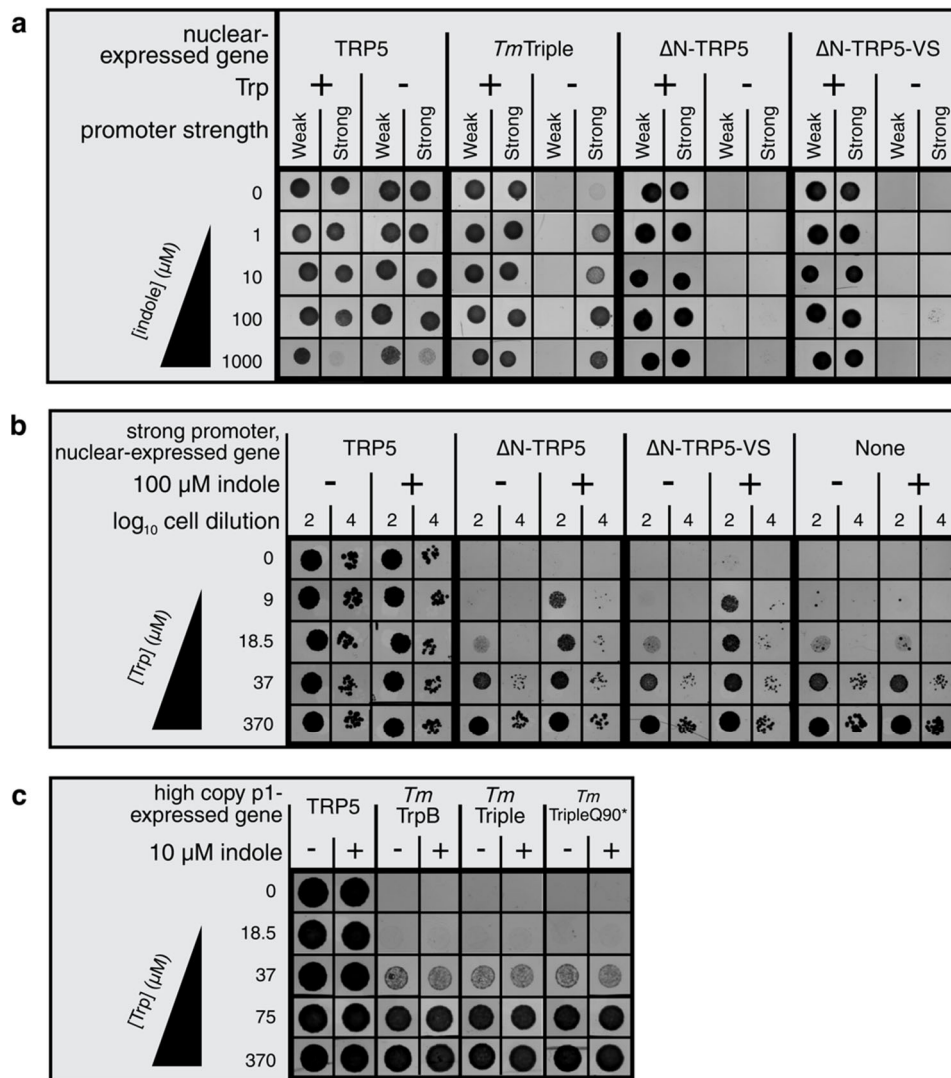


Figure C-1. Evaluation of indole-dependent TRP5 complementation of TrpB variants. (a)-(c) Spot plating assays for $\Delta trp5$ yeast strains expressing TrpB variants from a nuclear plasmid under two different promoter strengths (a), from a nuclear plasmid under a strong promoter (b), or from p1 at a high copy number (wt TP-DNAP1 expressed *in trans*) (c) grown on indicated growth medium.

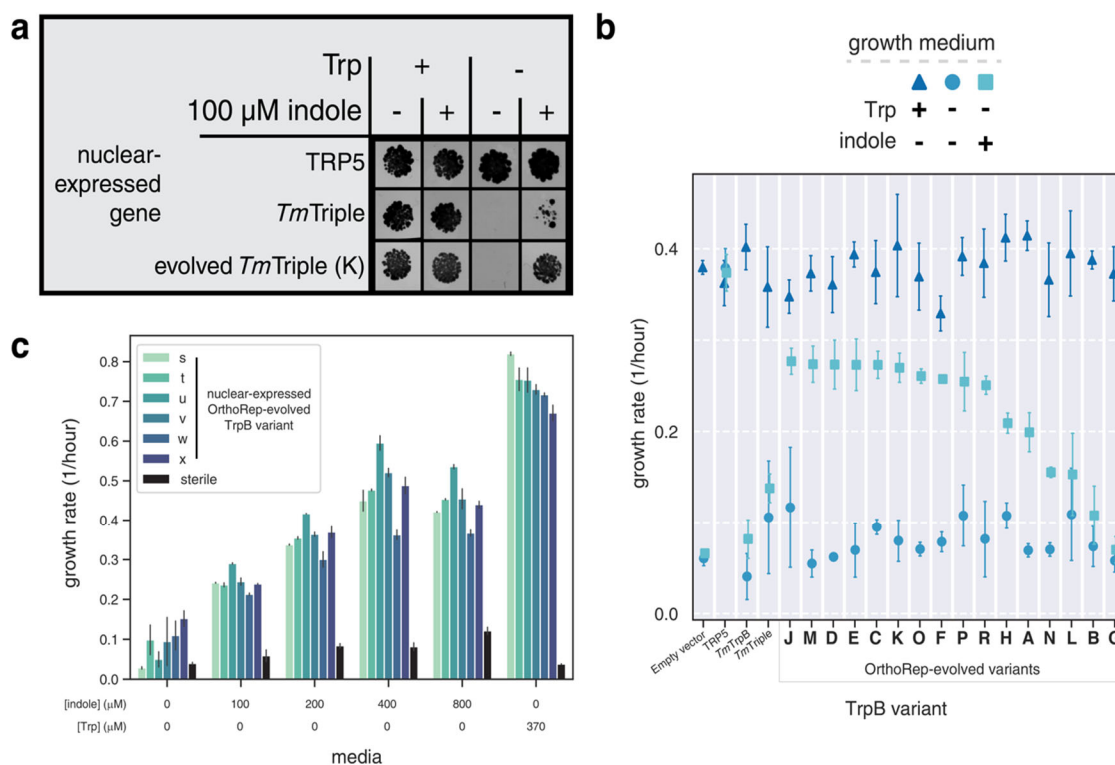


Figure C-2. In vivo Trp production by evolved TrpBs. (a) Spot plating assay for TRP5-deleted yeast with a nuclear plasmid expressing TRP5, *TmTriple*, or an individual OrthoRep-evolved *TmTrpB* variant driven by a promoter (pRNR2) that approximates expression of *TmTrpB* variants from OrthoRep's p1 plasmid, grown on indicated media. (b) Evaluation of TRP5 complementation by evolved variants through a growth rate assay. Maximum growth rates over a 24-hour period for $\Delta trp5$ yeast strains transformed with a nuclear plasmid expressing the indicated *TmTrpB* variant, grown in medium with or without 100 μ M indole, and with or without Trp, as indicated. Points and error bars represent mean \pm s.d. of four technical replicates, respectively. (c) Optimization of indole concentration. Growth rates during exponential growth for $\Delta trp5$ yeast with a nuclear plasmid expressing randomly chosen OrthoRep-evolved *TrpB* variants (exact sequence not determined), supplemented with one of four indole concentrations, Trp, or neither of these. Bars and error bars represent mean \pm s.d. of four technical replicates, respectively.

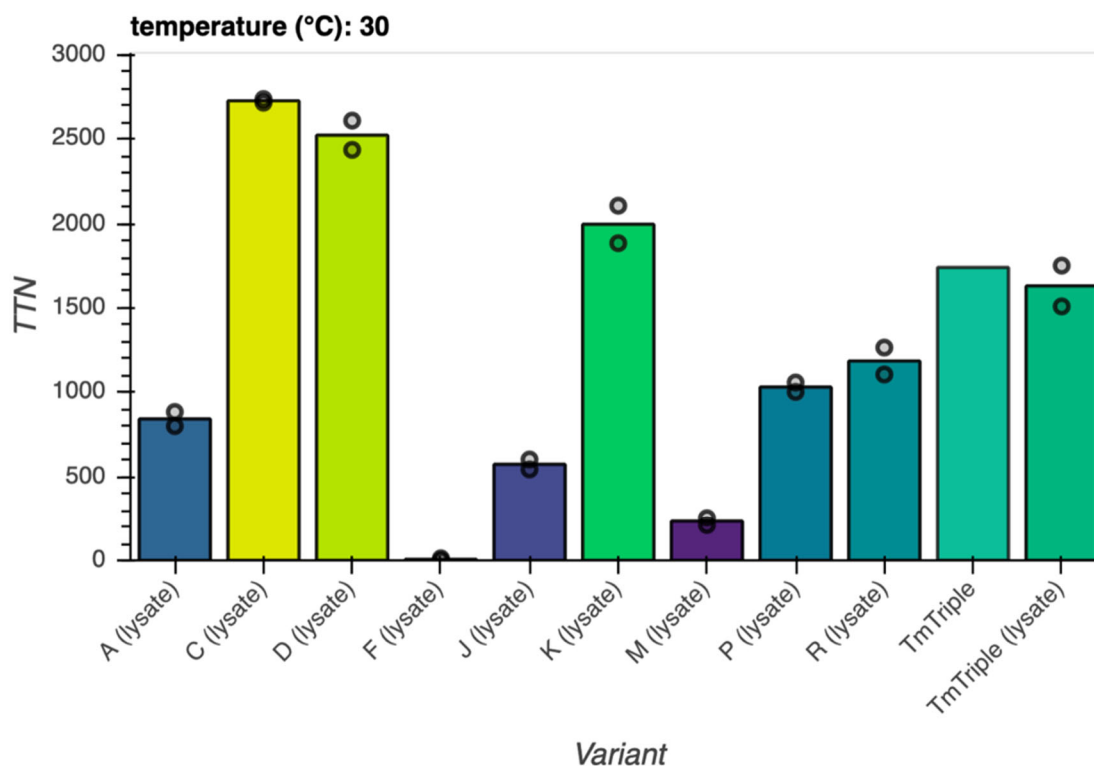


Figure C-3. *In vitro* Trp production by evolved TrpBs with heat-treated lysate. Trp production at 30 °C by indicated TmTrpB variants. Reactions with TmTriple were performed with both heat-treated lysate and with purified protein, while all other reactions were performed only with heat-treated lysate. TTN, total turnover number. Maximum TTN is 5,000. Points represent TTN for individual replicates, bars represent mean TTN for reactions with replicates, or TTN for a single replicate otherwise.

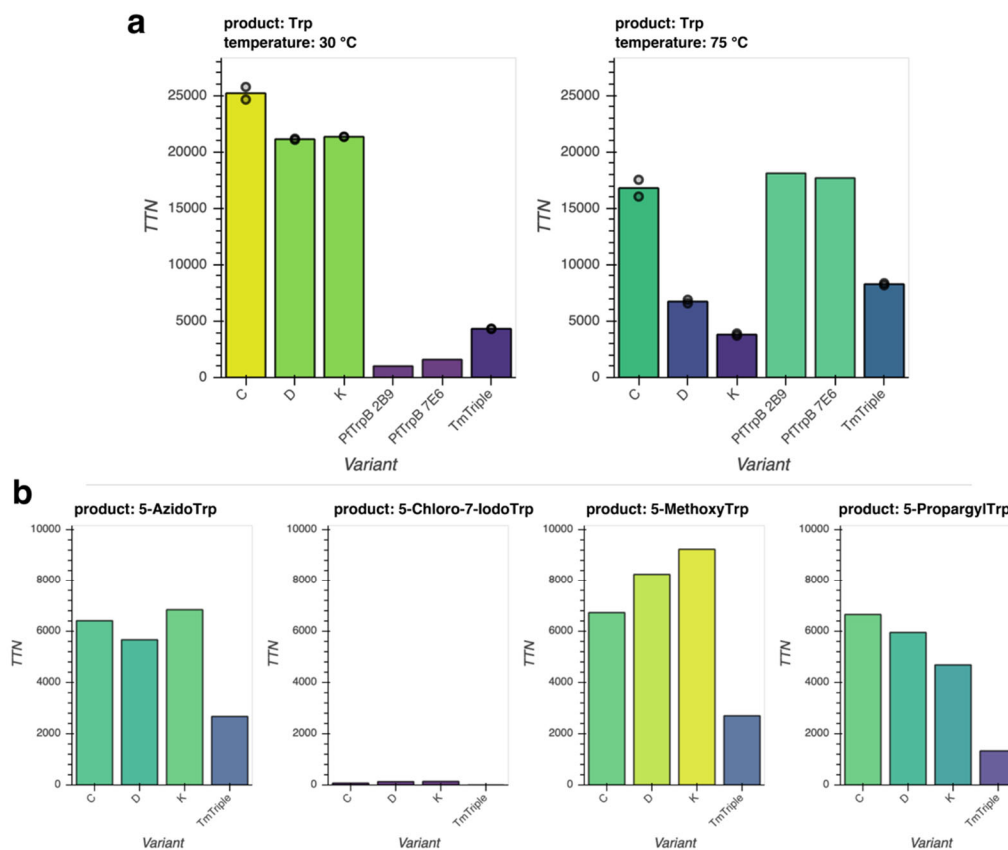


Figure C-4. *In vitro* Trp and Trp analog production with purified enzyme. (a)-(b) Production of (a) Trp at 30 °C or 75 °C with 40,000 maximum TTN or (b) indicated Trp analogs at 30 °C by column-purified TmTrpB variants. TTN, total turnover number with 10,000 as maximum TTN. Points represent TTN for individual replicates, bars represent mean TTN for reactions with replicates, or TTN for a single replicate otherwise.

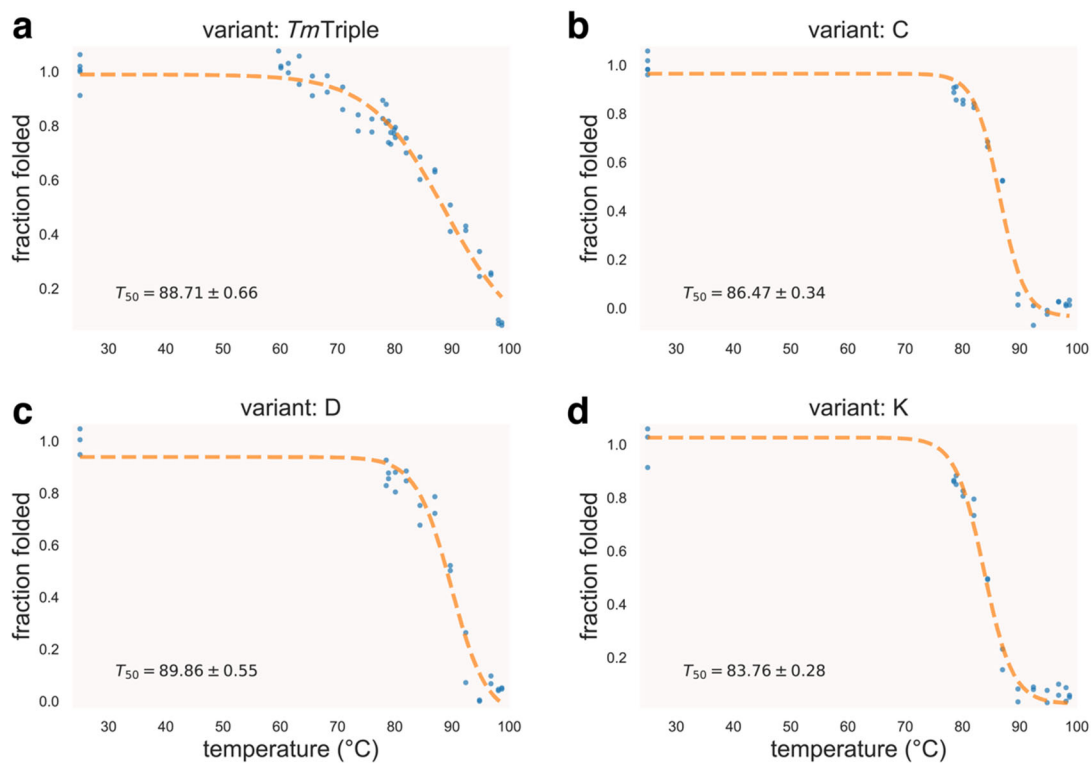


Figure C-5. Thermal shift assay on various *TmTrpBs*. Proportion of *TmTrpB* variants *TmTriple* (a), C (b), D (c), or K (d) that remain folded after incubation at indicated temperature for 1 hour, as measured by the fraction of Trp production relative to incubation at 25 °C. T_{50} , temperature at which 50% of enzyme is irreversibly inactivated, as estimated by best fit logistic model (dotted line). Each temperature tested in duplicate.

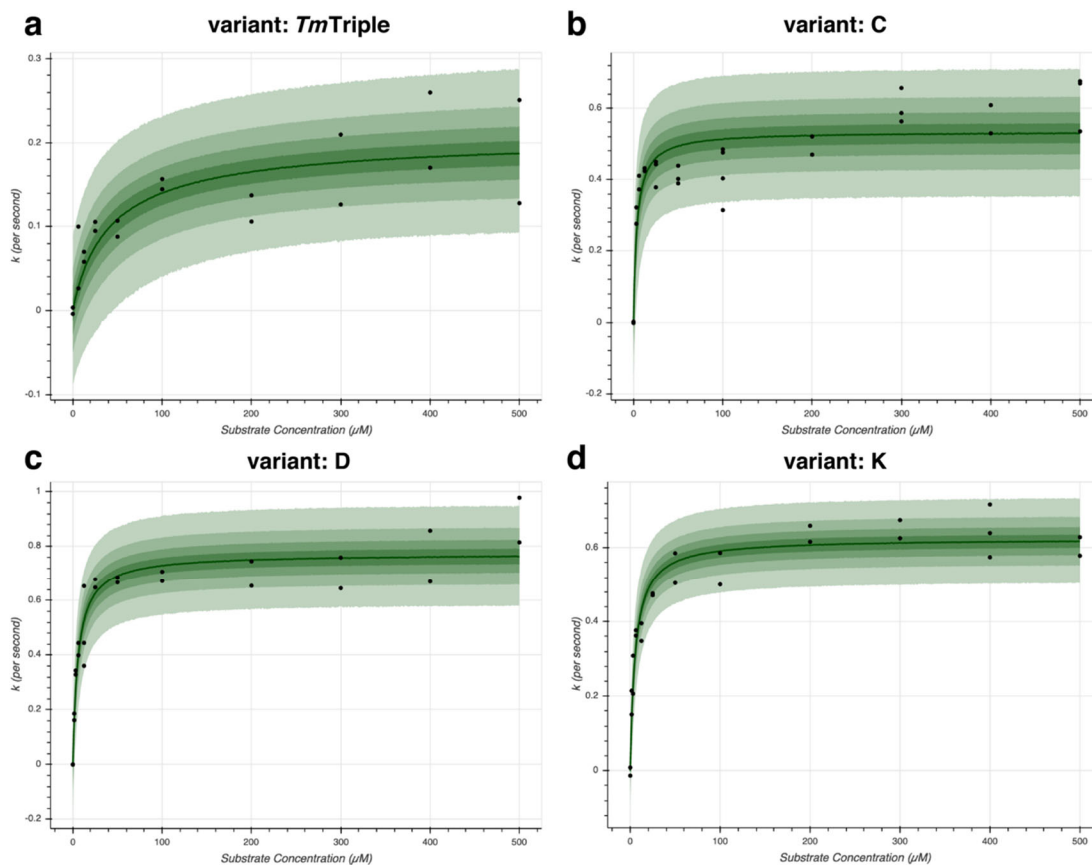


Figure C-6. Michaelis-Menten plots for rate of Trp production at saturating serine for evolved *TmTrpB* variants. Initial rate of Trp formation (k , per second) with TrpB variants *TmTriple* (a), C (b), D (c), or K (d) at saturating serine concentration (40 mM) vs. indole concentration. Points, median estimates for initial rate based on absorbance change over time (see **Methods Section B3**). The median estimated Michaelis-Menten curve is shown as a dark green line, with the 25, 50, 75, and 95% credible regions displayed from dark to light green, respectively. All measurements performed in at least duplicate.

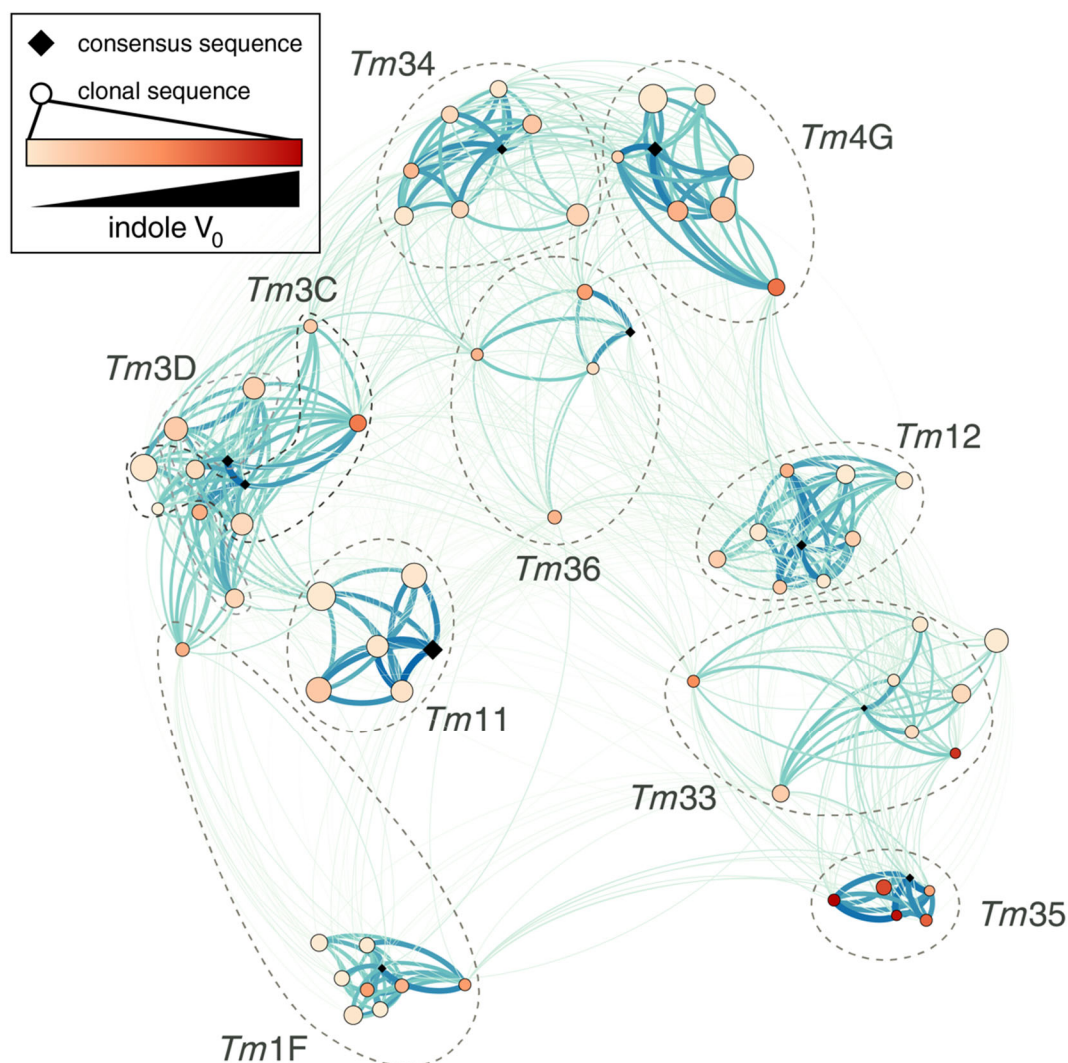


Figure C-7. Relatedness of TrpB panel sequences generated by OrthoRep evolution. Force directed graph where each node represents an individual sequence (all variants from set 2, and variants C, D, and K) or consensus sequence for one of the ten evolved populations. Edge weights are proportional to the number of shared mutations between two nodes. Higher edge weight yields a stronger attractive force between two nodes and is visualized as a darker color and a thicker line. Nodes for individual sequences are colored according to initial rate of Trp formation, similar to **Figure 4-5C**. Dotted lines are drawn around consensus sequences and individual sequences that are derived from the same evolved culture, if nodes are sufficiently clustered to allow it.

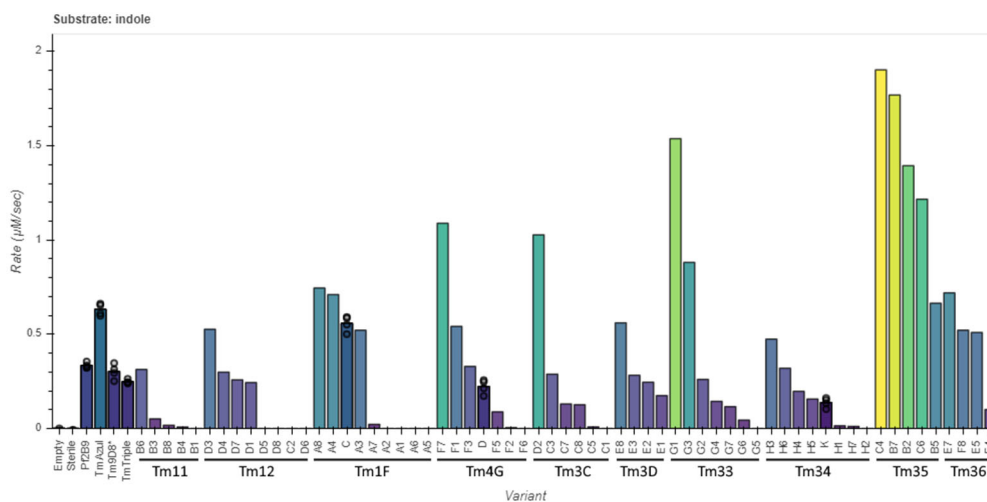
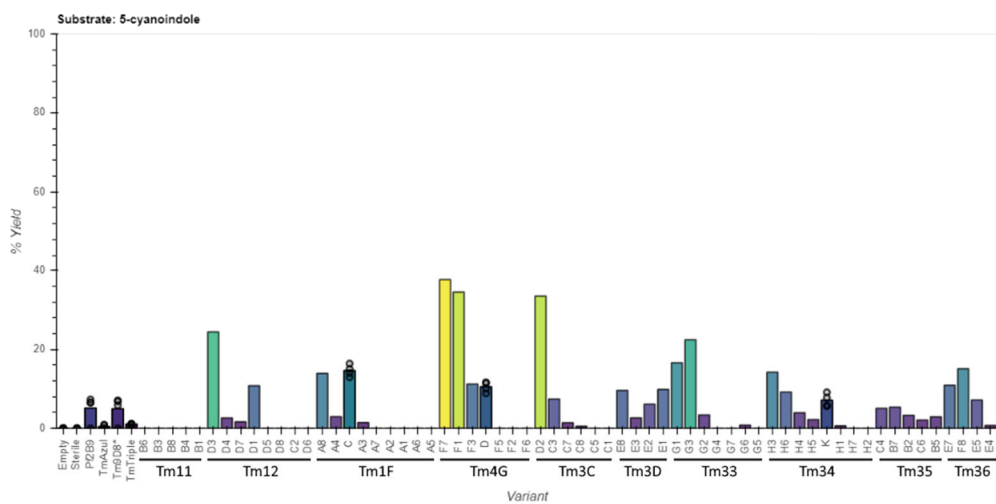
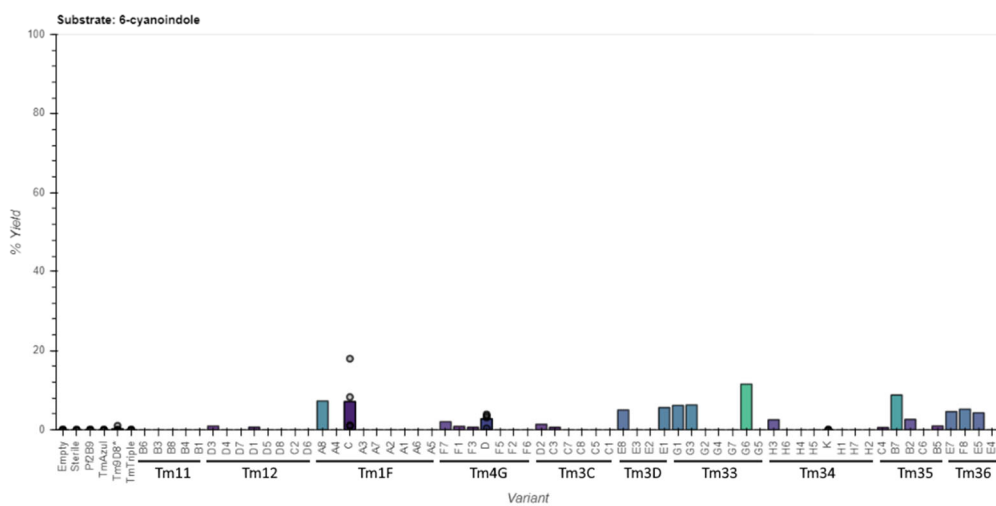
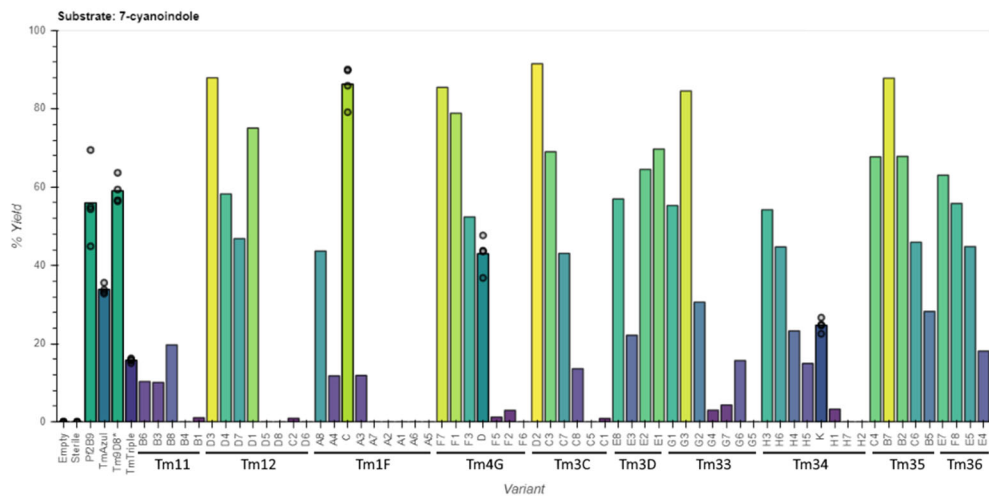


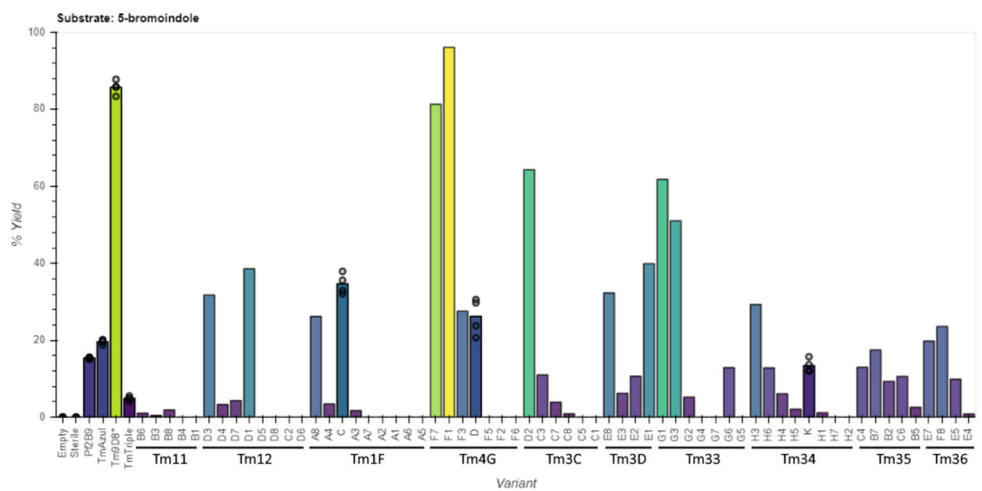
Figure C-8. TrpB panel indole activity by initial rate of Trp formation. Initial rate of Trp formation at saturating l-serine by UV-vis spectrophotometry. Points represent rate for individual replicates, bars represent mean rate for reactions with multiple replicates, or rate for a single replicate otherwise. OrthoRep-evolved variants are ordered first by the population from which they were derived, then by indole activity. Empty, expression vector without TrpB variant. Sterile, reaction master mix without heat-treated lysate added. Empty, Pf2B9, TmAzul, Tm9D8*, TmTriple, C, D, and K all performed in quadruplicate; sterile performed in duplicate; all other reactions performed in a single replicate.

a**b**

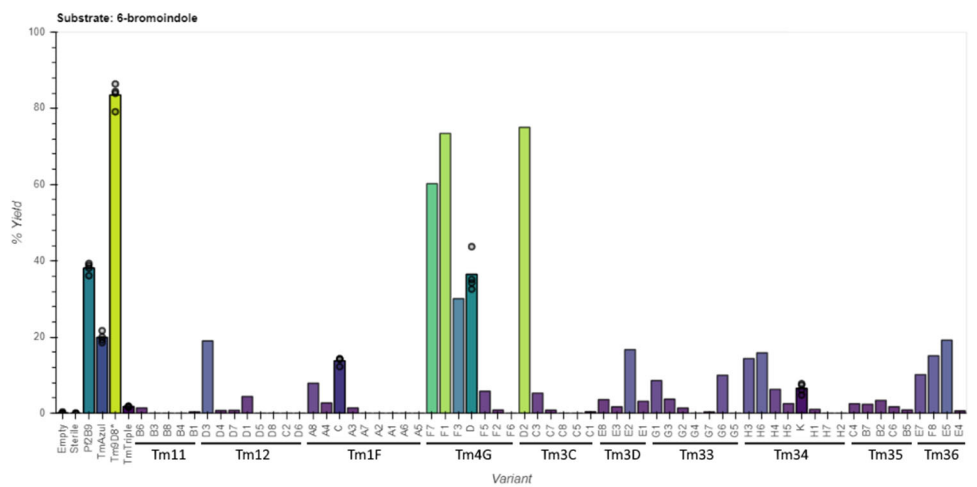
c



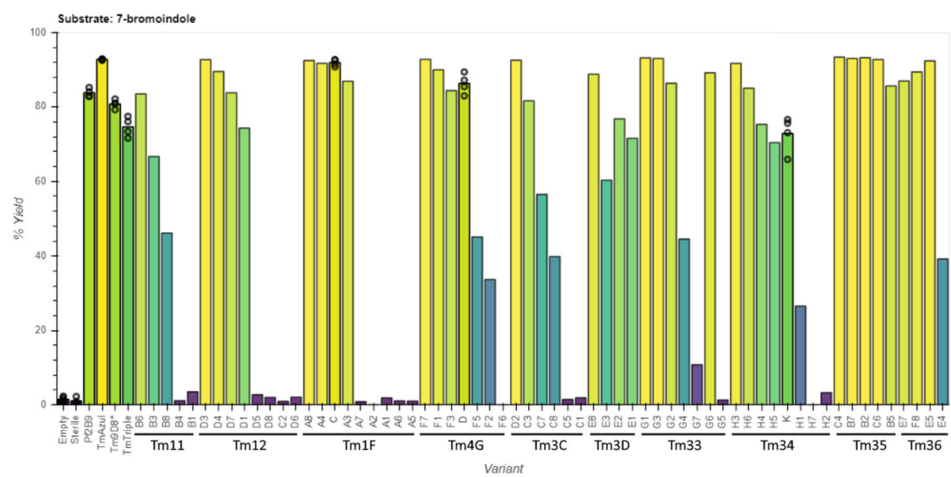
d



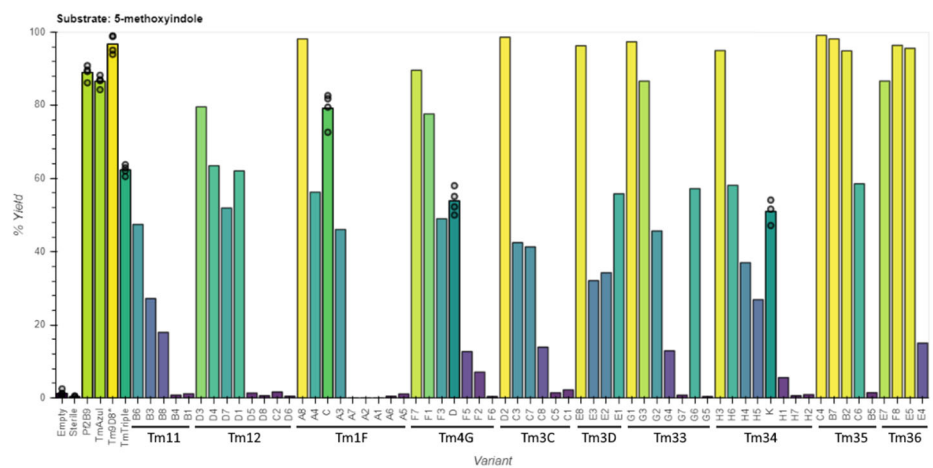
e



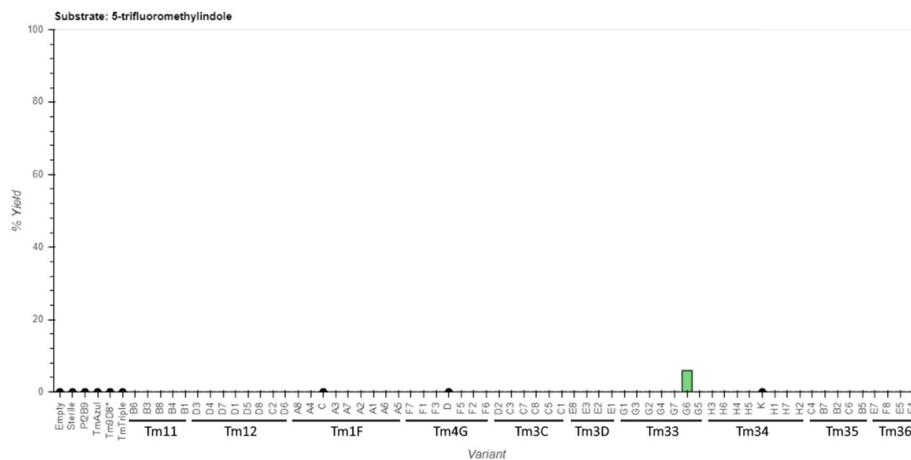
f



g



h



i

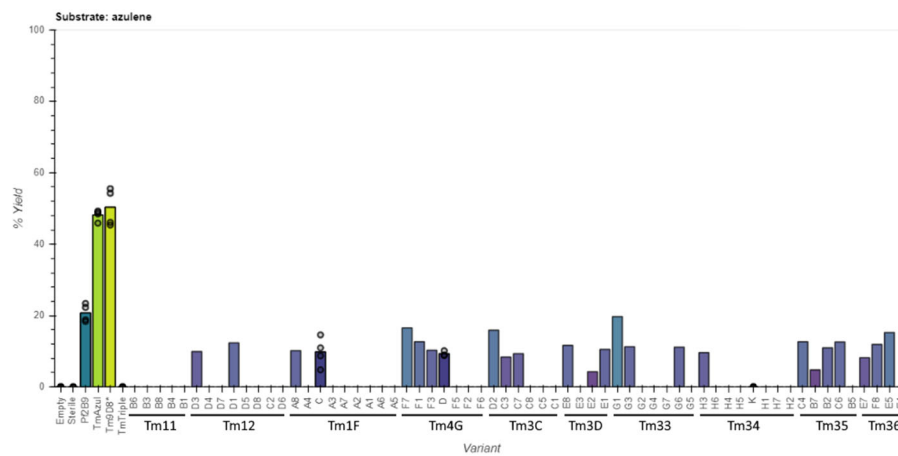


Figure C-9. TrpB panel activity with indole analogs by HPLC yield. (a)-(i) HPLC yield of (a) 5-cyanoTrp, (b) 6-cyanoTrp, (c) 7-cyanoTrp, (d) 5-bromoTrp, (e) 6-bromoTrp, (f) 7-bromoTrp, (g) 5-methoxyTrp, (h) 5-trifluoromethylTrp, and (i) β -(1-azulenyl)-L-alanine for indicated variants supplied with L-serine and each indole substrate. Points represent % yield for individual replicates, bars represent mean % yield for reactions with replicates, or % yield for a single replicate otherwise. Replicates and variant order are as in Fig. S8, and populations from which OrthoRep-evolved variants were derived are annotated.

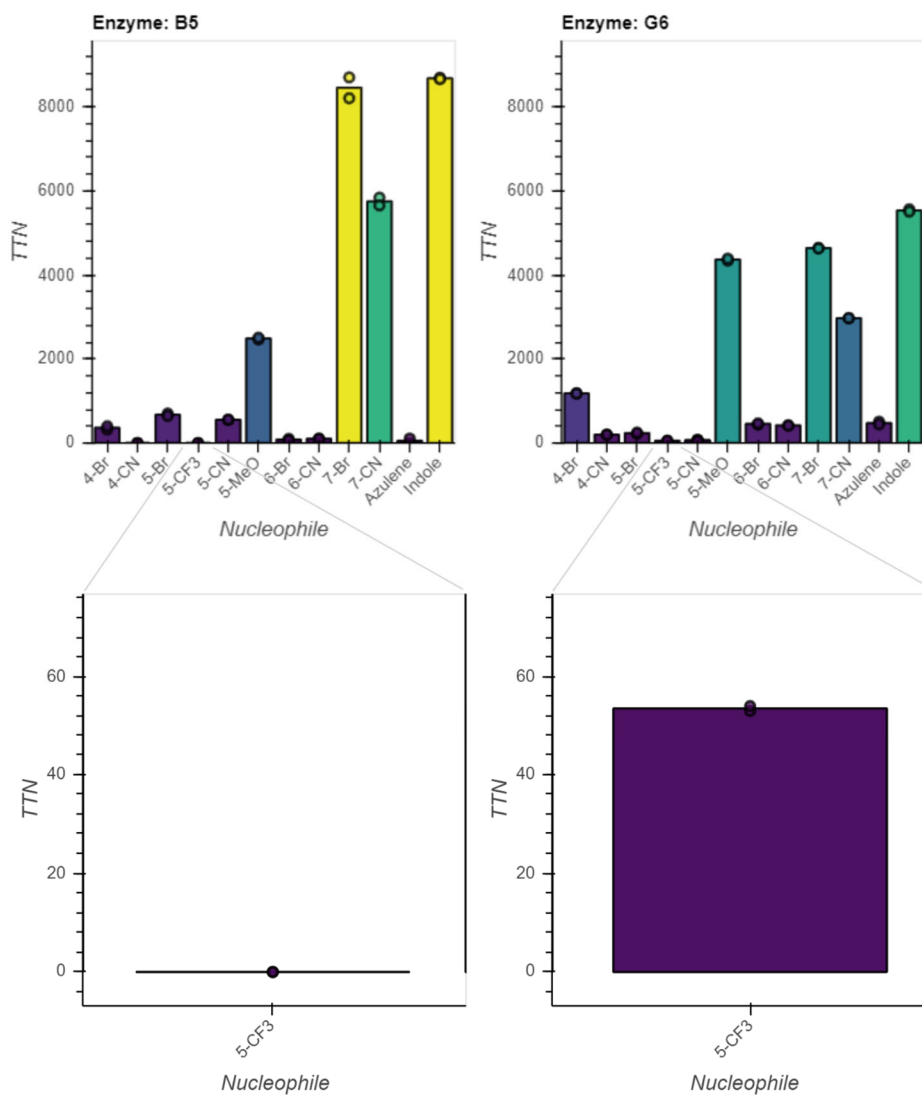


Figure C-10. Substrate activity profiles for large scale purification of variants B5 and G6. Total turnover number (TTN) for TmTrpB variants B5 and G6 purified at large scale (see **Section C.3 Methods**) and supplied with L-serine and the indicated indole analog, azulene, or indole (nucleophile), with a maximum TTN of 10,000. All reactions were performed in duplicate. Points represent TTN for individual replicates, bars represent mean for two replicates. Insets, TTN for 5-trifluoromethylindole with y-axis scale adjusted for clarity

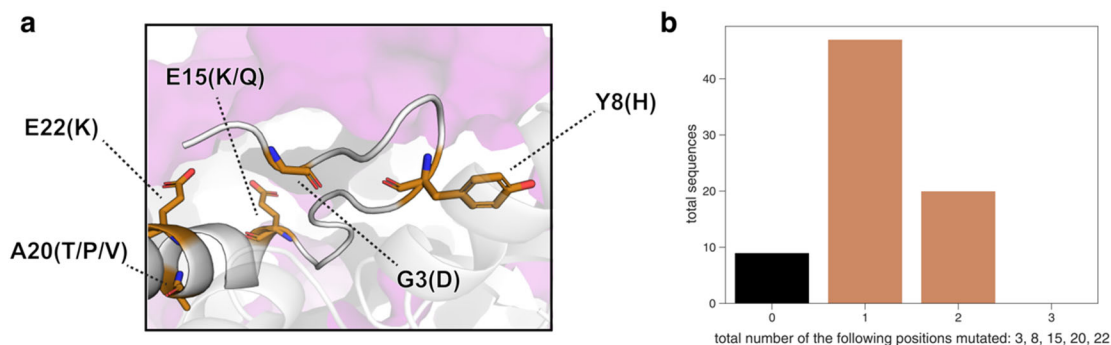


Figure C-11. Commonly observed mutations at the α -subunit interaction interface. (a) Homology-predicted TmTrpB structure (based on engineered stand-alone PfTrpB, PDB 6AM8), with commonly mutated TmTrpB residues located near the TrpA interaction interface highlighted. Solvent-exposed regions of TrpA (purple) (PDB 1WDW) are shown as a surface. Mutations are indicated by the wt residue and position, followed by any residues to which this wt residue is mutated in OrthoRep-evolved TrpB sequences. (b) Total number of sequences in both variant sets 1 and 2 that contain the indicated number of mutations to any of the residues highlighted in panel (a).

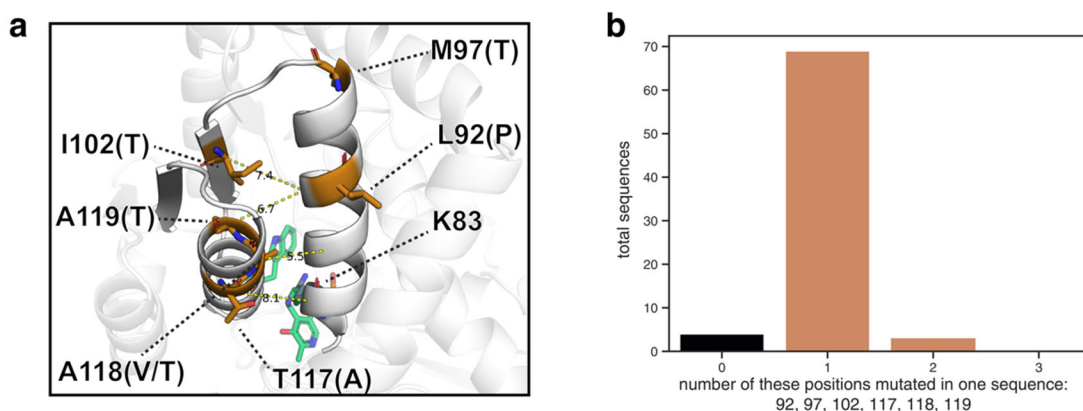


Figure C-12. Commonly observed mutations to residues near a catalytic α -helix. (a) Homology-predicted TmTrpB structure (aligned to engineered stand-alone PfTrpB, PDB 6AM8) with wt residues on or near the α -helix housing K83, which are commonly mutated in OrthoRep-evolved populations (orange). PLP (green) and Trp (green) are shown as sticks, and the catalytic lysine K83 (teal) is shown as spheres. Mutations are indicated by the wt residue and position, followed by any residues to which this wt residue is mutated in OrthoRep-evolved TrpB sequences. Dotted lines connect the α -carbon of residues not located on the K83 α -helix with the α -carbon of the nearest residue on the K83 α -helix, with the distance noted in Ångstroms. (b) Total number of sequences in both variant sets 1 and 2 that contain the indicated number of mutations to any of the residues highlighted in panel a.

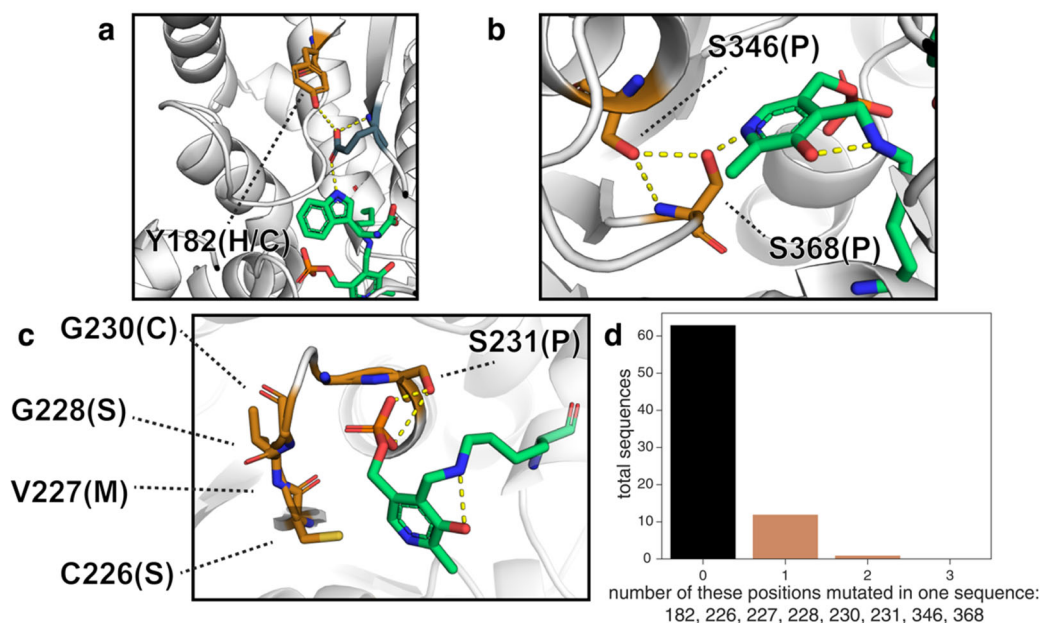


Figure C-13. First- and second-shell active site mutations. (a)-(c), Homology model of *TmTrpB* (aligned to *PfTrpB*, PDB: 6AM8) highlighting residues mutated in OrthoRep-evolved variants (orange) that may influence (a) indole charge, (b) PLP six-member ring binding, and (c) PLP-phosphate binding. Mutations are indicated by the wt residue and position, followed by any residues to which this wt residue is mutated in OrthoRep-evolved *TrpB* sequences. d) Total number of sequences in both variant sets 1 and 2 that contain the indicated number of mutations to any of the residues highlighted in panels (a), (b), or (c).

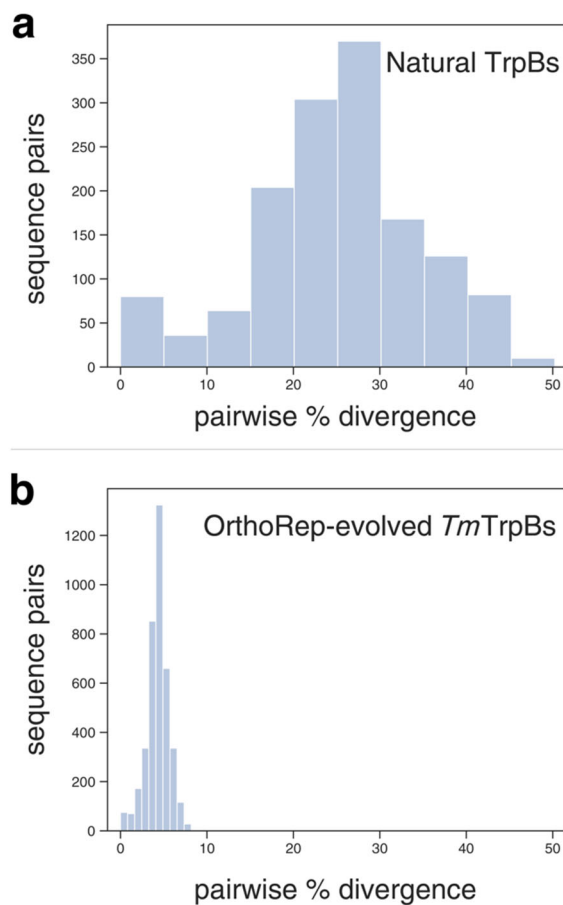


Figure C-14. Sequence divergence for natural and OrthoRep-evolved TrpBs. (a)-(b), Distributions of pairwise % amino acid sequence divergence for a diverse group of 38 naturally occurring mesophilic TrpB variants (a) and OrthoRep-evolved variant sets 1 and 2 (b).

C.3 Methods

C.3.1 DNA plasmid construction

All plasmids that were not generated in a previous study were constructed via Gibson assembly¹ from parts derived from the Yeast Toolkit,² from previously described OrthoRep integration cassette plasmids,³ from *E. coli* expression vectors for previously described TrpB variants,^{4,5} from synthesized oligonucleotides, from yeast genomic DNA, or from the standard *E. coli* expression vector, pET-22b(+). All DNA cloning steps and *E. coli* protein expression steps were performed in *E. coli* strains TOP10 and BL21(DE3), respectively. All

oligonucleotides used for PCR were purchased from IDT, and all enzymes and reagents used for cloning were purchased from NEC.

Parts used to generate yeast nuclear expression plasmids for testing the selection and p1 integration plasmids were PCR amplified from DNA sources listed above, Gibson assembled, used to transform *E. coli*, and plated onto selective LB agar plates. Individual clones were picked, grown to saturation in selective LB liquid media, minipreped, and sequence confirmed. Following evolution of TrpB, individual variants were assembled into new yeast or *E. coli* expression vectors through PCR amplification of purified DNA from evolved yeast cultures, bulk cloning into the appropriate expression vector, picking individual colonies, and confirming absence of any frameshift mutations by Sanger sequencing.

C.3.2 Yeast strains and media

All yeast strains used in this study are listed in Supplementary Table 6. Yeast were incubated at 30 °C, with shaking at 200 rpm for liquid cultures, and were typically grown in synthetic complete (SC) growth medium (20 g/L dextrose, 6.7 g/L yeast nitrogen base w/o amino acids (US Biological), 2 g/L SC dropout (US Biological) minus nutrients required for appropriate auxotrophy selection(s)), or were grown in YPD growth medium (10 g/L bacto yeast extract, 20 g/L bacto peptone, 20 g/L dextrose) with or without antibiotics, if no auxotrophic markers were being selected for. Media agar plates were made by combining 2X concentrate of molten agar and 2X concentrate of desired media formulation. Prior to all experiments, cells were grown to saturation in media selecting for maintenance of any plasmids present.

C.3.3 Yeast transformation

All yeast transformations were performed as described in Gietz and Shiestl.⁶ After all transformations, transformed cells were streaked onto selective media agar plates, and resulting single colonies were picked for all further uses. Transformations for integration onto p1 were performed as described previously:⁷ 2–4 µg of plasmid DNA with *ScaI* restriction sites adjacent to integration flanks was cut with *ScaI*-HF (NEB) and used to transform yeast harboring the wt p1 and p2 plasmids. Proper integration was validated by

miniprepping resulting clonal strains as previously described,⁷ visualizing the recombinant p1 band of the desired size by gel electrophoresis, and PCR and Sanger sequencing of the gene of interest integrated onto p1. Resulting strains were then transformed with either of two plasmids for nuclear expression of an OrthoRep terminal protein DNA polymerase 1 (TP-DNAP1) variant: wt TP-DNAP1 (pAR-Ec318) for evaluating trp5 complementation of TrpB variants without mutagenesis, or error-prone TP-DNAP1 (pAR-Ec633) for generating strains ready for TrpB evolution. These strains were passaged for ~40 generations in order to stabilize copy number of the recombinant p1 species, prior to any use in experiments.

Genomic deletions were made through co-transformation of a CRISPR/Cas9 plasmid targeting the region of interest and a linear DNA fragment comprised of two concatenated 50 bp homology flanks to the region of interest.⁸ Transformations were then plated on selective media agar, colonies were re-streaked onto nonselective media agar, and resulting colonies were grown to saturation in liquid media. The region of interest was PCR amplified and Sanger sequenced to confirm presence of desired modification.

C.3.4 Plating assays

Yeast strains expressing a TrpB variant either from a nuclear plasmid, or from p1, with wt OrthoRep polymerase (TP-DNAP1) expressed from a nuclear plasmid, were grown to saturation in SC –L or SC –LH, spun down, washed once with 0.9% NaCl, then spun down again, and the resulting pellet was resuspended in 0.9% NaCl. Washed cells were then diluted 1:100 (or 1:10,000 where indicated) in 0.9% NaCl, and 10 μ L of each diluted cell suspension were plated onto media agar plates in pre-marked positions. After three days of growth, cell spots were imaged (Bio-Rad ChemiDoc™). Resulting images were adjusted uniformly ('High' set to 40,000) to improve visibility of growth (Bio-Rad Image Lab™ Software). Figures utilizing these images (**Supplementary Figures C-1 and C-2**) were made by manually combining images of different plates, but all images of the same media condition within each figure panel were derived from the same image of a single plate.

C.3.5 *TmTrpB* evolution

Yeast strains with a nuclear plasmid expressing error-prone TP-DNAP1 and with wt *TmTrpB*, *TmTriple*, or *TmTripleQ90** encoded on p1 (GR-Y053, GR-Y055, and GR-Y057, Supplementary Table 6) were grown to saturation in SC –LH, prior to passaging for evolution. All cultures passaged for evolution of *TmTrpB* regardless of success are described in Supplementary Table 1. To provide enough indole substrate for sufficient Trp production, but not enough to induce toxicity, all growth media used for evolution of TrpB activity were supplemented with 100 μ M indole, as informed by results shown in Supplementary Fig. 1. All passages for evolution were carried out as 1:100 dilutions. In order to induce a growth defect but still allow for some growth, the first passage for each evolution culture was carried out in SC –LH media with 37 μ M Trp (7.6 mg/L). After two or three days of shaking incubation, if $OD_{600} > 1.0$ (Bio-Rad SmartSpec™ 3000) for 100-mL cultures, or if most wells in a 24-well block of 3-mL cultures were saturated to a similar degree by eye, cultures were passaged into fresh growth medium with a slightly reduced Trp concentration. If the level of growth was beneath this threshold, the culture was passaged into growth medium with the same Trp concentration. This process was continued until cultures were capable of growth in a Trp concentration of 3.7 μ M (or, in the sole case of WT-100-1, 4.7 μ M), at which point a passage into media lacking Trp was attempted, which typically resulted in successful growth. Resulting cultures were then passaged six additional times into growth medium lacking Trp.

C.3.6 Growth rate assays

Yeast strains containing nuclear plasmids encoding one of several OrthoRep-evolved TrpB variants, wt *TmTrpB*, *TmTriple*, or none of these (denoted ‘empty’) were grown to saturation in SC –L, washed as described above, then inoculated 1:100 into multiple media conditions in 96-well clear bottom plates, with four biological replicates per media/strain combination. Plates were then sealed with a porous membrane and allowed to incubate with shaking at 30 °C for 24 hours, with OD_{600} measurements taken automatically every 30 minutes (Tecan Infinite M200 Pro), according to a previously described protocol.⁹ Multiple 24 hour periods were required for each experiment, but empty controls were included in each individual 96-

well plate to ensure validity of growth in other cultures. Raw OD₆₀₀ measurements were fed into a custom MATLAB script,¹⁰ which carries out a logarithmic transformation to linearize the exponential growth phase, identifies this growth phase, and uses this to calculate the doubling time (T). Doubling time was then converted to growth rate by the equation $\ln(2)/T$.

C.3.7 Enzyme characterization: General experimental methods

Chemicals and reagents were purchased from commercial sources and used without further purification. All cultures were grown in Terrific Broth supplemented with 100 µg/mL carbenicillin (TB_{carb}). Cultures were shaken in a New Brunswick Innova 4000 shaker (shaking diameter 19 mm), with the exception of the 96-well deep-well plates (USA Scientific), which were shaken in a Multitron INFORS HT shaker (shaking diameter 50 mm). Lysis buffer was composed of 50 mM potassium phosphate, pH 8.0 (KPi buffer), supplemented with 100 or 200 µM pyridoxal 5'-phosphate (PLP). Heat lysis was performed in a 75 °C water bath (Fisher) for >1 hour. Protein concentrations were determined using a Pierce™ BCA Protein Assay Kit (Thermo Scientific). Reactions were performed in KPi buffer. Liquid chromatography/mass spectrometry (LCMS) was performed on an Agilent 1290 UPLC-LCMS equipped with a C-18 silica column (1.8 µm, 2.1 × 50 mm) using CH₃CN/H₂O (0.1% acetic acid by volume): 5% to 95% CH₃CN over 2 min; 1 mL/min. Liquid chromatography/mass spectrometry (LCMS) was also performed on an Agilent 1260 HPLC-MS equipped with Agilent InfinityLab Poroshell 120 EC-C18 column (2.7 µm, 4.6×50 mm): hold 5% CH₃CN for 0.5 min, 5-95% CH₃CN over 2 min; 1 mL/min.

TrpB variants selected for further characterization were cloned into a pET-22b(+) vector with a C-terminal 6X His-tag and used to transform *E. coli* BL21(DE3) cells (Lucigen).

C.3.8 Expression and characterization of set 1 variants

C.3.8.1 Large scale expression and lysis

A single colony containing the appropriate TrpB gene was used to inoculate 5 mL TB_{carb} and incubated overnight at 37 °C and 230 rpm. For expression, 0.5 mL of overnight culture were used to inoculate 50 mL TB_{carb} in a 250-mL flask and incubated at 37 °C and 250 rpm for 3 hours to reach an OD₆₀₀ of 0.6–0.8. Cultures were chilled on ice for 20 min and expression was induced with a final concentration of 1 mM isopropyl β-D-thiogalactopyranoside (IPTG). Expression proceeded at 25 °C and 250 rpm for approximately 20 hours. Cells were harvested by centrifugation at 5,000g for 5 min at 4 °C, and then the supernatant was decanted. The pellet was stored at –20 °C until further use or used immediately for whole cell transformations.

Pellets were lysed in 5 mL of lysis KPi buffer with 200 μM PLP, supplemented with 1 mg/mL lysozyme (HEWL, Sigma Aldrich), 0.02 mg/mL bovine pancreas DNase I, and 0.1X BugBuster (Novagen) and incubated at 37 °C for 30 minutes. Lysate was clarified by centrifugation at 5,000g for 10 min, divided into 1-mL aliquots, and stored at –20 °C until further use.

C.3.8.2 Expression and characterization of set 1 variants—lysate and whole cell small-scale reactions

Protein concentration in lysate was quantified by BCA. Lysate reactions were performed in 2-mL glass HPLC vials (Agilent) charged with indole (final conc. 20 mM) dissolved in DMSO (5% w/v), followed by the addition of lysate (final enzyme conc. 4 μM), and serine (final conc. 20 mM) to achieve a final volume of 200 μL. Whole-cell reactions were performed in 2-mL glass HPLC vials (Agilent) charged with indole (final conc. 20 mM) dissolved in DMSO (5% w/v), followed by the addition of cells diluted in KPi buffer (final OD₆₀₀=6), and serine (final conc. 20 mM) to achieve a final volume of 200 μL. Reactions were incubated at 30 °C for 24 hours, diluted with 800 μL 1:1 CH₃CN/1 M aq. HCl, and analyzed via UHPLC-MS.

C.3.8.3 Expression and characterization of set 1 variants—thermostability determination

Enzyme T_{50} measurements (the temperature at which 50% of the enzyme is irreversibly inactivated after a 1-hour incubation) were used to report on the thermostability of the enzyme. In a total volume of 100 μL , samples were prepared in KPi buffer with 1 μM enzyme in PCR tubes and either set aside (25 $^{\circ}\text{C}$) or heated in a thermal cycler on a gradient from 79–99 $^{\circ}\text{C}$ (OrthoRep-generated variants), or 59–99 $^{\circ}\text{C}$ (TmTriple), for 1 hour, with each temperature performed in duplicate. Precipitated protein was pelleted via centrifugation, and 75 μL of each sample were carefully removed and added to the wells of a 96-well UV-transparent assay plate containing 0.5 mM indole and 0.5 mM serine. Relative product formation was observed by measuring the change in absorbance at 290 nm to determine the temperature at which the sample had 50% residual activity compared to the 25 $^{\circ}\text{C}$ samples (modeled as a logistic function).

C.3.8.4 Enzyme kinetics

Enzymatic parameters, k_{cat} and K_{M} , for the conversion of indole to Trp were estimated via Bayesian inference assuming Michaelis-Menten behavior under saturating serine (40 mM) in KPi buffer. Briefly, initial velocities (v) were determined by monitoring Trp formation in a Shimadzu UV-1800 spectrophotometer at 30 $^{\circ}\text{C}$ for 1 min over a range of indole concentrations at 290 nm using the reported indole-Trp difference in absorbance coefficient ($\Delta\epsilon_{290} = 1.89 \text{ mM}^{-1} \text{ cm}^{-1}$).¹¹ These velocities were modeled using the equation:

$$\frac{V_{\text{max}}[\text{indole}]}{K_{\text{M}} + [\text{indole}]}$$

and estimates for v and V_{max} were converted to k and k_{cat} by normalizing for enzyme concentration. Parameter estimates are obtained as Hamiltonian Markov chain Monte Carlo (MCMC) posterior samples and reported as the median with their 95% credible regions (CR). The MCMC software used for sampling was Stan (pystan version 2.19.0.0). The sampling was performed with four separate chains, each starting with 2000 warm-up (disregarded) steps followed by 12000 posterior sampling steps. The priors chosen for k_{cat} and K_{M} were

lognormal distributions with means $\log(150)$ and $\log(500)$, standard deviations 2.5 and 1.5, with units of sec^{-1} and μM , respectively. This provided non-negative probability density that covered low-to-moderate values of the parameters for many known enzymes, but still had significant density out to very high values for each parameter. (In all cases, the data was shown to significantly inform the prior.) The code used to generate these estimates (along with example data) can be found at http://github.com/palmhjell/bayesian_kinetics.

C.3.9 Expression and characterization of set 2 variants

C.3.9.1 Small scale expression and lysis

Variants were arrayed into a 96-well deep-well plate along with *TmTriple*, *Tm9D8**, *TmAzul*, and *Pf2B9*. Individual colonies were grown in 600 μL TB_{carb} in 96-well polypropylene plates overnight at 37 °C, 250 rpm, 80% humidity. The following day, 20 μL of overnight culture were used to inoculate 630 μL TB_{carb} in deep-well 96-well plates and grown at 37 °C, 250 rpm. After 4 hours, cultures were chilled on ice for 20–30 min and protein expression was induced with 50 μL IPTG (final conc. 1 mM) diluted in TB_{carb} . Cultures were shaken at 20 °C, 250 rpm for 20–24 hours, after which they were subjected to centrifugation at 5,000g for 10 min. The cell pellets were frozen at –20 °C until further use or used immediately.

C.3.9.2 Indole rate measurements

Pellets were lysed in either 600 μL of KPi buffer with 100 μM PLP and heat treated at 75 °C for 1 hour, or in 600 μL of this buffer supplemented with 1 mg/mL lysozyme, 0.02 mg/mL bovine pancreas DNase I, and 0.1X BugBuster and incubated at 37 °C for 1 hour. Lysate from both conditions was clarified by centrifugation at 4,500g for 10 min and stored at 4 °C until further use.

Reaction master mix composed of 625 μM indole and 25 mM serine in KPi buffer was prepared and, before reactions, plates and master mix were incubated in a 30 °C water bath for 30 min. The microplate reader (Tecan Spark) was also pre-heated to 30 °C.

To UV-transparent 96-well assay plates (Caplugs, catalog # 290-8120-0AF), 160 μL pre-heated reaction master mix was added by 12-channel pipet followed by 40 μL of lysate from the pre-heated plate using a Microlab NIMBUS96 liquid handler (Hamilton). Plates were immediately transferred into the plate reader, shaken for 10 sec to mix and the absorbance of each well at 290 nm was recorded as rapidly as possible (~ 20 s between measurements) for 120 cycles. The rate of product formation was determined by finding the rate of absorbance change over time and converting to units of concentration using $\Delta\epsilon_{290} = 1.89 \text{ mM}^{-1} \text{ cm}^{-1}$ (see above) and a determined path length of 0.56 cm. We observed no systematic difference in activity between the two lysate preparations (Supplementary Fig. 15), suggesting that most enzyme variants retained sufficient thermostability for purification via heat treatment, and this method was used in subsequent experiments.

C.3.9.3 Substrate scope screen

Pellets were lysed in 300 μL KPi buffer with 200 μM PLP and clarified by centrifugation at 4,000g for 10 min. To a 96-well deep-well plate charged with 10 μL nucleophile dissolved in DMSO (final conc. denoted in the table directly below), 40 μL of the heat treated lysate were transferred using a Microlab NIMBUS96 liquid handler (Hamilton), followed by addition of 150 μL serine (final conc. 20 mM) with a 12-channel pipet. Reactions were sealed with 96-well ArctiSeal™ Silicone/PTFE Coating (Arctic White) and incubated in a 30 °C water bath for ~ 24 hours. Reactions were diluted with 600 μL 2:1 $\text{CH}_3\text{CN}/1 \text{ M}$ aq. HCl, subjected to centrifugation at 5,000g, and 400 μL were transferred to 2-mL glass HPLC vials (Agilent). Samples were analyzed by HPLC-MS. Azulene samples were further diluted 20X to avoid oversaturation of the UV-detector and analyzed via UHPLC-MS.

Table C-5. Nucleophiles tested in substrate scope screen

Nucleophile	Source	Catalog #	CAS	Final conc. (mM)
5-Cyanoindole	Chem Impex	21849	15861-24-2	20
6-Cyanoindole	Chem Impex	21181	15861-36-6	20
7-Cyanoindole	Sigma	CDS008484		20
5-Bromoindole	Sigma	B68607	10075-50-0	20
6-Bromoindole	Sigma	524344	52415-29-9	20
7-Bromoindole	Sigma	473723	51417-51-7	20
5-Methoxyindole	Combi-Blocks	IN0049	1006-94-6	20
5-Trifluoromethylindole	Sigma	701068	100846-24-0	10
Azulene	Alfa Aesar	L08271	275-51-4	20

All samples except those containing azulene were analyzed at 277 nm, representing the isosbestic point between indole and Trp and allowing estimation of yield by comparing the substrate and product peak areas for indole analogs.¹² Azulene yield was estimated as described previously.¹³ Nucleophile retention times were determined through injection of authentic standards and product retention times were identified by extracting their expected mass from the mass spectrum.

C.3.10 Characterization of Tri-100-3-F and Tri-100-1-G

C.3.10.1 Large-scale expression and purification

A single colony containing the appropriate TrpB gene was used to inoculate 5 mL TB_{carb} and incubated overnight at 37 °C and 230 rpm. For expression, 2.5 mL of overnight culture were used to inoculate 250 mL TB_{carb} in a 1-L flask and incubated at 37 °C and 250 rpm for 3 hours to reach OD₆₀₀ 0.6–0.8. Cultures were chilled on ice for 20 min and expression was induced with a final concentration of 1 mM IPTG. Expression proceeded at 25 °C and 250 rpm for approximately 20 hours. Cells were harvested by centrifugation at 5,000g for 5 min at 4 °C, and then the supernatant was decanted. The pellet was stored at –20 °C until further use.

Pellets were lysed in 25 mL KPi buffer with 200 μM PLP for >1 hour at 75 °C. Lysate was clarified by spinning 14,000g for 20 min at 4 °C (New Brunswick Avanti J-30I). Protein was

purified over hand-packed HisPur™ Ni-NTA Resin (Thermo Scientific, catalog # 88221), dialyzed into KPi buffer, and quantified by BCA.

C.3.10.2 Tri-100-3-F PLP-binding assay

Variant Tri-100-3-F did not exhibit the characteristic yellow color of PLP-bound TrpB variants after purification, however BCA indicated protein concentrations comparable to the Tri-100-1-G variant. We have previously observed that some TrpB variants lose binding affinity for PLP resulting in non-functional apoenzyme. We evaluated Trp formation of Tri-100-3-F supplemented with 0, 0.1, 0.25, 0.5, 1, 2, 5, and 100 μ M PLP via UV-Vis spectrophotometry. Serine (final conc. 25 mM) + PLP master mixes of the eight concentrations were prepared and dispensed into a 96-well UV-transparent plate. Enzyme dilutions (final conc. 1 μ M) with or without indole master mixes were prepared, and 100 μ L thereof were dispensed into the 96-well plate. The plate was immediately transferred into plate reader, shaken for 10 s to mix, and product formation was measured ~20 s for 120 cycles at 290 nm.

Only the 100 μ M PLP condition restored activity, supporting our hypothesis that the purified enzyme was apoprotein and binds PLP poorly, requiring supplementation of PLP to re-form a functional holoenzyme. Thus, we chose to supplement PLP in the subsequent purified protein reactions.

C.3.10.3 Small scale analytical reactions

Reactions were performed in 2-mL glass HPLC vials (Agilent) charged with nucleophile (final conc. 20 mM) dissolved in DMSO (5% w/v), followed by the addition of purified protein (final enzyme concentrations were either 2 μ M or 40 μ M), PLP (final conc. 100 μ M) and serine (final conc. 20 mM) to achieve a final volume of 200 μ L. Reactions were incubated at 30 °C for ~24 hours. Reactions were diluted with 600 μ L 2:1 CH₃CN/1 M aq. HCl, subjected to centrifugation at 5,000g, and 400 μ L transferred to 2-mL glass HPLC vials (Agilent). Samples were analyzed by HPLC-MS. Azulene samples were further diluted 20X to avoid oversaturation of the UV-detector and analyzed via UHPLC-MS.

Appendix C Bibliography

1. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
2. Lee, M. E., DeLoache, W. C., Cervantes, C., & Dueber, J. E. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth. Biol.* **4**, 975–986 (2015).
3. Ravikumar, A., Arzumanyan, G. A., Obadi, M. K. A., Javanpour, A. A., & Liu, C. C. Scalable, Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds. *Cell* **175**, 1946-1957.e13 (2018).
4. Boville, C. E., Romney, D. K., Almhjell, P. J., Sieben, M., & Arnold, F. H. Improved Synthesis of 4-Cyanotryptophan and Other Tryptophan Analogues in Aqueous Solvent Using Variants of TrpB from *Thermotoga maritima*. *J. Org. Chem.* **83**, 7447–7452 (2018).
5. Murciano-Calles, J., Romney, D. K., Brinkmann-Chen, S., Buller, A. R., & Arnold, F. H. A Panel of TrpB Biocatalysts Derived from Tryptophan Synthase through the Transfer of Mutations that Mimic Allosteric Activation. *Angew. Chemie - Int. Ed.* **55**, 11577–11581 (2016).
6. Gietz, R. D. & Schiestl, R. H. High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc.* **2**, 31–34 (2007).
7. Ravikumar, A., Arrieta, A., & Liu, C. C. An orthogonal DNA replication system in yeast. *Nat. Chem. Biol.* **10**, 175–177 (2014).
8. Ryan, O. W. & Cate, J. H. D. Multiplex engineering of Industrial Yeast Genomes Using CRISPRm. *Methods in Enzymology* **546**, (Elsevier Inc., 2014).
9. Jung, P. P., Christian, N., Kay, D. P., Skupin, A., & Linster, C. L. Protocols and programs for high-throughput growth and aging phenotyping in yeast. *PLoS One* **10**, (2015).
10. Zhong, Z. et al. Automated Continuous Evolution of Proteins in Vivo . *ACS Synth. Biol.* (2020). doi:10.1021/acssynbio.0c00135
11. Lane, A. N. & Kirschner, K. The Catalytic Mechanism of Tryptophan Synthase from *Escherichia coli*. *Eur. J. Biochem.* **129**, 571–582 (1983).
12. Romney, D. K., Murciano-Calles, J., Wehrmüller, J. E., & Arnold, F. H. Unlocking

Reactivity of TrpB: A General Biocatalytic Platform for Synthesis of Tryptophan Analogues. *J. Am. Chem. Soc.* **139**, 10769–10776 (2017).

13. Watkins, E. J., Almhjell, P. J., & Arnold, F. H. Direct Enzymatic Synthesis of a Deep-Blue Fluorescent Noncanonical Amino Acid from Azulene and Serine. *ChemBioChem* **21**, 80–83 (2020).

DEEP MUTATIONAL SCAN OF THE TRYPTOPHAN SYNTHASE
BETA SUBUNIT

E.J.W.D. and B.J.W. conceived of the project. E.J.W.D. designed and performed all experimental work. E.J.W.D. adapted code from B.J.W. for analysis of next generation sequencing data.

ABSTRACT

Machine learning has the potential to revolutionize how proteins are engineered, but requires large quantities of data to train models. In this chapter, I describe the development of a deep mutational scanning experiment of combinatorial site-saturation mutagenesis (SSM) libraries using the tryptophan synthase β -subunit (TrpB). The aim of the project is to generate a large dataset that maps enzyme sequence to function for the purpose of studying epistasis using machine learning. This project will generate triple- and quadruple-SSM libraries (8,000 and 160,000 unique variants, respectively) and sequence the populations using Next Generation Sequencing before and after a survival-based selection.

5.1 Introduction

The aim of this project is to collect a dataset on the combinatorial interactions of residues to study the nature of epistasis using machine-learning algorithms. We know that evolution generally works through the stepwise accumulation of single, beneficial mutations. Many mutations, however, are only beneficial in the context of another. This phenomenon, known as epistasis, has traditionally been studied by phylogenetic analysis, as so far, understanding general principles about epistasis has largely been constrained by a lack of data.¹ Intra-protein interactions can be explored by Combinatorial Site-Saturation Mutagenesis (CSSM), in which one or more positions in a protein are mutated to create a library of variants for screening. However, methods for collecting and parsing through the large datasets of DNA sequences were previously insufficient to measure combinatorial fitness landscapes. Breakthroughs in high-throughput sequencing technologies as well as the blossoming machine-learning field have poised the evolution community to begin to unfold complex protein phenomena.^{2,3}

Deep Mutational Scanning (DMS) has emerged as a method to evaluate large library selections. DMS utilizes Next Generation Sequencing (NGS) to characterize library populations before and after a selection.⁴ The frequencies of library variants represented in the sequencing before and after selection are compared to calculate a ‘functional score,’ or selection coefficient, corresponding to the fitness of the enzymes (**Figure 5-1**, right). The resultant calculated fitness landscape can then be analyzed via machine learning (ML) to uncover patterns that reveal complex interactions between the targeted protein residues which may be impossible to identify by human analysis alone. While many groups have used this technology to study epistasis in a context of binding,⁴ I wanted to generate a fitness landscape surrounding a much more difficult challenge: enzyme catalysis.

For this project, I proposed to perform a deep mutational scan to measure combinatorial fitness landscapes of a tryptophan synthase β -subunit (TrpB) in order to observe and predict evolutionary pathways to novel functions. Once collected, the data will be used to train machine-learning (ML) models to predict beyond the data that they have been given. I

believe using the combination of DMS and NGS with TrpB libraries will fulfill the need of the protein-engineering ML community and deliver large, quality datasets to develop generalized, predictive models so that we can begin to explore underlying principles that govern proteins.

5.2 Original design overview

For this DMS study, I planned to generate three- and four-site CSSM libraries of which TrpB with a selection on the variants' ability to accept indole to complement a Trp autotrophic yeast strain (**Figure 5-1**, left). The libraries would be sequenced before and after selection via NGS, and functional scores, otherwise known as selection coefficients, were to be calculated from the frequency of each variant at each timepoint (**Figure 5-1**, right). The 8,000-variant, three-site combinatorial library was planned to serve as the lower cost, pilot study to ensure our cloning, selection, and sequencing workflow are optimized for the more challenging, 160,000-variant, four-site library. This work is unfinished and will be continued by another Arnold Lab member. I developed the entire project to be performed in a yeast host, but technical issues associated with library cloning and DNA preparation proved to be extremely challenging. While the selection system will be re-developed using *Escherichia coli* to circumvent yeast-specific problems, the insights gained and the PCR methods developed from my work will serve as the foundation for the continuation of the project.

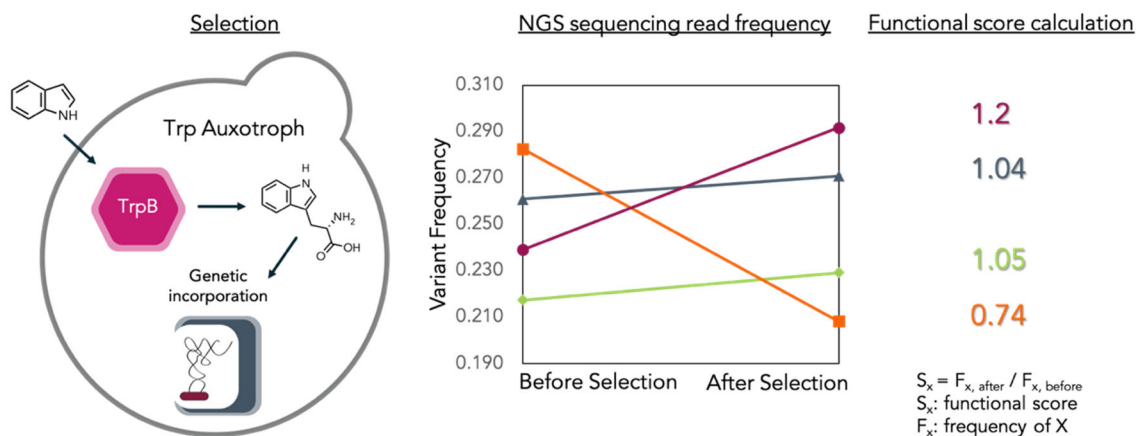


Figure 5-1. Deep mutational scan overview. Left: Deep mutational scanning scheme. Indole is provided to a strain of yeast that is auxotrophic for Trp, with each cell expressing a TrpB variant. If the TrpB variant is active, it will synthesize Trp, thus complementing the auxotrophy and allowing the yeast to survive. The TrpB variants better at complementation should grow faster and therefore be better represented in the population after the selection. Right: How functional scores are calculated. The frequency of each variant is measured before and after the selection. The functional score (S_x) is calculated by dividing the frequency of variant X after the selection, ($F_{x, \text{after}}$) by the frequency of variant X before the selection ($F_{x, \text{before}}$)

There are several key factors that are important to consider for collecting quality deep mutational scanning data.

- 1) To ensure adequate sampling of the library, at least ten times as many transformants as library members should be obtained.⁵
 - a. Although there is a possibility of 8,000 different unique amino-acid combinations for a three-site library, using NNK degenerate codons to construct the library will result in more codon combinations than variants, with some amino acid combinations occurring more than others. We can calculate the least frequently occurring variant for the three-site library as $1/32^3$. Since there are 32,768 possible codon combinations, the library size

will be approximately 3×10^4 variants. We can therefore estimate that if we need ten times the least frequently occurring variant, we will need 3×10^5 **transformants**.

b. For a four-site library, the least frequently occurring variant is $1/32^4$, so the library size is estimated to be 1×10^6 , and the library transformation efficiency must be 1×10^7 .

2) Each variant should be covered by at least 100 sequencing reads in the initial library.⁵

a. We must consider the quantity of DNA that we are able to extract from samples as well as the quantity necessary for preparation of the library for NGS. The DNA extraction efficiency will influence the size of the culture necessary for the selection. These values must be determined empirically.

3) Deep mutational scanning data can be noisy, so at least one biological replicate should be obtained.⁵

4) Extending experiment time and sampling more timepoints results in higher precision for measuring selection coefficients.⁶

5.3 Establishing the yeast selection

I began by obtaining the Trp autotrophic yeast strain (*Saccharomyces cerevisiae* BY4742 Δ Trp5) from the Chang Liu lab at UC Irvine. Yeast offers several advantages over our typical *E. coli* expression host. First, the Liu lab has already demonstrated TrpB's ability to complement the yeast autotrophic strain BY4742 Δ Trp5 and standard protocols have been developed to tune indole addition as selection pressure.⁷ Second, yeast affords the ability to

avoid false positives that could arise from library-cloning errors. By integrating TrpB into the genome of yeast via homologous recombination using constructs from the MoClo Yeast Tool kit,⁸ uniform gene copy number is ensured and any risk associated with multiple transformants in a single cell is removed. The TrpB genes were cloned into the pYTK096 integration plasmid that has homology to the *Ura3* region of the genome. I cloned the *I-SceI* “landing pad” into BY4742 Δ Trp5, which is recognized by the *I-SceI* restriction enzyme. Yeast cells can be co-transformed with a plasmid encoding *I-SceI* and the linearized pYTK096 plasmid, and the restriction enzyme is expressed *in vivo* to introduce a double stranded break at the recognition site, encouraging higher rates of recombination (**Figure 5-2**). The *I-SceI* step is optional and can be avoided if the cells are transformed with a larger quantity of library DNA.

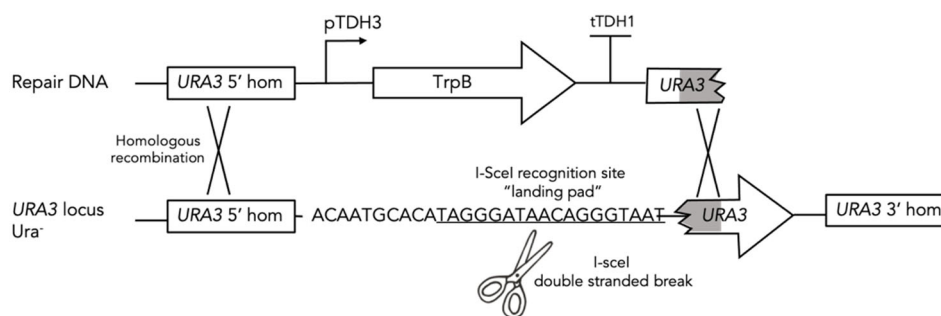


Figure 5-2. Genomic integration by homologous recombination.

To test the selection, I cloned five different *Tm*TrpB variants that were known to have a range of activity on indole into the auxotrophic yeast strain. *Tm*Triple was selected because it was a starting point for the previous OrthoRep evolution project⁷ and possesses the fewest mutations required to create a stand-alone TrpB,⁹ *Tm*9D8* was selected because of its known activity at 37 °C compared to other TrpB variants that function best at temperatures ranging from 55–75 °C,¹⁰ *Tm*9D8* E105G was selected because it is known that the E105G mutation dramatically decreases Trp activity¹¹ and could be compared to *Tm*9D8*; and finally, *Tm*1Ff and *Tm*4Ge were chosen because they were evolved in the BY4742 Δ Trp5 strain for auxotrophic complementation using the OrthoRep continuous evolution system.⁷

In a qualitative test, I grew the five different variants with and without the addition of 400 μM indole and monitored their growth over the course of several days (**Figure 5-3**). The concentration of 400 μM was determined previously by Gordon Rix to be the maximum concentration tolerable without causing indole toxicity and stunting yeast growth.⁷ The variants behaved exactly as expected during the selection, with the Orthorep evolved variants growing fastest and the *Tm9D8** E105G variant growing the slowest.

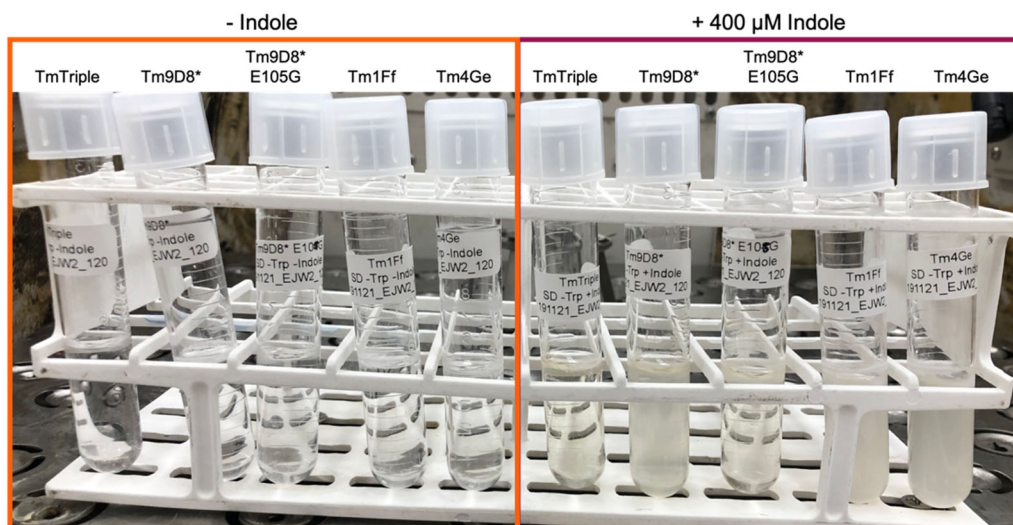


Figure 5-3. Pilot selection with TrpB variants integrated into yeast genome. 1:100 dilution from overnight culture, 48 h growth, Synthetic Defined (SD) Trp dropout minimal medium, 30 °C, 250 rpm (Innova 42). Five TrpB variants cloned into the yeast genome were grown in the absence (Left box) and presence (Right box) of 400 μM indole. The cells should not grow without indole (negative control) and should grow at varying rates that correspond to TrpB activity when in the presence of indole. The turbidity of the culture represents how well the culture grew. Variants better at Trp complementation (*Tm1Ff*, *Tm4Ge*, and *Tm9D8**) are very opaque, while variants that are not (*TmTriple* and *Tm9D8** E105G) remain transparent.

It is important that the cultures be maintained in the exponential regime of the growth curve during the selection so that the selection conditions are consistent throughout the entire experiment.⁶ The size of the culture and the frequency of the timepoints were yet to be determined because I needed to measure the quantity of DNA that could be extracted per optical density at 600 nm (OD_{600}) and volume of cells. Regardless, it was necessary to estimate the growth kinetics of the cells in SD-Trp + 400 μM indole media so that the cells

could be appropriately passaged to maintain growth in the exponential phase. An overnight *Tm9D8** culture in YPD media was diluted 1:20 and the growth curves measured over the course of 25 hours in 5-mL and 50-mL cultures. From the assay, the cultures reached stationary phase at approximately OD₆₀₀ of 7–8, so they should be passaged in an OD₆₀₀ range of 4–6 to maintain logarithmic growth (**Figure 5-4**).

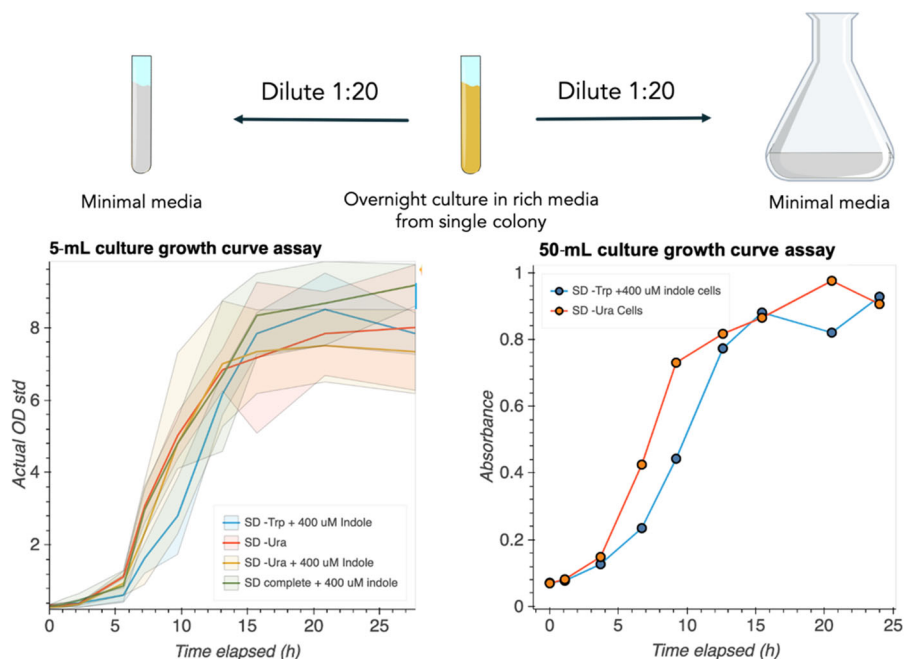


Figure 5-4. Growth curve assay of *BY4742 ΔTrp5* expressing *Tm9D8 in minimal media.** 1:20 dilution from overnight culture, 25 h growth, Synthetic Defined (SD) Trp dropout minimal medium supplemented with 400 μ M indole, 30 °C, 250 rpm (Innova 42). Shaded regions in 5-mL assay represent ± 2 standard deviations from the mean.

During a trial selection, I observed that the growth curve did not follow the typical behavior of a homogeneous population, most likely caused by the heterogeneity of TrpB variant activity. The initial density of cells is comprised of a range of slow to fast growing cells, and as the faster growing cells overtake the slow ones, the apparent doubling time of the entire culture accelerates. Since it was impossible to predict the time that the cells reached exponential phase, continuously monitoring the culture would prevent a situation where the cells might reach stationary phase while unsupervised. I arranged to use the Cell Growth

Quantifier technology from Aquila Biolabs, which reports cell density in real time using a detector that is placed underneath the shake flask during growth.

5.4 Cloning combinatorial libraries

Because the aim of the project is to study epistasis, it is paramount to make libraries with residues that are likely to interact epistatically. Residues that are spatially adjacent to one another are more likely to be epistatic than residues that are not.¹² Therefore, I chose to construct ten different libraries that contained spatially adjacent residues based on the *Pyrococcus furiosus* TrpB crystal structure (PDB: 6AMH),¹³ focusing primarily on active-site residues and ones that had been previously shown to influence activity (**Table 5-1** and **Figure 5-5**).

Table 5-1. Sites selected for triple-site combinatorial libraries (*Thermotoga maritima* scaffold numbering).

Library	Site 1	Site 2	Site 3
A	104	105	106
B	105	106	107
C	106	107	108
D	117	118	119
E	184	185	186
F	162	166	301
G	227	228	301
H	228	230	231
I	182	183	184

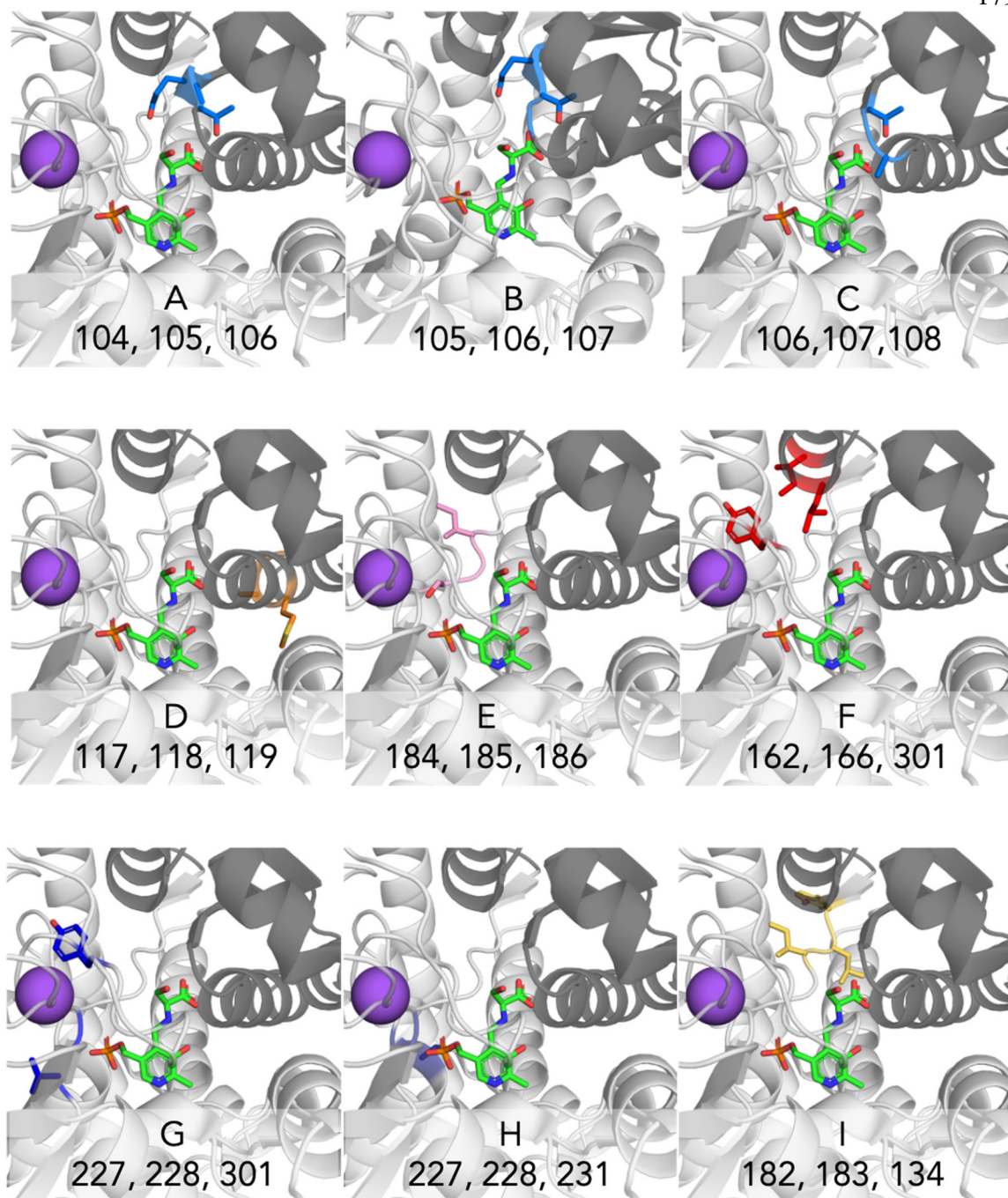


Figure 5-5. Site-saturation mutagenesis libraries highlighted in *Pyrococcus furiosus* scaffold (PDB: 6AMH). Green sticks: PLP with amino-acrylate bound. Purple sphere: Potassium cation ion binding site. Dark grey cartoon: TrpB COMM domain.

I initially designed standard site-saturation mutagenesis Gibson primers to clone the libraries and used Sanger sequencing to verify that there was good representation of each base at the randomized positions (**Figure 5-6A**). Unfortunately, when the libraries were later sequenced by MiSeq NGS, I discovered that they were strongly biased for the parent sequence. This was likely because the primer that contained the parent sequence would be thermodynamically favored during the library generation PCR since it possesses more base pairs matched to the template sequence than any other library primer. To circumnavigate this issue, I devised a two-step PCR scheme with the help of Bruce Wittmann that eliminated the possibility for bias (**Figure 5-6B** and **5-7**). The first PCR amplified the entire plasmid except the region where the mutations are being introduced (called the ‘gap PCR’), and then the reaction was *DpnI* digested to remove all template DNA. Afterwards, this new template was used with the library primers that introduce the genetic diversity at the gap region and the resultant linear PCR product was assembled via Gibson assembly as normally. Using the two-step gap PCR technique, most of the libraries were successfully constructed, and their codon representation verified by MiSeq NGS sequencing.

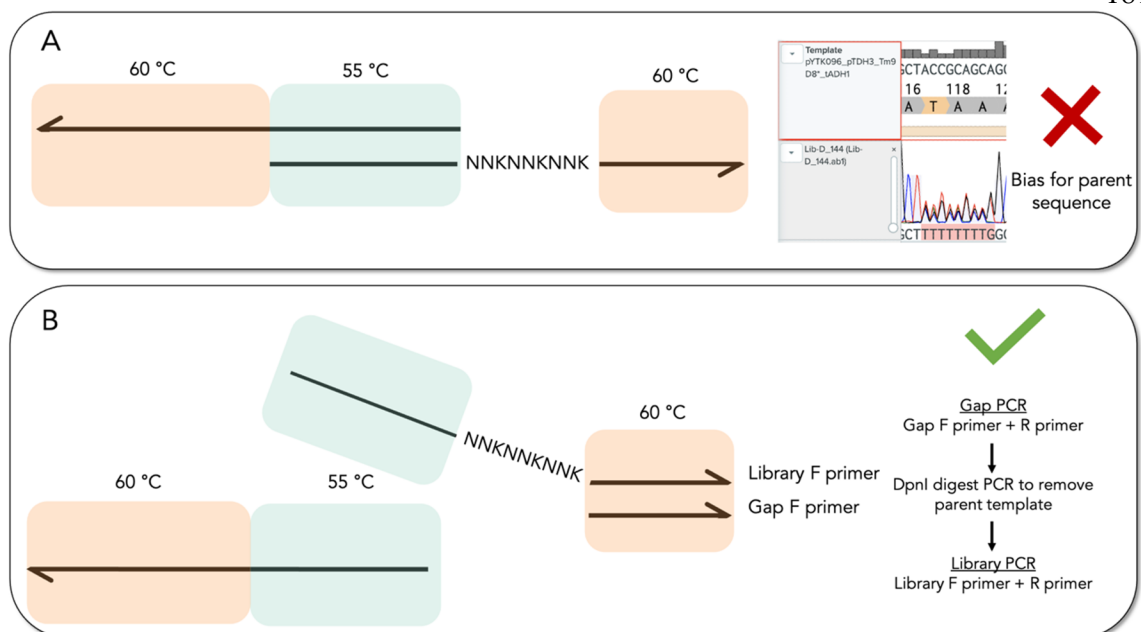


Figure 5-6. Combinatorial library primer design. A) Standard Gibson assembly primer design. The traces to the right represent the Sanger Sequencing results of a library. B) Two-step gap PCR technique. The first PCR amplifies everything except region of mutagenesis. The second PCR introduces genetic variation at region of mutagenesis without parent sequence present to bias binding.

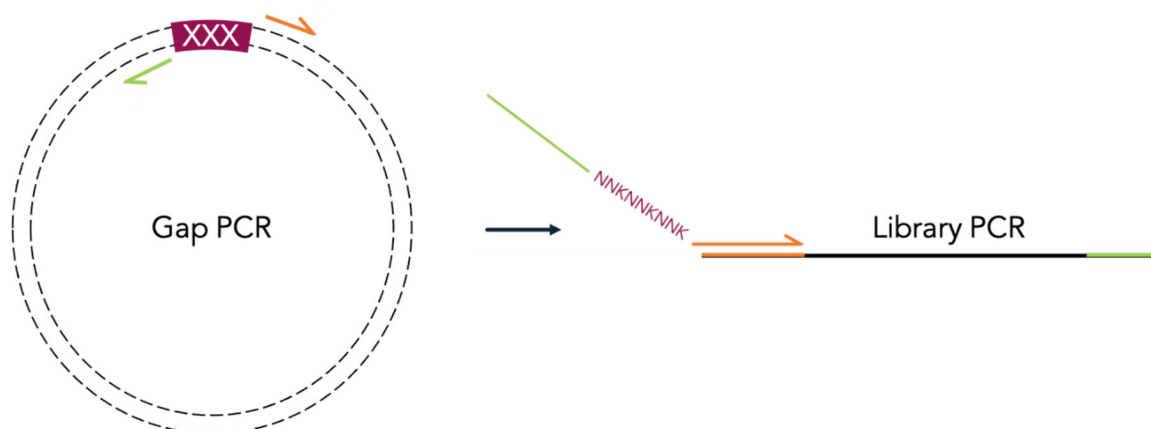


Figure 5-7. Triple SSM library cloning scheme.

Although I had previous experience at transforming yeast with the pYTK096 plasmid for genomic integration at high efficiencies in the Ellington Lab at the University of Texas at Austin, I encountered difficulty transforming them in the Arnold Lab. I tried numerous different transformation protocols,^{14,15} different libraries that I had constructed, and different DNA concentration techniques, and I was only able to successfully transform one library. After exhausting all of my own abilities, I recruited the help of Sabine Brinkmann-Chen, who also had prior experience with yeast transformations. Together, we still could not transform the libraries with multiple attempts, and even the positive control plasmids failed. An industry expert on yeast transformations suggested that quality of the agar plates can have a dramatic effect on whether colonies grow, so it was possible that the issue was with the growth conditions after the transformation. The recommended plate brand was ordered, but shipping was heavily delayed due to COVID.

5.5 Preparing library amplicons for NGS

Next generation sequencing is the key technology that enables the execution of deep mutational scans. Without the ability to sequence reads in parallel, it would be impractical to embark on projects that require sequencing thousands of variants. Since deep mutational scanning studies rely heavily on quality NGS data to accurately report variant frequencies, it is important to minimize the amount of error that is introduced in the sequencing runs. The error rate of NGS can be up to 1%, which can result in fictitious mutations that do not accurately reflect the library.⁵ This issue can be ameliorated by using paired-end reads, which generate double coverage of the sequence of interest, thus reducing the error rate and allowing the researcher to throw out any sequences whose forward and reverse sequences do not match at the positions of interest.

The low sequence diversity in a deep mutational scanning library presents two major challenges. First, during library preparation, incomplete PCR products can act as megaprimers for other templates, which causes a scrambling of variant identities (**Figure 5-8A**). This is not a large concern for libraries where the codons are immediately adjacent to one another, because scrambling is more likely to occur between sites that are further apart from

one another. Some of the libraries I designed did not have three consecutive sites mutated, and I wanted the method I developed to be as generalizable as possible, so it was prudent to treat all libraries as if this recombination were to occur. To minimize this effect, researchers try to keep the number of PCR cycles during library preparation as low as possible, reducing the number of chances that a recombination event could occur. While an effective solution, I imagined that this issue could be completely avoided by implementing an emulsion-based water-in-oil droplet PCR (emPCR) method that spatially separates the templates, thus eliminating any chance of different templates coming into contact with one another (**Figure 5-8B**). Emulsion PCRs have been used for NGS library preparation in similar ways before,¹⁶ and I had personally worked to optimize an emulsion technique for amplification of DNA in whole yeast,¹⁷ so I believed that this approach would be feasible.

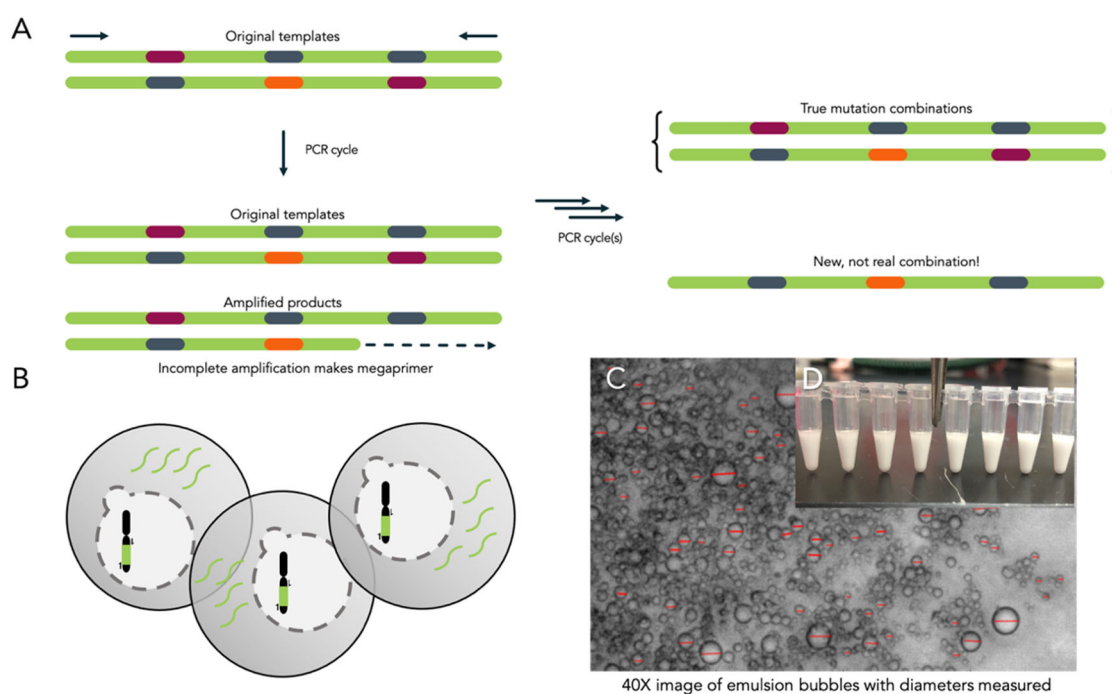


Figure 5-8. Template recombination and yeast emulsion scheme. (a) Recombination of templates results in false mutation combinations. (b) Abstraction of emulsion bubbles. Yeast are spheroplasted to make the membrane permeable to PCR reagents in buffer. Primers are designed to amplify the TrpB gene integrated into the genome. (c) 40X image of emulsion bubbles with diameters measured to estimate the number of bubbles in each reaction. (d) Image of emulsions post-PCR.

The libraries are to be sequenced using Illumina NGS Mi-Seq and Hi-seq technology. Briefly summarized, Illumina NGS technology works by attaching DNA to a flow cell with adaptors that have homology to oligos affixed to the surface of the flow cell, then amplifying each gene to generate small clusters of DNA, and flowing dNTPs that have fluorescent labels with color associated for each base attached to them. For each position that is sequenced, the fluorescent labels are excited, recorded, and then removed so that the next round of dNTPs can be attached to the following base (**Figure 5-9**).

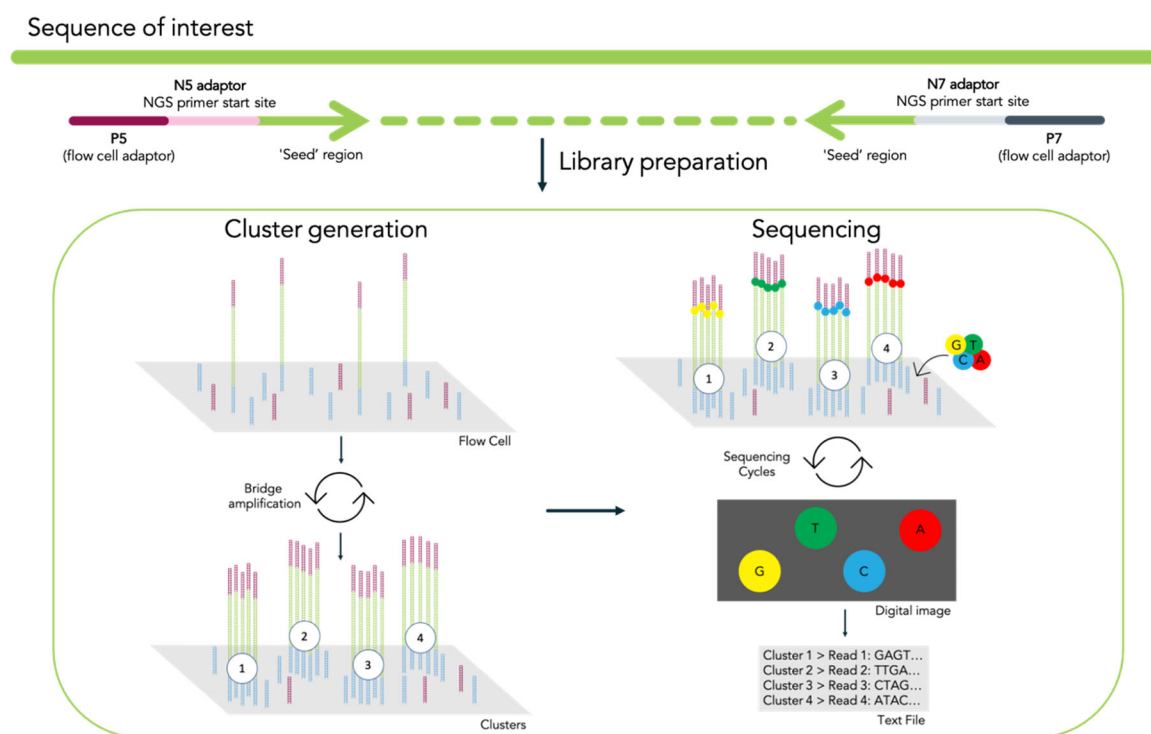


Figure 5-9. General overview of Illumina sequencing. The method of adaptor attachment can vary; here is shown how a targeted gene would be prepared for sequencing.

During sequencing, the second issue of low sequence diversity arises: the sequencing software uses the first 25 bases sequenced to differentiate between the clusters of DNA and will struggle to identify clusters that have the same exact beginning sequence.¹⁸ This is a problem because the cluster boundaries inferred are used to assign bases, otherwise known as base calling, and if mixed populations of DNA are assigned together, it will reduce the

quality of the calls. It is common to spike sequencing runs with sheared genomic DNA from the virus phiX174 to artificially introduce diversity and improve sequencing quality.¹⁹ However, there are other methods that have been developed to introduce more sequence diversity. For example, the ‘phasing amplicon sequencing approach’ (PAS) shifts sequencing frames among different samples by introducing varying numbers of bases (0–7) as ‘spacers.’²⁰ The issue can also be addressed by introducing random 10mer barcodes at the beginning of each sequencing read, which artificially introduces diversity that can be used for calling bases.²¹ The random 10mer was appealing due to its relative simplicity to the PAS approach, so it was integrated into my primer design. I also planned other strategies for introducing sequence diversity; multiple libraries could be pooled into one sample (**Figure 5-10A**), or different timepoints could have alternating directionality so that there would be at least two populations of different sequences in each run (**Figure 5-10B**). There was not an opportunity to test these ideas in the laboratory, so their effectiveness has yet to be demonstrated.

The library preparation was designed to be nested PCR, with both the external and internal primers combined in one emPCR reaction. (**Figure 5-10C**). The internal primer begins with a ‘seed region’ that has homology to the gene of interest, then a 3mer barcode that represents the timepoint, followed by a random 10mer region for diversity, and then homology to the i5/i7 adaptor region so that the external adaptor primers can be attached. The external primer is a standard Illumina Nextera primer.

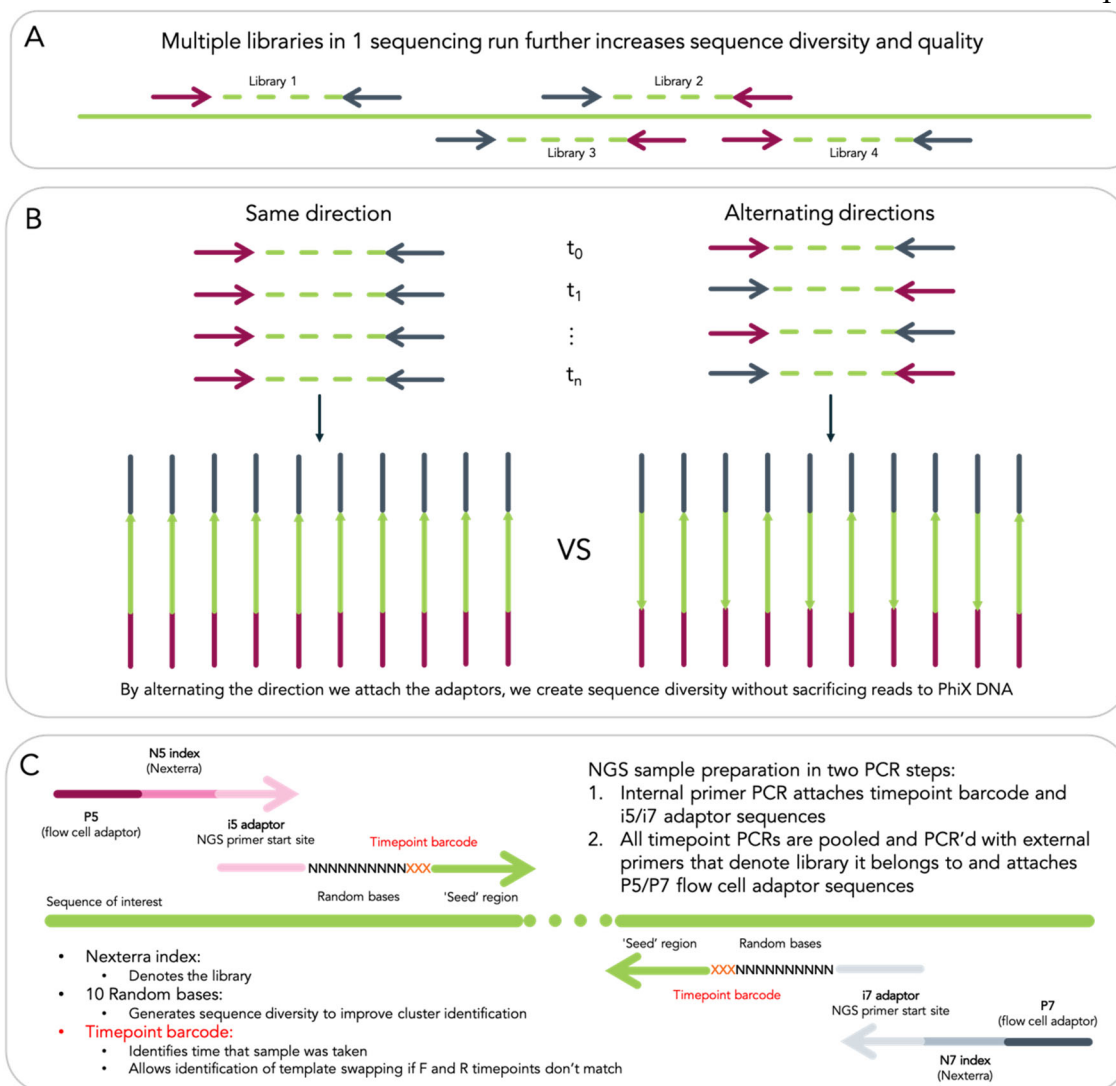


Figure 5-10. Library diversity strategies and amplicon primer design. (a) Including multiple libraries in one Illumina run can increase sequence diversity. (b) Alternating the directionality of the amplicons could also increase sequence diversity. (c) Final primer design.

Both emPCRs and nested PCRs can be finicky methods, and I experienced difficulty getting the nested PCR to work in emulsion conditions. In the interest of time, I decided to forego the emPCR and optimized the internal primer PCR for the minimal number of cycles on extracted DNA, as is the current standard method for minimizing recombination during deep mutational scans.⁵

5.6 Proposed *E. coli* DMS redesign

The benefits supposedly conferred by using yeast for this selection were nullified by the difficulties that were encountered transforming the yeast with the libraries and preparing the NGS amplicons. Instead, I propose that the project be continued in an auxotrophic *E. coli* strain with library variants constitutively expressed on a plasmid. The Arnold lab has experience achieving high transformation efficiencies with *E. coli*, and *E. coli* systems are generally easier to transform and work with than yeast genomic integration. Furthermore, the yeast genomic extraction kit I used resulted in low DNA yields, so it would have required that very large selection cultures be grown in order to have enough cells to passage and extract for NGS amplicon preparation. In contrast, it is very easy to extract large quantities of plasmid DNA from *E. coli*, so smaller cultures can be used.²² Finally, the protocol for *E. coli* emPCR is simpler and more robust than the one for *S. cerevisiae* because *E. coli* are easier to lyse than yeast, so it may be easier to implement the proposed emPCR method.²³ The ‘gap PCR’ library cloning technique and NGS amplicon protocols that I have developed are directly transferrable to the *E. coli* system, so only the selection will need to be re-established to complete the project.

5.7 Analysis of NGS data

When working with large datasets, it is important to establish an organized and automated workflow for data analysis (**Figure 5-11**). Next Generation Sequencing data is formatted as a .fastq file, which is a list of “information blocks” that consist of four lines: sequence identifier, sequence, quality score identifier line (consisting only of a ‘+’), and the quality scores (**Figure 5-11**, pink block). Bruce Wittmann had previously written a package to parse NGS data for another project and shared this package with me. I modified the package to suit the needs of this project. The inputs are the forward and reverse .fastq files and the reference sequence in string format with the mutated codons represented as ‘nnn.’ The package parses through all of the information blocks in the forward and reverse .fastq files and pairs together reads sharing the same sequence identifier. Orphan sequences that do not have a pair, sequences whose average quality scores over all bases are >30, and sequence pairs whose timepoint barcode does not match due to recombination during the NGS adaptor PCR are

discarded. After sequences are parsed and paired, the now ‘wrangled’ sequence pairs are aligned to the reference sequence and the codons at mutated positions are translated and recorded. The number of each unique amino acid combination is counted, and a dictionary containing the amino-acid sequence and counts is returned for each different timepoint. These counts can be then used to calculate the frequencies and functional scores.

```

                                .fastq file info block
Label ----- @<instrument>:<run number>:<flowcell ID>:<lane>:<tile>:<x-pos>:<y-pos> <read>:<is filtered>:<control number>:<sample number>
Base calls ----- TCGCACTCATGTCGCGCTGCATATGACAAGACAGAATC
Quality scores --- +
                   <>:##=><9=AAAAAAAAAAAAA9#:<#<;<<<?????##=

                                Data analysis pipeline pseudocode
1 RunNGSAnalysis(refseq='TCGCACTCAACGnnnTGCATAnnnCAAGACAGAATC', forward_reads.fastq, reverse_reads.fastq):
2
3     NGS_Wrangler(forward_reads.fastq, reverse_reads.fastq):
4         read in all info_blocks in forward_reads.fastq
5         for info_block in forward_info_blocks:
6             SeqPair(info_Block):
7                 return (sequence pair object with label, sequence, quality score attributes)
8         add label to list of forward_unique_labels
9         read in all info blocks for reverse_reads.fastq
10        for info_block in reverse_info_blocks:
11            SeqPair(info_Block):
12                return (seqpair_object with label, sequence, quality score attributes)
13        if reverse_label in list of forward_unique_labels:
14            seqpair_object.attach_partner()
15        remove seqpairs without partners (orphans)
16        remove seqpairs with timepoint barcodes that do not match
17        remove pair if average Q score for f or r read is <30
18        return (wrangled_seqpairs)
19
20 Alignment(refseq, wrangled_seqpairs):
21
22     align(refseq, wrangled_seqpairs):
23         for seqpair in wrangled_seqpairs:
24             align to refseq and add to list
25         return list_of_alignments
26
27     analyze_alignment(refseq, list_of_alignments)
28         for alignment in list_of_alignments:
29             identify codons at refseq positions specified as 'nnn'
30             add string of codons to list_of_variants)
31         return(list_of_variants and number of variant counts)

```

Output:

```

T1={'AA': 40, T2={'AA': 27,
'AC', 83,      'AC', 89,
'GR', 76,      'GR', 114,
...
'YY', 26}     'YY', 66}

```

Figure 5-11. Pseudocode of the data analysis package.

The software also creates histograms of the average Q-scores for each read, which can be used to identify if there were any issues during the sequencing run. (Figure 5-12)

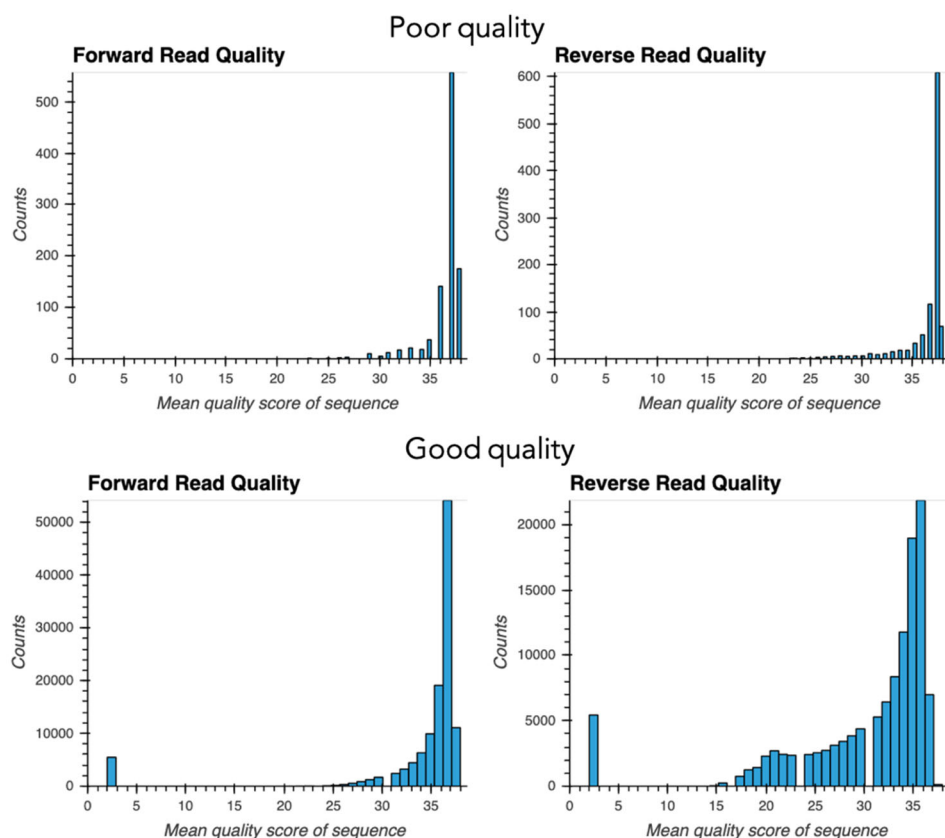


Figure 5-12. Example of poor and good quality forward and reverse read data. Poor quality read data has few overall reads compared to what is expected or can have the majority of reads with mean score below 30. Good quality data has few reads with mean score below 30 and thousands of high-quality reads. The expected quantity of reads will vary depending on type of sequencing performed.

Bibliography for Chapter V

1. Starr, T. N. & Thornton, J. W. Epistasis in protein evolution. *Protein Science* **25**, 1204–1218 (2016).
2. Yang, K. K., Wu, Z., & Arnold, F. H. Machine-learning-guided directed evolution for protein engineering. *Nat. Methods* **16**, 687–694 (2019).
3. Wittmann, B. J., Johnston, K. E., Wu, Z., & Arnold, F. H. Advances in machine learning for directed evolution. *Curr. Opin. Struct. Biol.* **69**, 11–18 (2021).
4. Fowler, D. M. & Fields, S. Deep mutational scanning: A new style of protein science.

- Nat. Methods* **11**, 801–807 (2014).
5. Fowler, D. M., Stephany, J. J., & Fields, S. Measuring the activity of protein variants on a large scale using deep mutational scanning. *Nat. Protoc.* **9**, 2267–2284 (2014).
 6. Matuszewski, S., Hildebrandt, M. E., Ghenu, A. H., Jensen, J. D., & Bank, C. A statistical guide to the design of deep mutational scanning experiments. *Genetics* **204**, 77–87 (2016).
 7. Rix, G. et al. Scalable continuous evolution for the generation of diverse enzyme variants encompassing promiscuous activities. *Nat. Commun.* **11**, 1–11 (2020).
 8. Lee, M. E., DeLoache, W. C., Cervantes, B., & Dueber, J. E. A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synth. Biol.* **4**, 975–986 (2015).
 9. Murciano-Calles, J., Romney, D. K., Brinkmann-Chen, S., Buller, A. R., & Arnold, F. H. A Panel of TrpB Biocatalysts Derived from Tryptophan Synthase through the Transfer of Mutations that Mimic Allosteric Activation. *Angew. Chemie - Int. Ed.* **55**, 11577–11581 (2016).
 10. Boville, C. E., Romney, D. K., Almhjell, P. J., Sieben, M., & Arnold, F. H. Improved Synthesis of 4-Cyanotryptophan and Other Tryptophan Analogues in Aqueous Solvent Using Variants of TrpB from *Thermotoga maritima*. *J. Org. Chem.* **83**, 7447–7452 (2018).
 11. Watkins, E. J., Almhjell, P. J., & Arnold, F. H. Direct enzymatic synthesis of a deep-blue fluorescent noncanonical amino acid from azulene and serine. *ChemBioChem* **21**, 80–83 (2020).
 12. Anishchenko, I., Ovchinnikov, S., Kamisetty, H., & Baker, D. Origins of coevolution between residues distant in protein 3D structures. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 9122–9127 (2017).
 13. Buller, A. R. et al. Directed Evolution Mimics Allosteric Activation by Stepwise Tuning of the Conformational Ensemble. *J. Am. Chem. Soc.* **140**, 7256–7266 (2018).
 14. Benatuil, L., Perez, J. M., Belk, J., & Hsieh, C.-M. An improved yeast transformation method for the generation of very large human antibody libraries. *Protein Eng. Des. Sel.* **23**, 155–159 (2010).
 15. Chao, G. et al. Isolating and engineering human antibodies using yeast surface display.

- Nat. Protoc.* **1**, 755–768 (2006).
16. Metzker, M. L. Sequencing technologies the next generation. *Nat. Rev. Genet.* **11**, 31–46 (2010).
 17. Gardner, E. C., Watkins, E. J., Gollihar, J., & Ellington, A. D. Emulsion-based directed evolution of enzymes and proteins in yeast. *Methods in Enzymology* (Elsevier Inc., 2020). doi:10.1016/bs.mie.2020.04.053
 18. Mitra, A., Skrzypczak, M., Ginalski, K., & Rowicka, M. Strategies for achieving high sequencing accuracy for low diversity samples and avoiding sample bleeding using Illumina platform. *PLoS One* **10**, 1–21 (2015).
 19. Lundberg, D. S., Yourstone, S., Mieczkowski, P., Jones, C. D., & Dangl, J. L. Practical innovations for high-throughput amplicon sequencing. **10**, (2013).
 20. Wu, L. et al. Phasing amplicon sequencing on Illumina Miseq for robust environmental microbial community analysis. *BMC Microbiol.* **15**, 1–12 (2015).
 21. Peng, Q., Vijaya Satya, R., Lewis, M., Randad, P., & Wang, Y. Reducing amplification artifacts in high multiplex amplicon sequencing by using molecular barcodes. *BMC Genomics* **16**, 1–12 (2015).
 22. Pronobis, M. I., Deutch, N., & Peifer, M. The Miraprep: A protocol that uses a Miniprep kit and provides Maxiprep yields. *PLoS One* **11**, 1–12 (2016).
 23. Ellefson, J. W. et al. Directed evolution of genetic parts and circuits by compartmentalized partnered replication. *Nat. Biotechnol.* **32**, 97–101 (2014).

SUPPLEMENTARY INFORMATION FOR CHAPTER V

D.1 Materials

Oligonucleotide primers were obtained from Integrated DNA Technologies (IDT DNA). PCRs were run with the Phusion® High-Fidelity PCR Kit (New England Biolabs). Gibson assembly mix is prepared with isothermal master mix in-house, and enzymes T5 exonuclease, Phusion® DNA polymerase, and Taq DNA ligase purchased from New England Biolabs. Gel extractions were performed with ZymoClean gel recovery kit.

D.2 Construction of triple-site saturation mutagenesis libraries

Libraries were constructed according to the cloning scheme described in **Section 5.4** and depicted in **Figure D-1**. The ‘gap’ PCR was set up using F Gap primer and R primer in **Table D-1**, PCR mix in **Table D-2**, and thermal cycler protocol in **Table D-3**. The PCR was then *DpnI* digested and analyzed by gel electrophoresis. The properly sized fragment was gel extracted and eluted in 30–50 μ L ddiH₂O. The library PCR was set up using the F library primer and R primer in **Table S-2**, PCR mix in **Table D-2**, and thermal cycler protocol in **Table D3**.

Table D-1. Gap/3X SSM Library PCR primers

Lib.	Primer	EJW Primer name	Sequence
A	F Gap	EJW_pr205_9D8st_106_gap_F	GGTGCTGGTCAGCACG
	F Library	EJW_pr174_DMS_104-105-106_F	AATGGGCAAAACCCGTATCATTNNKNNKNNKGGTGCTGG TCAGCACG
	R	EJW_pr177_DMS_104_R	AATGATACGGGTTTTGCCATTAGTTTTGCCAGCAGAACC TGGC
B	F Gap	EJW_pr218_9D8st_107_gap_F	GCTGGTCAGCACGGC
	F Library	EJW_pr175_DMS_105-106-107_F	AATGGGCAAAACCCGTATCATTGCTNNKNNKNNKGGTGCT CAGCACGGC
	R	EJW_pr177_DMS_104_R	AATGATACGGGTTTTGCCATTAGTTTTGCCAGCAGAACC TGGC
C	F Gap	EJW_pr219_9D8st_108_gap_F	GGTCAGCACGGCGTAG
	F Library	EJW_pr176_DMS_106-107-108_F	AATGGGCAAAACCCGTATCATTGCTGAAANNKNNKNNKGGT CAGCACGGCGTAG
	R	EJW_pr177_DMS_104_R	AATGATACGGGTTTTGCCATTAGTTTTGCCAGCAGAACC TGGC
D	F Gap	EJW_pr220_9D8st_119_gap_F	GCGCTGTTCCGGTATGGAAT
	F Library	EJW_pr178_DMS_117-118-119_F	ACGGCGTAGCAACTGCTNNKNNKNNKGGCCTGTTCCGGTAT GGAATGTGTAATC
	R	EJW_pr179_DMS_117_R	AGCAGTTGCTACGCCGTGCTGACCAGCACCCGTTTCAG
E	F Gap	EJW_pr221_9D8st_186_gap_F	GTGGTTGGTCCGCATCC
	F Library	EJW_pr180_DMS_184-185-186_F	CTGCAGACCACCTATTACGTNNKNNKNNKGTGGTTGGTC CGCATCCATATCC
	R	EJW_pr181_DMS_184_R	CACGTAATAGGTGGTCTGCAGGTTGTAATCCAGTCACG CAGAGCT
F*	F Gap	EJW_pr222_9D8st_166_gap_F	GACGAAGCTCTGCGTGAC
	F Library	EJW_pr182_DMS_162-166_F	GTAAAATCCGGTAGCCGTACCNNKAAAGACGCANNKGCAC GAAGCTCTG
	R	EJW_pr183_DMS_162_R	GGTACGGCTACCGGATTTTACCGGTACAACCTTAGCACCC AGCAG
G*	F Gap	EJW_pr223_9D8st_228_gap_F	GGTGGTTCTAACGCTGCC
	F Library	EJW_pr184_DMS_227-228_F	GGACTACATCGTTGCGTGCNNKNNKGGTGGTTCTAACGC TGCCGGTA
	R	EJW_pr185_DMS_227_R	GCACGCAACGATGTAGTCCGGCAGACGGCCTTCTTTTCT GG
H	F Gap	EJW_pr224_9D8st_231_gap_F	AACGCTGCCGGTATCTTCTAT
	F Library	EJW_pr186_DMS_228-230-231_F	GGACTACATCGTTGCGTGCNNKGGTNNKNNKAACGCT GCCGGTATCTTCTATCCG
	R	EJW_pr185_DMS_227_R	GCACGCAACGATGTAGTCCGGCAGACGGCCTTCTTTTCT GG
I	F Gap	EJW_pr225_9D8st_184_gap_F	GGCTCTGTGGTTGGTCC
	F Library	EJW_pr187_DMS_182-183-184_F	CCAACCTGCAGACCACCTATNNKNNKNNKGGCTCTGTGGT TGGTCCGC
	R	EJW_pr188_DMS_182_R	ATAGGTGGTCTGCAGGTTGGTAATCCAGTCACGCAGAGC TTCGT

*Make 301X library and use as template for these libraries

Table D-2. Gap/3X SSM PCR master mix

Component	Volume (μL)
primer_for (10 μ M)	1
primer_rev (10 μ M)	1
template	0.5
NTP mix (10 mM)	1
Phusion	0.5
HF buffer (5x)	10
water	33.5
DMSO	0.5
Total:	50

Table D-3. Gap/3X SSM PCR thermal cycler protocol

Temperature (C)	Time (S)	Cycles
98	30	1
98	10	
55	15	30
72	200	
72	600	1
10	Infinite	1

D.3 NGS amplicon preparation

To prepare NGS amplicons, extract DNA from host and quantify amount by NanoDrop. It is good to know roughly how many copies of DNA are being sampled to ensure adequate representation of all library variants. Perform the internal primer PCR using primers from **Table D-4**, PCR mix from **Table D-5**, and thermal cycler protocol from **Table D-6**. The internal primer PCR amplifies the region of DNA of interest (**Figure 5-6B** and **Figure 5-7**) and attaches the random 10mer and timepoint barcode to each amplicon. *DpnI* digest the PCR, run on analytical gel, and gel extract. Elute in 20 μ L. This can be directly sent to Laragen or can be used as template for external primer PCR to attach the Illumina Nexterra adaptors (Primers: **Table D-7**, PCR mix: **Table D-8**, thermal cycler protocol: **Table D-9**). The external primer PCR is not necessary as Laragen will attach its own adaptors but can be useful if more control over PCR cycle number is desired.

Table D-4. Internal NGS amplicon PCR primers

EJW Primer name	i5/i7	Barcode	Seed region	F/R	Sequence 5' → 3'
EJW_pr191_N501_TGT_266-285_F	i5	TGT	226-285	F	TCGTCGGCAGCGTCAGATGTGTATAA GAGACAGNNNNNNNNNN TGTGCCAG GTTCTGCTGGCAAAA
EJW_pr192_N501_GTG_266-285_F	i5	GTG	226-285	F	TCGTCGGCAGCGTCAGATGTGTATAA GAGACAGNNNNNNNNNN GTGCCAG GTTCTGCTGGCAAAA
EJW_pr193_N501_ACA_266-285_F	i5	ACA	226-285	F	TCGTCGGCAGCGTCAGATGTGTATAA GAGACAGNNNNNNNNNN ACAGCCAG GTTCTGCTGGCAAAA
EJW_pr198_N701_GTG_266-285_F	i7	GTG	226-285	F	GTCTCGTGGGCTCGGAGATGTGTATA AGAGACAGNNNNNNNNNN GTGCCA GTTCTGCTGGCAAAA
EJW_pr195_N701_TGT_415-394_R	i7	TGT	415-394	R	GTCTCGTGGGCTCGGAGATGTGTATA AGAGACAGNNNNNNNNNN GTCTCTG GCGGATCGTGCTTCTTC
EJW_pr196_N701_GTG_415-394_R	i7	GTG	415-394	R	GTCTCGTGGGCTCGGAGATGTGTATA AGAGACAGNNNNNNNNNN GTGTCTG GCGGATCGTGCTTCTTC
EJW_pr197_N701_ACA_415-394_R	i7	ACA	415-394	R	GTCTCGTGGGCTCGGAGATGTGTATA AGAGACAGNNNNNNNNNN ACATCTG GCGGATCGTGCTTCTTC
EJW_pr206_DMS_i5_TGT_448-473_F	i5	TGT	448-473	F	TCGTCGGCAGCGTCAGATGTGTATAA GAGACAGNNNNNNNNNN TGTGCTAA AGTTGTACCGGTAAAATCCGG
EJW_pr207_DMS_i5_GTG_448-473_F	i5	GTG	448-473	F	TCGTCGGCAGCGTCAGATGTGTATAA GAGACAGNNNNNNNNNN GTGCTAA AGTTGTACCGGTAAAATCCGG
EJW_pr208_DMS_i5_ACA_448-473_F	i5	ACA	448-473	F	TCGTCGGCAGCGTCAGATGTGTATAA GAGACAGNNNNNNNNNN ACAGCTAA AGTTGTACCGGTAAAATCCGG
EJW_pr209_DMS_i7_TGT_589-565_R	i7	TGT	589-565	R	GTCTCGTGGGCTCGGAGATGTGTATA AGAGACAGNNNNNNNNNN GTTCGAT AATCGGATATGGATGCCGACC
EJW_pr210_DMS_i7_GTG_589-565_R	i7	GTG	589-565	R	GTCTCGTGGGCTCGGAGATGTGTATA AGAGACAGNNNNNNNNNN GTGCGAT AATCGGATATGGATGCCGACC
EJW_pr211_DMS_i7_ACA_589-565_R	i7	ACA	589-565	R	GTCTCGTGGGCTCGGAGATGTGTATA AGAGACAGNNNNNNNNNN ACCGAT AATCGGATATGGATGCCGACC
EJW_pr215_DMS_i5_TGT_659-678_F	i5	TGT	659-678	F	TCGTCGGCAGCGTCAGATGTGTATAA GAGACAGNNNNNNNNNN TGTCCGAC TACATCGTTGCGTGC
EJW_pr216_DMS_i5_GTG_659-678_F	i5	GTG	659-678	F	TCGTCGGCAGCGTCAGATGTGTATAA GAGACAGNNNNNNNNNN GTCCGAC TACATCGTTGCGTGC
EJW_pr217_DMS_i5_ACA_659-678_F	i5	ACA	659-678	F	TCGTCGGCAGCGTCAGATGTGTATAA GAGACAGNNNNNNNNNN ACCGAC TACATCGTTGCGTGC
EJW_pr212_DMS_i7_TGT_929-911_R	i7	TGT	929-911	R	GTCTCGTGGGCTCGGAGATGTGTATA AGAGACAGNNNNNNNNNN GTTAGG CGTGTTCCGGACCG
EJW_pr213_DMS_i7_GTG_929-911_R	i7	GTG	929-911	R	GTCTCGTGGGCTCGGAGATGTGTATA AGAGACAGNNNNNNNNNN GTTAGG CGTGTTCCGGACCG
EJW_pr214_DMS_i7_ACA_929-911_R	i7	ACA	929-911	R	GTCTCGTGGGCTCGGAGATGTGTATA AGAGACAGNNNNNNNNNN ACATAGG CGTGTTCCGGACCG

Table D-5. Internal NGS amplicon PCR master mix

Component	Conc.	Volume (μL)
internal primer_for	10 μM	1
internal primer_rev	10 μM	1
Template	~2400 ng	
NTP mix	10 mM	1
Phusion		0.5
HF buffer (5x)		10
DMSO		0.5
water		Q.S. 50
Total:		50

Table D-6. Internal NGS amplicon PCR thermal cycler protocol

Step	Temperature (C)	Time (S)	Cycles
Initial denature	98	30	1
Denature	98	20	
Anneal	55	15	15
Extend (30 sec/kb)	72	Variable	
Hold	10	infinite	1

Table D-7. Nexterra adaptor ('external') primers

EJW Primer name	P5/P7	Nexterra index	Sequence 5' → 3'
EJW_pr151_DMS_N501_i5_adaptor	P5	N501	AATGATACGGCGACCACCGAGATCTACACTAGATCG CTCGTCGGCAGCGTCAGAT
EJW_pr152_DMS_N701_i7_adaptor	P7	N701	CAAGCAGAAGACGGCATAACGAGATTCGCCTTAGTCT CGTGGGCTCGGAGATG
EJW_pr155_DMS_N502_i5_adaptor	P5	N502	AATGATACGGCGACCACCGAGATCTACACCTCTCTA TTCGTCGGCAGCGTCAGAT
EJW_pr156_DMS_N702_i7_adaptor	P7	N702	CAAGCAGAAGACGGCATAACGAGATCTAGTACGGTCT CGTGGGCTCGGAGATG
EJW_pr157_DMS_N503_i5_adaptor	P5	N503	AATGATACGGCGACCACCGAGATCTACACTATCCTC TTCGTCGGCAGCGTCAGAT
EJW_pr158_DMS_N703_i7_adaptor	P7	N703	CAAGCAGAAGACGGCATAACGAGATTCTGCCTGTCT CGTGGGCTCGGAGATG

Table D-8. External NGS amplicon PCR

Component	Conc.	Volume (μL)
external primer_for	10 μ M	1
external primer_rev	10 μ M	1
Template (PCR product from step 1)		20
NTP mix	10 mM	1
Phusion		0.5
HF buffer (5x)		10
DMSO		0.5
water		16
Total:		50

Table D-9. External NGS amplicon PCR thermal cycler protocol

Step	Temperature (C)	Time (S)	Cycles
Initial denature	98	30	1
Denature	98	20	
Anneal	55	15	5
Extend (30 sec/kb)	72	Variable	

Table D-10. Sequencing primer combinations

Library	Site 1	Site 2	Site 3	F Seed	R Seed	Seq length (bp)
A	104	105	106	226-285	415-394	150
B	105	106	107	226-285	415-394	150
C	106	107	108	226-285	415-394	150
D	117	118	119	226-285	415-394	150
E	184	185	186	448-473	589-565	142
F	162	166	301	448-473	929-911	482
G	227	228	301	659-678	929-911	271
H	228	230	231	659-678	929-911	271
I	182	183	184	448-473	589-565	142



Figure D-1. Snapshot of TrpB gene with annotations of seed regions of internal primers and the amplicons that would be generated