



Discovery and Characterization of a Prominent Gut Microbial Glycyl Radical Enzyme Responsible for 4-Hydroxyproline Metabolism

Citation

Huang, Yue. 2019. Discovery and Characterization of a Prominent Gut Microbial Glycyl Radical Enzyme Responsible for 4-Hydroxyproline Metabolism. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

Link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:41121269>

Terms of use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Other Posted Material (LAA), as set forth at

<https://harvardwiki.atlassian.net/wiki/external/NGY5NDE4ZjgzNTc5NDQzMGIzZWZhMGFIOWI2M2EwYTg>

Accessibility

<https://accessibility.huit.harvard.edu/digital-accessibility-policy>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#)

**Discovery and characterization of a prominent gut microbial glycyl radical enzyme
responsible for 4-hydroxyproline metabolism**

A dissertation presented by

Yue Huang

to

the Committee on Higher Degrees in Chemical Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Chemical Biology

Harvard University

Cambridge, Massachusetts

October 2018

© 2018 – Yue Huang

All rights reserved

**Discovery and characterization of a prominent gut microbial glycyl radical enzyme
responsible for 4-hydroxyproline metabolism**

Abstract

The human gut is one of the most densely populated microbial habitat on Earth and the gut microbiota is extremely important in maintaining health and disease states. Advances in sequencing technologies have enabled us to gain a better understanding of microbiome compositions, but the majority of microbial genes are not functionally annotated. Therefore, the molecular basis by which gut microbes influence human health remains largely unknown. Mechanistic studies connecting sequences to functions remain a priority in this field. In this thesis, we describe the discovery and characterization of a new glycyl radical enzyme (GRE) responsible for anaerobic metabolism of an abundant host-derived amino acid. From our work on this enzyme, we uncovered a widely distributed metabolic capability for *trans*-4-hydroxy-L-proline (Hyp) dehydration in the gut microbiome.

The activity of this novel GRE was proposed based on its genomic context and high sequence similarities to characterized GRE eliminases. This GRE was predicted to be a Hyp dehydratase (HypD) that catalyzes the dehydration of Hyp to (*S*)- Δ^1 -pyrroline-5-carboxylate (P5C). Through *in vitro* reconstitution of the GRE and its activating enzyme, we experimentally validated its proposed function. Hyp metabolism was demonstrated specifically in HypD-encoding Clostridiales in culture-based experiments. HypD was found among common gut isolates and was

detected in all stool metagenomes analyzed. Overall, the work covered in this Chapter has expanded known reactivities in the GRE superfamily and revealed the enzyme responsible for anaerobic Hyp metabolism.

Chapter 3 details work from a close collaboration with the Drennan lab and help from the Raines lab at MIT. Structural elucidation of substrate-bound HypD was achieved through X-ray crystallography by Lindsey Backman (Drennan lab) and Hyp conformation was calculated by Dr. Brian Gold (Raines lab). The HypD crystal structure informed site-directed mutagenesis of conserved residues. Biochemical characterization of these protein mutants revealed the importance of these residues in activity. The structural and biochemical work provided insight into the molecular basis of Hyp dehydration. The unique features of this elimination reaction were highlighted and a mechanism was proposed for HypD.

Chapter 4 describes the phylogenetic analyses of HypD sequences along with analyses of their genome neighborhoods. Hyp-degrading bacterial communities from environmental samples were obtained through enrichment culturing. Determination of community compositions revealed species closely related to sequenced HypD-encoding isolates. Detection of *hypD* in these cultures provided support for HypD to be the principle enzyme responsible for anaerobic Hyp metabolism.

In Chapter 5, we present work toward identifying the physiological role of HypD among gut Clostridiales and Bacteroidales. We demonstrated that HypD is part of Stickland fermentation and upstream of L-proline reduction. With help from Prof. Aimee Shen (Tufts University), *Clostridioides difficile* deletion mutants were generated to show HypD is necessary for Hyp metabolism. From comparative transcriptomics of *C. difficile*, pathways were found to be differentially regulated by Hyp and Pro. The activity of HypD from *Bacteroides vulgatus* was verified *in vitro* so we attempted to detect Hyp metabolism in gut Bacteroidales. *Bacteroides*

genetics, growth experiments, and competitive colonization of mice part of a collaboration with Prof. Laurie Comstock (Harvard Medical School) are described. At the end of this Chapter, we discuss in depth potential roles of Hyp metabolism in microbial physiology and its impact on host biology.

Table of Contents

Abstract.....	iii
Table of Contents	vi
List of Figures.....	xv
List of Tables	xx
List of Abbreviations	xxii
Acknowledgements	xxv
1 Chapter 1. Introduction to the human gut microbiome, the glycyl radical enzyme superfamily, and 4-hydroxyproline metabolism	1
1.1 The gut microbiota influences human physiology	1
1.2 Glycyl radical enzymes are abundant in the gut microbiome	5
1.3 Mechanistic overview of glycyl radical enzymes.....	7
1.3.1 The general catalytic cycle of GREs	7
1.3.2 Recent mechanistic work on GRE activating enzymes	11
1.4 Shared structural features of glycyl radical enzymes	13
1.5 Glycyl radical enzymes catalyze key steps in primary metabolic pathways	18
1.5.1 Formate-lyases: PFL, TdcE	22
1.5.2 Class III ribonucleotide reductases: NrdD.....	24
1.5.3 Decarboxylases: HPAD, PAD, IAD	26

1.5.4	X-succinate synthases: BSS, XSS	30
1.5.5	Eliminases: GD, PD, CutC	31
1.6	<i>trans</i> -4-Hydroxy-L-proline is an abundant non-proteinogenic amino acid in the eukaryotic world.....	43
1.6.1	Sources of Hyp in the human gut	46
1.7	Human metabolism of <i>trans</i> -4-hydroxy-L-proline	47
1.8	Bacterial strategies for metabolizing <i>trans</i> -4-hydroxy-L-proline	50
1.8.1	Stickland fermentation.....	50
1.8.2	Catabolic metabolism of Hyp in aerobic bacteria	53
1.9	Chapter preview.....	55
1.10	References.....	56
2	Chapter 2. Discovery and characterization of a <i>trans</i>-4-hydroxy-L-proline dehydratase	71
2.1	Introduction	71
2.1.1	Discovery of a new glycyl radical enzyme predicted to be a <i>trans</i> -4-hydroxy-L-proline dehydratase	72
2.1.2	A putative role for HypD in Stickland fermentation	74
2.2	Results and discussions	75
2.2.1	Multiple sequence alignments and construction of a HypD homology model	75
2.2.2	Heterologous overexpression and purification of proteins in an <i>E. coli proC</i> deletion mutant.....	81

2.2.3	Characterization of P5CR activity using a Pro dehydrogenase assay	87
2.2.4	Purification and [4Fe-4S] cluster reconstitution of AE, a radical SAM enzyme 89	
2.2.5	Demonstration of <i>trans</i> -4-hydroxy-L-proline dehydration activity <i>in vitro</i>	93
2.2.6	HypD is highly specific toward <i>trans</i> -4-hydroxy-L-proline.....	98
2.2.7	Kinetic analysis of HypD in a spectrophotometric coupled enzyme assay ...	103
2.2.8	Analysis of 4-hydroxyproline structural analogs as inhibitors of HypD.....	105
2.2.9	Utilization of <i>trans</i> -4-hydroxy-L-proline by HypD-encoding Clostridiales..	113
2.2.10	Phylogenetic distribution of HypD in sequenced bacterial and archaeal genomes	115
2.2.11	HypD is highly abundant and prevalent in human stool metagenomes	118
2.3	Conclusions	125
2.4	Materials and methods.....	127
2.4.1	Construction of HypD homology model	128
2.4.2	Compilation of sequenced bacterial and archaeal genomes encoding putative HypDs	129
2.4.3	Construction of plasmids for overexpression of HypD, HypD-AE, and P5CR 129	
2.4.4	Generation of a P5CR deletion mutant in <i>E. coli</i>	131
2.4.5	Overexpression and purification of HypD.....	132
2.4.6	Overexpression and purification of MBP-AE	134

2.4.7	Overexpression and purification of P5CR.....	135
2.4.8	Visualization of AE on SDS-PAGE under different reducing conditions.....	136
2.4.9	Detection of [4Fe-4S] cluster(s) in HypD-AE through spectrophotometry ..	137
2.4.10	Determination of iron content in HypD-AE.....	137
2.4.11	Determination of sulfide content in HypD-AE	137
2.4.12	Glycyl radical detection and quantification by electron paramagnetic resonance (EPR) spectroscopy	137
2.4.13	Spectrophotometric assay for proline dehydrogenase activity of P5CR.....	139
2.4.14	End-point enzymatic assays for HypD activity	140
2.4.15	LC-MS/MS methods for detection of Pro and Hyp	141
2.4.16	Coupled enzyme spectrophotometric assay for HypD kinetics.....	142
2.4.17	4-Hydroxyproline derivatization and detection on chiral HPLC	143
2.4.18	Coupled enzyme spectrophotometric for inhibition of HypD	143
2.4.19	Clostridiales growth experiments and metabolite analyses.....	144
2.4.20	Construction of the GRE sequence similarity network	145
2.4.21	Determination of enzyme abundances in metagenomes and metatranscriptomes	
	146	
2.5	References	150
3	Chapter 3. Biochemical and structural studies of <i>trans</i>-4-hydroxy-L-proline dehydratase.....	156

3.1	Introduction	156
3.2	Results and discussions	159
3.2.1	Overall architecture of HypD is similar to other GREs.....	159
3.2.2	Key residues in the active site predicted to be important for catalysis.....	167
3.2.3	A distinct set of residues likely facilitates dehydration in HypD	170
3.2.4	Additional active site residues likely facilitate Hyp binding.....	177
3.3	Conclusions	181
3.4	Materials and methods.....	190
3.4.1	Vector constructs for the overexpression of HypD, HypD-AE, and P5CR ..	190
3.4.2	Site-directed mutagenesis and construction of overexpression vectors for HypD mutants	191
3.4.3	Overexpression and purification of proteins used for biochemical experiments	193
3.4.4	Overexpression and purification of HypD wild-type for crystallography studies	194
3.4.5	Glycyl radical detection and quantification by EPR spectroscopy.....	194
3.4.6	End-point activity assays with HypD mutants	194
3.4.7	Kinetic analysis of HypD variants using a coupled spectrophotometric assay	195
3.4.8	Crystallization of wild-type HypD	195
3.4.9	Structure determination of HypD	196

3.4.10	Density function theory calculations for Hyp conformation.....	199
3.5	References	200
4	Chapter 4. Anaerobic <i>trans</i>-4-hydroxy-L-proline utilization mediated by HypD is a widespread microbial metabolic activity	203
4.1	Introduction	203
4.2	Results and discussions	204
4.2.1	Phylogenetic analysis of HypD sequences	204
4.2.2	Anaerobic enrichment culturing of Hyp-degrading bacteria from sediment samples.....	212
4.2.3	Community composition of enrichment cultures by 16S rDNA sequencing	215
4.2.4	Detection of <i>hypD</i> in enrichment cultures through degenerate PCR.....	221
4.3	Conclusions	226
4.4	Materials and methods.....	229
4.4.1	Construction of a phylogenetic tree with representative HypD sequences ...	229
4.4.2	Enrichment culturing of Hyp-degrading microbes.....	230
4.4.3	Hyp derivatization and detection by HPLC-RID	230
4.4.4	Amplification of 16S rRNA gene from genomic DNA of enrichment cultures	231
4.4.5	Construction of a phylogenetic tree with 16S rRNA genes amplified from enrichment cultures	232
4.4.6	Design of degenerate primers for <i>hypD</i> genes.....	232

4.4.7	Degenerate PCR and sequencing of partial <i>hypD</i> fragments from enrichment cultures	233
4.5	References	235
5	Chapter 5. Toward an understanding of the physiological roles of anaerobic 4-hydroxyproline metabolism in the gut microbiota	238
5.1	Introduction	238
5.2	Results and discussions	239
5.2.1	The role of HypD in Stickland fermentation by Clostridiales.....	239
5.2.2	Generation of deletion mutants in <i>C. difficile</i> 630 Δ <i>erm</i>	244
5.2.3	Phenotypes of <i>C. difficile</i> 630 Δ <i>erm</i> Δ <i>hypD</i> and 630 Δ <i>erm</i> Δ <i>proC</i> strains.....	246
5.2.4	Comparative transcriptomics of <i>C. difficile</i> 630 Δ <i>erm</i> during growth on Hyp, Pro, and a rich medium using RNA-seq	251
5.2.5	Activity of HypD from <i>Bacteroides vulgatus</i> ATCC 8482	273
5.2.6	Growth of <i>Bacteroides</i> and <i>Parabacteroides</i> species in a minimal medium containing Pro or Hyp	280
5.2.7	Genetics in <i>Bacteroides</i> and <i>Parabacteroides</i> species for heterologous expression and deletion of BvHypD and BvAE	285
5.2.8	Experiments examining growth of <i>B. vulgatus</i> ATCC 8482 WT, Δ <i>hypD</i> , and Δ <i>AE</i> strains.....	288
5.2.9	Competitive colonization of germ-free mice by <i>B. vulgatus</i> WT and Δ <i>hypD</i> strains	290

5.3	Conclusions	292
5.3.1	Anaerobic Hyp metabolism by gut microbes	293
5.3.2	Hyp utilization may contribute to shaping the microbiota composition	296
5.3.3	Stickland fermentation of Hyp during <i>C. difficile</i> colonization	297
5.3.4	Breakdown of host collagen to access Hyp	298
5.3.5	The biological activity of the fermentation product 5-aminovalerate may influence host pathways	299
5.4	Materials and methods.....	300
5.4.1	Growth experiments and metabolite analyses for <i>C. difficile</i> 630 Δ <i>erm</i> and <i>T. glycolicus</i> DSM 1288.....	301
5.4.2	Construction of pMTL-YN3 plasmids for genetic deletions in <i>C. difficile</i> 630 Δ <i>erm</i> Δ <i>pyrE</i>	302
5.4.3	Construction of pMTL-YN1C plasmids for genetic complementation in <i>C. difficile</i> 630 Δ <i>erm</i> Δ <i>pyrE</i>	308
5.4.4	Construction of deletion mutants <i>C. difficile</i> 630 Δ <i>erm</i> Δ <i>hypD</i> and <i>C. difficile</i> 630 Δ <i>erm</i> Δ <i>proC</i>	308
5.4.5	Growth experiments and metabolite analyses for <i>C. difficile</i> 630 Δ <i>erm</i> WT, Δ <i>hypD</i> , and Δ <i>proC</i>	310
5.4.6	RNA-sequencing of <i>C. difficile</i> 630 Δ <i>erm</i> in MACC-Pro, MACC-Hyp, and RCM	310

5.4.7	Construction of pTrcHis2A plasmids for overexpression of HypD-AE gene clusters	312
5.4.8	Detection of HypD activity in <i>E. coli</i> overexpression strains in BHI	312
5.4.9	Construction of plasmids for overexpression of BvHypD and BvAE in <i>E. coli</i>	313
5.4.10	Overexpression and purification of BvHypD, BvAE, and CdP5CR.....	314
5.4.11	End-point and kinetic assays for BvHypD	314
5.4.12	Growth experiments and metabolite analyses for <i>Bacteroides</i> and <i>Parabacteroides</i> species	315
5.4.13	Construction of pFD340 vectors for protein expression in <i>Bacteroides</i> species	316
5.4.14	Construction of plasmids for genetic deletions in <i>B. vulgatus</i> ATCC 8482	316
5.4.15	Construction of expression strains in <i>Bacteroides</i> species using pFD340-BvHypDAE	317
5.4.16	Deletions of HypD and AE in <i>B. vulgatus</i> ATCC 8482.....	318
5.4.17	Competitive colonization of germ-free mice with <i>B. vulgatus</i> ATCC 8482 WT and $\Delta hypD$ strains	319
5.5	References	320

List of Figures

Figure 1.1 Metabolic crosstalk between host and gut microbes.	3
Figure 1.2 GRE-catalyzed transformations that produce metabolites linked to human health.	7
Figure 1.3 Catalytic cycle of a glycy radical enzyme.....	10
Figure 1.4 Structural features shared among glycy radical enzymes.	14
Figure 1.5 A multiple sequence alignment of Fe–S cluster binding motifs present in activating enzymes.....	17
Figure 1.6 Five classes of known GREs categorized by activity.....	20
Figure 1.7 Proposed mechanisms for diol dehydration are distinct for GRE and B ₁₂ -dependent dehydratases.....	36
Figure 1.8 Proposed mechanism for glycerol dehydration catalyzed by GD.	40
Figure 1.9 Proposed mechanism for propanediol dehydration catalyzed by PD.....	41
Figure 1.10 Proposed mechanism for TMA elimination catalyzed by CutC.....	42
Figure 1.11 Biosynthesis and sources of <i>trans</i> -4-hydroxy-L-proline in eukaryotes.	44
Figure 1.12 Biosynthesis and sources of <i>trans</i> -4-hydroxy-L-proline in bacteria and archaea.....	46
Figure 1.13 <i>trans</i> -4-Hydroxy-L-proline metabolism in the human body.....	49
Figure 1.14 Stickland fermentation carried out by species in the order Clostridiales.	52
Figure 1.15 Catabolic metabolism of <i>trans</i> -4-hydroxy-L-proline by bacteria.	54
Figure 2.1 A conserved gene cluster identified in the order Clostridiales and hypothesis for its activities.	73
Figure 2.2 Dehydration activity proposed for the new GRE parallels known reactivities of GRE eliminases.....	74
Figure 2.3 A multiple sequence alignment of putative HypDs with characterized GREs.....	77

Figure 2.4 A homology model of HypD supports its role as a dehydratase.	80
Figure 2.5 SDS-PAGE of purified enzymes used in biochemical studies.	82
Figure 2.6 End-point activity assay for HypD with contaminating P5CR activity.	83
Figure 2.7 Contaminating P5CR activity in purified proteins and <i>E. coli</i> cell lysates.	85
Figure 2.8 Proline dehydrogenase activity of P5CR.	88
Figure 2.9 SDS-PAGE visualization of purified AE as two protein bands.	91
Figure 2.10 Absorbance spectra of HypD-AE indicate the presence of a redox active [4Fe-4S] cluster.	93
Figure 2.11 Assay set up for activation of HypD by its cognate AE.	94
Figure 2.12 EPR spectrum of activated HypD.	95
Figure 2.13 End-point activity assay for HypD.	97
Figure 2.14 LC-MS/MS chromatograms of Hyp and Pro in end-point assays.	98
Figure 2.15 HypD activity toward hydroxyproline structural isomers.	100
Figure 2.16 LC-MS/MS chromatograms of Pro in assays testing activity toward structural isomers.	101
Figure 2.17 Commercial sources of 4-hydroxyprolines are enantiomerically pure as determined by chiral-HPLC.	103
Figure 2.18 Kinetic analysis of HypD from <i>C. difficile</i> 70-100-2010.	105
Figure 2.19 Assay designed to test HypD inhibition by Hyp structural analogs.	107
Figure 2.20 Preliminary results from HypD inhibition assays by substrate analogs.	109
Figure 2.21 Varying P5CR concentration in HypD kinetic assay.	112
Figure 2.22 Growth of <i>hypD</i> ⁺ and <i>hypD</i> ⁻ Clostridial strains in the presence of Hyp or Pro and metabolite analyses.	114

Figure 2.23 Distribution of HypD in sequenced bacterial and archaeal genomes.....	116
Figure 2.24 Sequence similarity network (SSN) of the GRE superfamily.....	120
Figure 2.25 Abundance of GREs in HMP metagenomes.	123
Figure 3.1 Hyp dehydration catalyzed by a GRE revealed a prominent metabolic pathway.	157
Figure 3.2 HypD shares the canonical GRE fold with conserved Gly and Cys loops.....	161
Figure 3.3 Hyp conformers generated from DFT calculations.	163
Figure 3.4 Conserved Gly and Cys residues in HypD involved in radical transfer.....	166
Figure 3.5 Active site residues in HypD selected for site-directed mutagenesis.....	168
Figure 3.6 Active site residues are conserved among all putative HypD sequences.....	170
Figure 3.7 SDS-PAGE of purified proteins used in this study.	172
Figure 3.8 Activation of HypD wild-type and mutants detected by EPR spectroscopy.....	173
Figure 3.9 Most HypD mutants did not have detectable activity.	175
Figure 3.10 Kinetic analysis of HypD-Y450F and HypD-T645A.....	179
Figure 3.11 A Phe340 potentially involved in aromatic-proline interaction.	181
Figure 3.12 Transformations catalyzed by radical enzymes proposed to involve an α -aminoalkyl radical intermediates.	183
Figure 3.13 Similar reactions catalyzed by other radical enzymes.....	185
Figure 3.14 Dihedral angles of substrates bound in GRE eliminases.....	188
Figure 3.15 Proposed mechanism for Hyp dehydration by HypD.....	189
Figure 4.1 A phylogenetic tree of representative HypD sequences from the UniProt database.	210
Figure 4.2 Workflow for enrichment culturing of Hyp-degrading microbes.	213
Figure 4.3 Hyp detection in enrichment cultures using HPLC-RID.....	215
Figure 4.4 A phylogenetic tree of 16S rDNA sequenced from enrichment cultures.....	219

Figure 4.5 Degenerate primers designed to amplify <i>hypD</i> partial fragments.	222
Figure 4.6 Amplification of partial <i>hypD</i> sequences using degenerate PCR.....	225
Figure 4.7 Hyp metabolism mediated by HypD interfaces with several amino acid metabolic pathways.	228
Figure 5.1 Stickland fermentation pathways for Hyp and Pro.....	241
Figure 5.2 Growth of two HypD-encoding Clostridiales and metabolite analysis.	243
Figure 5.3 Growth of <i>C. difficile</i> 630 Δ <i>erm</i> WT, Δ <i>hypD</i> , and Δ <i>proC</i> mutants in a rich medium.	247
Figure 5.4 Growth of <i>C. difficile</i> 630 Δ <i>erm</i> WT, Δ <i>hypD</i> , and Δ <i>proC</i> mutants in a minimal medium.	250
Figure 5.5 RNA-seq experimental set up.....	253
Figure 5.6 Hierarchical clustering of Euclidean distances for all samples.	255
Figure 5.7 Hierarchical clustering of highly variable genes across all samples.	258
Figure 5.8 Histograms of false discovery rates for all genes analyzed in the MACC-Hyp/MACC-Pro expression comparison.	260
Figure 5.9 A scatter plot comparing read counts from MACC-Hyp to MACC-Pro.....	262
Figure 5.10 Attempted detection of Hyp metabolism in <i>E. coli</i> overexpressing the HypD-AE gene cluster.....	276
Figure 5.11 SDS-PAGE of purified BvHypD and MBP-BvAE from <i>B. vulgatus</i> ATCC 8482.	277
Figure 5.12 BvHypD is active toward Hyp in a coupled enzyme kinetic assay.	279
Figure 5.13 End-point activity assay for BvHypD.	280
Figure 5.14 Growth of HypD-encoding Bacteroidales on M9 minimal medium supplemented with Hyp.....	284

Figure 5.15 Growth of <i>B. vulgatus</i> ATCC 8482 WT, $\Delta hypD$, and ΔAE strains in BHIS.	289
Figure 5.16 Proposed scheme for Hyp metabolism by the gut microbiota.....	296

List of Tables

Table 2.1 Distribution of HypD in sequenced bacterial and archaeal genomes at the phylum level.	117
Table 2.2 Primers used for cloning in this study.	130
Table 3.1 Cartesian coordinates for zwitterionic Hyp in <i>Cγ-exo</i> pucker calculated from DFT.	163
Table 3.2 Cartesian coordinates for zwitterionic Hyp in <i>Cγ-endo</i> pucker calculated from DFT.	164
Table 3.3 Glycyl radical quantification, activity, and kinetic parameters of HypD mutants.	180
Table 3.4 Primers used in site-directed mutagenesis of HypD.....	192
Table 3.5 Data collection statistics for crystallography.....	198
Table 3.6 Model refinement statistics for crystallography. ^{xiii}	198
Table 4.1 UniProt ID and genomic context for sequences used to generate the HypD phylogenetic tree.....	205
Table 4.2 Locations of samples collected to inoculate enrichment cultures.....	213
Table 4.3 Composition of the Hyp enrichment medium.....	214
Table 4.4 Primers used in this study.	216
Table 4.5 Clone libraries generated for 16S rDNA sequencing of enrichment cultures.	217
Table 5.1 Strains and plasmids used in studies on <i>C. difficile</i>	246
Table 5.2 Top 50 upregulated genes in <i>C. difficile</i> 630 Δ <i>erm</i> grown on MACC-Hyp compared to MACC-Pro.....	263
Table 5.3 Top 50 downregulated genes in <i>C. difficile</i> 630 Δ <i>erm</i> grown on MACC-Hyp compared to MACC-Pro.....	267
Table 5.4 Genes of interest and their fold changes in MACC-Hyp / MACC-Pro	272

Table 5.5 Summary of growth of Bacteroidales in M9 and BHIS.....	282
Table 5.6 Strains and plasmids used in studies on Bacteroidales.	286
Table 5.7 Primers used for cloning in this study.	303

List of Abbreviations

aa or AA	amino acid
AE	activating enzyme
α -KG	α -ketoglutarate
Ala, A	alanine
Am	apramycin
Amp	ampicillin
Asn, N	asparagine
5-AV	5-aminovaleric acid
BDE	bond dissociation energy
BHI	brain heart infusion
BME	β -mercaptoethanol
°C	degree Celsius
Cm	chloramphenicol
CoA	coenzyme A
Cys, C	cysteine
DTT	dithiothreitol
EFI-EST	Enzyme Function Initiative-Enzyme Similarity Tool
EPR	electron paramagnetic resonance
[4Fe-4S]	4 iron–4 sulfur cluster
FPLC	fast-protein liquid chromatography
g	gram
GI tract	gastrointestinal tract

GC-MS	gas chromatography-mass spectrometry
Gly, G	glycine
GRE	glycyl radical enzyme
h	hour
H	hydrogen
HMP	Human Microbiome Project
HPLC	high-performance liquid chromatography
Hyp	<i>trans</i> -4-hydroxy-L-proline
HypD	<i>trans</i> -4-hydroxy-L-proline dehydratase
IPTG	isopropyl β -D-1-thiogalactopyranoside
Kan	kanamycin
kb	kilobase
kDa	kilodalton
L	liter
LB	Luria-Bertani or Lysogeny broth
LC-MS/MS	liquid chromatography tandem mass spectrometry
M	molar
min	minute
MW	molecular weight
NAD(P) ⁺	oxidized nicotinamide adenine dinucleotide (phosphate)
NAD(P)H	reduced nicotinamide adenine dinucleotide (phosphate)
NaDT	sodium dithionite
NCBI	National Center for Biotechnology Information

Ni-NTA	nickel-nitrilotriacetic acid
NMR	nuclear magnetic resonance
OD	optical density
P5C	(<i>S</i>)- Δ^1 -pyrroline-5-carboxylate
P5CR	(<i>S</i>)- Δ^1 -pyrroline-5-carboxylate reductase
psi	pounds per square inch
PCR	polymerase chain reaction
PDB	Protein Data Bank
Pro, P	proline
RNA-seq	RNA sequencing
rpm	revolutions per minute
s	sec
SAM	<i>S</i> -adenosylmethionine
SDS-PAGE	sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SD	standard deviation
SSN	sequence similarity network
TCA	tricarboxylic acid
Tm	trimethoprim
Tris	2-amino-2-(hydroxymethyl)propane-1,3-diol
Tyr, Y	tyrosine
UV-Vis	ultraviolet visible
V	volt

Acknowledgements

I would like to first and foremost thank Prof. Emily Balskus, my thesis advisor, for her guidance and mentorship throughout graduate school. I am grateful for the rigorous training I have received in the past four years. Her tirelessness, enthusiasm, and rigor for science have been major sources of inspiration. I deeply appreciate the opportunities I received to attend conferences and present my work. I am especially grateful for her support in attending the Microbial Diversity research course early on during my graduate career. Outside of research, Emily has been extremely generous with her time to provide training and helpful advices on communication skills and career decisions. I have learned so much from Emily and could not have hoped for a better scientific training in graduate school.

The diverse research experiences I gained as an undergraduate were instrumental in my decision to pursue a PhD. From the labs I worked in, countless graduate students and postdocs were role models for the scientist I wanted to become. I especially want to thank Amélie Ménard and Shane Caldwell for serving as my longest mentors in Prof. Karine Auclair's and Prof. Albert Berghuis's labs at McGill University. I always admired how knowledgeable they were, their patience, and their enthusiasm for science.

I would like to thank everyone who served on my dissertation committee: Prof. Suzanne Walker, Prof. Wendy Garrett, Prof. Cathy Drennan, and Prof. Dan Kahne. Their scientific advices, especially toward prioritizing and identifying interesting research directions, have been invaluable. Dan has been an incredible mentor on multiple aspects of my grad school experience. I am grateful for the time he took to listen and give advices during our lengthy conversations. His support and encouragements have meant a lot to me. In addition to being on my thesis committee, Cathy has

been a collaborator on the HypD structural work. I am grateful for all of her expertise and insight into GREs.

Suzanne, Dan, and Jason Millberg have done an exceptional job at directing and managing the Chemical Biology PhD program. I am extremely grateful to have been accepted into the program, my experience in grad school would have been very different without the Chem Bio family. Dan, Suzanne, and Jason have gone out of their way to foster a welcoming and friendly culture in the program. I have made many close friends here and I will always treasure all the fun outings.

I am grateful to have had many scientific mentors within the Balskus lab and in the Boston community. I will try to thank everyone here. All Balskus lab members, past and present, have been amazing colleagues and provided a welcoming environment. Smaranda Bodea first introduced me to anaerobic techniques and GREs during my lab rotation. Ana Martínez-del Campo taught me much of cloning and microbiology, I am grateful to have collaborated with her on the phylogenetics and degenerate primers work. Spencer Peck and Benjamin Levin for all the helpful discussions on protein biochemistry and for their work on quantitative profiling of the GRE superfamily.

To the many collaborators I have had the pleasure of working with, I learned a lot from all of them. Thank you for making collaborative research such a great experience. I am indebted to Yifeng Wei for the conception of this project and for formulating the initial hypothesis on HypD activity. Being part of different labs, Yifeng provided a different perspective on GRE work and the relevant methods, which was immensely helpful during the early months of my work. Lindsey Backman's crystallography work on HypD was critical for the biochemical work and proposed mechanism reported in Chapter 3. I am grateful for everything I learned about structural

elucidation through this collaboration. Dr. Brian Gold for his calculations on Hyp conformations, which was crucial for substrate modelling in the HypD structure. I really enjoyed my conversations with Zhongyue Yang and learning about his computational work on HypD. It is always interesting to learn about a different field and I am excited to see where this work goes. I had the honour of working closely with Prof. Laurie Comstock and Prof. Aimee Shen to carry out genetics in *Bacteroides* species and *C. difficile*, respectively. I am very grateful for their enthusiasm, generosity with their time, technical expertise, and helpful scientific discussions. The bacterial mutants generated from these collaborations are pivotal to work reported in Chapter 5. I am thankful to Michael Luescher for helpful discussions on enzyme mechanisms and being a great resource for chemical knowledge. I am grateful to have mentored and worked with Chanan Sessler, who started in the lab as an undergraduate. Chanan's hard work and curiosity for science were instrumental in making a lot of progress toward identifying collagenases in *C. difficile*.

I would like to thank everyone at Harvard facilities who assisted me with various techniques and instruments. Bill Collins and Shaw Huang for troubleshooting issues with the EPR spectrometer. To those at the Small Molecule Mass Spectrometry Facility: Kelly Chatham for initial Pro detections by LC-MS/MS, Sunia Trauger for gel band proteomics, Jennifer Wang for GC-MS analyses, and Bogdan Budnik for solution proteomics at the Mass Spectrometry and Proteomics Resource Lab.

Attending the Microbial Diversity 2015 course at Marine Biological Laboratory was a transformative experience that shaped my research interests. I am grateful to have Prof. Dianne Newman and Prof. Jared Leadbetter as the course directors. Their enthusiasm for environmental microbiology and microbial metabolism were contagious. I had an amazing time exploring Woods Hole, hunting microbes, and learning from a diverse group of scientists. The exploratory nature of

the course's research component provided a unique and special experience. The side project I started during the course on enrichment culturing led to the work described in Chapter 4.

To all past and present members of bay 3: Stephen, Carolyn, Abe, Ben S., Doug, Paul, Maud, Ben W. You are the best bay mates I could have hoped for. I will miss all the sarcasm, sass, snacks, and music. Thank you for all the exciting scientific discussions, fun times, and the company especially during the late nights!

A major thank you to Ben S. and Lauren for editing my entire thesis and reading all hundreds of pages. Beverly and Ben L. for edits on some sections of the thesis. To my writing buddies: Cristina, Jennifer, and Tiffany, thank you for the company and for keeping me motivated these past few months.

To all the ladies in the Chatham household, Marina, Nienke, and Nastaran, thank you for being there for the past four years and for building a home. I will cherish all the times we spent together.

To the Harvard GSAS Dragon Boat team – this is where I met so many amazing people and made life-time friends. I will always treasure our time on the water, dinners at Ana's Taqueria, and the insane distances we travelled for races. I will miss watching the sun go down as we practice on the Charles and even the stormy days when we got soaked because we were on the boat together as a team. Thank you for being the team where I learned to steer and coach. I have grown and learned so much within this community.

To all the amazing friends I made here – Ana, Abe, Cristina, Guiping, Doug, Nathaniel, Nitzan, Matt, Benika, Jeannie, and everyone. I will always treasure the fun times and adventures we shared. Thank you for the wonderful memories and support throughout these years, especially

during the tough times. Thank you for all your generosity, kindness, guidance, humor, and patience. You are truly the best part of grad school.

Most importantly, I need to thank my mom and dad for raising me to be the person I am today. Words cannot describe how much I look up to you. I am reminded every day of the perseverance and hard work you have put into building a new life for us in Canada. You have provided and opened up so many opportunities for me. I am grateful for the sense of adventure and curiosity about the natural world you have nurtured in me since childhood. You have always encouraged me to ask why and approach problems logically, something I especially appreciated during grad school. Thank you for your unwavering support and understanding throughout the last few years.

1 Chapter 1. Introduction to the human gut microbiome, the glycol radical enzyme superfamily, and 4-hydroxyproline metabolism

Parts of this chapter were adapted from published work.¹

1.1 The gut microbiota influences human physiology

Communities of microorganisms that live within and on our bodies (the human microbiota) carry out metabolic processes that have profound impacts on their surrounding habitats and the host. It is estimated that trillions of bacterial cells make up the human microbiota, resulting in the same order of magnitude of bacterial cells as human cells.² The GI tract is the most densely populated habitat for microbes among human body sites with an estimated 10^{11} bacterial cells per gram.³ Decades worth of research using traditional cultivation methods have observed gut microbial metabolism of metabolites with biological activities, highlighting the potential for how gut microbes may influence human physiology. However, it was only in the past 10 to 20 years with advances in DNA sequencing technologies that extensive efforts have shed light on gut microbiome compositions.^{4,5} These sequencing-based studies have discovered correlations between changes in the gut microbiome composition and diseases such as metabolic disorders, cardiovascular diseases, autoimmune diseases, and neurological disorders.^{6,7} The majority of microbial pathways that might be responsible for these phenotypes have not been identified, due in part to the vast genetic diversity of microbes. Metabolic capabilities encoded within microbes are far more diverse than that of the host. Indeed, it is estimated that microbial genes outnumber human genes by 150-fold in the human body.⁸ Despite significant progress in the field, it is still extremely difficult to infer metabolic activities of the gut microbiome from sequencing data due to our poor understanding of the functions encoded by microbial genes.⁹⁻¹¹ Specifically, the precise

mechanisms by which gut microbial metabolites are produced and impact host physiology are largely unknown. Solving these problems remains a high priority in the field to advance our understanding of basic microbial metabolism and to inform therapeutic development efforts to treat diseases.

The GI tract is also the most anoxic site for microbial growth in the human body and thus is inhabited by a distinct community of facultative and obligate anaerobes. Anaerobic growth of these microbes depends on oxygen-sensitive enzymes that complement and expand the suite of transformations carried out by host enzymes. These distinct gut microbial metabolic pathways interface with host pathways in many ways. Host-microbe interactions include microbial metabolism of dietary components to affect substrate bioavailability for the host, production of essential nutrients, and synthesis of bioactive microbial metabolites linked to human diseases.^{6,12} Examples of these specific interactions will be covered in this section with an emphasis on complex polysaccharide, amino acid, and peptide metabolism by the gut microbiota.

One obvious way in which the host interacts with gut microbes is to provide host-derived substrates from the diet (Figure 1.1). Studies from several decades ago elucidated the important role of gut microbiota in nutrient acquisition by the host, both through the breakdown of host-indigestible dietary components such as complex polysaccharides and the biosynthesis of essential metabolites such as vitamin K.^{6,13} Most of these transformations were first observed in *ex vivo* incubations and studies using germ-free animals, and have not been linked to individual microbial isolates, enzymes, or genes. However, recent efforts to study gut microbial metabolism of certain dietary substrates have elucidated the genes and enzymes involved in these transformations as well as their importance to human health.

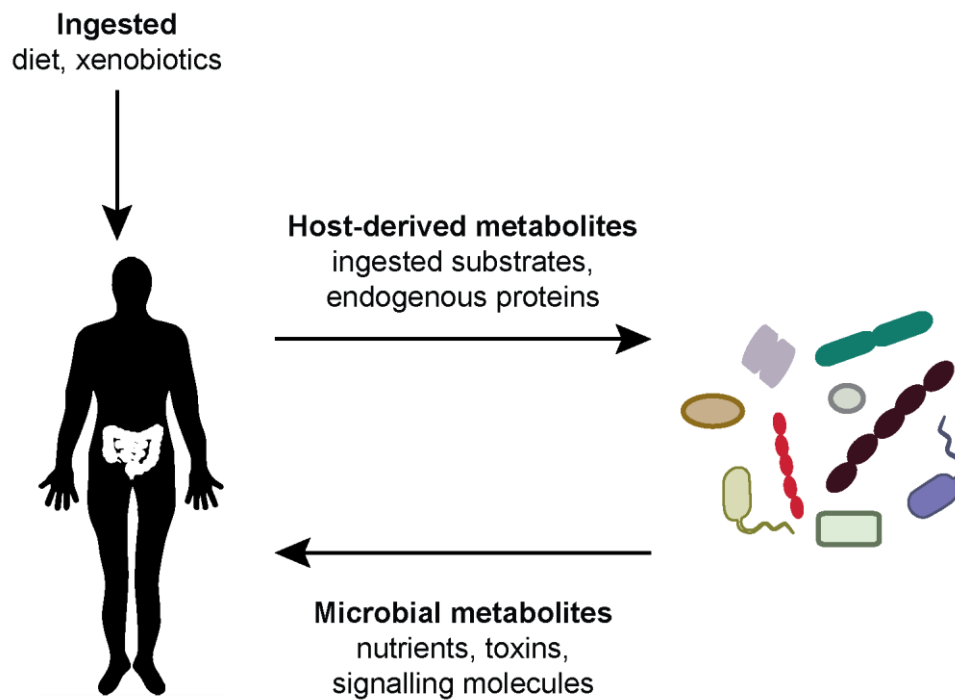


Figure 1.1 Metabolic crosstalk between host and gut microbes.

Perhaps the most well-known example of the gut microbiota complementing host metabolism is the breakdown of complex dietary polysaccharides that cannot be digested by the human body. The fermentation end products of these carbohydrates include short-chain fatty acids (SCFAs), which accumulate to high levels in the gut and influence a wide range of host processes.¹³ SCFAs have been appreciated to be a major source of energy for colonocytes in the GI tract and are estimated to account for up to 10% of daily caloric requirements in humans.¹⁴ In addition to dietary polysaccharides, mucins are an additional source of carbohydrates for gut microbes.¹⁵ These host proteins make up a major component of the mucosal layer in the GI tract and are heavily glycosylated, thereby providing diverse glycans that influence microbial colonization and composition in this habitat.¹⁵ To date, host-microbe co-metabolism of complex polysaccharides is

one of the most extensively studied area in the gut microbiome field. Yet only a small subset of enzymes that cleave glycosidic linkages to release monosaccharides from complex carbohydrates are currently known.

Proteins make up another major source of nutrients for microbial growth that originate from the diet and the human body. For the host, dietary proteins provide essential amino acids, which are not synthesized by human enzymes. The gut microbiota can also provide these amino acids through *de novo* synthesis or hydrolysis of proteins and peptides. Furthermore, microbial amino acid metabolism can also lead to the production of molecules known to have biological activities in host pathways. In particular, microbial metabolism of aromatic amino acids can lead to the production of immunomodulatory molecules, toxins, and neurotransmitters such as indolepropionic acid, tryptamine, *p*-cresol, indoxyl sulfate, and histamine.¹⁶⁻¹⁹ Using defined microbial consortia to colonize germ-free mice, Fischbach group has linked microbial production of some metabolites to host phenotypes.^{16,19} More recently, elucidating biological consequences that gut microbes have on the nervous system has become the focus of several research groups.^{20,21} It has become clear that the gut-brain axis is a true phenomenon and microbial metabolism of amino acids likely plays an important role, but the underlying molecular mechanisms remain to be determined.

In addition to the biological implications of amino acid metabolism, an important, but underappreciated, function of the gut microbiota may be the degradation of short peptides not susceptible to hydrolysis. Metabolomics studies on germ-free or antibiotic-treated rodents have detected differences in peptide levels in fecal and urine samples to support a role for gut microbes in peptide metabolism.²²⁻²⁴ Ingestion of proteins by human volunteers resulted in the accumulation of specific peptides to indicate that certain amino acid sequences make peptides more recalcitrant

to host peptidase activities (see section 1.7 for details).²⁵⁻²⁷ Overall, it is likely that peptide metabolism is yet another route through which microbes can alter the bioavailability of both peptides and its amino acid building blocks to the host. Intriguingly, many of these non-hydrolysable peptides have been found to have biological activities in human cell lines.²⁶ Release of free amino acids from these peptides by gut microbes would provide additional nutrients to the host and microbes, leading to the production of bioactive microbial metabolites. Therefore, metabolism of indigestible peptides could mirror that of indigestible polysaccharides involving both host and microbial enzymes.

The extensive metabolic crosstalk between host and its microbiome illuminated in the past decade has unveiled the importance of the gut microbiota in host biology. However, little is known about the physiological functions of most microbial genes, which illustrates a great need to investigate this uncharacterized genetic information.^{9,28} Therefore in order to understand the metabolic potential of the gut microbiome and inform mechanistic research, it is imperative that we link these genes to functions. One approach is to focus efforts toward characterization of chemical activities that are uniquely microbial and on functions relevant to the metabolism of an abundant substrate.

1.2 Glycyl radical enzymes are abundant in the gut microbiome

Glycyl radical enzymes (GREs) make up an evolutionarily ancient protein superfamily that is essential for anaerobic primary metabolism in bacteria and archaea.^{29,30} These enzymes utilize protein-centered radicals to accomplish chemically challenging transformations and are extremely oxygen-sensitive. Therefore, GREs are only found in facultative or obligate anaerobic microorganisms. Previous metagenomic and metaproteomic studies have detected GREs as one of the most abundant microbial protein families in the human gut.³¹⁻³³ Furthermore, certain

characterized GRE activities have connections to human health (Figure 1.2). For example, metabolism of the essential nutrient choline by choline trimethylamine-lyase (CutC), a GRE, produces trimethylamine (TMA),^{34,35} a metabolite that is positively correlated with incidences of heart and liver diseases^{36,37}. *p*-Hydroxyphenylacetate decarboxylase (HPAD) catalyzes the decarboxylation of *p*-hydroxyphenylacetate to yield *p*-cresol,³⁸ which can compete with host metabolism of drugs³⁹. This metabolite is found at elevated levels children with autism although the mechanism of action remains unknown.^{18,39} In addition, *p*-cresol has antibacterial activity against some gut microbes and is thought to mediate microbial competition within the community.⁴⁰ Despite the importance of these GREs in producing metabolites with biological activities, their distributions in human microbiomes remain largely unknown. Identification of specific GRE functions in sequencing datasets is difficult due to the unusually high amino acid sequence similarities within this superfamily.

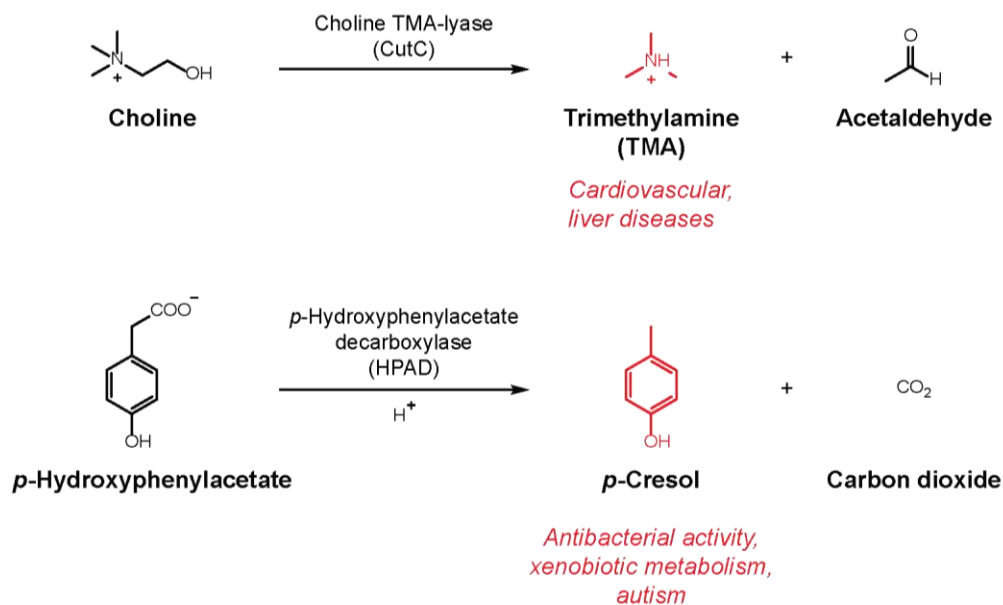


Figure 1.2 GRE-catalyzed transformations that produce metabolites linked to human health.

Work from our lab addressing some of these problems is described in detail in Chapter 2. Briefly, new bioinformatic tools have allowed us to differentiate sequences that likely encode distinct enzyme activities using a prior understanding of GRE biochemistry. This enabled us to profile the abundance of members of the GRE superfamily, both characterized and uncharacterized, in microbial metagenome datasets collected by the Human Microbiome Project (HMP) from healthy individuals. In the future, comparison of GRE abundances between healthy and disease cohorts is expected to help generate hypotheses regarding the contribution of GRE-mediated metabolic activities toward disease progression in the host.

1.3 Mechanistic overview of glycy radical enzymes

1.3.1 The general catalytic cycle of GREs

GREs catalyze chemically diverse and challenging transformations in microbial pathways using protein-based radical chemistry. All GREs must be post-translationally activated by a partner activating enzyme (AE) that installs a radical on the α -carbon of an invariant glycine residue on the GRE's C-terminal loop. AEs are radical *S*-adenosylmethionine (SAM) enzymes that contain at least one [4Fe-4S] cluster binding motif, CxxxCxxC, containing three cysteine residues to coordinate the cluster (Figure 1.3Figure 1.1). SAM cleavage catalyzed by AEs is presumed to follow the consensus mechanism determined for the radical SAM enzyme superfamily.⁴¹ First, the [4Fe-4S] cluster binds SAM by chelating its amino and carboxyl moieties to the unique Fe not coordinated by a cysteine residue. The reduced [4Fe-4S] cluster, in its +1 oxidation state, is thought to subsequently donate a single electron to SAM to induce its cleavage into L-methionine and a 5'-deoxyadenosyl (5'-dA) radical (Figure 1.3A). Although the 5'-dA radical product, a highly reactive intermediate, was proposed to directly abstract a hydrogen atom from the GRE, recent work provided evidence for the involvement of an organometallic intermediate (discussed in Section 1.3.2). The hydrogen atom abstraction from a conserved glycine residue on the GRE then generates the glycine-centered radical cofactor (Figure 1.3B).

The glycy radical is located within the GRE active site and is stabilized through the captodative effect, having an adjacent electron withdrawing (carbonyl) and electron donating (amide) group that allow delocalization of the unpaired electron along the peptide backbone. It is proposed that this radical is then transferred onto an invariant cysteine residue on the GRE through a hydrogen atom abstraction. The thiyl radical typically initiates catalysis by generating a substrate-based radical, which then undergoes rearrangement to form a product-based radical. Through a final hydrogen atom abstraction from the thiol of cysteine, the final product is formed and the thiyl radical is regenerated for the next catalytic cycle. In the absence of substrate binding,

the radical is thought to be stored on the glycine residue due to the higher stability of glycy radical compared to thiyl radicals. The glycy radical can be detected and quantified using electron paramagnetic resonance (EPR) spectroscopy through a characteristic signal from the hyperfine splitting of the unpaired electron by the adjacent proton nucleus. Experimentally, activated PFL was observed to maintain full catalytic activity for more than 7 hours at 25 °C, illustrating the high stability of the glycy radical.⁴² The regeneration of a stable protein-centered radical species at the end of each catalytic cycle allows GREs to undergo multiple turnovers without synthesis of radical cofactors.

Prior to the thesis work presented here, the maximum activation observed for any GRE was with PFL where up to 0.5 equivalents of glycy radical relative to protein monomer was detected by EPR.⁴³ Based on this observation and the dimeric nature of most GREs, these enzymes may exhibit half-site reactivity where the activation of one monomer precludes activation of the other monomer due to a conformational change induced by glycy radical formation. It is also possible that the incomplete GRE activation observed *in vitro* is a result of assay conditions that are less optimal than the intracellular environment. In addition, the physiological electron donors that reduce AE's [4Fe-4S] clusters *in vivo* are not known. Many questions remain to be addressed on the subject of GRE activation, including the stoichiometry of glycy radical formation by AE, protein-protein interactions between GRE and AE, radical stability, and regulation.

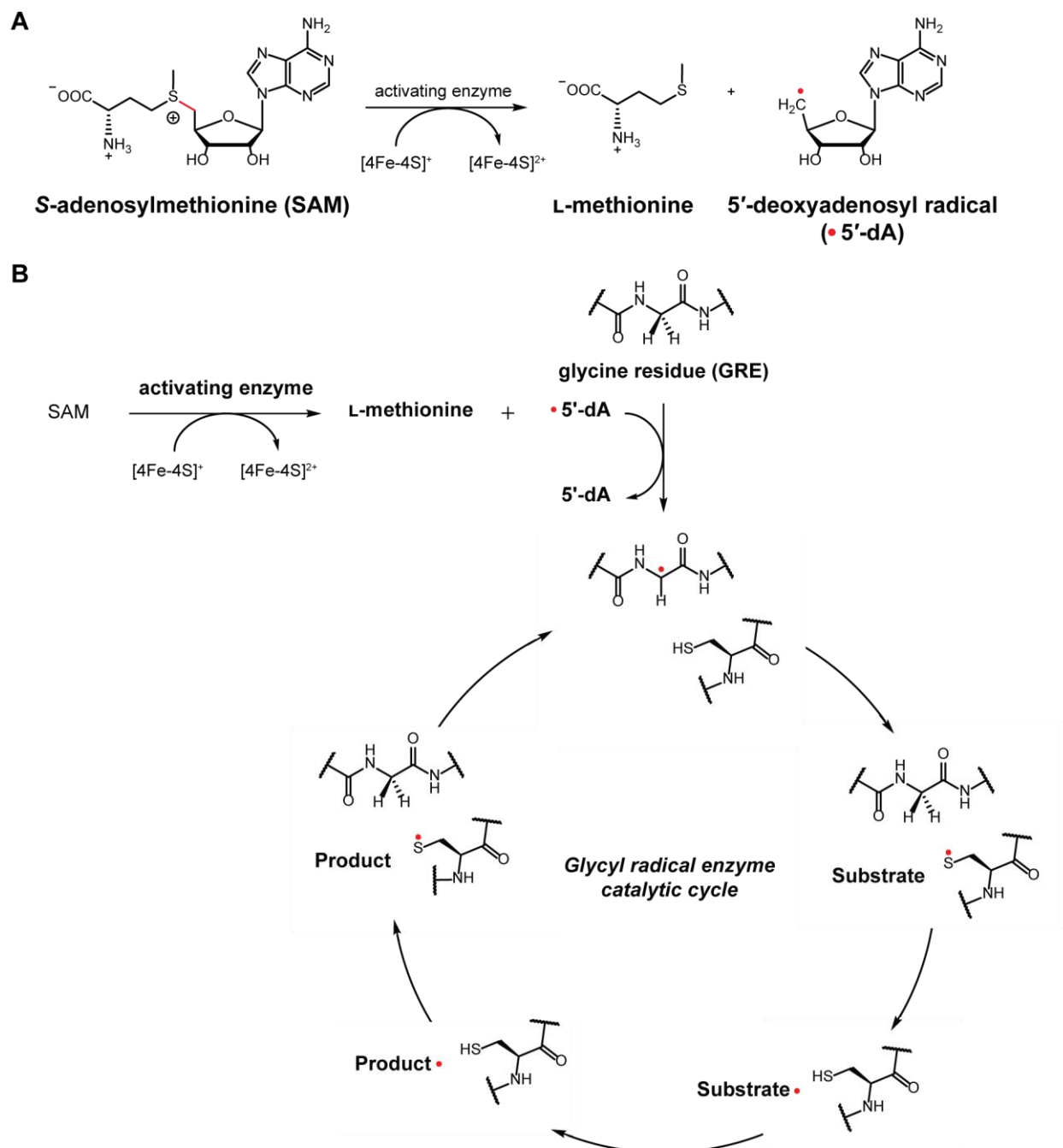


Figure 1.3 Catalytic cycle of a glycy radical enzyme.

(A) The reduced $[4Fe-4S]$ cluster on the AE transfers a single electron to SAM, causing reductive cleavage of SAM to generate a 5'-deoxyadenosyl radical ($\bullet$ 5'-dA), which is essential for GRE activation.

Figure 1.3 (Continued)

(B) GRE becomes catalytically active upon installation of a glycine-centered radical by the 5'-deoxyadenosyl radical through a hydrogen atom abstraction. Upon substrate binding, it is proposed that this radical is then transferred onto a conserved cysteine residue to generate a more reactive thiyl radical. This thiyl radical initiates catalysis with substrate to generate a substrate-based radical. Additional active site residues catalyze the rearrangement of substrate to product. The cysteine thiyl radical is regenerated on the conserved cysteine as the final step of product formation.

1.3.2 Recent mechanistic work on GRE activating enzymes

Although the 5'-dA radical intermediate is thought to be a product of one-electron reduction of SAM catalyzed by radical SAM enzymes, this highly reactive species has not been directly detected. In 2016, Broderick group reported efforts to capture this radical intermediate using rapid freeze-quench EPR and electron nuclear double-resonance (ENDOR) spectroscopies during PFL activation by its AE.⁴⁴ Unexpectedly, initial EPR signals of a transient radical intermediate did not match an organic radical presumed for 5'-dA radical formation or for such a radical coupled to a metal-ion center like [4Fe-4S] cluster. Additional ENDOR spectroscopic experiments with heavy isotope ⁵⁷Fe and ¹³C-enriched PFL-AE and SAM, respectively, established the existence of an organometallic intermediate Ω containing a covalent Fe–C bond. In a recent follow-up paper, work from the Broderick group demonstrated that intermediate Ω specifically contains a covalent linkage between a cluster Fe and [5'-C] of the 5'-dA fragment while the methionine fragment remains coordinated to the unique Fe post-cleavage.⁴⁵ This metal–carbon bond could then undergo

homolytic cleavage to generate a 5'-dA radical for GRE activation through a hydrogen atom abstraction. This study also showed that Ω is a central intermediate produced by several radical SAM enzymes covering a wide range of reactions and substrate scope. Interestingly, a similar study on a distinct class of radical SAM enzymes that cleave a different C-S bond in SAM to generate a 3-amino-3-carboxypropyl (ACP) radical also uncovered an organometallic intermediate.⁴⁶ This intermediate was found to contain an Fe-C bond between the unique iron of the [4Fe-4S] cluster and C γ of ACP analogous to intermediate Ω . Together, these recent publications encompass landmark studies in the discovery of novel bioorganometallic complexes critical to the diverse array of reactions catalyzed by the largest known enzyme superfamily.

Among radical enzyme superfamilies, the formation of Ω by radical SAM enzymes parallels the formation of adenosylcobalamin in coenzyme B₁₂-dependent radical enzymes. Both enzyme families form bioorganometallic complexes and then cleave the metal-carbon bond to generate a reactive 5'-dA radical to initiate catalysis. In B₁₂-dependent enzymes, the adenosylcobalamin cofactor contains a covalent bond between [5'-C] of 5'-dA and the Co metal center of cobalamin (coenzyme B₁₂). Yet another parallel between radical SAM enzymes and B₁₂-dependent enzymes was discovered recently by the Broderick group. In PFL-AE, a binding site for a monovalent cation was identified in the active site in proximity to the [4Fe-4S] cluster and SAM.⁴⁷ The presence of a cation in purified protein solutions resulted in greater activity and stability of AE. A preference for K⁺ was observed for PFL-AE and was thought to orchestrate SAM binding by interacting with its carboxyl moiety. This ion could also affect the reactivity of the previously described organometallic intermediate Ω in radical SAM enzymes and product release. Similarly, a K⁺ ion also binds in the B₁₂-dependent diol dehydratase active site and was shown to catalyze the homolytic cleavage of adenosylcobalamin to generate 5'-dA radical.^{48,49}

1.4 Shared structural features of glycyl radical enzymes

Based on size-exclusion chromatography and X-ray crystallography studies, most GREs appear to be homodimers, including PFL, class III RNR, GD, and CutC. HPAD and BSS are two exceptions with both composed of multiple subunits with a large subunit containing the canonical GRE domain and small subunits containing iron-sulfur cluster binding motifs.

HPAD crystallized as a $(\beta\gamma)_4$ heterotetramer with a large β subunit with the canonical GRE fold and the catalytic residues while the small γ subunit contains two $[4\text{Fe-4S}]$ clusters.⁵⁰ The functions of these additional clusters have not been established and their presence represents a new protein architecture among GREs. BSS makes up the other structurally characterized GRE to contain additional $[4\text{Fe-4S}]$ cluster-binding subunits.⁵¹ Two genes essential for activity are located adjacent to the GRE gene in the BSS-encoding operon and encode 7 and 9 kDa small subunit proteins.⁵¹ BSS forms an $(\alpha\beta\gamma)_2$ heterohexamer with the large α subunit containing the glycyl radical domain.⁵¹ The β and γ subunits each contain a $[4\text{Fe-4S}]$ cluster-binding motif and are similar in structure to the HPAD γ subunit.⁵¹ The BSS β subunit is positioned next to a putative substrate channel where it could play a role in substrate binding.⁵² Additionally, the cluster in BSS β is positioned near the glycyl radical domain, suggesting a possible role in glycyl radical installation.⁵¹

Structural elucidation of several GREs by X-ray crystallography has illustrated similar features conserved within the protein superfamily. Prior to the work described in this thesis, the structures of PFL⁵³, class III RNR⁵⁴, CutC⁵⁵, BSS⁵¹, HPAD⁵⁰, and GD⁵⁶ had been reported. GREs share high amino acid sequence similarities which is reflected in the common barrel architecture observed thus far in crystal structures. The canonical GRE fold consists of a ten-stranded α/β -barrel: antiparallel five-stranded sheets make up two halves of the barrel surrounded by α -helices

on the outside (Figure 1.4A). This barrel architecture surrounds a buried active site where catalytic residues glycine and cysteine (Figure 1.4B). These two residues are located in distant regions of the amino acid sequence, but are positioned in close proximity in the protein fold for radical relay to occur. The conserved glycine residue is located within the C-terminal glycy radical domain and part of the Gly loop with a consensus sequence motif of RVXG[Y/F/W].²⁹ The cysteine residue is poised in the active site adjacent to the hydrogen atom to be abstracted from the substrate.

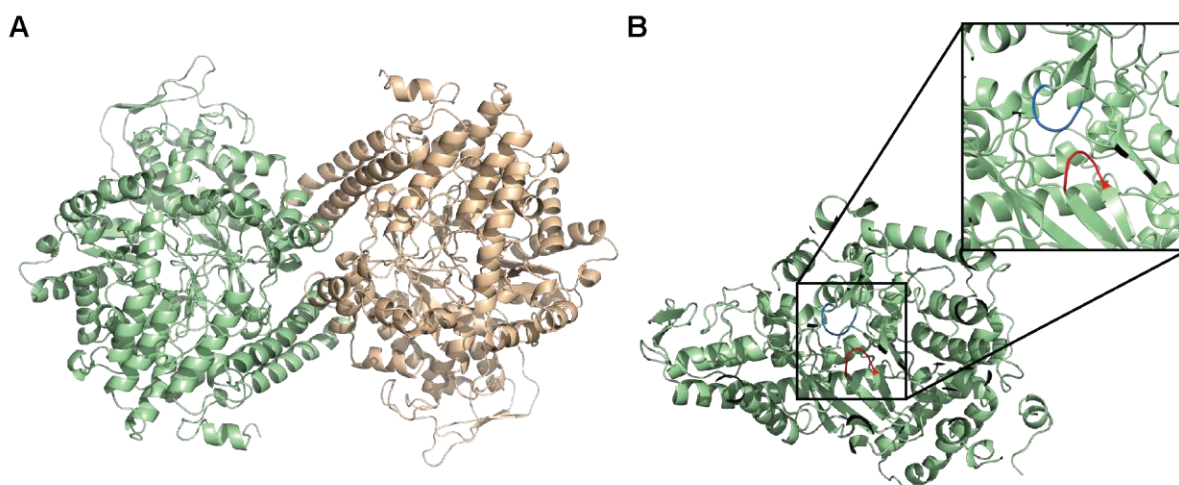


Figure 1.4 Structural features shared among glycy radical enzymes.

(A) Most GREs contain a single glycy radical domain and crystallize with extensive interactions between two monomers to support a physiological dimeric state. GREs share a 10-stranded α/β -barrel structure that surround a buried active site. (B) Two loops are positioned in close proximity in the active site encoding the essential residues glycine (blue) and cysteine (red) for radical transfer. PFL from *E. coli* is depicted here as an example (PDB: 2PFL).

The deeply buried active site in GREs is thought to be a protective mechanism to prevent radical species from being quenched by the solvent, similar to other radical enzymes such as

adenosylcobalamin or radical SAM enzymes.^{57,58} The buried Gly loop is predicted to undergo a drastic conformational change into an open state in order for the AE to access the Gly residue for GRE activation. Evidence for this comes from the co-crystallization of the PFL-AE with a small peptide mimic (7 amino acids) of the PFL Gly loop, which remains to be the only structure elucidated for AEs.⁵⁹ As expected, the glycine residue in the peptide bound adjacent (4.1 Å) to the SAM cofactor that suggests a large conformational change would be required for the Gly loop to swing outward from the buried GRE active site. Another study using circular dichroism and EPR concluded that the presence of PFL-AE increased the population of PFL in the open conformation as demonstrated by increased quenching of the glycyl radical by external reductants.⁶⁰

All AEs are members of the radical SAM enzyme superfamily and encode single radical SAM cluster binding motif.⁵⁹ With the exception of PFL-, RNR-, and PD-AEs, most characterized AEs contain a ferredoxin-like domain that encodes one to two CX₂₋₅CX₂₋₄CX₃C motifs predicted to coordinate auxiliary [4Fe-4S] clusters (Figure 1.5).^{41,61} Martins and coworkers showed that the radical SAM domain of the HPAD-AE was sufficient to activate HPAD using a HPAD-AE mutant with its ferredoxin-like domain truncated.⁶² Intriguingly, activation of HPAD by its wild-type AE resulted in higher levels and longer stability of glycyl radicals compared to the ferredoxin-domain mutant.⁶² This data potentially implicates the auxiliary clusters in electron transfer between reductants and the radical SAM cluster and/or protection of the Gly loop from radical quenching during conformational changes in HPAD. But the molecular details of their functions remain to be characterized. To date, the auxiliary clusters have only been studied in HPAD-AE, therefore it is unclear how conserved their functions may be. Future structural and biochemical studies of other AEs will likely elucidate similar or novel interactions important for GRE activation. In addition, a

structure of a GRE in complex with its AE has not been obtained. Crystallization of this protein complex will provide crucial information on the activation process of GREs.

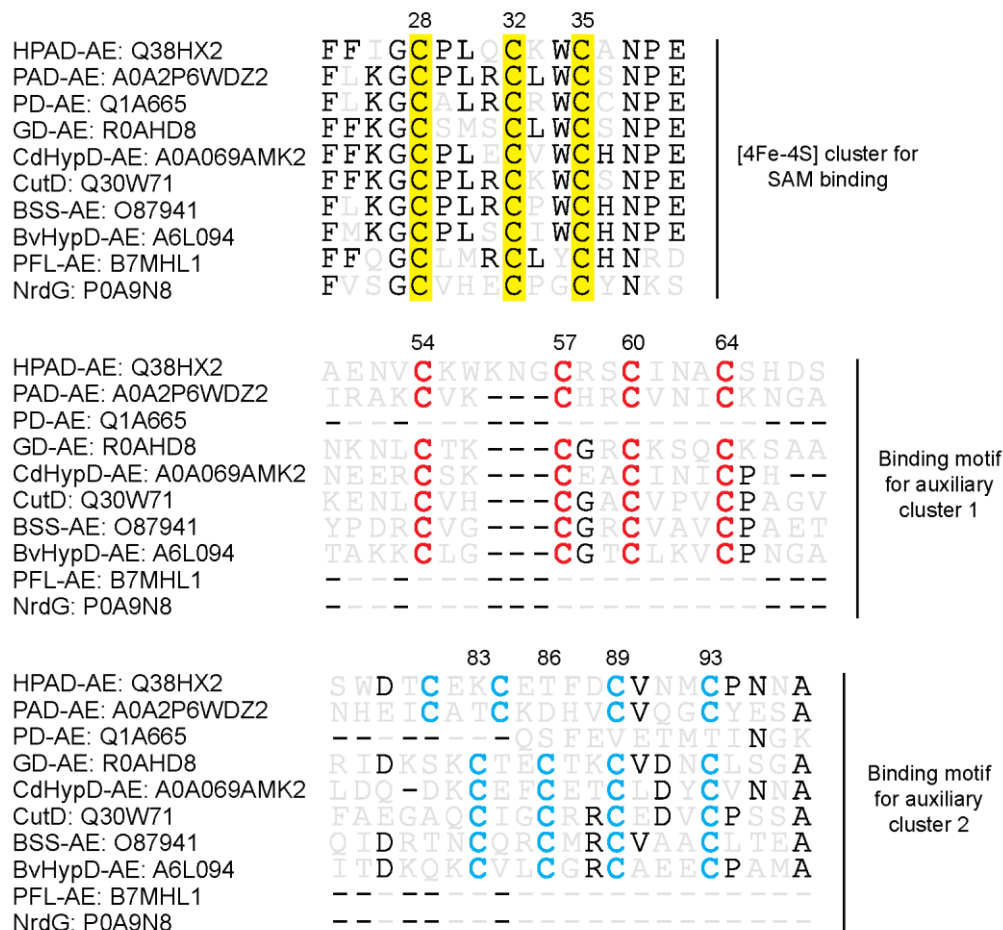


Figure 1.5 A multiple sequence alignment of Fe–S cluster binding motifs present in activating enzymes.

A radical SAM cluster binding motif is located near the N-terminus and is highlighted in yellow. Conserved cysteines part of a putative auxiliary cluster binding motif are highlighted in red. Conserved cysteines making up a second putative binding motif are highlighted in blue. Two HypD-AEs were included from different phyla: Cd – *Clostridioides difficile*, Bv – *Bacteroides vulgatus*. UniProt IDs of sequences used are displayed next to the protein name. Residue numbering is based on CdHypD-AE. Amino acid sequences were aligned using ClustalW⁶³ in Geneious 9.0.4⁶⁴.

1.5 Glycyl radical enzymes catalyze key steps in primary metabolic pathways

The GRE superfamily participates in evolutionarily ancient, anaerobic primary metabolism.⁶⁵ With both GREs and AEs deactivated by molecular oxygen, these proteins are only encoded in bacteria and archaea inhabiting anoxic environments. The choice of a glycine residue, the simplest amino acid, for radical storage is considered to be an evolutionarily ancient strategy that negates the need for metabolically expensive cofactors used by other radical enzyme families.²⁹ The two best-studied GREs are essential for life in the anoxic environment and are likely the most ancient members of this protein superfamily. *De novo* DNA synthesis from RNA is catalyzed by class III RNR, a GRE proposed to be the most primitive class of RNRs that predate oxygenation of the Earth.^{65,66} Other classes of RNRs (I and II) likely arose through divergent evolution to adapt to life in the presence of oxygen. In addition to class III RNR, PFL is also thought to be an ancient GRE as a key enzyme in glucose fermentation that produces acetyl-CoA.⁶⁷ Interestingly, RNR and PFL have really low sequence similarity (*E. coli*: 13% coverage, 24% ID) to each other and likely have undergone divergent evolution over a long period of time. Currently, all characterized GREs have high sequence similarities to PFL, leading to widespread misannotations of uncharacterized GREs as PFLs in sequence datasets.⁶⁸

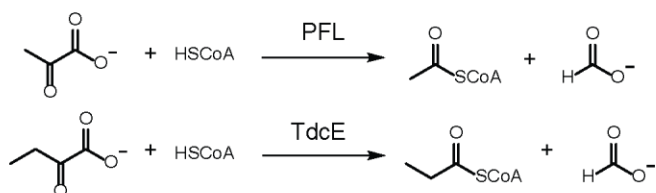
Using radical chemistry, GREs can access chemically challenging and diverse reactions central to many primary metabolic pathways. The superfamily can be divided into five broad classes based on biochemical activities: formate-lyases, ribonucleotide reductases, decarboxylases, X-succinate synthases, and eliminases (Figure 1.6).³⁰ Altogether, known GRE chemistry encompasses a wide range of rearrangements: C–C bond cleavage, C–C bond formation, C–O bond cleavage, and C–N bond cleavage.³⁰ Each class is described in detail in the next few sections (1.5.1–1.5.5). Additional mechanistic details are provided for reported GRE eliminases to provide

an overview of proteins with similar activities as the recently discovered *trans*-4-hydroxy-L-proline (Hyp) dehydratase (HypD). Altogether, well-characterized GREs participate in microbial pathways for glucose fermentation, DNA synthesis, and the metabolism of toluene, tyrosine, phenylalanine, tryptophan, choline, glycerol, and 1,2-propanediol.^{29,30} With the majority of GRE domain-containing proteins lacking a functional annotation, many new reaction types are likely to be discovered in the future.⁶⁸ In addition, a group of unusually short GREs of around 500 amino acids or less form a separate phylogenetic clade and have no known function.⁶⁹ Given the role of GREs in primary metabolism, new GREs likely have important roles in many key human-microbe interactions.

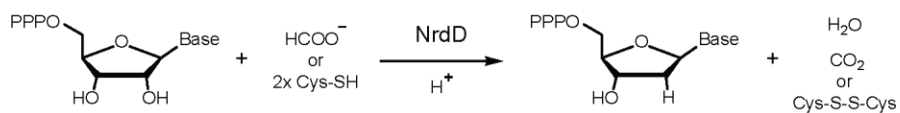
Figure 1.6 Five classes of known GREs categorized by activity.

PFL – pyruvate formate-lyase; TdcE – ketobutyrate formate-lyase; NRD – class III ribonucleotide reductase; HPAD – 4-hydroxyphenylacetate decarboxylase; PAD – phenylacetate decarboxylase; IAD – indoleacetate decarboxylase; BSS – benzylsuccinate synthase; XSS – X-succinate synthase; GD – glycerol dehydratase; CutC – choline trimethylamine-lyase; PD – 1,2-propanediol dehydratase; HypD – *trans*-4-hydroxy-L-proline dehydratase, the focus of this thesis work.

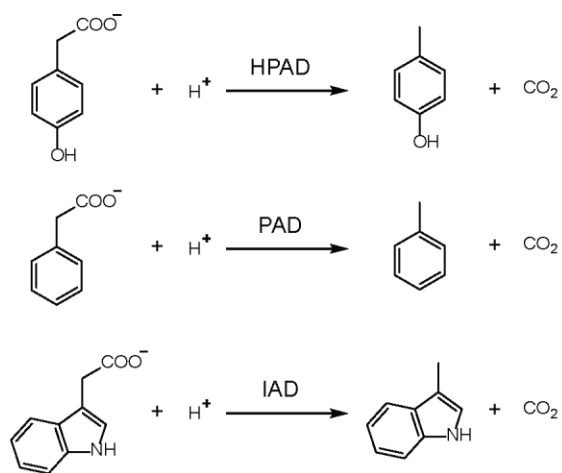
Formate-lyases



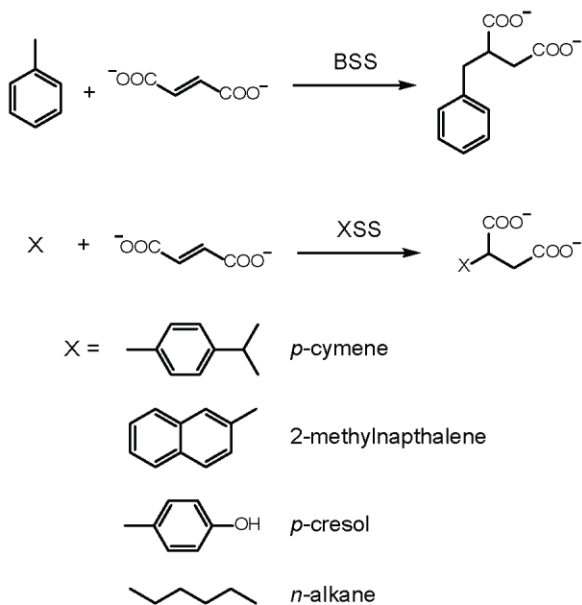
Ribonucleotide reductase



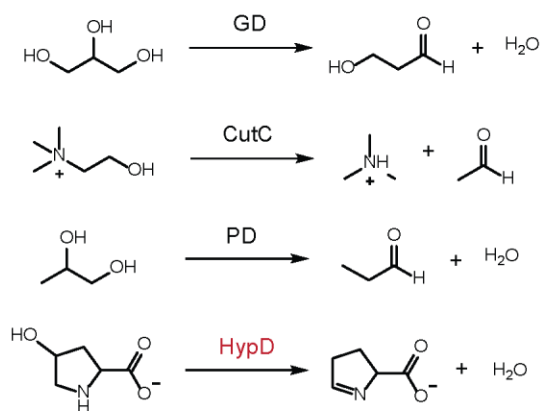
Decarboxylases



X-succinate synthases



Eliminases



1.5.1 Formate-lyases: PFL, TdcE

The formate-lyase class consists of GREs that catalyze C–C bond cleavage of keto acids to generate formate in a metabolic transformation part of carbohydrate and amino acid fermentation.^{29,30} PFL is one of the best-characterized GREs and is the prototype of this class. PFL catalyzes cleavage of pyruvate with co-substrate CoA to acetyl-CoA and formate (Figure 1.6). Acetyl-CoA can be used to generate ATP through substrate-level phosphorylation and can also provide carbon for anabolic pathways.⁷⁰ The *E. coli* PFL was the first GRE detected to carry an organic radical by EPR spectroscopy in pioneering work carried out by the Knappe group in 1984.⁷¹ Oxidative cleavage upon exposure to molecular oxygen resulted in fragmentation at the glycyl radical to release a C-terminal peptide.⁷² Mass spectrometry elucidation of this peptide led to identification of the Gly734 as the site for radical installation.⁷² Crystallographic structures obtained later revealed two conserved Cys residues poised adjacent to Gly734 to suggest the possibility of a radical relay between these three residues.^{53,73} In a co-crystal structure, pyruvate bound in the active site is positioned next to one of the Cys residues to support an initial hydrogen atom abstraction from a thiyl radical.⁷⁴

Due to PFL's essential role in energy production from glucose, it is highly expressed in the cell under anaerobic conditions and microbes have evolved a mechanism to preserve its activity and reduce costs associated with *de novo* protein synthesis.⁷⁵ Some facultative anaerobes encode small ~14 kDa proteins containing a C-terminal region with high sequence similarity to the glycyl radical domain including conserved sequences around Gly734 of PFL.⁷⁶ Knappe and coworkers provided the first biochemical evidence that these small proteins from *E. coli* and bacteriophage T4, annotated as YfiD and Y06I, respectively, function as autonomous glycyl radical cofactors to repair oxidatively cleaved PFLs.⁷⁶ YfiD binds the N-terminal fragment of PFL to form a protein

complex that can be activated by PFL-AE and has full catalytic activity toward pyruvate.⁷⁶ Under microaerophilic conditions, the expression of YfiD is upregulated in *E. coli*,⁷⁷ supporting its role as a repair mechanism for oxygenolytic damage of PFL. Currently, PFL is the only known GRE to interact with a separate glycyl radical domain containing protein acting as a ‘spare part’. The precise mechanism by which YfiD restores activity in PFL and the existence of repair mechanisms for other GREs remain unclear and are intriguing topics to pursue for the GRE superfamily.

Initially, an alternative mechanism to cope with GREs’ oxygen sensitivity was proposed to involve the quenching of the glycyl radical on PFL by a deactivase, AdhE, in *E. coli*.⁷⁸ A more recent study by the Broderick group, however, reported that the reduction in PFL activity was non-enzymatic and independent of AdhE as the effect was recapitulated by the addition of presumed cofactors NAD⁺, Fe(II), and CoA.⁴² They also demonstrated that small molecule reductants 2-mercaptoethanol and dithiothreitol could reversibly quench the glycyl radical on PFL *in vitro*.⁴² This could potentially be a mechanism used by cells to salvage activated GREs to reduce the need for *de novo* protein synthesis upon transient oxygen exposure, although experimental evidence is required to support this hypothesis.

The second known member of the formate-lyases, 2-ketobutyrate formate-lyase (TdcE), was discovered serendipitously in a complementation screen of an *E. coli pfl* mutant looking for rescue of growth defects on glucose.⁷⁰ TdcE has 82% amino acid identity to PFL in *E. coli* and is encoded in the *tdc* operon responsible for threonine (Thr) fermentation.⁷⁰ Intriguingly, unlike most GREs, no AE gene was identified adjacent to the TdcE gene in the genome.⁷⁹ Sawyers and coworkers showed that activation of TdcE is dependent on PFL-AE *in vivo*, providing the first example of an AE capable of converting two distinct GREs into their radical-bearing forms.⁷⁹ In crude protein extracts, TdcE accepted both pyruvate and 2-ketobutyrate as substrates equally

whereas PFL was selective for pyruvate.⁷⁹ *In vivo*, transcription of the *tdc* operon is repressed by glucose, and thus TdcE is likely to be physiologically important specifically for Thr fermentation. Discovery of additional formate-lyases may be impeded by the extremely high sequence similarities among members of this class. Identification of genomes encoding multiple copies of highly similar ‘PFLs’ and examination of their genomic neighborhoods, however, could help to disentangle substrate preferences among new keto acid formate-lyases.

1.5.2 Class III ribonucleotide reductases: NrdD

Class III ribonucleotide reductases (RNRs) are divergent from other GREs, which is reflected in their biochemical mechanism and structure.⁶⁸ Like most GREs, class III RNRs are homodimers, but contain allosteric sites that bind to various dNTPs and NTP effectors and a Zn-binding motif.⁸⁰ Class III RNRs catalyze dehydration and reduction of ribonucleotides to deoxyribonucleotides whereas GRE eliminases catalyze the elimination of a functional moiety from the substrate without a net change in the oxidation state. Overall, these distinctions in mechanism and activity warrant a separate class for NrdDs from the GRE eliminases.

RNRs catalyze the transformation of all four ribonucleotides (A, G, C, U) to 2'-deoxynucleotides to generate building blocks of DNA.⁸¹ This reductive transformation is coupled to the oxidation of either formate or a pair of active site cysteine thiols. RNRs catalyze the only known pathway for *de novo* dNTP biosynthesis and are strictly encoded in all organisms as well as some viruses.⁸¹ All RNRs catalyze the nucleotide reduction using radical-dependent redox chemistry via a Cys thiyl radical to abstract an hydrogen atom from the 3' position of the ribose substrate.⁸² RNRs are divided into three classes based on differences in metallo-cofactor requirements for the generation of an active site thiyl radical.^{81,82} Class I RNRs require molecular oxygen and thus are only found in aerobic organisms. These RNRs contain dimetallo-cofactors

made of different combinations of Fe and Mn and O₂. Class II RNRs rely on adenosylcobalamin to directly generate a thiyl radical in the active site. These enzymes are not O₂-dependent and thus sensitive to deactivation by oxygen and are found in both aerobic and anaerobic bacteria and archaea. Class III RNR (also termed NrdDs) is a GRE and thus is only present in facultative or strict anaerobes. Notably, all three classes of RNRs share the ten-stranded α/β barrel fold common to GREs with class I and II being the only non-GREs known to share this protein fold.³⁰ This supports the proposal that all RNRs evolved from a common ancestor that is more GRE-like and diversified over time to cope with the accumulation of atmospheric oxygen on Earth due to appearance of oxygenic photosynthesis. To deal with the diverse environments microbes encounter, multiple classes of RNRs are often encoded within a single species for differential expression and utilization.⁸³

Mechanistically, ribonucleotide reduction can be divided into two half reactions where the first half reaction consists of dehydration and the second half reaction being a reduction.⁸² The steps used to catalyze the first half reaction are conserved among all RNR classes. An active site Cys thiyl radical initiates catalysis to generate a 3'-nucleotide α -hydroxyalkyl radical. Subsequent elimination of the adjacent 2'-hydroxyl group as a water molecule is facilitated by deprotonation of the 3'-hydroxyl group to form a ketyl radical. These first steps are reminiscent of the proposed mechanism for GRE eliminases (PD, GD, and CutC). However, rather than a final hydrogen atom abstraction from the Cys residue to regenerate the catalytic thiyl radical as proposed for GRE eliminases, RNRs utilize various reducing equivalents during the second half reaction to generate the 3'-deoxynucleotide radical. In this step, an electron is gained from an active site thiyl or thiosulfuranyl radical through a proton coupled electron transfer with a reductant.⁸⁴ Finally, the

hydrogen atom that was initially abstracted from the 3'-position of the nucleotide is returned to the same carbon in the product.

Contrary to initial findings, class III RNRs appear to be mechanistically diverse with respect to the reduction step, which has been grouped into three subclasses.⁸⁴⁻⁸⁷ For many years, formate was thought to be the only reductant used by this class during the second half reaction.⁸⁸ Another widely encoded GRE, PFL, is the principle source of formate under fermentation conditions during microbial growth in some species.^{75,89} The observation that many species encoding class III RNRs do not possess a PFL or enzymes involved in other formate-producing pathways prompted investigations toward identifying alternative reducing systems used by class III RNRs.⁸⁶ The Stubbe lab reported a second subtype of class III RNR from *Neisseria bacilliformis* that uses the thioredoxin / thioredoxin reductase / NADPH system as the electron source instead of formate.⁸⁶ Additional work from the Stubbe group led to the characterization of a third subtype within class III RNRs from the methanogenic archaeon *Methanosarcina barkeri*. This RNR was found to use a reduction system consisted of ferredoxin, a ferredoxin:disulfide reductase, and a thioredoxin-like protein encoded within the operon.⁸⁷ The diversity observed within class III RNRs likely reflects evolutionary pressure driven by the availability of intracellular reductants. Indeed, the subtype of RNRs encoded in an organism was found to correlate with its anaerobic metabolic capabilities.⁹⁰

1.5.3 Decarboxylases: HPAD, PAD, IAD

4-Hydroxyphenylacetate decarboxylase (HPAD) is the prototype of GRE decarboxylases and has been structurally and biochemically characterized (Figure 1.6).^{38,50} Microbial decarboxylation of *p*-hydroxyphenylacetic acid to *p*-cresol and carbon dioxide was first described in 1985 in *Clostridioides difficile* (formerly known as *Clostridium difficile*) by Barker and D'Ari.⁹¹

In 2001, the GRE responsible for this activity was purified and identified from *C. difficile* by Selmer and Andrei.³⁸ HPAD is a multisubunit GRE that crystallized as a heterotetramer ($\beta\gamma$)₄ consisting of one large catalytic unit (β) that harbors the Gly-Cys dyad and one small subunit (γ) containing two [4Fe-4S] clusters.⁵⁰

The substrate *p*-hydroxyphenylacetate is bound in the active site with its carboxylate group in vicinity of the Cys residue, the site of the thiyl radical. This orientation supports the involvement of a Kolbe-type decarboxylation where one-electron oxidation of a carboxylate by the thiyl radical is proposed to occur. This proposed mechanism is distinct from all other GREs where a hydrogen atom abstraction has been accepted as the first step. The thiolate generated on the conserved Cys can undergo protonation by an adjacent, conserved Glu. The resulting substrate-based radical is resonance stabilized through the benzyl group. A conserved Glu residue is predicted to deprotonate the *para*-hydroxyl group to generate a ketyl radical anion that can decompose to yield carbon dioxide and the *p*-cresol product radical. This proposed mechanism is supported by crystallography,⁵⁰ biochemical characterization,⁹² and a computational investigation using continuum electrostatics and quantum chemical/molecular mechanical calculations⁹³.

In an early study that tested a small panel of Clostridiales, *C. difficile* was unique in its production of *p*-cresol, which was shown to be a product of tyrosine fermentation via the intermediate *p*-hydroxyphenylacetate.⁹⁴ The phenolic product *p*-cresol is a bacteriostatic compound that inhibits the growth of anaerobic bacteria.⁴⁰ Interestingly, the gut pathogen *C. difficile* can tolerate this compound at higher levels than most other Clostridial species and therefore *p*-cresol has been proposed to be a virulence factor that confers a fitness advantage.^{95,96} *p*-Cresol tolerance levels differ between different *C. difficile* strains with the hypervirulent strain R20291 being more tolerant and a better producer.⁹⁶ The tolerance mechanism remains unknown

and appears to be distinct from the HPAD operon. In addition to having antimicrobial properties, *p*-cresol interferes with human drug metabolism³⁹ and is elevated in children with autism¹⁸. Specifically, bacterially produced *p*-cresol in humans is *O*-sulfonated in the liver for excretion, a process that competes with the sulfonation of drugs like acetaminophen.³⁹ Sulfonation of xenobiotics modifies the physical properties of these compounds and affects drug bioavailability and bioactivity. Therefore, *p*-cresol is a bacterial metabolite from the gut microbiome that impacts human drug metabolism. Urinary *p*-cresol and *p*-cresylsulfate were detected at higher levels in autistic children.¹⁸ Intriguingly, gut microbiota dysfunctions are often associated with autism severity,⁷ which may be mediated through changes in microbial metabolites such as *p*-cresol. The molecular details of how this compound impacts host physiology remains unclear. Furthermore, it has been argued that *p*-cresol conjugates (*p*-cresylsulfate and *p*-cresylglucuronide) are likely to be the dominant and physiologically relevant metabolites in humans, therefore future studies on toxicities should focus on these conjugates rather than the parent compound *p*-cresol.⁹⁷

A second GRE decarboxylase was identified to mediate the last step in anoxic toluene biosynthesis, a microbial transformation that was first reported over three decades ago in lakes.⁹⁸ Due to difficulties in observing activity in bacterial monocultures, the Beller group anoxically enriched a microbial community capable of producing toluene from sewage sludge.⁹⁹ Labelled intermediates demonstrated that phenylacetate was decarboxylated to produce toluene (Figure 1.6). This activity was hypothesized to be catalyzed by a GRE enriched in partially purified cell extracts. In a follow-up paper, the Beller and Keasling groups confirmed the identity and sequence of the GRE through metagenomics and metaproteomics of enrichment cultures and partially purified protein fractions, respectively.¹⁰⁰ Phenylacetate decarboxylase (PAD) was purified and characterized *in vitro*, with substrate specificity toward phenylacetate and not *p*-

hydroxyphenylacetate. Interestingly, PAD appears to have a single large subunit and differs from its homolog HPAD. However, the rates of product formation obtained by the authors were extremely low. Possible contributing factors include protein instability or additional small redox-active subunits that could mediate electron transfer but have not been identified. Substrate preference between PAD and HPAD can potentially be explained by the absence of a conserved Glu residue in PAD that would be positioned next to the carbon in para position of the aromatic ring. The lack of a *p*-hydroxyl group in phenylacetate makes its decarboxylation more chemically difficult due to the absence of an electron-donating effect from this hydroxyl group, which would contribute to additional resonance stabilization.

Analogous to the pathway mediated by HPAD, phenylacetate is likely a product of phenylalanine fermentation, which could be encoded within the same strain expressing PAD or through cross-feeding between bacteria. The physiological role of microbial toluene biosynthesis is currently not known.¹⁰⁰ It may serve a similar role as *p*-cresol biosynthesis in possessing antibacterial properties that provide the producing strain a competitive advantage in the community. Another possible biological role for both decarboxylases is alkalinization of the intracellular environment via the consumption of a proton during decarboxylation. The products toluene and CO₂ are neutral and can diffuse through the cell membrane. This could provide the microorganism tolerance to low pH environments. In addition to having a novel chemical activity, PAD is one of few known enzymes to catalyze the biosynthesis of aromatic hydrocarbons, which are an industrially important source of fuel. Additional work is required, however, to improve the stability and rate of PAD for applications in renewable fuel production.

In addition to GRE decarboxylases involved in tyrosine and phenylalanine fermentation, groups had speculated that a GRE is also involved in tryptophan fermentation.³⁸ Some

Clostridiales have been shown to produce skatole as the end product of tryptophan fermentation.⁹⁴ Labelling experiments identified indole-3-acetate as an intermediate in the pathway performed by these Clostridial species.¹⁰¹ Using similar chemical reactivities catalyzed by GRE decarboxylases, indole-3-acetate could be converted to skatole. Additional evidence for the involvement of a new GRE comes from the instability and rapid deactivation by molecular oxygen in cell-free extracts from a ruminal *Lactobacillus* strain.¹⁰² Indeed, a recent publication described the discovery and biochemical characterization of a third GRE decarboxylase, indoleacetate decarboxylase (IAD), with this activity (Figure 1.6).¹⁰³ The authors demonstrated decarboxylation of indole-3-acetate to generate skatole using purified and activated IAD. Furthermore, they constructed a homology model to identify putative active site residues. The physiological roles for IAD remain unclear, but functions previously proposed for other GRE decarboxylases may apply to IAD as well. Specifically, this transformation could facilitate alkalization of the cytoplasm, production of a bacteriostatic compound to increase bacterial fitness, and maintenance of the proton motive force.

1.5.4 X-succinate synthases: BSS, XSS

GRE succinate synthases are C–C bond forming enzymes that activate a wide range of inert hydrocarbon substrates via an addition of an sp^3 -hybridized C–H bond to the double bond of the co-substrate fumarate (Figure 1.6). Benzylsuccinate synthase (BSS) is the best studied member of this class and catalyzes the addition of toluene to fumarate to generate benzylsuccinate.⁵¹ Homologs of BSS have been identified that accept other aryl- or alkyl-hydrocarbon substrates; these enzymes are termed X-succinate synthases (XSS) where X represents the hydrocarbon.^{104,105} Overall, this class of GREs has evolved as part of microbial strategies to recover carbon and energy through anaerobic biodegradation of hydrocarbon pollutants that can persist in the environment.¹⁰⁶⁻

BSS is a non-typical GRE containing multiple subunits that bind auxiliary [4Fe-4S] clusters.⁵¹ Unfortunately, BSS-AE has eluded purification attempts and thus *in vitro* reconstitution of BSS activity has not been possible, preventing in-depth biochemical characterizations.³⁰ Extensive structural studies by the Drennan group have revealed the positioning of toluene in a hydrophobic active site to orient the methyl group next to the Cys residue. An initial hydrogen atom abstraction at the methyl group is proposed to occur before radical addition can take place.

Other known XSSs can accept *p*-cymene,^{104,109} 2-methylnapthalene,¹¹⁰ *p*-cresol,¹¹¹ or *n*-alkanes^{105,112} as substrates with their activities initially detected in environmental bacterial isolates or complex microbial communities (Figure 1.6). The genes responsible for these transformations have been identified and form distinct phylogenetic branches grouped by activity.¹⁰⁴ Overall, there is great interest in harvesting the power of these enzymes for bioremediation of contamination sites.

1.5.5 Eliminases: GD, PD, CutC

This class of GREs catalyzes 1,2-elimination reactions, including some transformations that are also catalyzed by coenzyme B₁₂-dependent eliminases. Glycerol dehydratase (GD)⁵⁶, CutC^{35,55}, and propanediol dehydratase (PD)^{113,114} make up this class and have been biochemically and structurally characterized (Figure 1.6). HypD is the newest member of the eliminase class and catalyzes a unique transformation that does not have a known analogous reaction catalyzed by a coenzyme B₁₂-dependent enzyme. The discovery and characterization of HypD will be the topic of subsequent chapters. This section begins by describing the initial discoveries of GD, PD, and CutC and their roles in microbial physiology. Proposed mechanism for each catalyzed reaction are then discussed in detail. We highlight similarities and differences between these three GRE

eliminases to provide background for structural and biochemical studies of HypD reported in Chapter 3.

GD catalyzes the dehydration of glycerol to 3-hydroxypropionaldehyde and was initially discovered in *Clostridium butyricum* in 2003.¹¹⁵ This reaction is of industrial importance since glycerol is a major by-product of biodiesel fuel production.¹¹⁶ The product 3-hydroxypropionaldehyde can be reduced to 1,3-propanediol, which is an important monomer in industrial polymer synthesis.¹¹⁷ Therefore, there is interest in the field for biotechnological applications of GD toward 1,3-propanediol production.¹¹⁷ *In vitro*, GD can also catalyze the dehydration of 1,2-propanediol to propionaldehyde acting as a propanediol dehydratase, but the physiological relevance of this activity is unclear.⁵⁶ This enzyme was the first discovered GRE eliminase and has been the subject of multiple computational studies, which provided the first indications that this GRE class uses a direct elimination mechanism.¹¹⁸

The role of PD in L-fucose fermentation was first proposed from a 2006 transcription profiling study in the microbe *Roseburia inulinivorans*.¹¹⁹ A propanediol utilization (*pdu*) operon was upregulated during growth on L-fucose found to encode a GRE and AE instead of the previously characterized B₁₂-dependent propanediol dehydratase found in other propanediol utilization (*pdu*) operons.¹¹⁹ Its genomic context and high sequence similarity to GD provided strong evidence that the newly identified GRE was a PD. Its precise activity was only recently confirmed, a decade later, through *in vitro* studies carried out by Lanzilotta¹¹³ and our group⁶⁸. Both studies established that the GRE PD has a strong preference for the dehydration of (*S*)-1,2-propanediol over (*R*)-1,2-propanediol. These results are consistent with (*S*)-1,2-propanediol being the stereoisomer produced from microbial fermentation of L-fucose.¹²⁰ Dehydration of 1,2-

propanediol results in propionaldehyde formation, which can be reduced to propanol or oxidized to propionate in the microbe.

CutC is the only known enzyme to catalyze C–N bond cleavage of choline to form TMA and acetaldehyde. This enzyme was discovered by Prof. Emily Balskus through a BLAST search of a TMA-producing strain *Desulfovibrio desulfuricans* ATCC 27774 for homologs of proteins involved in ethanolamine catabolism.³⁴ Anaerobic choline metabolism was predicted to share similar biochemical logic as ethanolamine metabolism due to identical downstream intermediates. A gene cluster was identified containing protein homologs expected to catalyze these downstream steps, which was predicted to be responsible for choline metabolism. Instead of a coenzyme B₁₂-dependent ethanolamine ammonia-lyase (EAL), a putative GRE and its AE was found in this locus. The similar genomic organization of this cluster to the ethanolamine utilization operon along with the ability of species encoding this cluster to metabolize choline to TMA formed the basis for the proposed activity of this GRE to be a choline TMA-lyase. Extensive characterization by Dr. Smaranda Bodea and Dr. Ana Martínez-del Campo in the Balskus group led to validation of CutC's activity, an understanding of its mechanism and substrate scope, the role of additional genes encoded in the gene cluster, and its distribution in sequenced genomes.^{34,35,121,122}

In addition to the unique chemistry CutC performs, this enzyme catalyzes a uniquely microbial transformation associated with various human diseases.^{36,37,123} With the discovery of the genetic basis for this metabolic pathway, *E. coli* mutants containing single deletions of the essential genes have enabled gnotobiotic mouse studies.¹²² This work demonstrated that CutC is essential for choline metabolism and TMA production in the gut microbiome, which impact host physiology.¹²² In addition, there are interests to develop inhibitors to reduce TMA production in CutC-encoding microbes due to its disease relevance.^{124,125} Since TMA is a microbially produced

metabolite not accessible through human enzymes, the anaerobic choline metabolic pathway is an ideal candidate for inhibitor development.

A unique feature shared among GRE eliminases PD and CutC is their colocalization in gene cluster with genes encoding microcompartment shell proteins. Bacterial microcompartments are proteinaceous intracellular structures that encapsulate enzymes involved in a common pathway, often catabolism of a metabolite. Microcompartments can help to concentrate substrates as well as to prevent the release of volatile and/or reactive products that need to be further processed. In the case of PD and CutC, generation of reactive propionaldehyde and acetaldehyde, respectively, are thought to underlie the need for encapsulation.¹²⁶ Downstream enzymes involved in reduction and oxidation of these aldehydes are predicted to be microcompartment-associated as well, which supports this claim. The molecular details of how proteins, cofactors, and metabolites are assembled within and transported across the microcompartment shell remain mostly unclear.

B₁₂-dependent dehydratases that catalyze identical reactions as GRE GD and PD were discovered decades earlier with multiple studies on their mechanisms.^{127,128} It was established through labelling experiments that the B₁₂-dependent radical enzymes mediate a 1,2-migration of the hydroxyl group on C2 (Figure 1.7B).¹²⁷ However, it has been debated whether GRE GD and PD mechanisms involve a direct elimination or an intramolecular rearrangement of this hydroxyl group similar as the B₁₂-dependent enzymes (Figure 1.7).^{56,118} Prior to work from our lab, the sole evidence for GREs catalyzing a direct elimination came from a computational study on GD.¹¹⁸ Quantum mechanics/molecular mechanics (QM/MM) simulations revealed a much higher energy barrier (>20 kcal/mol) for the proposed 1,2-migration of hydroxyl group from C2 to C1 compared to a direct elimination of this leaving group (<6 kcal/mol).¹¹⁸ A heavy isotopic labelling experiment conducted by Benjamin Levin in our group was pivotal in supporting a direct elimination

mechanism for GRE PD.¹¹⁴ In this study, stereo- and enantio-selectively ^{18}O -labelled 1,2-propanediol substrates were incubated with both GRE and B₁₂-dependent PD. Percentage of the ^{18}O -label in the reduced product propanol was quantified by GC-MS. Approximately 50% labelled product was observed for all B₁₂-dependent PD reactions, which is in agreement with the previous labelling studies to support a migration of the hydroxyl group from C2 to C1 (Figure 1.7).¹²⁷ In stark contrast, the percentage of labelled product generated by the GRE PD depended on the site of labelling in the propanediol substrate. Labelling of the hydroxyl group on C2 in propanediol led to nearly complete loss of ^{18}O -atom in the product. From substrates with ^{18}O -labelled C1 hydroxyl group, almost all propanol generated retained the ^{18}O -label. These results strongly support a mechanism involving direct elimination for GRE PD (Figure 1.7A). Furthermore, a QM/MM study was published soon after and provided addition support for these two distinct mechanisms between B₁₂-dependent and GRE GD.¹²⁹

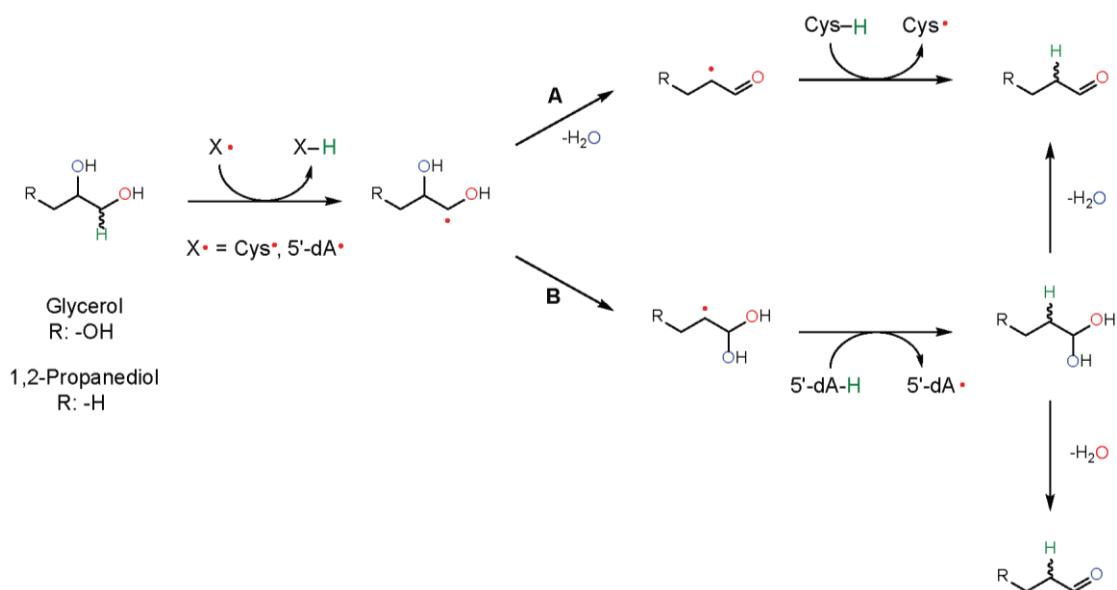


Figure 1.7 Proposed mechanisms for diol dehydration are distinct for GRE and B₁₂-dependent dehydratases.

Glycerol or 1,2-propanediol dehydration are catalyzed by both B₁₂-dependent enzymes and GREs. A 5'-dA radical serves as the radical initiator in B₁₂-dependent enzymes whereas a thiyl radical on a conserved Cys residue is used in GREs. Initial hydrogen atom abstraction is proposed to occur on C1 of substrate in both cases. (A) Direct elimination of the hydroxyl group on C2 is predicted to occur in GRE-catalyzed reactions. This generates an aldehyde product radical that undergoes hydrogen atom re-abstraction with the Cys residue to regenerate the thiyl radical. (B) A 1,2-migration rearrangement is proposed to occur during B₁₂-dependent enzyme-catalyzed dehydration. This generates a less stable 1,1-*gem* diol product radical, which then undergoes a final hydrogen atom abstraction with 5'-dA. The product then collapses to release a water molecule, yielding the final aldehyde product. O-atoms

Figure 1.7 (Continued)

are labelled by color to highlight a retention of -OH on C1 in mechanism A as opposed to scrambling of O-atoms in mechanism B.

Evolution of B₁₂-dependent enzymes and GREs to catalyze identical reactions through different mechanisms can be elegantly explained by differences in the reactivity of the radical cofactors involved (Figure 1.7). In B₁₂-dependent enzymes, a highly reactive 5'-dA radical catalyzes the initial hydrogen atom abstraction on C1 of substrate. The thiyl radical used by GREs is less reactive, but hydrogen bonding between C1 hydroxyl group and a conserved Glu residue PD and GD can decrease the BDE of the C1-H bond to facilitate homolytic cleavage (Figure 1.8, Figure 1.9). Differences in the stability of radical products generated by these mechanisms correlate to differences in the X-H bond strength of radical cofactors (Figure 1.7). 1,2-migration rearrangement generates a 1,1-diol product radical that is not stabilized through resonance. In contrast, direction elimination produces a radical aldehyde product that is in resonance with a ketyl radical, resulting in stabilization of the unpaired electron through delocalization. The greater reactivity of the 1,1-diol radical is likely important in facilitating hydrogen atom re-abstraction by cofactor 5'-dA (BDE of C-H ~100 kcal/mol)¹³⁰, which has a stronger bond compared to S-H in Cys (BDE ~87 kcal/mol)¹³¹. Overall, both experimental and computational studies strongly support an intramolecular rearrangement during dehydration catalyzed by B₁₂-dependent enzymes and a direct elimination catalyzed by GRE dehydratases.

Similarly for CutC, choline metabolism could proceed through a 1,2-migration of TMA functional group or a direct elimination. Crystallographic snapshots of several CutC mutants with altered active site residues in combination with their biochemical characterizations led to the

Balskus and Drennan groups to propose a mechanism that proceeds via direct elimination.⁵⁵ A major finding that supported this conclusion was the retention of activity in the double mutant Y208F/Y506F. These residues interact with the TMA functional group through CH–O hydrogen bonding, which is thought to be important in facilitating migration of this moiety. In contrast, mutagenesis of equivalent residues in EAL, a B₁₂-dependent enzyme that catalyzes a C–N bond cleavage through 1,2-migration of the leaving group resulted in a loss of activity.¹³² Additional support comes from the tight packing of choline in the CutC active site which may not accommodate a rearrangement of the bulky TMA moiety. However, crystal structures do not account for protein flexibility and dynamics. In conclusion, structural and biochemical experiments on CutC provided evidence to favor a direct elimination over a 1,2-migration for its mechanism (Figure 1.10).

In addition to mechanistic insight, structural elucidation of these GRE eliminases also revealed a set of residues with conserved roles during catalysis. As previously described, Gly and Cys residues involved in radical chemistry are absolutely conserved in GREs. In eliminases, Gly763 and Cys433 in GD, Gly817 and Cys438 in PD, and Gly821 and Cys489 in CutC participate in radical relay to the substrate. Mutagenesis of these residues in PD and CutC abolished activities as expected.^{55,68} Additionally, no detectable activation was observed in PD and CutC mutants for the Gly residue to validate this as the site of radical generation.

All three transformations catalyzed by these GRE eliminases result in the oxidation of a C–O bond to generate an aldehyde product. To achieve this, a conserved Glu residue is poised near the C1 hydroxyl group to catalyze its deprotonation. In addition, hydrogen bonding between the hydroxyl and Glu is proposed to lower the BDE of the C–H bond on C2 to facilitate initial hydrogen atom abstraction.^{129,133} Mutagenesis of this residue in PD and CutC abolished activity *in*

vitro, supporting an essential role in catalysis.^{55,68} This residue corresponds to Glu435 in GD (Figure 1.8), Glu440 in PD (Figure 1.9), and Glu491 in CutC (Figure 1.10) and is conserved among all putative GRE eliminases with these activities.⁶⁸ Overall, this Glu residue is thought to be the general base that deprotonates hydroxyl group on C1 to generate a ketyl radical intermediate for all three GRE eliminases.

During the elimination step, a general acid is expected to protonate the leaving group. A conserved His residue among GD and PD, and an Asp residue in CutC are thought to serve this catalytic role. In substrate-bound structures, His164 in GD and His166 in PD are in hydrogen bonding distances with the hydroxyl group on C2 (Figure 1.8, Figure 1.9).^{56,113} Protonation of this hydroxyl group makes it a better leaving group and thus facilitates the dehydration step. This is supported by a lack of detectable activity in the corresponding PD mutant and the absolute conservation of a His residue in all putative GRE dehydratases.⁶⁸ In CutC, the substrate is positively charged and thus protonation by Asp216 is predicted to occur post-elimination of the TMA moiety (Figure 1.10). TMA production was observed in CutC-D216N, demonstrating that this residue is not essential for C–N bond cleavage. To summarize, a His residue serves as the general base in GRE dehydratases while a conserved Asp residue in CutC is proposed to fulfill this role.

Both conserved His and Glu residues present in PD and GD are positioned adjacent to each other in the active site. A direct proton transfer between these two residues could occur to reset their protonation states for the next catalytic cycle. In CutC, however, the general acid (Glu491) and the general base (Asp216) are outside of hydrogen bonding distances from one another for proton transfer. Furthermore, a Thr residue is positioned between these residues (Figure 1.10). Mutagenesis of this residue drastically reduced activity in the CutC mutant.⁵⁵ This result and the

CutC structure led Bodea, Funk *et al.* to postulate that Thr502 may mediate proton transfer from Glu491 to Asp216, which would then catalyze protonation of TMA upon elimination.⁵⁵

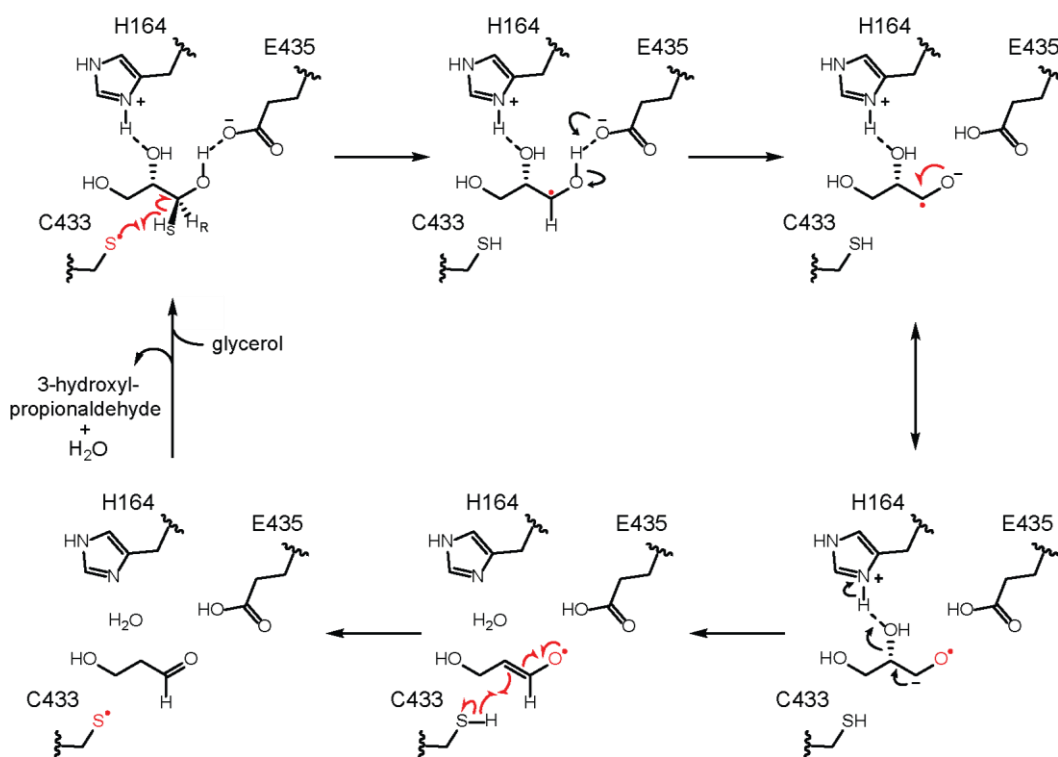


Figure 1.8 Proposed mechanism for glycerol dehydration catalyzed by GD.

Mechanism and residues important for catalysis were proposed based on structural and computational studies.^{56,118,129} Residue numbering is based on GD from *Clostridium butyricum* (UniProt ID: Q8GEZ8).

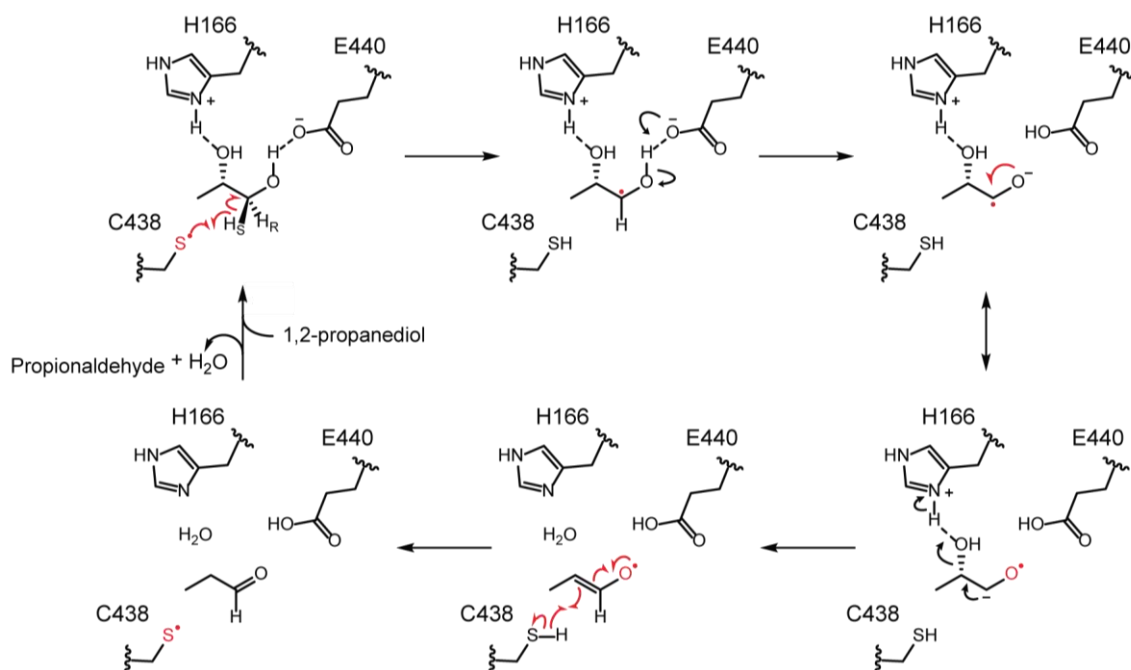


Figure 1.9 Proposed mechanism for propanediol dehydration catalyzed by PD.

Residues important for catalysis were identified from structural and mutagenesis experiments.^{68,113} Residue numbering is based on PD from *Roseburia inulinivorans* DSM 16841 (UniProt ID: Q1A666).

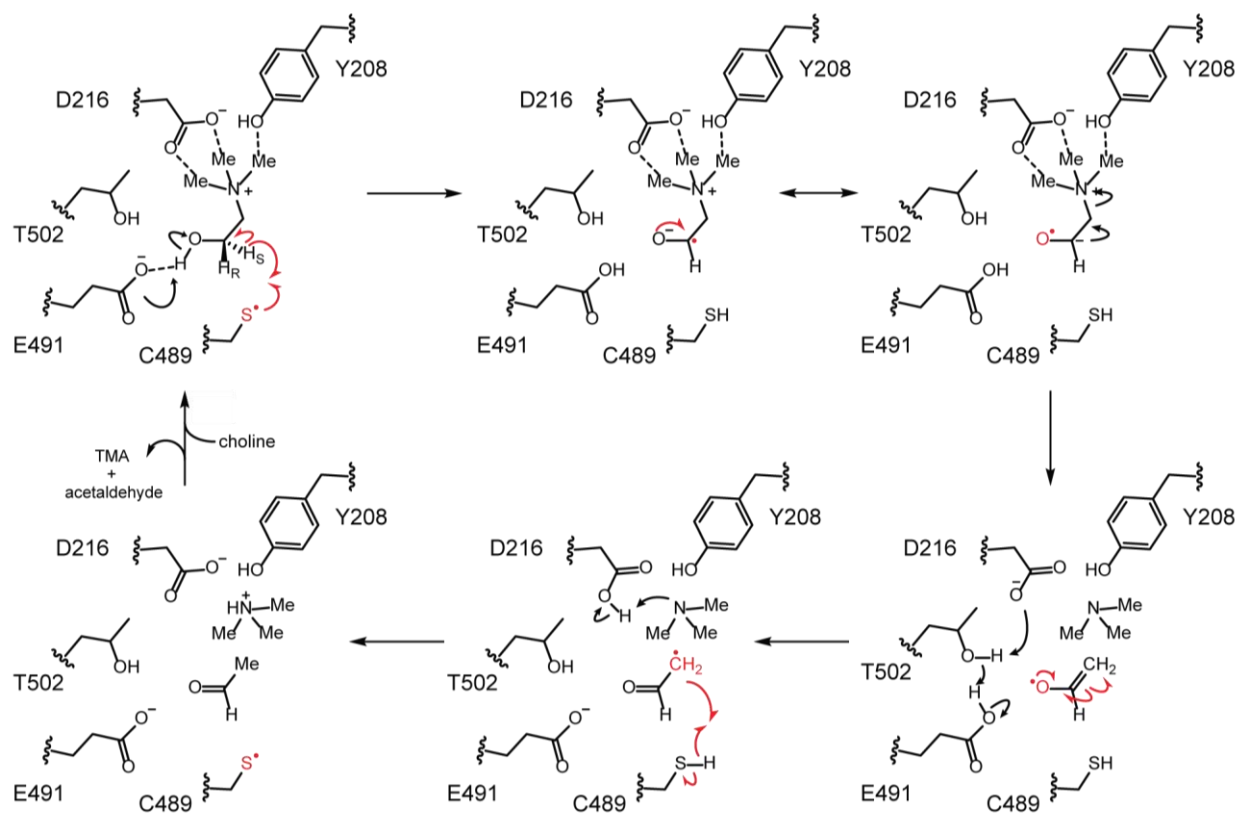


Figure 1.10 Proposed mechanism for TMA elimination catalyzed by CutC.

Residues important for catalysis were identified from structural and mutagenesis experiments.⁵⁵ Residue numbering is based on CutC from *Desulfovibrio alaskensis* G20 (UniProt ID: Q30W70).

In summary, this section presented the proposed mechanisms for characterized GRE eliminases while highlighting differences and similarities between diol dehydratases and CutC. All GRE mechanisms share the common feature of a direct elimination, which is strongly supported by computational, structural, and biochemical studies. Conserved residues have been identified in these enzymes to enable acid-base chemistry that is essential in elimination reactions. Subsequent chapters of this thesis describe in detail work toward the study of a new GRE, HypD,

that catalyzes dehydration of Hyp to (*S*)- Δ^1 -pyrroline-5-carboxylate (P5C). The amino acid Hyp and its metabolism are subjects of remaining sections in this chapter.

1.6 *trans*-4-Hydroxy-L-proline is an abundant non-proteinogenic amino acid in the eukaryotic world

As topics of work in Chapter 2 and 4, the recently characterized HypD is widespread in microbial genomes and metagenomes, indicating this transformation may be a core function in microbial primary metabolism. Its prevalence mirrors the distribution of its substrate Hyp. Hyp is an abundant non-proteinogenic amino acid generated by prolyl 4-hydroxylases and proline 4-hydroxylases. These enzymes belong in the family of mononuclear non-heme iron and α -ketoglutarate (α -KG) dependent dioxygenases, found in all domains of life.¹³⁴ Hydroxylation of proline on C4 is catalyzed using an Fe(II) metal center with α -KG and molecular oxygen as co-substrates. The requirement for molecular oxygen is reflected in the distribution of dioxygenases in facultative and strict aerobes.¹³⁴ Across the domains of life, hydroxylases responsible for Hyp synthesis are widespread among eukaryotes, especially plants and animals, whereas they are rarely found in bacteria and archaea. Therefore, eukaryotes are the primary source of Hyp with microbes contributing to a much lesser extent.

Eukaryotes encode prolyl 4-hydroxylases, which post-translationally modify Pro residues in proteins and peptides (Figure 1.11A).¹³⁴ In mammals, Hyp is almost exclusively found in collagen, which is the most abundant protein in the human body (Figure 1.11B).¹³⁴ Hyp residues are critical for the functions of additional structural and regulatory proteins, including elastin and hypoxia inducible factor.¹³⁴ Overall, Hyp is estimated to make up ~4% of the amino acid pool in the human proteome, making it more abundant than several proteinogenic amino acids.¹³⁴ In addition to animals, plants and algae also produce proteins enriched in Hyp that are important for

cell wall rigidity, growth, and signaling.^{134,135} Indeed, Hyp-rich glycoproteins (HRGPs) are a major component of the plant cell wall, with the 4-hydroxyl groups of Hyp residues serving as sites for *O*-glycosylation (Figure 1.11C).¹³⁶

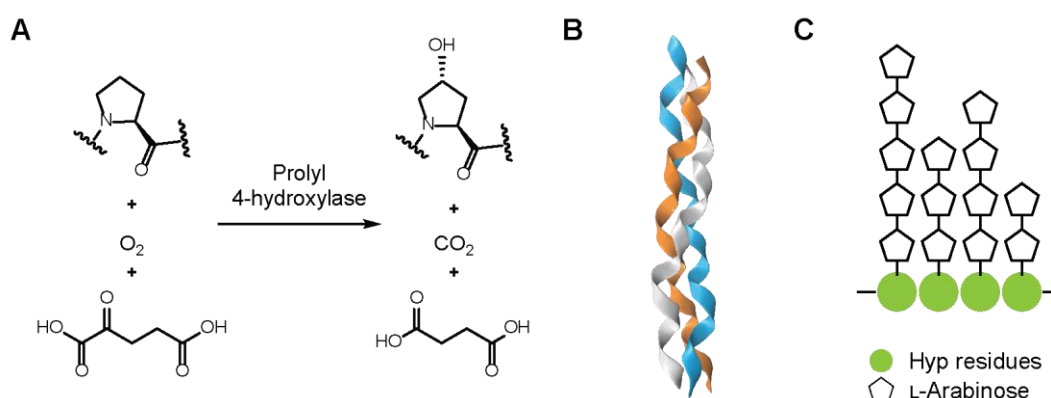


Figure 1.11 Biosynthesis and sources of *trans*-4-hydroxy-L-proline in eukaryotes.

(A) Prolyl 4-hydroxylase, a non-heme iron α -KG dependent dioxygenase, catalyzes hydroxylation of peptidyl proline at C4 of its side chain. (B) Mammalian structural protein, collagen, contains triplet repeats with Pro-Hyp-Gly being the most common across all collagen types. The first high-resolution crystal structure of a collagen triple helix is displayed as an example (PDB ID: 1CAG). Each peptide chain is colored differently. (C) Plant cell wall glycoproteins are enriched in Hyp as sites for *O*-glycosylation. Arabinose repeats in extensin are depicted as an example.

In contrast to its prevalence in eukaryotes, Hyp synthesis is uncommon in bacteria and archaea. Most Hyp-generating species use proline 4-hydroxylases to hydroxylate free Pro (Figure 1.12A).¹³⁴ Free Hyp is synthesized by halophiles and acts as an osmolyte under high salt

concentrations.¹³⁷ Hyp is also synthesized as a building block in the biosynthesis of non-ribosomal peptide natural products, including actinomycin and etamycin (Figure 1.12B).^{138,139} Despite known microbial examples of Hyp synthesis being predominantly catalyzed by proline 4-hydroxylases, there are two reported bacterial prolyl 4-hydroxylases that post-translationally modify peptidyl Pro.^{140,141} *Bacillus anthracis* encodes a prolyl 4-hydroxylase that can bind and hydroxylate collagen-like (Pro-Pro-Gly)₁₀ and (Pro-Pro-Gly)₅ peptides.^{141,142} The physiological substrate(s) of this enzyme, however, remained unknown until 2017 when Dey and coworkers identified elongation factor Tu as the substrate.¹⁴³ This supports a wide substrate scope for this enzyme and a putative role in O₂-sensing and translational regulation. Interestingly, the second example of a bacterial prolyl 4-hydroxylase, first characterized in *Pseudomonas aeruginosa*, also modifies elongation factor Tu as its substrate.¹⁴⁰ It has been proposed that bacterial prolyl 4-hydroxylases were the evolutionary origins of hypoxia sensing in eukaryotes.

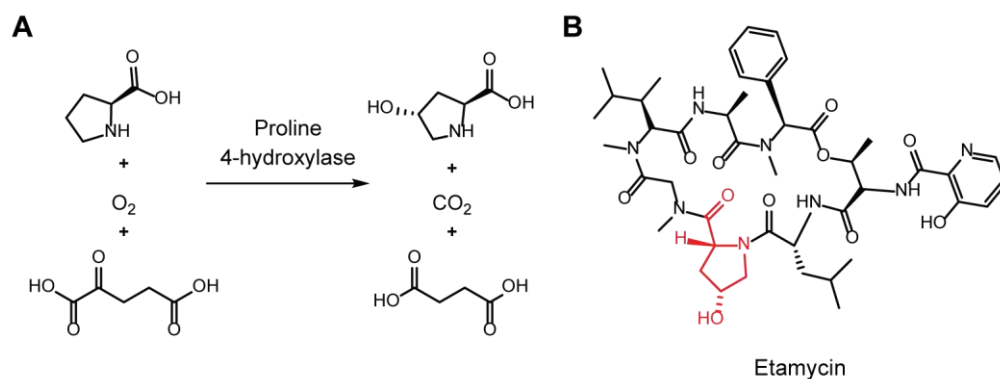


Figure 1.12 Biosynthesis and sources of *trans*-4-hydroxy-L-proline in bacteria and archaea.

(A) Proline 4-hydroxylase, a non-heme iron α -KG dependent dioxygenase, catalyzes hydroxylation of free amino acid proline at C4. (B) Etamycin is an example of a bacterial non-ribosomally synthesized peptide containing *cis*-4-hydroxy-L-proline (in red) derived from free Hyp.

1.6.1 Sources of Hyp in the human gut

Because Hyp is highly abundant in both animal and plant proteins, a wide variety of dietary sources contain this amino acid. In meat products, Hyp content correlates with the amount of connective tissue and can be detected at up to 1% in weight.^{144,145} Connective tissues are rich in collagen, which has a defining feature of a three amino acid repeating motif of Xaa-Yaa-Gly with Pro-Hyp-Gly being the most common triplet.¹⁴⁶ Hyp is also enriched in dietary fibers that contain plant cell wall polysaccharides as a major component. For example, Hyp makes up over 4% of the protein content in dietary fibers extracted from grape pomace.¹⁴⁷ Arabinogalactan proteins, a class of HRGPs, are also a major component in gum arabic, which is used as a thickening and stabilizing agent in the food industry and is found in products like candies, gum, and beverages.¹⁴⁸ In a metabolomics study, Hyp was identified as a top plasma biomarker for distinguishing omnivores

from vegans (due to its higher abundance in omnivores), lending support to the diet being an significant source of Hyp.¹⁴⁹ This finding is also in agreement with the higher Hyp content in meat products compared to plant material. However, the possibility that dietary changes affect the degree to which gut microbes and/or the host metabolize Hyp cannot be discounted.

In addition to dietary sources, it is estimated that endogenous collagen turnover in humans releases ~300 mg of Hyp daily.¹⁵⁰ Therefore, both dietary and endogenous collagen are candidate sources of Hyp for microbial metabolism in the human gut. Because Hyp is rarely included in amino acid measurements, Hyp concentrations within the GI tract remain unknown, with one study detecting free Hyp as the most abundant amino acid in the large intestine of one human subject.¹⁵¹ Considering this study, as well as the wide distribution of Hyp in dietary substrates and its high abundance in endogenous structural proteins, we propose that microbes likely encounters significant levels of Hyp in free and peptide forms in the gut.

1.7 Human metabolism of *trans*-4-hydroxy-L-proline

In the human body, the large quantity of Hyp released from endogenous collagen turnover cannot be reused for protein synthesis. Rather, the amino acid is degraded into smaller metabolites and half of its carbon content is excreted as waste. About 80-90% of free Hyp is efficiently metabolized to glyoxylate and pyruvate within the mitochondria of liver and kidney cells in an overall oxidative pathway (Figure 1.13A).^{150,152} Pyruvate is a central intermediate in many metabolic pathways, including the tricarboxylic acid (TCA) cycle and is a source of carbon and energy (Figure 1.13B). On the contrary, glyoxylate produced from host Hyp metabolism is reactive and typically considered as a waste product (Figure 1.13B). In the cytosol, glyoxylate is oxidized to oxalate, which cannot be further metabolized and can form calcium oxalate crystals at higher concentrations.¹⁵³ To prevent crystal formation, the majority of glyoxylate is reduced to glycolate

for detoxification and excretion.¹⁵³ Genetic mutations of enzymes involved in this pathway can cause a rare condition termed primary hyperoxaluria characterized by recurrent kidney and bladder stone formation, which can lead to organ failure.¹⁵³ Hyp is a major contributor to the glyoxylate pool in the human body¹⁵⁴ and research efforts have been directed toward inhibiting human enzymes involved in Hyp metabolism to help treat primary hyperoxaluria^{155,156}.

Due to efficient metabolism of the free amino acid, the vast majority (up to 95%) of Hyp excreted in urine is in the form of small peptides with Pro-Hyp as the major product.¹⁵⁰ Interestingly, this peptide is not efficiently hydrolyzed by host peptidases^{25,26} and may require microbial peptidases for complete release of free amino acids. Multiple studies have obtained evidence to suggest that in addition to amino acids, peptides can be directly absorbed by the host from the GI tract. For example, ingestion of various gelatin and collagen hydrolysates resulted in increased levels of collagen-derived short peptides in the blood of human subjects, with Pro-Hyp dipeptide being the major product in most cases and Hyp-Gly observed in some individuals at high levels.^{27,157} However, the possibility that these dietary substrates impacted endogenous collagen turnover by inducing elevated levels of collagen-derived peptides within the host cannot be ruled out.

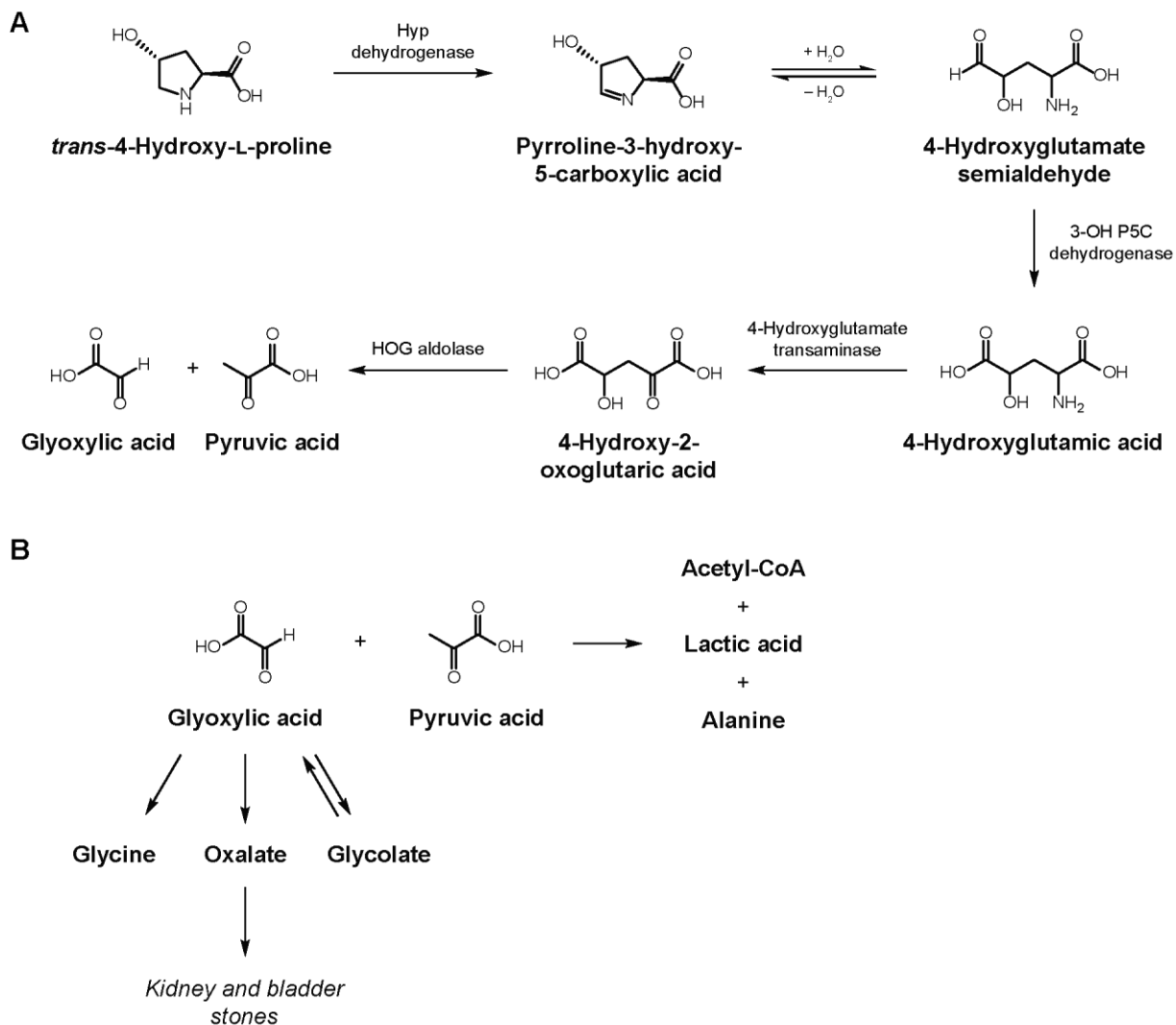


Figure 1.13 *trans*-4-Hydroxy-L-proline metabolism in the human body.

(A) In the only known Hyp metabolic pathway in mammals, Hyp is oxidized to a pyrroline intermediate. The pyrroline undergoes non-enzymatic hydrolysis and ring opening to form a semialdehyde which is then sequentially oxidized to a keto acid. C–C bond cleavage this intermediate yields two smaller metabolites, glyoxylate and pyruvate. (B) Downstream metabolites of glyoxylate and pyruvate in humans. Glyoxylate is mostly reduced to glycolate for excretion. When oxidation of

Figure 1.13 (Continued)

glyoxylate to oxalate is upregulated, this can lead to a disorder called hyperoxaluria where kidney and bladder stone formation occurs at high frequencies. Pyruvate is a central intermediate of the TCA cycle and anaerobic glycolysis.

1.8 Bacterial strategies for metabolizing *trans*-4-hydroxy-L-proline

1.8.1 Stickland fermentation

Anaerobic metabolism of Hyp was first observed in *Clostridium sporogenes*, a proteolytic species in the order Clostridiales, by L.H. Stickland in 1934.¹⁵⁸ This seminal work established a type of bacterial energy metabolism later coined Stickland fermentation wherein amino acids are used as electron donors and/or acceptors to generate ATP through substrate-level phosphorylation (Figure 1.14). In the original experiment, Stickland incubated washed suspensions of *C. sporogenes* with redox active dyes and individual amino acids to identify substrates that can cause oxidation or reduction of each dye. Pro and Hyp were found to cause rapid and complete oxidation of certain dyes, whereas Gly resulted in partial oxidation. Utilization of one or more of these amino acids as electron acceptors in anaerobic growth was later observed in many Clostridiales including *C. sticklandii*,¹⁵⁹ *C. difficile*,¹⁶⁰ and *C. sporogenes*¹⁶¹. Despite several decades of work since their discovery, only the biochemical basis underlying Pro and Gly reduction has been elucidated with the enzymes catalyzing each step characterized.^{160,162,163} Interestingly, Pro reduction does not involve ATP synthesis but has been proposed to couple proton export to generate energy by driving proton motive force (Figure 1.14B).¹⁶¹ An additional role of this pathway may be to generate oxidizing equivalents like NAD⁺ for oxidative pathways. Notably,

both key reductases essential for Stickland fermentation of Gly and Pro use selenocysteines encoded within subunits to access reductive C–N bond cleavage.^{160,164}

In 2006, Self and coworkers reported that Hyp can induce the expression of Pro reductase, which is responsible for Pro reduction in *C. difficile*, indicating that Pro may be an intermediate of in Hyp metabolism.¹⁶⁰ However, the enzymes and downstream metabolites involved in Stickland fermentation of Hyp remained unknown until work presented in this thesis.⁶⁸ Through our investigations to characterization a new GRE, we established that Hyp is reduced to Pro in two enzymatic steps, and this amino acid can then be further reduced to 5-aminovalerate by Pro reductase. Although 5-aminovalerate is considered a fermentation end product of Pro reduction, two Clostridial species were isolated that can further metabolize 5-aminovalerate for energy production.¹⁶⁵⁻¹⁶⁷ These species couple the reduction of 5-aminovalerate to its oxidation into short chain fatty acids for ATP synthesis (Figure 1.14B). Although a few enzymes involved in these pathways have been purified through crude fractionation with detected activities in work by Buckel and coworkers,¹⁶⁸⁻¹⁷⁰ their sequences have not been reported and thus the distribution of this pathway in microbes remains to be determined.

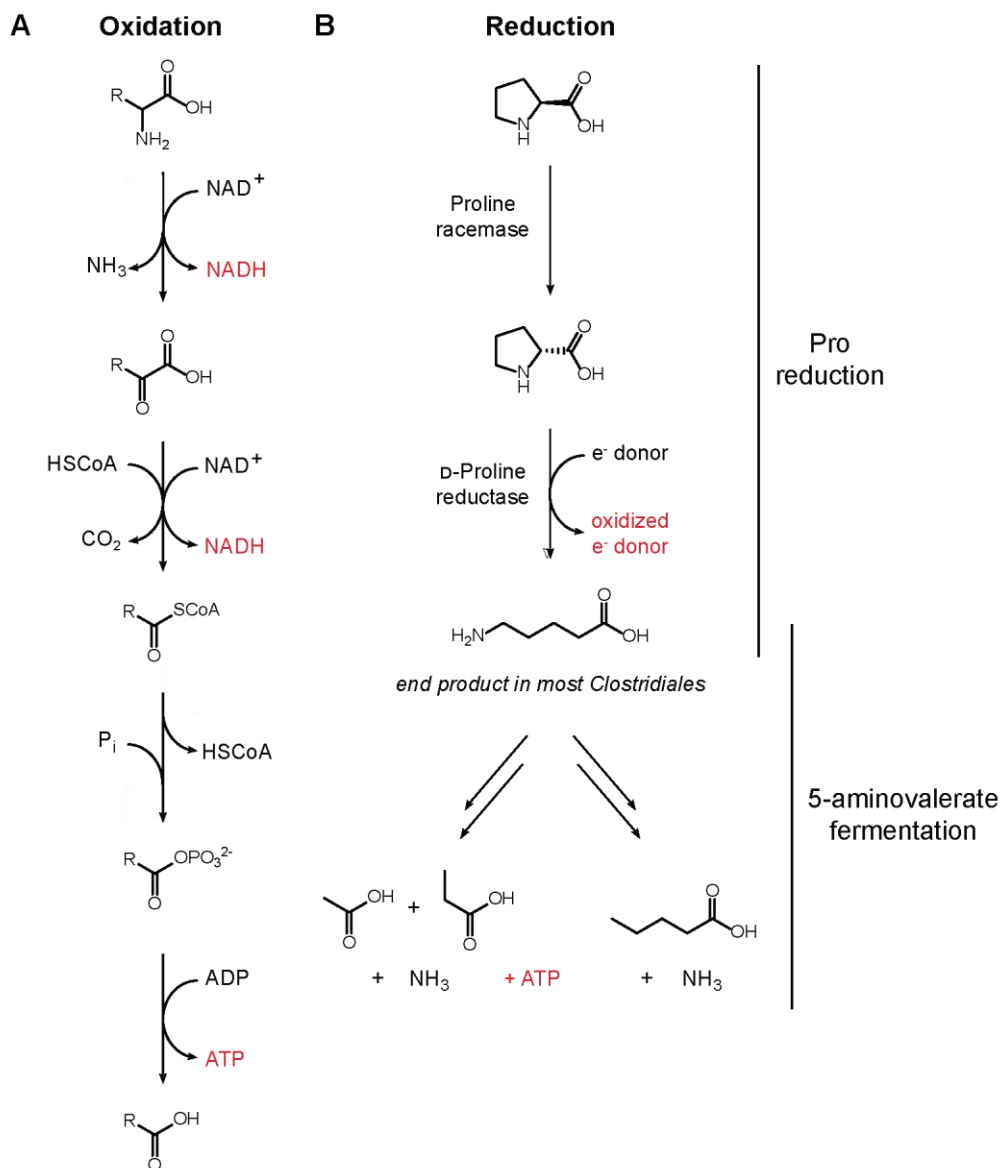


Figure 1.14 Stickland fermentation carried out by species in the order Clostridiales.

(A) A general scheme for the oxidative branch of Stickland fermentation with the generation of reducing equivalents, NADH, and the energy currency, ATP, which are highlighted in red. The first step involves oxidative deamination of the amino acid to yield an α -keto acid and ammonia followed by oxidative decarboxylation and

Figure 1.14 (Continued)

substrate-level phosphorylation to generate an additional reducing equivalent and ATP. Known amino acids utilized as electron donors in this pathway include alanine, valine, isoleucine, and leucine^{158,171,172} (B) Proline reduction in Stickland fermentation results in accumulation of 5-aminovalerate in most Clostridiales.^{161,171} However, certain Clostridial species can ferment 5-aminovalerate into ammonia, acetate, propionate, and valerate through coupled oxidative and reductive reactions.¹⁷⁰

1.8.2 Catabolic metabolism of Hyp in aerobic bacteria

In addition to Hyp fermentation, a different pathway for Hyp metabolism was uncovered in 1959 when E. Adams isolated a *Pseudomonas* strain from soil that could grow aerobically on Hyp as the sole source of carbon and nitrogen.¹⁷³ The enzymes involved in this pathway have been since biochemically characterized in *Pseudomonas* and *Sinorhizobium* species.^{174,175} Aerobic Hyp metabolism involves sequential oxidations of Hyp to generate α -KG and ammonia, providing carbon and nitrogen for anabolic pathways. *cis*-4-Hydroxy-D-proline dehydrogenase carries out a key step in this pathway and shares high sequence similarity with hydrogen cyanide synthase, which is predicted to use O₂ as the terminal electron acceptor.^{174,176} Given its distribution in sequenced genomes and the potential requirement for O₂, we hypothesize that this metabolism occurs only in aerobes. To assess the presence of this pathway in the human microbiota, we performed BLASTP¹⁷⁷ searches of the HMP reference genomes (NCBI BioProject: PRJNA28331) using two characterized *cis*-4-hydroxy-D-proline dehydrogenases¹⁷⁴ as query (NCBI reference ID: NP_743415.1 and NP_249957.1). This search identified hits mostly in organisms isolated from

the skin and respiratory tract.¹ Therefore, we concluded that this alternative pathway likely does not contribute significantly to Hyp consumption in the gut.

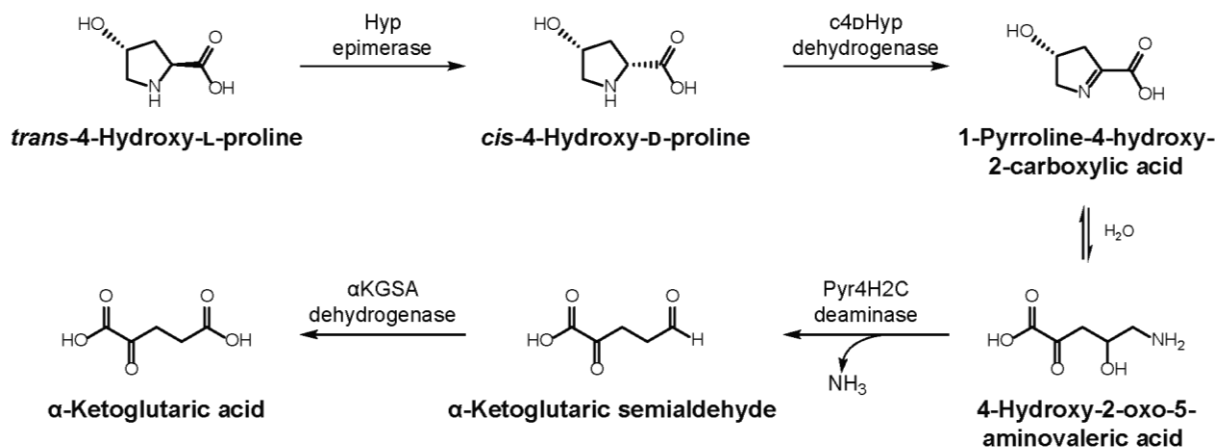


Figure 1.15 Catabolic metabolism of *trans*-4-hydroxy-L-proline by bacteria.

Hyp can serve as a source of carbon, nitrogen, and energy in this catabolic pathway encoded in, to the best of our knowledge, facultative and obligate aerobes.^{174,175} Hyp is first epimerized and then oxidized to undergo hydrolysis and ring opening. Deamination and oxidation yields α-KG, a metabolite with key roles in many pathways including the TCA cycle, amino acid synthesis, and oxidation reactions catalyzed by α-KG-dependent enzymes.

To summarize, there are two known pathways for Hyp metabolism encoded in microorganisms. Although we cannot rule out the possibility of additional, yet-to-be-discovered pathways, findings from culture-based studies have clearly implicated these two routes. The logic underlying these pathways is distinct, with the oxidative metabolism of Hyp providing essential nutrients for aerobic bacteria, while the anaerobic pathway uses Hyp as an electron sink to

regenerate NAD^+ as part of Stickland fermentation in Clostridiales, but likely has distinct roles in other species.

1.9 Chapter preview

This thesis project began with the discovery of a new GRE encoded within a gene cluster conserved in many Clostridiales. This enzyme was proposed to catalyze the dehydration of Hyp to P5C based on its genomic context and known GRE chemistry. *In vitro* biochemical investigations and validation of our hypothesis are described in **Chapter 2**. We confirmed that this GRE is a HypD as well as validated the gene annotations for the remaining genes in the gene cluster to be its AE and a P5C reductase. The distribution of HypD was profiled in sequenced microbial genomes as well as healthy human gut metagenomes to reveal anaerobic Hyp metabolism to be prevalent and abundant in the gut habitat.

After initial biochemical characterization of HypD from *C. difficile*, we sought to gain insight into the mechanism behind Hyp dehydration. In **Chapter 3**, we describe structural elucidation of HypD and biochemical characterization of HypD mutants. We assessed whether conserved active site residues are essential for activation and catalysis to probe their putative roles in the radical-mediated dehydration.

Given HypD's wide distribution in gut microbes and the high abundance of its substrate Hyp in eukaryotes, we wondered if HypD-encoding microbes could be enriched from the environment. To this end, **Chapter 4** focuses on the enrichment culture work and isolation of anaerobic microbial communities capable of Hyp degradation from sediment samples. Furthermore, the gene encoding HypD was detected in these communities, supporting the conclusion that HypD catalyzes the principle pathway for anaerobic Hyp metabolism.

In **Chapter 5**, we report studies toward understanding the role of anaerobic Hyp metabolic pathway in gut bacteria. The topics listed in this chapter can be broadly categorized into two sections based on studies in the phylogenetically distant bacterial orders, Clostridiales and Bacteroidales. Genetic deletions in *C. difficile* established HypD as essential for Hyp fermentation. Analysis of comparative transcriptomics of *C. difficile* grown on Pro or Hyp as a source of electron acceptor is provided. Finally, preliminary attempts at detecting Hyp metabolism in gut Bacteroidales are described. The putative physiological roles of this pathway in the context of microbial communities and host-microbe interactions are proposed and discussed along with potential directions for future research.

1.10References

1. YY Huang, A Martinez-Del Campo, EP Balskus. Anaerobic 4-hydroxyproline utilization: Discovery of a new glycyl radical enzyme in the human gut microbiome uncovers a widespread microbial metabolic activity. *Gut microbes*. **2018**, 9(5):437-51.
2. R Sender, S Fuchs, R Milo. Revised estimates for the number of human and bacteria cells in the body. *PLOS Biol*. **2016**, 14(8):e1002533.
3. C Tropini, KA Earle, KC Huang, JL Sonnenburg. The gut microbiome: Connecting spatial organization to function. *Cell Host & Microbe*. **2017**, 21(4):433-42.
4. THMP Consortium. Structure, function and diversity of the healthy human microbiome. *Nature*. **2012**, 486(207-14).
5. J Lloyd-Price, A Mahurkar, G Rahnavard, J Crabtree, J Orvis, AB Hall, A Brady, HH Creasy, C McCracken, MG Giglio, et al. Strains, functions and dynamics in the expanded Human Microbiome Project. *Nature*. **2017**, 550(7674):61-6.
6. G Sharon, N Garg, J Debelius, R Knight, Pieter C Dorrestein, Sarkis K Mazmanian. Specialized metabolites from the microbiome in health and disease. *Cell Metabolism*. **2014**, 20(5):719-30.
7. D-W Kang, JG Park, ZE Ilhan, G Wallstrom, J LaBaer, JB Adams, R Krajmalnik-Brown. Reduced incidence of *Prevotella* and other fermenters in intestinal microflora of autistic children. *PLOS ONE*. **2013**, 8(7):e68322.

8. J Qin, R Li, J Raes, M Arumugam, KS Burgdorf, C Manichanh. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*. **2010**, 464.
9. R Joice, K Yasuda, A Shafquat, Xochitl C Morgan, C Huttenhower. Determining microbial products and identifying molecular targets in the human microbiome. *Cell Metabolism*. **2014**, 20(5):731-41.
10. CA Hutchison, R-Y Chuang, VNI Noskov, N Assad-Garcia, TJ Deerinck, MH Ellisman, J Gill, K Kannan, BJ Karas, L Ma, et al. Design and synthesis of a minimal bacterial genome. *Science*. **2016**, 351(6280).
11. AM Schnoes, SD Brown, I Dodevski, PC Babbitt. Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS Comp Biol*. **2009**, 5(12):e1000605.
12. N Koppel, Emily P Balskus. Exploring and understanding the biochemical diversity of the human microbiota. *Cell chemical biology*. **2016**, 23(1):18-30.
13. I Rowland, G Gibson, A Heinken, K Scott, J Swann, I Thiele, K Tuohy. Gut microbiota functions: metabolism of nutrients and other food components. *European Journal of Nutrition*. **2018**, 57(1):1-24.
14. EN Bergman. Energy contributions of volatile fatty acids from the gastrointestinal tract in various species. *Physiological reviews*. **1990**, 70(2):567-90.
15. LE Tailford, EH Crost, D Kavanaugh, N Juge. Mucin glycan foraging in the human gut microbiome. *Frontiers in Genetics*. **2015**, 6(81).
16. D Dodd, MH Spitzer, W Van Treuren, BD Merrill, AJ Hryckowian, SK Higginbottom, A Le, TM Cowan, GP Nolan, MA Fischbach, et al. A gut bacterial pathway metabolizes aromatic amino acids into nine circulating metabolites. *Nature*. **2017**, 551(7682):648-52.
17. CM Thomas, T Hong, JP van Pijkeren, P Hemarajata, DV Trinh, W Hu, RA Britton, M Kalkum, J Versalovic. Histamine derived from probiotic *Lactobacillus reuteri* suppresses TNF via modulation of PKA and ERK signaling. *PLOS ONE*. **2012**, 7(2):e31951.
18. AM Persico, V Napolioni. Urinary *p*-cresol in autism spectrum disorder. *Neurotoxicol Teratol*. **2013**, 36(82-90).
19. AS Devlin, A Marcobal, D Dodd, S Nayfach, N Plummer, T Meyer, KS Pollard, JL Sonnenburg, MA Fischbach. Modulation of a circulating uremic solute via rational genetic manipulation of the gut microbiota. *Cell host & microbe*. **2016**, 20(6):709-15.
20. CA Olson, HE Vuong, JM Yano, QY Liang, DJ Nusbaum, EY Hsiao. The gut microbiota mediates the anti-seizure effects of the ketogenic diet. *Cell*. **2018**, 173(7):1728-41.e13.

21. TR Sampson, JW Debelius, T Thron, S Janssen, GG Shastri, ZE Ilhan, C Challis, CE Schretter, S Rocha, V Gradinaru, et al. Gut microbiota regulate motor deficits and neuroinflammation in a model of Parkinson's disease. *Cell*. **2016**, 167(6):1469-80.e12.
22. X Zheng, G Xie, A Zhao, L Zhao, C Yao, NHL Chiu, Z Zhou, Y Bao, W Jia, JK Nicholson, et al. The footprints of gut microbial–mammalian co-metabolism. *J Proteome Res*. **2011**, 10(12):5512-22.
23. RLP Jump, A Polinkovsky, K Hurless, B Sitzlar, K Eckart, M Tomas, A Deshpande, MM Nerandzic, CJ Donskey. Metabolomics analysis identifies intestinal microbiota-derived biomarkers of colonization resistance in clindamycin-treated mice. *PLOS ONE*. **2014**, 9(7):e101267.
24. M Matsumoto, R Kibe, T Ooga, Y Aiba, S Kurihara, E Sawaki, Y Koga, Y Benno. Impact of intestinal microbiota on intestinal luminal metabolome. *Sci Rep*. **2012**, 2(233).
25. PH Weiss, L Klein. The quantitative relationship of urinary peptide hydroxyproline excretion to collagen degradation. *J Clin Invest*. **1969**, 48(1):1-10.
26. D Liu, M Nikoo, G Boran, P Zhou, JM Regenstein. Collagen and gelatin. *Annu Rev Food Sci Technol*. **2015**, 6(1):527-57.
27. K Iwai, T Hasegawa, Y Taguchi, F Morimatsu, K Sato, Y Nakamura, A Higashi, Y Kido, Y Nakabo, K Ohtsuki. Identification of food-derived collagen peptides in human blood after oral ingestion of gelatin hydrolysates. *Journal of agricultural and food chemistry*. **2005**, 53(16):6531-6.
28. AM Schnoes, SD Brown, I Dodevski, PC Babbitt. Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLOS Comput Biol*. **2009**, 5(12):e1000605.
29. T Selmer, AJ Pierik, J Heider. New glycyl radical enzymes catalysing key metabolic steps in anaerobic bacteria. *Biol Chem*. **2005**, 386(10):981-8.
30. LRF Backman, MA Funk, CD Dawson, CL Drennan. New tricks for the glycyl radical enzyme family. *Crit Rev Biochem Mol Biol*. **2017**, 54(6):674-95.
31. K Kurokawa, T Itoh, T Kuwahara, K Oshima, H Toh, A Toyoda, H Takami, H Morita, VK Sharma, TP Srivastava, et al. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res*. **2007**, 14(4):169-81.
32. K Ellrott, L Jaroszewski, W Li, JC Wooley, A Godzik. Expansion of the protein repertoire in newly explored environments: Human gut microbiome specific protein families. *PLOS Comput Biol*. **2010**, 6(6):e1000798.

33. CA Kolmeder, M de Been, J Nikkilä, I Ritamo, J Mättö, L Valmu, J Salojärvi, A Palva, A Salonen, WM de Vos. Comparative metaproteomics and diversity analysis of human intestinal microbiota testifies for its temporal stability and expression of core functions. *PLOS ONE*. **2012**, 7(1):e29913.
34. S Craciun, EP Balskus. Microbial conversion of choline to trimethylamine requires a glyceryl radical enzyme. *Proc Natl Acad Sci USA*. **2012**, 109(52):21307-12.
35. S Craciun, JA Marks, EP Balskus. Characterization of choline trimethylamine-lyase expands the chemistry of glyceryl radical enzymes. *ACS Chem Biol*. **2014**, 9(7):1408-13.
36. Z Wang, E Klipfell, BJ Bennett, R Koeth, BS Levison, B DuGar, AE Feldstein, EB Britt, X Fu, Y-M Chung, et al. Gut flora metabolism of phosphatidylcholine promotes cardiovascular disease. *Nature*. **2011**, 472(7341):57-63.
37. M-E Dumas, RH Barton, A Toye, O Cloarec, C Blancher, A Rothwell, J Fearnside, R Tatoud, V Blanc, JC Lindon, et al. Metabolic profiling reveals a contribution of gut microbiota to fatty liver phenotype in insulin-resistant mice. *Proc Natl Acad Sci USA*. **2006**, 103(33):12511-6.
38. T Selmer, PI Andrei. *p*-Hydroxyphenylacetate decarboxylase from *Clostridium difficile*. A novel glyceryl radical enzyme catalysing the formation of *p*-cresol. *Eur J Biochem*. **2001**, 268(5):1363-72.
39. TA Clayton, D Baker, JC Lindon, JR Everett, JK Nicholson. Pharmacometabonomic identification of a significant host-microbiome metabolic interaction affecting human drug metabolism. *Proc Natl Acad Sci USA*. **2009**, 106(34):14728-33.
40. LF Dawson, RA Stabler, BW Wren. Assessing the role of *p*-cresol tolerance in *Clostridium difficile*. *J Med Microbiol*. **2008**, 57(6):745-9.
41. KA Shisler, JB Broderick. Glyceryl radical activating enzymes: Structure, mechanism, and substrate interactions. *Archives of Biochemistry and Biophysics*. **2014**, 546(64-71).
42. MR Nnyepi, Y Peng, JB Broderick. Inactivation of *E. coli* pyruvate formate-lyase: Role of AdhE and small molecules. *Archives of Biochemistry and Biophysics*. **2007**, 459(1):1-9.
43. U Volker, NF A, K Joachim. The free radical of pyruvate formate-lyase. *European Journal of Biochemistry*. **1989**, 184(3):723-8.
44. M Horitani, K Shisler, WE Broderick, RU Hutcheson, KS Duschene, AR Marts, BM Hoffman, JB Broderick. Radical SAM catalysis via an organometallic intermediate with an Fe – [5' -C]-deoxyadenosyl bond. *Science*. **2016**, 352(6287):822.
45. AS Byer, H Yang, EC McDaniel, V Kathiresan, S Impano, A Pagnier, H Watts, C Denler, AL Vagstad, J Piel, et al. Paradigm shift for radical *S*-adenosyl-L-methionine reactions: The

organometallic intermediate Ω is central to catalysis. *Journal of the American Chemical Society*. **2018**, 140(28):8634-8638.

46. M Dong, V Kathiresan, MK Fenwick, AT Torelli, Y Zhang, JD Caranto, B Dzikovski, A Sharma, KM Lancaster, JH Freed, et al. Organometallic and radical intermediates reveal mechanism of diphthamide biosynthesis. *Science*. **2018**, 359(6381):1247.

47. KA Shisler, RU Hutcheson, M Horitani, KS Duschene, AV Crain, AS Byer, EM Shepard, A Rasmussen, J Yang, WE Broderick, et al. Monovalent cation activation of the radical SAM enzyme pyruvate formate-lyase activating enzyme. *Journal of the American Chemical Society*. **2017**, 139(34):11803-13.

48. N Shibata, J Masuda, T Tobimatsu, T Toraya, K Suto, Y Morimoto, N Yasuoka. A new mode of B₁₂ binding and the direct participation of a potassium ion in enzyme catalysis: X-ray structure of diol dehydratase. *Structure*. **1999**, 7(8):997-1008.

49. PA Schwartz, PA Frey. Dioldehydrase: an essential role for potassium ion in the homolytic cleavage of the cobalt-carbon bond in adenosylcobalamin. *Biochemistry*. **2007**, 46(24):7293-301.

50. BM Martins, M Blaser, M Feliks, GM Ullmann, W Buckel, T Selmer. Structural basis for a Kolbe-type decarboxylation catalyzed by a glycyl radical enzyme. *Journal of the American Chemical Society*. **2011**, 133(37):14666-74.

51. MA Funk, ET Judd, ENG Marsh, SJ Elliott, CL Drennan. Structures of benzylsuccinate synthase elucidate roles of accessory subunits in glycyl radical enzyme activation and activity. *Proceedings of the National Academy of Sciences*. **2014**, 111(28):10161.

52. MA Funk, ENG Marsh, CL Drennan. Substrate-bound structures of benzylsuccinate synthase reveal how toluene is activated in anaerobic hydrocarbon degradation. *J Biol Chem*. **2015**, 290(37):22398-408.

53. V-M Leppänen, MC Merckel, DL Ollis, KK Wong, JW Kozarich, A Goldman. Pyruvate formate lyase is structurally homologous to type I ribonucleotide reductase. *Structure*. **1999**, 7(7):733-44.

54. DT Logan, J Andersson, B-M Sjöberg, P Nordlund. A glycyl radical site in the crystal structure of a class III ribonucleotide reductase
Science. **1999**, 283(5407):1499.

55. S Bodea, MA Funk, EP Balskus, CL Drennan. Molecular basis of C-N bond cleavage by the glycyl radical enzyme choline trimethylamine-lyase. *Cell chemical biology*. **2016**, 23(10):1206-16.

56. JR O'Brien, C Raynaud, C Croux, L Girbal, P Soucaille, WN Lanzilotta. Insight into the mechanism of the B₁₂-independent glycerol dehydratase from *Clostridium butyricum*: preliminary biochemical and structural characterization. *Biochemistry*. **2004**, 43:4635-45.

57. DP Dowling, AK Croft, CL Drennan. Radical use of Rossmann and TIM barrel architectures for controlling coenzyme B₁₂ chemistry. *Annual review of biophysics*. **2012**, 41:403-27.
58. DP Dowling, JL Vey, AK Croft, CL Drennan. Structural diversity in the AdoMet radical enzyme superfamily. *Biochim Biophys Acta*. **2012**, 1824(11):1178-95.
59. JL Vey, J Yang, M Li, WE Broderick, JB Broderick, CL Drennan. Structural basis for glycyl radical formation by pyruvate formate-lyase activating enzyme. *Proc Natl Acad Sci USA*. **2008**, 105(42):16137-41.
60. Y Peng, SE Veneziano, GD Gillispie, JB Broderick. Pyruvate formate-lyase: Evidence for an open conformation favored in the presence of its activating enzyme. *Journal of Biological Chemistry*. **2010**, 285(35):27224-31.
61. ND Lanz, SJ Booker. Auxiliary iron–sulfur cofactors in radical SAM enzymes. *Biochimica et Biophysica Acta*. **2015**, 1853(6):1316-34.
62. B Selvaraj, AJ Pierik, E Bill, BM Martins. The ferredoxin-like domain of the activating enzyme is required for generating a lasting glycyl radical in 4-hydroxyphenylacetate decarboxylase. *Journal of biological inorganic chemistry : JBIC : a publication of the Society of Biological Inorganic Chemistry*. **2014**, 19(8):1317-26.
63. F Sievers, A Wilm, D Dineen, TJ Gibson, K Karplus, W Li, R Lopez, H McWilliam, M Remmert, J Söding, et al. Fast, scalable generation of high - quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. **2011**, 7(539).
64. M Kearse, R Moir, A Wilson, S Stones-Havas, M Cheung, S Sturrock, S Buxton, A Cooper, S Markowitz, C Duran, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. **2012**, 28(12):1647-9.
65. P Reichard. The evolution of ribonucleotide reduction. *Trends Biochem Sci*. **1997**, 22(3):81-5.
66. D Lundin, G Berggren, DT Logan, B-M Sjöberg. The origin and evolution of ribonucleotide reduction. *Life*. **2015**, 5(1):604-36.
67. J Knappe, HP Blaschkowski, P GrÖBner, T Schmttt. Pyruvate formate-lyase of *Escherichia coli*: the acetyl-enzyme intermediate. *European Journal of Biochemistry*. **1974**, 50(1):253-63.
68. BJ Levin*, YY Huang*, SC Peck, Y Wei, A Martínez-del Campo, JA Marks, EA Franzosa, C Huttenhower, EP Balskus. A prominent glycyl radical enzyme in human gut microbiomes metabolizes *trans*-4-hydroxy-L-proline. *Science*. **2017**, 355(6325):eaai8386.
69. L Lehtiö, A Goldman. The pyruvate formate lyase family: sequences, structures and activation. *PEDS*. **2004**, 17(6):545-52.

70. G Sawers, C Heßlinger, N Muller, M Kaiser. The glycyl radical enzyme TdcE can replace pyruvate formate-lyase in glucose fermentation. *Journal of Bacteriology*. **1998**, 180(14):3509-16.
71. J Knappe, FA Neugebauer, HP Blaschkowski, M Ganzler. Post-translational activation introduces a free radical into pyruvate formate-lyase. *Proceedings of the National Academy of Sciences of the United States of America*. **1984**, 81(5):1332-5.
72. AF Wagner, M Frey, FA Neugebauer, W Schäfer, J Knappe. The free radical in pyruvate formate-lyase is located on glycine-734. *Proc Natl Acad Sci USA*. **1992**, 89(3):996-1000.
73. A Becker, K Fritz-Wolf, W Kabsch, J Knappe, S Schultz, AF Volker Wagner. Structure and mechanism of the glycyl radical enzyme pyruvate formate-lyase. *Nat Struct Mol Biol*. **1999**, 6(10):969-75.
74. A Becker, W Kabsch. X-ray structure of pyruvate formate-lyase in complex with pyruvate and CoA: How the enzyme uses the Cys-418 thiyl radical for pyruvate cleavage. *Journal of Biological Chemistry*. **2002**, 277(42):40036-42.
75. G Sawers, A Bock. Anaerobic regulation of pyruvate formate-lyase from *Escherichia coli* K-12. *J Bacteriol*. **1988**, 170(11):5330-6.
76. AFV Wagner, S Schultz, J Bomke, T Pils, WD Lehmann, J Knappe. YfiD of *Escherichia coli* and Y06I of bacteriophage T4 as autonomous glycyl radical cofactors reconstituting the catalytic center of oxygen-fragmented pyruvate formate-lyase. *Biochem Biophys Res Comm*. **2001**, 285(2):456-62.
77. NR Wyborn, SL Messenger, RA Henderson, G Sawers, RE Roberts, MM Attwood, J Green. Expression of the *Escherichia coli* yfiD gene responds to intracellular pH and reduces the accumulation of acidic metabolic end products. *Microbiology*. **2002**, 148(Pt 4):1015-26.
78. D Kessler, W Herth, J Knappe. Ultrastructure and pyruvate formate-lyase radical quenching property of the multienzymic AdhE protein of *Escherichia coli*. *Journal of Biological Chemistry*. **1992**, 267(25):18073-9.
79. C Heßlinger, A Fairhurst Shirley, G Sawers. Novel keto acid formate-lyase and propionate kinase enzymes are components of an anaerobic pathway in *Escherichia coli* that degrades L-threonine to propionate. *Molecular Microbiology*. **2002**, 27(2):477-92.
80. J Andersson, M Westman, A Hofer, B-M Sjöberg. Allosteric regulation of the class III anaerobic ribonucleotide reductase from bacteriophage T4. *Journal of Biological Chemistry*. **2000**, 275(26):19443-8.
81. J Stubbe. Ribonucleotide reductases in the twenty-first century. *Proceedings of the National Academy of Sciences*. **1998**, 95(6):2723.

82. R Lenz, B Giese. Studies on the mechanism of ribonucleotide reductases. *Journal of the American Chemical Society*. **1997**, 119(12):2784-94.
83. E Torrents, P Aloy, I Gibert, F Rodriguez-Trelles. Ribonucleotide reductases: divergent evolution of an ancient enzyme. *Journal of molecular evolution*. **2002**, 55(2):138-52.
84. Y Wei, G Mathies, K Yokoyama, J Chen, RG Griffin, J Stubbe. A chemically competent thiosulfuranyl radical on the *Escherichia coli* Class III ribonucleotide reductase. *Journal of the American Chemical Society*. **2014**, 136(25):9001-13.
85. O Aurelius, R Johansson, V Bågenholm, D Lundin, F Tholander, A Balhuizen, T Beck, M Sahlin, B-M Sjöberg, E Mulliez, et al. The crystal structure of *Thermotoga maritima* class III ribonucleotide reductase lacks a radical cysteine pre-positioned in the active site. *PLOS ONE*. **2015**, 10(7):e0128199.
86. Y Wei, MA Funk, LA Rosado, J Baek, CL Drennan, J Stubbe. The class III ribonucleotide reductase from *Neisseria bacilliformis* can utilize thioredoxin as a reductant. *Proceedings of the National Academy of Sciences*. **2014**, 111(36):E3756-E65.
87. Y Wei, B Li, D Prakash, JG Ferry, SJ Elliott, J Stubbe. A ferredoxin disulfide reductase delivers electrons to the *Methanosarcina barkeri* class III ribonucleotide reductase. *Biochemistry*. **2015**, 54(47):7019-28.
88. E Mulliez, S Ollagnier, M Fontecave, R Eliasson, P Reichard. Formate is the hydrogen donor for the anaerobic ribonucleotide reductase from *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America*. **1995**, 92(19):8759-62.
89. J Knappe, G Sawers. A radical-chemical route to acetyl-CoA: the anaerobically induced pyruvate formate-lyase system of *Escherichia coli*. *FEMS Microbiol Rev*. **1990**, 6(4):383-98.
90. Y Wei. Mechanistic investigations of class III anaerobic ribonucleotide reductases. *Department of Chemistry*. Massachusetts Institute of Technology. **2015**, PhD.
91. L D'Ari, HA Barker. p-Cresol formation by cell-free extracts of *Clostridium difficile*. *Arch Microbiol*. **1985**, 143(3):311-2.
92. B Selvaraj, W Buckel, BT Golding, GM Ullmann, BM Martins. Structure and function of 4-hydroxyphenylacetate decarboxylase and its cognate activating enzyme. *Journal of molecular microbiology and biotechnology*. **2016**, 26(1-3):76-91.
93. M Feliks, BM Martins, GM Ullmann. Catalytic mechanism of the glycy radical enzyme 4-hydroxyphenylacetate decarboxylase from continuum electrostatic and QC/MM calculations. *J Am Chem Soc*. **2013**, 135(39):14574-85.
94. SR Elsdén, MG Hilton, JM Waller. The end products of the metabolism of aromatic amino acids by clostridia. *Archives of Microbiology*. **1976**, 107(3):283-8.

95. S Hafiz, CL Oakley. *Clostridium difficile*: isolation and characteristics (Plate VIII). *Journal of Medical Microbiology*. **1976**, 9(2):129-36.
96. LF Dawson, EH Donahue, ST Cartman, RH Barton, J Bundy, R McNerney, NP Minton, BW Wren. The analysis of para-cresol production and tolerance in *Clostridium difficile* 027 and 012 strains. *BMC Microbiology*. **2011**, 11(1):86.
97. R Vanholder, B Bammens, H de Loor, G Glorieux, B Meijers, E Schepers, Z Massy, P Evenepoel. Warning: the unfortunate end of p-cresol as a uraemic toxin. *Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association*. **2011**, 26(5):1464-7.
98. F Jüttner, JJ Henatsch. Anoxic hypolimnion is a significant source of biogenic toluene. *Nature*. **1986**, 323:797.
99. K Zargar, R Saville, RM Phelan, SG Tringe, CJ Petzold, JD Keasling, HR Beller. *In vitro* characterization of phenylacetate decarboxylase, a novel enzyme catalyzing toluene biosynthesis in an anaerobic microbial community. *Scientific Reports*. **2016**, 6:31362.
100. HR Beller, AV Rodrigues, K Zargar, Y-W Wu, AK Saini, RM Saville, JH Pereira, PD Adams, SG Tringe, CJ Petzold, et al. Discovery of enzymes for toluene synthesis from anoxic microbial communities. *Nature Chemical Biology*. **2018**, 14(5):451-7.
101. TR Whitehead, NP Price, HL Drake, MA Cotta. Catabolic pathway for the production of skatole and indoleacetic acid by the acetogen *Clostridium drakei*, *Clostridium scatologenes*, and swine manure. *Applied and Environmental Microbiology*. **2008**, 74(6):1950-3.
102. DC Honeyfield, JR Carlson. Assay for the enzymatic conversion of indoleacetic acid to 3-methylindole in a ruminal *Lactobacillus* species. *Applied and Environmental Microbiology*. **1990**, 56(3):724-9.
103. D Liu, Y Wei, X Liu, Y Zhou, L Jiang, J Yin, F Wang, Y Hu, AN Nanjaraj Urs, Y Liu, et al. Indoleacetate decarboxylase is a glycyl radical enzyme catalysing the formation of malodorous skatole. *Nature Communications*. **2018**, 9(1):4224.
104. A Strijkstra, K Trautwein, R Jarling, L Wöhlbrand, M Dörries, R Reinhardt, M Drozdowska, BT Golding, H Wilkes, R Rabus. Anaerobic activation of p-cymene in denitrifying betaproteobacteria: Methyl group hydroxylation versus addition to fumarate. *Applied and Environmental Microbiology*. **2014**, 80(24):7592-603.
105. AV Callaghan, B Wawrik, SM Ní Chadhain, LY Young, GJ Zylstra. Anaerobic alkane-degrading strain AK-01 contains two alkylsuccinate synthase genes. *Biochemical and Biophysical Research Communications*. **2008**, 366(1):142-8.

106. T Biegert, G Fuchs, J Heider. Evidence that anaerobic oxidation of toluene in the denitrifying bacterium *Thauera aromatica* is initiated by formation of benzylsuccinate from toluene and fumarate. *Eur J Biochem.* **1996**, 238(3):661-8.
107. A Acosta-González, R Rosselló-Móra, S Marqués. Diversity of benzylsuccinate synthase-like (bssA) genes in hydrocarbon-polluted marine sediments suggests substrate-dependent clustering. *Applied and Environmental Microbiology.* **2013**, 79(12):3667-76.
108. A Acosta-Gonzalez, R Rossello-Mora, S Marques. Characterization of the anaerobic microbial community in oil-polluted subtidal sediments: aromatic biodegradation potential after the Prestige oil spill. *Environ Microbiol.* **2013**, 15(1):77-92.
109. G Harms, R Rabus, F Widdel. Anaerobic oxidation of the aromatic plant hydrocarbon p-cymene by newly isolated denitrifying bacteria. *Archives of Microbiology.* **1999**, 172(5):303-12.
110. D Selesi, N Jehmlich, M von Bergen, F Schmidt, T Rattei, P Tischler, T Lueders, RU Meckenstock. Combined genomic and proteomic approaches identify gene clusters involved in anaerobic 2-methylnaphthalene degradation in the sulfate-reducing enrichment culture N47. *J Bacteriol.* **2010**, 192(1):295-306.
111. L Wöhlbrand, H Jacob Jacob, M Kube, M Mussmann, R Jarling, A Beck, R Amann, H Wilkes, R Reinhardt, R Rabus. Complete genome, catabolic sub-proteomes and key-metabolites of *Desulfobacula toluolica* Tol2, a marine, aromatic compound-degrading, sulfate-reducing bacterium. *Environmental Microbiology.* **2012**, 15(5):1334-55.
112. O Grundmann, A Behrends, R Rabus, J Amann, T Halder, J Heider, F Widdel. Genes encoding the candidate enzyme for anaerobic activation of n-alkanes in the denitrifying bacterium, strain HxN1. *Environmental Microbiology.* **2007**, 10(2):376-85.
113. JW LaMattina, ND Keul, P Reitzer, S Kapoor, F Galzerani, DJ Koch, IE Gouvea, WN Lanzilotta. 1,2-propanediol dehydration in *Roseburia inulinivorans*; structural basis for substrate and enantiomer selectivity. *J Biol Chem.* **2016**, 291(30):15515-26.
114. BJ Levin, EP Balskus. characterization of 1,2-propanediol dehydratases reveals distinct mechanisms for B₁₂-dependent and glycyl radical enzymes. *Biochemistry.* **2018**, 57(23):3222-6.
115. C Raynaud, P Sarçabal, I Meynial-Salles, C Croux, P Soucaille. Molecular characterization of the 1,3-propanediol (1,3-PD) operon of *Clostridium butyricum*. *Proc Natl Acad Sci USA.* **2003**, 100(9):5010-5.
116. J C. Thompson, B B. He. Characterization of crude glycerol from biodiesel production from multiple feedstocks. *Applied Engineering in Agriculture.* **2006**, 22(2):261-5.
117. JZ Liu, W Xu, A Chistoserdov, RK Bajpai. Glycerol dehydratases: Biochemical structures, catalytic mechanisms, and industrial applications in 1,3-propanediol production by naturally

occurring and genetically engineered bacterial strains. *Applied biochemistry and biotechnology*. **2016**, 179(6):1073-100.

118. M Feliks, GM Ullmann. Glycerol dehydration by the B₁₂-independent enzyme may not involve the migration of a hydroxyl group: a computational study. *J Phys Chem B*. **2012**, 116(24):7076-87.

119. KP Scott, JC Martin, G Campbell, C-D Mayer, HJ Flint. Whole-genome transcription profiling reveals genes up-regulated by growth on fucose in the human gut bacterium *Roseburia inulinivorans*. *J Bacteriol*. **2006**, 188(12):4340-9.

120. N Reichardt, SH Duncan, P Young, A Belenguer, C McWilliam Leitch, KP Scott, HJ Flint, P Louis. Phylogenetic distribution of three pathways for propionate production within the human gut microbiota. *ISME J*. **2014**, 8(6):1323-35.

121. A Martínez-del Campo, S Bodea, HA Hamer, JA Marks, HJ Haiser, PJ Turnbaugh, EP Balskus. Characterization and detection of a widely distributed gene cluster that predicts anaerobic choline utilization by human gut bacteria. *mBio*. **2015**, 6(2):e00042-15.

122. KA Romano, A Martinez-del Campo, K Kasahara, CL Chittim, EI Vivas, D Amador-Noguez, EP Balskus, FE Rey. Metabolic, epigenetic, and transgenerational effects of gut bacterial choline consumption. *Cell Host & Microbe*. **2017**, 22(3):279-90.

123. RJ Mackay, CJ McEntyre, C Henderson, M Lever, PM George. Trimethylaminuria: Causes and Diagnosis of a Socially Distressing Condition. *The Clinical Biochemist Reviews*. **2011**, 32(1):33-43.

124. Z Wang, Adam B Roberts, Jennifer A Buffa, Bruce S Levison, W Zhu, E Org, X Gu, Y Huang, M Zamanian-Daryoush, Miranda K Culley, et al. Non-lethal inhibition of gut microbial trimethylamine production for the treatment of atherosclerosis. *Cell*. **2015**, 163(7):1585-95.

125. AB Roberts, X Gu, JA Buffa, AG Hurd, Z Wang, W Zhu, N Gupta, SM Skye, DB Cody, BS Levison, et al. Development of a gut microbe-targeted nonlethal therapeutic to inhibit thrombosis potential. *Nature Medicine*. **2018**, 24(9):1407-17.

126. J Zarzycki, O Erbilgin, CA Kerfeld. Bioinformatic characterization of glyceryl radical enzyme-associated bacterial microcompartments. *Applied and Environmental Microbiology*. **2015**, 81(24):8315-29.

127. J Rétey, A Umani-Ronchi, J Seibl, D Arigoni. Zum mechanismus der propandioldehydrase-reaktion. *Experientia* **1966**, 22:502-3.

128. B Zagalak, PA Frey, GL Karabatsos, RH Abeles. The Stereochemistry of the Conversion of D and L 1,2-Propanediols to Propionaldehyde. *Journal of Biological Chemistry*. **1966**, 241(13):3028-35.

129. B Kovacevic, D Baric, D Babic, L Bilic, M Hanzevacki, GM Sandala, L Radom, DM Smith. Computational tale of two enzymes: Glycerol dehydration with or without B₁₂. *J Am Chem Soc.* **2018**, 140(27):8487-96.
130. SD Wetmore, DM Smith, BT Golding, L Radom. Interconversion of (*S*)-glutamate and (2*S*,3*S*)-3-methylaspartate: a distinctive B₁₂-dependent carbon-skeleton rearrangement. *J Am Chem Soc.* **2001**, 123(33):7963-72.
131. Y-R Luo. Comprehensive handbook of chemical bond energies. CRC Press CRC Press 2007.
132. K Mori, T Oiwa, S Kawaguchi, K Kondo, Y Takahashi, T Toraya. Catalytic roles of substrate-binding residues in coenzyme B₁₂-dependent ethanolamine ammonia-lyase. *Biochemistry.* **2014**, 53(16):2661-71.
133. Y Liu, AA Gallo, J Florian, YS Liu, S Mora, W Xu. QM/MM (ONIOM) study of glycerol binding and hydrogen abstraction by the coenzyme B₁₂-independent dehydratase. *The journal of physical chemistry B.* **2010**, 114(16):5497-502.
134. KL Gorres, RT Raines. Prolyl 4-hydroxylase. *Crit Rev Biochem Mol Biol.* **2010**, 45(2):106-24.
135. M Hijazi, SM Velasquez, E Jamet, JM Estevez, C Albenne. An update on post-translational modifications of hydroxyproline-rich glycoproteins: toward a model highlighting their contribution to plant cell wall architecture. *Front Plant Sci.* **2014**, 5:395.
136. E Nguema-Ona, M Vicré-Gibouin, M Gotté, B Plancot, P Lerouge, M Bardor, A Driouich. Cell wall *O*-glycoproteins and *N*-glycoproteins: aspects of biosynthesis and function. *Front Plant Sci.* **2014**, 5(499).
137. KH Kim, B Jia, CO Jeon. Identification of *trans*-4-hydroxy-L-proline as a compatible solute and its biosynthesis and molecular characterization in *Halobacillus Halophilus*. *Front Microbiol.* **2017**, 8:2054.
138. E Katz, DJ Prockop, S Udenfriend. Precursors of the hydroxyproline and ketoproline in actinomycin. *J Biol Chem.* **1962**, 237(5):1585-8.
139. CC Lawrence, WJ Sobey, RA Field, JE Baldwin, CJ Schofield. Purification and initial characterization of proline 4-hydroxylase from *Streptomyces griseoviridis* P8648: a 2-oxoacid, ferrous-dependent dioxygenase involved in etamycin biosynthesis. *Biochem J.* **1996**, 313(Pt 1):185-91.
140. JS Scotti, IKH Leung, W Ge, MA Bentley, J Paps, HB Kramer, J Lee, W Aik, H Choi, SM Paulsen, et al. Human oxygen sensing may have origins in prokaryotic elongation factor Tu prolyl-hydroxylation. *Proc Natl Acad Sci USA.* **2014**, 111(37):13331-6.

141. NJ Schnicker, M Dey. *Bacillus anthracis* prolyl 4-hydroxylase modifies collagen-like substrates in asymmetric patterns. *J Biol Chem.* **2016**, 291(25):13360-74.
142. MA Culpepper, EE Scott, J Limburg. Crystal structure of prolyl 4-hydroxylase from *Bacillus anthracis*. *Biochemistry.* **2010**, 49(1):124-33.
143. NJ Schnicker, M Razzaghi, S Guha Thakurta, S Chakravarthy, M Dey. *Bacillus anthracis* prolyl 4-hydroxylase interacts with and modifies elongation factor Tu. *Biochemistry.* **2017**, 56(43):5771-85.
144. JP Wold, F Lundby, B Egelanddal. Quantification of connective tissue (hydroxyproline) in ground beef by autofluorescence spectroscopy. *J Food Sci.* **1999**, 64(3):377-83.
145. FA Vázquez - Ortíz, OE Morón - Fuenmayor, NF González - Méndez. Hydroxyproline measurement by HPLC: Improved method of total collagen determination in meat samples. *J Liq Chrom Relat Tech.* **2004**, 27(17):2771-80.
146. JAM Ramshaw, NK Shah, B Brodsky. Gly-X-Y tripeptide frequencies in collagen: A context for host-guest triple-helical peptides. *Journal of Structural Biology.* **1998**, 122(1):86-91.
147. C Valiente, E Arrigoni, RM Esteban, R Amado. Grape pomace as a potential food fiber. *J Food Sci.* **1995**, 60(4):818-20.
148. D Verbeken, S Dierckx, K Dewettinck. Exudate gums: occurrence, production, and applications. *Appl Microbiol Biotechnol.* **2003**, 63(1):10-21.
149. GD Wu, C Compher, EZ Chen, SA Smith, RD Shah, K Bittinger, C Chehoud, LG Albenberg, L Nessel, E Gilroy, et al. Comparative metabolomics in vegans and omnivores reveal constraints on diet-dependent gut microbiota metabolite production. *Gut.* **2016**, 65(1):63-72.
150. KI Kivirikko. Urinary excretion of hydroxyproline in health and disease. *Int Rev Connect Tissue Res.* **1970**, 5:93-163.
151. EA Smith, GT Macfarlane. Enumeration of amino acid fermenting bacteria in the human large intestine: effects of pH and starch on peptide metabolism and dissimilation of amino acids. *FEMS Microbiol Ecol.* **1998**, 25(4):355-68.
152. E Adams, L Frank. Metabolism of proline and the hydroxyprolines. *Annu Rev Biochem.* **1980**, 49(1):1005-61.
153. E Salido, AL Pey, R Rodriguez, V Lorenzo. Primary hyperoxalurias: Disorders of glyoxylate detoxification. *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease.* **2012**, 1822(9):1453-64.
154. J Knight, J Jiang, DG Assimos, RP Holmes. Hydroxyproline ingestion and urinary oxalate and glycolate excretion. *Kidney Int.* **2006**, 70(11):1929-34.

155. Candice B Summitt, Lynnette C Johnson, Thomas J Jönsson, D Parsonage, Ross P Holmes, WT Lowther. Proline dehydrogenase 2 (PRODH2) is a hydroxyproline dehydrogenase (HYPDH) and molecular target for treating primary hyperoxaluria. *Biochemical Journal*. **2015**, 466(2):273.
156. X Li, J Knight, S Fargue, B Buchalski, Z Guan, EW Inscho, A Liebow, K Fitzgerald, W Querbes, WT Lowther, et al. Metabolism of $^{13}\text{C}_5$ -hydroxyproline in mouse models of Primary Hyperoxaluria and its inhibition by RNAi therapeutics targeting liver glycolate oxidase and hydroxyproline dehydrogenase. *Biochimica et biophysica acta*. **2016**, 1862(2):233-9.
157. Y Shigemura, S Akaba, E Kawashima, EY Park, Y Nakamura, K Sato. Identification of a novel food-derived collagen peptide, hydroxyprolyl-glycine, in human peripheral blood by pre-column derivatisation with phenyl isothiocyanate. *Food chemistry*. **2011**, 129(3):1019-24.
158. LH Stickland. Studies in the metabolism of the strict anaerobes (genus *Clostridium*). I. The chemical reactions by which *C. sporogenes* obtains its energy. *Biochem J*. **1934**, 28(5):1746-59.
159. N Fonknechten, S Chaussonnerie, S Tricot, A Lajus, JR Andreesen, N Perchat, E Pelletier, M Gouyvenoux, V Barbe, M Salanoubat, et al. *Clostridium sticklandii*, a specialist in amino acid degradation: revisiting its metabolism through its genome sequence. *BMC Genomics*. **2010**, 11(555).
160. S Jackson, M Calos, A Myers, WT Self. Analysis of proline reduction in the nosocomial pathogen *Clostridium difficile*. *J Bacteriol*. **2006**, 188(24):8487-95.
161. RW Lovitt, DB Kell, JG Morris. Proline reduction by *Clostridium sporogenes* is coupled to vectorial proton ejection. *FEMS Microbiology Letters*. **1986**, 36(2-3):269-73.
162. UC Kabisch, A Gräntzdörffer, A Schierhorn, KP Rücknagel, JR Andreesen, A Pich. Identification of D-proline reductase from *Clostridium sticklandii* as a selenoenzyme and indications for a catalytically active pyruvoyl group derived from a cysteine residue by cleavage of a proprotein. *J Biol Chem*. **1999**, 274(13):8445-54.
163. DC Turner, TC Stadtman. Purification of protein components of the clostridial glycine reductase system and characterization of protein A as a selenoprotein. *Archives of Biochemistry and Biophysics*. **1973**, 154(1):366-81.
164. JR Andreesen. Glycine reductase mechanism. *Current Opinion in Chemical Biology*. **2004**, 8(5):454-61.
165. HA Barker, L D'Ari, J Kahn. Enzymatic reactions in the degradation of 5-aminovalerate by *Clostridium aminovalericum*. *J Biol Chem*. **1987**, 262(19):8994-9003.
166. JK Hardman, TC Stadtman. Metabolism of ω -amino acidS: II. Fermentation of δ -aminovaleric acid by *Clostridium aminovalericum* N. Sp. *J Bacteriol*. **1960**, 79(4):549-52.

167. W Buckel, PH Janssen, A Schuhmann, U Eikmanns, P Messner, U Sleytr, W Liesack. *Clostridium viride* sp. nov., a strictly anaerobic bacterium using 5-aminovalerate as growth substrate, previously assigned to *Clostridium aminovalericum*. *Archives of Microbiology*. **1994**, 162(6):387-94.
168. U Eikmanns, W Buckel. Crystalline green 5-hydroxyvaleryl-CoA dehydratase from *Clostridium aminovalericum*. *European Journal of Biochemistry*. **1991**, 197(3):661-8.
169. U Eikmanns, W Buckel. A green 2,4-pentadienoyl-CoA reductase from *Clostridium aminovalericum*. *European Journal of Biochemistry*. **1991**, 198(1):263-6.
170. W Buckel. Unusual enzymes involved in five pathways of glutamate fermentation. *Applied microbiology and biotechnology*. **2001**, 57(3):263-73.
171. HP de Vlardar. Amino acid fermentation at the origin of the genetic code. *Biology Direct*. **2012**, 7:6.
172. A Parthasarathy, N Chowdhury. Microbial degradation of amino acids in anoxic environments. In: D'Mello J, editors. *The Handbook of Microbial Metabolism of Amino Acids*. 2017. p. 398-417.
173. E Adams. Hydroxyproline metabolism: I. Conversion to α -ketoglutarate by extracts of *Pseudomonas*. *J Biol Chem*. **1959**, 234(8):2073-84.
174. S Watanabe, D Morimoto, F Fukumori, H Shinomiya, H Nishiwaki, M Kawano-Kawada, Y Sasai, Y Tozawa, Y Watanabe. Identification and characterization of D-hydroxyproline dehydrogenase and Δ^1 -pyrroline-4-hydroxy-2-carboxylate deaminase involved in novel L-hydroxyproline metabolism of bacteria. *J Biol Chem*. **2012**, 287(39):32674-88.
175. S Chen, CE White, GC diCenzo, Y Zhang, PJ Stogios, A Savchenko, TM Finan. L-Hydroxyproline and D-proline catabolism in *Sinorhizobium meliloti*. *J Bacteriol*. **2016**, 198(7):1171-81.
176. C Blumer, D Haas. Mechanism, regulation, and ecological role of bacterial cyanide biosynthesis. *Arch Microbiol*. **2000**, 173(3):170-7.
177. SF Altschul, W Gish, W Miller, EW Myers, DJ Lipman. Basic local alignment search tool. *J Mol Biol*. **1990**, 215(3):403-10.

2 Chapter 2. Discovery and characterization of a *trans*-4-hydroxy-L-proline dehydratase

Parts of this chapter were adapted from published work.^{1,2}

2.1 Introduction

Recent research in the gut microbiome field has highlighted the potential for this community to influence human health and disease.^{3,4} Unfortunately, we know surprisingly little about specific microbial biochemical reactions and the mechanisms by which this chemistry shapes the gut community and influences host biology.⁵ To gain an understanding of the metabolic potential of the gut microbiome and inform mechanistic research, it is imperative that we link genes to functions in this habitat. However, the majority of microbial genes remain unidentified, illustrating a great need to investigate this uncharacterized genetic information.^{6,7}

Metagenomic and metaproteomic studies have observed protein annotations for glycine radical domains and/or pyruvate formate-lyases to be enriched in the gut microbiome.⁸⁻¹⁰ Both of these annotations represent proteins part of the glycyl radical enzyme (GRE) superfamily. Because GREs have high sequence similarities, homology-based annotations overestimate pyruvate formate-lyase activity in sequencing studies. A major portion of these annotated genes likely encode GREs of unknown functions dissimilar from those catalyzed by formate-lyases. The few GRE activities known to date span a diverse range of chemically challenging transformations, enabled by the utilization of radical chemistry.^{11,12} Activation of GREs by cognate activating enzymes (AEs) is required for glycyl radical formation, the mechanism of which is provided in detail in Chapter 1. The glycine-centered radical in activated GREs and the [4Fe-4S] cluster coordinated in AEs are deactivated by stoichiometric amounts of molecular oxygen. As a result, GREs are strictly found in facultative and obligate anaerobic microorganisms and participate in

anaerobic primary metabolism. In the GRE superfamily, five types of reactivities have been discovered with most GREs having unknown functions. However, technical challenges of working with GREs and AEs often limit the study of these enzymes *in vitro*.

2.1.1 Discovery of a new glycyl radical enzyme predicted to be a *trans*-4-hydroxy-L-proline dehydratase

An uncharacterized GRE was first brought to our attention by Dr. Yifeng Wei from Prof. JoAnne Stubbe's lab at MIT.ⁱ He proposed this GRE to be a novel *trans*-4-hydroxy-L-proline (Hyp) dehydratase (HypD) based on initial insights gained from multiple sequence alignments and its genomic context in Clostridiales. A homology model was constructed to further support this hypothesis. This, in combination with a multiple sequence alignment of HypD with characterized GREs, revealed an active site resembling those of GRE eliminases GD, PD, and CutC (discussed in Section 2.2.1). Additionally, a gene annotated as (*S*)- Δ^1 -pyrroline-5-carboxylate (P5C) reductase (P5CR) is colocalized with the GRE gene in many sequenced genomes of Clostridiales (Figure 2.1A). P5CR catalyzes the reduction of P5C to L-proline (Pro) as the final step in Pro biosynthesis.^{13,14} Clustering of genes participating in the same pathway is a widespread feature of bacterial genomes that enables co-transcription under favorable conditions. Considering this genomic context, Yifeng considered the non-proteinogenic amino acid *trans*-4-hydroxy-L-proline (Hyp) as a putative substrate for the GRE. Given its high sequence similarity to the characterized GD (37% amino acid ID), this enzyme was hypothesized to catalyze dehydration of Hyp to P5C, which could then be reduced to Pro by P5CR (Figure 2.1B).

ⁱ Dr. Yifeng Wei formulated the hypothesis that this new GRE might catalyze dehydration of 4-hydroxyproline based on its genomic context in Clostridiales and conserved residues among GRE eliminases.

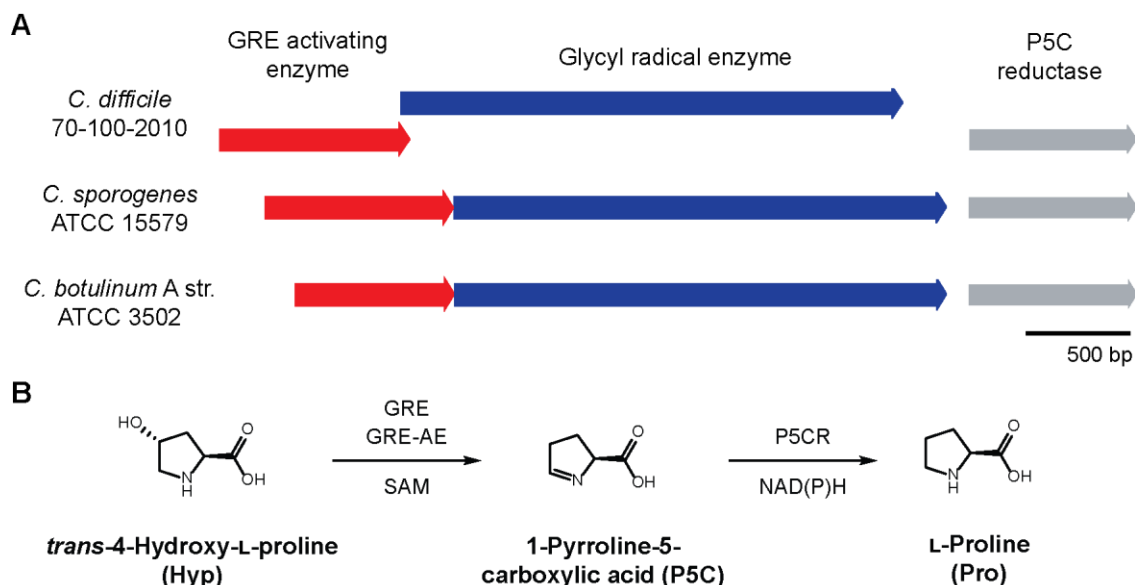


Figure 2.1 A conserved gene cluster identified in the order Clostridiales and hypothesis for its activities.

(A) An uncharacterized GRE was found in a gene cluster with a conserved genomic context across many species in the order Clostridiales. Genes encoding a putative AE and P5CR make up remaining genes in this cluster. (B) Proposed activities for genes in this conserved gene cluster.

Dehydration of Hyp was an unprecedented enzymatic activity that is chemically difficult to achieve solely through the acid-base chemistry performed by most enzymes. Hydroxyproline dehydratases characterized at the time were only known to accept 3-hydroxyproline as a substrate.^{15,16} The hydroxyl leaving group on 3-hydroxyproline is adjacent to a relatively acidic proton ($pK_a \sim 29$) on the α -carbon of the amino acid and therefore can be readily eliminated through deprotonation of this carbon by a general base. In contrast, there are no acidic protons on the carbon atoms adjacent to the hydroxyl substituent of 4-hydroxyproline. The use of radical chemistry to

mediate this transformation presents a solution to this chemical challenge. Furthermore, this type of elimination has precedence in the GRE superfamily with similar reactions catalyzed by GD, PD, and CutC (Figure 1.3, see Chapter 1 for background).

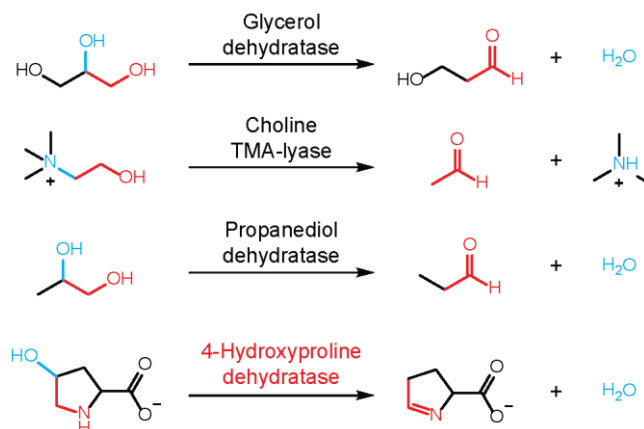


Figure 2.2 Dehydration activity proposed for the new GRE parallels known reactivities of GRE eliminases.

The part of substrates undergoing oxidation to drive elimination is highlighted in red. The eliminated functional group and the bond undergoing cleavage are highlighted in blue.

2.1.2 A putative role for HypD in Stickland fermentation

The proposed role of this gene cluster in the overall reduction of Hyp to Pro in Clostridiales has precedence in microbiology literature. Stickland fermentation is unique to mostly species in the order Clostridiales and involves the oxidation and reduction of amino acids to generate energy. Both Pro and Hyp are common electron acceptors used by Clostridiales, including *Clostridioides difficile* (formerly known as *Clostridium difficile*) and *Clostridium sporogenes*.^{17,18} While the enzymes responsible for Pro reduction had been studied, those involved in Hyp fermentation

remained unknown prior to our work.^{18,19} Validation of this GRE as a HypD uncovered the molecular details of Stickland fermentation of Hyp. The conversion of Hyp to Pro, a substrate known to upregulate expression of the Pro reduction operon, accounted for a previous observation that found Hyp to induce the expression of D-proline reductase, a key enzyme in Pro reduction, in *C. difficile*.¹⁸

2.2 Results and discussions

2.2.1 Multiple sequence alignments and construction of a HypD homology model

Before we sought to validate the new GRE as a HypD experimentally, we obtained additional evidence to support our hypothesis through a bioinformatic approach. A multiple sequence alignment of this protein against characterized GREs showed conservation of residues thought to be important in dehydration chemistry (Figure 2.3). Gly765 and Cys434 (UniProt A0A031WDE4, *C. difficile*) are conserved in all putative HypDs and predicted to mediate radical transfer to the substrate. HypD sequences contain additional residues that can act as a general base or acid and are thought to be essential for catalyzing the elimination step via a spin-center shift on a radical intermediate (Figure 2.3). PD contains residues Glu440 and His166 (UniProt Q1A666, *R. inulinivorans*), which are also conserved in GD. Glu440 likely deprotonates the C1-hydroxyl group while His166 is thought to protonate the departing C2-hydroxyl group (see Chapter 1 for a detailed description of PD and GD mechanisms).^{20,21} Site-directed mutagenesis of PD carried out by Benjamin Levin in our group confirmed these residues as critical for catalysis.¹ CutC, the only reported GRE eliminase that does not catalyze dehydration, has an Asp216 residue (UniProt B8J0I2, *D. desulfuricans*) at the position of the conserved His residue found in GRE dehydratases (see Chapter 1 for a description of CutC mechanism).²²

In HypD, the corresponding conserved residues are Glu436 and His160 (UniProt A0A031WDE4, *C. difficile*) (Figure 2.3). The presence of a His instead of Asp and the higher amino acid sequence identity to GD (UniProt Q8GEZ8) compared to CutC (UniProt Q30W70), 37% vs. 31%, indicated that this new GRE was likely a dehydratase. Since Glu436 and His160 are highly conserved in putative HypD sequences, we postulated that they may act as the general base and acid during catalysis (Figure 2.3).

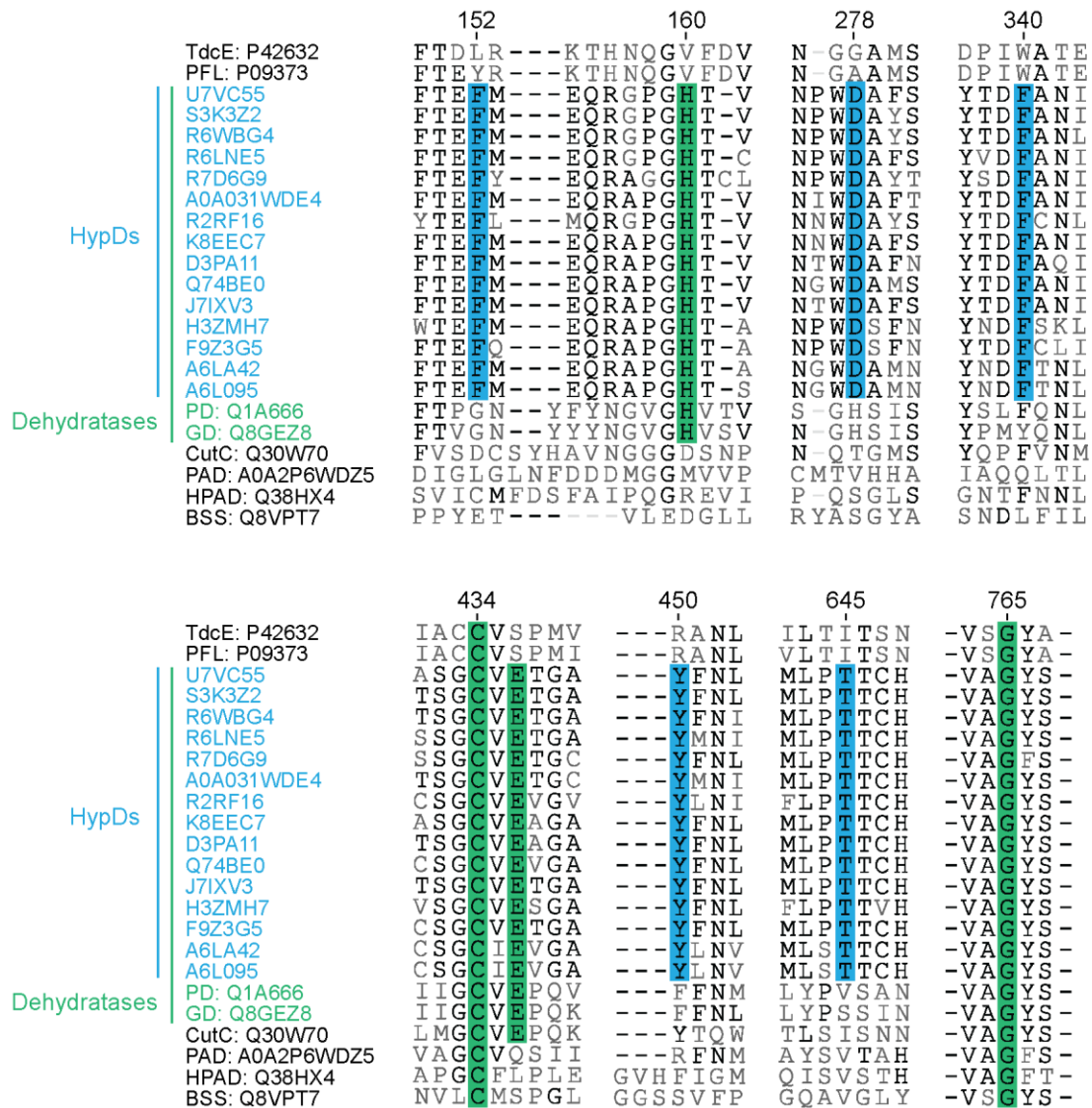


Figure 2.3 A multiple sequence alignment of putative HypDs with characterized GREs.

A multiple sequence alignment of characterized GREs and putative HypDs selected to cover a wide range of phylogenetic diversity. Residues conserved among GRE dehydratases are highlighted in green. Conserved residues observed to be in the active site of the HypD homology model (Figure 2.4) generated from sequences in the NCBI database (Figure 2.23) are highlighted in blue. Conserved amino acids are displayed

Figure 2.3 (Continued)

in bold. UniProt accession codes are listed for each sequence with residue numbering based on HypD from *C. difficile* (A0A031WDE4). Sequences were aligned by Clustal Omega.²³

To identify active site residues that might provide additional information regarding the activity of this GRE, we constructed a homology model. Using HHPred²⁴, the highest secondary structure scores were calculated for PD, GD, and CutC. The GD (PDB: 1R9D) structure²⁵ was used as the template in Modeller²⁶ to generate the HypD homology model, which exhibited the canonical 10-stranded α/β barrel architecture for GREs as expected (Figure 2.4A). The previously identified Gly765 and Cys434 are positioned within a buried active site adjacent to each other and overlay well with Gly763 and Cys433 in GD (Figure 2.4B). The acid-base residue pair (His160, Glu436) in HypD is located within the active site and aligns well with His164 and Glu435 in GD (Figure 2.4B). Furthermore, we identified a set of residues in the active site of the homology model that are absolutely conserved in putative HypD sequences. These residues may be important for catalysis and substrate binding and include F152, D278, F340, Y450, and T645 (Figure 2.3, Figure 2.4C). Some of these residues differ from those present in the active sites of known GRE eliminases, which implies the binding of a different substrate (Figure 2.4C). Because the HypD homology model was generated from a GD structure template, the spatial orientation of its active site residues is likely biased toward the arrangements found in GD. Therefore, the set of residues identified here was likely incomplete. Indeed, a crystal structure of HypD with Hyp bound obtained by Lindsey Backman in the Drennan lab was crucial in more precisely identifying the residues critical for substrate binding and catalysis (discussed in Chapter 3). In summary,

bioinformatic analyses provided additional support for this GRE as a new member of the GRE eliminase class catalyzing a dehydration reaction.

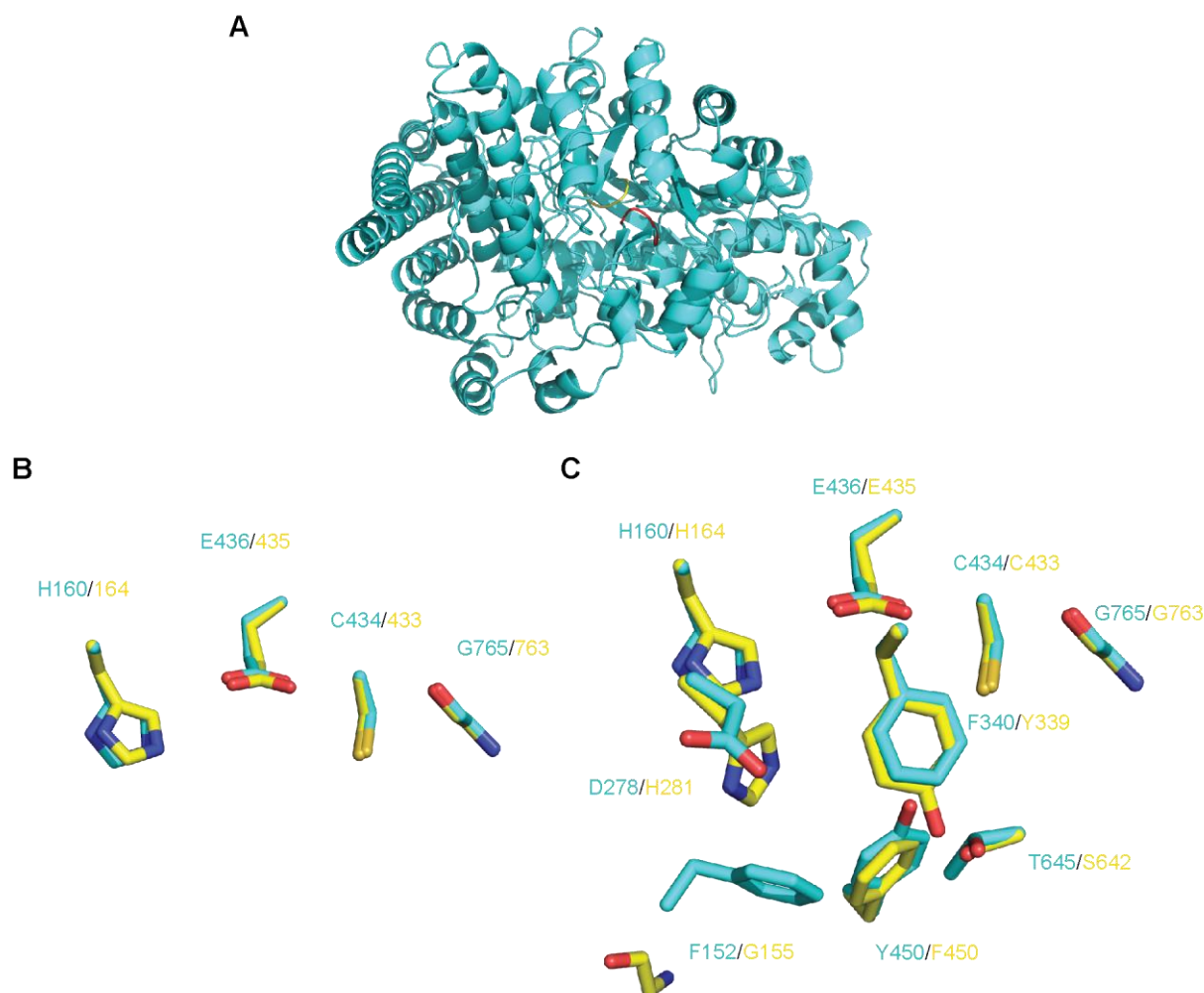


Figure 2.4 A homology model of HypD supports its role as a dehydratase.

(A) Homology model of HypD (cyan) constructed using HHPred and Modeller.^{27,28}

The Gly loop is highlighted in red and the Cys loop is highlighted in yellow. (B) An overlay of residues in the dehydratase motif of HypD and GD (yellow, PDB: 1R9D)²⁵. (C) An overlay of predicted active site residues in HypD (cyan) and the analogous residues in GD (yellow). Active site residues displayed here are conserved among all predicted HypD sequences compiled from the NCBI database (Table 2.1).

2.2.2 Heterologous overexpression and purification of proteins in an *E. coli proC* deletion mutant

In order to study activities of putative HypD, AE, and P5CR *in vitro*, we set out to obtain purified proteins. Plasmid vectors were constructed by Dr. Yifeng Wei (Stubbe lab) to individually overexpress each protein in *Escherichia coli* for protein purification.ⁱⁱ The corresponding genes were amplified from *C. difficile* 70-100-2010 for these studies. Overexpression of these proteins was tested among *E. coli* strains BL21(DE3), BL21-CodonPlus(DE3)-RIL, and Rosetta(DE3)pLysS at 15 °C or 25 °C overnight and soluble forms were best obtained in *E. coli* BL21-CodonPlus(DE3)-RIL for all proteins. Purification conditions for GRE and AE were optimized extensively to obtain consistent glycyl radical formation with improved yields (see Section 2.2.4 for activation experiments). GRE, AE, and P5CR were purified in soluble forms aerobically and then cycled through vacuum and N₂ or Ar gas to remove oxygen for storage and biochemical studies (Figure 2.5).

ⁱⁱ Overexpression vectors were constructed by Dr. Yifeng Wei. Initial overexpression and purification of GRE and AE were carried out by Dr. Yifeng Wei.

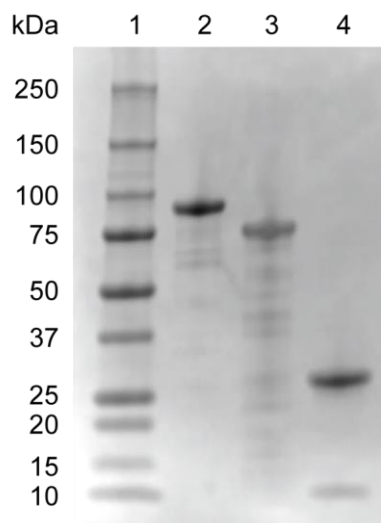


Figure 2.5 SDS-PAGE of purified enzymes used in biochemical studies.

Precision Plus Protein™ All Blue Standards (Bio Rad) (lane 1), HypD (lane 2), MBP-HypD-AE (lane 3), and P5CR (lane 4). Molecular weights of these proteins are as follows: HypD – 91.3 kDa; MBP-AE – 81.5 kDa, P5CR – 30.4 kDa.

Initial attempts at *in vitro* reconstitution of all three enzymes with cofactors SAM and NADH in a coupled enzyme end-point assay resulted in the full conversion of Hyp to Pro. However, complete formation of Pro was also observed in the negative control without addition of purified P5CR (Figure 2.6). One possible explanation for this result could be the presence of contaminating P5CR activity from the *E. coli* overexpression strain due to co-purification of low levels of *E. coli* P5CR during purification of HypD and AE. This potentially could be a result of protein-protein interactions between P5CR and one of the proteins. An alternative explanation could be that this GRE catalyzes the transformation of Hyp to Pro in a NADH-dependent manner, although no NADH-binding motif is predicted for this enzyme.

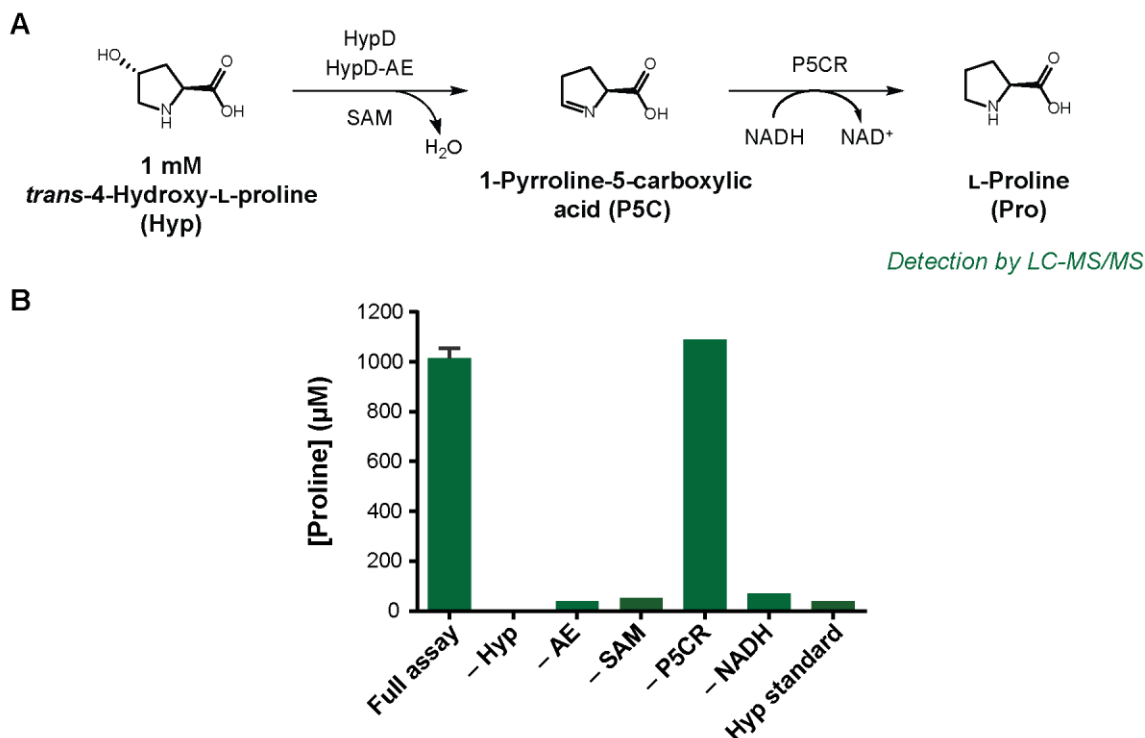


Figure 2.6 End-point activity assay for HypD with contaminating P5CR activity.

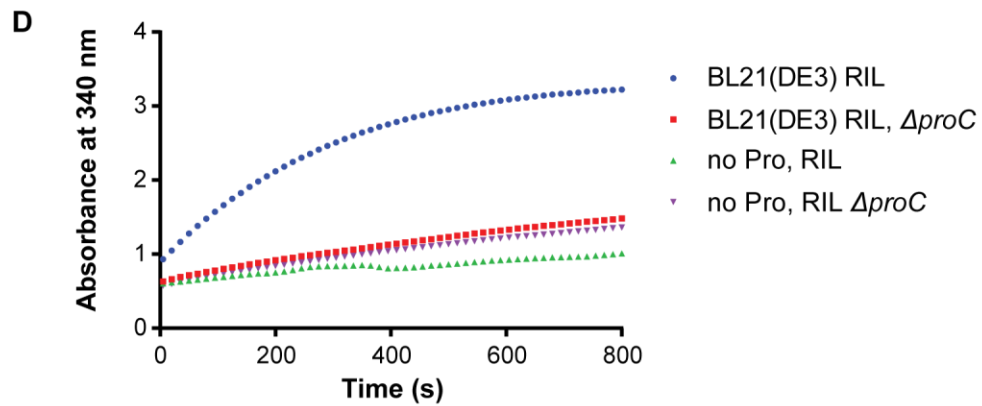
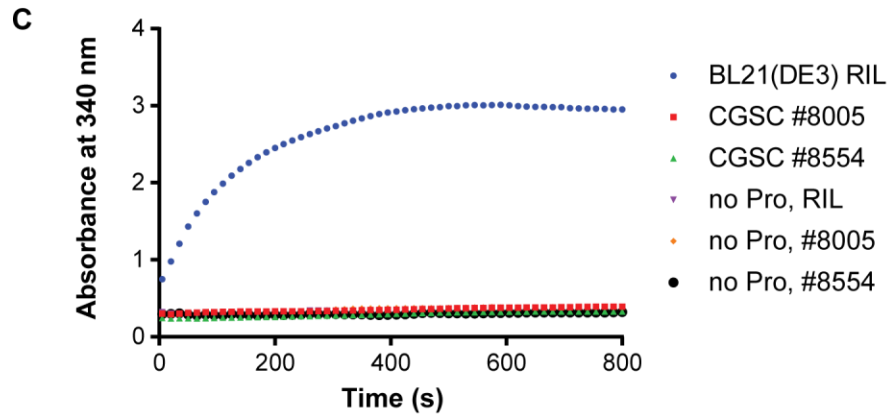
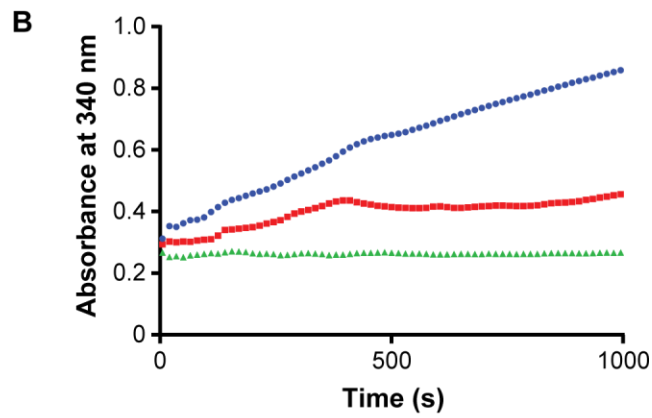
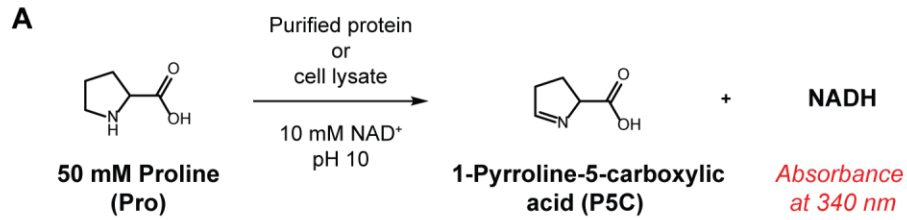
(A) Scheme depicting coupled enzyme assay used to test dehydration of Hyp by HypD. (B) Incubation of activated HypD, P5CR, NADH, and 1 mM Hyp resulted in full conversion of Hyp to Pro in the full assay. Individual components of the assay were excluded in the negative controls. Quantification of Pro was achieved by LC-MS/MS. Trace amounts of Pro observed in most negative controls were likely due to impurities in the commercial Hyp supply.

To study the precise function of each enzyme, we set out to test these two hypotheses using a previously described spectrophotometric assay used to study P5CRs under non-physiological conditions.^{14,29} Under high pH in the presence of high concentrations of Pro and cofactor NAD(P)⁺, P5CR can catalyze its physiological reaction in the reverse direction to oxidize Pro to P5C and

thus acting as a Pro dehydrogenase.^{14,29} This approach circumvents the need to synthesize P5C, an unstable metabolite that can hydrolyze to form glutamate-5-semialdehyde in aqueous solutions.³⁰ Indeed, contaminating P5CR activity in purified GRE and AE solutions was detected in the Pro dehydrogenase activity assay where accumulation of NADH was measured by monitoring absorbance at 340 nm over time (Figure 2.7A, B). In the same assay using cell lysates from the *E. coli* overexpression strain and two Keio deletion mutants (CGSC #8005 and #8554³¹) of *proC*, the P5CR-encoding gene, Pro dehydrogenase activity was absent in the deletion mutants only (Figure 2.7C). Given that genomic differences between K-12 (Keio mutants) and BL21(DE3) strains do not involve genes that participate in Pro biosynthesis, we concluded that *proC* is likely the sole gene responsible for the observed P5C dehydrogenase activity in the RIL overexpression strain. To remove contaminating P5CR from purified GRE and AE, deletion of *proC* in the RIL strain was achieved through P1 phage transduction to generate *E. coli* BL21-CodonPlus(DE3)-RIL $\Delta proC::aac(3)IV(Am^R)$. The previously observed Pro dehydrogenase activity was absent in cell lysate from the RIL $\Delta proC$ strain (Figure 2.7D), and thus this strain was used to overexpress and purify proteins used in all subsequent biochemical studies.

Figure 2.7 Contaminating P5CR activity in purified proteins and *E. coli* cell lysates.

(A) Scheme depicting assay conditions used to drive P5CR activity in the reverse direction to obtain Pro dehydrogenase activity. (B) Pro dehydrogenase activity was observed in purified HypD and AE solutions. (C) Pro dehydrogenase activity was observed in assays containing lysate from *E. coli* BL21-CodonPlus(DE3)-RIL, but not in lysates from two *proC* deletion mutants from the Keio collection (CGCS #8005, 8554). (D) Deletion of the *proC* gene from *E. coli* BL21-CodonPlus(DE3)-RIL removed Pro dehydrogenase activity in cell lysates. All curves represent single data points.



2.2.3 Characterization of P5CR activity using a Pro dehydrogenase assay

With purified proteins in hand, we first characterized the activity of P5CR. Biochemical and structural studies of bacterial P5CRs were previously reported, but P5CR from *C. difficile* had not been characterized.^{13,14,32} *C. difficile*, along with a few other Clostridiales possessing the HypD gene cluster, encode two copies of P5CR with one adjacent to the putative HypD and the second copy located in a distant region of the genome. Interestingly, the two P5CR genes found in *C. difficile* 70-100-2010 only have 67% amino acid sequence identity. This could be a result of either divergent evolution due to their roles in different pathways or a recent horizontal gene transfer of the GRE-encoding gene cluster from a different species. From a sequence alignment, we predicted that this P5CR likely prefers NADH over NADPH as its cofactor, due to the absence of a conserved arginine to interact with the phosphate group of NADPH.³² Furthermore, active site residues proposed to be involved in substrate binding are present in the *C. difficile* enzyme.³²

As described earlier, the non-physiological conversion of Pro to P5C is favored for P5CR at high pH and in the presence of high concentrations of Pro and cofactor NAD(P)⁺ (Figure 2.8A).¹⁴ We used a similar assay to investigate the dehydrogenase activity of purified *C. difficile* P5CR as well as its substrate and cofactor specificity. These experiments established that this P5CR is highly selective for NAD⁺ (Figure 2.8B) and L-proline (Figure 2.8C) over NADP⁺ and D-proline. Therefore, NADH was included as the reductant for coupled enzyme assays to study HypD biochemistry.

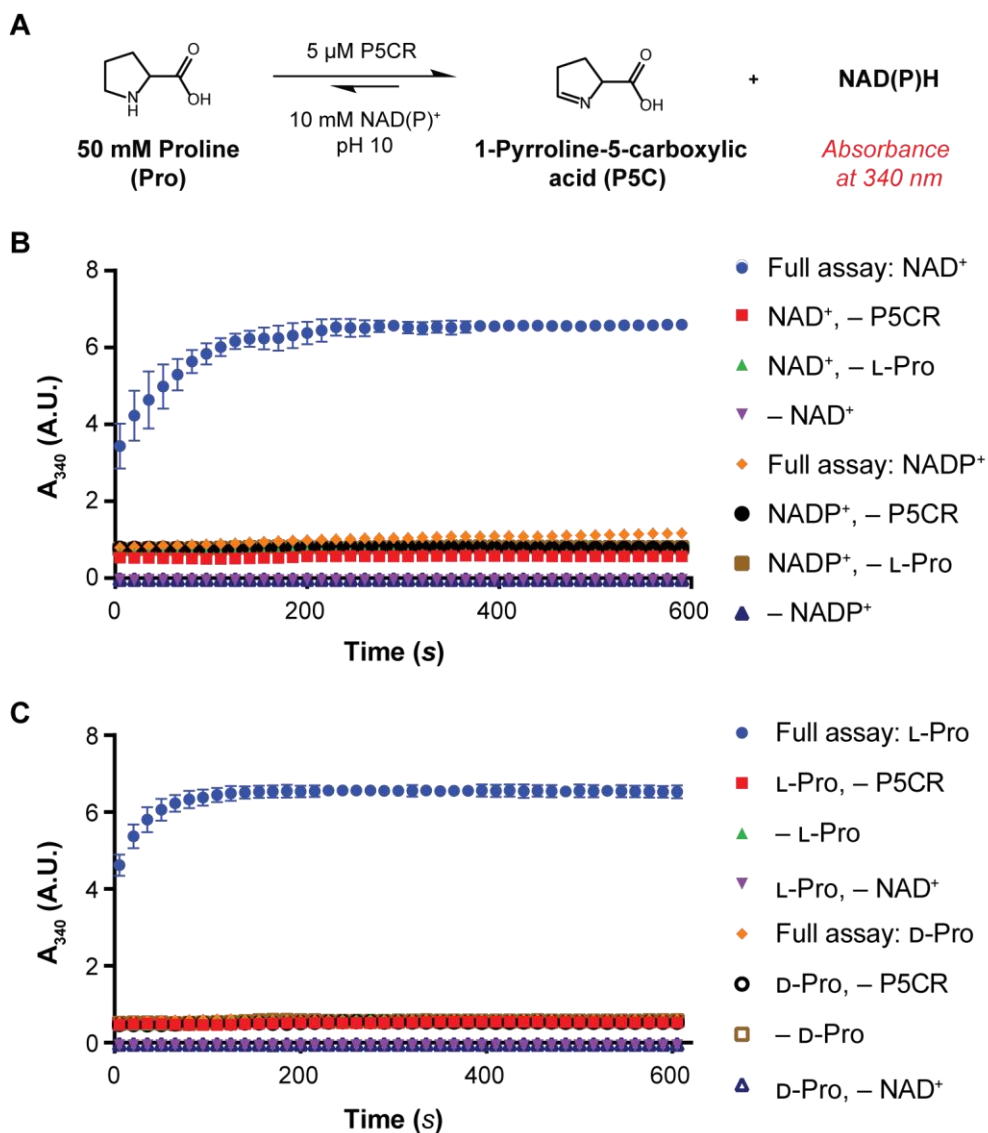


Figure 2.8 Proline dehydrogenase activity of P5CR.

(A) Scheme depicting assay conditions used to drive the P5CR-catalyzed reaction in the reverse direction. Pro dehydrogenase activity was monitored over time by measuring the absorbance of NADH or NADPH at 340 nm. (B) Pro dehydrogenase activity was observed only in full assays containing cofactor NADH and not NADPH. (C) Pro dehydrogenase activity was observed only in full assays containing substrate L-proline and not D-proline.

2.2.4 Purification and [4Fe-4S] cluster reconstitution of AE, a radical SAM enzyme

Different overexpression, purification, and reconstitution protocols were tested for maltose binding protein (MBP)-tagged AE (molecular weight: 81.5 kDa) to improve activation of HypD. HypD-AE was overexpressed and purified aerobically to obtain proteins containing partially assembled iron-sulfur clusters, as indicated by a brown colored solution. Purified HypD-AE was then rendered anaerobic and incubated with Fe(II) and sulfide overnight for chemical reconstitution of its [4Fe-4S] clusters. Attempts were made to co-express HypD-AE with the *E. coli isc* operon under anaerobic conditions to assemble [4Fe-4S] clusters intracellularly. However, this failed to yield good overexpression of AE and thus was abandoned in favor of *in vitro* reconstitution of [4Fe-4S] clusters. The N-terminal MBP tag on the HypD-AE did not appear to interfere with activation of HypD and thus was not removed post-purification.

Unexpectedly, upon elution of HypD-AE from TALON resin under aerobic conditions, two dark protein bands of approximately 60 and 80 kDa were observed by SDS-PAGE (Figure 2.9A). After anaerobic reconstitution of [4Fe-4S] clusters and buffer-exchange to remove excess iron and sulfide, HypD-AE migrated as a single band at ~80 kDa (Figure 2.9B). During optimizations of purification conditions, we noticed that purified HypD-AE behaved differently on SDS-PAGE and appeared either as a single band or double bands. This led us to wonder if the aerobically purified HypD-AE with incomplete [4Fe-4S] clusters contained additional disulfide bonds formed between free cysteine residues that would otherwise ligate to Fe in a completely assembled cluster. These intramolecular disulfide bonds would not be as readily reduced and denatured, which might lead to secondary structures that would affect its migration on SDS-PAGE. As part of routine protein visualization, samples were heat denatured and reduced with 2-mercaptoethanol (BME). To test our hypothesis, we screened a range of conditions to reduce and

denature a single solution of reconstituted AE that was stored aerobically. The higher molecular weight (MW) band was observed in all heat denatured samples along with non-heat denatured sample reduced with tris(2-carboxyethyl)phosphine (TCEP) (Figure 2.9C). In contrast, a faint high MW band and a dominant ~60 kDa band were observed in the untreated sample and non-heat denatured samples reduced with dithiothreitol (DTT) or 2-mercaptoethanol (BME) (Figure 2.9C). Appearance of the smaller band in samples that were less denatured supports the identity of this band to consist of AEs that contain some structure, which would migrate as a smaller sized protein. Heat-denatured and BME-reduced samples routinely collected during purifications visualized as double bands, which contradicts the single band observed in the screen under identical conditions (Figure 2.9A vs. C). This is likely due to a more oxidized state of AE upon aerobic purification compared to reconstituted AE that was then incubated aerobically overnight. TCEP is a more stable and effective reducing agent than DTT and BME,³³ which likely explains why heat was not required for full protein denaturation in the non-heat treated, TCEP-reduced sample. In conclusion, both protein bands consist of AE with the higher MW band representing a more denatured state whereas the lower MW band is indicative of a structured state likely containing disulfide bonds.

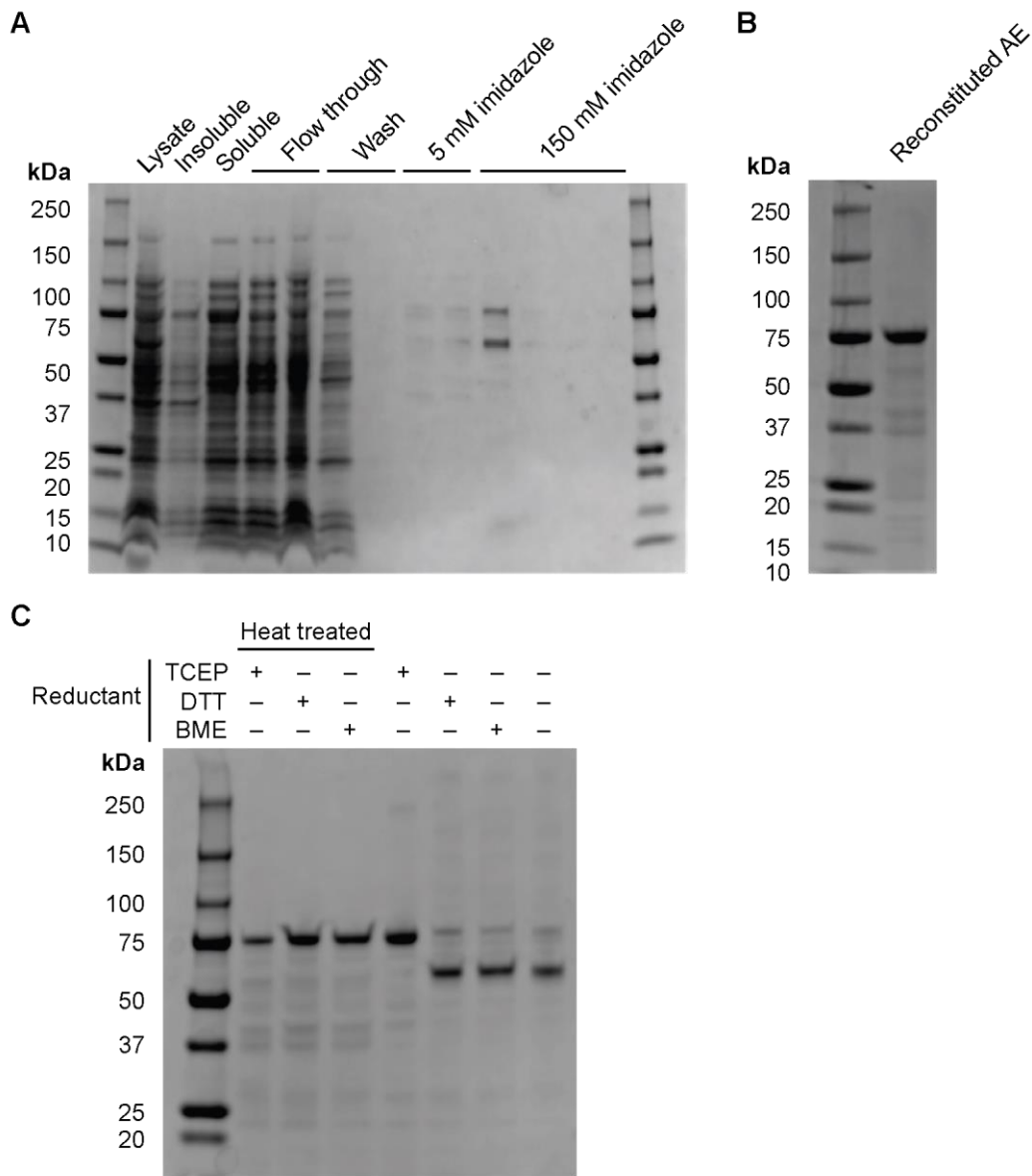


Figure 2.9 SDS-PAGE visualization of purified AE as two protein bands.

Precision Plus ProteinTM All Blue Standards (Bio Rad) was included as ladder on all SDS-PAGE gels. (A) A representative gel showing fractions collected from various steps of AE purification. (B) A representative gel of purified AE after [4Fe-4S] cluster reconstitution. (C) Reconstituted AE undergone different treatments in heat denaturation and chemical reduction.

A UV-Vis absorbance spectrum of reconstituted HypD-AE revealed an absorption peak around 410 nm, which is typical of $[4\text{Fe-4S}]^{2+}$ clusters (Figure 2.10).³⁴ As expected, reduction of this cluster to $[4\text{Fe-4S}]^+$ by sodium dithionite (NaDT) resulted in a loss of absorbance. Analyses of the iron and sulfide content of purified HypD-AE revealed a range of 6.3–9.9 moles of Fe and 4.7–7.4 moles of S per mole of enzyme from multiple purifications. This could indicate the presence of multiple $[4\text{Fe-4S}]$ clusters, including partially assembled clusters. It is notable that similar to most AEs, HypD-AE contains two Clostridial-type motifs that may bind auxiliary $[4\text{Fe-4S}]$ clusters (Figure 1.5). Unfortunately, the number of complete $[4\text{Fe-4S}]$ clusters assembled cannot be determined from these assays. We also cannot account for adventitious binding of iron and sulfide to purified proteins. Therefore, the precise number of these auxiliary clusters and their functions remain to be elucidated.

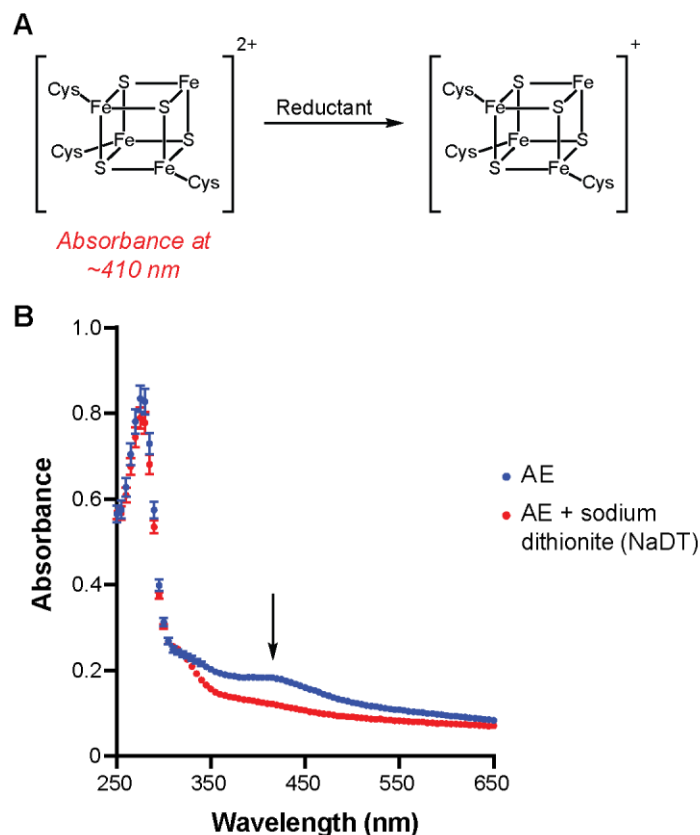


Figure 2.10 Absorbance spectra of HypD-AE indicate the presence of a redox active [4Fe-4S] cluster.

(A) Absorbance at ~410 nm is indicative of $[4\text{Fe-4S}]^{2+}$ clusters. (B) UV-Vis spectra of HypD-AE with and without reduction by sodium dithionite (NaDT). Data are shown as mean $\pm$ SD with $n = 3$ replicates.

2.2.5 Demonstration of *trans*-4-hydroxy-L-proline dehydration activity *in vitro*

In order to study the proposed dehydration activity, we first proceeded to obtain robust activation of the GRE by its AE. In an *in vitro* assay, we reduced the [4Fe-4S] cluster in purified AE with the reducing agent Acriflavine (Figure 2.11). Upon addition of SAM and HypD, we

observed the formation of a glycine-centered radical by electron paramagnetic resonance (EPR) spectroscopy. The first activation of HypD was achieved by Dr. Yifeng Wei with a yield of approximately 10% radical formation per protein, which established these enzymes as an active GRE and AE pair. In addition to optimization of purification conditions, activation assay parameters including buffer reagent, pH, SAM concentration, incubation time, and the GRE:AE stoichiometric ratio were screened to improve glycyl radical yield. Ultimately, we were able to achieve $51 \pm 1\%$ glycyl radical formation per protein monomer (Figure 2.12). This is the proposed theoretical maximum yield based on the observed half-site reactivity in other GREs.³⁵

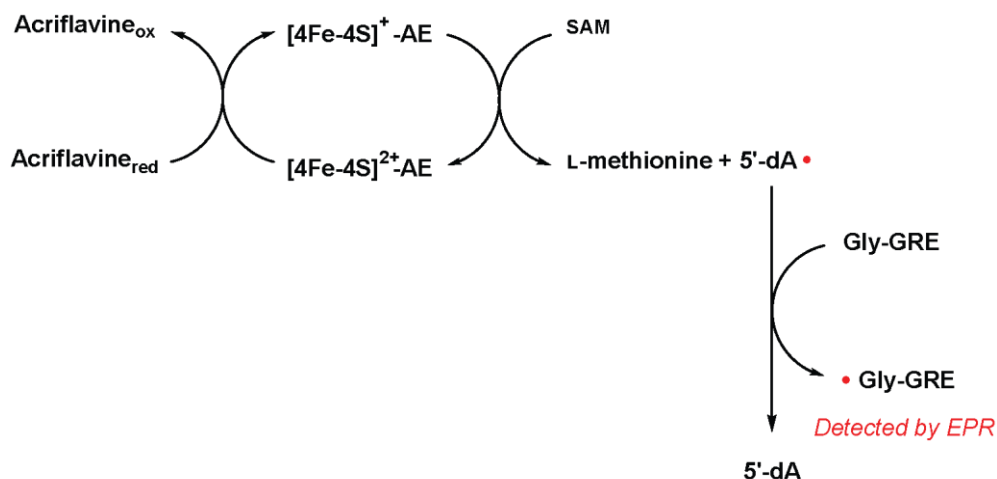


Figure 2.11 Assay set up for activation of HypD by its cognate AE.

A scheme depicting the assay used to activate HypD using its AE. AE (60 μM) was first incubated for 20 min with the photoreductant Acriflavine (0.1 mM) to reduce $[4\text{Fe-4S}]$ clusters. HypD (15 μM) along with SAM (1.5 mM) were added to reduced AE. Glycyl radical formation after 2 h was quantified by EPR spectroscopy using external standards of Frémy's salt (potassium nitrosodisulfonate).

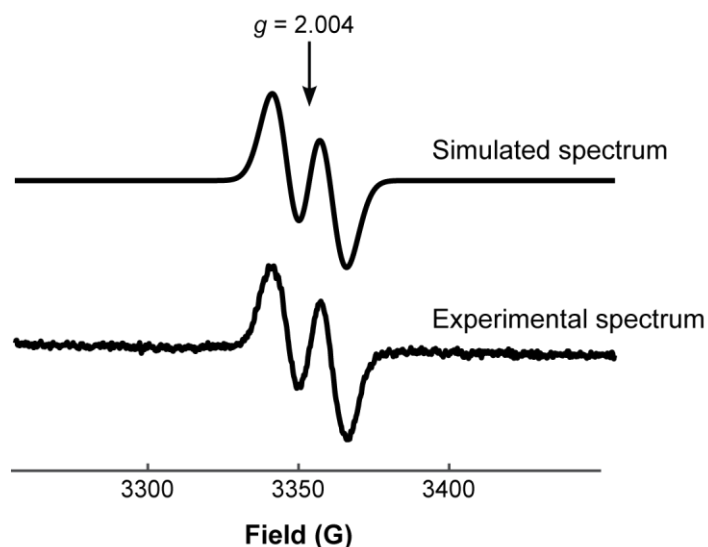


Figure 2.12 EPR spectrum of activated HypD.

A representative EPR spectrum of the glycy radical in activated HypD. An average of 0.51 ± 0.01 (mean $\pm$ SD) glycy radical per monomer was observed with hyperfine coupling constant $A = 1.44$ mT. Quantification of glycy radicals was calculated from an activation assay carried out in triplicate.

With the activation conditions optimized, we first tested HypD activity toward the most biologically abundant stereoisomer of 4-hydroxyproline: Hyp (see Chapter 1 for details on Hyp biosynthesis). This amino acid is synthesized by the diastereoselective prolyl 4-hydroxylase and proline 4-hydroxylase enzymes. A coupled enzyme end-point assay containing activated HypD, P5CR, NADH, and Hyp resulted in the complete conversion of Hyp to Pro (Figure 2.13). Hyp and Pro in assay mixtures were detected and quantified using liquid chromatography–tandem mass spectrometry (LC-MS/MS) (Figure 2.14). Each component of the full assay was essential for the production of Pro, although consumption of Hyp was also observed in assays lacking P5CR or NADH (Figure 2.13). This supports the activity of P5CR toward reduction of P5C as the second

step in this coupled enzyme assay and that HypD activity is not dependent on the downstream P5CR activity. The P5C intermediate presumably formed in this assay was not directly detected due to the lack of a chemical standard. Unexpectedly, trace amounts of Hyp and Pro were detected in the negative control that excluded Hyp. This may be caused by contamination of amino acids between samples during preparation or during runs on the LC-MS instrument.

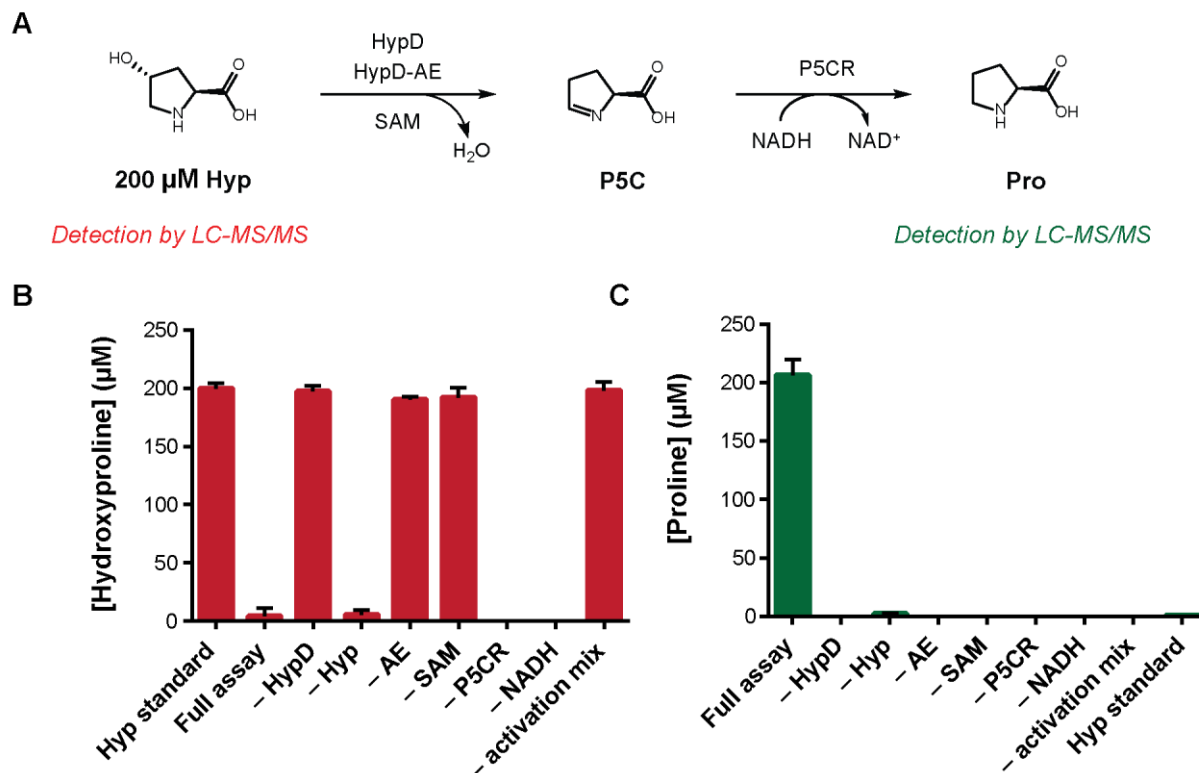


Figure 2.13 End-point activity assay for HypD.

(A) Scheme depicting the coupled enzyme assay used to validate enzyme activities involved in the overall transformation of Hyp to Pro. HypD was first activated with AE using the previously described EPR assay. Activated HypD was added to a final concentration of 0.3 μ M in a mixture containing 3 μ M P5CR, 0.4 mM NADH, and 0.2 mM Hyp to initiate reaction. Aliquots were removed for LC-MS/MS quantification of Pro and Hyp. Concentrations of (B) Hyp and (C) Pro were quantified using external standard curves. Data are shown as mean $\pm$ SD with $n = 3$ replicates. Assays were quenched after 1 h of incubation and data shown in both panels were obtained from the same experiment.

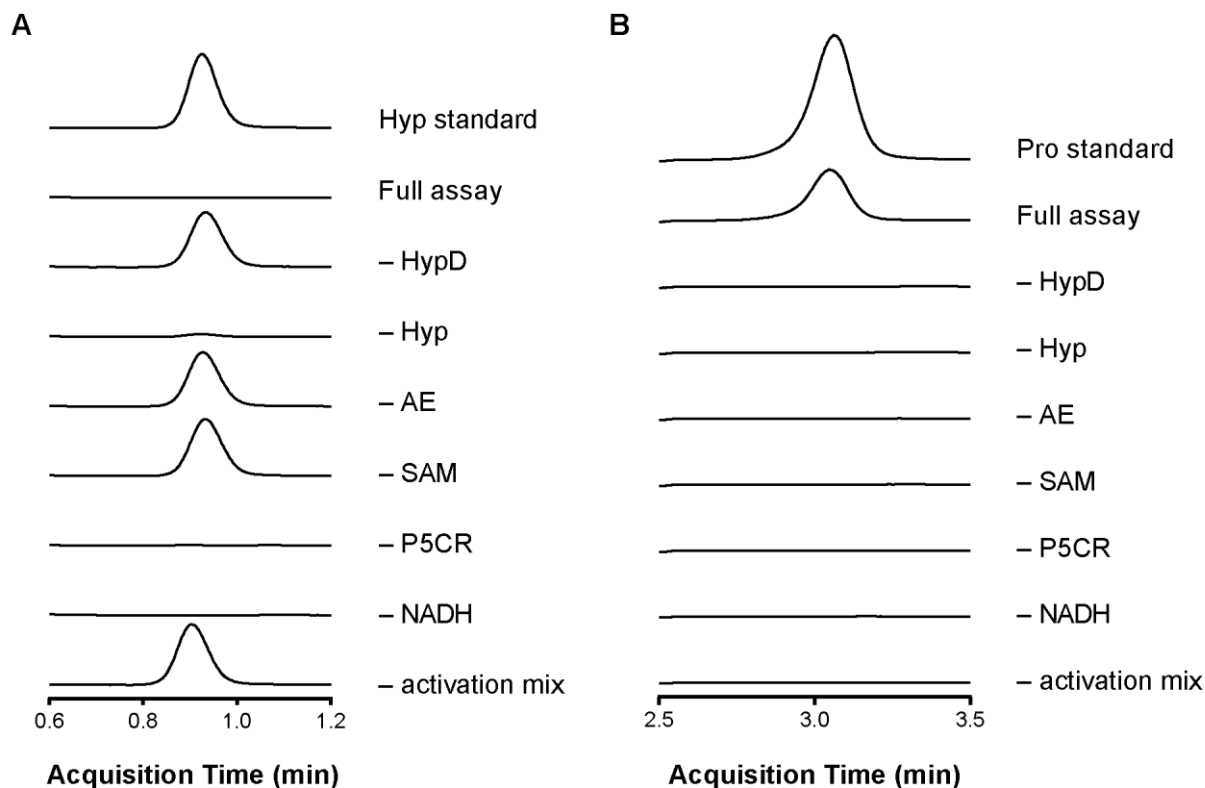


Figure 2.14 LC-MS/MS chromatograms of Hyp and Pro in end-point assays.

Representative chromatograms of multiple reaction monitoring (MRM) for (A) Hyp and (B) Pro detected by LC-MS/MS under each set of assay conditions shown in the previous figure.

2.2.6 HypD is highly specific toward *trans*-4-hydroxy-L-proline

We proceeded to determine if HypD has activity toward the remaining three stereoisomers of 4-hydroxyproline and a structural isomer, *cis*-3-hydroxy-L-proline (Figure 2.15A). We employed the previously described end-point assay with the exception that reductant NaBH₄ was used to chemically reduce any pyrroline product to proline since P5CR was previously shown to be stereospecific for L-proline and not D-proline (Figure 2.15B). Therefore, P5CR would accept

(*S*)-1-pyrroline-5-carboxylate as the physiologically-relevant substrate, which is consistent with P5CR from *C. sporogenes*.²⁹ In this end-point activity assay, HypD exhibited undetectable or greatly reduced activity toward the other substrates, as determined Pro formation (Figure 2.15C, Figure 2.16). Only low conversions of $11.4 \pm 0.3\%$ (mean $\pm$ SD) for *trans*-4-hydroxy-D-proline, the enantiomer of Hyp, were detected under conditions in which HypD catalyzed full conversion of Hyp to Pro.

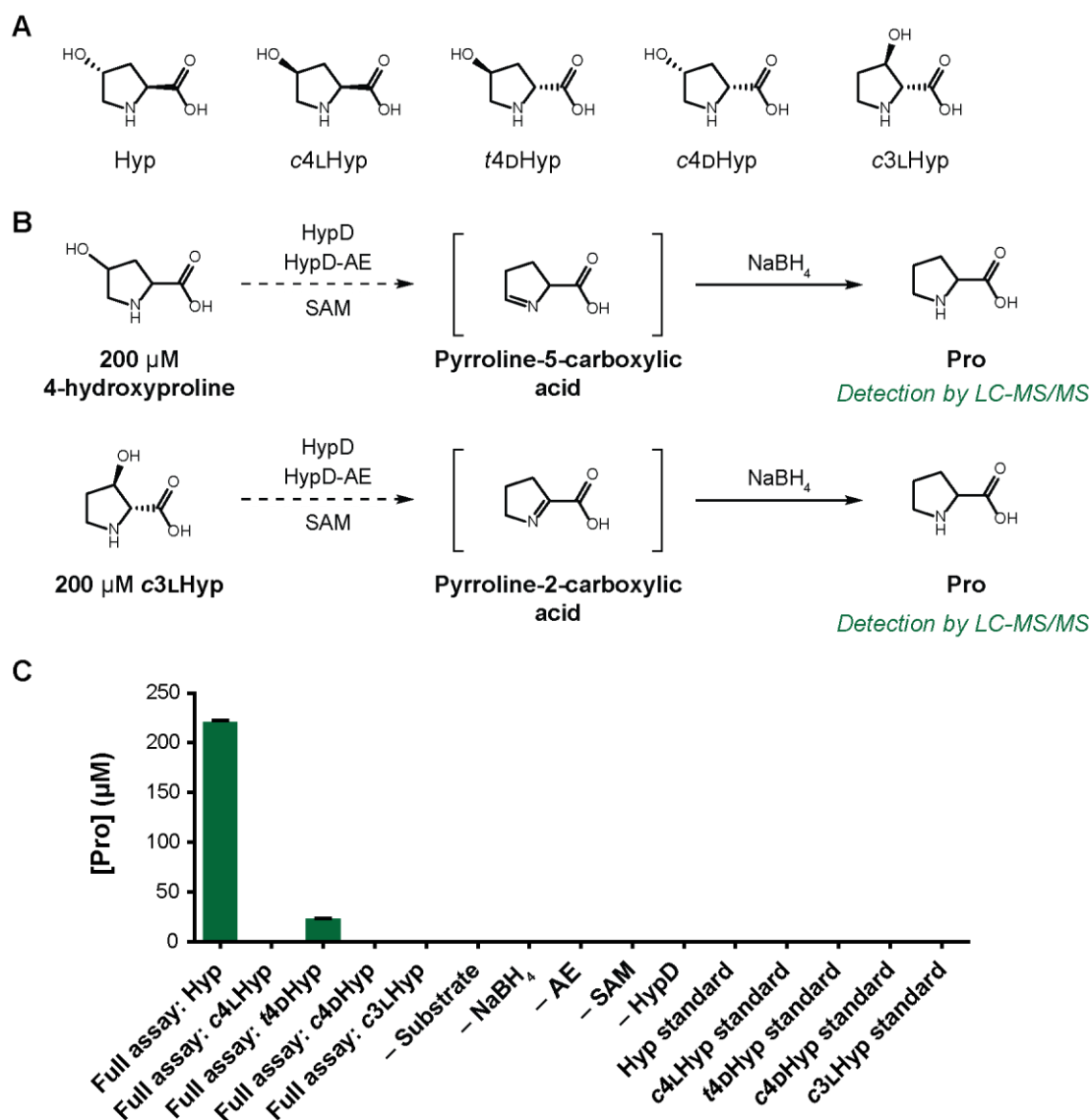


Figure 2.15 HypD activity toward hydroxyproline structural isomers.

(A) Five structural isomers of hydroxyproline were tested as substrates: *trans*-4-hydroxy-L-proline (Hyp), *cis*-4-hydroxy-L-proline (c4LHyp), *trans*-4-hydroxy-D-proline (t4DHyp), *cis*-4-hydroxy-D-proline (c4DHyp), *cis*-3-hydroxy-L-proline (c3LHyp). (B) 200 μ M of each hydroxyproline substrate was incubated with activated HypD. Dehydration of any 4-hydroxyproline stereoisomer by HypD would yield

Figure 2.15 (Continued)

either (*S*)- or (*R*)-pyrroline-5-carboxylate. Dehydration of *cis*-3-hydroxy-L-proline is expected to result in the formation of pyrroline-2-carboxylate. Both pyrroline products can be chemically reduced by NaBH₄ to proline for detection by LC-MS/MS. All assay mixtures were quenched after 1 h of incubation. (C) Final proline concentrations measured in each assay condition. Data are shown as mean $\pm$ SD and *n* = 3 replicates.

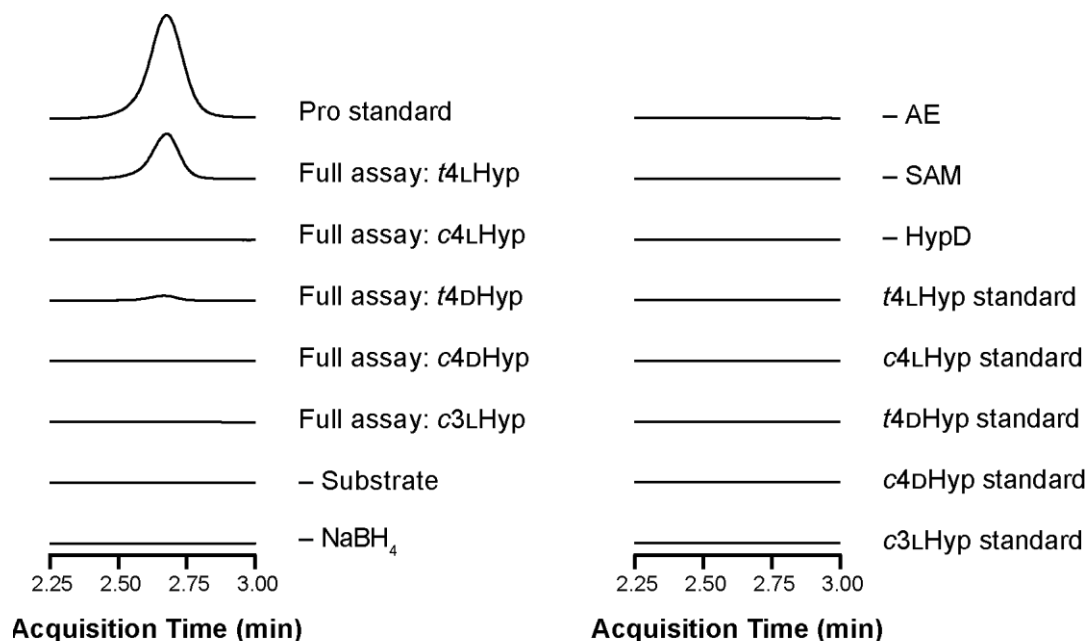


Figure 2.16 LC-MS/MS chromatograms of Pro in assays testing activity toward structural isomers.

Representative chromatograms of multiple reaction monitoring (MRM) for Pro detected by LC-MS/MS under each set of assay conditions shown in the previous figure.

The absence of a Hyp impurity in the commercial supply of *trans*-4-hydroxy-D-proline was verified using chiral chromatography. Each 4-hydroxyproline stereoisomer obtained commercially was derivatized with N α -(2,4-dinitro-5-fluorophenyl)-L-alaninamide (Marfey's reagent) and separated on chiral HPLC (Figure 2.17). This analysis showed each of these compounds as a single stereoisomer (>99% ee), thereby validating HypD's activity toward *trans*-4-hydroxy-D-proline. HypD's low activity toward *trans*-4-hydroxy-D-proline was unexpected due to a lack of known biological source of this amino acid. Rather, this observation may be explained by an absence of selection pressure for HypD to exclude binding of this substrate because of its rare occurrence in Nature.

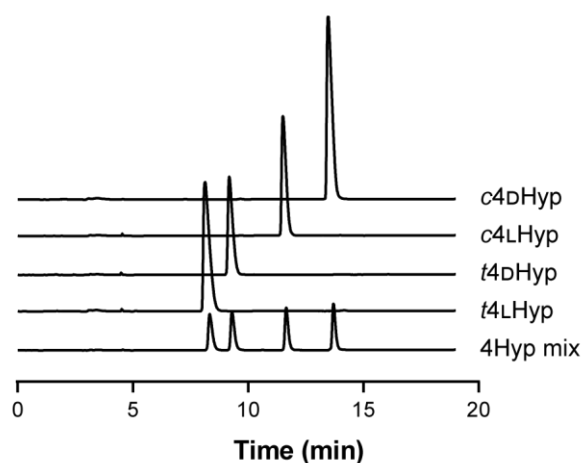


Figure 2.17 Commercial sources of 4-hydroxyprolines are enantiomerically pure as determined by chiral-HPLC.

Stock solutions from commercial supplies of 4-hydroxyproline isomers used in the stereoselectivity assay were derivatized using Marfey's reagent. Derivatized compounds were injected onto the HPLC for separation on Lux® Amylose-1 column (5 μm , 100 $\text{\AA}$, 100 $\times$ 4.6 mm, Phenomenex). Single peaks with no detectable 4-hydroxyproline impurities were observed for all stereoisomer standards.

2.2.7 Kinetic analysis of HypD in a spectrophotometric coupled enzyme assay

In a coupled enzyme assay where production of P5C was coupled to its reduction catalyzed by P5CR, the rate of NADH consumption was monitored spectrophotometrically to measure kinetic parameters of HypD. P5CR was added in excess (30-fold greater than HypD) and was not the rate-determining step of the observed NADH consumption during the optimization of assay conditions. The calculated kinetic parameters for HypD, $k_{\text{cat}} = 45 \pm 1 \text{ s}^{-1}$, $K_{\text{m}} = 1.2 \pm 0.1 \text{ mM}$, $k_{\text{cat}}/K_{\text{m}} = 3.8 \pm 0.3 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$, are within expected ranges to be physiologically relevant (Figure

2.18). Compared to kinetic values obtained for other GREs, HypD's turnover number is in the lower range.³⁶ HypD also has a lower binding affinity for its substrate Hyp compared to most GREs except for PFL (K_m of ~ 2 mM).^{36,37}

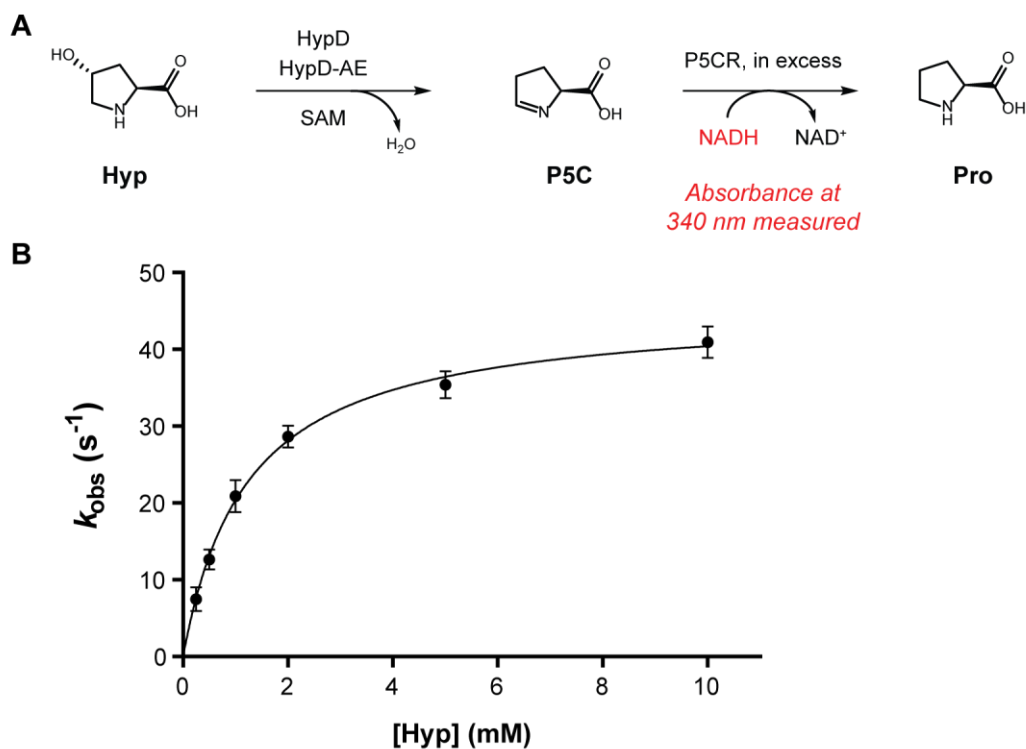


Figure 2.18 Kinetic analysis of HypD from *C. difficile* 70-100-2010.

(A) Conversion of Hyp to P5C by HypD was coupled to reduction of P5C to L-proline by P5CR. Absorbance of NADH at 340 nm was continuously monitored to calculate initial rates. (B) Kinetic data of HypD was fit to the Michaelis–Menten equation using nonlinear regression in Graphpad Prism 7.00. Initial rates were converted to k_{obs} assuming $51 \pm 1\%$ activation of HypD monomers as determined by EPR spectroscopic assays. Background rate in the absence of Hyp was subtracted from other data points. Kinetic parameters were calculated to be $K_{\text{m}} = 1.2 \pm 0.1$ mM, $k_{\text{cat}} = 45 \pm 1$ s⁻¹, $k_{\text{cat}}/K_{\text{m}} = 3.8 \pm 0.3 \times 10^4$ M⁻¹s⁻¹. Kinetic parameters are displayed as mean $\pm$ SE. Data are shown as mean $\pm$ SD with $n = 3$ replicates.

2.2.8 Analysis of 4-hydroxyproline structural analogs as inhibitors of HypD

In Section 2.2.6, we determined that this GRE is highly specific toward the naturally abundant amino acid Hyp. Intriguingly, HypD exhibited some activity toward *trans*-4-hydroxy-D-proline (Figure 2.15). This raised the question of whether HypD can bind some of these isomers with affinities high enough to inhibit Hyp dehydration. Five compounds were tested for HypD inhibition using the previously described kinetic assay with some modifications (Figure 2.19A). The compound of interest was pre-mixed with the substrate Hyp and then added to reaction mixtures to initiate catalysis. Compounds included in the inhibition assay were *trans*-4-hydroxy-D-proline (*t*4DHyp), *cis*-4-hydroxy-D-proline (*c*4DHyp), *cis*-4-hydroxy-L-proline (*c*4LHyp), *cis*-3-hydroxy-L-proline (*c*3LHyp), and L-proline (Pro). Although HypD showed some activity toward *t*4DHyp in the end-point assay, its dehydration would yield (*R*)-1-pyrroline-5-carboxylate, which cannot be reduced by P5CR. Therefore, any HypD activity toward *t*4DHyp in the kinetic assay is not expected to directly contribute to decreases in NADH absorbance.

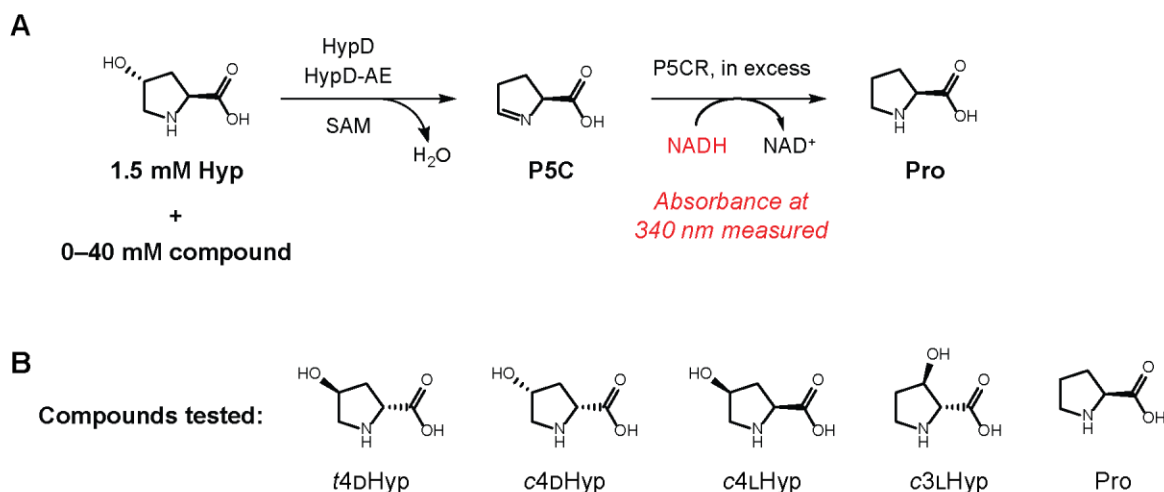


Figure 2.19 Assay designed to test HypD inhibition by Hyp structural analogs.

(A) HypD activity was measured using kinetic assay conditions previously described except that Hyp concentration was held constant at 1.5 mM. This concentration was selected based on the K_m value of 1.2 mM for HypD. A second compound was added to a final concentration of 0–40 mM. Hyp and the compound of interest were pre-mixed and added to assay mixtures to initiate catalysis. (B) Five compounds were tested for HypD inhibition: *t*4DHyp, *c*4DHyp, *c*4LHyp, *c*3LHyp, and Pro.

Although high concentrations in the millimolar range were tested for all compounds, only *c*4LHyp and Pro exhibited inhibition of HypD in this assay (Figure 2.20A). Therefore, *c*4LHyp and Pro appear to be weak inhibitors of HypD. Because HypD activity is coupled to P5CR activity in this assay, it is unclear whether the decreases in turnover rates were due to inhibition of HypD, P5CR, or a combination of both. Particularly because P5CR can bind and oxidize Pro to P5C under non-physiological conditions, we repeated the assay with a higher concentration of P5CR to see if an effect on the kinetic readouts could be observed. An increase in the rate of NADH consumption

would support inhibitory effects of Pro toward P5CR, although this does not preclude inhibition of HypD as well. When P5CR concentration was increased 10-fold, a slight increase in initial rates was observed at the lowest Pro concentration tested (Figure 2.20B). This may indicate product inhibition of P5CR by Pro, but doesn't preclude the possibility of HypD inhibition. To determine if HypD is inhibited by Pro or *c*4LHyp, direct detection of Hyp consumption or P5C production would circumvent the caveats associated with our coupled-enzyme assay. Regardless of the enzymatic step(s) affected, neither amino acids are potent inhibitors (Figure 2.20).

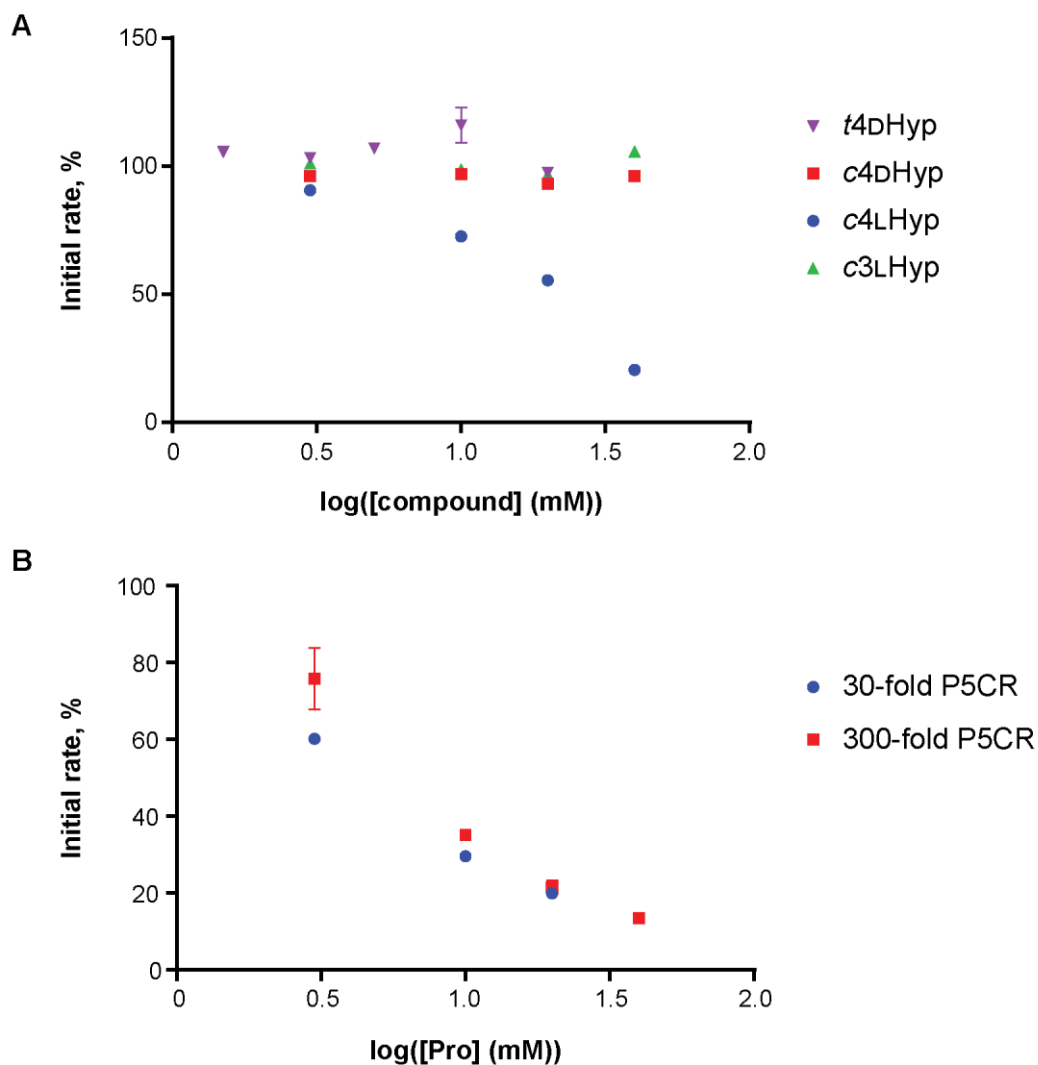


Figure 2.20 Preliminary results from HypD inhibition assays by substrate analogs.

(A) Dose-response curves for inhibition of HypD by structural isomers of Hyp. Inhibition was observed for $c4LHyp$ at concentrations of 10-40 mM. No inhibitory effect was observed with the other compounds. (B) Inhibition of HypD by Pro was observed at concentrations in the millimolar range. P5CR concentrations at 30- and 300-fold excess of HypD concentration were tested. All initial rates are normalized from control conditions lacking the compound being tested. All data points are shown

Figure 2.20 (Continued)

as mean $\pm$ SD with $n = 3$ replicates.

Inhibition observed in the presence of high Pro concentrations seemed to contradict previous observations that an increase in P5CR concentration in the kinetic assay did not affect measured initial rates. This is likely accounted for by an upper range of 400 μ M Pro that can accumulate, which is limited by the 400 μ M NADH included in the assay. Only readings from the first minute of catalysis were included to calculate initial rates and therefore P5C and Pro concentrations are expected to be $< 400 \mu$ M. To be thorough, we revisited the kinetic assay parameters to ensure that dehydration of Hyp was being measured as the rate-limiting step. Kinetic assays were repeated under identical conditions with P5CR concentration held constant or increased by 4-fold (Figure 2.21), which corresponds to final concentrations of 30- or 120-fold excess of HypD. Two kinetic experiments were conducted to compare the rates obtained when this parameter was altered. A higher P5CR concentration resulted in similar results in one experiment, but reduced rates at higher substrate concentrations in a second experiment (Figure 2.21B). k_{cat} and K_{m} values calculated from each set of kinetic data contained some variability, but the overall catalytic efficiencies ($k_{\text{cat}}/K_{\text{m}}$) ranging $4.4\text{--}9.3 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$ are similar to the previously reported value of $3.8 \pm 0.3 \times 10^4 \text{ M}^{-1} \text{ s}^{-1}$ (Figure 2.21B). The slight differences observed could potentially be accounted for by variabilities between protein purifications and experiments. Among all conditions in these experiments, k_{cat} values of $64\text{--}73 \text{ s}^{-1}$ and K_{m} values of $0.7\text{--}1.6 \text{ mM}$ were obtained. A greater standard deviation was observed for the data point collected at the lowest substrate concentration for the kinetic assay with the unusually low K_{m} value of 0.7 mM . In addition, fewer Hyp concentrations were included in this experiment and do not sufficiently cover

the lower range of concentrations making fitting to the Michaelis–Menten equation likely less accurate for these kinetic curves. Overall, no consistent increase in rates was observed with a 4-fold increase in P5CR concentration, which strongly suggests that P5C reduction is not rate-limiting in our assay conditions. We verified that Pro does not accumulate to sufficient levels to inhibit P5CR activity within the timeframe measured in our kinetic assay. Further efforts, however, are required to investigate the possibility of HypD inhibition by its product P5C.

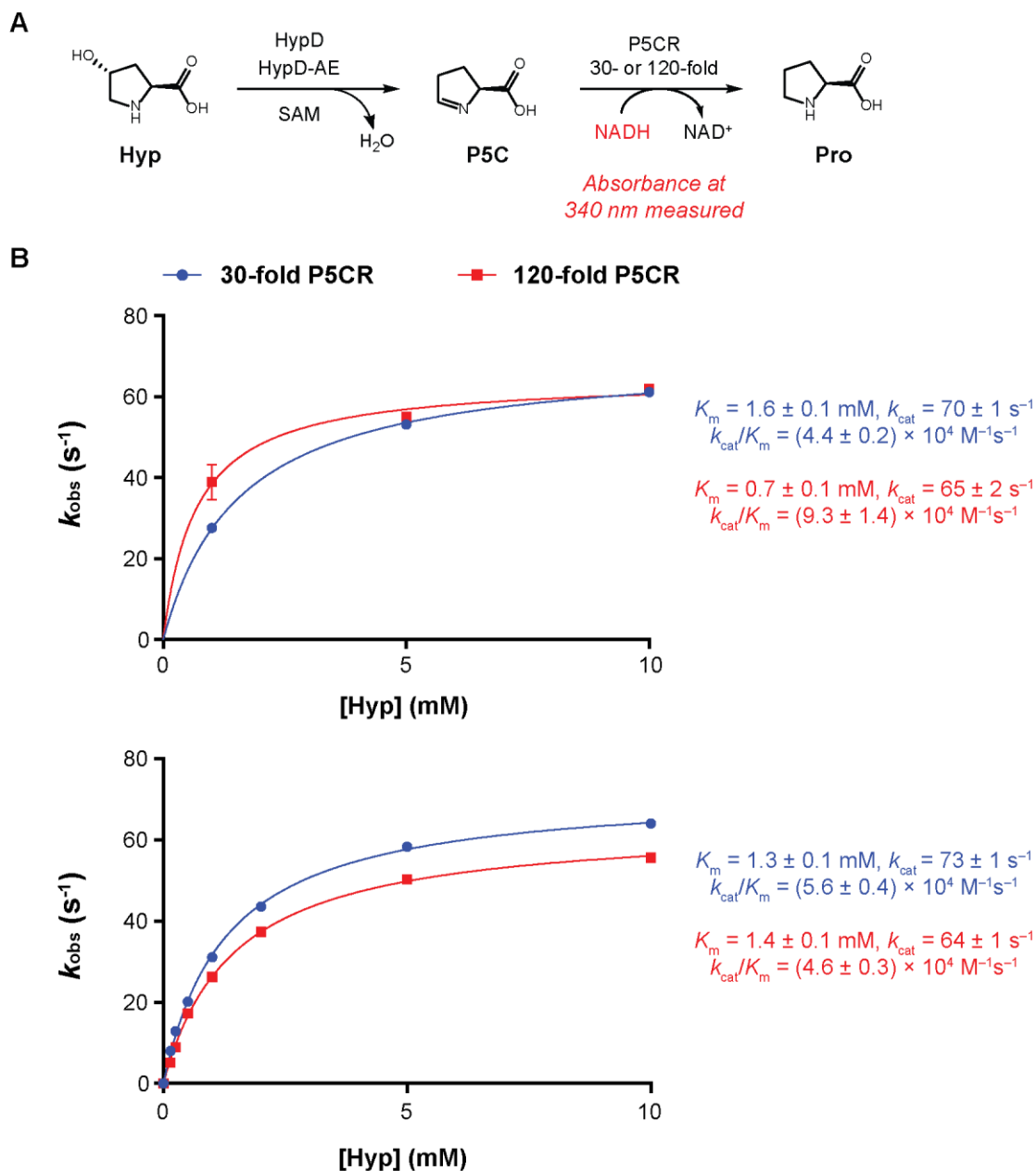


Figure 2.21 Varying P5CR concentration in HypD kinetic assay.

(A) Conversion of Hyp to P5C by HypD was coupled to reduction of P5C to L-proline by P5CR. Absorbance of NADH at 340 nm was continuously monitored to calculate initial rates. P5CR concentrations in 30- and 120-fold excess of HypD concentration (30 nM) were tested. P5CR was added to 30-fold excess in all other kinetic assays.

Figure 2.21 (Continued)

(B) Kinetic data of HypD activity was fit to the Michaelis–Menten equation using nonlinear regression in Graphpad Prism 7.00. Initial rates were converted to k_{obs} assuming 51% HypD activation quantified for HypD from previous purifications. Kinetic parameters are shown next to each plot and displayed as mean $\pm$ SE. Data points are shown as mean $\pm$ SD with $n = 3$ replicates. Each plot consists of experiments carried out on the same day. Aliquots of proteins from the same purifications were used for both experiments.

2.2.9 Utilization of *trans*-4-hydroxy-L-proline by HypD-encoding Clostridiales

We set out to obtain further evidence of HypD's participation in bacterial Hyp metabolism using a culture-based approach. We tested the growth of four sequenced Clostridial strains possessing or lacking HypD in a defined medium designed to be deficient in electron acceptors. All three Clostridiales encoding HypD showed enhanced growth in Hyp-supplemented medium compared to the basal medium lacking Hyp (Figure 2.22A–C). In all cases, strains reached exponential phase sooner and higher maximum optical densities in the presence of Hyp. Growth of these strains was also accompanied by the consumption of Hyp. In contrast, Hyp supplementation did not improve growth of *Clostridium sticklandii*, which does not encode HypD (Figure 2.22D). Correspondingly, Hyp was not consumed in these cultures at the end of growth of this isolate. All four strains selected for this experiment encode the Pro reduction operon to use Pro as an electron acceptor and thus a Pro-supplemented medium was included for comparison. Addition of Pro enhanced growth similar to that observed upon Hyp supplementation in all strains except for *C. difficile*. *C. difficile* reached the maximum optical density more quickly on medium

supplemented with Hyp compared to Pro (Figure 2.22C). Together, these experiments provide support for the hypothesis that HypD is essential for Stickland fermentation of Hyp as an electron acceptor in Clostridiales.

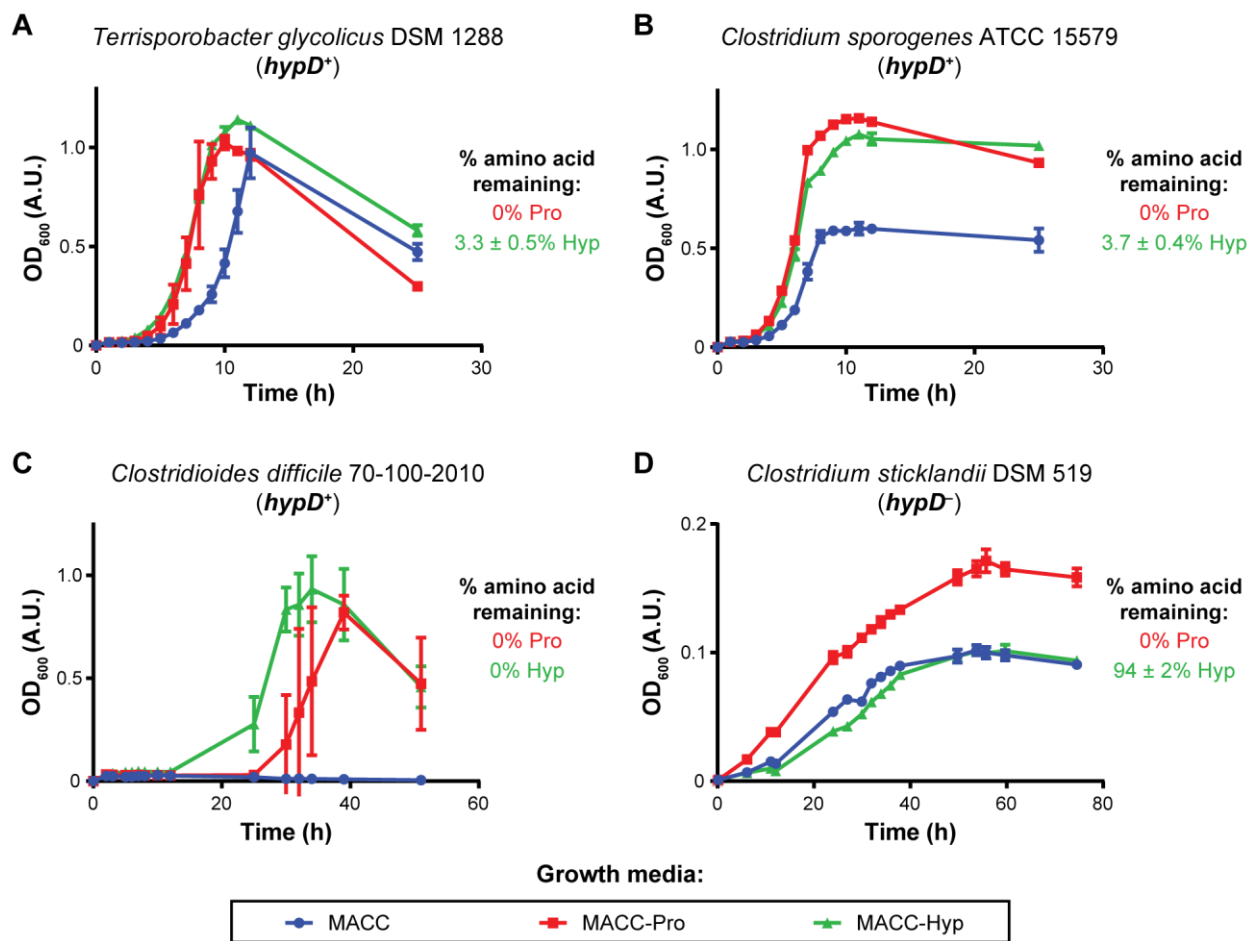


Figure 2.22 Growth of *hypD*⁺ and *hypD*⁻ Clostridial strains in the presence of Hyp or Pro and metabolite analyses.

Four Clostridial strains were grown on a phosphate and carbonate-based defined medium containing 20 mM glucose (MACC). MACC-Pro was supplemented with 20 mM Pro, and MACC-Hyp was supplemented with 20 mM Hyp. Data are shown

Figure 2.22 (Continued)

as mean $\pm$ SD with $n = 3$ replicates with the exception of *C. sticklandii* growth on MACC-Hyp where $n = 2$ replicates. Pro and Hyp were quantified in spent cultures using LC-MS/MS. (A) Addition of Pro and Hyp enhanced growth of *T. glycolicus* DSM 1288, which encodes HypD. (B) Addition of Pro and Hyp enhanced growth of *C. sporogenes* ATCC 15579, which encodes HypD. (C) Addition of Pro and Hyp enhanced growth of *C. difficile* 70-100-2010, which encodes HypD. (D) Addition of Pro enhanced growth of *C. sticklandii* DSM 519, which lacks HypD, whereas addition of Hyp did not affect growth. Hyp was not significantly metabolized by this strain.

2.2.10 Phylogenetic distribution of HypD in sequenced bacterial and archaeal genomes

Given the biological relevance of HypD in amino acid fermentation carried out by Clostridiales, we next sought to determine the distribution of this enzyme among all sequenced organisms. To assess its phylogenetic distribution, a set of representative HypD amino acid sequences generated from a sequence similarity network (SSN) (see next section) was used as search query against the National Center for Biotechnology Information (NCBI) and the InterPro databases (as of February, 2016). A BLASTP³⁸ search located putative HypDs in over 850 sequenced bacterial and archaeal genomes (Figure 2.23, Table 2.1). The top 500 hits of non-redundant amino acid sequences were analyzed and included the highly similar GRE GD. Predicted GDs were therefore manually removed based on the absence of predicted HypD active site residues. Interestingly, HypD-encoding species include several prominent gut and oral commensals (*Parabacteroides* spp., *Bacteroides* spp., and Clostridiales) and human pathogens

such as *C. difficile* and *C. botulinum*. Notably, greater than 97% of sequenced *C. difficile* isolates (NCBI database) seem to possess HypD, implicating Hyp metabolism as a core biological function of this pathogen.

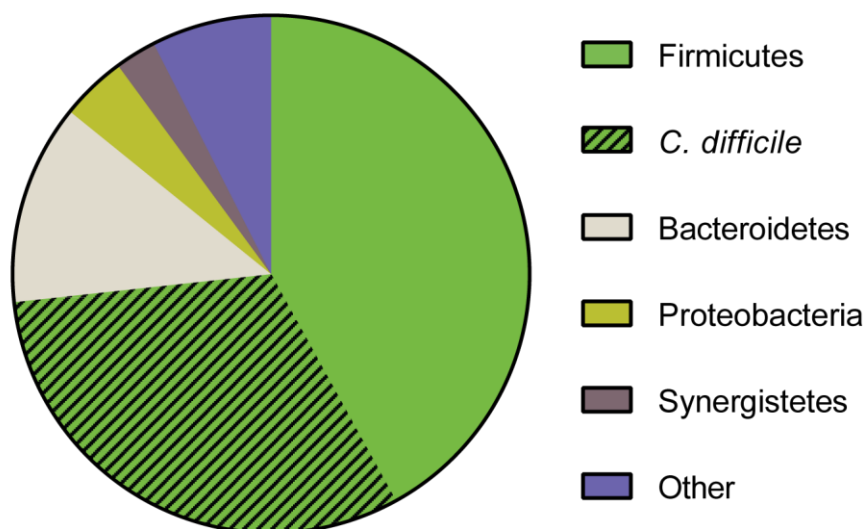


Figure 2.23 Distribution of HypD in sequenced bacterial and archaeal genomes.

A total of 856 genomes from the NCBI and InterPro databases encode HypD. Genomes are grouped at the phylum level with *C. difficile* strains making up close to half of the sequenced genomes from Firmicutes. Examples of phyla in the ‘Other’ category include Euryarchaeota, Spirochaetes, Deferribacteres, Actinobacteria, Thermotogae, and unclassified organisms. See Table 2.1 for a complete breakdown of HypD’s distribution across phyla.

Table 2.1 Distribution of HypD in sequenced bacterial and archaeal genomes at the phylum level.

Values used to generate Figure 2.23.

Phylum	Number of genomes
Firmicutes (<i>C. difficile</i>)	625 (267)
Bacteroidetes	107
Proteobacteria	35
Synergistetes	22
Unknown	18
Euryarchaeota	15
Spirochaetes	7
Deferribacteres	5
Actinobacteria	4
Chloroflexi	4
Thermotogae	3
Fusobacteria	3
Planctomycetes	3
Gemmatimonadetes	3
Latescibacteria	1
Candidatus Acetothermia	1

At the phylum level, HypD seems to be most prevalent among Firmicutes and Bacteroidetes, which are the two most abundant phyla in the human gut microbiome. It is important to note that this analysis biases distribution toward well-studied and highly sequenced human pathogenic isolates, as highlighted with *C. difficile* (Figure 2.23). Likewise, isolates from the human gut have been extensively cultured and sequenced in recent years, which may bias the distribution of HypD toward Firmicutes and Bacteroidetes. In addition to being present in 15 bacterial phyla, HypD was found in 15 genomes from the archaeal phylum Euryarchaeota with

most species in the genera *Thermococcus* and *Aciduliprofundum*. In contrast to HypD being prevalent among bacterial gut isolates, HypD was not found in any human-associated archaea. Instead, these archaeal species are thermophiles and were isolated from extreme environments including hydrothermal vents and oil reservoirs. The presence of HypD among species isolated from a highly diverse range of environments suggests a wide distribution of the substrate Hyp and its importance in microbial growth and survival.

2.2.11 HypD is highly abundant and prevalent in human stool metagenomes

In addition to HypD in sequenced genomes, we also wondered if this enzyme is prevalent in the human microbiome. Our lab recently developed a pipeline termed to separate sequences from a protein superfamily into groups with predicted identical functions and quantification of these functions in metagenomic data. Benjamin Levin and Dr. Spencer Peck used this workflow to profile the distribution of the GRE superfamily in Human Microbiome Project (HMP) metagenomes.¹ⁱⁱⁱ This workflow combines clustering of protein sequences according to their predicted functions using an SSN³⁹ with a quantitative metagenomic tool called ShortBRED (Short, Better Representative Extract Dataset)⁴⁰. Given the unusually high sequence similarity of members within this superfamily, applying our knowledge of GRE biochemistry was crucial for defining a network that best grouped isofunctional sequences together. Sequences were quantitatively profiled against raw reads of metagenomes using ShortBRED to obtain abundance data. This approach allowed us to assess the distribution and abundance of individual members of the GRE

ⁱⁱⁱ : Construction of SSNs for the GRE superfamily was carried out by Dr. Spencer Peck, a former postdoctoral researcher in the Balskus lab. ShortBRED analyses of sequences from the GRE SSNs were performed Benjamin Levin, a graduate student in the Balskus lab. Statistical calculations were completed by both Spencer and Benjamin.

superfamily, both known and uncharacterized, in 378 publicly available HMP metagenomes collected from six body sites.¹

First, 6343 amino acid sequences from InterPro family IPR004184 were retrieved to build an SSN using the web-based Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST).³⁹ These sequences encompass enzymes containing a “PFL domain” and all characterized GREs except the phylogenetically distinct class III ribonucleotide reductases and sequences shorter than 500 amino acids. SSN parameters were iteratively refined to separate biochemically characterized GREs into distinct clusters (Figure 2.24). Each node in the SSN represents sequences that are >95% identical at the amino acid level. Edges connect nodes that have pairwise sequence identities of $\geq 62\%$ and nodes connected by edges make up a cluster in the SSN. Based on prior knowledge of well-studied GREs, conserved active site residues were used to assign each cluster to its predicted activity. Additionally, the genomic contexts of characterized GREs, such as the presence of colocalized microcompartment shell protein-encoding genes, were critical to verify the separation of highly similar GRE eliminases. The final SSN successfully resolves known GRE eliminases, GD, PD, CutC, and HypD, supporting that these optimized parameters mostly cluster sequences with identical biochemical activities together. It is notable that a vast majority of the GRE superfamily contains unexplored activities with 195 of the 241 clusters having no obvious assignable function based on experimental studies (Figure 2.24).

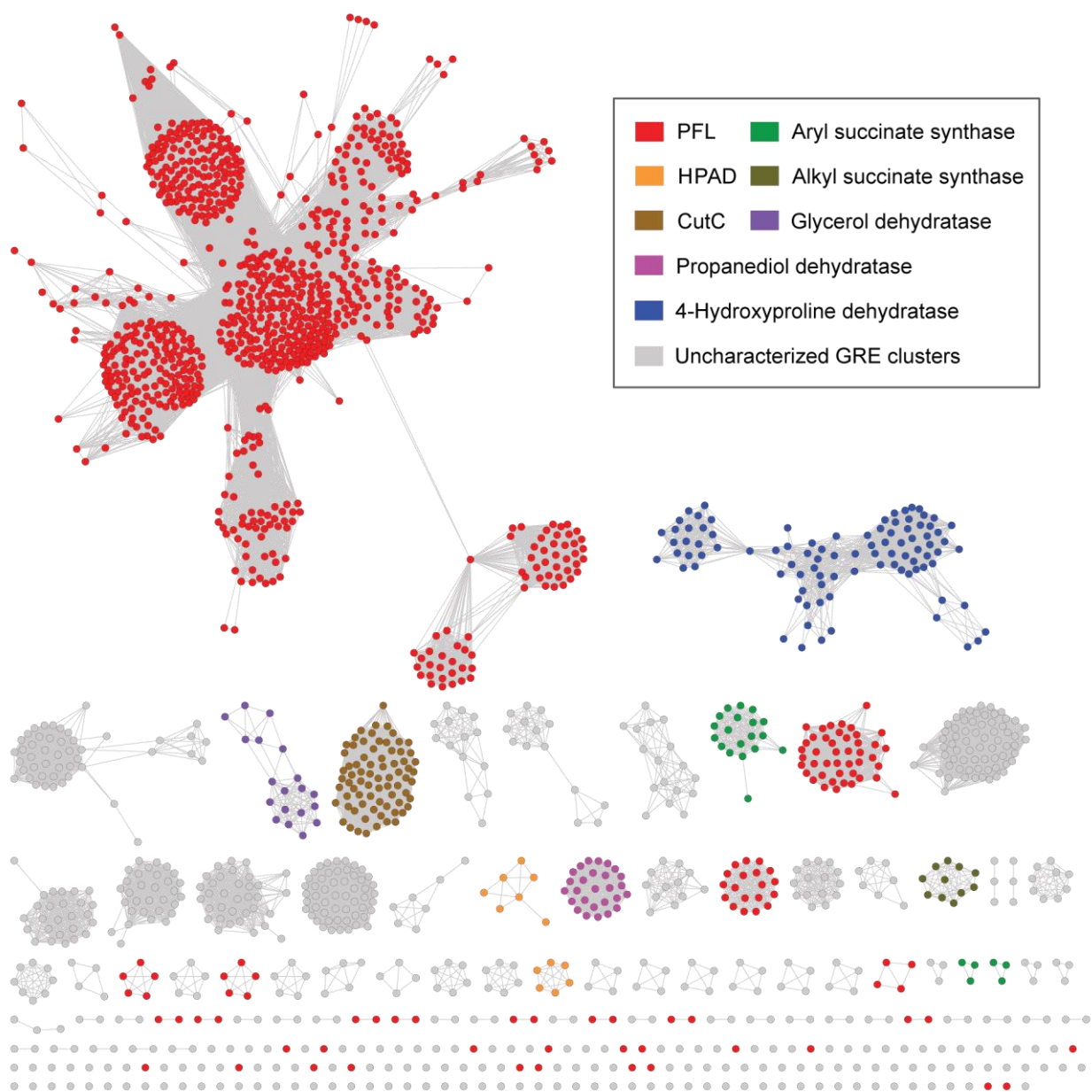


Figure 2.24 Sequence similarity network (SSN) of the GRE superfamily.^{iv}

An SSN of the GRE superfamily (InterPro version 53.0; IPR004184, PFL domain) was constructed such that nodes are connected by an edge if the pairwise sequence identity is $\geq 62\%$ ID. Each of the 1843 nodes within this SSN contains sequences with $>95\%$ amino acid identity. This SSN does not include class III

Figure 2.24 (Continued)

ribonucleotide reductases nor sequences shorter than 500 amino acids. An updated version of this SSN was recently published by the Gerlt group.⁴¹

Sequences in the SSN were then used as query to profile GRE abundance in 378 HMP metagenomes sequenced from healthy participants.⁴² These metagenomes consist of samples collected from six body sites: stool (representative of the GI tract), buccal mucosa (oral), supragingival plaque (oral), tongue dorsum (oral), anterior nares (skin), and posterior fornix (vaginal). ShortBRED first generated short protein markers for GREs grouped at a high sequence similarity of 85% ID. These markers were selected from protein sequence regions that can distinguish each group of proteins. Sequences encoding these unique markers were then quantified in unassembled metagenomic reads by ShortBRED. Sequences from 75 of 241 SSN clusters were detected in the HMP datasets. The read abundances of these markers were mapped to each cluster, tabulated, and normalized by using previously calculated microbial genome sizes for these metagenomes (Figure 2.25A).⁴³

This analysis revealed that GREs are highly enriched in stool metagenomes, which is reflective of the lower GI tract (Figure 2.25A). This result was expected due to the high sensitivity of GREs to molecular oxygen and the gut being the most anoxic environment of the body sites analyzed here. This also matches previous detections of GREs as an abundant protein family in stool metagenomes and metaproteomes.⁸⁻¹⁰ Oral sites also host many GRE-encoding microbes due to its microaerobic environment and the presence of anaerobes in these microbial communities. A few PFLs are widely spread across both aerobic and anaerobic body sites, which can be explained

^{iv} Prof. Emily Balskus and Dr. Spencer Peck generated the original figure, a modified version of which is shown here.

by its presence in facultative anaerobes and existence of a repair mechanism to salvage activity from oxidatively cleaved enzymes.^{44,45}

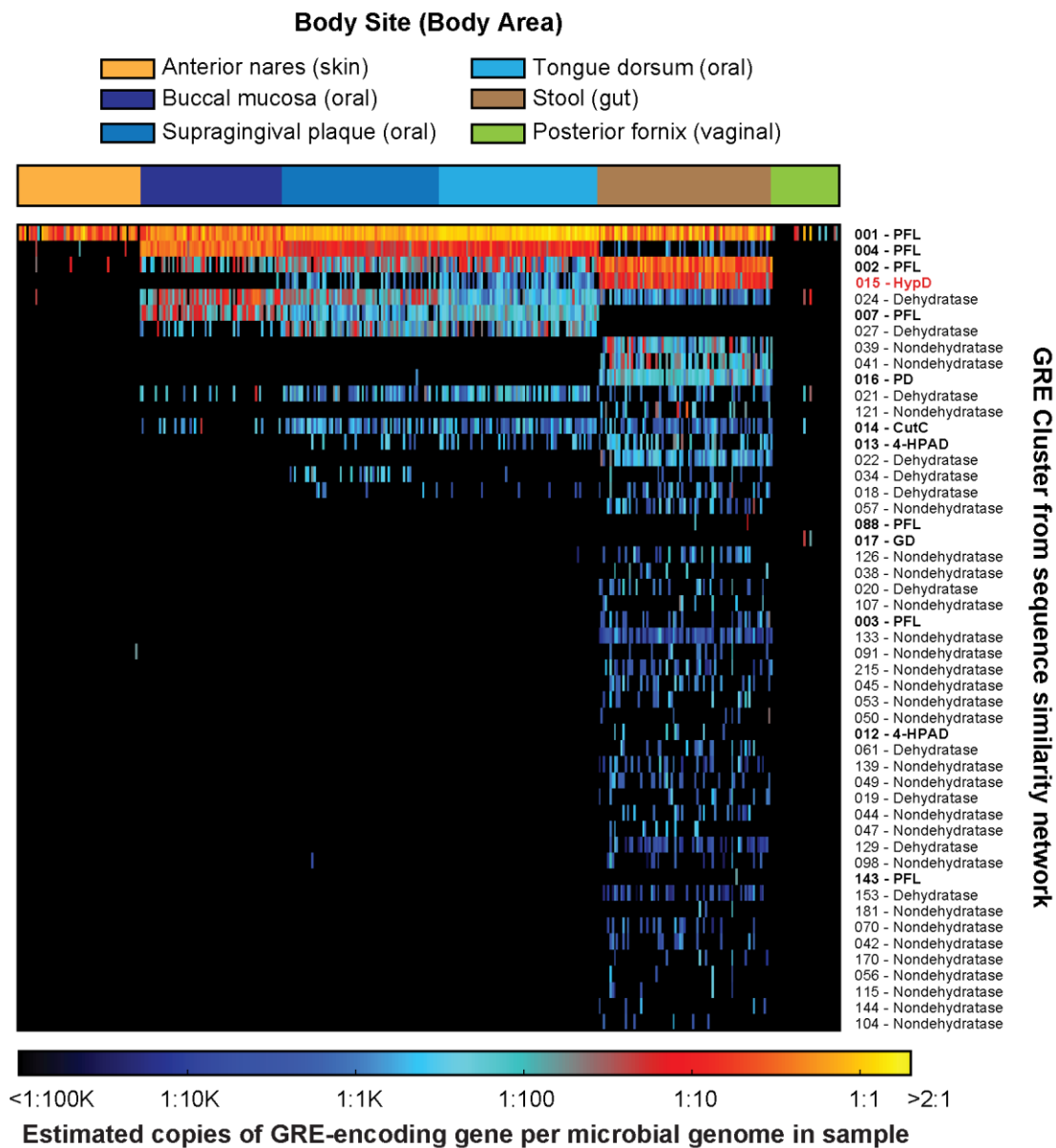
Interestingly, the second most abundant GRE detected in HMP metagenomes was HypD. Paired stool metagenomes and metatranscriptomes from eight healthy human subjects were profiled to reveal that HypD was present and transcribed in these samples.⁴⁶ This demonstrated that the presence of *hypD* in metagenomes likely correspond to the presence of active enzymes in the gut habitat. HypD was found to be universally prevalent in the human gut microbiome, being present in all 80 stool metagenomes included in the analysis (Figure 2.25B). It is also highly abundant in the gut with a median of 0.11 *hypD* copies per microbial genome. Two oral sites, tongue dorsum and supragingival plaque, are also inhabited by HypD-encoding species. The wide distribution of HypD among HMP metagenomes reflects the large number of sequenced gut isolates previously found to encode HypD. The essential role of HypD in anaerobic Hyp metabolic pathway and its presence in all stool metagenomes analyzed strongly support its activity as a core function of the gut microbiota. However, the analysis described here is limited to sequencing datasets obtained from healthy individuals in Western populations and therefore it remains unclear if HypD is universally distributed in other populations.

Figure 2.25 Abundance of GREs in HMP metagenomes.^v

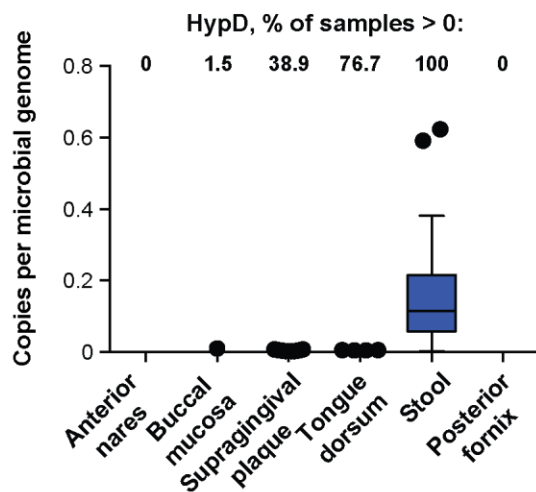
(A) Heatmap showing the abundance and distribution of the 50 most abundant GRE clusters in 378 HMP metagenomes⁴² from six body sites as quantified using ShortBRED⁴⁰. Previously characterized GRE clusters are shown in bold type. HypD, the subject of this thesis work, is highlighted in red along with the recently characterized PD. (B) Abundance (copies per microbial genome) and prevalence (%) of Hyp across the six body sites sampled for the HMP metagenomes.

^v Benjamin Levin generated the original figures, modified versions of which are shown here.

A



B



2.3 Conclusions

The work described in this Chapter led to the discovery of a new enzymatic activity, Hyp dehydration, catalyzed by the GRE HypD. This reaction results in elimination of a water molecule and the oxidation of a C–N bond that is unique among GREs. The requirement for molecular oxygen in the synthesis of Hyp by non-heme iron-dependent dioxygenases is notable given the oxygen-sensitivity of GREs and AEs. Therefore, from an evolutionary perspective, HypD may have emerged after the oxygenation of Earth's atmosphere in response to the evolution of this abundant post-translational modification in eukaryotes. The high stereospecificity of HypD toward Hyp over other stereoisomers of 4-hydroxyproline is perhaps unsurprising given the diastereoselectivity of prolyl and proline 4-hydroxylases responsible for hydroxylation of peptidyl or free Pro.⁴⁷

Anaerobic Hyp degradation was first observed several decades ago in Clostridiales as part of Stickland fermentation for energy production, but the enzymes responsible for this pathway remained unknown until our work. Interestingly, a search in sequenced genomes revealed the presence of HypD across several microbial phyla. Most of these HypD-encoding species have not been reported to metabolize Hyp anaerobically. Given the diversity of these organisms, future work will likely uncover HypD-mediated pathways that are distinct from amino acid fermentation.

In the context of the human microbiome, HypD mediates a previously unappreciated host-microbe metabolic interaction. As discussed in Chapter 1, the substrate Hyp is abundant and commonly found in host and dietary proteins. Dietary sources of Hyp include meat products, which are high in collagen content, and plant-based substrates such as hydroxyproline-rich glycoproteins.^{47,48} Although Hyp is the product of a common post-translational modification of Pro in eukaryotic proteins, this amino acid is rarely synthesized by bacteria.⁴⁷ Therefore, Hyp is

an abundant host-derived amino acid and a major source of nutrient for gut microbes. Intriguingly, the K_m of HypD (1.2 ± 0.1 mM) is within range of the >1 mM concentration of free Hyp detected in the large intestinal content of a human individual.⁴⁹ However, since these measurements came from one subject, which may not be representative of the population and does not account for concentration of Hyp in peptides. Additional quantitative studies are needed to determine the range of free and peptidyl Hyp present in the gut.

The human body breaks down Hyp oxidatively into smaller molecules, pyruvate and glyoxylate, without regenerating Pro as an intermediate.⁵⁰ In contrast, the combined activities of HypD and P5CR enable microbes to reverse a post-translational modification to regenerate Pro from Hyp. In addition to having a role in Stickland fermentation where Hyp is used as an electron acceptor, the products P5C and Pro could be a source of carbon and nitrogen for the gut microbiota. P5C is also a central metabolite in amino acid biosynthetic pathways and thus could be converted into different amino acids to supply building blocks for protein synthesis (further discussed in Chapter 5). Finally, microbial conversion of Hyp to P5C or Pro could affect bioavailability of these key metabolites for the host. Hyp metabolism has implications in the levels of glyoxylate, which can be oxidized to oxalate, a metabolite that can contribute to kidney and bladder stone formation. Furthermore, human Pro metabolism is known to impact cell stress responses, autophagy, apoptosis, and tumor cell survival.⁵¹⁻⁵³

Importantly, the discovery of HypD has allowed us to identify hundreds of new bacterial isolates capable of Hyp metabolism and to study the distribution of this activity in sequence datasets. Its universal prevalence and high abundance in HMP stool metagenomes and sequenced genomes strongly suggest that HypD is an important function for microbes in the human gut. Interestingly, its presence in both gut commensals and pathogens hints at a role in competition for

nutrient and colonization resistance within microbial communities. Further experiments need to be carried out to explore the many potential biological implications of this activity.

2.4 Materials and methods

All chemicals and solvents were purchased from Sigma-Aldrich, except where otherwise noted. Clostridial strains were purchased from ATCC and DSMZ. Luria-Bertani Lenox (LB) medium was purchased from EMD Millipore or Alfa Aesar. DifcoTM Reinforced Clostridial Medium (RCM) was purchased from BD Difco. DNA sequencing results and multiple sequence alignments were analyzed with Geneious Pro⁵⁴ or Clustal Omega²³. Multiple sequence alignments were visualized with Jalview⁵⁵ or Geneious Pro⁵⁴. Primers were purchased from Integrated DNA Technologies (Coralville, IA). PCR was performed with a MyCycler gradient cycler (Bio-Rad) or a C1000 Gradient Cycler (Bio-Rad). All PCR amplifications were analyzed by 1% agarose gel electrophoresis with ethidium bromide staining in 1×TAE buffer. PCR products and digested DNA were purified using an Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare). The identities of all plasmids were confirmed by DNA sequencing (Beckman Coulter Genomics). All restriction enzymes, ligases, polymerases, and PCR mixes were obtained from New England Biolabs. Protein solutions were routinely denatured at 90 °C for 10 min in equal volume Laemmli sample buffer (BioRad) prior to visualization by SDS-PAGE (4-15% Tris-HCl gel, Bio-Rad). 2-Mercaptoethanol was added at a final concentration of 355 mM to Laemmli sample buffer. Fractions from protein purifications were routinely visualized by SDS-PAGE following staining (Biosafe Coomassie, Bio-Rad). Isopropyl β-D-1-thiogalactopyranoside (IPTG) was obtained from Teknova. Ni-NTA and TALON resin were obtained from Qiagen and Clontech, respectively. All absorbance measurements in 96-well plates were carried out using a PowerWave HT Microplate Spectrophotometer (Biotek) inside of an anaerobic chamber (MBraun) for GRE kinetic assays or

outside of a chamber for Pro dehydrogenase assays. All absorbance data shown for kinetic assays and growth experiments were obtained with pathlength corrected to 1 cm.

Samples were rendered anaerobic as follows. Solids were brought into anaerobic chambers (MBraun and Coy Laboratory) in perforated 1.7 mL microcentrifuge tubes. Protein solutions with volumes greater than 1 mL were made anaerobic on a Schlenk line with 3 cycles of evacuation on vacuum (5 min) followed by filling with argon (5 min). Buffers and other solutions were rendered anaerobic in 1 to 20 mL volumes by sparging argon or nitrogen through the liquid for about 30 min. Media solutions of 200 to 300 mL in volume were first microwaved until boiling, followed by sparging nitrogen through the liquid for 1 h. Inside an anaerobic chamber, aliquots were dispensed in smaller volumes into 16×125 mm Hungate tubes, which were then autoclaved. Unless otherwise stated, routine culturing of all anaerobes was carried out at 37 °C in an anaerobic chamber (Coy Laboratory) under an atmosphere of 5% H₂/95% N₂. All glycerol stocks of bacterial strains were stored in rich media at –80 °C.

2.4.1 Construction of HypD homology model

The HypD amino acid sequence from *C. difficile* 70-100-2010 (locus tag: HMPREF9945_00950),) was used as the query for an HHpred analysis to identify templates in the PDB database (pdb70_03Jun16) for homology modelling.^{24,28} The multiple sequence alignment generation was carried out using the HHBlits method.^{28,56}

Hits with the highest secondary structure scores were characterized GREs PD (PDB ID: 5I2A)⁵⁷, glycerol dehydratase (GD) (PDB ID: 1R9D)²⁵, and CutC (PDB ID: 5A0U)²². A homology model of HypD was constructed using Modeller v9.16^{27,28} with GD from *Clostridium butyricum* (1R9D) as the template. The HypD homology model and the template GD structure are highly similar (RMSD = 0.141 Å, C_α alignment of 789 (HypD) vs 786 (GD) residues).

2.4.2 Compilation of sequenced bacterial and archaeal genomes encoding putative HypDs

UCLUST was employed^{vi} to generate a set of representative sequences of HypD with clustering at 60% ID using all HypD sequences from the SSN (>62% ID) as the input. Representative sequences were queried against the NCBI non-redundant protein sequence database using the BLASTP algorithm (performed on 2016-02-16).³⁸ The top 500 unique sequence hits were aligned using the Clustal Omega alignment tool in Geneious.⁵⁴ A total of 391 sequences were identified from the alignment that contained all of the predicted active site residues present in the HypD homology model. A total of 853 deposited genomes encoding these sequences were retrieved from the NCBI database. Three additional isolates encoding HypD were identified from the HypD cluster in the SSN (InterPro database) and were added to those identified from the NCBI database to yield a total of 856 genomes.

2.4.3 Construction of plasmids for overexpression of HypD, HypD-AE, and P5CR^{vii}

Genomic DNA was purified from *C. difficile* 70-100-2010 (BEI Resources, USA) and was used for PCR amplification of genes encoding HypD (locus tag: HMPREF9945_00950), HypD-AE (locus tag: HMPREF9945_00949), and P5CR (locus tag: HMPREF9945_00951). Primers used to amplify these genes are shown in Table 2.2. PCR was carried out using Phusion-HF polymerase according to the manufacturer's protocol. PCR products of HypD and P5CR were purified and ligated into pET28a to yield N-terminal His₆-tagged constructs. HypD-AE was ligated into pSV272-PfMBP⁵⁸ to yield an N-terminal His₆-tagged maltose-binding protein (MBP) fusion

^{vi} Benjamin Levin performed the clustering using UCLUST.

^{vii} Dr. Yifeng Wei carried out the cloning work to construct the overexpression plasmids for HypD, HypD-AE, and P5CR.

construct, PfMBP-CdHypDAE. The NdeI cleavage site (CATATG) at the 5' start codon of the His₆ tag in pSV272-PfMBP had been mutated to CACATG by site-directed mutagenesis to prevent digestion by NdeI and removal of the His₆-tagged MBP during cloning. Both pET28a and pSV272-PfMBP were linearized with NdeI and EagI. All vectors were constructed by Gibson isothermal assembly according to the manufacturer's protocol to yield pET28a-CdHypD, pET28a-CdP5CR, and pSV-PfMBP-CdHypDAE. Plasmids were transformed into chemically competent *E. coli* BL21-CodonPlus (DE3)-RIL *ΔproC::aac(3)IV* (Apramycin^R or Am^R). All plasmid sequences are deposited on Addgene (#110288: pET28a-CdHypD, #110295: pET28a-CdP5CR, #110296: pSV-PfMBP-CdHypDAE).

Table 2.2 Primers used for cloning in this study.

Primer	Sequence (5'–3')
<i>EcP5CR-aac(3)IV</i> -Fwd	GTATACACCCCCTCCCCGGATAAAGTCGCCGCCCTCCATGTGTAGGCT GGAGCTGCTTC
<i>EcP5CR-aac(3)IV</i> -Rev	AATGGTGGTGCCTCCCGGTGAGCAGACCATATCTTTCAGCCGGCCTTT GAATGGGTTCATG
<i>CdHypD</i> -Fwd	CCTGGTGCCGCGCGGCAGCCATATGGCAAGAGGAACCTTTTGAGAGAA CTAAAAAATTAAG
<i>CdHypD</i> -Rev	GGTGCTCGAGTGCGGCCGCAAGCTTTAGAATGTTTGCTCAGTTCTTCCT ATTATCTCATC
<i>CdHypD</i> AE-Fwd	CCTGTATTTTCAGGGCGCCCATATGAATCCATTAGTTATAAACTTACA AAAATGTAGC
<i>CdHypD</i> AE-Rev	GCCCGTTTGATCTCGAGTGCGGCCGTTACCCTCCAATCTTAGTATTGAA ATTACTTGC
<i>CdP5CR</i> -Fwd	CCTGGTGCCGCGCGGCAGCCATATGAAAACCTTTAGGATTTATTGGTTC AGGG
<i>CdP5CR</i> -Rev	GGTGCTCGAGTGCGGCCGCAAGCTTATTTACTCATATCTTTAGACTTAT CTATACAAGC

2.4.4 Generation of a P5CR deletion mutant in *E. coli*

Deletion of *proC* from the BL21-CodonPlus(DE3)-RIL strain used for protein overexpression was necessary to remove contaminating *E. coli* P5CR activity in purified HypD and HypD-AE solutions. *E. coli* BW25113/pKD46 (Coli Genetic Stock Center, #7739) was first mutated to *E. coli* BW25113 $\Delta proC::aac(3)IV(Am^R)$ to be used as the donor strain in P1 transduction. *E. coli* BL21-CodonPlus(DE3)-RIL was used as the recipient strain to generate *E. coli* BL21-CodonPlus(DE3)-RIL $\Delta proC::aac(3)IV(Am^R)$ for protein overexpression.

The apramycin resistance marker, *aac(3)IV*, was PCR amplified from vector pIJ773⁵⁹ and contain regions flanking the *proC* gene in *E. coli* BW25113 for homologous recombination (Table 2.2). PCR mixes contained Phusion-HF PCR Master Mix, primers, and purified pIJ773. PCR product was excised and purified from 1% agarose gel. The purified solution (50 μ L) was digested with 1 μ L DpnI and 6 μ L CutSmart buffer (New England Biolabs) for 1 h at 37 °C. The digested PCR product was purified and stored at –20 °C.

A frozen glycerol stock of *E. coli* BW25113/pKD46 was streaked onto LB agar plate containing 100 μ g mL⁻¹ ampicillin (Amp-100) and incubated at 30 °C overnight. A single colony was inoculated into 2 mL LB-Amp100 and grown overnight at 30 °C. This overnight culture was diluted 1:100 into 5 mL LB-Amp100 and 10 mM arabinose to prepare electrocompetent cells. This culture was incubated at 30 °C for 4 h with shaking at 220 rpm before harvesting by centrifugation (3,220 g, 10 min, 4°C). The cell pellet was washed twice with 10 mL and then once with 5 mL chilled 10% (v/v) glycerol, with pelleting in between (3,220 g, 10 min, 4°C). 110 ng *aac(3)IV* PCR product was added to 100 μ L cell resuspension and incubated on ice for 1 min. Cells were electroporated in a 1 mm cuvette at 1.8 kV for 5.1 ms. Electroporated cells were resuspended in 1 mL LB medium and recovered at 37 °C for 3 h. The recovered cells were plated onto LB agar

plates containing 25 $\mu\text{g mL}^{-1}$ Am and grown at 37 °C. Colonies were screened for successful homologous recombination by colony PCR.

P1 transduction was performed to generate the *E. coli* BL21-CodonPlus(DE3)-RIL $\Delta\text{proC}::\text{aac}(3)\text{IV}(\text{Am}^{\text{R}})$ strain. An *E. coli* BW25113 $\Delta\text{proC}::\text{aac}(3)\text{IV}(\text{Am}^{\text{R}})$ starter culture was diluted 1:100 into 2 mL LB medium containing 5 mM CaCl_2 , 10 mM MgCl_2 , and 0.2% (w/v) glucose, and incubated at 37 °C, 200 rpm for 2 h. Two drops of P1 lysate were added to the culture, which was incubated at 37 °C with shaking at 200 rpm until the culture clarified. 50 μL of chloroform was added to the tube and the tube was vortexed. Cells were pelleted (3,220 g, 10 min) and P1 lysate supernatant was collected and stored at 4 °C. Cells from an overnight culture of *E. coli* BL21-CodonPlus(DE3)-RIL were resuspended in 550 μL of buffer (100 mM MgCl_2 , 5 mM CaCl_2). 10^{-1} to 10^{-4} serial dilutions of P1 lysate of *E. coli* BW25113 $\Delta\text{proC}::\text{aac}(3)\text{IV}(\text{Am}^{\text{R}})$ were prepared with LB containing 5 mM CaCl_2 . 100 μL of resuspended CodonPlus cells was mixed with 100 μL of P1 lysate at each dilution and all samples were incubated at 37 °C for 20 min. 1 mL LB medium with 50 mM sodium citrate was added to each solution and cells were recovered at 30 °C for 2 h. Cells were harvested and washed with 100 μL LB containing 50 mM sodium citrate twice. Resuspended cells were plated on LB-Cam25-Am25 agar plates and were incubated at 37 °C overnight. Colonies were restreaked on LB-Cam25-Am25 agar plates three times to remove any remaining P1 phage. The *proC* deletion was confirmed by colony PCR and sequencing.

2.4.5 Overexpression and purification of HypD

A frozen stock of *E. coli* BL21-CodonPlus(DE3)-RIL $\Delta\text{proC}::\text{aac}(3)\text{IV}(\text{Am}^{\text{R}})$ transformed with pET28a-CdHypD was streaked onto an LB-Kan50-Cam25-Am25 agar plate. A single colony was inoculated into 100 mL LB-Kan50-Cam25-Am25 and was grown overnight at 37 °C with shaking at 175 rpm. The starter culture was diluted 1:50 into 4 L Erlenmeyer baffled flasks

containing 2 L LB-Kan50-Cam25. Cultures were incubated at 37 °C at 175 rpm until OD600 = 0.6. HypD overexpression was induced by addition of 0.1 mM IPTG and cultures were incubated at 25 °C for 20 h at 200 rpm. Cells were harvested by centrifugation (7,900 g, 20 min, 4 °C). Cell pellets from 2 L cultures were transferred into 50 mL conical tubes, flash frozen with N₂ (*l*), and stored at –80 °C prior to protein purification.

All purification steps were carried out at 4 °C. 30 mL HypD lysis buffer (20 mM Tris-HCl pH 7.5, 200 mM KCl, 5 mM 2-mercaptoethanol (BME), 1 mM phenylmethylsulfonyl fluoride (PMSF), 0.2 mg mL⁻¹ lysozyme, 1% (w/v) streptomycin sulfate) was added to each frozen cell pellet. Cell pellets were incubated with lysis buffer for 30-60 min until completely thawed on a mixer. Resuspended cells were lysed by passaging three times through a cell disruptor (Avestin EmulsiFlex-C3) at 10,000 psi. Cell debris was removed by centrifugation (20,000 g, 20 min). Clarified lysate was incubated with 10 mL TALON metal affinity resin (Clontech) for 30 min with mixing. The resin was transferred onto a column and the flow through was passed through the resin again. The resin was washed with 40 mL HypD wash buffer (20 mM Tris-HCl pH 7.5, 200 mM KCl, 5 mM BME). Protein was eluted using 3 mL HypD wash buffer supplemented with 5 mM and 60 mL HypD wash buffer supplemented with 150 mM imidazole. Samples were collected at every step of the purification for analysis of purity by SDS-PAGE. Fractions containing protein, typically both the 5 mM and 150 mM imidazole fractions, were combined, concentrated, and dialyzed twice in HypD dialysis buffer (25 mM Tris-HCl pH 7.5, 50 mM KCl, 5 mM DTT) overnight. The dialyzed protein solution was concentrated using a Spin-X® UF 20 mL 30 kDa MWCO concentrator (3,220 g, 4 °C). Concentrated protein solution was incubated on ice and rendered anaerobic using a Schlenk line as described in the general materials and methods section. In an anaerobic chamber (Coy Labs), purified protein solution was passed through a 0.22 µm pore-

size Acrodisc syringe filter to remove precipitates and aliquoted into 0.5 mL cryogenic vials that were then placed in 18×150 mm Hungate tubes that were sealed with butyl stoppers and crimped aluminum seals. Hungate tubes were frozen with N₂ (l) and stored at –80 °C. An average yield of 60 mg L⁻¹ culture was obtained for HypD based on protein concentration determined by Nanodrop. An extinction coefficient of 78,300 M⁻¹ cm⁻¹ was calculated using ProtParam tool.⁶⁰

2.4.6 Overexpression and purification of MBP-AE

A frozen stock of *E. coli* BL21-CodonPlus(DE3)-RIL $\Delta proC::aac(3)IV(Am^R)$ transformed with pSV272-CdHypDAE was streaked onto an LB-Kan50-Cam25-Am25 agar plate. A single colony was inoculated into 85 mL LB-Kan50-Cam25-Am50 and grown overnight at 37 °C with shaking at 175 rpm. The starter culture was diluted 1:100 into 4 L Erlenmeyer baffled flasks each containing 2 L LB-Kan50-Cam25. Cultures were incubated at 37 °C with shaking until OD₆₀₀ = 0.6. HypD-AE overexpression was induced by addition of 0.1 mM IPTG and cultures were incubated at 15 °C for 16 h. Cells were harvested by centrifugation (7,900 g, 20 min, 4 °C).

All purification steps were carried out at 4 °C under aerobic conditions until [4Fe-4S] cluster reconstitution. 30 mL HypD lysis buffer was added to each cell pellet harvested from a 2 L culture. Cell pellets were incubated with lysis buffer for 30 min on a mixer until homogeneous. Resuspended cells were lysed by passaging two times through a cell disruptor (Avestin EmulsiFlex-C3) at 10,000 psi. Cell debris was removed by centrifugation (20,000 g, 20 min, 4 °C). Clarified lysate was passed over 7 mL TALON metal affinity resin by gravity, and the resin was washed with 35 mL HypD wash buffer. Protein was eluted using 35 mL HypD wash buffer supplemented with 5 mM imidazole and 80 mL HypD wash buffer supplemented with 150 mM imidazole. Samples were collected at every step of purification for analysis of purity by SDS-PAGE. Fractions containing protein, typically the 150 mM imidazole eluates, were combined,

concentrated, and dialyzed twice in HypD-AE dialysis buffer (20 mM Tris-HCl pH 7.5, 100 mM KCl, 5 mM DTT) overnight. The dialyzed protein solution was concentrated using 20 mL Spin-X® UF 10 kDa MWCO concentrators (3,220 g, 4 °C). The concentrated protein solution was rendered anaerobic using a Schlenk line as described in the general materials and methods section. In a Coy Lab anaerobic chamber, HypD-AE was incubated with 10 mM DTT, 12 equiv Na₂S • 9 H₂O to protein, and 12 equiv Fe(NH₄)₂(SO₄)₂ • 6 H₂O to protein for 12 h to reconstitute the [4Fe-4S] clusters. Reconstituted MBP-AE was buffer exchanged into anoxic storage buffer (25 mM Tris-HCl pH 7.5, 100 mM KCl, 5 mM DTT) by repeatedly concentrating and diluting in 6 mL Spin-X® UF 30 kDa MWCO concentrators (3,220 g, 4 °C). The concentrated protein solution was passed through a 0.22 µm pore-size Acrodisc syringe filter to remove precipitates. HypD-AE was frozen and stored as previously described for HypD. An average yield of 6.5 mg L⁻¹ culture was obtained for MBP-AE based on protein concentration determined by Nanodrop. An extinction coefficient of 109,550 M⁻¹ cm⁻¹ was calculated using ProtParam tool.⁶⁰

2.4.7 Overexpression and purification of P5CR

A frozen stock of *E. coli* BL21-CodonPlus(DE3)-RIL *ΔproC::aac(3)IV*(Am^R) transformed with pET28a-CdP5CR was streaked onto an LB -Kan50-Cam25-Am25 agar plate. A single colony was inoculated into 5 mL LB-Kan50-Cam25-Am25 and was grown overnight at 37 °C with shaking at 175 rpm. This overnight culture was diluted 1:100 into 50 mL LB-Kan50-Cam25-Am25 and was grown overnight at 37 °C with shaking at 175 rpm. The starter culture was diluted 1:100 into 4 L Erlenmeyer baffled flasks containing 2 L LB-Kan50-Cam35-Am25. Cultures were incubated at 37 °C at 190 rpm until OD₆₀₀ = 0.5. P5CR overexpression was induced by addition of 0.5 mM IPTG and cultures were incubated at 25 °C for 16 h. Cells were harvested by centrifugation

(7,900 g, 20 min, 4 °C). Cell pellets from 2 L cultures were transferred into 50 mL conical tubes, flash frozen with N₂ (*l*), and stored at –80 °C prior to protein purification.

All purification steps were carried out at 4 °C. 30 mL HypD lysis buffer was added to each frozen cell pellet. Cell pellets were incubated with lysis buffer for 30–60 min on a mixer until homogeneous. Resuspended cells were lysed by passaging three times through a cell disruptor (Avestin EmulsiFlex-C3) at 10,000 psi. Cell debris was removed by centrifugation (20,000 g, 20 min). Clarified lysate was slowly passed over 5 mL TALON metal affinity resin. Resin was washed with 35 mL HypD wash buffer. Protein was eluted using 25 mL HypD wash buffer supplemented with 5 mM imidazole and 60 mL HypD wash buffer supplemented with 150 mM imidazole. Samples were collected at every step of purification for analysis of purity by SDS-PAGE. Fractions containing protein, the 150 mM imidazole eluates, were combined, concentrated, and dialyzed twice in P5CR dialysis buffer (20 mM Tris-HCl pH 7.5, 100 mM KCl, 5 mM DTT) overnight. Dialyzed protein solution was concentrated using Spin-X® UF 20 mL 10 kDa MWCO concentrators (3,220 g, 4 °C). The concentrated protein solution was rendered anaerobic using a Schlenk line as described in the general materials and methods section. Purified P5CR was frozen and stored as previously described for HypD. An average yield of 39 mg L⁻¹ culture was obtained for P5CR based on protein concentration determined by Bradford assay using BSA as an external standard.

2.4.8 Visualization of AE on SDS-PAGE under different reducing conditions

Purified and reconstituted AE was stored aerobically at 4 °C overnight prior to visualization by SDS-PAGE. Laemmli sample buffer was added with 5 mM TCEP, 50 mM DTT, or 355 mM BME as reducing agent. Equal volume of Laemmli buffer was mixed with a diluted protein solution. Heat denaturation was carried out at 90 °C for 10 min.

2.4.9 Detection of [4Fe-4S] cluster(s) in HypD-AE through spectrophotometry

A 10 μ M solution of purified HypD-AE in buffer (25 mM Tris-HCl pH 7.5, 100 mM KCl) was reduced with 50 μ M of sodium dithionite (NaDT) for 25 minutes at room temperature in an anaerobic chamber. Absorbance measurements were taken over a range of 250-800 nm with increments of 5 nm in an UV-transparent 96-well plate (UV-Star®). Spectra were also obtained for non-reduced HypD-AE solutions. Each condition was prepared in triplicate.

2.4.10 Determination of iron content in HypD-AE

Iron contents of 6 μ M HypD-AE solutions in 100 μ L volumes were determined using a Ferene-based colorimetric assay.⁶¹ The protocol was carried out with modifications as follows. Reaction mixtures were incubated at 37 °C for 1 h. 200 μ L of each reaction mixture was transferred to a 96-well plate for absorbance measurements. Quantification was carried out using a standard curve of ammonium iron (II) sulfate (0-70 μ M).

2.4.11 Determination of sulfide content in HypD-AE

Sulfide contents of 6 μ M HypD-AE solutions in 200 μ L volumes were determined using a methylene blue formation colorimetric assay.⁶² The protocol was carried out with a few modifications. Assays were performed in microcentrifuge tubes and vortexed between additions of reagents. Mixtures were incubated for 20 min at room temperature after the addition of NaOH. 200 μ L of each reaction mixture was transferred to a 96-well plate for absorbance measurements at 670 nm. Quantification was carried out using a standard curve of sodium sulfide (0-200 μ M).

2.4.12 Glycyl radical detection and quantification by electron paramagnetic resonance (EPR) spectroscopy

HypD was prepared for EPR spectroscopy as follows, with all listed values representing final concentrations. Anoxic HypD and HypD-AE aliquots were brought into an anaerobic chamber (MBraun). Acriflavine (0.1 mM) and HypD-AE (60 μ M) were incubated together in buffer (20 mM Tris-HCl pH 7.5, 100 mM KCl, 50 mM bicine) for 20 min. HypD (15 μ M total monomer concentration) and SAM (1.5 mM) were then added, and the solution was incubated for 2 h. Samples at both steps were incubated at about 10 inches from the MBraun glovebox light. The entire 220 μ L sample was used for glycy radical measurement.

Perpendicular mode X-band EPR spectra were recorded on a Bruker ElexSysE500 EPR instrument fitted with a quartz dewar (Wilma Lab-Glass) for measurements at 77 K. All samples were loaded into EPR tubes with 4 mm outer diameter and 8" length (Wilma Lab-Glass, 734-LPV-7), sealed, and frozen in N₂ (*l*). Data acquisition was performed with Xepr software (Bruker). The magnetic field was calibrated with an external standard of α,γ -bisdiphenylene- β -phenylallyl (BDPA), $g = 2.0026$ (Bruker). The experimental spectra for the glycy radicals were modeled with EasySpin (Version 5.0.22) for MATLAB (MathWorks) to obtain g values, hyperfine coupling constants, and line widths. Spin concentration measurements were performed by numerically calculating the double integral of the simulated spectra and comparing the area to that of a K₂(SO₃)₂NO standard (raw spectra without simulation). This standard was prepared before each set of EPR measurements by dissolving solid K₂(SO₃)₂NO under anaerobic conditions in anoxic 0.5 M KHCO₃ and diluting to a final concentration of 0.3–0.7 mM. To account for any decomposition during dissolution, the concentration was measured at 248 nm ($\epsilon = 1,690 \text{ M}^{-1} \text{ cm}^{-1}$), using a NanoDrop 2000 UV-Vis Spectrophotometer.⁶³ EPR spectra represent the average of 1 to 25 scans and were recorded under the following conditions: temperature, 77 K; center field, 3350 Gauss; sweep width, 200 Gauss; microwave power, 20 μ W; microwave frequency, 9.45 MHz;

modulation amplitude, 0.4 mT; modulation frequency, 100 kHz; time constant, 20.48 ms; conversion time, 20.48 ms; scan time, 20.97 s; receiver gain, 60 dB (for enzymatic assays) or 30 dB (for standards). Normalization for the difference in receiver gain was performed by the spectrometer.

All EPR assays were performed in triplicate. Negative controls in which HypD, HypD-AE, or SAM was excluded, were also carried out. No radical signal was observed in these samples.

2.4.13 Spectrophotometric assay for proline dehydrogenase activity of P5CR

Assay mixtures contained 200 mM sodium bicarbonate pH 10, 50 mM proline, and 10 mM NAD⁺ or NADP⁺ in a final volume of 200 μ L. Assays were initiated by addition of NAD⁺ or NADP⁺ cofactor. The pathlength-corrected absorbance at 340 nm was measured every 15 s over a period of 10 min in a 96-well plate.

Purified HypD or AE were added to a final concentration of 200 μ M in assay mixtures whereas 20 μ L of crude lysates were included in assays looking for contaminating P5CR activity. These assay conditions were carried out once. 5 μ M of purified P5CR was included in assays assessing its substrate and cofactor specificity. Absorbance of buffer blanks was subtracted from absorbance data collected from all assay conditions tested within the same experiment. All assays using purified P5CR were carried out in triplicate.

Crude extracts were obtained using the following protocol. *E. coli* BW13635 (CGSC# 8005) and JW0377-1 (CGSC# 8554) strains were purchased from the Coli Genetic Stock Center (Yale University). Each *E. coli* starter culture was inoculated 1:100 into 100 mL LB-Tet20 (CGSC# 8005) or LB-Kan50 (CGSC# 8554) and incubated at 37 °C for 2 h, 175 rpm; temperature was lowered to 25 °C for incubation overnight. Cells were pelleted through centrifugation and resuspended in 2 mL of lysis buffer (50 mM Tris HCl pH 8, 1 mM PMSF, 0.03% Triton X, 0.2 mg/ml lysozyme).

Cells were lysed through a freeze-thaw cycle with incubation at -80°C for 15 minutes and then thawed slowly at room temperature. 5 mg of deoxyribonuclease I was added to each lysate and incubated on ice for 30 min. 13 mL of buffer (20 mM Tris pH 7.5, 5 mM BME) and streptomycin sulfate (final concentration of 1% w/v) were added and mixed. Lysates were clarified by centrifugation at 20,000 g for 10 min at 4°C . Ammonium sulfate was added to achieve 70% saturation to clarify lysates, which were incubated on a nutating mixer for 20 min at 4°C . Protein precipitates were harvested by centrifugation, dissolved, and buffer-exchanged into 20 mM Tris pH 7.5, 5 mM DTT using a 500 μL 30kDa MWCO concentrator. Lysates were concentrated down to final volumes of 50 to 100 μL .

2.4.14 End-point enzymatic assays for HypD activity

All assays were prepared and incubated inside an anaerobic chamber (MBraun). P5CR coupled enzymatic assays contained 20 mM Tris-HCl pH 7.5, 100 mM KCl, 0.4 mM NADH, 3 μM P5CR, 0.2 mM Hyp, and 0.3 μM HypD. HypD was first activated under conditions described for EPR spectroscopic assays. Activated HypD mix was then diluted to a final concentration of 0.3 μM protein monomer into buffer containing P5CR and NADH. All assays were carried out in triplicate and were initiated by adding Hyp into 2 mL LC-MS vials. Vials were capped immediately and allowed to incubate for 1 h at 22°C (mixtures incubated for 23 h yielded similar results). Upon removal from the anaerobic chamber, reactions were quenched with a $2\times$ volume of methanol and protein precipitates were removed by centrifugation (15,200 g , 10 min). Supernatants were further diluted with water 60-fold for proline detection and 12-fold for hydroxyproline detection by LC-MS/MS.

End-point assays containing *t*4LHyp, *t*4DHyp, *c*4LHyp, *c*4DHyp, and *c*3LHyp as substrates contained 20 mM Tris-HCl pH 7.5, 100 mM KCl, 0.2 mM substrate, and 0.3 μM HypD. HypD

was activated and diluted into assay mixtures as previously described. Assays were initiated by adding the substrate last and were incubated in the anaerobic chamber (MBraun) for 1 h at 22 °C (mixtures incubated for 23 h yielded similar results). All assays were carried out in triplicate and *t4LHyp* was used as the substrate in negative controls. Aliquots were removed from the chamber and NaBH₄ was added to 100 mM followed by a 20 min incubation at rt. Reactions were quenched with a 2× volume of methanol and protein precipitates were removed by centrifugation (15,200 g, 10 min). Supernatants were further diluted with water. A final dilution of 65.5-fold was carried out for proline detection by LC-MS/MS.

2.4.15 LC-MS/MS methods for detection of Pro and Hyp

LC-MS/MS analyses were performed on an Agilent 6410 Triple Quadrupole LC-MS instrument (Agilent Technologies). The proline detection method was performed as described previously using a Luna SCX column (5 µm, 100 Å, 50 × 2.0 mm, Phenomenex) with the following modifications.⁶⁴ The capillary voltage was set to 4000 V, the collision energy to 13 eV, and the fragmentor voltage to 80 V. The drying gas temperature was maintained at 300 °C with a flow rate of 12 L min⁻¹ and a nebulizer pressure of 35 psi.

Proline was monitored using an isocratic flow of 30 mM ammonium acetate (solvent A) and 5% acetic acid (solvent B) at a ratio of 15:85 with a flow rate of 0.4 mL min⁻¹ for 6 min. The hydroxyproline detection method used 200 mM ammonium acetate as solvent A with the other parameters identical to the proline detection method. The mass spectrometer was operated in multiple reaction monitoring (MRM) mode with positive ionization monitoring. Precursor and product ions of m/z 116.1 → m/z 70.1 were monitored for proline and m/z 132.1 → m/z 86.1 were monitored for hydroxyproline. On the Agilent MassHunter Workstation Data Acquisition software, the MS¹ resolution was set to “wide” for proline and “unit” for hydroxyproline. MS² resolution

was set to “unit” for both amino acids. The injection volume for all samples was 3 μ L. Data analysis was performed with Agilent MassHunter Qualitative Analysis software. Amino acid standards were dissolved in water and diluted to various concentrations prior to sample injections. A standard curve was used to calculate proline and hydroxyproline concentrations in samples based on peak integrations. For the stereospecificity assays, standards consisted of aqueous solutions of the corresponding hydroxyproline isomer (200 μ M) and the peak areas of standards were used to calculate percentage of remaining starting material.

2.4.16 Coupled enzyme spectrophotometric assay for HypD kinetics

HypD was first activated using conditions described for EPR spectroscopic assays. Activated HypD was used for coupled enzyme kinetic assays. Kinetic assay mixtures contained 20 mM Tris-HCl pH 7.5, 50 mM bicine pH 7.5, 100 mM KCl, 400 μ M NADH, 900 nM P5CR, and 30 nM HypD (total monomer concentration). Assays were initiated by addition of Hyp to a final concentration of 0, 0.25, 0.5, 1, 2, 5, or 10 mM. The absorbance at 340 nm was measured every 10 s over a period of 5 min in a 96-well plate. Initial rates were calculated from absorbance measurements that decreased linearly. Pathlengths were corrected to 1 cm and absorbance values were converted to concentrations assuming $\epsilon_{340} = 6,220 \text{ M}^{-1} \text{ cm}^{-1}$ for NADH. The average initial rate from 0 mM Hyp assays was subtracted from assays containing Hyp. Similar initial rates were recorded in assays containing higher concentrations of P5CR, confirming that HypD activity was being directly measured. Data used for calculations and the Michaelis–Menten kinetics plot came from assays carried out in triplicate of a single experiment. Additional kinetic curves obtained from multiple experiments using proteins purified from different days resulted in kinetic parameters within the same order of magnitude. Data were fit simultaneously to the Michaelis–Menten equation using nonlinear regression in Graphpad Prism 7.00. The k_{obs} parameter was

calculated based on $51 \pm 1\%$ (mean $\pm$ SE) activation of HypD monomers as determined by EPR spectroscopic assays.

2.4.17 4-Hydroxyproline derivatization and detection on chiral HPLC

Analysis of commercial supplies of hydroxyproline stereoisomers (Sigma-Aldrich) was performed using HPLC on a Dionex Ultimate 3000 instrument (Thermo Scientific). *t*4LHyp, *t*4DHyp, *c*4LHyp, and *c*4DHyp were individually derivatized with N_{α} -(2,4-dinitro-5-fluorophenyl)-L-alaninamide (FDAA) as previously reported except FDAA was added to a final concentration of 3.2 mM in derivatization mixtures.⁶⁵ Derivatized amino acids were diluted 2-fold before injection on the HPLC.

A Lux® Amylose-1 column (5 μ m, 100 Å, 100 $\times$ 4.6 mm, Phenomenex) was used to separate each derivatized stereoisomer using a flow gradient of 0-15% solvent A over 15 min at a flow rate of 1 mL min⁻¹. Solvent A consisted of 10 mM aqueous ammonium acetate and solvent B consisted of acetonitrile with no additive. Derivatized amino acids were monitored by absorbance at 340 nm. An injection volume of 1 μ L was used for all samples.

2.4.18 Coupled enzyme spectrophotometric for inhibition of HypD

HypD was first activated using conditions described for EPR spectroscopic assays. Kinetic assay conditions were identical to those described previously with the following exceptions. Hyp concentration was kept constant at 1.5 mM for all reaction mixtures. Hyp was pre-mixed with the compound of interest and added to assay mixtures to initiate catalysis. Compounds *trans*-4-hydroxy-D-proline (*t*4DHyp), *cis*-4-hydroxy-D-proline (*c*4DHyp), *cis*-4-hydroxy-L-proline (*c*4LHyp), *cis*-3-hydroxy-L-proline (*c*3LHyp), and L-proline (Pro) were tested for inhibition of HypD. These compounds were tested at a final concentration of 0, 1.5, 3, 5, 10, 20, or 40 mM.

Each assay condition was carried out in triplicate within a single experiment for each compound. Turnover rates were calculated assuming $51 \pm 1\%$ (mean $\pm$ SE) activation of HypD. Activity was calculated by normalizing initial rates against rates in conditions excluding a second compound but containing 1.5 mM Hyp. Concentrations of compound tested were plotted on a logarithmic scale against normalized activity using Graphpad Prism 7.02.

2.4.19 Clostridiales growth experiments and metabolite analyses

Terrisporobacter glycolicus DSM 1288, *Clostridium sporogenes* ATCC 15579, *Clostridioides difficile* 70-100-2010, and *Clostridium sticklandii* DSM 519 were cultivated at 37 °C in an anaerobic chamber (Coy Laboratory) under an atmosphere of 5% H₂/95% N₂. *T. glycolicus* DSM 1288, *C. sporogenes* ATCC 15579, and *C. difficile* 70-100-2010 encode HypD, and *C. sticklandii* DSM 519 does not encode HypD. All media were based on an amino acid complete (MACC) medium previously designed for *Clostridium sporogenes* NCIB 8053, except that ATCC® trace mineral supplement and ATCC® vitamin supplement were used.⁶⁶ MACC is a phosphate and carbonate-based medium that contains 10 amino acids but does not contain Pro and Hyp. MACC-Pro was supplemented with 20 mM Pro and MACC-Hyp was supplemented with 20 mM Hyp. All three MACC media formulations were supplemented with 20 mM glucose.

Clostridial strains were streaked out from glycerol stocks onto RCM agar plates inside an anaerobic chamber and incubated at 37 °C. A single colony of each strain was inoculated into 5 mL RCM broth and incubated at 37 °C overnight. These starter cultures were diluted 1:100 into 5 mL of MACC in capped Hungate tubes (16×125 mm). All strains were grown in each medium in triplicate with the exception of *C. sticklandii* for which one MACC-Hyp culture was excluded due to lack of growth. OD₆₀₀ measurements were taken using a Genesys™ 20 Visible Spectrophotometer.

After the final OD₆₀₀ measurement, spent medium was collected from each culture by centrifugation (3,220 *g* × 10 min). Proteins were precipitated by addition of 2× volume of methanol followed by vortexing and centrifugation (3,220 *g* × 10 min). Clarified supernatants were diluted either 500- or 1,000-fold with water for Pro and Hyp detection via LC-MS/MS following methods described in previous sections. Uninoculated media served as blanks. The percentages of Pro and Hyp remaining in spent media were normalized to media blanks using peak area integrations.

2.4.20 Construction of the GRE sequence similarity network^{viii}

All sequence similarity networks in the text were created as described previously.³⁹ In brief, the InterPro family⁶⁷ IPR004184 (pyruvate formate lyase domain, version 53.0, accessed on October 9, 2015) was used as the input for option B of the Enzyme Function Initiative's Enzyme Similarity Tool (found at: <http://efi.igb.illinois.edu/efi-est>). IPR004184 includes every characterized GRE except for ribonucleotide reductase class III (RNR). RNR shares little sequence homology with PFL and other GREs and its uniqueness led to its exclusion from IPR004184 and our analysis.^{68,69} Networks were generated at an initial score value of E-300 with a minimum length of 500 amino acids. To reduce the file size of the networks, the E-300 networks were downloaded as 95% representative node networks in which collections of proteins with ≥95% ID to one another are represented by single nodes. Nodes with similarity greater than the threshold score are connected by an edge. For the E-300 network, this initial edge value was 65 ± 2% ID (mean ± SD). Due to the sheer number of edges within each network, there are still numerous edges that fall well below the initial average edge value (*e.g.*, down to 55% ID in the E-300 network). Cytoscape 3.2.0

^{viii} The method section for generating the GRE SSN was written by Dr. Spencer Peck, a condensed and modified version of which is shown here.

was used to visualize the networks and further refine the threshold scores by raising the minimum identity required for two nodes to be connected.⁷⁰ By scouring the literature and sequence databases, we connected reported GRE activities to sequences within our SSN. Moreover, all functionally characterized formate-lyases (PFLs; >20 instances) contain a catalytically essential Cys-Cys active site motif. Therefore, any sequence containing these residues in a multiple sequence alignment was also annotated as a PFL.

We used multiple sequence alignments in tandem with construction of homology models for HypD and PD to predict the residues that occupied the active sites of these GREs. Only by raising the stringency threshold of our network to 62% ID do GRE dehydratases split apart into different clusters. One cautionary note is that we cannot be certain that all of the clusters in our network are isofunctional with the edge score set to 62% ID. To gain further evidence either way, we constructed multiple sequence alignments of the sequences that are grouped with structurally-characterized GREs including PFL, choline trimethylamine-lyase (CutC), 4-hydroxyphenylacetate decarboxylase (HPAD), benzylsuccinate synthase (BSS), and GD. We identified functionally important residues in the active sites of these crystal structures and checked to see if they were conserved in other sequences from these clusters. In nearly all cases, the residues present in the crystal structure were universally conserved in the other sequences from the same cluster of the SSN. Deviations invariably resulted in conservative mutations. We therefore think it quite likely that groups of proteins within our E-300/62% ID network catalyze the same reaction.

2.4.21 Determination of enzyme abundances in metagenomes and metatranscriptomes^{ix}

^{ix} The method section for ShortBRED analysis was written by Benjamin Levin, a condensed and modified version of which is shown here.

ShortBRED was used to determine the abundance of GREs in metagenomes.⁴⁰ For a set of proteins of interest, ShortBRED groups the proteins at a specified amino acid similarity threshold (here, 85% ID) to identify non-redundant representative sequences. These representative sequences are then compared to a comprehensive, non-redundant protein catalog to identify distinguishing peptide markers among the proteins of interest. Metagenomic reads are then mapped to the identified markers to profile the abundance of their corresponding proteins with high specificity. To find markers to profile the GRE family, the InterPro family IPR004184 was prepared for study by removing redundant (100% amino acid ID) entries and all sequences containing fewer than 500 amino acids were discarded. The remaining sequences were used as input for ShortBRED-Identify and UniRef90 (downloaded on September 22, 2015) was used as the comprehensive protein reference catalog.⁷¹ ShortBRED-Identify was run with the default parameters to identify markers.

Once markers were obtained, ShortBRED-Quantify was used to search for these markers in HMP metagenomic datasets. Shotgun sequencing reads for 378 metagenomes were obtained from the HMP website (www.hmpdacc.org).^{42,72} These samples had previously passed quality control assessment and originated from six different body sites: anterior nares, posterior fornix, buccal mucosa, supragingival plaque, tongue dorsum, and stool. The selected metagenomes were restricted to subjects' first sampling visits, and thus we included no duplicate samples for a given subject within a given body site. ShortBRED-Quantify was also run with the default parameters. The markers file used was the file generated by running ShortBRED-Identify according to the parameters described above.

By default, ShortBRED reports protein abundance in RPKM units (reads mapped per kilobase of coding sequence per million sample reads). We converted these values to units of “copies per

microbial genome” in the following way. First, each ShortBRED RPKM value (C) was converted to an equivalent coverage value (S_{cov}). If H is the number of hits to a given sequence of length L , and R is the average read length for the sample, then coverage can be estimated as:

$$S_{\text{cov}} = \frac{H \times R}{L}$$

Similarly, the average coverage of a genome in the corresponding sample (G_{cov}) can be estimated. Taking N as the total number of reads in a sample and AGS as the sample’s average genome size, the coverage can be expressed as:

$$G_{\text{cov}} = \frac{N \times R}{AGS}$$

The AGS for all HMP samples used in this study were previously computed.⁴³ The ratio of $S_{\text{cov}}/G_{\text{cov}}$ then serves as a measure of the relative copy number of a sequence of interest among genomes in the sample (“copies per microbial genome”). The computation of this ratio reduces considerably to:

$$\frac{S_{\text{cov}}}{G_{\text{cov}}} = \frac{\left(\frac{H \times R}{L}\right)}{\left(\frac{N \times R}{AGS}\right)} = AGS \frac{H}{L \times N} = AGS \frac{H}{\left(\frac{L}{10^3}\right) \times \left(\frac{N}{10^6}\right)} 10^{-9} = AGS \times C \times 10^{-9}$$

The factor of 10^{-9} arises because L and N as used by ShortBRED are in units of kb and million reads respectively.

This analysis outputs a C value for each representative protein sequence. These representatives are chosen by ShortBRED by clustering the input sequences at 85% amino acid identity, so some of these sequences will have the same function. By summing the C values for all protein sequences that belong to the same GRE cluster on the SSN, we obtain C values that measure the abundances of each isofunctional group of GREs. As C and $S_{\text{cov}}/G_{\text{cov}}$ are related by the equation

above, we can determine the total number of GREs with a particular function per microbial genome in a metagenome.

In addition to the metagenomes from the HMP, the distribution and abundance of GREs were measured in metagenomes and metatranscriptomes in stool samples from eight healthy individuals.⁴⁶ The metagenomes were analyzed in the same way as above, except that because the average genome sizes of those samples were unavailable, the RPKM value (C) was used instead of $S_{\text{cov}}/G_{\text{cov}}$. Metatranscriptomes can be processed by ShortBRED in the same manner as metagenomes and the RPKM value (C) was used to determine if particular GREs were being transcribed in those datasets. The heatmap shown in Figure 2.25 was assembled with a custom-written script in hclust2 (<https://bitbucket.org/nsegata/hclust2>).

2.5 References

1. BJ Levin*, YY Huang*, SC Peck, Y Wei, A Martínez-del Campo, JA Marks, EA Franzosa, C Huttenhower, EP Balskus. A prominent glycyl radical enzyme in human gut microbiomes metabolizes *trans*-4-hydroxy-L-proline. *Science*. **2017**, 355(6325):eaai8386.
2. YY Huang, A Martinez-Del Campo, EP Balskus. Anaerobic 4-hydroxyproline utilization: Discovery of a new glycyl radical enzyme in the human gut microbiome uncovers a widespread microbial metabolic activity. *Gut microbes*. **2018**, 9(5):437-51.
3. VB Young. The role of the microbiome in human health and disease: an introduction for clinicians. *BMJ*. **2017**, 356:j831.
4. G Sharon, N Garg, J Debelius, R Knight, Pieter C Dorrestein, Sarkis K Mazmanian. Specialized metabolites from the microbiome in health and disease. *Cell Metabolism*. **2014**, 20(5):719-30.
5. N Koppel, Emily P Balskus. Exploring and understanding the biochemical diversity of the human microbiota. *Cell chemical biology*. **2016**, 23(1):18-30.
6. R Joice, K Yasuda, A Shafquat, Xochitl C Morgan, C Huttenhower. Determining microbial products and identifying molecular targets in the human microbiome. *Cell Metabolism*. **2014**, 20(5):731-41.
7. AM Schnoes, SD Brown, I Dodevski, PC Babbitt. Annotation error in public databases: misannotation of molecular function in enzyme superfamilies. *PLoS Comp Biol*. **2009**, 5(12):e1000605.
8. K Kurokawa, T Itoh, T Kuwahara, K Oshima, H Toh, A Toyoda, H Takami, H Morita, VK Sharma, TP Srivastava, et al. Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes. *DNA Res*. **2007**, 14(4):169-81.
9. CA Kolmeder, M de Been, J Nikkilä, I Ritamo, J Mättö, L Valmu, J Salojärvi, A Palva, A Salonen, WM de Vos. Comparative metaproteomics and diversity analysis of human intestinal microbiota testifies for its temporal stability and expression of core functions. *PLOS ONE*. **2012**, 7(1):e29913.
10. K Ellrott, L Jaroszewski, W Li, JC Wooley, A Godzik. Expansion of the protein repertoire in newly explored environments: Human gut microbiome specific protein families. *PLOS Comput Biol*. **2010**, 6(6):e1000798.
11. T Selmer, AJ Pierik, J Heider. New glycyl radical enzymes catalysing key metabolic steps in anaerobic bacteria. *Biol Chem*. **2005**, 386(10):981-8.
12. LRF Backman, MA Funk, CD Dawson, CL Drennan. New tricks for the glycyl radical enzyme family. *Crit Rev Biochem Mol Biol*. **2017**, 54(6):674-95.

13. AH Deutch, CJ Smith, KE Rushlow, PJ Krelscher. *Escherichia coli* Δ^1 -pyrroline-5-carboxylate reductase: gene sequence, protein overproduction and purification. *Nucleic Acids Res.* **1982**, 10(23):7701-14.
14. J Kenklies, R Ziehn, K Fritsche, A Pich, JR Andreesen. Proline biosynthesis from L-ornithine in *Clostridium sticklandii*: purification of Δ^1 -pyrroline-5-carboxylate reductase, and sequence and expression of the encoding gene, *proC*. *Microbiology.* **1999**, 145:819-26.
15. WF Visser, NM Verhoeven-Duif, TJ de Koning. Identification of a human *trans*-3-Hydroxy-L-proline dehydratase, the first characterized member of a novel family of proline racemase-like enzymes. *Journal of Biological Chemistry.* **2012**, 287(26):21654-62.
16. X Zhang, R Kumar, MW Vetting, S Zhao, MP Jacobson, SC Almo, JA Gerlt. A unique *cis*-3-hydroxy-L-proline dehydratase in the enolase superfamily. *Journal of the American Chemical Society.* **2015**, 137(4):1388-91.
17. LH Stickland. Studies in the metabolism of the strict anaerobes (genus *Clostridium*). I. The chemical reactions by which *C. sporogenes* obtains its energy. *Biochem J.* **1934**, 28(5):1746-59.
18. S Jackson, M Calos, A Myers, WT Self. Analysis of proline reduction in the nosocomial pathogen *Clostridium difficile*. *J Bacteriol.* **2006**, 188(24):8487-95.
19. UC Kabisch, A Gräntzdörffer, A Schierhorn, KP Rücknagel, JR Andreesen, A Pich. Identification of D-proline reductase from *Clostridium sticklandii* as a selenoenzyme and indications for a catalytically active pyruvoyl group derived from a cysteine residue by cleavage of a proprotein. *J Biol Chem.* **1999**, 274(13):8445-54.
20. M Feliks, GM Ullmann. Glycerol dehydration by the B₁₂-independent enzyme may not involve the migration of a hydroxyl group: a computational study. *J Phys Chem B.* **2012**, 116(24):7076-87.
21. BJ Levin, EP Balskus. characterization of 1,2-propanediol dehydratases reveals distinct mechanisms for B₁₂-dependent and glycyl radical enzymes. *Biochemistry.* **2018**, 57(23):3222-6.
22. S Bodea, MA Funk, EP Balskus, CL Drennan. Molecular basis of C-N bond cleavage by the glycyl radical enzyme choline trimethylamine-lyase. *Cell chemical biology.* **2016**, 23(10):1206-16.
23. F Sievers, A Wilm, D Dineen, TJ Gibson, K Karplus, W Li, R Lopez, H McWilliam, M Remmert, J Söding, et al. Fast, scalable generation of high - quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* **2011**, 7(539).
24. A Hildebrand, A Remmert, A Biegert, J Söding. Fast and accurate automatic structure prediction with HHpred. *Proteins.* **2009**, 77(S9):128-32.

25. JR O'Brien, C Raynaud, C Croux, L Girbal, P Soucaille, WN Lanzilotta. Insight into the mechanism of the B₁₂-independent glycerol dehydratase from *Clostridium butyricum*: preliminary biochemical and structural characterization. *Biochemistry*. **2004**, 43:4635-45.
26. B Webb, A Sali. Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinformatics*. **2014**, 47(5.6):1-32.
27. A Sali, L Potterton, F Yuan, H van Vlijmen, M Karplus. Evaluation of comparative protein modelling by MODELLER. *Proteins*. **1995**, 23:318-26.
28. V Alva, S-Z Nam, J Söding, AN Lupas. The MPI bioinformatics toolkit as an integrative platform for advanced protein sequence and structure analysis. *Nucleic Acids Res*. **2016**, 44(W1):W410-W5.
29. RN Costilow, D Cooper. Identity of proline dehydrogenase and Δ^1 -pyrroline-5-carboxylic acid reductase in *Clostridium sporogenes*. *Journal of Bacteriology*. **1978**, 134(1):139-46.
30. I Williams, L Frank. Improved chemical synthesis and enzymatic assay of Δ^1 -pyrroline-5-carboxylic acid. *Analytical Biochemistry*. **1975**, 64(1):85-97.
31. T Baba, T Ara, M Hasegawa, Y Takai, Y Okumura, M Baba, KA Datsenko, M Tomita, BL Wanner, H Mori. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*. **2006**, 2:2006.0008.
32. B Nocek, C Chang, H Li, L Lezondra, D Holzle, F Collart, A Joachimiak. Crystal structures of Δ^1 -pyrroline-5-carboxylate reductase from human pathogens *Neisseria meningitidis* and *Streptococcus pyogenes*. *J Mol Biol*. **2005**, 354(1):91-106.
33. JA Burns, JC Butler, J Moran, GM Whitesides. Selective reduction of disulfides by tris(2-carboxyethyl)phosphine. *The Journal of Organic Chemistry*. **1991**, 56(8):2648-50.
34. SR Weeksler, S Stoll, H Tran, OT Magnusson, SP Wu, D King, RD Britt, JP Klinman. Pyrroloquinoline quinone biogenesis: demonstration that PqqE from *Klebsiella pneumoniae* is a radical S-adenosyl-L-methionine enzyme. *Biochemistry*. **2009**, 48(42):10151-61.
35. U Volker, NF A, K Joachim. The free radical of pyruvate formate-lyase. *European Journal of Biochemistry*. **1989**, 184(3):723-8.
36. S Craciun, JA Marks, EP Balskus. Characterization of choline trimethylamine-lyase expands the chemistry of glycyl radical enzymes. *ACS Chem Biol*. **2014**, 9(7):1408-13.
37. J Knappe, HP Blaschkowski, P GröBner, T Schmttt. Pyruvate formate-lyase of *Escherichia coli*: the acetyl-enzyme intermediate. *European Journal of Biochemistry*. **1974**, 50(1):253-63.
38. SF Altschul, W Gish, W Miller, EW Myers, DJ Lipman. Basic local alignment search tool. *J Mol Biol*. **1990**, 215(3):403-10.

39. JA Gerlt, JT Bouvier, DB Davidson, HJ Imker, B Sadkhin, DR Slater, KL Whalen. Enzyme Function Initiative-Enzyme Similarity Tool (EFI-EST): a web tool for generating protein sequence similarity networks. *BBA-Proteins Proteom.* **2015**, 1854(8):1019-37.
40. J Kaminski, MK Gibson, EA Franzosa, N Segata, G Dantas, C Huttenhower. High-specificity targeted functional profiling in microbial communities with ShortBRED. *PLOS Comput Biol.* **2015**, 11(12):e1004557.
41. R Zallot, NO Oberg, JA Gerlt. 'Democratized' genomic enzymology web tools for functional assignment. *Current Opinion in Chemical Biology.* **2018**, 47:77-85.
42. THMP Consortium. Structure, function and diversity of the healthy human microbiome. *Nature.* **2012**, 486:207-14.
43. S Nayfach, KS Pollard. Average genome size estimation improves comparative metagenomics and sheds light on the functional ecology of the human microbiome. *Genome Biol.* **2015**, 16(1):1-18.
44. J Knappe, G Sawers. A radical-chemical route to acetyl-CoA: the anaerobically induced pyruvate formate-lyase system of *Escherichia coli*. *FEMS Microbiol Rev.* **1990**, 6(4):383-98.
45. AFV Wagner, S Schultz, J Bomke, T Pils, WD Lehmann, J Knappe. YfiD of *Escherichia coli* and Y06I of bacteriophage T4 as autonomous glycyl radical cofactors reconstituting the catalytic center of oxygen-fragmented pyruvate formate-lyase. *Biochem Biophys Res Comm.* **2001**, 285(2):456-62.
46. EA Franzosa, XC Morgan, N Segata, L Waldron, J Reyes, AM Earl, G Giannoukos, MR Boylan, D Ciulla, D Gevers, et al. Relating the metatranscriptome and metagenome of the human gut. *Proc Natl Acad Sci USA.* **2014**, 111(22):E2329-E38.
47. KL Gorres, RT Raines. Prolyl 4-hydroxylase. *Crit Rev Biochem Mol Biol.* **2010**, 45(2):106-24.
48. M Hijazi, SM Velasquez, E Jamet, JM Estevez, C Albenne. An update on post-translational modifications of hydroxyproline-rich glycoproteins: toward a model highlighting their contribution to plant cell wall architecture. *Front Plant Sci.* **2014**, 5:395.
49. EA Smith, GT Macfarlane. Enumeration of amino acid fermenting bacteria in the human large intestine: effects of pH and starch on peptide metabolism and dissimilation of amino acids. *FEMS Microbiol Ecol.* **1998**, 25(4):355-68.
50. E Adams, L Frank. Metabolism of proline and the hydroxyprolines. *Annu Rev Biochem.* **1980**, 49(1):1005-61.
51. JM Phang, W Liu, O Zabirnyk. Proline metabolism and microenvironmental stress. *Annu Rev Nutr.* **2010**, 30(1):441-63.

52. O Olivares, JR Mayers, V Gouirand, ME Torrence, T Gicquel, L Borge, S Lac, J Roques, M-N Lavaut, P Berthezène, *et al.* Collagen-derived proline promotes pancreatic ductal adenocarcinoma cell survival under nutrient limited conditions. *Nature Communications*. **2017**, 8:16031.
53. JM Phang, W Liu, CN Hancock, JW Fischer. Proline metabolism and cancer: emerging links to glutamine and collagen. *Current opinion in clinical nutrition and metabolic care*. **2015**, 18(1):71-7.
54. M Kearse, R Moir, A Wilson, S Stones-Havas, M Cheung, S Sturrock, S Buxton, A Cooper, S Markowitz, C Duran, *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. **2012**, 28(12):1647-9.
55. AM Waterhouse, JB Procter, DMA Martin, M Clamp, GJ Barton. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics*. **2009**, 25(9):1189-91.
56. M Remmert, A Biegert, A Hauser, J Söding. HHblits: Lightning-fast iterative protein sequence searching by HMM-HMM alignment. *Nat Methods*. **2012**, 9:173-5.
57. JW LaMattina, ND Keul, P Reitzer, S Kapoor, F Galzerani, DJ Koch, IE Gouvea, WN Lanzilotta. 1,2-propanediol dehydration in *Roseburia inulinivorans*; structural basis for substrate and enantiomer selectivity. *J Biol Chem*. **2016**, 291(30):15515-26.
58. Y Wei, B Li, D Prakash, JG Ferry, SJ Elliott, J Stubbe. A ferredoxin disulfide reductase delivers electrons to the *Methanosarcina barkeri* class III ribonucleotide reductase. *Biochemistry*. **2015**, 54(47):7019-28.
59. B Gust, GL Challis, K Fowler, T Kieser, KF Chater. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc Natl Acad Sci USA*. **2003**, 100(4):1541-6.
60. E Gasteiger, A Gattiker, C Hoogland, I Ivanyi, RD Appel, A Bairoch. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*. **2003**, 31(13):3784-8.
61. MC Kennedy, TA Kent, M Emptage, H Merkle, H Beinert, E Munck. Evidence for the formation of a linear [3Fe-4S] cluster in partially unfolded aconitase. *J Biol Chem*. **1984**, 259(23):14463-71.
62. H Beinert. Semi-micro methods for analysis of labile sulfide and of labile sulfide plus sulfane sulfur in unusually stable iron-sulfur proteins. *Analytical Biochemistry*. **1983**, 131(2):373-8.
63. JH Murib, DM Ritter. Decomposition of nitrosyl disulfonate ion. I. Products and mechanism of color fading in acid solution. *J Am Chem Soc*. **1952**, 74(13):3394-8.

64. B Thiele, N Stein, M Oldiges, D Hofmann. Direct analysis of underivatized amino acids in plant extracts by LC-MS-MS. *Methods Mol Biol.* **2012**, 828:317-28.
65. T Langrock, N García-Villar, R Hoffmann. Analysis of hydroxyproline isomers and hydroxylysine by reversed-phase HPLC and mass spectrometry. *J Chromatogr B.* **2007**, 847:282-8.
66. RW Lovitt, JG Morris, DB Kell. The growth and nutrition of *Clostridium sporogenes* NCIB 8053 in defined media. *J Appl Bacteriol.* **1987**, 62:71-80.
67. A Mitchell, H-Y Chang, L Daugherty, M Fraser, S Hunter, R Lopez, C McAnulla, C McMenamin, G Nuka, S Pesseat, et al. The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res.* **2015**, 43(D1):D213-D21.
68. V-M Leppänen, MC Merckel, DL Ollis, KK Wong, JW Kozarich, A Goldman. Pyruvate formate lyase is structurally homologous to type I ribonucleotide reductase. *Structure.* **1999**, 7(7):733-44.
69. L Lehtiö, A Goldman. The pyruvate formate lyase family: sequences, structures and activation. *PEDS.* **2004**, 17(6):545-52.
70. P Shannon, A Markiel, O Ozier, NS Baliga, JT Wang, D Ramage, N Amin, B Schwikowski, T Ideker. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **2003**, 13(11):2498-504.
71. BE Suzek, Y Wang, H Huang, PB McGarvey, CH Wu, tU Consortium. UniRef clusters: a comprehensive and scalable alternative for improving sequence similarity searches. *Bioinformatics.* **2015**, 31(6):926-32.
72. THMP Consortium. A framework for human microbiome research. *Nature.* **2012**, 486:215-21.

3 Chapter 3. Biochemical and structural studies of *trans*-4-hydroxy-L-proline dehydratase

This chapter represents a preliminary draft of a manuscript in preparation for submission.ⁱ

3.1 Introduction

The discovery and biochemical characterization of *trans*-4-hydroxy-L-proline (Hyp) dehydratase (HypD) were described in detail in Chapter 2. It is a glycyl radical enzyme (GRE) that catalyzes the transformation of Hyp to (*S*)- Δ^1 -pyrroline-5-carboxylate (P5C) and water (Figure 3.1A). In Clostridiales, a colocalized P5C reductase (P5CR) reduces the product P5C to Pro, which undergoes further reduction in Stickland fermentation as demonstrated in Chapter 5. However, most HypD-encoding microbes are not Stickland fermenters, and it remains unclear which additional pathways P5C participates in. Since P5C is a central metabolite in amino acid biosynthetic pathways, possible physiological roles include providing building blocks for protein synthesis or sources of carbon and nitrogen (Figure 3.1A).

Our work on HypD has revealed a new enzymatic activity and expanded known chemistry within the GRE superfamily. GREs rely on a glycine-centered radical cofactor for catalysis as discussed in Chapter 1. Briefly, these enzymes are activated by a cognate activating enzyme (AE) to generate protein-based radicals on conserved Gly and Cys residues to catalyze transformations using radical chemistry (Figure 3.1B). Upon product formation, the thiyl radical is regenerated and can be transferred to the Gly residue for storage.

ⁱ Results and discussions, and methods and materials sections were co-written with Lindsey R.F. Backman and edited by Prof. Catherine L. Drennan and Prof. Emily P. Balskus.

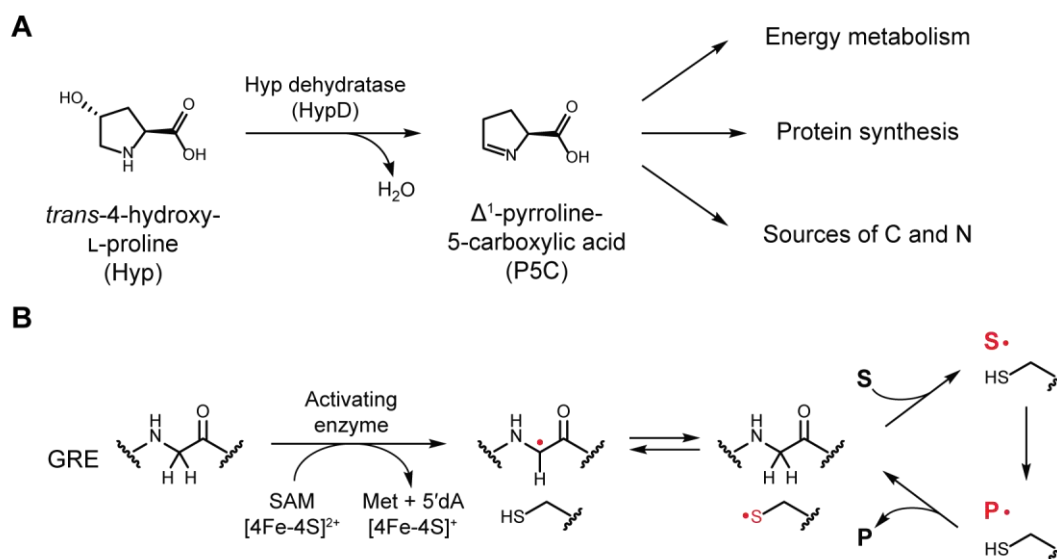


Figure 3.1 Hyp dehydration catalyzed by a GRE revealed a prominent metabolic pathway.

(A) Anaerobic microbial metabolism of *trans*-4-hydroxy-L-proline (Hyp) is catalyzed by Hyp dehydratase (HypD), a glycyl radical enzyme (GRE). The product of this transformation, Δ^1 -pyrroline-5-carboxylate (P5C), is an intermediate in many primary metabolic pathways. Hyp can be used to generate ATP for energy metabolism, converted to other amino acids for protein synthesis, and catabolized to form sources of carbon and nitrogen. (B) General mechanism proposed for GREs. A radical *S*-adenosylmethionine (SAM) activating enzyme (AE) generates a radical species on a conserved glycine residue in the GRE. A thiyl radical on a conserved cysteine is proposed in most GREs to initiate catalysis by abstracting a hydrogen atom from the substrate (S). Upon product (P) formation, the thiyl radical is regenerated to complete the catalytic cycle.

HypD is part of the eliminase class of GREs, which consists of three other reported enzymes: glycerol dehydratase (GD), propanediol dehydratase (PD), and choline trimethylamine-lyase (CutC).¹⁻⁴ The activities and proposed mechanisms for these enzymes were discussed in detail in Chapter 1. Interestingly, GD, PD, and CutC catalyze transformations with reactivities that are also accessible by adenosylcobalamin (B₁₂)-dependent enzymes, a different radical enzyme superfamily. Dehydration of 1,2-propanediol and glycerol were first discovered in the 1960s as an activity catalyzed by a B₁₂-dependent propanediol dehydratase.^{5,6} Choline cleavage by CutC resembles the activity of ethanolamine ammonia-lyase, a B₁₂-dependent enzyme that catalyzes an analogous C–N bond cleavage of ethanolamine using radical chemistry.^{7,8} In contrast, dehydration of Hyp to P5C is unique among GRE eliminases in not having a known corresponding reactivity catalyzed by other radical enzyme families.

As a new enzymatic activity, Hyp dehydration is chemically interesting. Specifically, the amino acid Hyp is structurally unique among GRE eliminase substrates and consists of a restrained pyrrolidine ring. The oxidation of a C–N bond was unprecedented among GREs since all other eliminases catalyze the oxidation of a C–O bond to generate aldehyde products. Collectively, many structural and biochemical studies have been conducted on GD, PD, and CutC, which were highlighted in Chapter 1. Proposed mechanisms for these enzymes share the common feature of utilizing a spin-center shift to drive a direct elimination of the leaving group (mechanisms shown in Chapter 1 – Figure 1.8, Figure 1.9, Figure 1.10). We expect HypD to utilize a similar mechanism to catalyze the elimination of water from Hyp with one key difference. Hyp dehydration likely proceeds through an α -aminoalkyl radical intermediate instead of the α -hydroxyalkyl radical intermediate proposed for other GRE eliminases.

In Chapter 2, we had reported the construction of a HypD homology model, which led to the identification of conserved residues within a motif proposed to be important for dehydrating GREs. Briefly, a conserved Glu residue is proposed to deprotonate the C1 hydroxyl group of PD and GD substrates. A conserved His residue is predicted to protonate the leaving hydroxyl group on C2 to facilitate elimination. In the HypD homology model, it was difficult to dock Hyp in an orientation such that both the amino and hydroxyl groups are adjacent to the Glu and His residues. Biases from using GD as the template to generate this homology model may have mispositioned these residues. Alternatively, a different set of catalytic residues may be responsible for dehydration in HypD. Therefore, we set out to elucidate the structure of HypD to identify active site residues that may be important for catalysis. In this Chapter, we present a Hyp-bound structure of HypD from *Clostridioides difficile* 70-100-2010 along with biochemical analyses of mutant proteins to provide mechanistic insight into a new type of GRE dehydratase.

3.2 Results and discussionsⁱⁱ

3.2.1 Overall architecture of HypD is similar to other GREs

The structure of HypD from *C. difficile* 70-100-2010 to 2.05 Å resolutionⁱⁱⁱ was solved by molecular replacement ($R_{\text{work}} = 0.166$, $R_{\text{free}} = 0.193$), using the GRE homolog CutC (PDB: 5FAU)⁹ as the search model. During model refinement, we observed electron density resembling glycerol in the active site. Glycerol was used as a cryoprotectant during crystallization and therefore was excluded in subsequent crystallization conditions. To obtain a substrate-bound structure, Hyp was

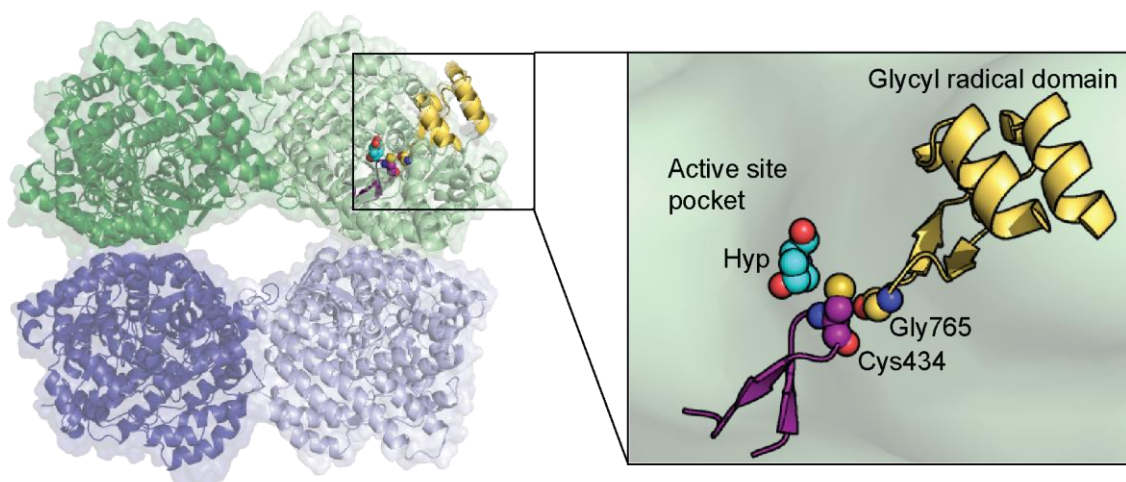
ⁱⁱ All work related to the structural elucidation of HypD was carried out by Lindsey Backman in Prof. Catherine Drennan's lab at MIT.

ⁱⁱⁱ HypD structures described in this Chapter is not in the final refined stage. Therefore, distances between residues and figures reported are preliminary.

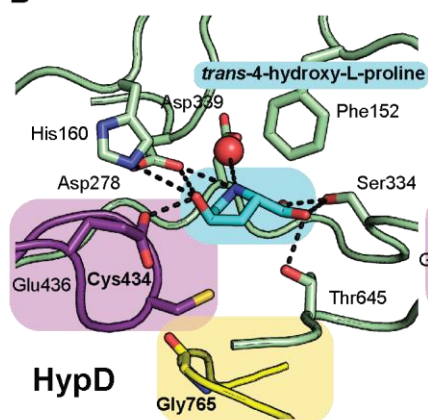
included in the crystallization buffer and the cryoprotectant solution. A second structure was solved to 2.52 Å resolution by molecular replacement ($R_{\text{work}} = 0.186$, $R_{\text{free}} = 0.224$), using the glycerol-bound HypD. During model refinement, Hyp was observed to have bound in the active site and thus this Hyp-bound HypD structure was used for all analyses described in this Chapter.

HypD crystallized as two loosely associated tetramers in the asymmetric unit, with the tetramers resembling those observed in CutC crystal structures (Figure 3.2A).⁹ In agreement with the canonical architecture observed in all characterized GREs, the active site is buried within two five-stranded half β -barrels, anti-parallel to one another and surrounded by α -helices. The active site is buried within the center of the barrel, which is thought to protect radical species from quenching by the solvent. Two loops essential for catalysis are juxtaposed in the active site: the Gly loop and the Cys loop (Figure 3.2A). The Gly loop contains a universally conserved catalytic Gly residue (Gly765 in HypD) and is part of the C-terminal glycyl radical domain found in all GREs. The universally conserved catalytic Cys residue (Cys434 in HypD) is encoded in the Cys loop (Figure 3.2B).

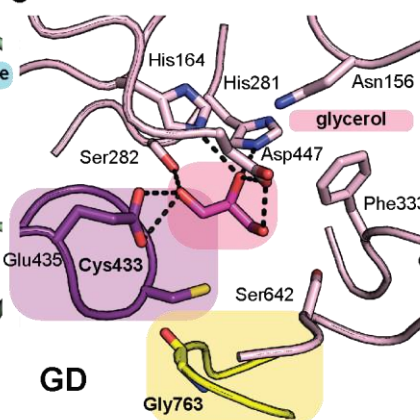
A



B



C



D

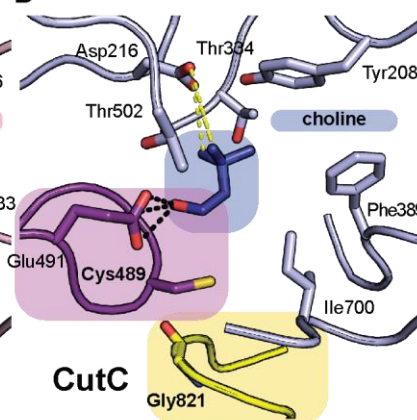


Figure 3.2 HypD shares the canonical GRE fold with conserved Gly and Cys loops.^{iv}

(A) HypD crystallized as two tetramers in the asymmetric unit. Within each tetramer, two homodimers can be discerned based on more extensive interactions between two monomers as highlighted by color. The substrate Hyp is bound in the active site of all monomers within the asymmetric unit near residues Gly765 and Cys434 conserved in all GREs. Conserved Gly and Cys loops in addition to active site residues are displayed for (B) HypD, (C) GD, and (D) CutC. PDB-deposited

Figure 3.2 (Continued)

structures for GD (1R9D)² and CutC (5FAU)⁹ were used to generate this figure. The Cys loop is highlighted in purple and the Gly loop is highlighted in yellow. Residues potentially acting as the general acid and base to facilitate elimination in HypD include Glu436, His160, Asp278, Asp339, and a water molecule. Ser334 and Thr645 are within hydrogen bonding distance to the carboxylate of Hyp and might be important for substrate binding. The equivalent residues in GD and CutC are shown in panels C and D.

Although the electron density of Hyp was observed in the active site, it was not possible to distinguish between the different conformers of the pyrrolidine ring to definitively model Hyp. To approach this problem, we collaborated with Dr. Brian Gold from the Raines lab at MIT, who calculated the optimal conformations of both puckers. Density Function Theory (DFT) calculations were performed for the zwitterionic Hyp using B3LYP/6-31G* theory¹⁰ in the gas phase (Figure 3.3). These calculations revealed that the *C γ -exo* pucker is more energetically favorable than the *C γ -endo* pucker by 2.7 kcal mol⁻¹. Optimizations were also calculated for the anionic Hyp containing a neutral amine. Similarly, the *C γ -exo* pucker was found to be more energetically favorable by 2.9 kcal mol⁻¹. Furthermore, deprotonation of the amino group did not significantly affect the conformation of each pucker. Based on these results, zwitterionic Hyp in the *C γ -exo* pucker state was used to fit into the electron density during model refinement. Bond angles and lengths in the Hyp structure were computationally restrained. A preference for *C γ -exo*

^{iv} HypD structural figures were generated by Lindsey Backman.

pucker in Hyp is further supported by previous work on free Hyp and peptidyl Hyp in the context of collagen structure.^{11,12}

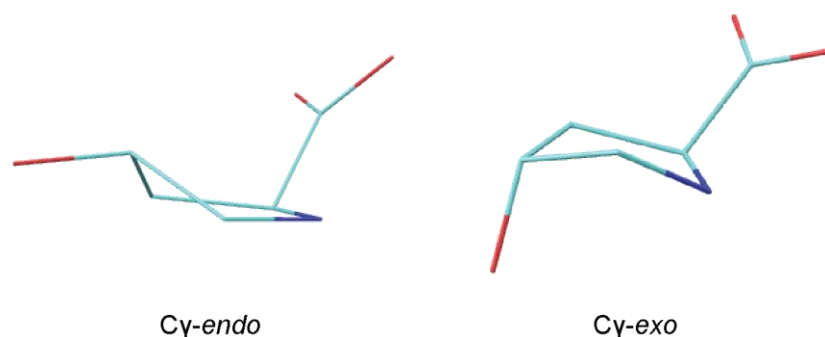


Figure 3.3 Hyp conformers generated from DFT calculations.

Structures of zwitterionic Hyp obtained for *endo*- and *exo*- puckered states. Cartesian coordinates are listed in Table 3.1 and Table 3.2.

Table 3.1 Cartesian coordinates for zwitterionic Hyp in C γ -*exo* pucker calculated from DFT.^v

Coordinates of zwitterionic Hyp structure used to fit into the HypD crystal structure.

N	0.206357	1.105352	-0.52589
C	1.064827	1.041448	0.701676
C	-0.40746	-0.28177	-0.66796
H	0.391256	1.120155	1.557756
H	1.79178	1.854673	0.707412
C	1.693408	-0.34196	0.589719
C	0.522207	-1.2255	0.118162
H	-0.48364	-0.5171	-1.73089
C	-1.86028	-0.11771	-0.04063
H	-0.034	-1.65295	0.957941

^v Coordinates were generated by Dr. Brian Gold in Prof. Ronald Raines lab, MIT

Table 3.1 (Continued)

H	0.886365	-2.06147	-0.48846
O	-2.44922	-1.17391	0.189775
O	-2.17093	1.098951	0.133994
H	0.754198	1.395124	-1.33745
O	2.700451	-0.20189	-0.41598
H	2.12723	-0.67123	1.543191
H	3.044807	-1.08426	-0.62427
H	-0.64106	1.647375	-0.37211

Table 3.2 Cartesian coordinates for zwitterionic Hyp in *C γ -endo* pucker calculated from DFT.^{vi}

N	1.13874	-0.0408	0.25613
C	1.57462	1.27676	-0.2141
C	-0.1842	-0.0141	0.85481
H	1.70693	1.31226	-1.2974
H	2.51431	1.57238	0.26137
C	0.44347	2.22313	0.19105
C	-0.355	1.46598	1.24533
H	-0.218	-0.6733	1.72183
C	-1.2582	-0.5082	-0.1216
H	-1.3919	1.79439	1.27969
H	0.09256	1.63335	2.22759
O	-1.0889	-0.7573	-1.2838
O	-2.4366	-0.6271	0.51955
H	1.78891	-0.3581	0.94647
H	-0.1746	2.44543	-0.6536
O	0.97937	3.42749	0.74533
H	1.39813	3.41357	1.60907
H	1.11531	-0.6725	-0.5414

In the active site, the substrate Hyp is positioned above the Gly and Cys loops similar to other substrate-bound GRE structures (Figure 3.2A). Cys434 is positioned below Hyp with the

^{vi} Coordinates were generated by Dr. Brian Gold in Prof. Ronald Raines lab, MIT

closest proximity to C5 at a distance of 3.8 Å. The *exo* pucker of Hyp orientates the C5 *pro-S* H-atom toward Cys434, prompting us to propose an abstraction of this H-atom by the thiyl radical as the first step of catalysis. Interestingly, the *pro-S* H-atom had been proposed to participate in this step for all reported GRE eliminases based on structural studies.^{2,3,9} Gly765 is positioned 7.5 Å away from the substrate with Cys434 positioned at an intermediate distance of 3.9 Å. This organization is consistent with a radical transfer from Gly765 upon HypD activation to Cys434 to generate a transient thiyl radical for catalysis initiation. Site-directed mutagenesis of Gly765 and Cys434 was performed to confirm their essential roles in catalysis (Figure 3.7). As expected, glycyl radical formation was not detected on HypD-G765A activated by HypD-AE in an EPR spectroscopic assay reported in Chapter 2 (Figure 3.8). The HypD-C434S mutant was activated as observed by glycyl radical formation, which is consistent with having an intact Gly residue for radical installation (Figure 3.8). Both G765A and C434S mutants were inactive in Hyp dehydration in an overnight endpoint assay (Figure 3.9). This result is in agreement with the absolute conservation of Gly and Cys active site residues in GREs and their essential roles in mediating the radical-based chemistry performed by this protein superfamily.

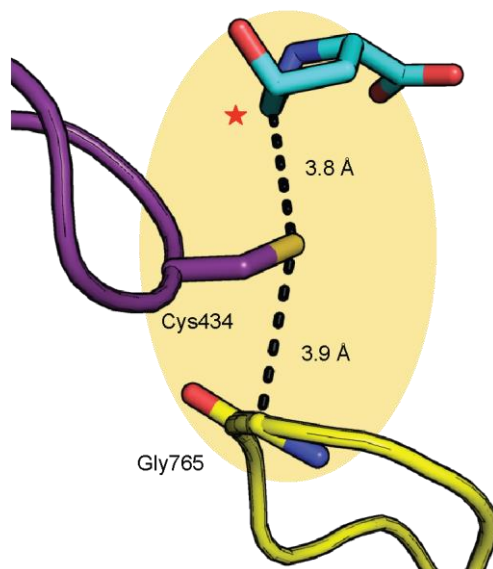


Figure 3.4 Conserved Gly and Cys residues in HypD involved in radical transfer.^{vii}

HypD is activated by its partner AE to install a protein-centered radical on Gly765. Cys434 is proposed to be the thiyl radical essential for H-atom abstraction from C5 of Hyp to initiate catalysis.

The active site of HypD lies in the upper region of the β -barrel, a similar arrangement is observed in GD and CutC structures (Figure 3.2B–D). Compared to these eliminases, the Cys loop in HypD is also positioned similarly (Figure 3.2B–D). The overall backbones of these active sites are alike with subtle variations in key active site residues that accommodate substrates different in size and functional groups. These residues also dictate the respective chemical reactions catalyzed by each GRE. Residues that could act as the general acid and base are positioned within hydrogen

^{vii} HypD structural figures were generated by Lindsey Backman.

bonding distances of functional groups undergoing oxidation or elimination. Furthermore, due to Hyp's bulkier size, residues (Ser334, Thr645) located farther away can interact with the carboxylate through hydrogen bonding.

3.2.2 Key residues in the active site predicted to be important for catalysis

We proceeded to identify residues that may be important for catalysis or substrate-binding by considering residues in close proximity to Hyp in the active site (Figure 3.5A). The importance of these residues was examined through site-directed mutagenesis. We first sought to identify residues that may mediate acid-base chemistry as previously observed in GD, PD, and CutC. From sequence alignments, two residues conserved in HypD (Glu436 and His160) make up a previously identified motif specific to GRE dehydratases.¹³ Unexpectedly, the hydroxyl group of Hyp is oriented towards both Glu436 (2.7 Å) and His160 (3.6 Å) while is too distant for any direct proton transfer (Figure 3.5B). Therefore, the conserved Glu residue in HypD appears to have a distinct role compared to a general base as previously observed for all other eliminases. The amino group of Hyp is within hydrogen bonding distances of 3.1 Å with Asp278 and a distance of 3.3 Å with an ordered water molecule (Figure 3.5B). This ordered water molecule is positioned between Asp339, Asp278, and the amino group of Hyp and is present in all eight monomers of the asymmetric unit, forming an extended hydrogen bonding network (Figure 3.5B). A multiple sequence alignment revealed that Asp339 is uniquely conserved in HypDs but not in other GRE eliminases (Chapter 2, Figure 2.3). Both Asp residues and the ordered water molecule can potentially act as a general acid to deprotonate the Hyp amino group. Overall, we have identified a group of residues likely to be important in catalyzing elimination through acid-base chemistry. Notably, these residues are conserved among putative HypDs (Figure 2.3).

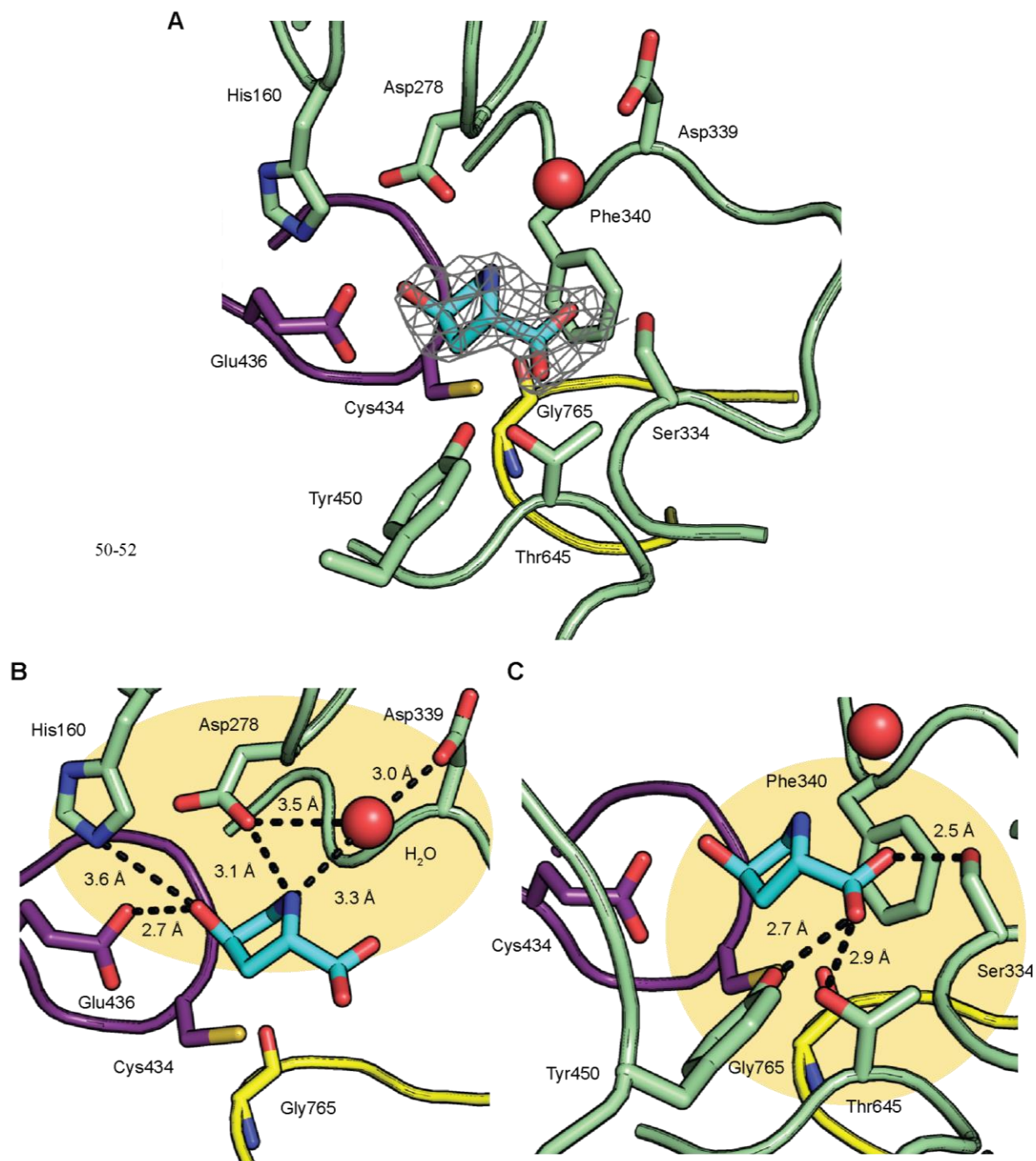


Figure 3.5 Active site residues in HypD selected for site-directed mutagenesis.^{viii}

(A) Residues thought to be important in radical transfer, acid-base chemistry, or substrate binding through hydrogen bonds are labelled in the active site. These

Figure 3.5 (Continued)

residues were selected for mutagenesis. These residues are absolutely conserved in a previously curated list of putative HypD sequences (Chapter 2). (B) Residues predicted to be important in elimination chemistry and their distances to the respective functional groups in Hyp. (C) Residues within hydrogen bonding distances with the carboxylate group of Hyp to stabilize substrate binding.

^{viii} HypD structural figures were generated by Lindsey Backman.

	160	278	334	339	340
	★	★	★	★	★
<i>Treponema maltophilum</i> ATCC 51939 (S3K3Z2)	P G H T V	P W D A F Y	K E S S T Y T	T D F A N	A N N
<i>Cetobacterium somerae</i> ATCC BAA-474 (U7VC55)	P G H T V	P W D A F	K E S S T Y T	T D F A N	A N N
<i>Johnsonella ignava</i> ATCC 51276 (G5GJG7)	P G H T V	P W D A F	K E S S T Y T	T D F A N	A N N
<i>Oscillibacter valericigenes</i> DSM 18026 (G4KZA5)	P G H T V	P W D A Y	K E S S T Y T	T D F A N	A N N
<i>Clostridioides difficile</i> 70-100-2010 (A0A031WDE4)	P G H T V	I W D A F	K E S S T Y T	T D F A N	A N N
<i>Dorea longicatena</i> DSM 13814 (A6BHN6)	P G H T C	P W D A F	K E S A T Y V	T D F A N	A N N
<i>Coprococcus comes</i> ATCC 27758 (C0BAQ4)	P G H T C	P W D A F	K E S A T Y V	T D F A N	A N N
<i>Stomatobaculum longus</i> (H1HV67)	P G H T A	P W D A F	K E S G T Y V	T D F S N	A N N
<i>Collinsella</i> sp. CAG 289 (R7D6G9)	G G H T C	P W D A Y	K E S G T Y S	T D F A N	A N N
<i>Enterococcus raffinosus</i> ATCC 49464 (R2RF16)	P G H T V	N W D A Y	K E S A T Y T	T D F C N	A N N
<i>Desulfotomaculum ruminis</i> ATCC 23193 (F6DS33)	P G H T V	N W D S F	K E S G T Y T	T D F A N	A N N
<i>Thermanaerovibrio velox</i> DSM 12556 (H0UP67)	P G H T A	G W D A M	Q E S G T Y N	T D F T Q	A N N
<i>Desulfomicrobium baculatum</i> DSM 4028 (C7LR40)	P G H T A	G W D A M	K E S G T Y N	T D F T N	A N N
<i>Parabacteroides distasonis</i> DSM 20701 (A6LA42)	P G H T A	G W D A M	K E S G T Y N	T D F T N	A N N
<i>Bacteroides vulgatus</i> DSM 1447 (A6L095)	P G H T S	G W D A M	K E S G T Y N	T D F T N	A N N
<i>Thermovirga lienii</i> ATCC BAA-1197 (G7V7P2)	P G H T T	G W D A M	A E S G T Y N	T D F T N	A N N
<i>Geobacter sulfurreducens</i> ATCC 51573 (Q74BE0)	P G H T V	G W D A M	A E S G T Y T	T D F A N	A N N
<i>Desulfosporosinus youngiae</i> DSM 17734 (H5Y5V0)	P G H T V	T W D A F	E E S G T Y T	T D F A N	A N N
<i>Thermococcus litoralis</i> DSM 5473 (H3ZMH7)	P G H T A	P W D S F	E E S F T Y N	T D F S K	A N N
<i>Odoribacter splanchnicus</i> ATCC 29572 (F9Z3G5)	P G H T A	P W D S F	E E S N T Y T	T D F C L	A N N
<i>Deferribacter desulfuricans</i> DSM 14783 (D3PA11)	P G H T V	T W D A F	K E S A T Y T	T D F A Q	A N N

	434	436	450	645	765
	★	★	★	★	★
<i>Treponema maltophilum</i> ATCC 51939 (S3K3Z2)	S G C V E T G	T G Y F N	L P T T C	V A G Y S	
<i>Cetobacterium somerae</i> ATCC BAA-474 (U7VC55)	S G C V E T G	T G Y F N	L P T T C	V A G Y S	
<i>Johnsonella ignava</i> ATCC 51276 (G5GJG7)	S G C V E T G	T G Y F N	L P T T C	V A G Y S	
<i>Oscillibacter valericigenes</i> DSM 18026 (G4KZA5)	S G C V E T G	T G Y F N	L P T T C	V A G Y S	
<i>Clostridioides difficile</i> 70-100-2010 (A0A031WDE4)	S G C V E T G	T G Y M N	L P T T C	V A G Y S	
<i>Dorea longicatena</i> DSM 13814 (A6BHN6)	S G C V E T G	T G Y L N	L P T T C	V A G Y S	
<i>Coprococcus comes</i> ATCC 27758 (C0BAQ4)	S G C V E T G	T G Y M N	L P T T C	V A G Y S	
<i>Stomatobaculum longus</i> (H1HV67)	S G C T E T G	T G Y V N	L P T T C	V A G Y S	
<i>Collinsella</i> sp. CAG 289 (R7D6G9)	S G C V E T G	T G Y F N	L P T T C	V A G F S	
<i>Enterococcus raffinosus</i> ATCC 49464 (R2RF16)	S G C V E V G	T G Y L N	L P T T C	V A G Y S	
<i>Desulfotomaculum ruminis</i> ATCC 23193 (F6DS33)	S G C V E T G	T G Y L N	L P T T C	V A G Y S	
<i>Thermanaerovibrio velox</i> DSM 12556 (H0UP67)	S G C I E T G	H G Y L N	L S T T C	V A G Y S	
<i>Desulfomicrobium baculatum</i> DSM 4028 (C7LR40)	S G C I E T G	T G Y L N	L S T T C	V A G Y S	
<i>Parabacteroides distasonis</i> DSM 20701 (A6LA42)	S G C I E V G	T G Y L N	L S T T C	V A G Y S	
<i>Bacteroides vulgatus</i> DSM 1447 (A6L095)	S G C I E V G	T G Y L N	L S T T C	V A G Y S	
<i>Thermovirga lienii</i> ATCC BAA-1197 (G7V7P2)	S G C V E T G	T G Y L N	L P T T C	V A G Y S	
<i>Geobacter sulfurreducens</i> ATCC 51573 (Q74BE0)	S G C V E V G	T G Y F N	L P T T C	V A G Y S	
<i>Desulfosporosinus youngiae</i> DSM 17734 (H5Y5V0)	S G C V E T G	T G Y F N	L P T T C	V A G Y S	
<i>Thermococcus litoralis</i> DSM 5473 (H3ZMH7)	S G C V E S G	T G Y F N	L P T T V	V A G Y S	
<i>Odoribacter splanchnicus</i> ATCC 29572 (F9Z3G5)	S G C V E T G	T G Y F N	L P T T C	V A G Y S	
<i>Deferribacter desulfuricans</i> DSM 14783 (D3PA11)	S G C V E A G	T G Y F N	L P T T C	V A G Y S	

Figure 3.6 Active site residues are conserved among all putative HypD sequences.

A Clustal Omega¹⁴ multiple sequence alignment of representative HypD protein sequences. Active site residues used for site-directed mutagenesis are marked with an asterisk (H160, D278, S334, D339, F340, C434, E436, Y450, T645, G765).

Residues are numbered based on HypD sequence from *C. difficile* 70-100-2010.

UniProt accession codes are displayed in parentheses.

3.2.3 A distinct set of residues likely facilitates dehydration in HypD

Site-directed mutagenesis was performed to study the importance of residues proposed to facilitate dehydration (Figure 3.5B). A series of HypD mutants were constructed to disrupt key interactions with the C4 hydroxyl group of Hyp (E436Q, H160Q, D278N), the amino group of Hyp (D278N), and the ordered water molecule (D339N) (Figure 3.7). EPR spectroscopy confirmed that mutation of these residues did not abolish activation by HypD-AE (Figure 3.8). However, these mutants were all activated at substantially reduced levels, suggesting that these amino acids may participate in interactions that affect glycy radical formation and/or stability.

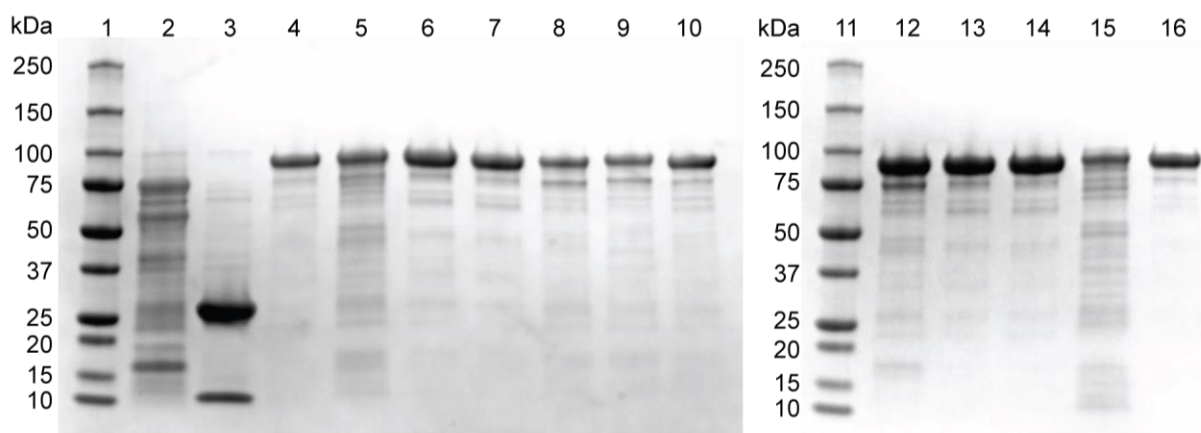


Figure 3.7 SDS-PAGE of purified proteins used in this study.

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) of the proteins used in biochemical assays. Samples were loaded as follows: lane 1 and 11 – Precision Plus Protein All Blue Standards (Bio-Rad), lane 2 – MBP-AE, lane 3 – P5CR, lane 4 – HypD-wild-type, lane 5 – HypD-H160Q, lane 6 – HypD-D278N, lane 7 – HypD-S334A, lane 8 – HypD-D339N, lane 9 – HypD-F340A, lane 10 – HypD-C434S, lane 12 – HypD-E436Q, lane 13 – HypD-Y450F, lane 14 – HypD-T645A, lane 15 – HypD-G765A, lane 16 – HypD-Y450F/T645A. All proteins were cloned from *C. difficile* 70-100-2010. HypD mutants exhibited similar stability and purification yields as wild-type.

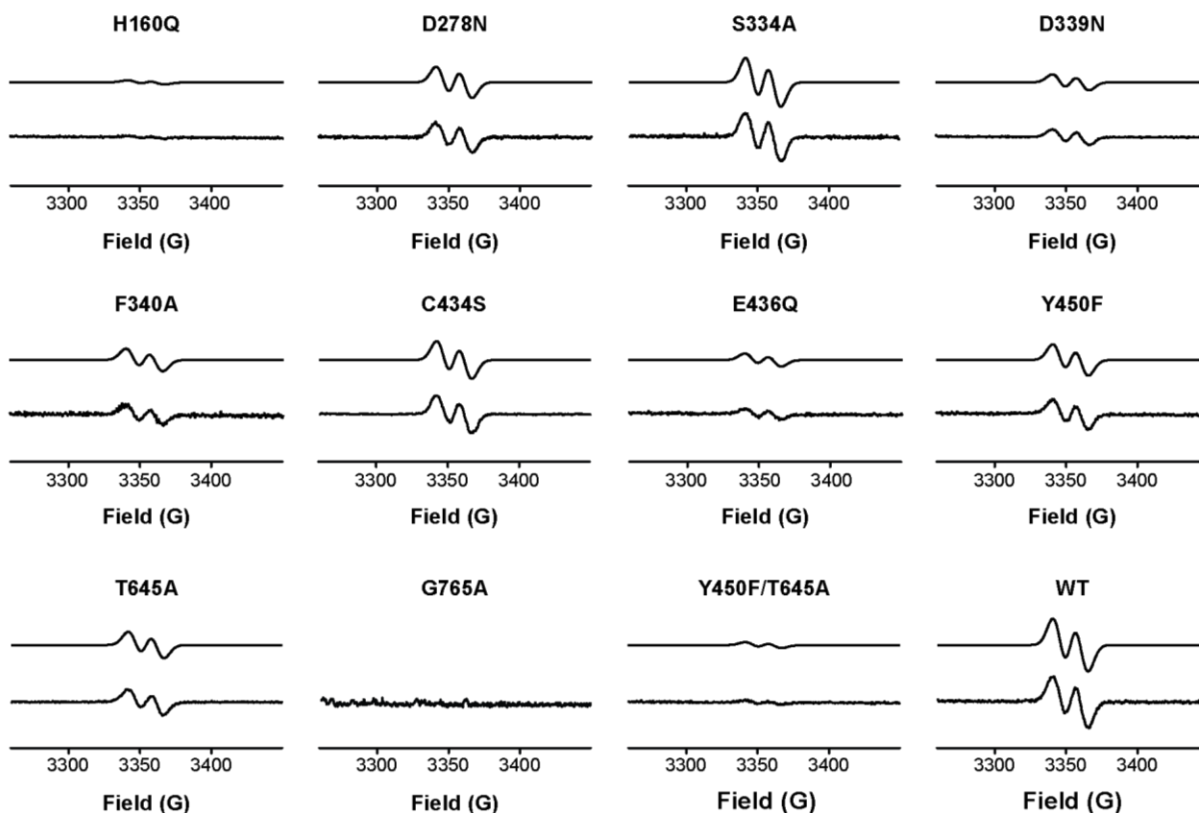


Figure 3.8 Activation of HypD wild-type and mutants detected by EPR spectroscopy.

Glycyl radical formation was quantified by EPR spectroscopy for all HypD mutants and wild-type. Representative EPR spectra are shown here for each mutant. Both simulated (top trace) and experimental (bottom trace) spectra are displayed for each HypD protein. HypD mutants are listed by residue order.

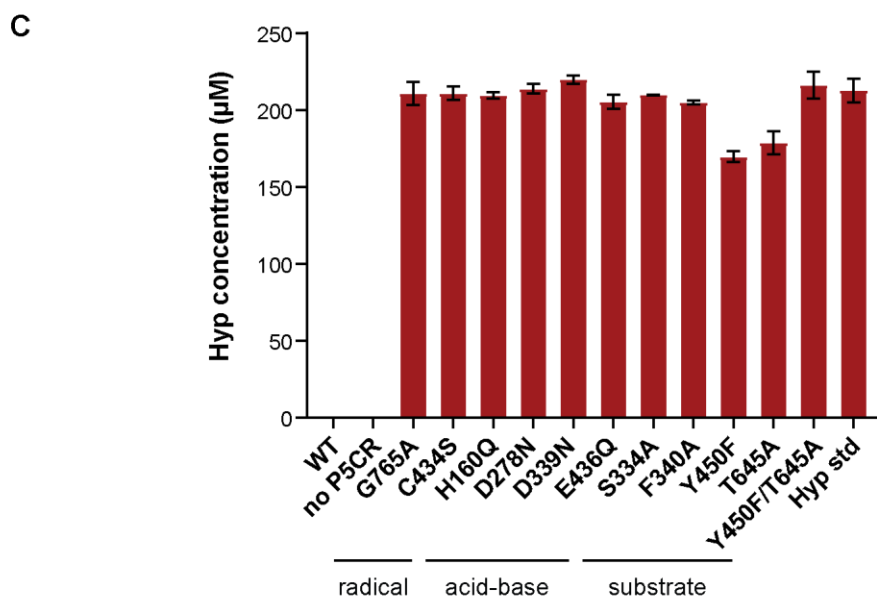
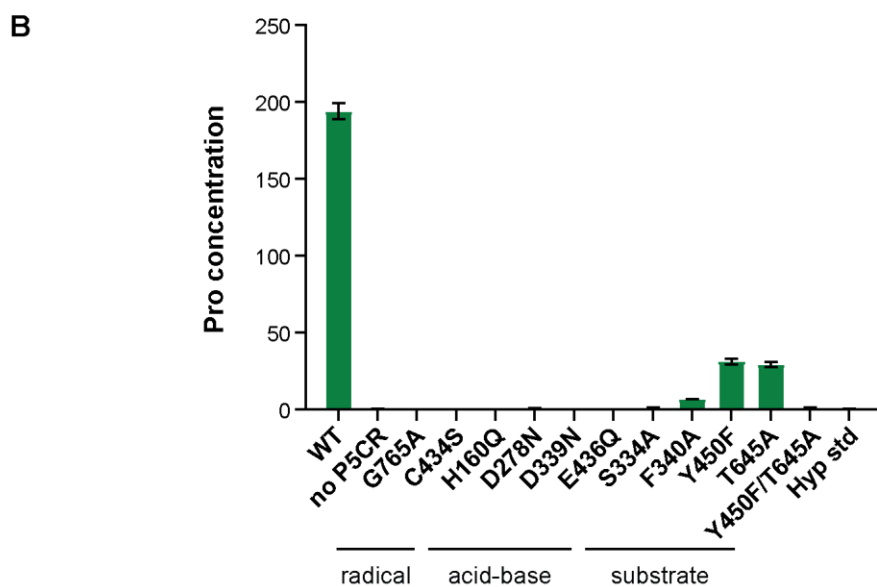
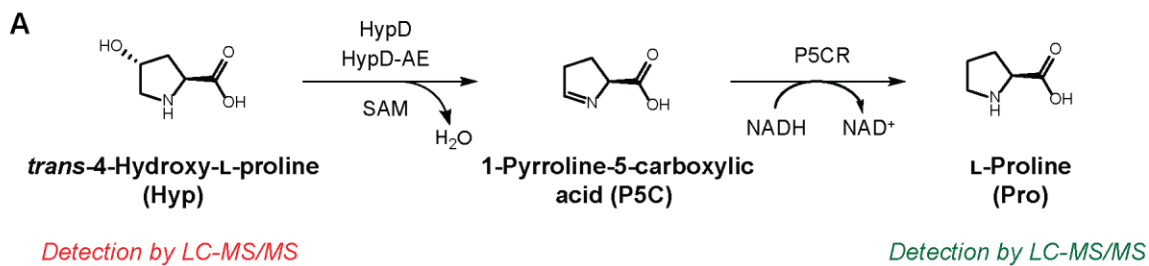
E436Q, H160Q, D278N, and D339N mutants exhibited no detectable activity in an endpoint assay, which strongly indicates that these residues are essential for catalysis (Figure 3.9). As described previously, Glu436 and His160 have unexpected orientations relative to Hyp in the active site compared to GD and PD. Here, both residues are located adjacent to the departing C4 hydroxyl group (Figure 3.5B). Due to the close proximity to one another (2.5 Å) and the

mutagenesis results, we propose that these residues work together as a catalytic dyad. His160 can mediate a proton transfer to Glu436, which can protonate the Hyp hydroxyl group to facilitate elimination and thus act as a general acid. Neither of these residues is close enough to the amino group to act as the general base for its deprotonation. Therefore, rather than acting as the general base, as in GD and PD, Glu436 is predicted to be a general acid in HypD. Interestingly, a distinct Glu or Asp residue is conserved in GD and PD, respectively, are thought to help facilitate departure of the C2 hydroxyl group through protonation.^{3,15} It is possible that the function of Glu436 in promoting dehydration in HypD parallels that of the extra Glu/Asp residue in diol dehydratases.

The loss of detectable activity in HypD-D278N and HypD-D339N suggest the hydrogen bonding network of these two residues with the ordered water molecule is important for catalysis. The close proximity of both Asp278 and the water molecule to Hyp amino group indicates that either one could be a general base responsible for substrate deprotonation. Asp278 could also be important for substrate binding and stabilization by forming an electrostatic interaction with the protonated HypD-amino group, which has a pK_a of ~ 9.7 .¹⁶ The lack of activity in HypD-D339N is intriguing, as this residue was not initially predicted to be involved in catalysis due to its position >7 Å from Hyp. The importance of this residue in HypD function is evident by the observed loss of activity upon mutation and also by its conservation in all putative HypD sequences (Figure 2.3). One potential role of Asp339 could be to deprotonate the water molecule, thereby activating it as a general base. Overall, the residues that could potentially serve the role of a general base (Asp278, Asp339) are distinct from the Glu residue used by other GRE eliminases.

Figure 3.9 Most HypD mutants did not have detectable activity.

(A) An *in vitro* coupled enzyme endpoint assay was used to measure activity of HypD mutants. P5C generated from HypD activity was reduced to Pro by P5CR in assay mixtures. Pro and Hyp were quantified using LC-MS/MS. (B) Pro concentrations in assay mixtures after incubation for 21 hours. (C) Hyp concentrations in assay mixtures after incubation for 21 hours. Data points represent mean $\pm$ SD with n = 3 replicates.



Intriguingly, the residues predicted to mediate acid-base chemistry (Glu436, His160, Asp278, H₂O, Asp339) form a hydrogen bonding network around the hydroxyl and amino groups of Hyp (Figure 3.5B). Interactions between these residues suggest a plausible route for transfer of the proton originating from the amino group of Hyp through these residues to the general acid (Glu436 or His160). A role in proton transfer may contribute to the lack of activity observed when any one of these residues was altered. The presence of a similar network was also reported in CutC.⁹ Overall, the loss of detectable activity in these mutants strongly supports catalytic roles for these residues His160, Asp278, Asp339, and Glu436.

3.2.4 Additional active site residues likely facilitate Hyp binding

Residues predicted to be important for substrate binding were also identified for mutagenesis. Three residues (Ser340, Tyr450, and Thr645) are within hydrogen bonding distances (2.5–2.9 Å) to the carboxylate of Hyp, suggesting a role in substrate stabilization (Figure 3.5C). These residues are absolutely conserved in HypDs, suggesting they are important features of HypD (Figure 2.3). Therefore, we constructed additional mutants S334A, Y450F, and T645A. Glycyl radical formation was detected in these three mutants, indicating successful activation by HypD-AE (Figure 3.8). Strikingly, different outcomes were observed in these three mutants. No activity was observed in HypD-S334A, indicating this amino acid likely plays a critical role in mediating substrate binding by providing a H-bond donor to the carboxylate moiety (Figure 3.9). In contrast, activity was detected in HypD-Y450F and HypD-T645A but at much lower levels compared to the wild-type protein (Figure 3.9).

Kinetic assays were performed for HypD-Y450F and HypD-T645A to examine effects of these mutations on catalysis. The catalytic efficiency for both mutants are 2–3 orders of magnitude lower than that of wildtype HypD due to increased K_m and decreased k_{cat} values (Figure 3.10, Table

3.3). Given that Tyr450 and Thr645 participate in hydrogen bonding interactions with the carboxylate group, increased K_m values support a role for these residues in substrate binding. Decreases in k_{cat} could be explained by changes in Hyp orientation that reduced the catalytic rate due to suboptimal distances between the amino and hydroxyl groups and the catalytic residues. Finally, conformational changes caused by these mutations may lead to protein destabilization and thus could also contribute to the reduction in k_{cat} . This could be tested using thermal melt assays. Since both Tyr450 and Thr645 interact with the same carboxylate oxygen atom, we generated a double mutant, HypD-Y450F/T645A, to test if these two residues have redundant functions. HypD-Y450F/T645A was activated by HypD-AE as detected using EPR spectroscopy, but showed no detectable activity in the end point assay (Figure 3.8, Figure 3.9). Overall, complete disruption of hydrogen bonding between active site residues and any one of the oxygen atoms of Hyp abolished activity.

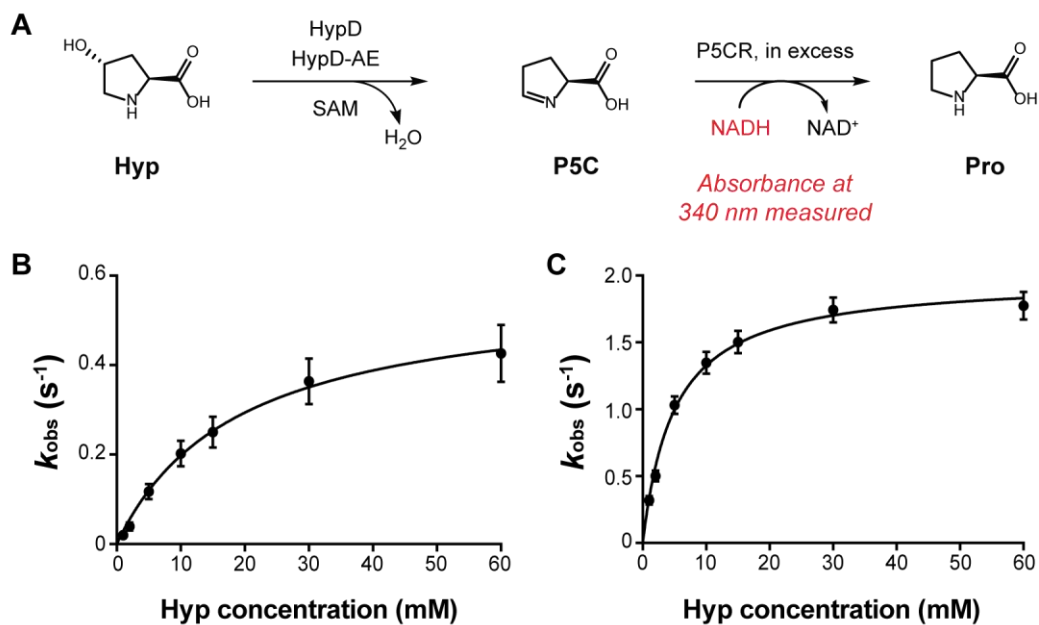


Figure 3.10 Kinetic analysis of HypD-Y450F and HypD-T645A.

(A) HypD activity was coupled to P5CR and absorbance at 340nm was measured to calculate initial rates for NADH consumption. (B) Michaelis–Menten kinetic curve for HypD-Y450F. (C) Michaelis–Menten kinetic curve for HypD-T645A. Data points represent the mean of 3 replicates and error bars represent SD.

Table 3.3 Glycyl radical quantification, activity, and kinetic parameters of HypD mutants.

HypD	Radical per monomer (%)	Activity detected by quantification of proline	K_m (mM)	Normalized k_{cat} (s ⁻¹)	Catalytic efficiency (M ⁻¹ s ⁻¹)
Wildtype ¹³	51 ± 1	Yes	1.2 ± 0.1	45 ± 1	3.8 ± 0.3 × 10 ⁴
G765A	0	No	ND	ND	
C434S	34 ± 8	No	ND	ND	
E436Q	12.4 ± 0.5	No	ND	ND	
H160Q	4.4 ± 0.8	No	ND	ND	
D278N	16 ± 4	No	ND	ND	
D339N	18 ± 8	No	ND	ND	
S334A	50 ± 19	No	ND	ND	
Y450F	29 ± 4	Yes	19 ± 3	0.57 ± 0.04	30 ± 6
T645A	19 ± 1	Yes	4.9 ± 0.4	1.98 ± 0.04	400 ± 30
Y450F/T645A	3.4 ± 0.8	No	ND	ND	
F340A	23 ± 5	No	ND	ND	

In addition to residues within hydrogen bonding distances with the Hyp carboxylate, an aromatic residue Phe340 likely stabilizes the substrate through an aromatic-proline interaction. This conserved residue is poised near Hyp such the shortest distance is observed between C5 and the Phe ring (3.2 Å) (Figure 3.11). We constructed a HypD-F340A mutant to see if removal of this aromatic ring would impact activity. While HypD-F340A was activated by HypD-AE, it displayed very little activity toward Hyp dehydration (Figure 3.8, Figure 3.9). Aromatic-proline interactions are similar to cation- π interactions and result from the polarized C–H bonds of the pyrrolidine ring interacting with the negatively charged π face of the aromatic ring.¹⁷ Interestingly, a conserved aromatic residue can be found adjacent to the substrate in CutC (Phe395), PD (Phe344), and GD (Tyr339). In CutC, Phe395 was hypothesized to contribute to substrate binding through a cation- π interaction with C2 of choline, which has a partial positive charge due to the adjacent quaternary ammonium group.⁹ Unfortunately, the role of this residue could not be examined due to a lack of

glycyl radical formation in the mutant.⁹ The drastically reduced activity in HypD-F340A supports a role for Phe340 in substrate stabilization, which is likely a conserved feature among GRE eliminases.

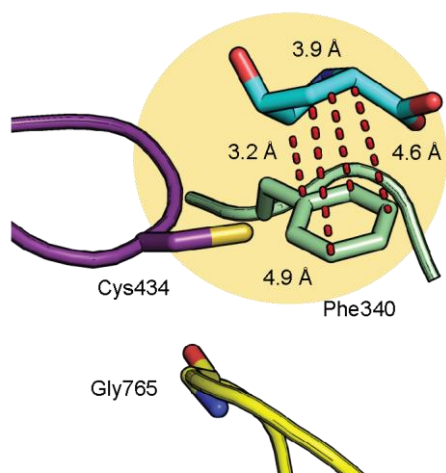


Figure 3.11 A Phe340 potentially involved in aromatic-proline interaction.^{ix}

An aromatic-proline interaction thought to be important for substrate stabilization.

Hyp is positioned with C5 in closest proximity to the aromatic ring of Phe.

3.3 Conclusions

Hyp dehydration is not only unique among GREs, but also represents a novel activity broadly among enzymes. The oxidation of a C–N bond to drive elimination through a spin-center shift appears to be unprecedented in enzyme chemistry and is proposed to involve an α -aminoalkyl radical intermediate (Figure 3.15II). α -Aminoalkyl radicals are most often observed in amino acid-based intermediates generated during GRE activation, amino acid epimerization, and thioether

^{ix} HypD structural figures were generated by Lindsey Backman.

bond formation in biosynthesis of ribosomally synthesized and post-translationally modified peptides (RiPPs).¹⁸ The C α -centered radical part of an amino acid is stabilized by delocalization of the unpaired electron onto the amino and the carboxyl groups.¹⁹ The frequent occurrence of these amino acid radicals in enzyme chemistry can be accounted for by their stability. For example, glutamate mutase is a B₁₂-dependent enzyme predicted to generate a glycy radical intermediate during the isomerization of L-glutamate to (2*S*, 3*S*)-3-methylaspartate (Figure 3.12).²⁰

While C α -centered amino acid radicals are commonly proposed in enzyme mechanisms, only a few enzymes have been proposed to generate intermediates containing an α -aminoalkyl radical without an adjacent carboxyl group. HppE has been proposed to generate α -aminoalkyl radical intermediate prior to imine formation in a substrate analog (Figure 3.12).²¹ However, this intermediate is not relevant for its endogenous substrate and physiological function.²¹ Historically, oxidation of amines catalyzed by flavoprotein monoamine oxidases was proposed to proceed by radical mechanisms and thereby would involve generating α -aminoalkyl radicals on the substrate and/or the flavin cofactor (Figure 3.12).²² However, the nucleophilic-based mechanism has gained greater support and is currently accepted by the field.²² While α -aminoalkyl radicals have been rarely reported in enzymatic chemistry, these radicals are often generated as reactive intermediates to drive substitution reactions in organic chemistry.^{23,24} Given their wide utility to synthetic chemists, it is tempting to speculate that these radicals are more common among enzyme reactions that have yet to be discovered.

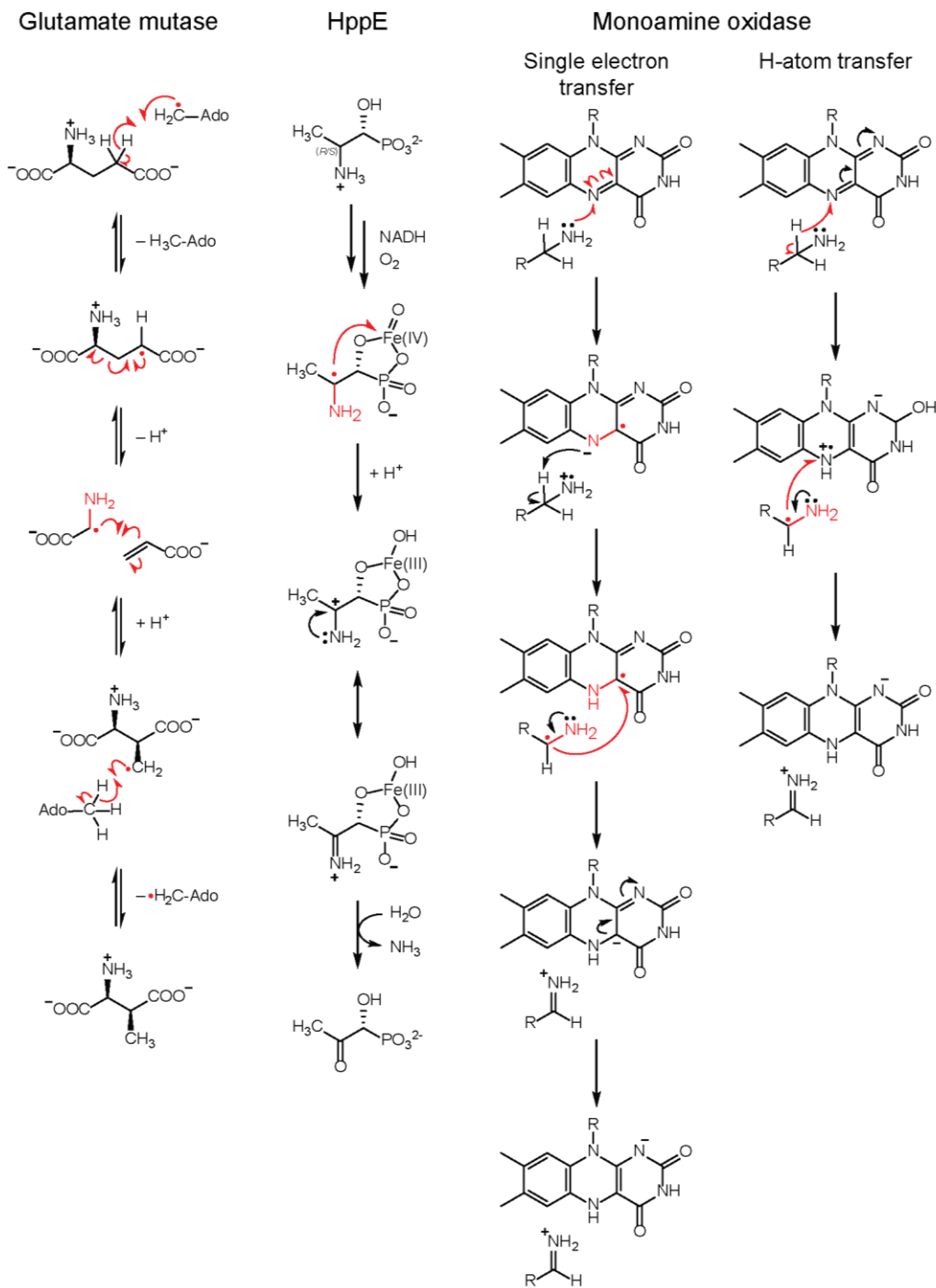


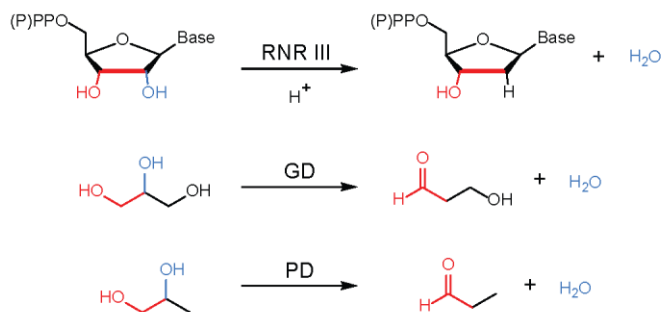
Figure 3.12 Transformations catalyzed by radical enzymes proposed to involve an α -aminoalkyl radical intermediates.

Figure 3.12 (Continued)

Proposed enzymatic mechanisms that involve intermediates containing an α -aminoalkyl radical moiety, which is highlighted in red.

Another notable feature of HypD is its use of a restrained substrate that lacks a rotatable C_α – C_β bond, in contrast to enzymes that accept 1,2-diols and choline. Among GREs, the first dehydration step catalyzed by class III RNRs (see Chapter 1) resembles Hyp dehydration in that both enzymes accept substrates containing 5-membered heterocyclic rings (Figure 3.13). In the RNR-catalyzed reaction, elimination occurs through a ketyl radical adjacent to the leaving hydroxyl group. Unlike in Hyp dehydration, this step does not involve an oxidation of a bond within the ring scaffold itself. Therefore, the functional groups being modified in the RNR reaction more closely resemble those invoked in diol dehydratases (Figure 3.13). Recently, the radical SAM enzyme viperin was shown to catalyze dehydration of the ribose-containing substrate cytidine triphosphate (CTP) (Figure 3.13).²⁵ This reaction is reminiscent of Hyp dehydration because it involves the elimination of a hydroxyl group on the carbon position β to the C–X bond within the 5-membered ring.²⁵ Instead of an nitrogen atom in the pyrrolidine ring of Hyp, viperin's substrate contains an oxygen atom in its ribose ring. Based on labelling experiments, a mechanism for viperin activity was proposed to involve a direct elimination coupled with oxidation of the furanose ring.²⁵

Dehydration of 1,2-diol-containing substrates



Dehydration facilitated by oxidation of a 5-membered ring

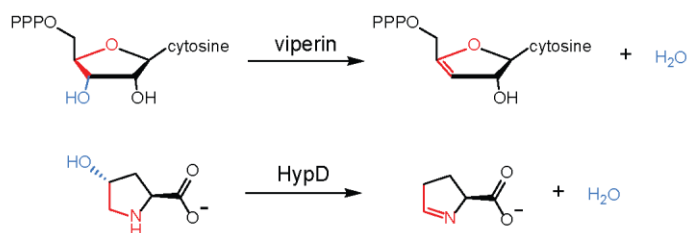


Figure 3.13 Similar reactions catalyzed by other radical enzymes.

Transformations catalyzed by other GREs and viperin similar to Hyp dehydration are shown. The leaving hydroxyl group is highlighted in blue and the remaining portion of the substrate involved in catalysis are highlighted in red. The colored C_α-C_β bond in GD and PD substrates are rotatable while the equivalent bond in RNR, viperin, and HypD substrates are constrained as a part of 5-membered rings. The first step of dehydration catalyzed by class III RNR involves the 1,2-diol moiety of the ribose ring and thus is analogous of reactions catalyzed by diol dehydratases. Both viperin and HypD oxidize a C-O or C-N bond part of a 5-membered ring, respectively, as part of their dehydration chemistry.

Extensive experimental and computational investigations have provided strong support for direct elimination as the mechanism used by GRE eliminases. These mechanistic studies on GD,

PD, and CutC were described in detail in Chapter 1. On the basis of these previous studies, we propose a mechanism for HypD that involves a direct elimination of the hydroxyl group on Hyp to generate P5C (Figure 3.15).

First, we predict that Hyp binds as a zwitterion in the active site based on its protonation state under neutral pH and the presence of Asp residues positioned near the amino group (pK_a of ~ 9.7). Deprotonation of the amino group by either Asp278 or the ordered water molecule can then occur to generate anionic Hyp. The nitrogen lone pair of the Hyp amino group would stabilize the adjacent radical upon H-atom abstraction from C5 through conjugative delocalization (Figure 3.15II, III).¹⁹ The initial H-atom abstraction in GRE-catalyzed transformations has been predicted to have a high free energy of activation.^{15,26} In other GRE eliminases, the deprotonation of the adjacent hydroxyl group by the conserved Glu residue is proposed to lower the energy barrier to this step.²⁷ Therefore, we hypothesize that deprotonation of Hyp amino group to its neutral state is essential prior to formation of a substrate-based radical. This is supported by our mutagenesis data where both Asp residues were essential for activity in the endpoint assay.

Once Hyp is in its anionic state, a thiyl radical on the conserved Cys can then initiate catalysis. Based on Cys434's position relative to Hyp, we postulate this thiyl radical to abstract the *pro-S* H-atom on C5 to generate an α -aminoalkyl radical intermediate (Figure 3.15I, II). Notably, the radical stabilization energy (RSE) associated with having an amino substituent adjacent to a carbon-centered radical is greater than stabilization from an alcohol or an alkoxide substituent.^{28,29} A difference of ~ 2.6 kcal mol⁻¹ has been reported between α -aminoalkyl radicals and α -hydroxyalkyl radicals.^{28,29} The α -aminoalkyl radical intermediate of Hyp is in resonance with an aminyl radical, which is analogous to the ketyl radical species postulated for other GRE eliminases (Figure 3.15II, III).

A direct elimination of the hydroxyl group on C4 is expected to proceed from this substrate-based radical. This can be facilitated through protonation of the departing hydroxyl group by the adjacent Glu436-His160 catalytic dyad (Figure 3.15III, IV). Structural and mutagenesis data provide strong support for these residues playing an important role in catalysis. The CXE motif is conserved among GRE eliminases, but the proposed role of Glu as a general acid in HypD is unprecedented among these enzymes. The same conserved Glu is thought to function as a general base to mediate the formation of a ketyl radical species in GD, PD, and CutC. However, in 4-hydroxyphenylacetate decarboxylase (HPAD), the Glu in the CXE motif is postulated to act as an acid to protonate the thiolate on Cys formed upon a single-electron oxidation of the substrate. Therefore, this conserved Glu is likely to have different catalytic roles depending on the active site environment and selective tuning of its pK_a value.

Hyp is distinct from other substrates in that it contains a restrained 5-membered ring and thus has decreased conformational flexibility. Although we were not able to elucidate the conformation of Hyp from the crystal structure, we determined the energetically favorable Hyp pucker conformation through DFT calculations. In the *exo* pucker conformation used in our structural analyses, the dihedral angle between the amino group and the hydroxyl group is 77°. Upon H-atom abstraction from C5 of the substrate, an antiperiplanar geometry would be observed between the partially occupied p-orbital on C5 and the C4 hydroxyl group. This geometry establishes an overlap between the p-orbital on C5 and the C–O σ^* orbital on C4 to promote dehydration. This also provides additional support for the *exo* pucker conformation, since orbital overlap is reduced in the *endo* pucker where the dihedral angle is 165° for the amino and hydroxyl groups. The resulting angle between the *pro-S* H-atom on C5 and the departing hydroxyl group would make dehydration more challenging for Hyp in its *endo* pucker. Analogously, similar

dihedral angles are found in substrates bound in other GRE eliminase crystal structures (Figure 3.14). This strongly support that these enzymes preferentially binds substrates in a conformation to facilitate elimination by optimizing orbital overlap.

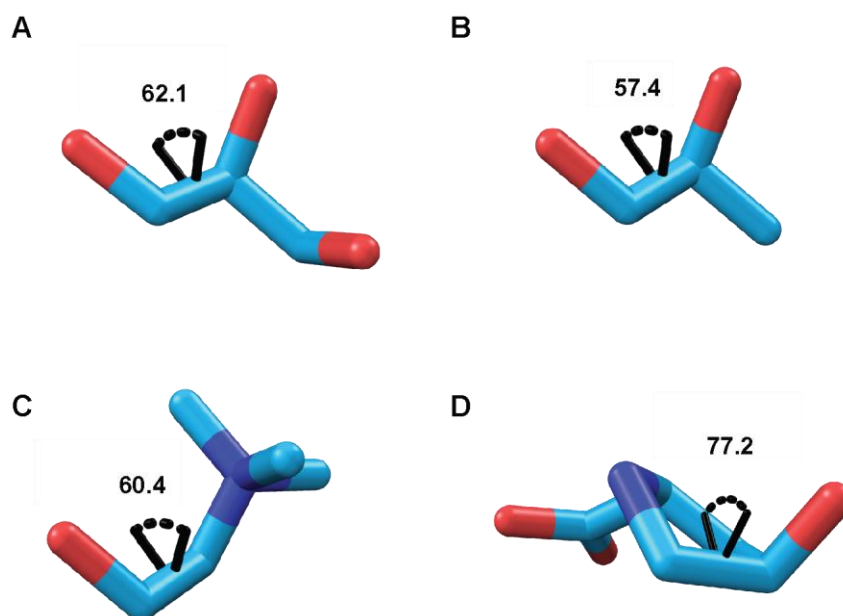


Figure 3.14 Dihedral angles of substrates bound in GRE eliminases.

The dihedral angle between the departing group and the C1 hydroxyl group is displayed for (A) glycerol (GD, PDB: 1R9D), (B) (*S*)-1,2-propanediol, (PD, PDB: 5I2G), (C) choline (CutC, PDB: 5FAU), and (D) Hyp in *exo* pucker (HypD).

Upon elimination of the C4 hydroxyl group, the final step involves re-abstraction of the H-atom from Cys434 by the product-based radical to regenerate the thiyl radical on HypD (Figure 3.15V, VI). As previously reported for other GRE eliminases, the carbon-centered radical in the product is likely reactive enough to generate the relatively stable thiyl radical (S–H bond dissociation energy (BDE) of $\sim 87 \text{ kcal mol}^{-1}$)³⁰. The product P5C has a pK_a of ~ 6.5 and thus could

undergo deprotonation prior to release from the enzyme active site.³¹ Resetting of protonation states of active sites residues is likely mediated by water molecules observed in the second sphere in the crystal structure.

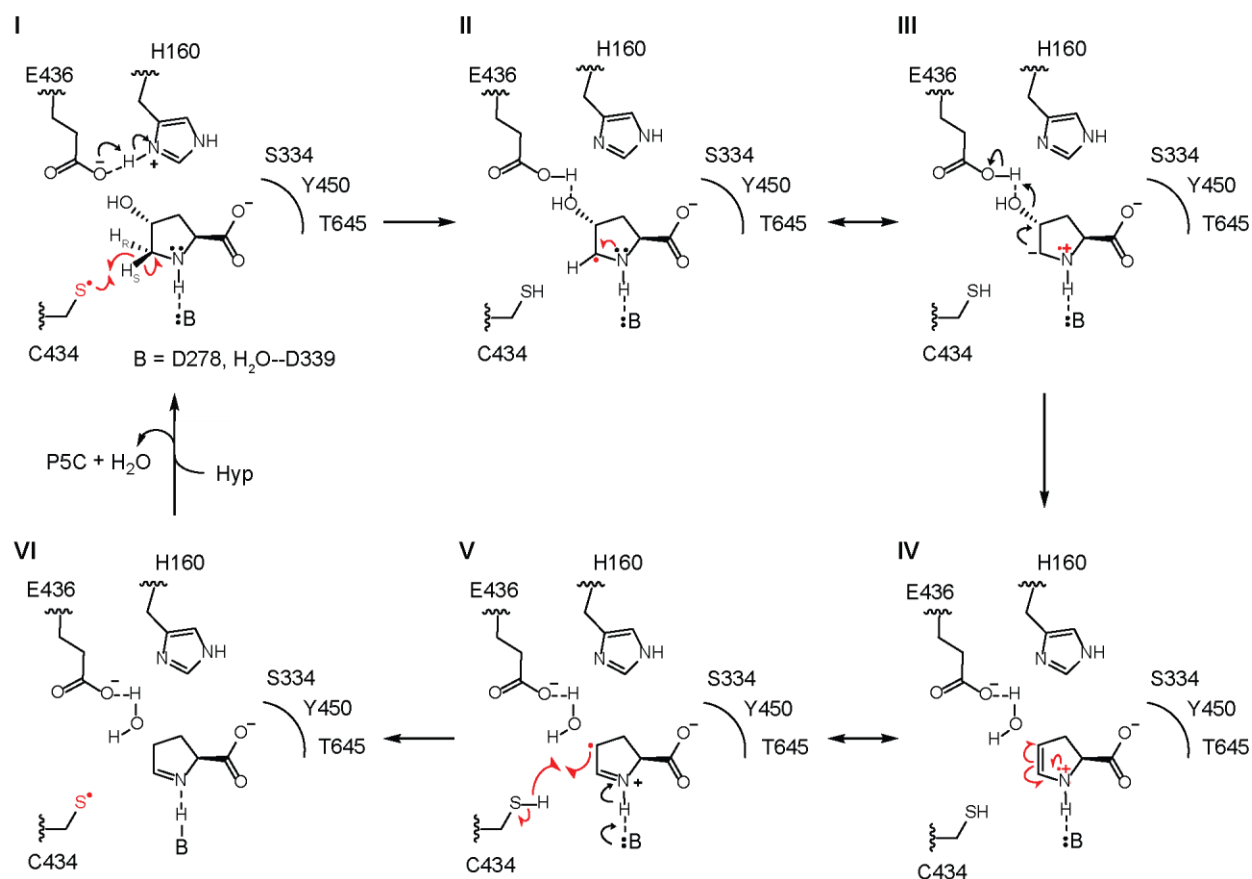


Figure 3.15 Proposed mechanism for Hyp dehydration by HypD.

The general base is depicted as “B”. In the active site structure, Asp278 and an ordered water molecule are within direct hydrogen bonding distances to the amino group. Asp399 is 3.0 Å from the water molecule and thus could indirectly participate in the deprotonation step.

In conclusion, we have identified a key set of residues important for catalysis and substrate binding in HypD through structural and biochemical experiments. From the data presented here and mechanistic studies on other GRE eliminases, we propose that Hyp dehydration involves a direct elimination of the C4 hydroxyl group. The proposed mechanism is interesting and unique from multiple aspects, which were highlighted throughout this section. We expect investigations using computational tools and labelling experiments will provide additional support for the mechanism proposed here. For example, using C5-deuterated Hyp in an activity assay may result in deuterium incorporation at C4 of the product, which would provide evidence for an H-atom abstraction at C5 and an H-atom re-abstraction at C4. Furthermore, detection of a kinetic isotope effect using this C5-labelled substrate would provide support for the H-atom abstraction step as rate-limiting.

3.4 Materials and methods

All chemicals, solvents, and reagents were purchased from vendors as previously detailed in Chapter 2's materials and methods section with some modifications. DNA sequencing results and multiple sequence alignments were analyzed with Geneious 9.0.4.³² PCR was performed with a C1000 Gradient Cycler (Bio-Rad). All plasmid constructs were verified by DNA sequencing through Eton Biosciences. Rad). All absorbance measurements in 96-well plates were carried out using a PowerWave HT Microplate Spectrophotometer (Biotek). Samples were made anaerobic as previously described in Chapter 2.

3.4.1 Vector constructs for the overexpression of HypD, HypD-AE, and P5CR

HypD (UniProt ID: A0A031WDE4), HypD-AE (UniProt ID: A0A069AMK2), and P5CR (UniParc ID: UPI000235AE56) were amplified from *C. difficile* 70-100-2010 genomic DNA as

previously reported in Chapter 2. Vectors were previously constructed to afford pET28a-CdHypD, pET28a-CdP5CR, and pSV272-PfMBP-CdHypDAE.

3.4.2 Site-directed mutagenesis and construction of overexpression vectors for HypD mutants

Single residue mutations were introduced in pET28a-CdHypD through site-directed mutagenesis using the corresponding primers listed in Table 3.4. The following residue changes were made: G765A, C434S, E436Q, H160Q, D278N, F340A, Y450F, S334A, T645A, D339N. A double mutant for CdHypD was constructed by introducing Y450F mutation into the vector pET28a-CdHypD T645A. PCR was carried out using Phusion-HF or Q5 polymerase according to the manufacturer's protocol in a total reaction volume of 25 μ L. An extension time of 200 s at 72 °C was used for the PCR protocol and the annealing temperature for each primer pair is listed in Table 3.4. Template plasmid was removed by digesting with DpnI (NEB) in all PCR mixtures at 37 °C for 1 h. 2 μ L of each PCR digestion was used to transform chemically competent *E. coli* TOP10 cells. Sequenced vectors were then transformed into chemically competent *E. coli* BL21-CodonPlus(DE3)-RIL $\Delta proC::aac(3)IV(Am^R)$ cells for protein overexpression.

Table 3.4 Primers used in site-directed mutagenesis of HypD

Nucleotides mutated are indicated in small letters.

Primer	Sequence (5' to 3')	Annealing temperature, °C
pET28a-CdHypD-G765A-fwd	GACTTAATAGTTAGAGTTGCAGcATATAGTGACCA TTTC	66
pET28a-CdHypD-G765A-rev	CTACTTAAATTATTGAAATGGTCACTATATgCTGCA ACTCTAAC	66
pET28a-CdHypD-C434S-fwd	AACCAGTGGTTcTGTTGAAACTGGATG	58
pET28a-CdHypD-C434S-rev	CAGTTTCAACAgAACCACTGGTTCCACC	58
pET28a-CdHypD-E436Q-fwd	CAGTGGTTGTGTTcAAACTGGATGTTTTGG	60
pET28a-CdHypD-E436Q-rev	ACATCCAGTTTgAACACAACCACTGGTTC	60
pET28a-CdHypD-H160Q-fwd	AGCCCCAGGACAgACAGTTTGTGGAGATAC	60
pET28a-CdHypD-H160Q-rev	ACAAACTGTcTGTCCTGGGGCTCTTTGTTC	60
pET28a-CdHypD-D278N-fwd	GAACTTAATATATGGaATGCTTTTACTCCAGGAAGA CTTGACC	66
pET28a-CdHypD-D278N-rev	CCTGGAGTAAAAGCATtCCATATATTAAGTTCAGTA GTAACCCC	66
pET28a-CdHypD-F340A-fwd	GAAAGTAGCACATATACAGATgcTGCAAATATAAA C	54
pET28a-CdHypD-F340A-rev	GATTTATTCCACCAGTGTTTATATTTGCAgcATCTGT ATATG	54
pET28a-CdHypD-Y450F-fwd	GTTTTGGTAAAGAAGCATATGTTCTAACTGGATtTA TGAACATTCC	66

Table 3.4 (Continued)

pET28a- CdHypD- Y450F-rev	GTATTTTGGGAATGTTTCATAaATCCAGTTAGAACAT ATGCTTCTTTACC	66
pET28a- CdHypD- S334A-fwd	GTTGGTATAACATTAAAAGAAgcTAGCACATATAC AGATTTTGC	60
pET28a- CdHypD- S334A-rev	CTGTATATGTGCTAgcTTCTTTTAATGTTATACCAAC TTTTGG	60
pET28a- CdHypD- T645A-fwd	ATGTTACCAgCAACTTGTCATATATACTTTGGAGAA ATTATGGG	66
pET28a- CdHypD- T645A-rev	TATGACAAGTTGcTGGTAACATATCTACTCTGTATT CTCCACC	66
pET28a- CdHypD- D339N-fwd	CATTAAGAAAGTAGCACATATACAaATTTTGCA AATATAAACTGG	66
pET28a- CdHypD- D339N-rev	GGATTTATTCCACCAGTGTTTATATTTGCAAAATtT GTATATGTGCTAC	66

3.4.3 Overexpression and purification of proteins used for biochemical experiments

All proteins used in biochemistry studies were overexpressed and purified as previously described in Chapter 2. Briefly, all proteins were overexpressed in *E. coli* BL21-CodonPlus(DE3)-RIL *ΔproC::aac(3)IV*(Am^R) and purified in 25 mM Tris buffer pH 7.5 with 50 or 100 mM KCl. HypDs and HypD-AE were purified using TALON metal affinity resin while P5CR was purified using Ni-NTA resin. All protein solutions were made anaerobic prior to freezing and storage at –80 °C. Protein concentrations were calculated using Abs280 measurements from Nanodrop as previously reported. A molar extinction coefficient of 80,680 M⁻¹cm⁻¹ was used for Y450F and Y450F/T645A mutants, and 82,170 M⁻¹cm⁻¹ was used for remaining HypD variants. All proteins

were visualized by denaturing polyacrylamide gel electrophoresis to confirm high purity (Figure 3.7).

3.4.4 Overexpression and purification of HypD wild-type for crystallography studies^x

HypD was overexpressed in *E. coli* BL21-CodonPlus(DE3)-RIL. Protein used for crystallography studies was purified in 20 mM HEPES buffer pH 8.0, 100 mM NaCl, and 0.5 mM TCEP and was purified using Ni-NTA resin. All protein solutions were frozen and stored at -80°C.

3.4.5 Glycyl radical detection and quantification by EPR spectroscopy

HypD wild-type and mutants were activated using HypD-AE under assay conditions reported in Chapter 2. X-band EPR spectroscopy and spin concentration calculations were carried out as previously described using EasySpin (Version 5.0.22)³³ on MATLAB (MathWorks). All EPR assays with wild-type and mutants were performed in triplicate.

3.4.6 End-point activity assays with HypD mutants

All assays were prepared as previously described in Chapter 2. Briefly, assays contained 20 mM Tris-HCl pH 7.5, 100 mM KCl, 0.8 mM NADH, 3 μ M P5CR, 0.2 mM Hyp, and 0.3 μ M HypD. All assays were carried out in triplicate and were initiated by adding Hyp into Eppendorf tubes. Reaction mixtures were incubated for 21 h at 22 °C. Upon removal from the anaerobic chamber, reactions were quenched with a 2 $\times$ volume of methanol and protein precipitates were removed by centrifugation (15,200 g, 10 min). Supernatants were further diluted with water 30-

^x HypD used in crystallization studies was purified by Dr. Michael Funk in Prof. Catherine Drennan's lab, MIT.

fold for proline detection and 7.5-fold for hydroxyproline detection by LC-MS/MS using methods described in Chapter 2.

3.4.7 Kinetic analysis of HypD variants using a coupled spectrophotometric assay

Activated HypD mutant was used for coupled enzyme kinetic assays as previously described in Chapter 2 with a few modifications. All kinetic assays contained 20 mM Tris-HCl pH 7.5, 50 mM bicine pH 7.5, 100 mM KCl, and 400 μ M NADH. 3 μ M of HypD-Y450F (total monomer) and 0.75 μ M of HypD-T765A (total monomer) were used along with 2 $\times$ concentration of P5CR for each assay. Assays were initiated by addition of Hyp to a final concentration of 0, 1, 2, 5, 10, 15, 30, and 60 mM. Each data point represents conditions carried out in triplicate during a single experiment. The data was fit simultaneously to the Michaelis-Menten equation using nonlinear regression in Graphpad Prism 7.00. The k_{obs} parameter was calculated based on $29 \pm 4\%$ (mean $\pm$ SD) activation of HypD-F340F and $19 \pm 1\%$ activation of HypD-T645A as determined by EPR spectroscopic assays.

3.4.8 Crystallization of wild-type HypD^{xi}

Initial screening of unactivated wild-type HypD protein with the intact N-terminal His₆ tag was performed aerobically with the aid of an Art Robbins Phenix micro-pipetting robot and Formulatrix Rock Imager. Initial conditions were found using the Qiagen Protein Complex screen, with optimization yielding a well solution containing 14% (w/v) polyethylene glycol (PEG) 3350, 100 mM potassium chloride, and 100 mM HEPES pH 7.5. Diffraction-quality crystals were

^{xi} This section was written by Lindsey Backman in Prof. Catherine Drennan's lab, MIT. Crystallization of HypD was conducted by Lindsey Backman.

optimized in hanging drop vapor diffusion trays at 21 °C. Protein at 100 μM (9 mg mL⁻¹) in buffer containing 20 mM HEPES buffer pH 8.0, 100 mM NaCl, and 0.5 mM TCEP was mixed with well solution in a 1:1 ratio. Plate-like crystals formed and grew to maximum size after 1–2 days of equilibration. Crystals were cryoprotected by soaking for 1 min in solution containing 15% (v/v) glycerol, 14% (w/v) PEG 3350, 100 mM potassium chloride, and 100 mM HEPES pH 7.5.

The following modifications for crystallization of substrate-bound HypD. 4 mM Hyp was added to the crystallization buffer prior to mixing with protein at a 1:1 ratio. The 15% (v/v) glycerol was excluded in the cryoprotectant solution in which 25% (v/v) dimethyl sulfoxide (DMSO) and 100 mM Hyp were added.

3.4.9 Structure determination of HypD^{xii}

Crystals were indexed in space group $P2_1$ with cell edges $a = 100.3 \text{ \AA}$, $b = 341.7 \text{ \AA}$, $c = 122.6 \text{ \AA}$, $\beta = 107.1^\circ$, and diffraction images were collected at the Advanced Photon Source beamline 24ID-C at a wavelength of $0.9795 \text{ \AA}$ on a Pilatus 6M detector (Dectris) (Table 3.5). Data were indexed, integrated, and scaled in XDS.^{34,35} The first structure obtained for HypD contained a glycerol bound in the active site and was solved by molecular replacement in the Phenix implementation of Phaser.³⁶ CutC was used as a search model (PDB: 5FAU)⁹ after trimming of side chains in Phenix Ensembler.³⁷ A solution with eight molecules per asymmetric unit was found with an initial R_{free} of 0.52 at $2.05 \text{ \AA}$ resolution. Several rounds of initial refinement in phenix.refine³⁷ with tight NCS restraints and optimization of group B factors were sufficient to reduce R_{free} values below 0.4 (

^{xii} This section was written by Lindsey Backman in Prof. Catherine Drennan's lab, MIT. Determination of two HypD structures was conducted by Lindsey Backman.

Table 3.6). Subsequently modeling in side chains, and eventually modeling in water molecules, further reduced R_{free} to ~ 0.3 . NCS restraints were removed after initial refinement. Positional and individual B factor refinement continued at the full resolution until the model was complete.

The Hyp-bound HypD structure was later solved using this glycerol-bound HypD structure as a molecular replacement model. Initial molecular replacement of the Hyp-bound HypD structure in an initial R_{free} of 0.31 at 2.52 Å resolution (Table 3.5). Simulated annealing was performed after molecular replacement to decrease biases from the glycerol-bound structural. Positional and individual B factor refinement continued at the full resolution until the model was complete (

Table 3.6). Hyp was fit into difference densities and verified with simulated annealing composite omit maps. Substrate restraints were calculated by density function theory (DFT) as described below. Parameter files for Hyp were generated in Phenix eLBOW for fitting into the crystal structure.³⁸ Water molecules were placed automatically after ligands were refined and verified manually. No density is observed for residues 1–14 and the His₆ affinity tag. Structural figures were made in PyMOL v2.0.7 (The PyMOL Molecular Graphics System, Version 2.0 Schrodinger, LLC).

Table 3.5 Data collection statistics for crystallography.^{xiii}

Data collection and processing statistics for glycerol-bound and Hyp-bound HypD structures.

Values in parentheses denote highest resolution bin.

	HypD with glycerol bound	HypD with Hyp bound
Space group	P2 ₁	P2 ₁
Unit cell	100.3, 341.7, 122.6, 90.0, 107.1, 90.0	101.2, 350.2, 124.5, 90.0, 105.7, 90.0
Resolution (Å)	50–2.05 (2.09–2.05)	50–2.52 (2.59–2.52)
R _{sym}	16.8 (75.7)	20.4 (97.5)
CC _{1/2}	99.0 (58.8)	99.4 (72.1)
<I/σ>	8.40 (1.82)	10.75 (2.12)
Completeness (%)	99.0 (98.3)	16.0 (85.4)
Unique reflections	486251 (24062)	278476 (44812)
Total reflections	1626596 (168674)	1944676 (294711)
Redundancy	7.07 (7.01)	6.98 (6.58)

Table 3.6 Model refinement statistics for crystallography.^{xiii}

	HypD with glycerol bound	HypD with Hyp bound
R _{work} /R _{free}	0.166/0.193	0.186/0.224
RMSD bond length (Å)	0.007	0.008
RMSD bond angles (°)	0.86	0.966
Ramachandran analysis		
Favored (%)	98.16	97.71
Allowed (%)	1.71	1.99
Disallowed (%)	0.13	0.3
Rotamer outliers (%)	1.46	3.27
Average B factors		
Protein (Å ²)	21	35.8
Water (Å ²)	27.2	27.7
Hyp / Glycerol (Å ²)	22.3	31.2

^{xiii} Crystallography data collection and refinement were conducted by Lindsey Backman in Prof. Catherine Drennan's lab at MIT.

3.4.10 Density function theory calculations for Hyp conformation^{xiv}

Optimization of Hyp structure was performed at the B3LYP/6-31G(d) level of theory using Gaussian 16.¹⁰ The zwitterionic structure was enforced by freezing the N–H bond length during optimization. Without this constraint, the proton transferred to the carboxylate during optimization. Cartesian coordinates for the *exo* pucker are listed in Table 3.1. This structure was used to model the substrate into the electron density of HypD crystal structure. Cartesian coordinates for the *endo* pucker are listed in Table 3.2.

^{xiv} The work described in this section was conducted by Dr. Brian Gold in Prof. Ronald Raines lab, MIT

3.5 References

1. LRF Backman, MA Funk, CD Dawson, CL Drennan. New tricks for the glycyl radical enzyme family. *Crit Rev Biochem Mol Biol.* **2017**, 54(6):674-95.
2. JR O'Brien, C Raynaud, C Croux, L Girbal, P Soucaille, WN Lanzilotta. Insight into the mechanism of the B₁₂-independent glycerol dehydratase from *Clostridium butyricum*: preliminary biochemical and structural characterization. *Biochemistry.* **2004**, 43:4635-45.
3. JW LaMattina, ND Keul, P Reitzer, S Kapoor, F Galzerani, DJ Koch, IE Gouvea, WN Lanzilotta. 1,2-propanediol dehydration in *Roseburia inulinivorans*; structural basis for substrate and enantiomer selectivity. *J Biol Chem.* **2016**, 291(30):15515-26.
4. G Kalnins, J Kuka, S Grinberga, M Makrecka-Kuka, E Liepinsh, M Dambrova, K Tars. Structure and function of CutC choline lyase from human microbiota bacterium *Klebsiella pneumoniae*. *J Biol Chem.* **2015**, 290(35):21732-40.
5. B Zagalak, PA Frey, GL Karabatsos, RH Abeles. The Stereochemistry of the Conversion of D and L 1,2-Propanediols to Propionaldehyde. *Journal of Biological Chemistry.* **1966**, 241(13):3028-35.
6. J Rétey, A Umani-Ronchi, J Seibl, D Arigoni. Zum mechanismus der propandioldehydrase-reaktion. *Experientia* **1966**, 22:502-3.
7. S Craciun, EP Balskus. Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. *Proc Natl Acad Sci USA.* **2012**, 109(52):21307-12.
8. K Mori, T Oiwa, S Kawaguchi, K Kondo, Y Takahashi, T Toraya. Catalytic roles of substrate-binding residues in coenzyme B₁₂-dependent ethanolamine ammonia-lyase. *Biochemistry.* **2014**, 53(16):2661-71.
9. S Bodea, MA Funk, EP Balskus, CL Drennan. Molecular basis of C-N bond cleavage by the glycyl radical enzyme choline trimethylamine-lyase. *Cell chemical biology.* **2016**, 23(10):1206-16.
10. MJFaGWTaHBSaGESaMARaJRCa G. (2016). Gaussian 16, Revision A.03. Gaussian Inc. Wallingford CT.
11. MD Shoulders, RT Raines. Collagen structure and stability. *Annu Rev Biochem.* **2009**, 78(1):929-58.
12. A Lesarri, EJ Cocinero, JC López, JL Alonso. Shape of 4(*S*)- and 4(*R*)-hydroxyproline in gas phase. *Journal of the American Chemical Society.* **2005**, 127(8):2572-9.

13. BJ Levin*, YY Huang*, SC Peck, Y Wei, A Martínez-del Campo, JA Marks, EA Franzosa, C Huttenhower, EP Balskus. A prominent glycyl radical enzyme in human gut microbiomes metabolizes *trans*-4-hydroxy-L-proline. *Science*. **2017**, 355(6325):eaai8386.
14. F Sievers, A Wilm, D Dineen, TJ Gibson, K Karplus, W Li, R Lopez, H McWilliam, M Remmert, J Söding, et al. Fast, scalable generation of high - quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. **2011**, 7(539).
15. M Feliks, GM Ullmann. Glycerol dehydration by the B₁₂-independent enzyme may not involve the migration of a hydroxyl group: a computational study. *J Phys Chem B*. **2012**, 116(24):7076-87.
16. DD Perrin. The effect of temperature on pK values of organic bases. *Australian Journal of Chemistry*. **1964**, 17(4):484-8.
17. NJ Zondlo. Aromatic–proline interactions: Electronically tunable CH/ π interactions. *Accounts of Chemical Research*. **2013**, 46(4):1039-49.
18. A Benjdia, C Balty, O Berteau. Radical SAM enzymes in the biosynthesis of ribosomally synthesized and post-translationally modified peptides (RiPPs). *Frontiers in Chemistry*. **2017**, 5(87).
19. HG Viehe, Z Janousek, R Merenyi, L Stella. The captodative effect. *Accounts of Chemical Research*. **1985**, 18(5):148-54.
20. B Beatrix, O Zelder, FK Kroll, G Örlýgsson, BT Golding, W Buckel. Evidence for a mechanism involving transient fragmentation in carbon skeleton rearrangements dependent on coenzyme B₁₂. *Angewandte Chemie International Edition in English*. **1995**, 34(21):2398-401.
21. W-c Chang, M Dey, P Liu, SO Mansoorabadi, S-J Moon, ZK Zhao, CL Drennan, H-w Liu. Mechanistic studies of an unprecedented enzyme-catalysed 1,2-phosphono-migration reaction. *Nature*. **2013**, 496:114–8.
22. H Gaweska, PF Fitzpatrick. Structures and mechanism of the monoamine oxidase family. *Biomolecular concepts*. **2011**, 2(5):365-77.
23. P Renaud, L Giraud. 1-Amino- and 1-amidoalkyl radicals: Generation and stereoselective reactions. *Synthesis*. **1996**, 1996(08):913-26.
24. K Nakajima, Y Miyake, Y Nishibayashi. Synthetic utilization of α -aminoalkyl radicals and related species in visible light photoredox catalysis. *Accounts of Chemical Research*. **2016**, 49(9):1946-56.
25. AS Gizzi, TL Grove, JJ Arnold, J Jose, RK Jangra, SJ Garforth, Q Du, SM Cahill, NG Dulyaninova, JD Love, et al. A naturally occurring antiviral ribonucleotide encoded by the human genome. *Nature*. **2018**, 558(7711):610-4.

26. Y Liu, AA Gallo, J Florian, YS Liu, S Mora, W Xu. QM/MM (ONIOM) study of glycerol binding and hydrogen abstraction by the coenzyme B₁₂-independent dehydratase. *The journal of physical chemistry B*. **2010**, 114(16):5497-502.
27. Y Zhongyue, M Rimsha, W Mengyi, Q Helena W., S Adam H., K Heather. Revealing quantum mechanical effects in enzyme catalysis with large-scale electronic structure simulation. *ChemRxiv Preprint*. **2018**.
28. Z H. Radical stability—a theoretical perspective. In: A. G, editors. Radicals in Synthesis I. Springer, Berlin, Heidelberg; **2006**.
29. DJ Henry, CJ Parkinson, PM Mayer, L Radom. Bond dissociation energies and radical stabilization energies associated with substituted methyl radicals. *The Journal of Physical Chemistry A*. **2001**, 105(27):6750-6.
30. Y-R Luo. Comprehensive handbook of chemical bond energies. CRC Press CRC Press **2007**.
31. BW Arentson, N Sanyal, DF Becker. Substrate channeling in proline metabolism. *Frontiers in bioscience : a journal and virtual library*. **2012**, 17:375-88.
32. M Kearse, R Moir, A Wilson, S Stones-Havas, M Cheung, S Sturrock, S Buxton, A Cooper, S Markowitz, C Duran, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. **2012**, 28(12):1647-9.
33. S Stoll, A Schweiger. EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. *J Magn Reson*. **2006**, 178(1):42-55.
34. W Kabsch. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica Section D*. **2010**, 66(2):133-44.
35. W Kabsch. XDS. *Acta Crystallogr D Biol Crystallogr*. **2010**, 66(Pt 2):125-32.
36. AJ McCoy, RW Grosse-Kunstleve, PD Adams, MD Winn, LC Storoni, RJ Read. Phaser crystallographic software. *Journal of applied crystallography*. **2007**, 40(Pt 4):658-74.
37. PD Adams, PV Afonine, G Bunkoczi, VB Chen, IW Davis, N Echols, JJ Headd, LW Hung, GJ Kapral, RW Grosse-Kunstleve, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*. **2010**, 66(Pt 2):213-21.
38. NW Moriarty, RW Grosse-Kunstleve, PD Adams. electronic Ligand Builder and Optimization Workbench (eLBOW): a tool for ligand coordinate and restraint generation. *Acta Crystallogr D Biol Crystallogr*. **2009**, 65(Pt 10):1074-80.

4 Chapter 4. Anaerobic *trans*-4-hydroxy-L-proline utilization mediated by HypD is a widespread microbial metabolic activity

Parts of this chapter were adapted from published work.¹

4.1 Introduction

As described in Chapter 2, we discovered and characterized a new glyceryl radical enzyme (GRE), 4-hydroxyproline dehydratase (HypD). The *hypD* gene cluster found in Clostridiales encodes three genes responsible for anaerobic metabolism of *trans*-4-hydroxy-L-proline (Hyp) to L-proline (Pro). The structural and biochemical characterization of HypD was reported in Chapter 3, which provided insights into the molecular basis for this activity. Quantitative profiling of GREs across Human Microbiome Project (HMP) metagenomes revealed a universal distribution of *hypD* in gut microbiomes of healthy subjects. Furthermore, a search of this gene in sequenced genomes uncovered many common gut microbes as well as environmental isolates that encode HypD. Altogether, this work elucidated the biochemistry of a prominent metabolic pathway in the gut microbiome.

HypD's presence in hundreds of sequenced bacterial and archaeal genomes prompted us to examine the phylogenetic relationship of representative HypD sequence, which is reported in this Chapter. We further searched the gene neighborhoods of *hypD* for annotations that may provide clues about how different organisms may transform the product of HypD, metabolite P5C. This body of work identifies possible roles for anaerobic Hyp metabolism in evolutionarily distant groups of organisms. In addition to human-associated microbes, HypD is present in a diverse range of environmental isolates. This motivated our work on the enrichment culturing of Hyp-degrading bacterial communities from sediment samples, which is also described in this Chapter. We sought

to link HypD to the observed Hyp-degrading activity in these cultures through amplification of *hypD* using degenerate PCR. Overall, this Chapter describes studies of HypD phylogeny that revealed distinct clades of sequences from diverse organisms and enrichment of HypD-degrading environmental microbes uncovered novel HypDs and isolates that have not been sequenced.

4.2 Results and discussions

4.2.1 Phylogenetic analysis of HypD sequences

A particularly striking feature of HypD is its wide distribution in sequenced microbial genomes. Previous BLASTP² searches revealed HypD in >850 genomes from the NCBI and InterPro databases, with the majority of hits found in Firmicutes and Bacteroidetes (described in Chapter 2).³ In addition to being widespread across bacterial phyla, HypD is also present in several archaeal genomes. To infer evolutionary relationships between HypD homologs from different bacteria and archaea, we constructed a phylogenetic tree using 84 representative protein sequences from the UniProt database⁴ (Figure 4.1, Table 4.1). One sequence from each species was included for analysis to cover the majority of taxonomic diversity observed for HypDs in this database. Briefly, a maximum-likelihood tree was constructed using MEGA7⁵ from a multiple sequence alignment generated using MAFFT⁶.

Table 4.1 UniProt ID and genomic context for sequences used to generate the HypD phylogenetic tree.

Strain name	UniProt ID	P5C reductase^b	P5C dehydrogenase^b
Acetomicrobium mobile (ATCC BAA-54 / DSM 13181 / NGA)	I4BZ30		
Aciduliprofundum boonei (DSM 19572 / T469)	B5IEV9		
Aciduliprofundum sp. MAR08-339	L0HNV0		
Alistipes sp. CAG:831	R5I1N0		
Alkaliphilus metalliredigens QYMF	A6TKU3		
Aminomonas paucivorans DSM 12260	E3CV25		
Bacteroides dorei DSM 17855	B6VYJ0		
Bacteroides massiliensis dnLKV3	R9IBD5		
Bacteroides vulgatus DSM 1447	A6L095		
Bacteroidetes sp. oral taxon 272	S8GPK1		
Cetobacterium somerae ATCC BAA-474	U7VC55	+	
Clostridioides difficile 630	Q180C6	+	
Clostridium bifermentans ATCC 638	T4VHU8	+	
Clostridium botulinum Loch Maree/A3	B1L0Y3	+	
Clostridium butyricum VPI 1718, GD	Q8GEZ8 ^a		
Clostridium carboxidivorans P7	C6PPT9	+	
Clostridium cf saccharolyticum K10	D6DEN7		
Clostridium hiranonis DSM 13275	B6FXX5	+	
Clostridium hylemonae DSM 15053	C0BZI1	+	
Clostridium innocuum 2959	N9V368	+	
Clostridium kluyveri ATCC 8527	A5N4Z6	+	
Clostridium scindens ATCC 35704	B0NDW5		
Clostridium sp. CAG:242	R6D016		+
Clostridium sp. CAG:43	R7AIS9		
Clostridium sp. D5	F0Z4X0	+	
Clostridium sp. DL-VIII	G7MD03		
Clostridium sp. MSTE9	J0N1R7	+	
Clostridium sporogenes ATCC 15579	J7TB00	+	
Clostridium tyrobutyricum DIVETGP	W6N7H0		
Collinsella sp. CAG:289	R7D6G9	+	

Table 4.1 (Continued)

Coprococcus comes ATCC 27758	C0BAQ4		
Coprococcus sp. HPP0048	S2YN60	+	
Coprothermobacter proteolyticus (ATCC 35245 / DSM 5265 / BT)	B5Y6E3		
Deferribacter desulfuricans DSM 14783	D3PA11		
Denitrovibrio acetiphilus DSM 12809	D4H8D6	+	+
Desulfobacterium autotrophicum ATCC 43914	C0QIN7		
Desulfomicrobium baculatum DSM 4028	C7LR40		
Desulfosporosinus acidiphilus DSM 22704	I4D335		+
Desulfosporosinus meridiei ATCC BAA-275	J7IXV3		+
Desulfosporosinus orientis ATCC 19365	G7WE24		+
Desulfosporosinus sp. OT	G2G0P1		
Desulfosporosinus youngiae DSM 17734	H5Y5V0		+
Desulfotalea psychrophila LSv54	Q6AQM8		
Desulfotomaculum hydrothermale Lam5	K8EEC7		
Desulfotomaculum reducens MI-1	A4J8R5		
Desulfotomaculum ruminis ATCC 23193	F6DS33		
Desulfovibrio alaskensis G20, CutC	Q30W70 ^a		
Desulfurella acetivorans A63	W0JIJ0		
Dorea longicatena DSM 13814	A6BHN6		
Dorea sp. 5-2	R9NEA8	+	
Dorea sp. CAG:317	R6WBG4	+	
Draconibacterium orientale DSM 25947	X5DGL1		
Enterococcus avium ATCC 14025	S0L759		
Enterococcus gilvus ATCC BAA-350	R2XU16		
Enterococcus malodoratus ATCC 43197	R2R5K9		
Enterococcus pallens ATCC BAA-351	R2SUA9		
Enterococcus raffinosus ATCC 49464	R2RF16		
Geobacter bemidjensis DSM 16622	B5EGM4		
Geobacter sp. M18	E8WN65		
Geobacter sulfurreducens ATCC 51573	Q74BE0		
Hipaea maritima ATCC 700847	F2LTL4		
Johnsonella ignava ATCC 51276	G5GJG7	+	
Candidatus Korarchaeum cryptofilum OPF8	B1L491		
Natronaerobius thermophilus DSM 18059	B2A1A1		
Odoribacter laneus YIT 12061	H1DHL3		

Table 4.1 (Continued)

Odoribacter sp. CAG:788	R5P085		
Odoribacter splanchnicus ATCC 29572	F9Z3G5		
Oscillibacter valericigenes DSM 18026	G4KZA5		+
Parabacteroides distasonis DSM 20701	A6LA42		
Parabacteroides goldsteinii CL02T12C30	K5YCC2		
Parabacteroides johnsonii DSM 18315	B7BAT2		
Parabacteroides merdae ATCC 43184	A7AGL9		
Parabacteroides sp. CAG:409	R7JLM7		
Pelosinus fermentans JBW45	I9NLE5		
Peptostreptococcus anaerobius CAG:621	R5JGP5	+	
Pseudotherrmotoga lettingae (ATCC BAA-301 / DSM 14385 / NBRC 107922 / TMO)	A8F6C2		
Stomatobaculum longum	H1HV67		+
Synergistes sp. 3 1 syn1	G9PSU2		
Tepidanaerobacter acetatoxydans (DSM 21804 / JCM 16047 / Re1)	F4LRJ5	+	+
Thermanaerovibrio acidaminovorans DSM 6589	D1B6M6		
Thermanaerovibrio velox DSM 12556	H0UP67		
Thermococcus kodakarensis (ATCC BAA-918 / JCM 12380 / KOD1)	Q5JFY6	+	
Thermococcus litoralis (ATCC 51850 / DSM 5473 / JCM 8560 / NS-C)	H3ZMH7		
Thermococcus sibiricus (MM 739 / DSM 12597)	C6A251	+	
Thermovirga lienii ATCC BAA-1197	G7V7P2		
Treponema lecithinolyticum ATCC 700332	U2LWK4	+	
Treponema maltophilum ATCC 51939	S3K3Z2	+	
Vagococcus lutrae LBD1	V6Q678		

^a *GD* and *CutC* were included as outgroups.

^b Presence of gene within 25 kbp upstream or downstream neighborhood of *hypD*.

In agreement with our previous analysis of HypD distribution in sequenced genomes, most representative HypD sequences included in the phylogenetic tree are encoded in species that have human-associated isolates (Figure 4.1). HypD is found in many common oral and gut species from genera *Treponema*, *Clostridium*, *Bacteroides*, *Parabacteroides*, and *Enterococcus*. This led us to

propose that HypD is likely responsible for the anaerobic Hyp degradation previously observed in human fecal and gingival samples.^{7,8} In addition to human commensals, HypD is also encoded in several notable human pathogens including *C. difficile* and *Clostridium botulinum*.

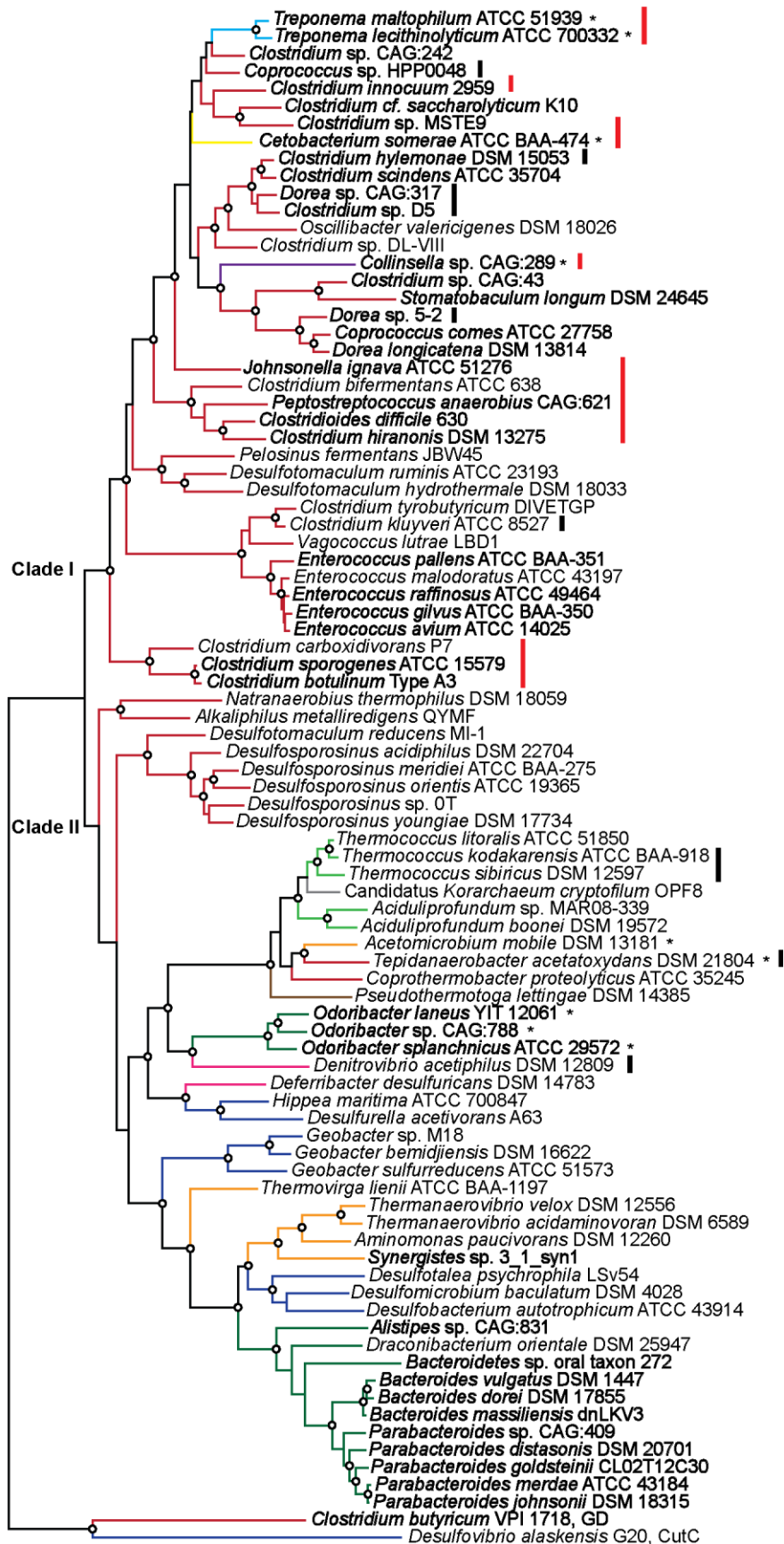
In the phylogenetic tree, the evolutionary relationships of HypD sequences broadly reflect the taxonomic relatedness of their host species. Two major clades can be discerned from the tree with Clade I consisting of mostly sequences from human-associated Firmicutes and Clade II comprised of sequences from environmental Firmicutes and most of the other phyla. Furthermore, Clade I is enriched with isolates encoding a gene annotated as *proC* nearby, with this gene most often located immediately adjacent to *hypD* suggesting co-transcription (Table 4.1, Figure 4.1). The vast majority of HypDs in Clade I are from Firmicutes except for four sequences from *Cetobacterium somerae*, *Collinsella* sp., and *Treponema* species representing three distinct phyla (Figure 4.1Figure 1.1). Interestingly, *hypD* sequences in these bacteria are colocalized with an adjacent *proC*, sharing the genomic organization commonly present in Clostridiales and within this clade. This shared genomic context could indicate a recent horizontal gene transfer event. Within Clade II, most isolates are of environmental origins except for the Bacteroidetes, which were mostly isolated from the human body. HypDs from *Odoribacter* species form a distinct clade from HypDs present in other Bacteroidetes. Interestingly, archaeal HypDs are evolutionary close to each other as expected, but these sequences form a clade with HypDs from thermophilic bacteria as well. It is possible that a recent horizontal gene transfer event occurred between bacteria and archaea that inhabit similar habitats. These sequences form a deep-branching clade and therefore are evolutionary distant from most bacterial HypDs. An examination of the genomic contexts of Clade II sequences revealed that *proC* is rarely found within the gene neighborhoods (25 kbp upstream and downstream) of these *hypD* genes. Instead, some of these genomes encode a putative

P5C dehydrogenase nearby indicating that P5C may be oxidized to L-glutamate in these species (Table 4.1).

The majority of representative HypD sequences used for this analysis are encoded within human-associated isolates, although this may be a result of biases toward cultivation and sequencing of clinically relevant microbes. Since our retrieval of HypD sequences from databases in 2016, advances in genome assembly have yielded genome sequences for thousands of novel, uncultivated bacterial and archaeal species.⁹⁻¹¹ A search for HypD in these genomes will likely expand the currently known distribution of this protein to additional phyla. The presence of HypD in 11 phyla identified in our analysis suggests that Hyp utilization may be an important metabolic function in anaerobes. However, in most species, the precise pathways by which Hyp is metabolized cannot be readily inferred from genomic context. The presence of *proC* within gene neighborhoods of *hypD* (25 kbp upstream and downstream) is only present in a small portion of genomes (Figure 4.1, Table 4.1). Altogether, this supports a wide range of roles for Hyp metabolism among highly diverse microorganisms.

Figure 4.1 A phylogenetic tree of representative HypD sequences from the UniProt database.

A maximum likelihood tree of 86 full-length HypD amino acid (aa) sequences from the UniProt database. The two major clades described in the text are indicated at the corresponding branches as Clade I and Clade II. HypD sequences from species that clade differently from 16S rDNA phylogeny are indicated with an asterisk. HypDs with a *proC* in their gene neighborhoods are indicated with a red or black vertical bar. An immediately adjacent *proC* in the genome is represented by a red bar whereas a *proC* that is farther away, but within 25 kbp upstream or downstream of *hypD* is represented by a black bar. The phylum of the isolate encoding each HypD is indicated by the color of its branch as indicated in the figure legend. Bootstrap values of 70 to 100% are indicated by open circles. Choline trimethylamine-lyase (CutC) and glycerol dehydratase (GD) are included as outgroups. See Table 4.1 for a list of UniProt IDs of sequences used to construct this phylogenetic tree. This table also indicates the presence of a putative P5C reductase or P5C dehydrogenase gene in the gene neighborhoods of each HypD.



4.2.2 Anaerobic enrichment culturing of Hyp-degrading bacteria from sediment samples

In addition to human-associated microbes, we located HypD in the genomes of many environmental bacteria mostly from the phyla Firmicutes, Proteobacteria, and Synergistetes, with these sequences making up the second major clade in the phylogenetic tree (Figure 4.1). Almost all of these genomes lack a *proC* gene near *hypD* and thus these bacteria likely do not metabolize Hyp to Pro as part of Stickland fermentation (Table 4.1). The wide distribution of anaerobic Hyp utilization in these environmental isolates suggests that this metabolic activity is important in a diverse range of microbial habitats beyond the human gut.

To gain insights into Hyp metabolism in environmental bacteria, we sought to examine the presence of this metabolic capability in hypoxic sediments through enrichment culturing (Figure 1.3). Community compositions of enrichment cultures observed to degrade Hyp anaerobically could be determined by 16S rDNA sequencing. Any *hypD* genes in these communities could be amplified and detected through degenerate PCR to support an essential role of HypD in anaerobic Hyp metabolism. We envisioned this workflow would enable us to identify Hyp-metabolizing microbes from sediment environments as well as establish HypD as the likely enzyme responsible for this activity. In addition, sequencing data from degenerate PCR could reveal additional diversity among HypD homologs that is currently not covered in sequence databases.

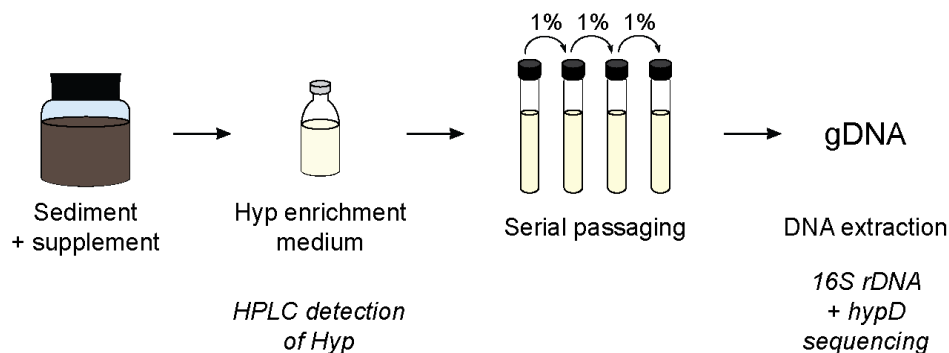


Figure 4.2 Workflow for enrichment culturing of Hyp-degrading microbes.

Environmental samples were collected from various wetlands in Woods Hole, MA, USA (Table 4.2). An enrichment medium was designed to include Hyp as a source of carbon, nitrogen, and electron acceptor (Table 4.3). Genomic DNA (gDNA) was extracted from four enrichment cultures after their 4th passage for amplicon sequencing.

Table 4.2 Locations of samples collected to inoculate enrichment cultures.

All samples were collected at various locations around Woods Hole, MA, USA.

Sampling site	Sample type	Supplement*	Passaged and sequenced cultures
Trunk River	Sediment	Beef extract	TR Beef
		Peptone	TR Peptone
		Hyp	TR Hyp
Cedar Swamp	Mud		Ced. Swamp
Little Sippwissett Salt Marsh	Microbial mat		
School Street Marsh	Mud		

* Samples were supplemented with Hyp or various Hyp-containing sources and incubated at room temperature before inoculation into enrichment medium.

First, a minimal medium was designed for enrichment culturing that included Hyp as an electron acceptor and a source of carbon and nitrogen (Table 4.3). Since most HypD-possessing environmental isolates are non-Clostridiales, the medium was designed to contain multiple sources of electron acceptors to enable microbial growth on Hyp as a source of carbon and nitrogen as well. Sediment samples were collected from various wetlands around Woods Hole, MA, USA (

Table 4.2). A subset of these samples was first incubated in the presence of Hyp or Hyp-containing supplements before inoculation into the defined medium (

Table 4.2). Interestingly, varying degrees of Hyp consumption were observed in all initial enrichment cultures via High Performance Liquid Chromatography-Refractive Index Detector (HPLC-RID). Four fast-growing Hyp-metabolizing enrichment cultures were selected for serial passaging to reduce community complexity (

Table 4.2, Figure 4.3). Genomic DNA (gDNA) was extracted from these passaged cultures for amplicon sequencing to analyze the community composition and to detect the presence of *hypD*.

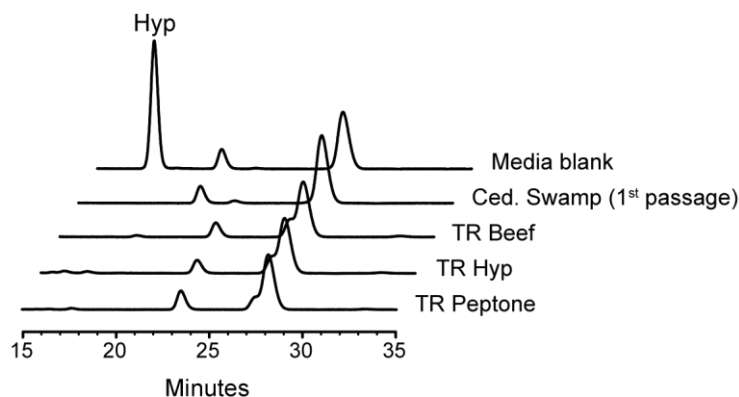
Table 4.3 Composition of the Hyp enrichment medium.

Component	Concentration
NaCl	100 g/L
MgCl ₂ •6H ₂ O	40 g/L
CaCl ₂ •2H ₂ O	10 g/L
KCl	50 g/L
NH ₄ Cl	10 mM
KH ₂ PO ₄	4 g/L
K ₂ HPO ₄	12.7 g/L
MOPS buffer, pH 7.2	10 mM
Sodium formate	40 mM
Sodium acetate	1 mM
Hyp	20 mM
Thiosulfate	0.25 mM
Sodium sulfide	0.25 mM

Table 4.3 (Continued)

Sodium 2-bromoethanesulfonate*	10 mM
ATCC trace mineral	1%
ATCC vitamin solution	1%
Resazurin	0.000001%

* added after passaging 2–3 times

**Figure 4.3 Hyp detection in enrichment cultures using HPLC-RID.**

High Performance Liquid Chromatography-Refractive Index Detector (HPLC-RID) traces for the four enrichment cultures further passaged for sequencing. Traces of cultures that were inoculated directly with sediment samples are shown for TR Beef, TR Hyp, and TR Peptone cultures. Cedar Swamp trace is shown for the first passaged enrichment culture. All cultures were turbid at the time of sampling. Consumption of Hyp was observed in all four cultures shown here.

4.2.3 Community composition of enrichment cultures by 16S rDNA sequencing

To identify the bacteria present in the passaged enrichment cultures, 16S rDNA sequences were amplified using universal primers 8F and 1510R (Table 4.4).¹² Clone libraries were generated using the amplified DNA and 24 clones from each enrichment culture library were sequenced

(Figure 4.4). A total of 75 high quality sequences were obtained from 96 clones. Redundant sequences (100% nucleotide ID) obtained within each enrichment culture and chimeric sequences were removed to yield a final list of 28 sequences for analysis. A range of 1.2 to 7.7-fold redundancy was obtained from the clone libraries, demonstrating good coverage of most community compositions (Table 4.5). Unexpectedly, the TR Beef community was much more complex compared to other cultures with 14 unique 16S rDNA sequences and sequencing of additional clones would likely reveal new isolates. Additional steps that could have been incorporated to reduce the number of isolates within each community include further serial passaging and transfer into defined media designed to constrain Hyp utilization to specific metabolic roles. Examples of the latter strategy include removal of formate and acetate as potential carbon sources in the enrichment medium after a few initial passages.

Table 4.4 Primers used in this study.

Target gene	Primer name	Primer sequence (5' to 3')	Amplicon size (bp)	Annealing temperature (°C)
16S rRNA	8F	AGAGTTTGATCCTGGCTCAG	~1500	50
16S rRNA	1510R	ACGGCTACCTTGTTACGACTT	~1500	50
<i>hypD</i>	HypD F149aa	TTYACIGARTTYATGGARCA	357	50
<i>hypD</i>	HypD R261aa	TGNACRAACCARTACATYTG	357	50

Table 4.5 Clone libraries generated for 16S rDNA sequencing of enrichment cultures.

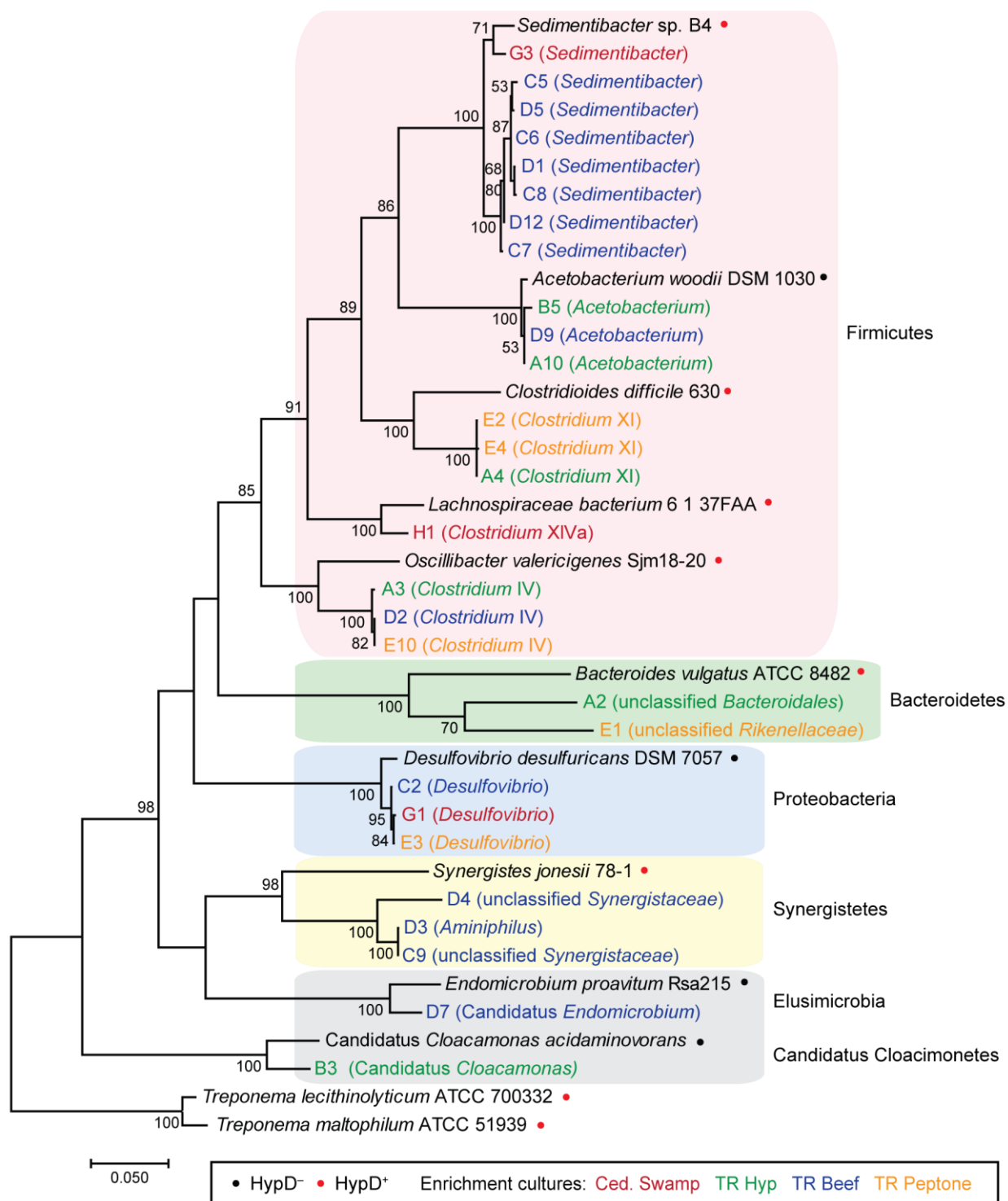
Enrichment culture	Ced. Swamp	TR Beef	TR Hyp	TR Peptone
Quality sequences	23	19	20	13
Unique sequences	3	16	6	5
Redundancy (quality / unique sequences)	7.7	1.2	3.3	2.6
Chimeric sequences	0	2	0	0
Included for analysis	3	14	6	5

Phylogenetic analysis of these 16S rDNA sequences revealed that the enrichments were dominated by Firmicutes (Figure 4.4). Interestingly, only a single 16S rDNA sequence was identical among all communities even though three of the enrichments were inoculated with the same environmental sample (Trunk River, TR). The same isolate from genus *Clostridium* IV was present in TR Peptone and TR Beef cultures (Figure 4.4). The enrichment inoculated with sediment from Cedar Swamp consisted of two species highly similar to the Trunk River enrichments by 16S analysis and one unique species in *Clostridium* XIVa (Figure 4.4). None of the 16S rDNA sequences yielded identical matches in the NCBI non-redundant nucleotide database when uncultured and environmental isolates were excluded in the search.¹³ Similarly, no hits among cultivated isolates were found in the Ribosomal Database Project (RDP) database. Therefore, these enrichment cultures consisted of uncultivated isolates. Single colonies were isolated from each culture and could be a source of interesting novel species for future sequencing efforts. In particular, two sequences are part of recently identified phyla (Candidatus Cloacimonete and Elusimicrobia) that have no or very few cultivated representatives.^{14,15} Most of the species present in these Hyp enrichments are part of genera or class containing sequenced isolates that encode HypD, indicating they are likely to possess this enzyme (Figure 4.4). Interestingly, a few *Acetobacterium* spp. and *Desulfovibrio* spp. were present in these communities, but no sequenced isolates from these two

genera are known to encode HypD. Since these 16S rDNA sequences are unique from 16S rRNAs present among sequenced genomes, these enriched species are novel strains that could potentially encode HypD. Alternatively, these isolates could be utilizing other components of the media such as formate and sulfate as electron acceptors for growth through previously studied pathways. *Acetobacterium* species are known acetogens that reduce CO₂ to acetate via formate whereas *Desulfovibrio* species are sulfate-reducing bacteria.^{16,17} Overall, the presence of Firmicutes, Bacteroidetes, and Synergistetes in these enrichment cultures is consistent with the distribution of HypD previously observed in sequenced genomes.³

Figure 4.4 A phylogenetic tree of 16S rDNA sequenced from enrichment cultures.

A maximum likelihood tree of 40 full-length 16S rRNA gene nucleotide sequences. 28 genes amplified from each enrichment culture and 12 genes from reference species were included. Chimeric, low-quality, and redundant sequences from each enrichment cultures were excluded in the analysis. Sequences were aligned using MUSCLE¹⁸ and manually trimmed to yield a final length of 1371 bp. The tree was constructed using MEGA7.⁵ Bootstrap values from 50 to 100% are displayed. The highest similarity score calculated by RDP SeqMatch (September, 2017) was used to assign the lowest classification level for each isolate, which are listed in parentheses.¹⁹



4.2.4 Detection of *hypD* in enrichment cultures through degenerate PCR

Detection of both Hyp degradation and the presence of isolates related to HypD-encoding organisms in passaged cultures led us to propose that HypD is responsible for the observed anaerobic Hyp metabolism in these samples. We sought to obtain further evidence to support this hypothesis by directly identifying *hypD* genes in these enrichments using degenerate PCR. Dr. Ana Martínez-del Campo designed degenerate primers to target conserved regions present in all HypD sequences from the UniProt database (Figure 4.5A, Table 4.4).ⁱ Amplification conditions were optimized using gDNA from Clostridiales and Bacteroidales. One degenerate primer pair was selected for degenerate PCR of gDNA from enriched communities based on its selectivity and robust amplification of *hypD* (Figure 4.5B). A fragment at the expected size for the targeted *hypD* partial sequence was amplified from all enrichment cultures serially passaged four times (Figure 4.6A). A clone library was constructed for each sample for *hypD* amplicon sequencing. Nucleotide and translated amino acid sequences that are redundant within each microbial community were excluded from analyses.

ⁱ Degenerate primer design, optimization of degenerate PCR conditions, and amplification of *hypD* from enrichment culture gDNA were conducted by Dr. Ana Martínez-del Campo. Clone libraries of 16S rDNA and *hypD* fragments were prepared in conjunction with Dr. Ana Martínez-del Campo for sequencing.

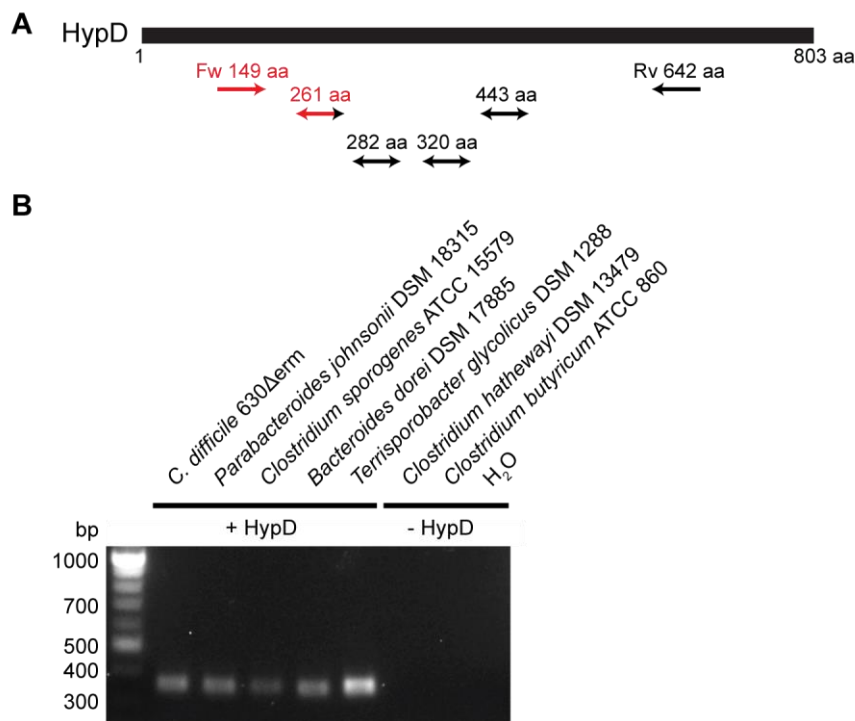


Figure 4.5 Degenerate primers designed to amplify *hypD* partial fragments.

(A) Regions of HypD that were targeted by degenerate primers. Amino acid numbering used to name primers is based on HypD from *C. difficile* 70-100-2010 (UniProt ID: EHJ39707). The primer pair used to optimize degenerate PCR conditions and to amplify *hypD* from enrichment culture DNA is highlighted in red.

(B) Degenerate primer pair F149aa and R261aa amplified an expected 357 bp product from gDNA of HypD-encoding isolates. No amplification product was observed using the gDNA of species lacking HypD.

A BLASTN² search of these amplicon sequences yielded hits of low similarity (<80% ID) in the NCBI nucleotide database, further demonstrating that the Hyp-metabolizing species in these enrichments have not been sequenced. To confirm that these amplicons do represent *hypD* fragments, a BLASTP² search was performed using translated amino acid sequences of HypD

partial genes unique within each enrichment as query against all GREs from our previously constructed sequence similarity network.³ All top 50 hits for each amplicon mapped back to nodes within the HypD cluster of the sequence similarity network, with most of the enrichment sequences resembling HypD sequences from Firmicutes. This result is consistent with the high abundance of Firmicutes in these communities as determined from the 16S rDNA analysis. Furthermore, the top hits from a BLASTP search against the non-redundant protein sequences in the NCBI database contained all important active site residues identified from our HypD mutagenesis studies (discussed in Chapter 3). There are two active site residues (*C. difficile* HypD: F152, H160)³, initially identified from the homology model discussed in Chapter 2, that are encoded within the region of *hypD* amplified by our degenerate primers. These two residues are conserved among all *hypD* amplicons sequenced from enrichment cultures (Figure 4.6B). While H160 is also conserved among GRE dehydratases (PD and GD), F152 is uniquely conserved in putative HypDs (Figure 2.3) as discussed in Chapter 2. Together, these results strongly support the presence of HypD in the four sediment enrichment cultures and suggest that this enzyme was responsible for the observed Hyp consumption. Lastly, the discovery of new, partial HypD sequences points to a much broader distribution of this activity across microbial diversity than is currently captured in sequenced genomes. Construction of a phylogenetic tree using these nucleotide sequences and the corresponding region in known HypDs will help to infer the evolutionary relationships of these new HypDs. Isolation of individual strains from these microbial communities that retain Hyp degradation capability and detection of *hypD* will provide strong support for HypD mediating the anaerobic Hyp metabolic pathway. This will also enable follow-up physiological studies on previously uncultivated isolates including the two isolates in phyla Candidatus Cloacimonete and Elusimicrobia. Additional serial passaging of the enrichments into more minimal media could

reduce community complexity to allow genome assemblies from metagenomic sequencing, which would provide an overview of metabolic capabilities and sequences of full-length HypDs.

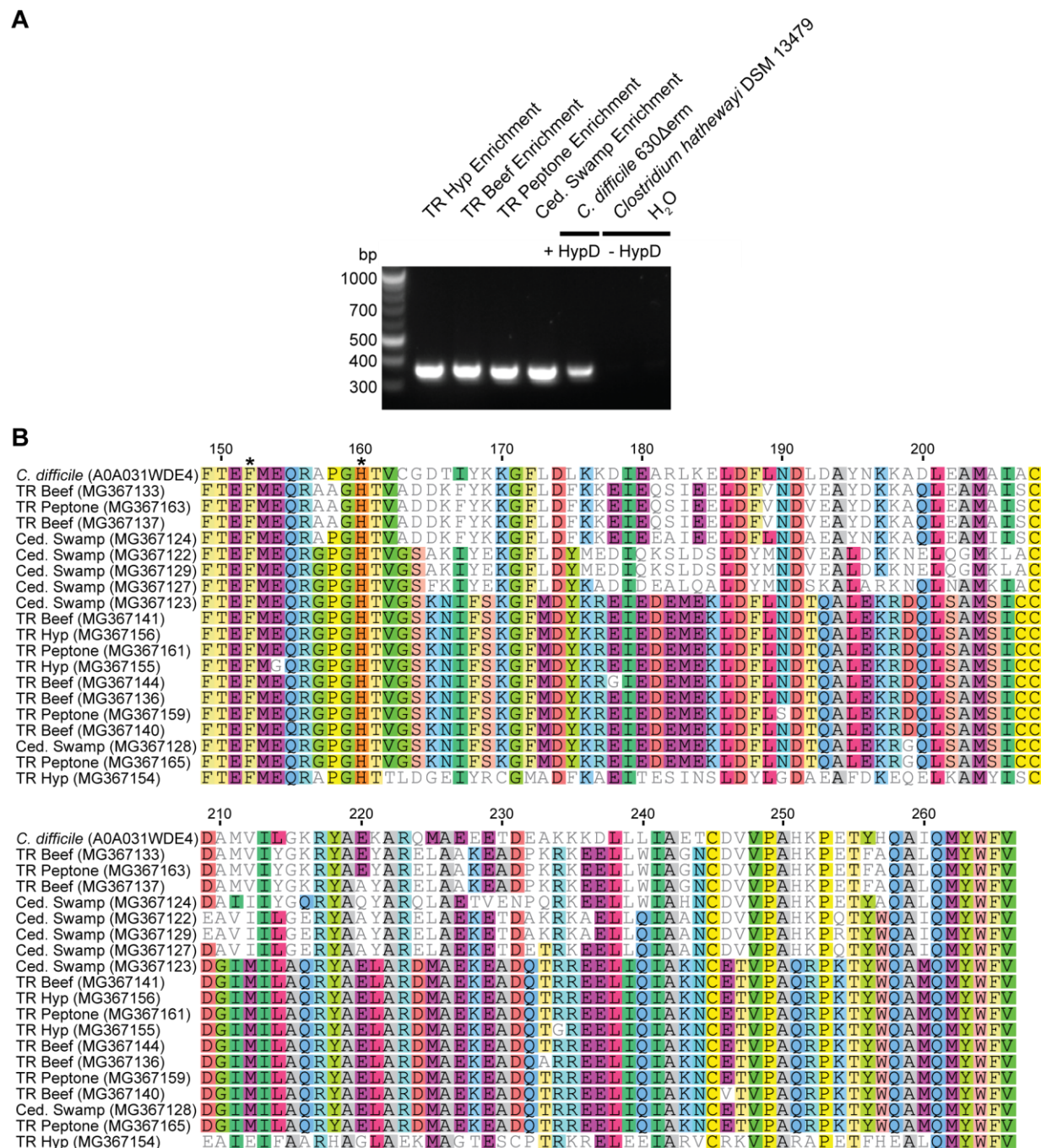


Figure 4.6 Amplification of partial *hypD* sequences using degenerate PCR.

(A) PCR products of ~360 bp were observed in all four enrichment cultures using primers HypD F149aa and R261aa. High quality nucleotide sequences that are non-redundant within each community were deposited on GenBank (accession:

Figure 4.6 (Continued)

MG367122–MG367168). (B) Non-redundant translated amino acid sequences from each enrichment culture were aligned using Clustal Omega²⁰ along with the biochemically characterized HypD from *C. difficile* 70-100-2010 (UniProt ID: A0A031WDE4). Two active site residues encoded within this amplicon are present in all enrichment sequences and are indicated by asterisks. Residue numbering is based on HypD sequence from *C. difficile* 70-100-2010.

4.3 Conclusions

In this chapter, we analyzed the phylogenetic relationships of putative HypDs from representative species and illustrated that HypD is present across a diverse set of anaerobes. Experimentally, we enriched microbial communities capable of degrading Hyp anaerobically from environmental samples, highlighting the presence of this metabolic activity in habitats distinct from the human gut. The few enrichment cultures that were analyzed encoded copies of *hypD*, supporting the hypothesis that HypD is responsible for the anaerobic Hyp metabolism observed in these samples. Overall, the prevalence of this metabolic capability in both human stool metagenomes and anoxic sediments correlates with the high abundance of Hyp in animal and plant structural proteins.

Despite being an electron acceptor in Stickland fermentation that was discovered several decades ago, the biological function of Hyp metabolism remains largely unknown in most HypD-encoding microbes. Certain Clostridial genomes contain a conserved gene cluster encoding HypD, AE, and P5C reductase, which were characterized in Chapter 2, and would metabolize Hyp to Pro in an overall reductive pathway (Table 4.1, Figure 4.7). In a small number of genomes, a putative

P5C dehydrogenase gene near *hypD* suggests that the product P5C may undergo oxidation instead (Table 4.1, Figure 4.7). Additionally, microbes can convert P5C into several other amino acids (Figure 4.7).²¹⁻²⁴ Interfacing HypD with these pathways could allow microbes to access building blocks for protein synthesis, substrates for fermentation, terminal electron acceptor for respiration, and sources of carbon and nitrogen. Future studies on Hyp utilization pathways in non-Stickland fermenters will likely uncover distinct purposes for the metabolism of this abundant amino acid. Ultimately, a n understanding of these pathways will allow us to better predict metabolic dynamics of nutrient competition within complex microbial communities.

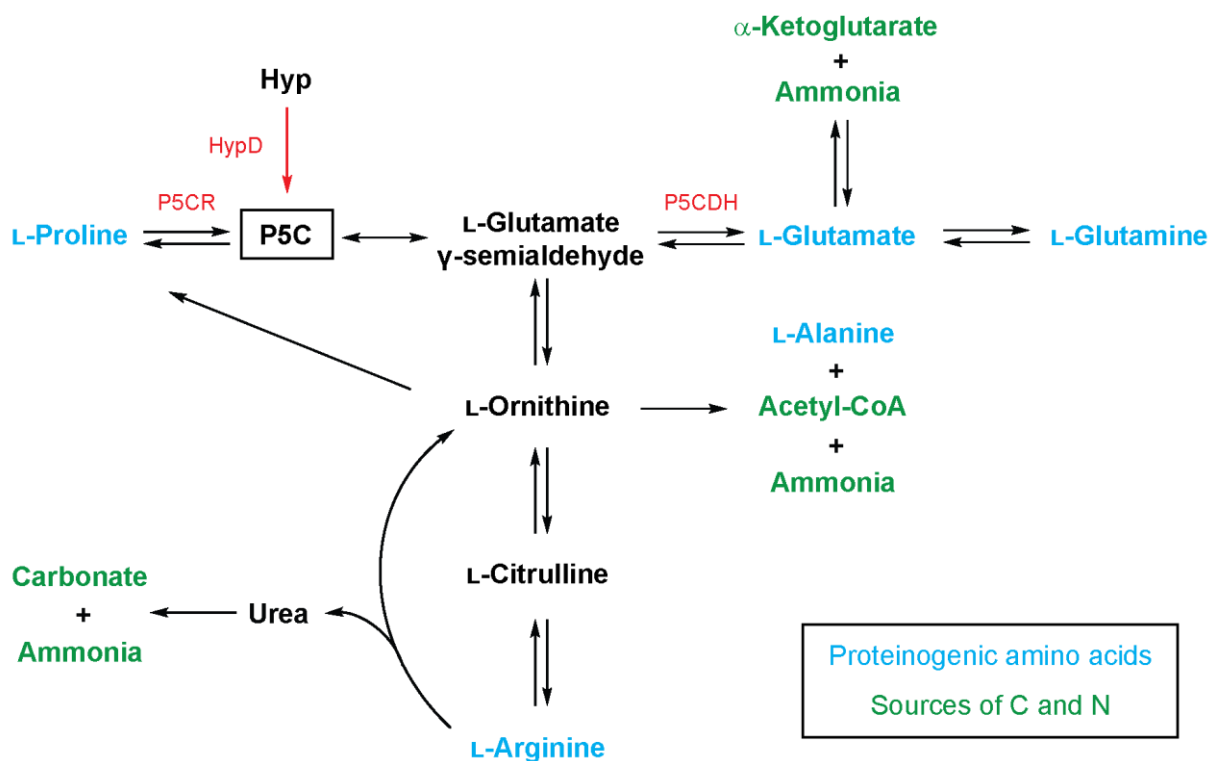


Figure 4.7 Hyp metabolism mediated by HypD interfaces with several amino acid metabolic pathways.

P5C is a central intermediate in amino acid metabolism. The downstream metabolites α -KG, carbonate, acetyl-CoA, and ammonia can serve as sources of carbon and nitrogen. These metabolites and amino acids can be intermediates in energy metabolic pathways as well. HypD, P5CR, and P5C dehydrogenase (P5CDH) are highlighted in red. Some arrows represent multiple steps, and only key metabolites are shown.

4.4 Materials and methods

Bacterial strains were purchased from ATCC and DSMZ. Luria-Bertani (LB) Lenox medium was purchased from EMD Millipore or Alfa Aesar. DNA sequencing results and multiple sequence alignments were analyzed with Geneious Pro 7.1.5 or Clustal Omega.^{20,25} Primers were purchased from Integrated DNA Technologies (Coralville, IA). PCR was performed with a MyCycler gradient cycler (Bio-Rad) or a C1000 Gradient Cycler (Bio-Rad). All PCR amplifications were analyzed by 1% agarose gel electrophoresis with ethidium bromide staining in 1×TAE buffer. PCR products were purified using an Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare). gDNA was purified from cultures using an UltraClean Microbial DNA isolation Kit (MO BIO). TOPO TA Cloning Kit (Invitrogen) was used to construct clone libraries for sequencing. DNA sequencing was carried out by Beckman Coulter Genomics. All restriction enzymes, ligases, polymerases, and PCR mixes were obtained from New England Biolabs. All absorbance measurements in 96-well plates were carried out using a PowerWave HT Microplate Spectrophotometer (Biotek) inside of an anaerobic chamber (MBraun) and were obtained with pathlength corrected to 1 cm. Media were made anaerobic after autoclaving by sparging nitrogen through in a sterile environment.

4.4.1 Construction of a phylogenetic tree with representative HypD sequences

86 full-length HypD amino acid (aa) sequences from the UniProt database were selected to include one sequence per species. All sequences contain the set of putative active site residues identified from a previously generated HypD homology model³ as well as all residues selected for site-directed mutagenesis based on the crystal structure described in Chapter 3. Sequences were aligned with MAFFT using the G-INS-I refinement method⁶ and the alignment was manually

trimmed to yield a final length of 783 aa. The maximum-likelihood tree was constructed in MEGA 7⁵ using the Jones–Taylor–Thorton matrix-based model²⁶. Statistical support was obtained by bootstrapping 100 iterations. Choline trimethylamine-lyase (CutC) and glycerol dehydratase (GD) were included as outgroups on the HypD reference tree.

4.4.2 Enrichment culturing of Hyp-degrading microbes

Samples used for initial enrichment cultures were collected from different wetlands around Woods Hole, MA, USA (

Table 4.2). Trunk River sediments were directly collected in glass jars and sealed with rubber septa. Hyp-containing supplements were added to Trunk River sediment samples (

Table 4.2) and incubated at room temperature for a few days before inoculation into the Hyp enrichment medium (Table 4.3). Remaining wetland samples were collected under aerobic conditions and directly inoculated into the enrichment medium. All enrichment cultures were incubated at 30 °C in sealed vessels in the dark. Turbidity and Hyp degradation were observed in all inoculated cultures before passaging into fresh anaerobic enrichment media. Serial passaging was achieved by diluting growing cultures 1:100 into 5 mL of anaerobic enrichment medium in 18×150 mm Hungate tubes (Chemglass). Hungate tubes were sealed with butyl stoppers and aluminum seals. Four enrichment cultures were serially passaged four times before gDNA extraction and sequencing.

4.4.3 Hyp derivatization and detection by HPLC-RID

Samples were derivatized through the van Slyke reaction with nitrosonium, generated from potassium nitrite and hydrogen chloride using a previously reported protocol.²⁷ 70 µL of 1 M potassium nitrite was added to 350 µL of Hyp standard or samples in 1.7 mL Eppendorf tubes.

Reactions were started by the addition of 14 μL of 12 N HCl and quenched with 28 μL of 5 N NaOH. Derivatization samples were carried out at 45 °C for 90 min. Derivatized solutions were mixed with 5 N H_2SO_4 at a 9:1 ratio prior to injecting on the HPLC. The *N*-nitroso derivative of Hyp was separated on an Aminex HPX-87P column.²⁷ The previously reported method was modified to include an isocratic elution of 5 mM H_2SO_4 at 0.4 mL min⁻¹ at 60 °C for 40 minutes. Detection of derivatized Hyp was achieved using a refractive index detector (RID). Derivatized Hyp eluted at 18 min under these conditions.

4.4.4 Amplification of 16S rRNA gene from genomic DNA of enrichment culturesⁱⁱ

The 16S rRNA gene was PCR amplified from gDNA isolated from enrichment cultures using primers 8F (5'-AGAGTTTGATCCTGGCTCAG-3') and 1510R (5'-ACGGCTACCTTGTTACGACTT-3') (Table 4.4).¹² PCR mixtures contained 1 $\times$ PCR buffer, 0.2 mM dNTPs, 100 ng of DNA template, 2 μM of each primer, and 0.4 U of Takara Taq Hot Start Version (Takara Bio Inc., Otsu, Japan) in a total volume of 50 μL . Thermocycling was carried out with an initial denaturation period of 5 min at 95 °C, followed by 35 cycles of 10 s at 98 °C, 30 s at 50 °C, and 2 min at 72 °C, with a final extension time of 10 min at 72 °C. Corresponding PCR products were gel purified and cloned into pCR4-TOPO vectors using the TOPO TA Cloning Kit to construct a clone library for sequencing. Sequencing reads for each vector were aligned using ClustalW to assemble full-length 16S rDNA sequences. Only high-quality reads were used for alignments and subsequent analyses.

ⁱⁱ PCR amplification of 16S rRNA gene from enrichment culture gDNA and construction of a clone library were conducted by Dr. Ana Martínez-del Campo.

4.4.5 Construction of a phylogenetic tree with 16S rRNA genes amplified from enrichment cultures

Chimeric 16S rRNA gene amplification products were detected using Uchime2²⁸ and excluded from analyses. Redundant sequences within each enrichment culture were also removed. 28 sequences from enrichment cultures and 12 reference sequences were aligned with MUSCLE.¹⁸ The alignment was manually trimmed to yield a final length of 1371 bp and a maximum likelihood tree was constructed in MEGA7⁵ using the Tamura–Nei method of substitution.²⁹ Statistical support was obtained by bootstrapping 100 iterations. The highest similarity score calculated by RDP SeqMatch was used to assign the lowest classification level for each enrichment isolate.¹⁹ RDP SeqMatch analysis was carried out on 2017-09. The full-length sequences obtained in this analysis have been deposited in GenBank (accession: MG367094–MG367121).

4.4.6 Design of degenerate primers for *hypD* genesⁱⁱⁱ

A multiple sequence alignment of 152 HypD sequences and three different GREs (glycerol dehydratase, propanediol dehydratase, and choline trimethylamine-lyase) was generated using ClustalW. Based on this alignment, degenerate primers were designed to target amino acid stretches with high conservation among HypD sequences and low identity to the other GREs. Fourteen degenerate primer pairs that amplify different fragments of *hypD* were selected and screened using gDNA isolated from strains that carry *hypD* in their genomes (*Bacteroides dorei* DSM 17885, *Parabacteroides johnsonii* DSM 18315, *Terrisporobacter glycolicus* DSM 1288, *Clostridioides difficile* 630Δerm and *Clostridium sporogenes* ATCC 15579) and strains that lack

ⁱⁱⁱ Degenerate primers were designed by Dr. Ana Martínez-del Campo. Degenerate PCR conditions were optimized by Dr. Ana Martínez-del Campo.

hypD but encode other GREs (*Clostridium hathewayi* DSM 13479 and *Clostridium butyricum* ATCC 860) as negative controls. Primer pair HypD F149aa (5'-TTYACIGARTTYATGGARCA-3') and HypD R261aa (5'-TGNACRAACCARTACATYTG-3') was chosen for further experiments since it showed good amplification with all *hypD* containing-strains and no amplification with strains that lack *hypD* (Table 4.4). These primers amplify a 357 bp region of the *hypD* gene that include two conserved residues predicted to be in the active site based on our previously published homology model.³

PCRs on gDNA isolated from single strains contained 1× PCR buffer, 0.2 mM dNTPs, 50 ng of DNA template, 2 µM of each primer, and 0.4 U of Takara Taq Hot Start Version (Takara Bio Inc., Otsu, Japan) in a total volume of 50 µL. Thermocycling was carried out with an initial denaturation period of 5 min at 95 °C, followed by 35 cycles of 30 s at 95 °C, 30 s at 50 °C, and 60 s at 72 °C, with a final extension time of 10 min at 72 °C. PCR reactions were analyzed by agarose gel electrophoresis with ethidium bromide staining.

4.4.7 Degenerate PCR and sequencing of partial *hypD* fragments from enrichment cultures^{iv}

A clone library for *hypD* was constructed using degenerate PCR with gDNA isolated from enrichment cultures passaged four times and the degenerate primers HypD F149aa and HypD R261aa (Table 4.4). The PCR and thermocycling conditions used are as described above, except that 100 ng of template DNA was used. These reactions were first examined by agarose gel electrophoresis with ethidium bromide staining. Next, DNA bands of approximately 357 bp were

^{iv} Degenerate PCR was conducted by Dr. Ana Martínez-del Campo. The clone library and sequencing were conducted with help from Dr. Ana Martínez-del Campo.

excised and purified by gel extraction. 1 μ L of each PCR product was cloned into the pCR4-TOPO vector using the TOPO TA cloning kit for sequencing (Invitrogen). The resulting plasmids were transformed into chemically competent *E. coli* TOP10 cells. 24 clones from each enrichment culture were randomly selected for sequencing. These clones were sent to Beckman Coulter Genomics (Danvers, MA) for plasmid purification and sequencing. Low-quality, unsuccessful, and no-insert sequencing results were discarded. High quality sequences that are non-redundant within each culture were deposited on GenBank (accession: MG367122–MG367168).

4.5 References

1. YY Huang, A Martinez-Del Campo, EP Balskus. Anaerobic 4-hydroxyproline utilization: Discovery of a new glycyl radical enzyme in the human gut microbiome uncovers a widespread microbial metabolic activity. *Gut microbes*. **2018**, 9(5):437-51.
2. SF Altschul, W Gish, W Miller, EW Myers, DJ Lipman. Basic local alignment search tool. *J Mol Biol*. **1990**, 215(3):403-10.
3. BJ Levin*, YY Huang*, SC Peck, Y Wei, A Martínez-del Campo, JA Marks, EA Franzosa, C Huttenhower, EP Balskus. A prominent glycyl radical enzyme in human gut microbiomes metabolizes *trans*-4-hydroxy-L-proline. *Science*. **2017**, 355(6325):eaai8386.
4. T UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Research*. **2018**, 46(5):2699.
5. S Kumar, G Stecher, K Tamura. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol*. **2016**, 33(7):1870-4.
6. K Katoh, J Rozewicki, KD Yamada. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform*. **2017**, bbx108.
7. EA Smith, GT Macfarlane. Dissimilatory amino acid metabolism in human colonic bacteria. *Anaerobe*. **1997**, 3(5):327-37.
8. SE Mergenhagen, HW Scherp. Lysis of reconstituted collagen and catabolism of products of collagenolysis by the oral microbiota. *Arch Oral Biol*. **1960**, 1:333-8.
9. LA Hug, BJ Baker, K Anantharaman, CT Brown, AJ Probst, CJ Castelle, CN Butterfield, AW Hernsdorf, Y Amano, K Ise, *et al*. A new view of the tree of life. *Nature Microbiology*. **2016**, 1:16048.
10. K Zaremba-Niedzwiedzka, EF Caceres, JH Saw, D Bäckström, L Juzokaite, E Vancaester, KW Seitz, K Anantharaman, P Starnawski, KU Kjeldsen, *et al*. Asgard archaea illuminate the origin of eukaryotic cellular complexity. *Nature*. **2017**, 541:353-8.
11. A Crits-Christoph, S Diamond, CN Butterfield, BC Thomas, JF Banfield. Novel soil bacteria possess diverse genes for secondary metabolite biosynthesis. *Nature*. **2018**, 558:440-4.
12. DJ Lane. 16S/23S rRNA sequencing. In: Stackebrandt E, Goodfellow M, editors. Nucleic acid techniques in bacterial systematics. Chichester, United Kingdom: John Wiley and Sons; 1991. p. 115-75.
13. NR Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. **2016**, 44:D7-D19.

14. E Pelletier, A Kreimeyer, S Bocs, Z Rouy, G Gyapay, R Chouari, D Rivière, A Ganesan, P Daegelen, A Sghir, *et al.* “*Candidatus* Cloacamonas Acidaminovorans”: Genome sequence reconstruction provides a first glimpse of a new bacterial division. *J Bacteriol.* **2008**, 190(7):2572-9.
15. H Zheng, C Dietrich, R Radek, A Brune. *Endomicrobium proavitum*, the first isolate of Endomicrobia class. nov. (phylum Elusimicrobia) – an ultramicrobacterium with an unusual cell cycle that fixes nitrogen with a Group IV nitrogenase. *Environ Microbiol.* **2016**, 18(1):191-204.
16. SW Ragsdale, E Pierce. Acetogenesis and the Wood–Ljungdahl pathway of CO₂ fixation. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics.* **2008**, 1784(12):1873-98.
17. TA Hansen. Metabolism of sulfate-reducing prokaryotes. *Antonie van Leeuwenhoek.* **1994**, 66(1):165-85.
18. RC Edgar. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, 32(5):1792-7.
19. JR Cole, Q Wang, JA Fish, B Chai, DM McGarrell, Y Sun, CT Brown, A Porras-Alfaro, CR Kuske, JM Tiedje. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Res.* **2014**, 42(D1):D633-D42.
20. F Sievers, A Wilm, D Dineen, TJ Gibson, K Karplus, W Li, R Lopez, H McWilliam, M Remmert, J Söding, *et al.* Fast, scalable generation of high - quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol.* **2011**, 7(539).
21. S Watanabe, D Morimoto, F Fukumori, H Shinomiya, H Nishiwaki, M Kawano-Kawada, Y Sasai, Y Tozawa, Y Watanabe. Identification and characterization of D-hydroxyproline dehydrogenase and Δ^1 -pyrroline-4-hydroxy-2-carboxylate deaminase involved in novel L-hydroxyproline metabolism of bacteria. *J Biol Chem.* **2012**, 287(39):32674-88.
22. E Adams, L Frank. Metabolism of proline and the hydroxyprolines. *Annu Rev Biochem.* **1980**, 49(1):1005-61.
23. JR Brown, WF Doolittle. Archaea and the prokaryote-to-eukaryote transition. *Microbiol Mol Biol Rev.* **1997**, 61(4):456-502.
24. M Falb, K Müller, L Königsmaier, T Oberwinkler, P Horn, S von Gronau, O Gonzalez, F Pfeiffer, E Bornberg-Bauer, D Oesterhelt. Metabolism of halophilic archaea. *Extremophiles.* **2008**, 12(2):177-96.
25. M Kearse, R Moir, A Wilson, S Stones-Havas, M Cheung, S Sturrock, S Buxton, A Cooper, S Markowitz, C Duran, *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics.* **2012**, 28(12):1647-9.

26. DT Jones, WR Taylor, JM Thornton. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics*. **1992**, 8(3):275-82.
27. D Pleissner, R Wimmer, NT Eriksen. Quantification of amino acids in fermentation media by isocratic HPLC analysis of their α -hydroxy acid derivatives. *Analytical Chemistry*. **2011**, 83(1):175-81.
28. R Edgar. UCHIME2: improved chimera prediction for amplicon sequencing. *bioRxiv*. **2016**,
29. K Tamura, M Nei. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol*. **1993**, 10(3):512-26.

5 Chapter 5. Toward an understanding of the physiological roles of anaerobic 4-hydroxyproline metabolism in the gut microbiota

This chapter consists of mostly unpublished, preliminary work that might serve as references for future studies.

5.1 Introduction

In Chapter 2 and 3, we described in detail biochemical and structural characterization of HypD from *Clostridioides difficile* 70-100-2010. The discovery of this enzyme illuminated the only known pathway for anaerobic Hyp metabolism. Predicted HypDs were found to be widely distributed in stool metagenomes, sequenced genomes, and environmental samples as described in Chapter 2 and 4. However, many questions remain regarding the physiological activity and role of HypD in different microbial phylogenetic groups. Here in this Chapter, we present results obtained from experiments attempting to resolve some of these questions.

Topics covered in this Chapter can be broadly categorized into two themes connected by the primary goal of studying metabolic pathways relevant for anaerobic Hyp metabolism in gut microbes. First, we examined whether Hyp dehydration catalyzed by HypD is a transformation upstream of the previously known Pro reduction pathway. Stickland fermentation of Hyp and Pro were studied using culture-based experiments and genetic knockouts in *C. difficile*. Comparative transcriptomics was conducted through RNA-sequencing of *C. difficile* cultures grown in the presence of either Pro or Hyp as the principle source of electron acceptor as well as the rich medium Reinforced Clostridial Medium (RCM). We compared transcript counts among the three media conditions and calculated overall similarities. From sequence read counts, we highlight pathways that are differentially regulated by Hyp and Pro.

The second and final section in this Chapter focuses on Hyp metabolism in gut species within the order Bacteroidales. These species are not Stickland fermenters and are common members of the human gut microbiome, which prompted us to examine the role of HypD in these species phylogenetically distant from Clostridiales. We performed *in vitro* biochemical assays to examine the activity of a putative HypD from the common gut microbe *Bacteroides vulgatus* ATCC 8482. In addition, we attempted to detect Hyp metabolism in several HypD-encoding *Bacteroides* and *Parabacteroides* species through culturing and Hyp quantification. Deletion mutants and expression strains in *Bacteroides* species were generated to study the physiological role of Hyp metabolism in these organisms. We endeavored to detect changes in host colonization through a competition experiment in germ-free mice with *B. vulgatus* WT and $\Delta hypD$ strains. However, these studies aimed to examine Hyp metabolism in non-Stickland fermenting microbes failed to clearly identify a physiological role for this transformation. We discuss in detail possible explanations for these observations and propose experiments that could address these issues.

Within each section, additional assays are proposed to advance more in-depth investigations. Collectively, many plasmids and mutant strains were generated from the work reported in this Chapter. We hope these resources will enable future studies on the topics presented here. In the conclusion section, we discuss potential ways by which anaerobic Hyp metabolism could impact the gut microbiome composition and host physiology.

5.2 Results and discussions

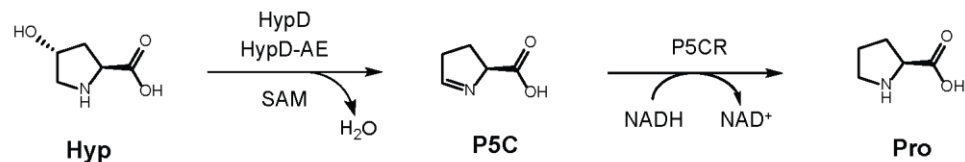
5.2.1 The role of HypD in Stickland fermentation by Clostridiales

The most apparent role of Hyp metabolism in microbial physiology is the conversion of Hyp to Pro catalyzed by the *hypD* gene cluster initially discovered in Clostridiales for use in

Stickland fermentation (Figure 5.1). We previously demonstrated that the enzymes encoded by this gene cluster, HypD, HypD-AE, and P5CR from *C. difficile* 70-100-2010, perform this sequence of transformations.¹ As described in Chapter 1, anaerobic Hyp metabolism was first observed as part of Stickland fermentation in *Clostridium sporogenes*.² Having uncovered the molecular basis for Hyp fermentation, we revealed that this reductive pathway generates one equivalent of NAD⁺ from reduction of P5C to Pro by P5CR. Some Clostridiales encode a second gene annotated as *proC* in genome regions distant to the *proC* gene near *hypD*. The presence of two *proC* genes in these Clostridiales supports distinct functions and regulation of the two P5CRs, with one being important for Stickland reduction of Hyp. The role of the second P5CR, however, cannot be readily inferred by genomic context as most adjacent genes are annotated as hypothetical proteins. Notably, ~65% of the Clostridiales species that possess the *prd* operon, which catalyzes Stickland fermentation of Pro (described in Chapter 1),^{3,4} also encode HypD (Figure 5.1). However, the *prd* operon and the *hypD* gene cluster are located in distant genomic regions, which prevented prior identification of Hyp-metabolizing enzymes.

Based on these observations, we propose that the Pro synthesized from Hyp is further reduced to 5-aminovalerate as part of Stickland fermentation. This is corroborated by a previous observation that Prd protein levels increased in the presence of Hyp, presumably from an accumulation of Pro.³ HypD is also present in species previously demonstrated to use both Hyp and Pro as electron acceptors (*C. difficile*, *C. sporogenes*).^{2,3} Therefore, we sought to test if the Stickland fermentation pathway for Hyp is upstream of the pathway for Pro reduction in two Clostridiales species.

Stickland fermentation of Hyp catalyzed HypD



Stickland fermentation of Pro catalyzed by the *prd* operon

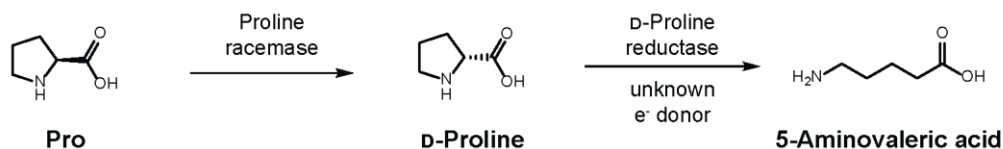


Figure 5.1 Stickland fermentation pathways for Hyp and Pro.

Hyp is metabolized to Pro by the concerted actions of HypD, HypD-AE, and P5CR.

The resulting Pro produce can be further reduced by enzymes encoded within the *prd* operon to 5-aminovalerate, an end product unique to Pro fermentation in anaerobic metabolism.

We selected two strains encoding both the three-gene cluster containing *hypD* and the *prd* operon (*C. difficile* 630Δ*erm* and *Terrisporobacter glycolicus* DSM 1288) for this experiment. Concerted actions of enzymes encoded in the *hypD* gene cluster would convert Hyp to Pro, which could then be metabolized to 5-aminovalerate by Pro racemase and reductase enzymes (Figure 5.1). Indeed, Pro fermentation in *C. difficile* had been previously observed to result in a 5-aminovalerate accumulation.⁵ We sought out to test if Hyp metabolism would also result in the production of 5-aminovalerate. Metabolites Hyp, Pro, and 5-aminovalerate were quantified at multiple time points during growth of these two strains in defined media supplemented with Hyp or Pro (Figure 5.2). For both strains grown in MACC-Pro, Pro levels decreased over time while 5-aminovalerate levels increased. This result had been previously observed for anaerobic

fermentation of Pro.^{3,5} Similarly, in MACC-Hyp cultures, Hyp was consumed while both Pro and 5-aminovalerate levels increased correspondingly (Figure 5.2). This result strongly suggested that Pro generated from Hyp metabolism is further reduced to 5-aminovalerate and that the HypD-catalyzed pathway is upstream of Pro reduction in *C. difficile*. In some species, 5-aminovalerate can be further metabolized for energy production as described in Chapter 1.^{6,7} The two Clostridial strains used in our study likely do not encode enzymes for 5-aminovalerate metabolism since this metabolite accumulated in the cultures.

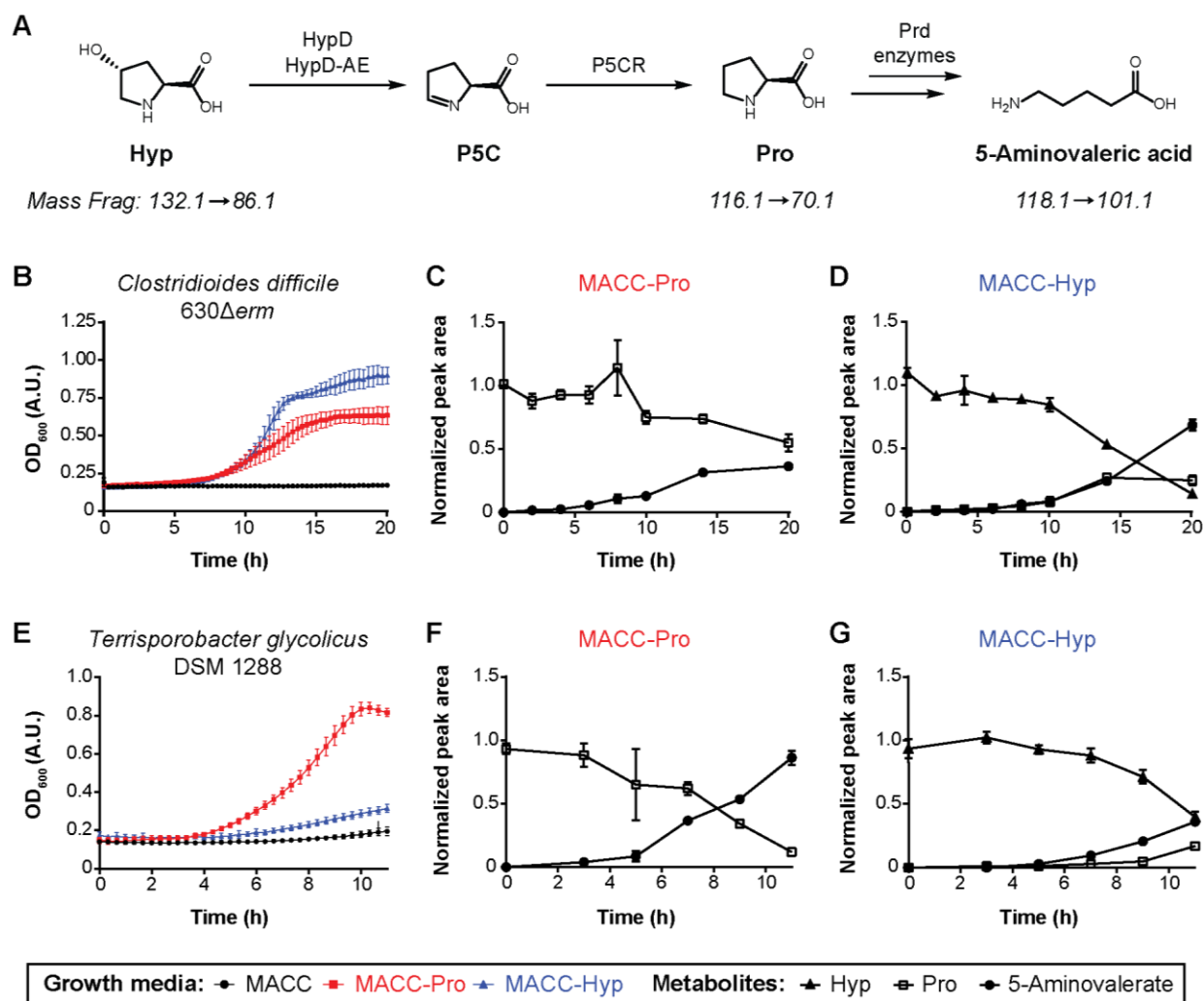


Figure 5.2 Growth of two HypD-encoding Clostridiales and metabolite analysis.

(A) In Clostridiales, Hyp is first metabolized to Pro via HypD and P5CR, and then reduced to 5-aminovalerate by the reductive Stickland pathway involving Pro epimerase and Pro reductase. Hyp, Pro, and 5-aminovalerate concentrations were monitored over time using LC-MS/MS. (B, E) Enhanced growth was observed for both *C. difficile* 630 Δ erm and *T. glycolicus* DSM 1288 upon supplementation of the basal medium with either Pro or Hyp (20 mM final concentration). Downstream metabolites were quantified at various timepoints in cultures supplemented with Pro

Figure 5.2 (Continued)

(C, F) or Hyp (D, G). Sample peak areas are normalized to the corresponding standard. Data points are shown as mean $\pm$ standard deviation (SD) with $n = 5$.

5.2.2 Generation of deletion mutants in *C. difficile* 630 Δ erm

In recent publications, Stickland fermentation was found to be an important pathway for *C. difficile* colonization in mouse models.^{8,9} Notably, both Pro and Hyp were identified as key nutrients consumed by *C. difficile* in these experiments, suggesting that these alternative electron acceptors are important for successful colonization.^{8,9} We sought out to test this hypothesis by generating single gene deletions in the *hypD*-encoding gene cluster using the *pyrE*-based allelic exchange system for *C. difficile*.¹⁰ By constructing a Δ *hypD* mutant, we will be able to establish whether HypD is solely responsible for anaerobic Hyp metabolism in *C. difficile* in culture-based experiments. Furthermore, genetic knockouts of this species will enable the study of Hyp utilization by *C. difficile* in the context of competition within microbial communities and host colonization.

Vectors were constructed for clean deletions of *hypD* (locus tag: CD630_32820) and *proC* (locus tag: CD630_32810) in the parental strain *C. difficile* 630 Δ erm Δ *pyrE* (Table 5.1).ⁱ The low toxigenic 630 Δ erm strain was selected over the hypervirulent R20291 strain for its ability to colonize mice without immediately causing lethal infections. Although 630 Δ erm is the model strain used to study factors involved in host colonization, its low toxin production provides a limitation to examine changes in toxin expression. In collaboration with Prof. Aimee Shen at Tufts

ⁱ All plasmids and strains used for *C. difficile* genetics were provided by Prof. Aimee Shen at Tufts University.

University, two mutant strains, $\Delta hypD$ and $\Delta proC$, were successfully constructed using these vectors (Table 5.1).ⁱⁱ Prof. Aimee Shen complemented and restored the *pyrE* locus in these deletion mutants to obtain strains 630 $\Delta erm\Delta hypD$ and 630 $\Delta erm\Delta proC$. Full-length *hypD* and *proC* genes, including upstream regions encoding predicted promoters were each cloned into pMTL-YN1C for complementation into the corresponding deletion mutant. In summary, we have constructed two *C. difficile* 630 Δerm deletion mutants as well as generated vectors to complement the gene back into each deletion mutant. Future work could involve the construction of a second set of vectors to include a transcriptional terminator upstream of the complemented gene to prevent transcriptional readthroughs during *pyrE* transcription. This would reduce differences in transcriptional regulation between the complemented gene and the native gene in WT strain. Generation of complemented strains will be crucial for demonstrating the sufficiency of each gene to recover activity in *C. difficile*.

ⁱⁱ *C. difficile* deletion mutants were generated together with Prof. Aimee Shen. Complementation of *pyrE* gene into *C. difficile* 630 $\Delta erm\Delta pyrE$ deletion mutants was conducted by Prof. Aimee Shen.

Table 5.1 Strains and plasmids used in studies on *C. difficile*.

Strain or plasmid	Description or features	Reference /source
<i>E. coli</i> strains		
TOP10	Cloning and plasmid storage	
HB101/pRK24	Conjugation with Bacteroides species	Shen lab
<i>C. difficile</i> strains		
630 Δ erm Δ pyrE	A derivative of 630 Δ erm strain with a deletion of <i>pyrE</i> , uracil protrophic	10
630 Δ erm Δ pyrE Δ hypD	Deletion in <i>pyrE</i> and the HypD gene, CD630_32820	This study
630 Δ erm Δ pyrE Δ proC	Deletion in <i>pyrE</i> and the P5CR gene, CD630_32810	This study
630 Δ erm	Parent strain with same genomic background as deletion mutants	
630 Δ erm Δ hypD	Deletion in HypD gene, CD630_32820	This study
630 Δ erm Δ proC	Deletion in P5CR gene, CD630_32810	This study
Plasmids		
pMTL-YN3	Deletion vector for allelic exchange in <i>C. difficile</i> 630 Δ erm Δ pyrE	10
pMTL-YN3- Δ HypD	Carrying ~1 kb flanking regions around <i>hypD</i> , CD630_32820	This study
pMTL-YN3- Δ P5CR	Carrying ~1 kb flanking regions around <i>proC</i> , CD630_32810	This study
pMTL-YN1C	Complementation vector for restoration of Δ pyrE locus to generate <i>C. difficile</i> 630 Δ erm	10
pMTL-YN1C-HypD	Carrying ~330 promoter region + HypD gene	This study
pMTL-YN1C-P5CR	Carrying ~230 promoter region + P5CR gene	This study

5.2.3 Phenotypes of *C. difficile* 630 Δ erm Δ hypD and 630 Δ erm Δ proC strains

Growth of *C. difficile* 630 Δ erm WT was first compared to growth of the two mutant strains in BHI, a medium rich in protein digests as sources of Hyp (Figure 5.3). Although some variations in the length of lag phase were observed between strains, they all reached similar maximum optical densities during stationary phase. The small discrepancies in the lag period between inoculation and growth are likely caused by differences in growth stage of bacterial cells between inocula.

Overall, deletion of *hypD* or *proC* in *C. difficile* 630 Δ *erm* did not impact growth in a nutrient-rich environment, which was expected given the abundance of other amino acids and nutrients available to support bacterial growth in BHI.

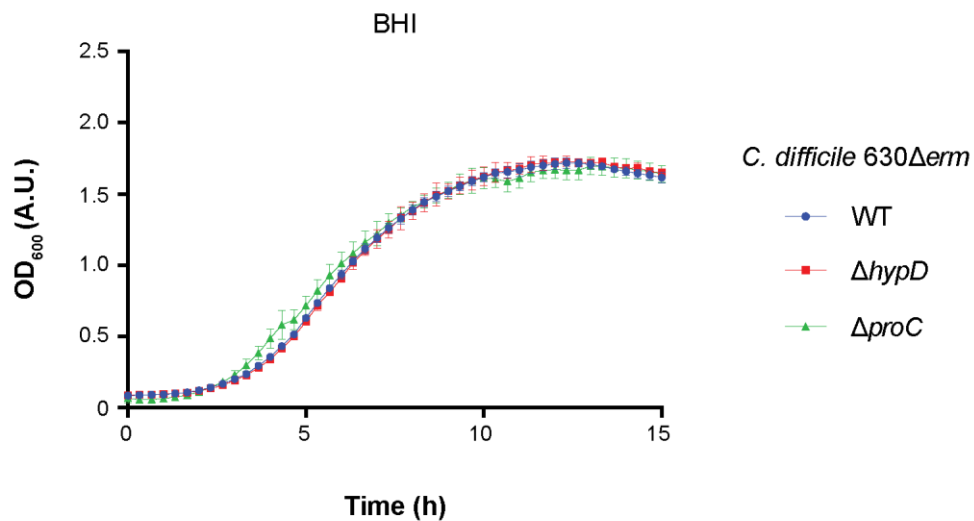


Figure 5.3 Growth of *C. difficile* 630 Δ *erm* WT, Δ *hypD*, and Δ *proC* mutants in a rich medium.

WT and deletion mutants of *C. difficile* 630 Δ *erm* were grown in BHI. Data from a representative growth experiment are displayed where similar growth was observed among all three strains. Data are shown as mean $\pm$ SD with $n = 3$ replicates and normalized to media controls

Next, we compared the growth of these strains in the minimal medium MACC supplemented with 20 mM Pro or 20 mM Hyp as described in Chapter 2. Hyp, Pro, and 5-aminovalerate were quantified at the end of each experiment to determine if the Δ *hypD* and Δ *proC* mutants are capable of metabolizing Hyp. Growth experiments were replicated four times to obtain

preliminary data on growth and amino acid metabolism. In MACC-Pro, some variations were observed for the length of lag phase, but all three strains consistently metabolized Pro to 5-aminovalerate completely. In MACC-Hyp, overall trends supported reduced growth of both mutants compared to WT (Figure 5.4A). As expected, Hyp was not metabolized by the $\Delta hypD$ mutant, supporting the growth data (Figure 5.4B). Interestingly, this mutant did exhibit some growth on MACC-Hyp without being able to access the NAD^+ generated from Hyp metabolism. Initially, this result seemed to contradict the previously observed lack of growth of the WT strain in MACC (Figure 5.1). However, low levels of Pro were detected by LC-MS/MS in MACC-Hyp media blanks and are likely from impurities in the commercial supply of Hyp. This small amount of Pro might have supported $\Delta hypD$ mutant's growth in MACC-Hyp. This hypothesis can be tested by growing the WT strain in MACC supplemented with similar low levels of Pro. In summary, we demonstrated here that HypD is essential for Hyp metabolism in *C. difficile* and is important for growth in a nutrient-limiting condition.

Metabolism of Hyp to 5-aminovalerate was observed in the $\Delta proC$ mutant, but to a lesser degree compared to WT strain in most experiments (Figure 5.4B). It is important to note that *C. difficile* 630 Δerm encodes a second putative *proC* gene (locus tag: CD630_14950) outside of the *hypD* gene cluster. Prior to conducting this experiment, we wondered whether this second P5CR would be expressed under the conditions tested and thus complement the *proC* deletion. Our data suggest that the expression and/or activity of this second P5CR differ from the P5CR encoded by the deleted *proC* gene, resulting in reduced growth and Hyp metabolism in the $\Delta proC$ mutant. Quantification of these metabolites at multiple time points throughout growth will more precisely define the metabolic differences between WT and $\Delta proC$ strains. For example, we might observe a decrease in Hyp without a corresponding increase in Pro or 5-aminovalerate until a later

time point, which would indicate a lag time in the expression of the second P5CR. Notably, genes in the neighborhood of the second *proC* gene are mostly annotated as hypothetical proteins and transcriptional regulators. The lack of putative functions related to amino acid transformations among these genes suggests that this second *proC* gene may not encode a true P5CR. Therefore, the possibility remains that a promiscuous enzyme reduced P5C to Pro overtime likely at a slower rate.

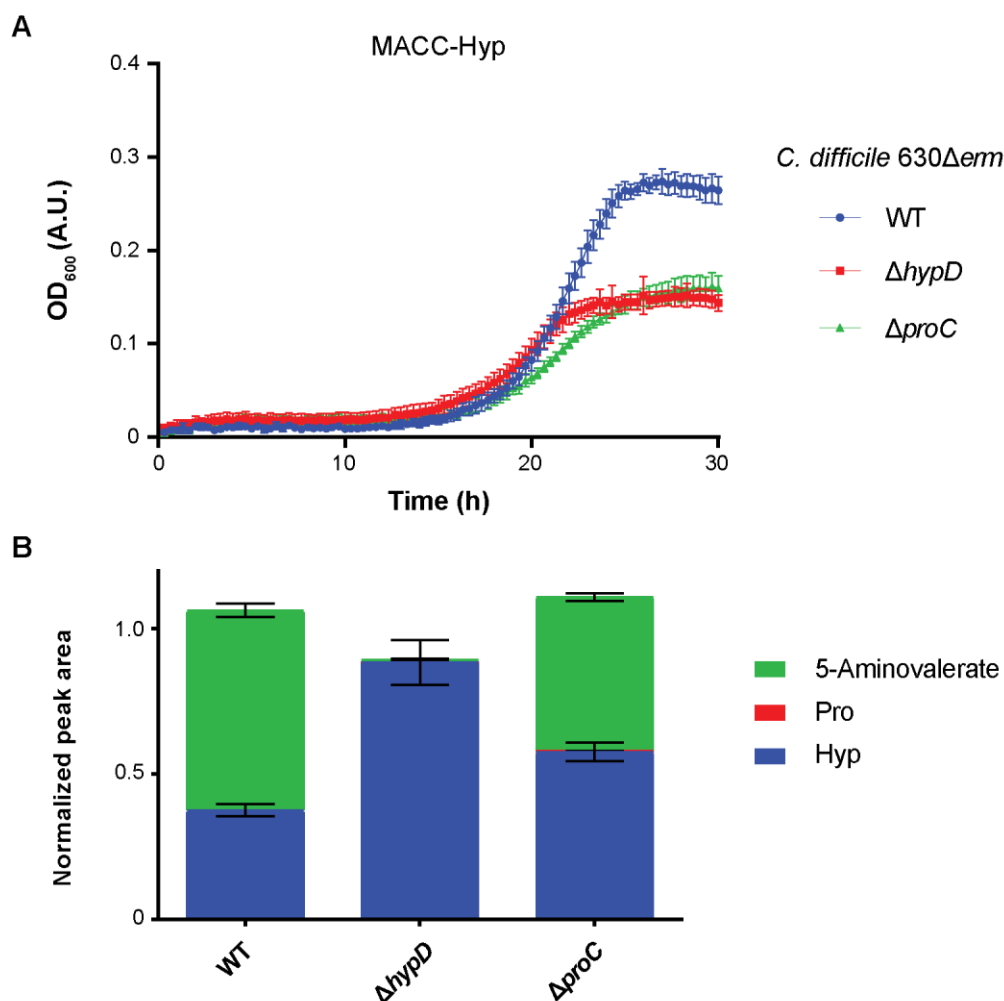


Figure 5.4 Growth of *C. difficile* 630Δerm WT, ΔhypD, and ΔproC mutants in a minimal medium.

(A) WT and deletion mutants of *C. difficile* 630Δerm were grown in the previously described MACC-Hyp medium. Data from a representative growth experiment are displayed here, in which the WT strain grew to a higher maximum optical density than both deletion mutants. Data are shown as mean ± SD with n = 5 replicates and normalized to media controls. (B) Metabolites Hyp, Pro, and 5-aminovalerate were quantified in spent media at the end of growth. For each metabolite, peak areas

Figure 5.4 (Continued)

obtained from cultures were normalized to 20 mM standards added to the basal medium MACC. Quantification was done by LC-MS/MS.

It is important to note that in these growth experiments, *C. difficile* 630 Δ *erm* did not reach similar maximum optical densities (0.6–0.9) to those previously observed using the same assay setup (Figure 5.2). The poor growth observed in MACC-Hyp was likely due to low levels of oxygen observed in the anaerobic chamber when these experiments were conducted. These growth experiments will need to be repeated along with mutants with the deleted gene complemented back to demonstrate that *hypD* is necessary and sufficient for Hyp metabolism.

It will be interesting to use these genetic mutants in animal studies to examine the importance of Hyp metabolism in the context of host colonization. A double deletion of the genes encoding HypD and Pro reductase might be required to detect a significant fitness defect in *C. difficile*. Previously studies had highlighted both Hyp and Pro as nutrients that were significantly metabolized during pathogenic colonization and expansion in mouse models.^{8,9} Therefore, access to Pro as an electron acceptor might be sufficient to promote colonization by the *C. difficile* 630 Δ *erm* Δ *hypD* mutant. The contribution of these two connected pathways toward host colonization is unknown and therefore studying single deletion mutants such as Δ *hypD* and Δ *prdB* alongside the double mutant Δ *hypD* Δ *prdB* will be critical.

5.2.4 Comparative transcriptomics of *C. difficile* 630 Δ *erm* during growth on Hyp, Pro, and a rich medium using RNA-seq

The *prd* operon and the *hypD* gene cluster are not colocalized in *C. difficile*, suggesting distinct transcriptional regulation of the encoded genes. Since *C. difficile* can utilize both Hyp and

Pro as electron acceptors, we employed RNA-sequencing (RNA-seq) to study the differences in gene transcription during growth on each substrate. In particular, we wondered if Stickland fermentation of Hyp differentially impacts regulation of other metabolic pathways in comparison to Pro. *C. difficile* 630 Δ *erm* was cultured in MACC medium (supplemented with either Pro or Hyp) and the rich medium RCM, encompassing three growth conditions (Figure 5.5). MACC without any supplement was not included due to a lack of growth previously observed for *C. difficile* 630 Δ *erm* in this experimental set up. Interestingly, addition of 20 mM glycine, an alternative electron acceptor utilized by *C. difficile*, did not support growth in MACC. Identification of other nutrients that could enable growth in this basal medium would provide a better control for comparative transcriptomics in future studies. Cultures were grown to exponential phase (average of ODs ~0.4), and cell pellets were then harvested and submitted for sample preparation and sequencing at the Broad Institute Microbial Omics Core (MOC). Briefly, total RNA was extracted from which gDNA and rRNA were depleted. A cDNA library was constructed using a modified version of RNAtag-seq¹¹ and sequenced using an Illumina HiSeq 2500 system. Sequencing reads were aligned to the reference genome of *C. difficile* CD196 (NCBI: NC_013315.1) using Bowtie2¹² and read counts were quantified using R package Rsubread¹³.ⁱⁱⁱ Total read counts were used for differential expression analysis using DESeq2¹⁴ through the R programming language (version 3.5.1).^{15,16}

ⁱⁱⁱ Alignment, mapping, and counting of sequencing reads to reference genome were conducted by the Broad Institute MOC.

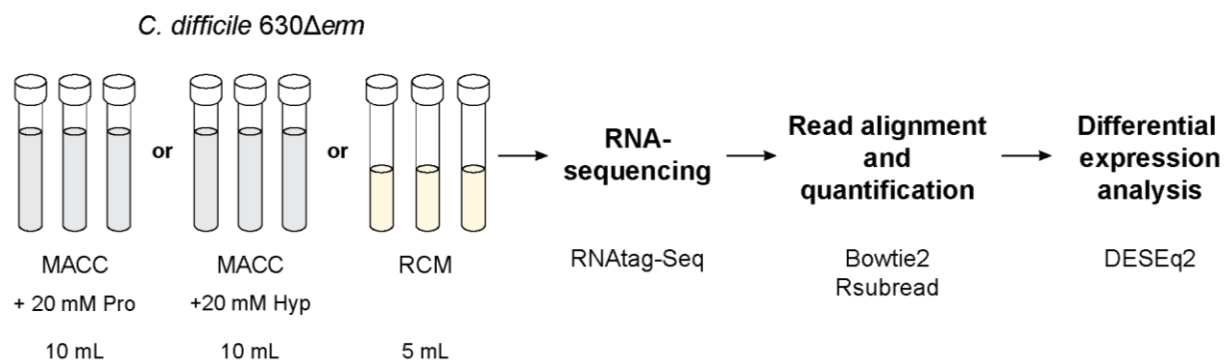


Figure 5.5 RNA-seq experimental set up.

C. difficile 630 Δ *erm* was grown in MACC supplemented with 20 mM Pro or 20 mM Hyp and RCM. Three cultures for each media condition were grown to exponential phase. Total RNA was extracted from cell pellets and sequenced using a modified version of the RNAtag-Seq.¹¹ Reads were aligned to the *C. difficile* reference genome (NC_013315.1) using Bowtie2¹² and quantified using Rsubread.¹³ Differential expression analysis was performed using the R package DESeq2.¹⁷

As a first step, we assessed the overall similarities among all samples to see if gene expression was more similar between which media conditions. For this analysis, read counts were first normalized to account for differences in sequencing depth between samples. A variance stabilization transformation (VST)¹⁴ was calculated using the dataset containing raw read counts. Normalized count values were then transformed to yield VST-transformed values to obtain more similar variances along the entire range of mean values (approximately homoskedastic). Euclidean distances between samples were calculated from VST-transformed counts and hierarchical clustering was performed using the complete-linkage function (Figure 5.6).¹⁸ As expected, cultures grown under identical conditions clustered together making up three clades representing each media type. Notably, MACC-Pro and MACC-Hyp samples have shorter distance scores and form

a clade of their own separate from RCM samples. This is consistent with the identical composition of the basal medium used for MACC-Pro and MACC-Hyp, while RCM is extremely different and consists of complex substrates (beef extract, peptone, and yeast extract). Interestingly, among the MACC samples, those grown in MACC-Pro were slightly more similar to RCM samples as determined by the higher distance scores. Therefore, the overall transcriptomes of *C. difficile* grown on Pro rather than on Hyp more closely resembled those obtained for the nutrient-rich environment in RCM.

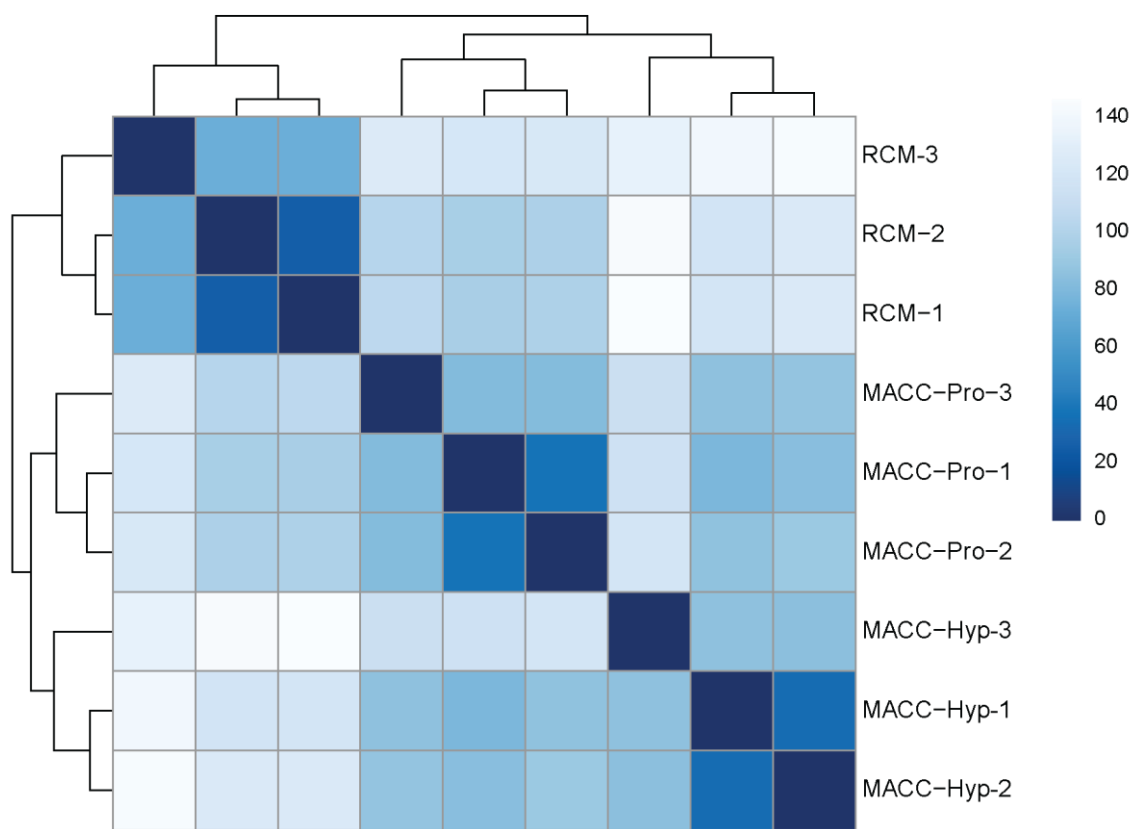


Figure 5.6 Hierarchical clustering of Euclidean distances for all samples.

Raw read counts were normalized and transformed by VST. Euclidean distances were computed between all samples using VST-transformed values to assess overall similarities. Hierarchical clustering by distances is displayed as a dendrogram. A color scale for the calculated distance values between samples is shown. The heatmap was generated using R packages pheatmap¹⁸ and RColorBrewer¹⁹.

In addition to sample clustering by Euclidean distances as a proxy for overall similarities, we also performed a similar analysis for a subset of genes with high variances. The variance among all samples for each gene was calculated from the VST-transformed count values.²⁰ Genes with the 50 highest variances were included for the following hierarchical clustering analysis. Briefly,

the deviation of each gene from its mean was computed for each sample and visualized on a heatmap (Figure 5.7). Using these deviation values, Euclidean distances were calculated between conditions as described above to generate the previous heatmap. Euclidean distances were also calculated between all 50 genes to generate a second dendrogram for gene deviations.

Intriguingly, clustering based on this subset of genes and their deviation values resulted in a split of the MACC conditions into separate clades in the conditions dendrogram. RCM and MACC-Pro samples formed one major clade while MACC-Hyp samples exhibited a drastically distinct expression profile. As observed previously, sample replicates from each media condition formed their own clades to indicate the greatest similarities. In the genes dendrogram, three major clades can be observed with counts being higher in clade 1 and lower in clade 3 when comparing MACC-Hyp samples to the other media conditions. Clustering of MACC-Hyp samples into a distinct clade in the conditions dendrogram is consistent with these major differences in expression profiles.

The putative gene functions making up clade 1 and 3 provided us with potential explanations for the observed differences in gene expression. Expression of most genes in clade 1 was greatly reduced in RCM samples and increased in MACC-Hyp samples. Annotations implicate these genes to encode transporters, vitamin biosynthetic enzymes, and chaperone proteins. Greater expression of these genes is likely required during growth under nutrient-poor conditions, such as in MACC-Hyp, but is less essential in RCM. Genes that exhibited reduced expression in MACC-Hyp as part of clade 3 likely make up a single operon as indicated by their annotations and consecutive positions in the genome (locus tag: CD196_0380 to 0387). The 2-hydroxyisocaproyl-CoA dehydratase encoded within this operon has been characterized and is known to catalyze a dehydration step in the reductive branch of leucine fermentation in

C. difficile.^{21,22} Growth in MACC-Hyp seems to strongly reduce the expression of this pathway relative to growth in RCM and MACC-Pro. One possible explanation for this observation is that Stickland fermentation of Hyp consumes an additional equivalent of NADH in comparison to Pro reduction, which presumably lowers the requirement for other reductive pathways.

In both MACC-Hyp and RCM samples, one of the three samples contained unusually high counts for genes making up clade 2. The differences observed among these genes appear to be outliers since they are not organized within the same genomic region and are mostly annotated as hypothetical proteins. It is intriguing that among the most differentially expressed genes, transcriptomes of *C. difficile* grown on RCM and MACC-Pro share greater similarities than between MACC-Pro and MACC-Hyp. One contributing factor may be that RCM contains a higher concentration of Pro compared to Hyp in the complex protein digests part of the media recipe.

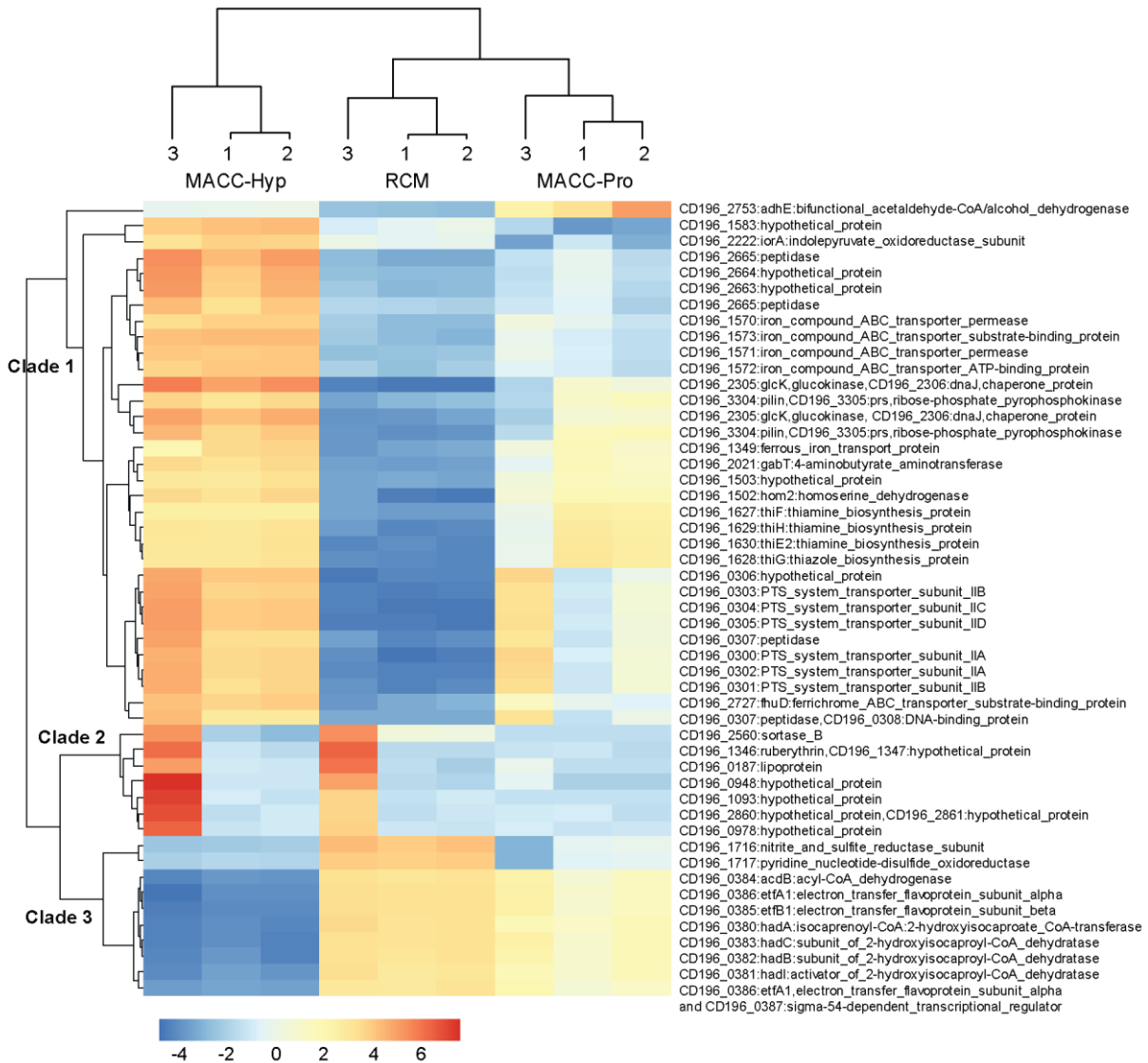


Figure 5.7 Hierarchical clustering of highly variable genes across all samples.

Genes with the 50 highest variance values across all samples were selected based on VST-transformed count values. The heatmap shows each gene's deviation from mean across all samples. The color scale is displayed at the bottom for deviation values. Locus tags and gene annotations are displayed to the right. The clustering and heatmap were generated through the R package pheatmap.¹⁸

Finally, we directly compared the transcriptomes of *C. difficile* grown on MACC-Hyp to MACC-Pro. Differential gene expression analysis was carried out using the R package DESeq2, which uses a negative binomial distribution model.¹⁷ Briefly, raw read counts were normalized to account for differences in sequencing depths between samples. Dispersion of each gene was estimated to obtain the parameter used for data fitting into a generalized linear model for differential expression testing. The logarithmic fold change ($\log_2(\text{fold change})$) and adjusted *p*-value or false discovery rate (FDR) were calculated for each gene. Sequencing reads were assigned to a total of 7,265 genomic regions. This value is almost 2-fold greater than the number of predicted genes in the *C. difficile* genome and is accounted for by assignments to coding sequences (CDSs), intergenic regions (IGRs), and antisense strands. To obtain significant changes in gene expression, an FDR threshold of 0.05 was selected as cutoff, resulting in a subset of 1,455 genes for analyses (Figure 5.8).

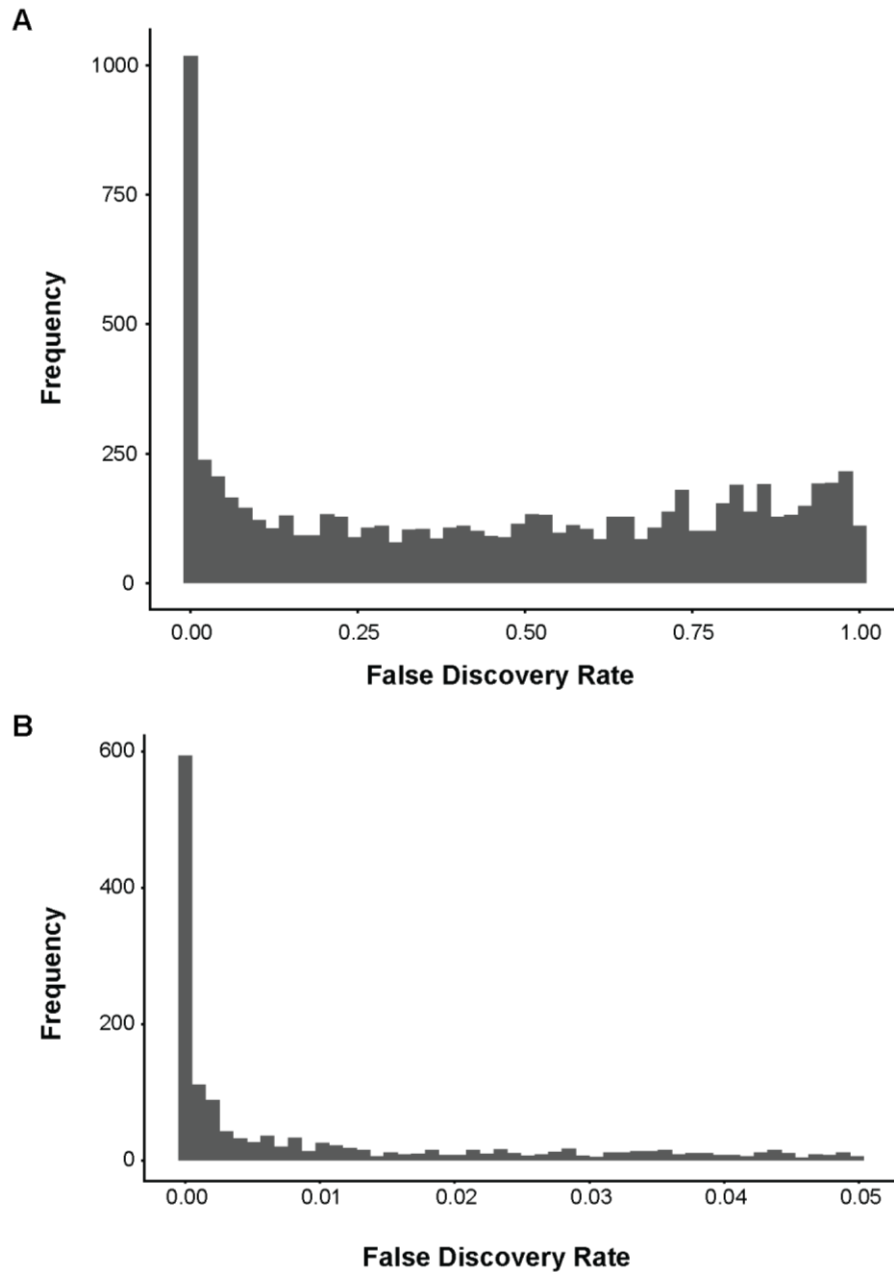


Figure 5.8 Histograms of false discovery rates for all genes analyzed in the MACC-Hyp/MACC-Pro expression comparison.

The FDR distribution of all genes is plotted as histograms. (A) Distribution of all genes analyzed for differential expression between MACC-Hyp and MACC-Pro. (B) Genes with $FDR < 0.05$ are displayed and considered to have significant fold change

Figure 5.8 (Continued)

values for expression analysis. Histograms were generated using the R package `ggplot2` with the number of bins set to 50.²³

Hundreds of gene transcripts were detected at different levels between MACC-Hyp and MACC-Pro with a range of -7.2 to 7.92 in $\log_2(\text{fold change})$ values (Figure 5.9). This demonstrates that substitution of a single media component between these two growth conditions led to a dramatic impact on the metabolic wiring of *C. difficile* at the time point sampled. This can be observed by the numerous genes with $\log_2(\text{fold change})$ values less than -2 or greater than 2 in Figure 5.9.

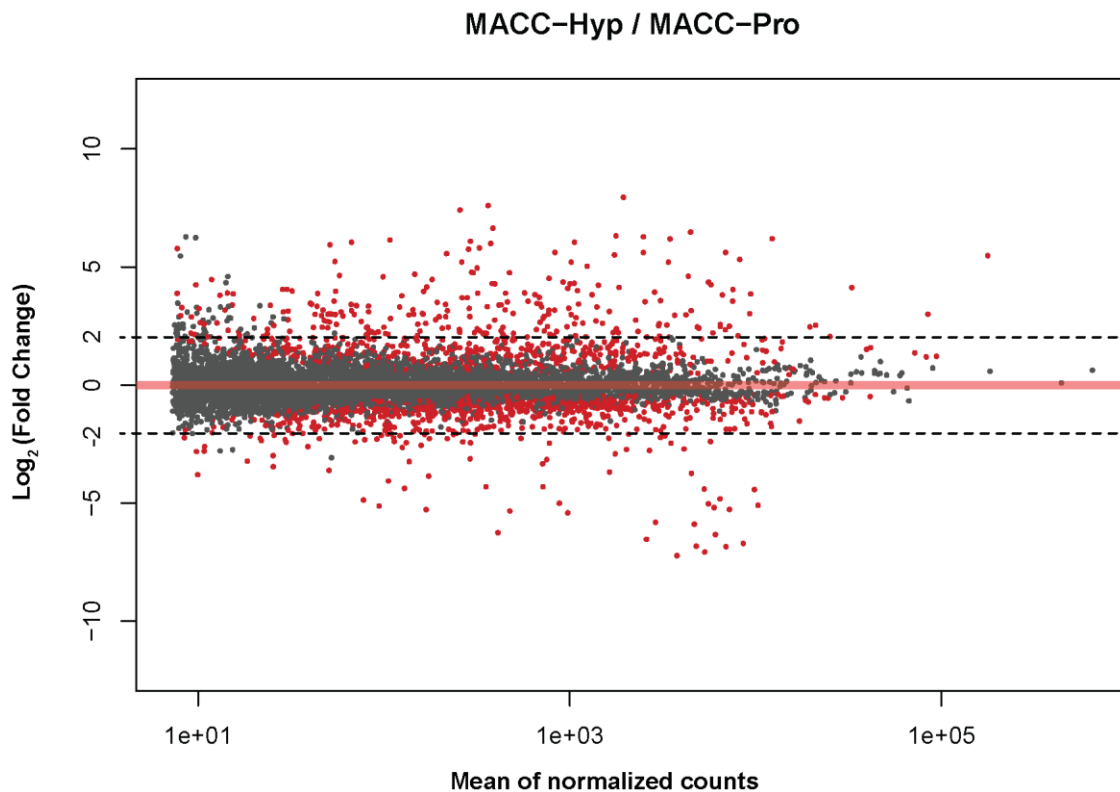


Figure 5.9 A scatter plot comparing read counts from MACC-Hyp to MACC-Pro.

A scatter plot (MA-plot) of $\log_2$ fold changes in normalized counts against the mean of normalized counts for each gene across all 6 samples. Fold changes were calculated for conditions MACC-Hyp compared to MACC-Pro. Data points represent count ratios calculated for each gene when comparing normalized counts between condition MACC-Hyp and MACC-Pro. Data points with a FDR <0.05 are highlighted in red. Dashed lines are displayed to highlight $\log_2(\text{fold change})$ values at thresholds of -2 and 2, which correspond to a 4-fold decrease or increase in gene transcripts. This MA-plot was generated using the `plotMA` function in the R package DESeq2.¹⁷

Genes with the 50 highest and 50 lowest logarithmic fold changes are listed in Table 5.2 and Table 5.3. Redundancy in gene annotations due to assignment to intergenic regions or antisense strands was observed. If a single gene was assigned multiple times to read counts, the regions representing intergenic regions and/or antisense strands were manually removed from the list.

Table 5.2 Top 50 upregulated genes in *C. difficile* 630 Δ erm grown on MACC-Hyp compared to MACC-Pro.

Genes are listed in order of greatest to least fold change (FC). FDR <0.05 was used as cutoff for significance. Multiple locus tags and annotations indicate assignment to an intergenic region.

Locus tag	Gene annotation	log₂ (FC)	FDR	FC	Pathway
CD196_1583	hypothetical protein	7.95	1.29E-24	247.	
CD196_2665, CD196_2666	peptidase, hypothetical protein	7.60	1.80E-13	194.	
CD196_1582, CD196_1583	cation-transporting_ATPase, hypothetical protein	7.41	6.90E-16	170.	
CD196_2665	peptidase	6.64	7.18E-25	99.8	
CD196_2663	hypothetical protein	6.32	2.49E-26	79.7	
CD196_2664	hypothetical protein	6.28	1.36E-30	77.8	
CD196_2222	indolepyruvate oxidoreductase subunit	6.20	2.82E-22	73.5	
CD196_0353, CD196_0354	bifunctional_acetaldehyde-CoA/alcohol_dehydrogenase, hypothetical protein	6.14	2.82E-22	70.6	

Table 5.2 (Continued)

CD196 _2662, CD196 _2663	transposase, hypothetical protein	6.09	5.71E -21	68.1
CD196 _2221	indolepyruvate oxidoreductase subunit beta	6.05	1.07E -17	66.2
CD196 _1882, CD196 _1883	flavodoxin, major intracellular serine protease	5.81	1.65E -15	56.0
CD196 _1572	iron_compound_ABC_transporter_ATP-binding protein	5.62	1.33E -32	49.1
CD196 _2223	aspartate aminotransferase	5.62	9.02E -18	49.1
CD196 _1882	flavodoxin	5.62	1.23E -46	49.0
CD196 _1582	cation-transporting ATPase	5.57	1.13E -09	47.4
CD196 _1573	iron_compound_ABC_transporter_substrate-binding protein	5.51	9.30E -30	45.7
CD196 _2305, CD196 _2306	glucokinase, chaperone protein	5.48	2.08E -17	44.6
CD196 _2220	butyrate kinase	5.32	9.58E -37	39.9
CD196 _1663	sarcosine_reductase_complex_component_B_s ubunit beta	5.21	1.29E -30	37.0
CD196 _1571	iron_compound ABC transporter permease	5.21	1.07E -28	36.9
CD196 _1881	hypothetical protein	5.20	1.19E -16	36.9
CD196 _1352, CD196 _1353	hypothetical protein, hypothetical protein	5.03	4.40E -10	32.7
CD196 _1662	sarcosine_reductase_complex_component_B_s ubunit alpha	4.96	2.99E -25	31.1
CD196 _2448, CD196 _2449	amino acid permease, peptidase	4.78	7.82E -07	27.4

Table 5.2 (Continued)

CD196_2219,					
CD196_2220	hypothetical protein, butyrate kinase	4.76	3.93E-12	27.1	
CD196_0490	hypothetical protein	4.76	1.93E-22	27.0	
CD196_1437,	acetolactate synthase large subunit,				
CD196_1438	prophage_lambdaba042C_site-specific recombinase	4.59	0.00021	24.0	
CD196_2190	Xaa-Pro dipeptidase	4.52	2.20E-73	22.9	
CD196_1357,					
CD196_1358	hypothetical protein, lipoprotein	4.47	0.008561	22.1	
CD196_1570	iron compound ABC transporter permease	4.36	1.55E-15	20.5	
CD196_1352	hypothetical protein	4.36	2.03E-17	20.5	
CD196_0353	bifunctional acetaldehyde-CoA/alcohol dehydrogenase	4.31	9.35E-24	19.9	
CD196_3095,					
CD196_3096	transcription_antiterminator, pyrroline-5-carboxylate reductase	4.30	2.52E-22	19.7	
CD196_1581,					
CD196_1582	glycine dehydrogenase subunit_2, cation-transporting ATPase	4.29	1.39E-09	19.6	
CD196_3097	formate acetyltransferase (HypD)	4.25	2.21E-50	19.0	Hyp metabolism
CD196_2722,					
CD196_2723	signaling protein, drug/sodium antiporter	4.24	5.99E-05	18.9	
CD196_3096	pyrroline-5-carboxylate reductase (P5CR)	4.23	5.72E-52	18.7	Hyp metabolism
CD196_1991	hypothetical protein	4.16	2.94E-58	17.8	

Table 5.2 (Continued)

CD196_2543	branched-chain_amino_acid_transport_system_carrier_protein	4.14	1.23E-32	17.6
CD196_1351	ferrous_iron_transport_protein_B	4.13	1.20E-23	17.5
CD196_1579, CD196_1580	hypothetical_protein, bifunctional_glycine_dehydrogenase/aminomethyl_transferase_protein	4.09	5.73E-20	17.1
CD196_2190, CD196_2191	Xaa-Pro_dipeptidase, glycylsarcosine/betaine_reductase_complex_component_C_subunit_alpha	4.05	0.006455	16.6
CD196_0395	cell_surface_protein	4.00	1.92E-43	16.0
CD196_2876, CD196_2877	glycerate_kinase, alpha-mannosidase	3.98	1.45E-10	15.8
CD196_0491	spore_peptidoglycan_hydrolase	3.97	0.003471	15.7
CD196_1573, CD196_1574	iron_compound_ABC_transporter_substrate-binding_protein, signaling_protein	3.97	8.66E-12	15.6
CD196_2595	hypothetical_protein	3.94	1.55E-15	15.3
CD196_2344	hypothetical_protein	3.90	5.15E-15	14.9
CD196_2015	glutamyl-aminopeptidase	3.89	6.21E-13	14.8
CD196_2372	aminotransferase	3.88	0.044025	14.8

Table 5.3 Top 50 downregulated genes in *C. difficile* 630 Δ erm grown on MACC-Hyp compared to MACC-Pro

Genes are listed in order of greatest to least fold change (FC). FDR <0.05 was used as cutoff for significance. Multiple locus tags and annotations indicate assignment to an intergenic region.

Locus tag	Gene annotation	log₂ (FC)	FDR	FC	Pathway
CD196_0380	isocaprenoyl-CoA:2-hydroxyisocaproate CoA-transferase	-7.22	2.64 E-90	149.4	Leu fermentation ^{21,22}
CD196_0383	subunit_of_oxygen-sensitive_2-hydroxyisocaproyl-CoA dehydratase	-7.07	3.09 E-56	134.3	
CD196_0385	electron transfer flavoprotein subunit beta	-6.84	2.24 E-50	114.7	
CD196_0382	subunit_of_oxygen-sensitive_2-hydroxyisocaproyl-CoA dehydratase	-6.82	6.95 E-64	113.1	
CD196_0386	electron_transfer_flavoprotein_subunit_alpha	-6.71	4.91 E-44	104.4	
CD196_0381	activator_of_2-hydroxyisocaproyl-CoA dehydratase	-6.53	1.14 E-54	92.5	
CD196_0384	acyl-CoA dehydrogenase	-6.33	1.23 E-43	80.7	
CD196_0379	2-hydroxyisocaproate dehydrogenase	-5.89	1.32 E-80	59.3	
CD196_0386, CD196_0387	electron_transfer_flavoprotein_subunit_alpha; sigma-54-dependent transcriptional regulator	-5.81	4.25 E-40	56.2	
CD196_0779	hypothetical protein	-5.41	6.49 E-72	42.5	
CD196_0935	3-hydroxybutyryl-CoA dehydrogenase	-5.26	9.67 E-14	38.4	Butyrate formation ²⁴
CD196_0931	butyryl-CoA dehydrogenase	-5.19	4.69 E-11	36.5	
CD196_0936	acetyl-CoA acetyltransferase	-5.09	3.91 E-12	34.0	
CD196_0932	electron transfer flavoprotein subunit beta	-5.03	8.22 E-11	32.8	
CD196_0936, CD196_0937	acetyl-CoA_acetyltransferase, pseudouridylate synthase	-5.01	2.18 E-08	32.1	

Table 5.2 (Continued)

CD196 _2187	hypothetical protein	-4.87	1.22 E-08	29.2	
CD196 _0933	electron_transfer_flavoprotein_subunit_alph a	-4.82	7.72 E-10	28.2	Butyrate formation ²⁴
CD196 _2753	bifunctional_acetaldehyde- CoA/alcohol_dehydrogenase	-4.43	4.91 E-12	21.5	
CD196 _0934	3-hydroxybutyryl-CoA_dehydratase	-4.40	2.36 E-08	21.2	Butyrate formation ²⁴
CD196 _2186	succinyl-CoA:coenzyme_A_transferase	-4.38	7.14 E-09	20.8	Succinate metabolism ²⁵
CD196 _2177	PTS_system_mannitol- specific_transporter_subunit_IIBC	-4.30	0.00 0105	19.7	
CD196 _2185	succinate-semialdehyde_dehydrogenase	-3.85	4.10 E-08	14.5	Succinate metabolism ⁴
CD196 _2199, CD196 _2200	oligoendopeptidase, hydrolase	-3.79	0.02 5048	13.9	
CD196 _0051	prolyl-tRNA_synthetase	-3.74	1.55 E-08	13.4	
CD196 _0930, CD196 _0931	regulatory_protein, CoA_dehydrogenase	butyryl- -3.69	1.43 E-06	12.9	
CD196 _2177, CD196 _2178	PTS_system_mannitol- specific_transporter_subunit_IIBC, inosine_5- monophosphate_dehydrogenase	-3.61	0.01 1366	12.2	
CD196 _2464	hypothetical protein	-3.46	1.07 E-06	11.0	
CD196 _2176	transcription_antiterminator	-3.33	0.00 2385	10.1	
CD196 _2752, CD196 _2753	signaling_protein, bifunctional_acetaldehyde- CoA/alcohol_dehydrogenase	-3.24	0.00 1154	9.4	
CD196 _2462, CD196 _2463	hypothetical protein, hypothetical protein	-3.16	3.27 E-11	8.9	
CD196 _1407, CD196 _1408	3-mercaptopyruvate_sulfurtransferase, ferredoxin- NADP(+)_reductase_subunit_alpha	-2.91	7.88 E-05	7.5	

Table 5.2 (Continued)

CD196_2989,					
CD196_2990	dihydroorotate_dehydrogenase_catalytic_subunit, xanthine permease	-2.82	0.02	4811	7.0
CD196_2990,					
CD196_2991	xanthine_permease, chlorohydrolase/aminohydrolase	-2.80	0.01	2636	6.9
CD196_0050	prolyl-tRNA synthetase	-2.77	7.83E-05		6.8
CD196_0783	Isocitrate/3-isopropylmalate dehydrogenase	-2.73	5.15E-11		6.6
CD196_2463	hypothetical protein	-2.71	1.23E-27		6.5
CD196_2175	PTS_system_mannitol-specific transporter subunit IIA	-2.64	0.01358		6.2
CD196_3345,					
CD196_3346	tRNA-dihydrouridine_synthase, pantothenate kinase	-2.60	0.017994		6.1
CD196_0378,					
CD196_0379	hypothetical_protein, hydroxyisocaproate dehydrogenase	2- -2.60	0.001918		6.1
CD196_0780,					
CD196_0781	ATP-dependent_RNA_helicase, trans-homoaconitate synthase	-2.55	9.21E-05		5.9
CD196_0782	aconitate hydratase	-2.46	6.98E-09		5.5
CD196_0844	iron-dependent_hydrogenase	-2.45	1.94E-09		5.4
CD196_0846	metallo-beta-lactamase	-2.44	1.97E-08		5.4
CD196_0801,					
CD196_0802	MarR_family_transcriptional_regulator, oligopeptide ABC transporter permease	-2.41	3.11E-05		5.3
CD196_1868,					
CD196_1869	cell_surface_protein, tryptophan transport protein	-2.40	5.40E-07		5.3

Table 5.2 (Continued)

CD196_1409	oxidoreductase	-2.39	2.42 E-05	5.2
CD196_2990	xanthine permease	-2.39	1.21 E-13	5.2
CD196_1865, CD196_1866	NUDIX_family_protein, hypothetical_protein	-2.35	0.00 0348	5.1
CD196_0821	maltose_O-acetyltransferase	-2.32	1.43 E-06	5.0
CD196_1783	Ethanolamine/propanediol_utilization_ protein	-2.31	3.03 E-06	5.0

Genes with the most negative fold change values are predicted to catalyze leucine fermentation (Table 5.3). These genes were drastically downregulated in MACC-Hyp compared to MACC-Pro, which is consistent with our previous analysis of genes with high variances (Figure 5.7). Genes encoding part of the butyrate production pathway were also among the most reduced read counts. The second most downregulated operon consists of genes that catalyze the overall transformation of acetyl-CoA to butyrate (CD196_0931 to 0937) (Table 5.3).²⁴ Notably, *C. difficile* had previously been shown to metabolize succinate to butyrate during infection in a mouse study and the deletion of succinate transporter led to a reduction in colonization.²⁵ Interestingly, two strongly downregulated genes (CD196_2185, CD196_2186) are part of a gene cluster that metabolizes succinate to crotonyl-CoA, an intermediate in the butyrate producing pathway.²⁶ From this data, metabolism of succinate to butyrate appears to be suppressed in cultures grown in MACC-Hyp relative to cultures grown in MACC-Pro. The link between Hyp metabolism and butyrate formation is particularly intriguing given butyrate's role as a major source of energy for colonocytes.^{27,28} Although *C. difficile* is not expected to be abundant in the healthy gut microbiota, HypD is encoded within other butyrate-producing gut isolates including *Clostridium*

sporosphaeroides, *Clostridium innocuum*, and *Clostridium tyrobutyricum*.²⁷ In addition to its role in energy metabolism, butyrate has beneficial effects for human health by promoting gut development, suppressing inflammatory responses, and inducing apoptosis in cancer cells.^{27,28}

In contrast, the functions for genes most upregulated in MACC-Hyp relative to MACC-Pro are difficult to predict. These genes are mostly annotated as hypothetical proteins or with general functions such as peptidase, cation-transporting ATPase, and transposase (Table 5.2). We can infer that the genes with the highest fold change values make up two operons: CD196_1582 to 1583 and CD196_2662 to 2666 (Table 5.2).

A broad search among annotations revealed 8 incidences of “peptidase” or “protease” in the upregulated genes compared to the single occurrence in the downregulated genes. This indicates that free Hyp might induce the expression of proteases, which is of interest since Pro and Hyp often occur immediately adjacent to one another in the sequences of highly abundant proteins such as collagen. It is possible that these peptidases act on proteins or peptides that are selectively enriched in Hyp over Pro. Additionally, these proteases may be selective for peptide bonds between Hyp and non-Pro residues. However, we cannot exclude contributions from different downstream metabolites that accumulated during growth between MACC-Hyp and MACC-Pro. Most of the annotations in both lists point to functions in primary metabolism catalyzed by redox-active enzymes. Genes encoding HypD and P5CR were among the top 50 overexpressed genes in MACC-Hyp with fold change of 19 and 18.7, respectively. The gene encoding AE, which is essential for HypD activation, was also upregulated with a fold change value of 10.8, demonstrating that transcription of the *hypD* gene cluster is induced by free Hyp to a greater degree compared to free Pro (Table 5.4). However, the possibility remains that Pro does influence

transcription of the *hypD* gene cluster since the basal medium MACC was not included in this experiment for comparative analysis with MACC-Pro.

Table 5.4 Genes of interest and their fold changes in MACC-Hyp / MACC-Pro

Locus tag	Gene annotation	log ₂ (FC)	FDR	FC	Pathway
CD196_3098	pyruvate_formate-lyase_3_activating_enzyme	3.4317	1.16E-28	10.8	Hyp fermentation
CD196_3097	formate_acetyltransferase	4.2463	2.21E-50	19.0	
CD196_3096	pyrroline-5-carboxylate_reductase	4.2265	5.72E-52	18.7	
CD196_3051	proline_racemase	-1.5247	0.004219	0.3	Pro fermentation
CD196_3052	hypothetical_protein	-1.8478	0.000379	0.3	
CD196_3053	hypothetical_protein	-1.8750	0.000545	0.3	
CD196_3054	hypothetical_protein	-1.7531	0.000511	0.3	
CD196_3055	prdB:proline_reductase	0.156	0.8523		
CD196_3055	prdB:proline_reductase	0.06449	0.9318		
CD196_3056	hypothetical_protein	0.1492	0.8297		
CD196_3057	prdA:proline_reductase_subunit_proprotein	-0.2775	0.6737		
CD196_3058	sigma-54-dependent_transcriptional_activator, prdR	-0.4084	0.21		
CD196_3059	prdC:electron transfer protein	-0.7895	0.1051		
CD196_0599	tcdD, putative transcriptional regulator	n/a			Toxins A and B, PaLoc
CD196_0600	toxin_B	-0.142	0.954		
CD196_0601	cell_wall_hydrolase_protein	n/a			
CD196_0602	toxin_A	-0.066	0.906		
CD196_2443	response_regulator	0.181	0.673		Binary toxin, CdtLoc
CD196_2444	ADP-ribosyltransferase_subunit_CdtA	0.7043	0.138		

CD196			
2445	ADP-ribosyltransferase binding protein	0.5278	0.196

In addition to sampling the top 100 upregulated and downregulated genes, we compiled a list of fold changes for genes of interest that are relevant in Stickland fermentation of Hyp or Pro, and virulence. Transcripts mapping to the *prd* operon with FDRs <0.05 were detected at slightly lower levels in MACC-Hyp compared to MACC-Pro (Table 5.4). Remaining genes in this operon are associated with high FDRs, indicating a lack of significant changes in *prd* transcription. Expression of *prd* proteins was shown to be induced by the regulator PrdR in the presence of Pro.²⁹ Therefore, the small differences observed can be attributed to the production of Pro in MACC-Hyp by coordinated actions of HypD and P5CR, which would induce the expression of *prd* genes to similar levels observed in MACC-Pro. To conclude, the *hypD* gene cluster was transcribed at a higher level in the presence of Hyp while expression of the *prd* operon did not change drastically.

Toxin productions in *C. difficile* has been associated with increasing severity in clinical outcome.³⁰ Well-studied toxins are encoded within genomic regions termed the pathogenicity locus (PatLoc) and the Cdt locus (CdtLoc).³⁰ Although the regulation of these toxins is not well-understood, studies have shown that growth on amino acids impact toxin production.^{5,29,31,32} Therefore, we wondered if toxin-encoding genes were selectively induced in the presence of Hyp or Pro. Fold changes among genes part of PatLoc and CdtLoc operons were extremely low between the two MACC conditions with high FDR values (Table 5.4). From this data, we can conclude that the utilization of Hyp or Pro as an alternative electron acceptor did not selectively alter transcriptional regulation of these genomic loci.

5.2.5 Activity of HypD from *Bacteroides vulgatus* ATCC 8482

Given HypD's distribution in highly diverse microbial species, we sought to gain insight into Hyp metabolism in non-Clostridial species. We were particularly intrigued by the high abundances of *hypD* found among Human Microbiome Project (HMP) stool metagenomes, which was described in Chapter 2. Specifically, most reads identified in these metagenomes mapped onto nodes in the sequence similarity network (SSN) containing Bacteroidetes HypD sequences. This prompted us to explore the role of Hyp metabolism in gut Bacteroidetes, a metabolic activity previously unknown for these species. First, we sought to validate the functional assignment of these GREs as HypDs based on the conserved active site residues tested in Chapter 3. At the time, HypD from *C. difficile* 70-100-2010 was the only HypD to be purified and characterized *in vitro*. Therefore, we proceeded to verify the activity of a putative HypD from a common gut isolate *Bacteroides vulgatus* ATCC 8482 (UniProt ID: A6L095, 52% amino acid identity to the characterized *C. difficile* HypD).

We initially attempted to obtain evidence of enzyme activity by detecting Hyp consumption in *E. coli* strains overexpressing the *B. vulgatus* HypD and its AE partner. The genomic region encoding HypD and AE was amplified in one step from *B. vulgatus* and cloned into pTrcHis2A for overexpression. As a positive control, pTrcHis2A vector containing an insert of HypD and AE from *C. difficile* 70-100-2010 was constructed as well. An empty pTrcHis2A vector was included as the negative control. Three *E. coli* strains (TOP10, BL21(DE3), and BL21(DE3)-CodonPlus-RIL) were transformed with these vectors (Figure 5.10). Protein overexpression was induced with isopropyl β -D-1-thiogalactopyranoside (IPTG) during exponential phase under anaerobic conditions. Hyp concentrations were quantified in overnight cultures by LC-MS/MS. No significant decrease in Hyp levels was detected in any culture compared to media controls (Figure 5.10). A few possible explanations can be provided for the

lack of activity in strains overexpressing the previously characterized *C. difficile* HypD and AE. Proteins expressed under the conditions tested may have been unstable or misfolded, resulting in low conversions not detectable by our LC-MS/MS method. Indeed, a similar experiment carried out with CutC and its AE, CutD, yielded <1.5% conversion of trimethylamine (TMA) from choline in *E. coli* cultures overexpressing these two enzymes.³³ The use of deuterated choline enabled detection of d₉-TMA at a high sensitivity to directly link its production to CutC activity, which was not possible in our experiments using unlabeled Hyp. Furthermore, TMA is an end product in choline metabolism, while P5C is an intermediate in several pathways and likely undergoes additional transformations, which could preclude detection. In conclusion, we did not obtain evidence for activity of *B. vulgatus* HypD and AE through this experiment.

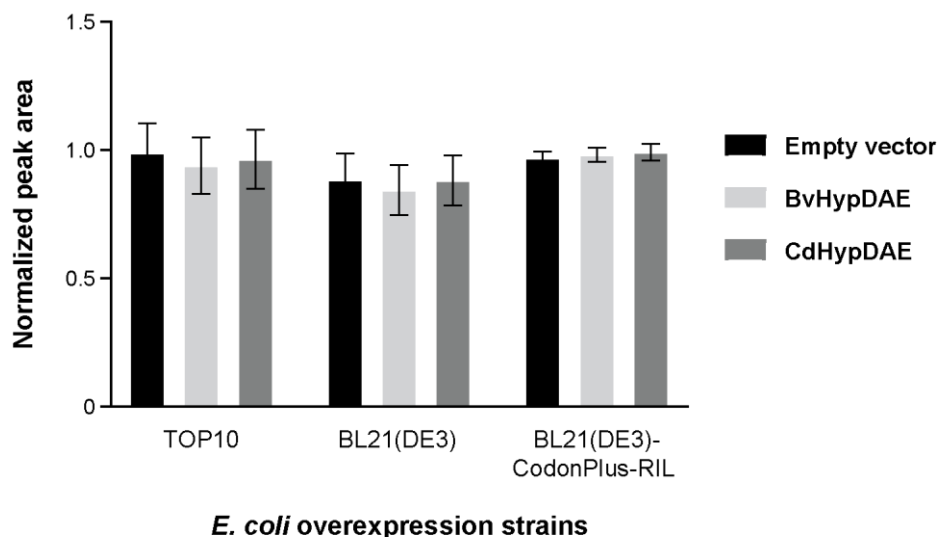


Figure 5.10 Attempted detection of Hyp metabolism in *E. coli* overexpressing the HypD-AE gene cluster.

E. coli strains TOP10, BL21(DE3), and BL21(DE3)-CodonPlus-RIL were transformed with an empty pTrcHis2A vector, pTrcHis2A-BvHypDAE, or pTrcHis2A-CdHypDAE. Bv indicates HypD and AE cloned from *B. vulgatus* ATCC 8482 whereas Cd indicates HypD and AE cloned from *C. difficile* 70-100-2010. All resulting *E. coli* strains were grown anaerobically at 37 °C in BHI supplemented with 1 mM Hyp and the corresponding antibiotics. Protein overexpression was induced with IPTG during exponential phase. Hyp was quantified in all cultures and media controls after incubation overnight. Hyp peak areas were normalized with media controls. Data are shown as mean $\pm$ SD with $n = 3$.

After failing to detect activity in the culture-based assay, we proceeded to purify HypD and AE individually from *B. vulgatus*. We employed the end-point and kinetic assays first reported in Chapter 2 for *in vitro* biochemical characterization. All protocols for protein overexpression and

purification, GRE activation, end-point activity assay, and kinetic assay used for *C. difficile* HypD were applied to the *B. vulgatus* proteins without any alteration. HypD (UniProt ID: A6L095) and AE (UniProt ID: A6L094) were cloned from *B. vulgatus* ATCC 8482 and are referred to as BvHypD and BvAE. Proteins were overexpressed and purified from *E. coli* BL21-CodonPlus(DE3)-RIL $\Delta proC::aac(3)IV(Am^R)$ to high homogeneity (Figure 5.11).

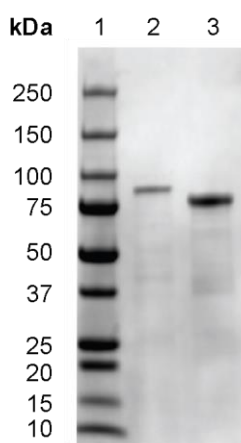


Figure 5.11 SDS-PAGE of purified BvHypD and MBP-BvAE from *B. vulgatus* ATCC 8482.

Precision Plus ProteinTM All Blue Standards (Bio Rad) (lane 1), BvHypD (lane 2, UniProt ID: A6L095), and MBP-BvAE (lane 3, UniProt ID: A6L094). BvAE was purified with a maltose binding protein fused at the N-terminus, which is abbreviated as MBP-BvAE. Molecular weights of these proteins are as follows: BvHypD – 92.4 kDa; MBP-BvAE – 80.7 kDa.

A coupled enzyme assay was used to detect activity of purified BvHypD by coupling Hyp dehydration to P5C reduction by *C. difficile* P5CR (CdP5CR) (Figure 5.12A). Without

optimization of purification and assay conditions, we expected low turnover rates for BvHypD and therefore selected a higher range of substrate concentrations for the kinetic assay than the assays performed with *C. difficile* enzymes. Preliminary data showed that BvHypD does generate P5C as detected by the decreasing NADH concentrations from P5C reduction by CdP5CR (Figure 5.12B). To confirm the finding from the kinetic assay, we also tested BvHypD's activity using an end-point activity assay (Figure 5.13A). Incubation of BvHypD, BvAE, CdP5CR, and Hyp in addition to cofactors SAM and NADH yielded Pro. In contrast to *C. difficile* HypD, which catalyzed full conversion of Hyp to Pro within one hour, BvHypD required a longer period of incubation (Figure 5.13B). Partial formation of Pro was observed in assays incubated for one hour whereas incubation for 21 hours resulted in completion of the transformation (Figure 5.13B). The slower rates could reflect biochemical differences between HypDs from phylogenetically diverse bacteria and/or suboptimal conditions for BvHypD reconstitution. Since no activation data was collected for BvHypD, another possibility is a lower activation of this enzyme. Although these experiments will need to be repeated with optimized assay conditions and with negative controls excluding individual assay components, these results strongly suggest these Bacteroides proteins are true HypDs. Collectively, results from the *in vitro* assays reported here provided evidence for the correct annotation of Bacteroidetes HypDs.

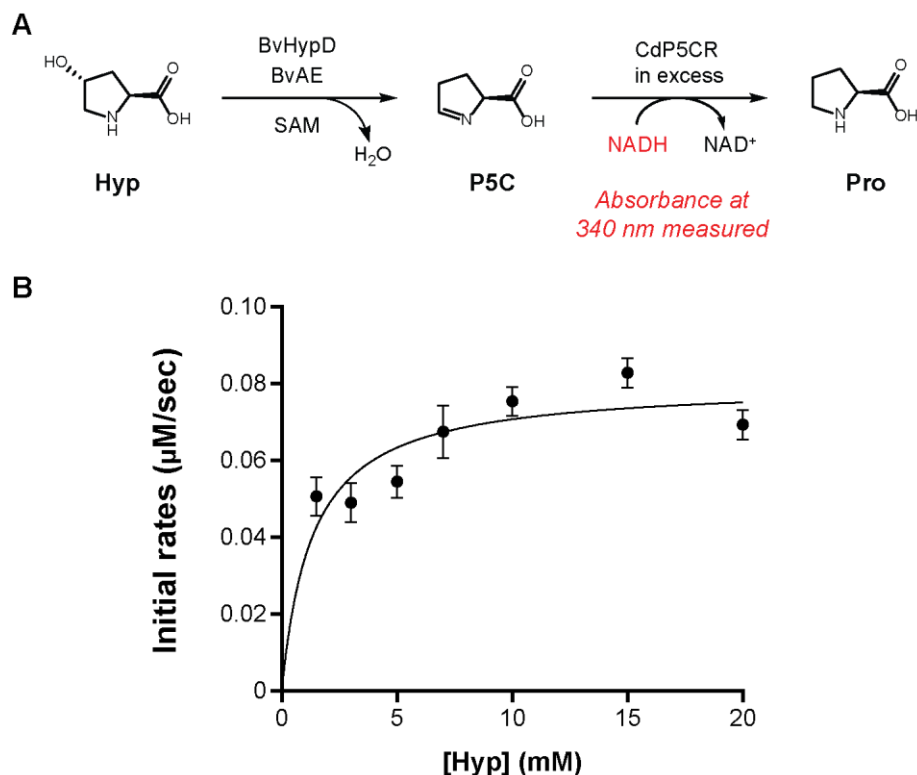


Figure 5.12 BvHypD is active toward Hyp in a coupled enzyme kinetic assay.

(A) Conversion of Hyp to P5C by BvHypD was coupled to reduction of P5C to L-proline by P5CR. Absorbance of NADH at 340 nm was continuously monitored to calculate initial rates. (B) Kinetic data of HypD was fit to the Michaelis–Menten equation using nonlinear regression in Graphpad Prism 7.02. Initial rates are not normalized to enzyme concentration due to a lack of activation data for BvHypD. Therefore, rates are reported in $\mu\text{M}/\text{sec}$ for substrate turnover rate. A low background rate in the absence of Hyp was subtracted from all data points. Preliminary data are shown as mean $\pm$ SD with $n = 3$ replicates from a single experiment.

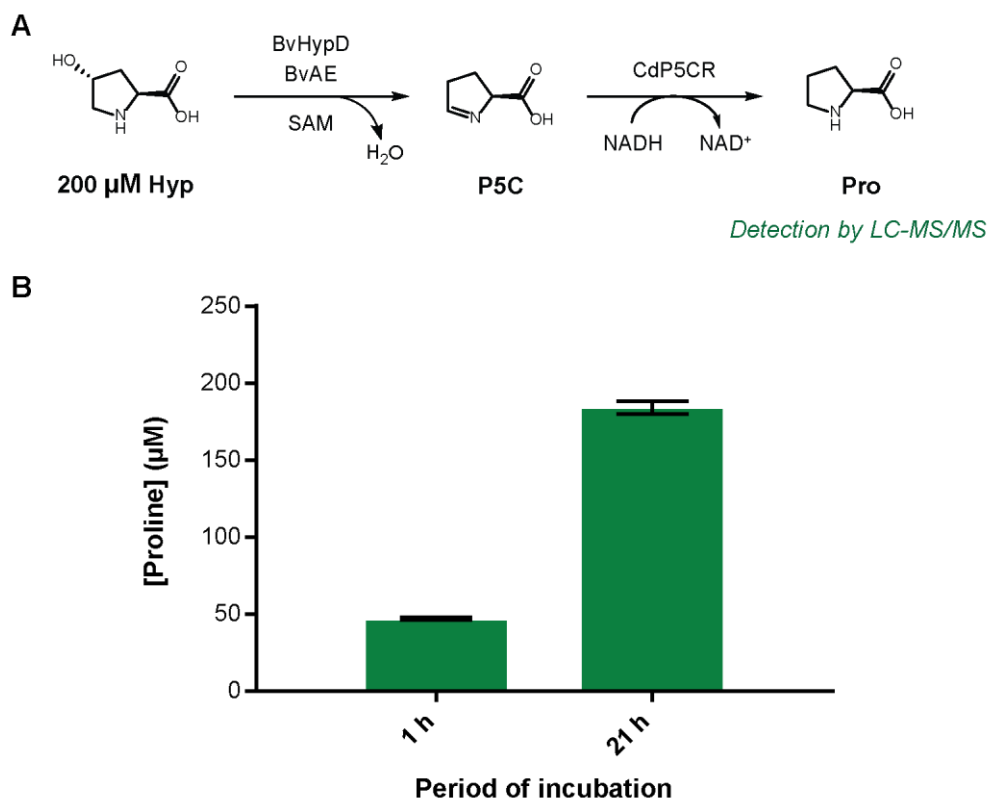


Figure 5.13 End-point activity assay for BvHypD.

(A) Scheme depicting the coupled enzyme assay used to test if BvHypD is can convert Hyp to P5C *in vitro*. BvHypD was first activated with BvAE. Activated BvHypD was added to a final concentration of 0.3 μM in a mixture containing 3 μM P5CR, 0.4 mM NADH, and 0.2 mM Hyp to initiate the reaction. (B) Concentrations of Pro were quantified in end-point activity assays using external standard curves. Data are shown as mean $\pm$ SD with $n = 3$ replicates. Assays were quenched after 1 hour or 21 hours of incubation.

5.2.6 Growth of *Bacteroides* and *Parabacteroides* species in a minimal medium containing Pro or Hyp

Having demonstrated BvHypD activity *in vitro*, we proceeded to study the physiological relevance of this enzyme in Bacteroidales. In contrast to Clostridiales, gut isolates within the order Bacteroidales have not been reported to perform Hyp or Pro fermentation. Additionally, many of these species are not auxotrophic for amino acids as demonstrated by growth in minimal media containing glucose and ammonium chloride as the carbon and nitrogen source.³⁴ Despite a lack of prior evidence, we asked if differences in growth and/or amino acid utilization could be detected among strains that encode or lack HypD. With this goal in mind, a panel of 9 *Bacteroides* and *Parabacteroides* species were screened for growth on M9 minimal medium. Supplementation with either Pro or Hyp (20 mM or 40 mM) was included to test for metabolism and growth enhancement. *Bacteroides caccae* ATCC 43185 (HypD⁻), *Bacteroides dorei* CL02T12C06 (HypD⁺), *Bacteroides fragilis* ATCC 25285 (HypD⁻), *Bacteroides ovatus* ATCC 8483 (HypD⁻), *Bacteroides thetaiotaomicron* VPI 5482 (HypD⁻), *B. vulgatus* ATCC 8482 (HypD⁺), *Parabacteroides distasonis* ATCC 8503 (HypD⁺), *Parabacteroides goldsteinii* CL02T12CL30 (HypD⁺), and *Parabacteroides johnsonnii* DSM 18315 (HypD⁺) were tested and results are summarized in Table 5.5.

Table 5.5 Summary of growth of Bacteroidales in M9 and BHIS.

Strain	HypD	Media	Supplement	Growth
<i>Bacteroides caccae</i> ATCC 43185	–			–
<i>Bacteroides dorei</i> CL02T12C06	+			–
<i>Bacteroides fragilis</i> ATCC 25285	–		None,	–
<i>Bacteroides ovatus</i> ATCC 8483	–		20 mM or	slight
<i>Bacteroides thetaiotaomicron</i> VPI 5482	–	M9	40 mM Pro,	slight
<i>Bacteroides vulgatus</i> ATCC 8482	+		20 mM or	+
<i>Parabacteroides distasonis</i> ATCC 8503	+		40 mM Hyp	slight
<i>Parabacteroides goldsteinii</i> CL02T12CL30	+			slight
<i>Parabacteroides johnsonii</i> DSM 18315	+			slight
<i>Bacteroides ovatus</i> ATCC 8483	–			+
<i>Bacteroides ovatus</i> ATCC 8483 pFD340-HypDAE	+			+
<i>Bacteroides thetaiotaomicron</i> VPI 5482	–			+
<i>Bacteroides thetaiotaomicron</i> VPI 5482 pFD340-HypDAE	+	BHIS	1 mM or	+
<i>Bacteroides vulgatus</i> ATCC 8482	+		10 mM Hyp	+
<i>Bacteroides vulgatus</i> ATCC 8482 $\Delta hypD$	–			+
<i>Bacteroides vulgatus</i> ATCC 8482 ΔAE	–			+

A lack of growth or poor growth was observed for most strains in all media conditions. Four strains lacking HypD were tested, of which two displayed poor growth while the remaining two showed no growth in any media. From the five HypD-encoding strains, four strains grew in all M9 media conditions (M9 + Hyp conditions shown in Figure 5.14). Overall, for strains that grew on M9, no significant difference was observed between M9 basal medium compared to M9 supplemented with an amino acid. The presence of 40 mM Hyp resulted in a slight decrease in growth of *B. vulgatus* and *P. distasonis* (Figure 5.14). Inhibition of growth in other bacteria had been previously observed for different amino acids under high concentrations as well.^{35,36} All cultures exhibiting bacterial growth were collected for Hyp quantification using the LC-MS/MS

method described in Chapter 4. No significant decrease in Hyp concentrations was observed in any culture and therefore no metabolic activity toward Hyp can be assigned to any of the Bacteroidales tested here. Possible explanations for this observation and proposed work to follow up on these initial experiments are discussed at the end of Section 5.2.8.

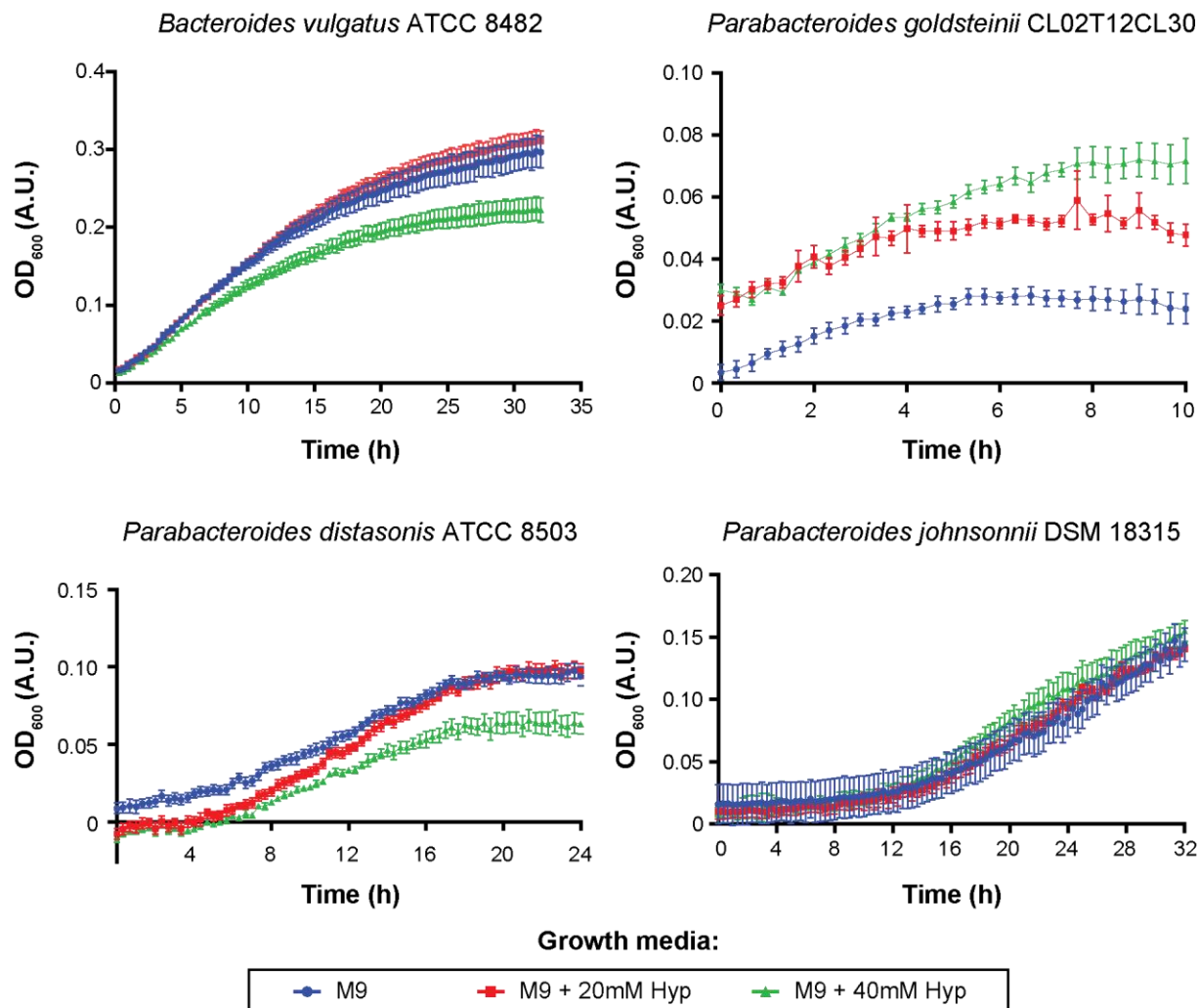


Figure 5.14 Growth of HypD-encoding Bacteroidales on M9 minimal medium supplemented with Hyp.

Four Bacteroidales strains were grown on M9 minimal medium supplemented with 20 mM or 40 mM Hyp. *Bacteroides vulgatus* ATCC 8482, *Parabacteroides goldsteinii* CL02T12CL30, *Parabacteroides distasonis* ATCC 8503, and *Parabacteroides johnsonnii* DSM 18315 did not exhibit enhanced growth upon supplementation with Hyp. All four strains encode a copy of *hypD*. Addition of 40 mM Hyp slightly inhibited growth of *B. vulgatus* and *P. distasonis*. Absorbance

Figure 5.14 (Continued)

at 600 nm was continuously measured during growth and normalized to the corresponding media controls. Data are shown as mean $\pm$ SD with n = 3 to 5 replicates.

5.2.7 Genetics in *Bacteroides* and *Parabacteroides* species for heterologous expression and deletion of BvHypD and BvAE

Genetic manipulations of Bacteroidales have been possible for the past few decades and have enabled mechanistic studies linking genes to functions in this particular group of gut microbes.³⁷⁻³⁹ Although Hyp metabolism was not detected in the WT strains tested, we wondered if a difference in fitness could be observed between WT and isogenic mutants for HypD and AE. In addition, we sought to test for a gain of function by constructing *Bacteroides* expression strains as an alternative approach from the heterologous expression experiments described in *E. coli*. Here, we describe the construction of deletion mutants and expression strains in gut isolates, providing tools for further investigations of Hyp metabolism in Bacteroidales. This work was done in collaboration with Prof. Laurie Comstock at Harvard Medical School.

The vectors and strains constructed as part of this work are listed in Table 5.6. Since protein expression in phylogenetically related host strains usually leads to better yields and activities, we also proceeded to generate expression vectors for *Bacteroides* species. A single region encoding BvHypD (locus tag: Bvu_1421) and BvAE (locus tag: Bvu_1420) was amplified from *B. vulgatus* ATCC 8482 and cloned into pFD340, which contains a constitutive *Bacteroides* promoter.⁴⁰ The resulting vector, pFD340-BvHypDAE, was conjugated into three *Bacteroides* species that lack a *hypD* gene, *B. ovatus* ATCC 8483, *B. fragilis* ATCC 25285, and *B. thetaiotaomicron* VPI 5482.

To generate vectors for genetic deletion through allelic exchange, regions 5'- and 3'- of genes encoding BvHyp and BvAE were amplified from *B. vulgatus* ATCC 8482. The corresponding flanking regions were joined and ligated into pBlueScript for colony screening and sequencing. Low transformation efficiencies and mutations within insert fragments were observed, which was likely a result of product toxicity from these sequences. The ligated flanking regions for BvHyp and BvAE were cloned into pKNOCK-*bla-ermGb*³⁸ to generate suicide vectors for clean deletions. Each verified plasmid was conjugated into *B. vulgatus* ATCC 8482 using *E. coli* S17-1 λ pir, and this was carried out by Prof. Laurie Comstock.^{iv} Colony PCR amplifying the gene of interest was used to identify bacterial clones with successful gene deletions. Two strains, *B. vulgatus* ATCC 8482 Δ *hypD* and Δ *AE*, were generated through this approach. Initial attempts to detect phenotypic differences between *B. vulgatus* WT and mutants are detailed in the following two sections.

Table 5.6 Strains and plasmids used in studies on Bacteroidales.

Strain or plasmid	Description or features	Reference /source
<i>Escherichia coli</i> strains		
BL21(DE3)	Overexpression of HypDAE cluster	Invitrogen
BL21(DE3)-CodonPlus-RIL	Overexpression of HypDAE cluster	Invitrogen
BL21(DE3)-CodonPlus-RIL Δ <i>proC</i>	Heterologous protein overexpression	¹
TOP10	Cloning and plasmid storage	ThermoFisher
S17-1 λ pir	Conjugation with <i>Bacteroides</i> species	Comstock lab
<i>Bacteroides</i> or <i>Parabacteroides</i> strains		
<i>Bacteroides caccae</i>	ATCC 43185	ATCC
<i>Bacteroides dorei</i>	CL02T12C06	Broad Institute

^{iv} Conjugation, selection for homologous recombination, serial passaging, and colony PCR to identify successful deletion of *hypD* or *AE* in *B. vulgatus* ATCC 8482 were conducted by Prof. Laurie Comstock at Harvard Medical School.

Table 5.6 (Continued)

<i>Bacteroides fragilis</i>	ATCC 25285	ATCC
	pFD340-BvHypDAE	This study
<i>Bacteroides ovatus</i>	ATCC 8483	ATCC
	pFD340-BvHypDAE	This study
<i>Bacteroides thetaiotaomicron</i>	VPI 5482	41
	pFD340-BvHypDAE	This study
<i>Bacteroides vulgatus</i>	ATCC 8482	ATCC
<i>Bacteroides vulgatus</i>	ATCC 8482 Δ hypD	This study
<i>Bacteroides vulgatus</i>	ATCC 8482 Δ AE	This study
<i>Parabacteroides goldsteinii</i>	CL02T12CL30	Comstock lab
<i>Parabacteroides distasonis</i>	ATCC 8503	ATCC
<i>Parabacteroides johnsonnii</i>	DSM 18315	DSMZ
Plasmids		
pTrcHis2A	Protein overexpression in <i>E. coli</i>	
pTrcHis2A-CdHypDAE	Cluster containing HypD and AE from <i>C. difficile</i> 70-100-2010	This study
pTrcHis2A-BvHypDAE	Cluster containing HypD and AE from <i>B. vulgatus</i> ATCC 8482	This study
pET28a-BvHypD	Overexpression of HypD from <i>B. vulgatus</i> ATCC 8482, Bvu_1421	This study
pSV272-PfMBP-BvAE	Overexpression of AE from <i>B. vulgatus</i> ATCC 8482, Bvu_1420	This study
pFD340	Expression vector for <i>Bacteroides</i> species	40
pFD340-BvHypDAE	Cluster containing HypD and AE from <i>B. vulgatus</i> ATCC 8482	This study
pBlueScript SK	Vector for blue-white colony screen and ligation of inserts	Comstock lab
pBlueScript- Δ HypD	Vector carrying 2 to 2.2 kb regions 5' and 3' of Bvu_1421	This study
pBlueScript- Δ AE	Vector carrying 2 to 2.5 kb regions 5' and 3' of Bvu_1420	This study
pKNOCK- <i>bla-ermGb</i>	Suicide vector for <i>Bacteroides</i> genetics by allelic exchange	38
pKNOCK- <i>bla-ermGb</i> - Δ BvHypD	Vector carrying 2 to 2.2 kb flanking regions around Bvu_1421	This study
pKNOCK- <i>bla-ermGb</i> - Δ BvAE	Vector carrying 2 to 2.5 kb flanking regions around Bvu_1420	This study

5.2.8 Experiments examining growth of *B. vulgatus* ATCC 8482 WT, $\Delta hypD$, and ΔAE strains

Here, we describe experiments studying growth and Hyp utilization in *B. vulgatus* ATCC 8482 WT, $\Delta hypD$, and ΔAE strains. These strains were grown in the rich medium BHIS supplemented with 1 mM or 10 mM Hyp (Figure 5.15). No drastic difference in growth was observed among these strains. We then quantified Hyp levels in cultures at stationary phase and did not detect significant decrease in Hyp compared to media blanks. Possible explanations for the discrepancy between our *in vitro* enzyme work and the growth experiment on *B. vulgatus* WT and deletion mutants in addition to culture-based studies on other Bacteroidales are discussed below.

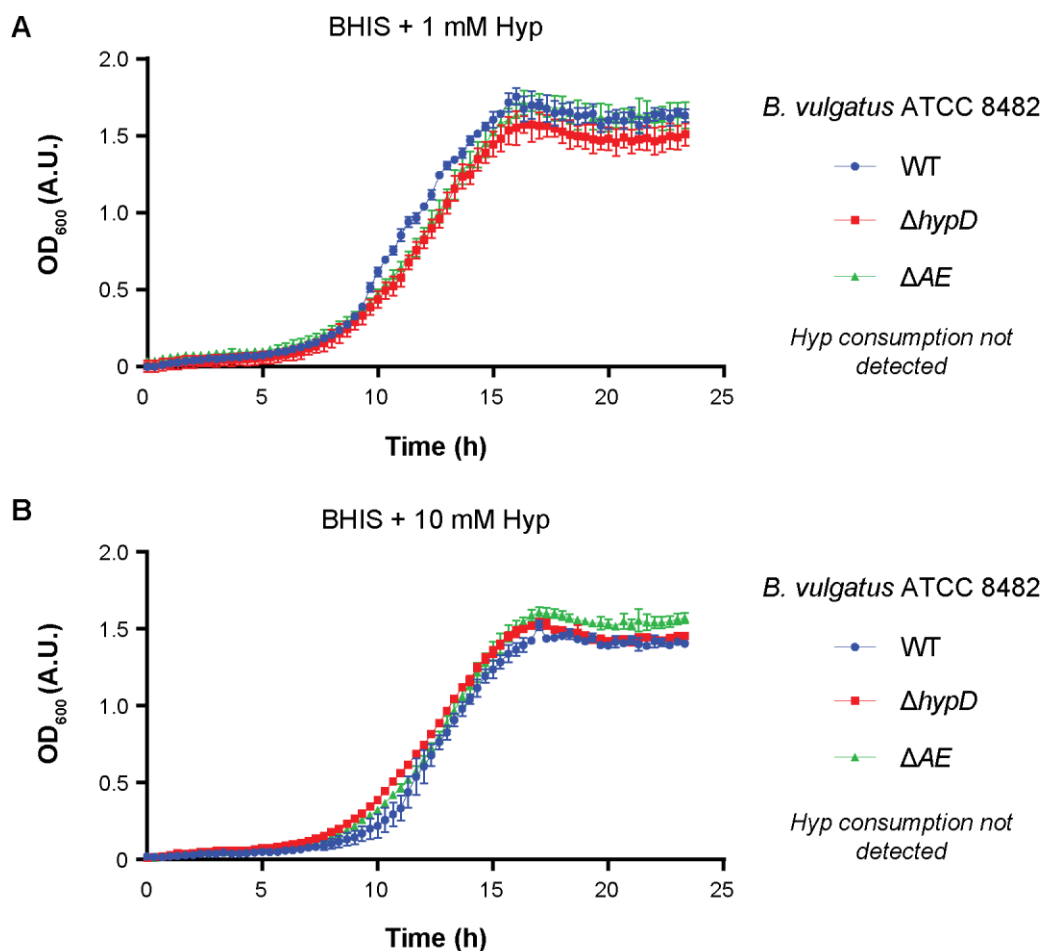


Figure 5.15 Growth of *B. vulgatus* ATCC 8482 WT, $\Delta hypD$, and ΔAE strains in BHIS.

Three *B. vulgatus* ATCC 8482 strains were grown in the rich medium BHIS supplemented with (A) 1 mM or (B) 10 mM Hyp. All strains exhibited similar growth in both media conditions. Absorbance at 600 nm was continuously measured during growth and normalized to the corresponding media controls. Data points are shown as mean $\pm$ SD with $n = 3$ replicates from a single experiment. Hyp was quantified in spent media by LC-MS/MS. No significant decrease in Hyp was detected in any culture condition.

In summary, we were unable to obtain evidence to support an *in vivo* activity for HypD in Bacteroidales from either culture-based experiments reported in Sections 5.2.6 or the experiments described in this section. However, we cannot exclude the possibility that Hyp was consumed at a level below the detection limit of the MS/MS method. Since Bacteroidales are not known to carry out Hyp fermentation as part of energy generating pathways, this amino acid likely has distinct function(s) such as providing substrates for protein synthesis. Therefore, Hyp metabolism is not expected to occur to a high degree if the pathways involved are not growth-limiting under the conditions tested. An additional major caveat of these experiments is a lack of transcriptional data. HypD might not be expressed during Bacteroidales growth on M9, which would explain the lack of a phenotype and Hyp degradation. Presence of *hypD* transcripts could be detected by reverse transcription polymerase chain reaction (RT-PCR) investigate this possibility. It is likely that regulation of HypD in Bacteroidales involves a mechanism distinct from Clostridiales, in which free Hyp was demonstrated to be a transcriptional activator (described in Section 5.2.4). In the conclusion section, we provide recommendations for experiments that may direct future efforts in connecting HypD activity to its physiological role in gut Bacteroidales.

5.2.9 Competitive colonization of germ-free mice by *B. vulgatus* WT and Δ *hypD* strains^v

Due to a lack of insight regarding the physiological activity of HypD from monoculture experiments, we wondered if Hyp metabolism might be important for host colonization. HypD activity could potentially confer a fitness advantage to the WT strain during competition with the Δ *hypD* mutant. Prof. Comstock tested this hypothesis in a preliminary experiment involving co-

^v All experiments described in this section were designed and conducted by Prof. Laurie Comstock at Harvard Medical School.

colonization of both strains in germ-free Swiss Webster mice. The two *B. vulgatus* ATCC 8482 strains were administered to germ-free mice via oral gavage at a starting ratio of 17:83 (WT to $\Delta hypD$). After one week of colonization, fecal samples were collected from all three mice and plated to obtain colonies. Strain types were quantified by colony PCR and % WT from each mouse was determined to be 12.5%, 14.6%, and 8.3%. These values show that WT did not out compete $\Delta hypD$ and appears to have decreased in population size relative to the mutant strain. Overall, both strains successfully colonized germ-free mice and did not outcompete each other within the duration of the experiment.

A lack of detectable difference in fitness between these strains might suggest either low quantities or an absence of Hyp-containing substrates accessible by *B. vulgatus* in the mouse gut. Once an understanding of the transcriptional regulation of *hypD* in *Bacteroides* is achieved, the corresponding substrates can be administered through drinking water or diet in future studies. One conclusion that could be drawn from this experiment is that host-derived peptides containing Hyp are not a major source of nutrient for *B. vulgatus*. It is also possible that proteolysis of these peptides requires enzymatic actions from other microbes that were absent in the germ-free mice. By identifying the source of Hyp for Bacteroidetes, we will be able to better design experiments to include relevant substrates. Plant-derived substrates that are rich in Hyp, including arabinogalactan, lectins, and certain peptide motifs,⁴²⁻⁴⁴ can be supplemented to the mouse diet when repeating this colonization competition experiment. Furthermore, colonization of *B. vulgatus* WT and mutant strains alongside a defined microbial community lacking HypD would better represent the complexity of the gut microbiota. In future experiments, colonization time can be prolonged in addition to taking measurements of strain populations at multiple time points. More rigorous efforts are required to gain insight into the transcriptional regulation of HypD, the source

of Hyp in the gut, and the physiological function of this pathway in Bacteroidetes. We cannot exclude the possibility that HypD is an evolutionary relic and does not serve any purpose in this group of microbes. But it is tempting to discount this based on its prevalence in among Bacteroidales and the high abundance of Hyp in dietary substrates and in the human proteome.

5.3 Conclusions

In summary, we reported studies toward elucidating HypD's role in Stickland fermentation carried out by Clostridiales and understanding HypD function in Bacteroidales. In the first part of this Chapter, we showed that Stickland fermentation of Hyp catalyzed by HypD is upstream of Pro reduction in Clostridiales that can utilize both Hyp and Pro as electron acceptors. Furthermore, we demonstrated that HypD is essential for Hyp utilization in *C. difficile* using a $\Delta hypD$ mutant. The deletion mutant will be useful for the study of this pathway's impact on host infection and disease outcome by *C. difficile*.

In non-Clostridiales, we observed Hyp dehydration activity in *in vitro* assays of HypD purified from the human commensal *B. vulgatus* ATCC 8482. In contrast, no Hyp metabolism was detected in cultures of several HypD-encoding *Bacteroides* or *Parabacteroides* strains. Since downstream transformations of P5C remain unknown in these microbes, we propose using stable-isotope labeled Hyp to identify metabolites enriched in the corresponding isotope. This could help elucidate pathways downstream of Hyp dehydration. In addition, repeating growth experiments with labeled Hyp will improve MS/MS method sensitivity by detecting product formation instead of substrate consumption.

Future efforts should focus on finding a culture condition or a substrate that induces HypD expression in *Bacteroides* species since a lack of expression may explain the lack of activity observed in the assays reported in this Chapter. Specifically, a panel of Hyp-containing substrates

can be used to supplement a defined medium, and *hypD* transcription can be measured with quantitative RT-PCR (RT-qPCR). Hyp-rich glycopeptides would be particularly interesting to test since Bacteroidetes are prolific metabolizers of complex, plant-derived polysaccharides. In addition to transcriptional regulation of *hypD*, these larger substrates would provide additional nutrients to support or enhance bacterial growth. Altogether, these proposed experiments will hopefully help to elucidate inducers of HypD expression and the physiological role of this enzyme in Bacteroidales. Despite the work reported here, many questions remain to be answered concerning this topic.

As described in Chapters 2 and 4, Hyp metabolism was found to be prominent and conserved among human gut microbiomes, leading us to believe this is an important metabolic function in this habitat. Here, we discuss in detail the possible roles of this pathway and implications of this activity in shaping microbiome compositions and human health. We also highlight exciting future directions for exploring the hypotheses proposed throughout the remaining sections.

5.3.1 Anaerobic Hyp metabolism by gut microbes

The conditions under which Hyp metabolism occurs in the human gut are not clear. However, metabolomics experiments in animal models provide evidence that gut microbes do metabolize this amino acid.⁴⁵⁻⁴⁷ Specifically, free Hyp and Hyp-containing peptides (Pro-Hyp, Gly-Pro-Hyp) were detected at higher levels in the urine, fecal, and colonic contents of antibiotic-treated or germ-free rodents compared to control groups with undisturbed gut microbiotas.⁴⁵⁻⁴⁷ Other studies showed that microbial mixtures grown from inoculated human fecal and gingival samples can degrade Hyp anaerobically.^{48,49} Collectively, data from these studies suggest that

significant amounts of free Hyp and Hyp-containing peptides reach the colon and that the gut microbiota plays a role in Hyp metabolism.

Major sources of Hyp for gut microbes include gelatin-containing products, meats, and vegetables from the diet as well as turnover of host collagen (Figure 5.16). Intriguingly, collagen itself is often recommended as a dietary supplement, and there is some evidence that collagen-derived peptides have beneficial effects on host cells, including antioxidant, anticancer, and proliferative activities.⁵⁰ Collagen-derived peptides have been observed to accumulate in plasma after ingestion of collagen or gelatin.⁵⁰ These peptides are rich in Pro and Hyp residues, which introduce conformational restraints and confer resistance to proteolysis by most host enzymes.^{50,51} Microbial metabolism of these host-recalcitrant peptides would impact their availability for the host. In addition, we propose that the gut microbiota could complement host protein metabolism by hydrolyzing these Pro/Hyp-rich short peptides to release free amino acids for further use by both microbes and host (Figure 5.16). This process would parallel the well-studied roles of gut microbes in metabolizing indigestible polysaccharides. Therefore, this may be another example of the gut microbiome improving bioavailability of nutrients in the host.

In addition to collagen-derived peptides from meat products, Hyp-rich glycopeptides or glycoproteins from plants are likely to be abundant in the gut as well (Figure 5.16). Bacteroidetes are known for their ability to catabolize complex dietary and host-derived polysaccharides, which are often covalently linked to Hyp residues in plant glycoproteins.⁵² Given that HypD is present in many common Bacteroidales, some of these species could release free Hyp from Hyp-rich glycoproteins through proteases and hydrolases to further metabolize this amino acid using HypD. Although much research has focused on enzymes that break down the glycan portion of glycoproteins,^{38,52} little is known about enzymes that cleavage glycosidic bonds between peptides

and polysaccharides. The discovery of gut bacterial enzymes that cleave amide bonds in Hyp/Pro-rich peptides and *O*-glycosidic linkages in Hyp-rich glycoproteins should be a priority for future investigations. Access to even small glycopeptides requires laborious extractions from plant materials or non-trivial chemoenzymatic syntheses and thus obtaining pure substrates is extremely difficult. These challenges have greatly inhibited structural elucidation of glycosylated proteins and thereby discovery of enzymes involved in their degradation. Advances in the characterization of their biosynthetic enzymes will help to identify glycosidic linkages in these complex substrates. Biosynthetic enzymes in combination with new synthetic methods will help to access some of these carbohydrates.

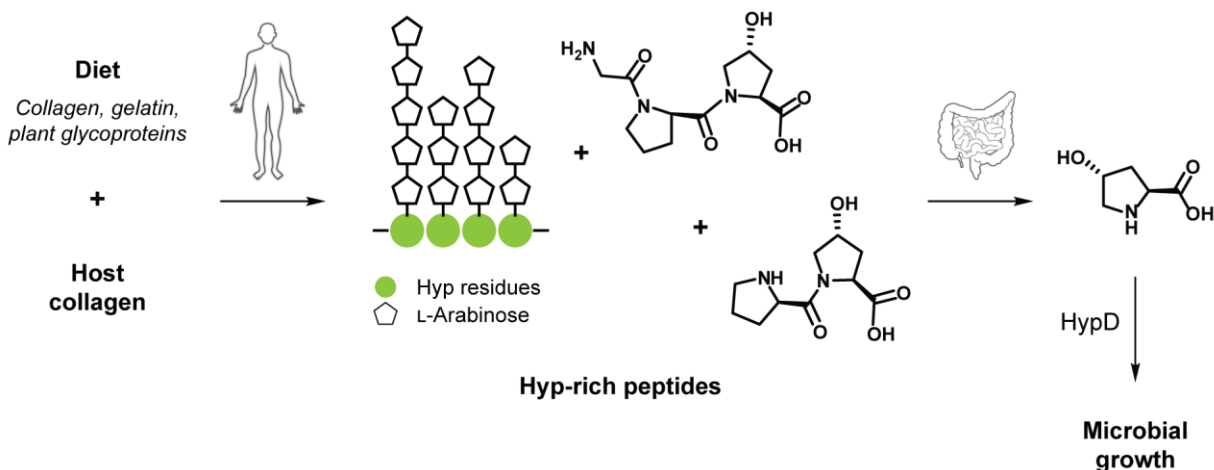


Figure 5.16 Proposed scheme for Hyp metabolism by the gut microbiota.

Sources of Hyp in the human gut could originate from dietary components or endogenous collagen turnover. Major collagen-derived peptides and Hyp repeats from extensin, a plant cell wall glycoprotein, are shown as examples. Short peptides recalcitrant to host peptidase activities may be hydrolyzed by gut microbial peptidases to release free Hyp. HypD-encoding microbes can then access this abundant free amino acid for growth and colonization.

5.3.2 Hyp utilization may contribute to shaping the microbiota composition

Increased Hyp uptake from the diet or increased turnover of endogenous collagen may promote the growth and expansion of gut microbes possessing HypD. Because the product of HypD, P5C, can be metabolized in several ways, we expect that the fitness advantages conferred upon HypD-encoding species depends on the overall nutrient landscape of the gut environment. For example, Stickland fermentation of Hyp is an energy-yielding process and could provide a significant growth advantage to Clostridiales over other HypD-encoding species when alternative

electron acceptors are scarce. In the context of pathogen expansion, consumption of Hyp by gut commensals could keep the levels of this amino acid low and help to establish colonization resistance against HypD-encoding pathogens. The metabolic pathways in which HypD participates need to be identified in order to better understand the importance of Hyp metabolism under different conditions for various microbes.

5.3.3 Stickland fermentation of Hyp during *C. difficile* colonization

In addition to supporting the growth of commensal microbes, Hyp metabolism may be important for colonization of opportunistic pathogens and toxin production. For example, *C. difficile* can utilize both Hyp and Pro as electron acceptors.³ Although the regulation of toxin production in *C. difficile* is not well understood, fermentation of Pro and other amino acids is thought to indirectly influence this process by modulating the intracellular NADH/NAD⁺ ratio.³¹ Since metabolism of Hyp to Pro generates NAD⁺, this transformation may also affect virulence in *C. difficile*. Recently, metabolomics and metatranscriptomics experiments in mouse models for *C. difficile* infection have concluded that Stickland pathways are critical for host colonization.^{8,9,53} Of particular interest, Pro reduction was highly ranked among the pathways thought to be important for *C. difficile* colonization.⁸ Hyp and Pro were efficiently metabolized while 5-aminovalerate levels increased in antibiotic-treated mice upon *C. difficile* infection.^{8,9} Notably, increased transcription of *hypD* was observed early during colonization, supporting HypD's role in facilitating access to Hyp as an electron acceptor.⁹ Many studies have indicated that a diverse gut microbiota can confer colonization resistance against *C. difficile*.⁵⁴ One mechanism by which this may occur is through restricting the metabolic niches available for the pathogen.^{8,9} We propose that HypD-encoding commensals in the healthy human gut may consume Hyp, reducing its availability to *C. difficile*. Disruption of the gut microbiota with antibiotics could reduce Hyp

metabolism, increasing concentrations of this key substrate. This hypothesis can be tested by studying colonization resistance in mice using a defined microbiota containing or lacking HypD. Alternatively, deletion mutants of *C. difficile* in both Hyp and Pro reduction can be tested for differences in colonization efficiency and virulence compared to wild-type strains. Understanding the role of Hyp utilization in *C. difficile* could have implications for efforts to develop defined microbial consortia as therapeutics in patients with recurrent *C. difficile* infection.

5.3.4 Breakdown of host collagen to access Hyp

It is likely that gut microbial acquisition of Hyp from host collagen could also affect host biology. Collagen is the principle source of Hyp in the human body and a major component of the extracellular matrix in connective tissues, including the basement membrane underlying the gut epithelium.⁵⁵ In the context of various diseases, the host inflammatory response and certain bacterial pathogens can degrade collagen, resulting in a breach of this barrier.⁵⁶ The resulting release of peptides and amino acids could increase nutrient availability for microbes and potentially alter the composition and functions of the microbiota. The most common amino acid repeat making up collagen is Pro-Hyp-Gly as previously discussed.⁵⁵ Therefore, bacterial degradation of collagen would release large amounts of Pro, Hyp, and Gly, which are common electron acceptors in Stickland fermentation. Clostridiales are well-studied for their proteolytic properties and perhaps co-evolved the unique ability to access these abundant amino acids for energy metabolism. Even though only a few microbial collagenases have been biochemically characterized, they are thought to be virulence factors important for diseases such as periodontitis.^{57,58} Identification and characterization of collagenases from gut microbes will enable efforts to understand the contribution of collagen degradation in diseases involving impaired gut barrier function, including inflammatory bowel disease. However, extraction of single collagen

types from animal tissues that maintain their native fold is difficult to achieve, which has drastically limited progress in this field.

5.3.5 The biological activity of the fermentation product 5-aminovalerate may influence host pathways

In addition to biological implications of the upstream pathways affecting Hyp metabolism, the end products of Hyp degradation may also interact with host cells. Notably, 5-aminovalerate, a known product of Stickland fermentations involving Hyp and Pro, has been detected in the cecum and feces of mice, suggesting that Pro reduction is an active pathway in the gut microbiota.^{45,47,59} Although an alternative bacterial pathway, L-lysine catabolism, is known to generate 5-aminovalerate, this is not expected to be a source of this metabolite in the anaerobic environment of the gut. Lys catabolism was first described in *Pseudomonas* species and involves an oxidative decarboxylation as its first step.^{60,61} This transformation is dependent on molecular oxygen as a co-substrate and is catalyzed by lysine monooxygenase.⁶⁰ Therefore, 5-aminovalerate is likely uniquely derived from Stickland fermentation of Hyp and Pro in the human gut.

Intriguingly, 5-aminovalerate is a chemical analog of γ -aminobutyric acid (GABA) and a weak antagonist of GABA_B receptor.⁶² These inhibitory receptors are activated upon GABA binding. In animals, GABA_B receptors are expressed in the enteric nervous system and play a role in regulating gastrointestinal motility.⁶³ They are also expressed in immune cells, and recent studies have linked GABAergic signaling to inflammation.⁶³ The prominence of Hyp metabolism in the gut raises the intriguing possibility that HypD-mediated Stickland fermentation may contribute significantly to 5-aminovalerate production, potentially altering gut motor and immune functions.

Overall, the many exciting research directions highlighted throughout the previous sections illustrate the importance of incorporating enzyme discovery efforts into microbiota research. We anticipate that future investigations of these hypotheses will reveal a significant contribution of anaerobic Hyp metabolism in shaping gut microbiota composition and influencing host physiological and pathological states.

5.4 Materials and methods

All chemicals, solvents, and reagents were purchased from vendors as previously detailed in Chapter 2's Materials and Methods section with some modifications. *Bacteroides* and *Parabacteroides* species were obtained from ATCC, DSMZ, or Prof. Laurie Comstock at Harvard Medical School. *E. coli* S17-1 λ pir, pBlueScript SK, pKNOCK-*bla-ermGb*, and pFD340 were generously provided by Prof. Laurie Comstock. *C. difficile* 630 Δ *erm* Δ *pyrE*, *E. coli* HB101/pRK24, pMTL-YN3, and pMTL-YN1C were generously provided by Prof. Aimee Shen at Tufts University. BBL™ Brain Heart Infusion (BHI) broth was purchased from BD Difco. Primers were purchased from Sigma-Aldrich. Unless otherwise stated, PCRs were carried out using Phusion-HF PCR Master Mix according to the manufacturer's protocol (NEB). The identities of all plasmids were confirmed by DNA sequencing (Eton Bioscience Inc.). All absorbance measurements in 96-well plates were carried out using a PowerWave HT Microplate Spectrophotometer (Biotek) inside of an anaerobic chamber (MBraun) for kinetic assays. Samples were made anaerobic as described previously in Chapter 2. Except autoclaved media solutions were either transferred into an anaerobic chamber to pre-reduce overnight prior to use or made anaerobic by sparging with nitrogen for 1 h under a sterile environment.

5.4.1 Growth experiments and metabolite analyses for *C. difficile* 630 Δ erm and *T. glycolicus* DSM 1288

Components of the defined medium MACC used to culture *C. difficile* and *T. glycolicus* were listed in Chapter 2. Briefly, this medium is a phosphate- and carbonate-based medium that does not contain proline nor 4-hydroxyproline. MACC-Pro was supplemented with 20 mM L-proline and MACC-Hyp was supplemented with 20 mM Hyp. All three MACC media formulations were supplemented with 20 mM glucose.

A single colony of each strain was inoculated into 5 mL RCM broth and incubated at 37 °C overnight. These starter cultures were diluted 1:100 into 200 μ L cultures for growth in 96-well plates. OD600 measurements were taken every 20 minutes using a PowerWave HT Microplate Spectrophotometer (BioTek). Samples were collected at various time points for metabolite quantification by liquid chromatography tandem-mass spectrometry (LC-MS/MS). 5 μ L of each culture was diluted into 195 μ L of methanol, which was vortexed and clarified by centrifugation (15,200 g for 10 min). Clarified supernatants were diluted 40-fold with water for detection of Hyp and 5-aminovalerate for a total dilution of 1,600-fold. Supernatants were further diluted 10-fold for the detection of Pro, resulting in a total dilution of 16,000-fold. The injection volume for all diluted samples was 3 μ L.

LC-MS/MS analyses were performed on an Agilent 6410 Triple Quadrupole LC-MS instrument (Agilent Technologies) as described in Chapter 2. The column and detection method used for proline were described previously.¹ For detection of 4-hydroxyproline and 5-aminovalerate, the previous method was modified as follows.¹ The collision energy was set to 13 eV for 4-hydroxyproline and 8 eV for 5-aminovalerate. Both metabolites were monitored using the same gradient flow with 200 mM ammonium acetate as solvent A and 5% acetic acid as

solvent B. A flow rate of 0.4 mL min⁻¹ was maintained for 9 min for each run. The LC conditions were: a gradient of 95 to 85% B for 4 min, 85% B for 1.5 min, a gradient increasing to 95% B for 0.01 min, and 95% B for 3.49 min. Precursor and product ions of m/z 132.1 $\rightarrow$ m/z 86.1 were monitored for 4-hydroxyproline and m/z 118.1 $\rightarrow$ m/z 101.1 were monitored for 5-aminovalerate. On the Agilent MassHunter Workstation Data Acquisition software, the MS¹ and MS² resolutions were set to “unit” for 5-aminovalerate. Blank MACC media supplemented with 20 mM of proline, 4-hydroxyproline, or 5-aminovalerate were included on culture plates as external standards. Sample peak areas were normalized to the corresponding standard peak areas. All growth conditions and metabolite quantifications were carried out in replicates of 5.

5.4.2 Construction of pMTL-YN3 plasmids for genetic deletions in *C. difficile*

630 Δ erm Δ pyrE

gDNA isolated from *C. difficile* 630 Δ erm was used as template for PCR amplifications. Flanking regions 5'- and 3'- of genes encoding HypD (CD630_32820) and P5CR(CD630_32810) were amplified to obtain ~1 kb genomic fragments using primers listed in Table 5.7. Annealing temperatures used for each primer pair are listed in Table 5.7. PCR products were purified from a 1% agarose gel.

The flanking regions of each target gene were joined using splicing by overlap extension (SOE) PCR. The overlap PCR step was performed using 30 ng of 5'-region and 30 ng of 3'-region for each gene in final volumes of 10 μ L. Thermocycling was carried out with an initial denaturation period of 30 s at 98 °C, followed by 15 cycles of 30 s at 98 °C, 150 s at 72 °C, and with a final extension time of 2 min at 72 °C. Extension PCR was then carried out using 7 μ L of overlap PCR mixtures as template DNA in final volumes of 20 μ L. Primers 5'-CdHypD-Fwd and 3'-CdHypD-Rev were used to amplify ligated flanking regions for *hypD*. Primers 5'-CdP5CR-Fwd and 3'-

CdP5CR-Rev were used to amplify ligated flanking regions for *proC*. Each primer was added 0.75 μ M for each PCR mixture to obtain ~2 kb insert fragments. Thermocycling for extension PCR was carried out with an initial denaturation period of 30 s at 98 °C, followed by 30 cycles of 30 s at 98 °C, 30 s at 61 °C, and 60 s at 72 °C, with a final extension time of 2 min at 72 °C.

The resulting PCR products were purified from agarose gel and digested with *AscI* and *SbfI*-HF at 37 °C. Vector pMTL-YN3 was linearized by digestions with *AscI* and *SbfI*-HF at 37 °C followed by gel-purification. Each insert fragment was cloned into linearized pMTL-YN3 using T4 DNA ligase in 10 μ L reactions at a 1:9 volume ratio (vector:insert). Entire ligation reactions were transformed into chemically competent *E. coli* TOP10 cells and plated onto LB-Cm25 agar. The plasmids generated, pMTL-YN3- Δ HypD and pMTL-YN3- Δ P5CR, were verified by sequencing of insert regions. Plasmids were transformed into chemically competent *E. coli* HB101/pRK24 for conjugation with *C. difficile* 630 Δ erm Δ pyrE.

Table 5.7 Primers used for cloning in this study.

Primer	Description	Sequence (5'–3')	Annealing Temperature (°C)
pMTL-YN3 vectors for deletion in <i>C. difficile</i> 630ΔermΔpyrE			
5'-CdHypD-F	~1kb region 5' of CdHypD (CD630_32820)	AATGGCGCGCCTGGAATGGATATT ATTTAGGGGTG	61
5'-CdHypD-R	~1kb region 5' of CdHypD (CD630_32820)	CTATTATCTCATCTTGTAAGTTCT ACTGTGTGGTTCTGCATTAATACTT TC	
3'-CdHypD-F	~1kb region 3' of CdHypD (CD630_32820)	GAAAGTATTAATGCAGAACCACAC AGTAGAACACTACAAGATGAGATA ATAG	61
3'-CdHypD-R	~1kb region 3' of CdHypD (CD630_32820)	AATCCTGCAGGACCCCATTAAGTGC TTGAG	

Table 5.7 (Continued)

5'- CdP5CR- F	~1kb region 5' of CdP5CR (CD630_32810)	AATGGCGCGCCCGATATGGATACA TTGCTTAAAGC	61
5'- CdP5CR- R	~1kb region 5' of CdP5CR (CD630_32810)	CTCATGTCTTTAGACTTATCTATAC AAGCCCCTGAACCAATAAATCCTA AAG	61
3'- CdP5CR- F	~1kb region 3' of CdP5CR (CD630_32810)	CTTTAGGATTTATTGGTTCAGGGG CTTGTATAGATAAGTCTAAAGACA TGAG	61
3'- CdP5CR- R	~1kb region 3' of CdP5CR (CD630_32810)	AATCCTGCAGGCCGCCATTTGGAT AAAGTAC	61
scr- CdHypD- F	cPCR screen, internal region of CdHypD (CD630_32820)	AGCAGTACTTATGACAGAAGC	60
scr- CdHypD- R	cPCR screen, internal region of CdHypD (CD630_32820)	GGTCACTATATCCTGCAAC	60
scr- CdP5CR- F	cPCR screen, internal region of CdP5CR (CD630_32810)	CTGCTGCTCTAAAGCCTTC	60
scr- CdP5CR- R	cPCR screen, internal region of CdP5CR (CD630_32810)	GTTCCCTGCTGGAGAACAAAC	60
pMTL-YN1C vectors for complementation in <i>C. difficile</i> 630ΔermΔpyrE			
5'-CdAE- F	~330 bp 5' of CdAE (CD630_32830)	CATGGAGATCTCGAGGAGATAGAG ACTACATGGTAAAGTGTAG	60
5'-CdAE- R	~330 bp 5' of CdAE (CD630_32830)	CTTGCCATGATTTTTCACCCCTAAA TAATATC	60
CdHypD- F	Full-length CdHypD (CD630_32820)	GTGAAAAATCATGGCAAGAGGAA CTTTTG	60
CdHypD- R	Full-length CdHypD (CD630_32820)	CTTGCATGTCTGCAGGCTAGAATG TTTGCTCAGTTCTTC	60
5'_CdP5 CR-F	~230 bp upstream + CdP5CR (CD630_32820)	CATGGAGATCTCGAGGAGATTAAT TATGAGCTAACTTTCTAAATAG	60
5'_CdP5 CR-R	~230 bp upstream + CdP5CR (CD630_32820)	CTTGCATGTCTGCAGGTTATTTACT CATGTCTTTAGACTTATC	60

pTrcHis2A expression vectors for HypDAE regions

Table 5.7 (Continued)

CdHypD AE-F	Gibson assembly of HMPREF9945_00949 +00950 into pTrcHis2A	TTAAATAAGGAGGAATAAACCATG AATCCATTAGTTATAAACTTACAA AAATGTAGCATACATG	60
CdHypD AE-R	Gibson assembly of HMPREF9945_00949 +00950 into pTrcHis2A	GCAGATCTCGAGCTCGGATCCTAG AATGTTTGCTCAGTTCTTCCTATTA TCTC	60
pTrcHis- Cd-F	Gibson assembly of HMPREF9945_00949 +00950 into pTrcHis2A	GTTTATTCCTCCTTATTTAATCGAT ACATTAATATATACC	56
pTrcHis- Cd-R	Gibson assembly of HMPREF9945_00949 +00950 into pTrcHis2A	CATGGATCCGAGCTCGAGATCTGC	56
BvHypD AE-F	Gibson assembly of Bvu_1420+1421 into pTrcHis2A	AATAAGGAGGAATAAACCATGAA AACAACAATCAACG	57
BvHypD AE-R	Gibson assembly of Bvu_1420+1421 into pTrcHis2A	CGAGCTCGGATCCATTTATCCACC TATGGTTGC	57
pTrcHis- Bv-F	Gibson assembly of Bvu_1420+1421 into pTrcHis2A	ATGGATCCGAGCTCGAGATC	57
pTrcHis- Bv-R	Gibson assembly of Bvu_1420+1421 into pTrcHis2A	GGTTTATTCCTCCTTATTTAATC	57

Overexpression vectors for BvHypD and BvAE

BvHypD- F	Gibson assembly of Bvu_1421 into pET28a(+)	CCGCGCGGCAGCCATATGAAAACA ACAATCAACG	57
BvHypD- R	Gibson assembly of Bvu_1421 into pET28a(+)	CAGTCATGCTAGCCATTCAGAAGG TTTCTTGTGC	57
pET28a- F	Gibson assembly of Bvu_1421 into pET28a(+)	ATGGCTAGCATGACTGGTGGACAG C	68
pET28a- R	Gibson assembly of Bvu_1421 into pET28a(+)	ATGGCTGCCGCGCGGCAC	68

Table 5.7 (Continued)

BvAE-F	Gibson assembly of Bvu_1422 into pSV272-PfMBP	TTTCAGGGCGCCCATATGAGTCTG ATTTTCGAC	57
BvAE-R	Gibson assembly of Bvu_1422 into pSV272-PfMBP	TCTCGAGTGC GGCCGTTATCCACC TATGGTTGC	57
pSV272-F	Gibson assembly of Bvu_1422 into pSV272-PfMBP	CGGCCGCACTCGAGATCA	68
pSV272-R	Gibson assembly of Bvu_1422 into pSV272-PfMBP	ATGGGCGCCCTGAAAATAC	68

Expression vectors for BvHypDAE in *Bacteroides* spp.

BvHypD AE-F	Gibson assembly of Bvu_1421+1420 into pFD340	AAATGGATCCATTAATCCGCCCTT ATGGAAAT	60
BvHypD AE-R	Gibson assembly of Bvu_1421+1420 into pFD340	ATGAGGATCCGGTCCTTTTACTAC AGAACGATGA	60

pBlueScript vectors for insert fragment screens

5'- BvHypD-F	Gibson Assembly to generate pBlueScript- ΔHypD, Bvu_1421	GCGGCCGCCACCGCGGTGGAGCTG AGGATTTTACGCAGTCTG	56
5'- BvHypD-R	Gibson Assembly to generate pBlueScript- ΔHypD, Bvu_1421	CATATTGTAAAGTCTGTCATACCG TTGATTG	56
3'- BvHypD-F	Gibson Assembly to generate pBlueScript- ΔHypD, Bvu_1421	TATGACAGACTTTACAATATGACG TAATAGCCC	60
3'- BvHypD-R	Gibson Assembly to generate pBlueScript- ΔHypD, Bvu_1421	AGGGAACAAAAGCTGGAGCTACTC CTTCTTTGGAACGATTATG	60
5'-BvAE-F	Gibson Assembly to generate pBlueScript- ΔAE, Bvu_1420	GCGGCCGCCACCGCGGTGGAGCTA AAGAATGGAAAGAAGCTTACG	60
5'-BvAE-R	Gibson Assembly to generate pBlueScript- ΔAE, Bvu_1420	GCACTGCTGTTGTCATAGCGTTTG ATGTCGAAAATC	60
3'-BvAE-F	Gibson Assembly to generate pBlueScript- ΔAE, Bvu_1420	AAACGCTATGACAACAGCAGTGCA TACAAATC	70

Table 5.7 (Continued)

3'-BvAE-R	Gibson Assembly to generate pBlueScript- Δ AE, Bvu_1420	AGGGAACAAAAGCTGGAGCTATG AATACCCCAACCGTTTAT	70
pBlueScr ipt-F	Gibson Assembly to generate vectors for Δ HypD and Δ AE	CCACCGCGGTGGCG	60
pBlueScr ipt-R	Gibson Assembly to generate vectors for Δ HypD and Δ AE	AGCTCCAGCTTTTGTTCCTTTAGT G	60
pKNOCK-<i>bla-erm</i>Gb vectors for deletions in <i>B. vulgatus</i> ATCC 8482			
BvHypD-del-F	Gibson Assembly to generate pKNOCK- <i>bla-erm</i> Gb- Δ HypD	AGAACTAGTGGATCCCAGCTGAGG ATTTTACGCAG	60
BvHypD-del-R	Gibson Assembly to generate pKNOCK- <i>bla-erm</i> Gb- Δ HypD	ATTCTGCAGCCCGGAGCTACTCC TTCTTTGGAAC	60
BvAE-del-F	Gibson Assembly to generate pKNOCK- <i>bla-erm</i> Gb- Δ AE	AGAACTAGTGGATCCCAGCTAAAG AATGGAAAGAAGC	60
BvAE-del-R	Gibson Assembly to generate pKNOCK- <i>bla-erm</i> Gb- Δ AE	ATTCTGCAGCCCGGAGCTATGAA TACCCCAA	60
pKNOC K- <i>bla-erm</i> Gb-F	Gibson Assembly to generate vectors for Δ HypD and Δ AE	CCGGGCTGCAGGAATTCTG	67
pKNOC K- <i>bla-erm</i> Gb-R	Gibson Assembly to generate vectors for Δ HypD and Δ AE	GGGATCCACTAGTTCTAGAGCG	67
scr-BvHypD-F	cPCR screen for HypD (Bvu_1421) deletion	ACAAACAGTAATTTTCGGCCATTAT	
scr-BvHypD-R	cPCR screen for HypD (Bvu_1421) deletion	GGGACATACTTTCAGACAAGTTCC	
scr-BvAE-F	cPCR screen for AE (Bvu_1420) deletion	AACATGTTGTCTACCACCTGTCAC	
scr-BvAE-R	cPCR screen for AE (Bvu_1420) deletion	GGGTACGAAGGTCCTTTTACTACA	

5.4.3 Construction of pMTL-YN1C plasmids for genetic complementation in

C. difficile 630 Δ erm Δ pyrE

gDNA isolated from *C. difficile* 630 Δ erm was used as template for PCR amplifications. Upstream regions predicted to encode the promoter for genes encoding HypD and P5CR were amplified for vector construction in addition to the full-length genes. A region ~330 bp upstream of CdAE (locus tag: CD630_32830) was amplified to be ligated to *hypD*. The full-length *hypD* (CD630_32820) and *proC* (CD630_32810), including the upstream ~230 bp were amplified by PCR. Primers and annealing temperatures are listed in Table 5.7. PCR products were gel-purified. pMTL-YN1C was linearized through a restriction digest with StuI and purified.

Promoter region of *hypD* and *hypD* inserts were cloned into pMTL-YN1C to generate pMTL-YN1C-HypD. Insert fragment containing promoter region and *proC* was cloned into pMTL-YN1C to generate pMTL-YN1C-P5CR. Gibson Assembly reactions contained 50 ng of linearized pMTL-YN1C to construct pMTL-YN1C-HypD and 100 ng of vector to construct pMTL-YN1C-P5CR. 3-fold molar excess of each insert was included in reaction volumes of 10 μ L. Gibson reactions were incubated at 50 °C for 1 h in a thermocycler. 5 μ L of each reaction was transformed into chemically competent *E. coli* TOP10 and plated onto LB-Cm25 agar. Colonies were picked for plasmid purification, which were confirmed by sequencing.

5.4.4 Construction of deletion mutants *C. difficile* 630 Δ erm Δ hypD and *C. difficile*

630 Δ erm Δ proC^{vi}

^{vi} Construction of deletion mutants in *C. difficile* was conducted together with Prof. Aimee Shen in the Shen lab at Tufts University. Complementation of *pyrE* in *C. difficile* deletion mutants was conducted by Prof. Aimee Shen.

All *C. difficile* strains were grown on RCM-agar plates supplemented with 0.1% w/v taurocholate (TA) when inoculated from glycerol stocks. Gene deletions in *C. difficile* 630 Δ erm Δ pyrE were achieved through an allelic exchange system using a previously described protocol.⁶⁴ Briefly, conjugation of pMTL-YN3- Δ HypD and pMTL-YN3- Δ P5CR into *C. difficile* 630 Δ erm Δ pyrE was carried out using *E. coli* HB101/pRK24. 2.5 mL of the *E. coli* HB101 donor strain in mid- to late-exponential phase in LB-Amp50-Cm20 was pelleted by centrifugation at 2,500 rpm for 5 min. In the anaerobic chamber, 1 mL of turbid *C. difficile* culture grown for 5–6 hours in BHIS (BHI supplemented with 5 g L⁻¹ yeast extract and 0.1% cysteine) was used to resuspend each *E. coli* donor pellet. Seven 100 μ L aliquots were spotted onto BHIS-agar plates and grown overnight at 37 °C. Cell mass from each conjugation culture was entirely resuspended in 1 mL of phosphate-buffered saline (PBS). 100 μ L of each mixture was spread out onto five BHIS-agar plates containing 10 μ g mL⁻¹ thiamphenicol, 50 μ g mL⁻¹ kanamycin, and 16 μ g mL⁻¹ cefoxitin. Transconjugants appeared two days later and were passaged onto BHIS-agar plates containing 15 μ g mL⁻¹ thiamphenicol, 50 μ g mL⁻¹ kanamycin, 16 μ g mL⁻¹ cefoxitin, and 5 μ g mL⁻¹ uracil. The quickest growing colonies of each strain were passaged 2–3 times after which single colonies were streaked onto CDDM-agar plates containing 2 mg mL⁻¹ 5-fluoroorotic acid and 5 μ g mL⁻¹ uracil. Single colonies that appeared were patched onto another CDDM plate for screening by cPCR.

For cPCR, each colony was resuspended into 100 μ L of 1 $\times$ Phusion buffer with 1 μ L Proteinase K (NEB). Colonies were lysed by incubating solutions at 60 °C for 1 h, followed by deactivation of Proteinase K at 95 °C for 15 min. Solutions were clarified by centrifugation. 3 μ L of colony solution supernatant was added as template to 19 μ L of PCR master mix. Primers targeting internal regions of *hypD* and *proC* were designed to screen colonies. Colonies containing

deletions were identified by the absence of an amplification product. A positive control reaction containing gDNA from the WT strain was included in all cPCRs. Overall, *C. difficile* 630 Δ erm Δ pyrE Δ hypD and *C. 630* Δ erm Δ pyrE Δ proC strains were constructed. Each deletion mutant was inoculated into BHIS to make glycerol stocks for long-term storage.

For complementation of *pyrE* into deletion mutants, *E. coli* HB101/pRK24 donor strains transformed with pMTL-YN1C-HypD or pMTL-YN1C-P5CR were grown in LB-Amp50-Cm20 at 37 °C. The conjugation protocol described above was followed. Transconjugants were directly streaked onto CDDM-agar plates. Single colonies were restreaked onto a CDDM-agar plate before inoculation into BHIS to generate glycerol stocks.

5.4.5 Growth experiments and metabolite analyses for *C. difficile* 630 Δ erm WT, Δ hypD, and Δ proC

Each strain was inoculated into 1.5 mL RCM broth and incubated at 37 °C overnight. Starter cultures were normalized by centrifugation and resuspension in MACC basal medium to obtain OD600 values of approximately 1. These cultures were diluted 1:100 into 200 μ L MACC-Hyp in a 96-well plate. Starter cultures were directly diluted 1:100 into 200 μ L BHI for the growth experiment on rich media. Growth was measured as previously described using OD600. Hyp, Pro, and 5-aminovalerate were quantified from all MACC-Hyp cultures by LC-MS/MS as described in Section 5.4.1. Each culture condition had 5 replicates.

5.4.6 RNA-sequencing of *C. difficile* 630 Δ erm in MACC-Pro, MACC-Hyp, and RCM^{vii}

^{vii} RNA extraction, cDNA library construction, sequencing, read alignments, and read count quantifications were conducted by the Broad Institute Microbial Omics Core (MOC) in Cambridge, MA.

Overnight, turbid starter cultures of *C. difficile* 630 Δ *erm* in RCM were inoculated 1.5:100 into 10 mL of MACC-Pro or MAC-Hyp, and 5 mL of RCM for growth at 37 °C. All media conditions were inoculated in triplicate. Cultures reached exponential phase after 24 h with an OD600 range of 0.44–0.46 for MACC-Hyp. Cultures grown in MACC-Pro reached OD600 values of 0.16, 0.40, and 0.41. Cultures grown in RCM reached an OD600 range of 0.31 to 0.40 after 12 h. Cell pellets were harvested by centrifugation, re-suspended in 500 μ L of Trizol reagent (ThermoFisher), and stored at –80 °C prior to RNA isolation. Cells were lysed by bead beating and total RNA was extracted using Zymo Research Direct-Zol RNA MiniPrep Plus kit according to the manufacturer’s protocol. Illumina cDNA libraries were generated using a modified version of the RNAtag-Seq protocol, which is described here.¹¹ Total isolated RNA (500 ng) was first fragmented, depleted of gDNA, and dephosphorylated. Purified RNA was then ligated to DNA adapters carrying 5'-AN8-3' barcodes containing a 5' phosphate and a 3' blocking group to be individually barcoded with respect to each sample. Barcoded RNAs were then pooled and ribosomal RNA was depleted using the RiboZero rRNA depletion kit (Epicentre). Illumina cDNA libraries were generated from pooled RNAs in three main steps. First, reverse transcription of RNA was achieved using a primer targeting the constant region of the barcoded adapter. A second adapter was added to the 3' end of the cDNA during reverse transcription using SmartScribe RT (Clontech) as previously described.⁶⁵ Finally, primers targeting the constant regions of the 5' and 3' adapters were used to PCR amplify the full sequence of each fragment including the Illumina sequencing adapters. The constructed cDNA libraries were sequenced on an Illumina HiSeq 2500.

For RNAtag-Seq data analysis, reads from each sample were grouped based on their barcode using custom scripts. Up to one mismatch in the barcode region was allowed to account for sequencing error, except when this resulted in assignment to more than one barcode sequence.

Barcode sequences were removed from reads, which were then aligned to the *C. difficile* genome (NCBI reference sequence: NC_013315.1) through Burrows-Wheeler transform⁶⁶ using Bowtie2¹². Read counts were tabulated using Rsubread.¹³ Differential expression analysis was carried out using DESeq2.¹⁷ Data analysis and visualization were performed using various packages in RStudio (version 1.0.143).¹⁶ R packages used to generate figures for data visualization and statistical calculations are cited throughout results section.

5.4.7 Construction of pTrcHis2A plasmids for overexpression of HypD-AE gene clusters

gDNA was purified from *C. difficile* 70-100-2010 and *B. vulgatus* ATCC 8482 and used to amplify CdHypDAE and BvHypDAE fragments, respectively. Primer sequences and annealing temperatures used for PCR amplifications are listed in Table 5.7. Insert fragments were purified then cloned into pTrcHis2A vector through Gibson Assembly. The pTrcHis2A vector was PCR amplified using the corresponding primers for each insert. Linearized vectors were treated with DpnI and then purified. Gibson Assembly reactions contained either 50 ng or 100 ng of purified pTrcHis2A with 2-fold molar excess of insert fragment to construct vectors pTrcHis2A-CdHypDAE, and pTrcHis2A-BvHypDAE. A total volume of 10 μ L was used for reactions containing 50 ng of vector and 20 μ L for reactions containing 100 ng of vector. Gibson reactions were incubated at 50 °C for 1 h in a thermocycler. 1 μ L of 50 ng and 2 μ L of 100 ng vector-containing reactions were transformed into chemically competent *E. coli* TOP10 and plated onto LB-Amp100 agar. Colonies were picked for plasmid purification and confirmed by sequencing.

5.4.8 Detection of HypD activity in *E. coli* overexpression strains in BHI

pTrcHis2A, pTrcHis2A-CdHypDAE, and pTrcHis2A-BvHypDAE were transformed into *E. coli* strains TOP10, BL21(DE3), and BL21(DE3)-CodonPlus-RIL. Transformed TOP10 and

BL21(DE3) strains were inoculated into 2 mL BHI-Amp100. Transformed BL21(DE3)-CodonPlus-RIL was inoculated into 2 mL BHI-Amp100-Cm25. All cultures described here were grown anaerobically at 37 °C. Overnight, turbid cultures were diluted 1:100 into 180 µL BHI supplemented with 1 mM Hyp and the corresponding antibiotics. Cultures were inoculated in triplicate in a 96-well plate. After 3 h of growth, protein expression was induced with 100 µM IPTG and cultures were incubated overnight at 37 °C. Hyp was quantified by normalizing area peaks to media blanks using the previously described LC-MS/MS method. Cultures were diluted 75-fold prior to injection on the LC-MS/MS.

5.4.9 Construction of plasmids for overexpression of BvHypD and BvAE in *E. coli*

The genes encoding BvHypD (locus tag: Bvu_1421) and BvAE (locus tag: Bvu_1420) were individually amplified from *B. vulgatus* ATCC 8482 gDNA. Primer sequences used to amplify these genes are shown in Table 5.7. PCRs were carried out using Q5-HF PCR Master Mix according to the manufacturer's protocol using the annealing temperature listed in Table 5.7. The PCR product of BvHypD was purified and ligated into pET28a to yield pET28a-BvHypD containing N-terminal His₆-tagged HypD. PCR product of BvAE was purified and ligated into pSV272-PfMBP⁶⁷ to yield vector pSV-PfMBP-BvAE. The NdeI cleavage site (CATATG) at the 5' start codon of the His₆ tag in pSV272-PfMBP had been mutated to CACATG by site-directed mutagenesis to prevent digestion by NdeI and removal of the His₆-tagged MBP during cloning. Both vectors were constructed using Gibson Assembly. Both empty vectors were linearized through PCR amplification using corresponding primers for each insert. PCR products for vectors were treated with DpnI to remove template DNA and then purified. Gibson Assembly reactions contained either 50 ng or 100 ng of purified vector backbone with 2-fold molar excess of BvHypD or BvAE insert fragment. A total volume of 10 µL was used for reactions containing 50 ng of

vector and 20 μ L for reactions containing 100 ng of vector. Gibson reactions were incubated at 50 °C for 1 h in a thermocycler. 1 μ L of 50 ng and 2 μ L of 100 ng vector-containing reactions were transformed into chemically competent *E. coli* TOP10. Transformed cells were plated out onto LB-Kan50 agar plates, colonies were picked for plasmid purification and confirmed by sequencing. Verified plasmids were transformed into chemically competent *E. coli* BL21-CodonPlus (DE3)-RIL *ΔproC::aac(3)IV* (Apramycin^R or Am^R) for overexpression and purification experiments.

5.4.10 Overexpression and purification of BvHypD, BvAE, and CdP5CR

BvHypD and PfMBP-BvAE were overexpressed in *E. coli* BL21-CodonPlus(DE3)-RIL *ΔproC::aac(3)IV*(Am^R), a strain reported in Chapter 2. Both proteins were purified using the same procedure as previously described in Materials and Methods 2.4.5 for BvHypD and Materials and Methods 2.4.6 for BvAE. Cell pellets from 1 L cultures were harvested for purification. From a single purification, a yield of 8 mg L⁻¹ culture was obtained for BvHypD and a yield of 15 mg L⁻¹ culture was obtained for MBP-BvAE based on protein concentrations determined by Nanodrop. An extinction coefficient of 85,150 M⁻¹ cm⁻¹ for BvHypD and 108,180 M⁻¹ cm⁻¹ for PfMBP-BvAE were calculated using ProtParam tool.⁶⁸

CdP5CR was purified from *E. coli* BL21-CodonPlus(DE3)-RIL *ΔproC::aac(3)IV*(Am^R) as described in Chapter 2 under Materials and Methods 2.4.7.

5.4.11 End-point and kinetic assays for BvHypD

BvHypD was first activated with BvAE using the previously described method for CdHypD in the Chapter 2's Materials and Methods 2.4.12. All concentrations and assay conditions were kept constant except purified BvHypD and BvAE were used. Glycyl radical formation was

not quantified by electron paramagnetic resonance (EPR) spectroscopy. The activated protein solution was directly used for the assays described in this section.

End-point enzymatic assays were carried out as previously described in Chapter 2 Materials and Methods 2.4.14 with the following exceptions. Activated BvHypD was included to detect Hyp dehydration, which was coupled to CdP5CR activity. Aliquots of assay mixtures were removed and quenched for analysis after 1 h and 21 h of incubation. Proline detection was carried out using LC-MS/MS as previously described in Chapter 2. Area peaks from samples were normalized to 200 μ M Pro standards.

The kinetic assay for activated BvHypD was conducted as previously described in Chapter 2's Materials and Methods 2.4.16. Assays were initiated by addition of Hyp to a final concentration of 0, 1, 3, 5, 7, 10, 15, or 20 mM. Kinetic data analysis was performed similarly as described before. However, since activation data were not collected for BvHypD, rates are reported in μ M sec⁻¹ for substrate turnover rate and not normalized to active protein concentrations.

5.4.12 Growth experiments and metabolite analyses for *Bacteroides* and *Parabacteroides* species

Species within the order Bacteroidales were routinely grown at 37 °C in supplemented BHI (BHIS) where BBL™ BHI was supplemented with 0.5 mg L⁻¹ hemin and 2.5 μ L L⁻¹ vitamin K₁. M9 minimal medium consisted of M9 salts (6 g L⁻¹ Na₂HPO₄, 3 g L⁻¹ KH₂PO₄, 1 g L⁻¹ NH₄Cl, 0.5 g L⁻¹ NaCl), 1 mM MgSO₄, 0.1 mM CaCl₂, 5 g L⁻¹ glucose, 0.5 g L⁻¹ L-cysteine, 5 mg L⁻¹ hemin, 2.5 μ L L⁻¹ vitamin K₁, 2 mg L⁻¹ FeSO₄•7H₂O, and 5 mg L⁻¹ vitamin B₁₂. M9 was supplemented with either Pro or Hyp with a final concentration of 20 mM or 40 mM. BHIS was supplemented with 1 mM or 10 mM Hyp for studies on deletion mutants and expression strains.

Growth experiments were carried out similarly for all strains. Turbid cultures grown in BHIS were diluted 1:100 into 200 μ L M9 minimal media or BHIS-Hyp in a 96-well plate. Growth curves were obtained through OD600 measurements as previously described. Hyp was quantified from all cultures at the end of the experiment by LC-MS/MS as described in Section 5.4.1. Each culture condition was tested with 3–5 culture replicates.

5.4.13 Construction of pFD340 vectors for protein expression in *Bacteroides* species

The genomic region encoding both BvHypD (locus tag: Bvu_1421) and BvAE (locus tag: Bvu_1420) was amplified in a single step from *B. vulgatus* ATCC 8482 gDNA. Primers used for PCRs are reported in Table 5.7. PCR product was gel-purified. Both insert fragment and pFD340 vector were restriction digested using BamHI. Digested insert and linearized vector were purified for ligation. BvHypDAE fragment was cloned into pFD340 using T4 DNA ligase according to manufacturer's protocol (1:3 molar ratio). The entire volume of ligation reaction was transformed into chemically competent *E. coli* TOP10 cells. Constructed plasmid, pFD340-BvHypDAE, was verified by sequencing and transformed into *E. coli* S17-1 λ pir for conjugation into *Bacteroides*.

5.4.14 Construction of plasmids for genetic deletions in *B. vulgatus* ATCC 8482^{viii}

Flanking regions 5'- and 3'- of gene encoding HypD (Bvu_1421) and AE (Bvu_1420) were amplified to obtain ~2–2.5 kb genomic fragments using primers listed in Table 5.7. PCR products were purified from PCR mixtures. Flanking regions for each gene were first ligated into pBlueScript vector using Gibson Assembly for colony screening. pBlueScript vector was

^{viii} Conjugation of pKNOCK vectors into *B. vulgatus*, serial passaging, plating and cPCR to identify deletion mutants were conducted by Prof. Laurie Comstock.

linearized by PCR amplification and then restriction digested with DpnI. Digested vector solutions were purified. Gibson Assembly reactions contained either 50 ng or 100 ng of vector with 2-fold molar excess of each insert in a total volume of 10 μ L for 50 ng vector reaction and 20 μ L for 100 ng vector reaction. Gibson reactions were incubated at 50 °C for 1 h in a thermocycler. 1 μ L of 50 ng and 2 μ L of 100 ng vector-containing reactions were transformed into chemically competent *E. coli* TOP10. Transformed cells were plated onto LB-Amp100 agar supplemented with 40 μ L per plate of X-Gal (5-bromo-4-chloro-3-indolyl-beta-d-galacto-pyranoside, Thermo Fisher) for blue-white colony screening to detect insert. White colonies were picked for plasmid purification and sequencing.

Ligated flanking regions verified in pBlueScript vectors were then amplified in a single step using Q5-HF Master Mix according to the manufacturer's protocol. PCR products were gel-purified. The pKNOCK-*bla-ermGb* vector was linearized by PCR amplification. PCR mixtures were restriction digested with DpnI and purified. Gibson Assembly reactions contained 100 ng of amplified vector with 2-fold molar excess of each insert in a total volume of 20 μ L. Gibson reactions were incubated at 50 °C for 1 h in a thermocycler. 2 μ L of each Gibson reaction was transformed into electrocompetent *E. coli* S17-1 λ pir. Transformed cells were plated onto LB-Amp100-Tp(trimethoprim)100 agar. Plasmids were purified and verified with sequencing to generate pKNOCK-*bla-ermGb*- Δ BvHypD and pKNOCK-*bla-ermGb*- Δ BvAE.

5.4.15 Construction of expression strains in *Bacteroides* species using pFD340-BvHypDAE

E. coli S17-1 λ pir pFD340-BvHypDAE was inoculated into 5 mL LB-Amp100-Tp100 and grown aerobically until the end of exponential phase for conjugation. Turbid starter cultures of *B. fragilis* ATCC 25285, *B. ovatus* ATCC 8483, and *B. thetaiotaomicron* VPI 5482 were inoculated 1:100 into 50 mL of Basal medium and incubated anaerobically at 37 °C. Basal medium

consisted of 20 g L⁻¹ proteose peptone, 5 g L⁻¹ yeast extract, 5 g L⁻¹ NaCl, 5 g L⁻¹ glucose, 5 g L⁻¹ K₂HPO₄, 0.5 g L⁻¹ L-cysteine, 5 mg L⁻¹ hemin, and 2.5 µL L⁻¹ vitamin K₁. When each *Bacteroides* culture reached early exponential phase (OD₆₀₀ ~0.2), *E. coli* culture was added aerobically and mixed well. Cells were harvested from these combined cultures by centrifugation. Supernatants were removed and pellets were suspended in ~100 µL of supplemented Basal medium. The entire volume of resuspended cells was plated as a puddle on BHIS-agar plates (no antibiotics) and incubated aerobically overnight at 37 °C. The entire cell mass on each plate was spread onto BHIS-Gm(gentamicin)200- Erm(erythromycin)5 (BHIS-GE) plates and incubated anaerobically at 37 °C. Colonies formed were streaked onto fresh BHIS-GE plates. Single colonies from these plates were inoculated into Basal medium with antibiotics Gm200 and Erm5 and incubated at 37 °C. Turbid cultures were used to make glycerol stocks. pFD340-BvHypDAE was purified from all cultures and verified by restriction digest with BamHI-HF to obtain DNA bands at the expected sizes. Any *E. coli* contamination was ruled out by lack of growth aerobically in LB.

5.4.16 Deletions of HypD and AE in *B. vulgatus* ATCC 8482^{ix}

E. coli S17-1 λ pir pKNOCK-*bla-ermGb*- Δ HypD and pKNOCK-*bla-ermGb*- Δ AE strains were inoculated into 10 mL LB-Amp100-Tp100 and grown aerobically until the end of exponential phase. Each *E. coli* culture was mixed with a 100 mL culture of *B. vulgatus* ATCC 8482 in early exponential phase for conjugation. Cells were harvested by centrifugation and pellets were resuspended in ~100 µL of supplemented Basal medium. The entire volume of resuspended cells was plated as a puddle on BHIS-agar plates (no antibiotics) and incubated

^{ix} Conjugation of pKNOCK vectors into *B. vulgatus*, serial passaging, plating and cPCR to identify deletion mutants were conducted by Prof. Laurie Comstock.

aerobically overnight at 37 °C. Cell mass on each plate was spread onto BHIS-GE agar plates and incubated anaerobically at 37 °C. Colonies were streaked onto fresh BHIS-GE plates. Single colonies from these plates were passaged in liquid Basal medium six times without antibiotics and then plated onto BHIS plates. Colonies from these non-selective plates were replica plated onto BHIS-GE plates to identify colonies that fail to grow on BHIS-GE plates. cPCR of the flanking gene regions was carried out to identify colonies with a clean deletion of *hypD* or *AE*. Screening primers are listed in Table 5.7. PCR products were also confirmed by sequencing.

5.4.17 Competitive colonization of germ-free mice with *B. vulgatus* ATCC 8482 WT and $\Delta hypD$ strains^x

B. vulgatus ATCC 8482 WT and $\Delta hypD$ cultures in exponential phase were mixed and plated to determine the ratio through colony PCR of 47 colonies. Germ-free mice were colonized with *B. vulgatus* ATCC 8482 at a ratio of 17:83, WT to $\Delta hypD$. Three 5.5 week old, female Swiss Webster mice were inoculated through oral gavage for this experiment and fed on a regular chow diet. After one week of inoculation, two stool samples were collected from each mouse and plated to obtain single colonies. 48 colonies from each mouse were picked for cPCR of the *hypD* gene to differentiate between WT and deletion mutants. Screening primers are listed in Table 5.7.

^x The colonization competition experiment in germ-free mice was conducted by Prof. Laurie Comstock.

5.5 References

1. BJ Levin*, YY Huang*, SC Peck, Y Wei, A Martínez-del Campo, JA Marks, EA Franzosa, C Huttenhower, EP Balskus. A prominent glycyl radical enzyme in human gut microbiomes metabolizes *trans*-4-hydroxy-L-proline. *Science*. **2017**, 355(6325):eaai8386.
2. LH Stickland. Studies in the metabolism of the strict anaerobes (genus *Clostridium*). I. The chemical reactions by which *C. sporogenes* obtains its energy. *Biochem J*. **1934**, 28(5):1746-59.
3. S Jackson, M Calos, A Myers, WT Self. Analysis of proline reduction in the nosocomial pathogen *Clostridium difficile*. *J Bacteriol*. **2006**, 188(24):8487-95.
4. UC Kabisch, A Gräntzdörffer, A Schierhorn, KP Rücknagel, JR Andreesen, A Pich. Identification of D-proline reductase from *Clostridium sticklandii* as a selenoenzyme and indications for a catalytically active pyruvoyl group derived from a cysteine residue by cleavage of a proprotein. *J Biol Chem*. **1999**, 274(13):8445-54.
5. M Neumann-Schaal, JD Hofmann, SE Will, D Schomburg. Time-resolved amino acid uptake of *Clostridium difficile* 630 Δ erm and concomitant fermentation product and toxin formation. *BMC Microbiology*. **2015**, 15:281-93.
6. HA Barker, L D'Ari, J Kahn. Enzymatic reactions in the degradation of 5-aminovalerate by *Clostridium aminovalericum*. *J Biol Chem*. **1987**, 262(19):8994-9003.
7. JK Hardman, TC Stadtman. Metabolism of ω -amino acidS: II. Fermentation of δ -aminovaleric acid by *Clostridium aminovalericum* N. Sp. *J Bacteriol*. **1960**, 79(4):549-52.
8. ML Jenior, JL Leslie, VB Young, PD Schloss. *Clostridium difficile* colonizes alternative nutrient niches during infection across distinct murine gut microbiomes. *mSystems*. **2017**, 2(4):e00063-17.
9. JR Fletcher, S Erwin, C Lanzas, CM Theriot. Shifts in the gut metabolome and *Clostridium difficile* transcriptome throughout colonization and infection in a mouse model. *mSphere*. **2018**, 3(2):e00089-18.
10. YK Ng, M Ehsaan, S Philip, MM Collery, C Janoir, A Collignon, ST Cartman, NP Minton. Expanding the repertoire of gene tools for precise manipulation of the *Clostridium difficile* genome: allelic exchange using *pyrE* alleles. *PLOS ONE*. **2013**, 8(2):e56051.
11. AA Shishkin, G Giannoukos, A Kucukural, D Ciulla, M Busby, C Surka, J Chen, RP Bhattacharyya, RF Rudy, MM Patel, et al. Simultaneous generation of many RNA-seq libraries in a single reaction. *Nature methods*. **2015**, 12(4):323-5.
12. B Langmead, SL Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature methods*. **2012**, 9(4):357-9.

13. Y Liao, GK Smyth, W Shi. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **2013**, 41(10):e108.
14. S Anders, W Huber. Differential expression analysis for sequence count data. *Genome Biology.* **2010**, 11(10):R106.
15. R_Core_Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org>
16. RStudio_Team (2016). RStudio: Integrated development environment for R. RStudio, Inc., Boston, MA. <http://www.rstudio.com>
17. MI Love, W Huber, S Anders. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology.* **2014**, 15(12):550.
18. R Kolde (2018). pheatmap: Pretty Heatmaps. R package version 1.0.10. <https://CRAN.R-project.org/package=pheatmap>
19. E Neuwirth (2014). RColorBrewer: ColorBrewer Palettes. R package version 1.1-2. <https://CRAN.R-project.org/package=RColorBrewer>
20. R. Gentleman, V. Carey, W. Huber, F Hahne (2018). genefilter: methods for filtering genes from high-throughput experiments. R package version 1.62.0.
21. J Kim, D Darley, W Buckel. 2-Hydroxyisocaproyl-CoA dehydratase and its activator from *Clostridium difficile*. *The FEBS journal.* **2005**, 272(2):550-61.
22. SH Knauer, W Buckel, H Dobbek. Structural basis for reductive radical formation and electron recycling in (R)-2-hydroxyisocaproyl-CoA dehydratase. *J Am Chem Soc.* **2011**, 133(12):4342-7.
23. H Wickham (2016). ggplot2: Elegant graphics for data analysis. Springer-Verlag New York, <http://ggplot2.org>
24. E-H Aboulnaga, O Pinkenburg, J Schiffels, A El-Refai, W Buckel, T Selmer. Effect of an oxygen-tolerant bifurcating butyryl coenzyme a dehydrogenase/electron-transferring flavoprotein complex from *Clostridium difficile* on butyrate production in *Escherichia coli*. *Journal of Bacteriology.* **2013**, 195(16):3704-13.
25. JA Ferreyra, KJ Wu, AJ Hryckowian, DM Bouley, BC Weimer, JL Sonnenburg. Gut microbiota-produced succinate promotes *C. difficile* infection after antibiotic treatment or motility disturbance. *Cell Host Microbe.* **2014**, 16(6):770-7.
26. P Louis, HJ Flint. Formation of propionate and butyrate by the human colonic microbiota. *Environmental Microbiology.* **2016**, 19(1):29-41.

27. SE Pryde, SH Duncan, GL Hold, CS Stewart, HJ Flint. The microbiology of butyrate formation in the human colon. *FEMS Microbiology Letters*. **2006**, 217(2):133-9.
28. A Bedford, J Gong. Implications of butyrate and its derivatives for gut health and animal production. *Animal Nutrition*. **2018**, 4(2):151-9.
29. L Bouillaut, WT Self, AL Sonenshein. Proline-dependent regulation of *Clostridium difficile* Stickland metabolism. *Journal of Bacteriology*. **2013**, 195(4):844-54.
30. DN Gerding, S Johnson, M Rupnik, K Aktories. *Clostridium difficile* binary toxin CDT: Mechanism, epidemiology, and potential clinical importance. *Gut microbes*. **2014**, 5(1):15-27.
31. L Bouillaut, T Dubois, AL Sonenshein, B Dupuy. Integration of metabolism and virulence in *Clostridium difficile*. *Res Microbiol*. **2015**, 166(4):375-83.
32. S Karlsson, LG Burman, T Åkerlund. Suppression of toxin production in *Clostridium difficile* VPI 10463 by amino acids. *Microbiology*. **1999**, 145(7):1683-93.
33. S Craciun, EP Balskus. Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme. *Proc Natl Acad Sci USA*. **2012**, 109(52):21307-12.
34. MK Bacic, CJ Smith. Laboratory maintenance and cultivation of *Bacteroides* species. *Current protocols in microbiology*. **2008**, 9(1):13C.1.1-C.1.21.
35. J Gordon, JW M'Leod. Inhibition of bacterial growth by some amino-acids and its bearing on the use of tryptic digests as culture media. *The Journal of Pathology and Bacteriology*. **1926**, 29(1):13-25.
36. L Dienes, PC Zamecnik. Transformation of bacteria into L forms by amino acids. *Journal of Bacteriology*. **1952**, 64(5):770-1.
37. AA Salyers, N Shoemaker, A Cooper, J D'Elia, JA Shipman. Genetic methods for *Bacteroides* species. In: Smith MCM, Sockett RE, editors. *Methods in Microbiology*. Academic Press; 1999. p. 229-49.
38. NM Koropatkin, EC Martens, JI Gordon, TJ Smith. Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure*. **2008**, 16(7):1105-15.
39. MJ Coyne, AO Tzianabos, BC Mallory, VJ Carey, DL Kasper, LE Comstock. Polysaccharide biosynthesis locus required for virulence of *Bacteroides fragilis*. *Infection and Immunity*. **2001**, 69(7):4342-50.
40. CJ Smith, MB Rogers, ML McKee. Heterologous gene expression in *Bacteroides fragilis*. *Plasmid*. **1992**, 27(2):141-54.

41. J Xu, MK Bjursell, J Himrod, S Deng, LK Carmichael, HC Chiang, LV Hooper, JI Gordon. A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. *Science*. **2003**, 299(5615):2074-6.
42. M Hijazi, SM Velasquez, E Jamet, JM Estevez, C Albenne. An update on post-translational modifications of hydroxyproline-rich glycoproteins: toward a model highlighting their contribution to plant cell wall architecture. *Front Plant Sci*. **2014**, 5:395.
43. E Nguema-Ona, M Viché-Gibouin, M Gotté, B Plancot, P Lerouge, M Bardor, A Driouich. Cell wall O-glycoproteins and N-glycoproteins: aspects of biosynthesis and function. *Front Plant Sci*. **2014**, 5(499).
44. B Thiele, N Stein, M Oldiges, D Hofmann. Direct analysis of underivatized amino acids in plant extracts by LC-MS-MS. *Methods Mol Biol*. **2012**, 828:317-28.
45. X Zheng, G Xie, A Zhao, L Zhao, C Yao, NHL Chiu, Z Zhou, Y Bao, W Jia, JK Nicholson, et al. The footprints of gut microbial-mammalian co-metabolism. *J Proteome Res*. **2011**, 10(12):5512-22.
46. RLP Jump, A Polinkovsky, K Hurless, B Sitzlar, K Eckart, M Tomas, A Deshpande, MM Nerandzic, CJ Donskey. Metabolomics analysis identifies intestinal microbiota-derived biomarkers of colonization resistance in clindamycin-treated mice. *PLOS ONE*. **2014**, 9(7):e101267.
47. M Matsumoto, R Kibe, T Ooga, Y Aiba, S Kurihara, E Sawaki, Y Koga, Y Benno. Impact of intestinal microbiota on intestinal luminal metabolome. *Sci Rep*. **2012**, 2(233).
48. EA Smith, GT Macfarlane. Dissimilatory amino acid metabolism in human colonic bacteria. *Anaerobe*. **1997**, 3(5):327-37.
49. SE Mergenhagen, HW Scherp. Lysis of reconstituted collagen and catabolism of products of collagenolysis by the oral microbiota. *Arch Oral Biol*. **1960**, 1:333-8.
50. D Liu, M Nikoo, G Boran, P Zhou, JM Regenstein. Collagen and gelatin. *Annu Rev Food Sci Technol*. **2015**, 6(1):527-57.
51. PH Weiss, L Klein. The quantitative relationship of urinary peptide hydroxyproline excretion to collagen degradation. *J Clin Invest*. **1969**, 48(1):1-10.
52. EC Martens, NM Koropatkin, TJ Smith, JI Gordon. Complex glycan catabolism by the human gut microbiota: The Bacteroidetes Sus-like paradigm. *J Biol Chem*. **2009**, 284(37):24673-7.
53. C Janoir, C Denève, S Bouttier, F Barbut, S Hoys, L Caleechum, D Chapetón-Montes, FC Pereira, AO Henriques, A Collignon, et al. Adaptive strategies and pathogenesis of *Clostridium difficile* from *in vivo* transcriptomics. *Infect Immun*. **2013**, 81(10):3757-69.

54. RA Britton, VB Young. Role of the intestinal microbiota in resistance to colonization by *Clostridium difficile*. *Gastroenterology*. **2014**, 146(6):1547-53.
55. MD Shoulders, RT Raines. Collagen structure and stability. *Annu Rev Biochem*. **2009**, 78(1):929-58.
56. P Biancheri, A Di Sabatino, GR Corazza, TT MacDonald. Proteases and the gut barrier. *Cell Tissue Res*. **2013**, 351(2):269-80.
57. AS Duarte, A Correia, AC Esteves. Bacterial collagenases – A review. *Crit Rev Microbiol*. **2016**, 42(1):106-26.
58. DJ Harrington. Bacterial collagenases and collagen-degrading enzymes and their potential role in human disease. *Infect Immun*. **1996**, 64(6):1885-91.
59. F-PJ Martin, N Sprenger, I Montoliu, S Rezzi, S Kochhar, JK Nicholson. Dietary modulation of gut functional ecology studied by fecal metabonomics. *J Proteome Res*. **2010**, 9(10):5284-95.
60. H Takeda, S Yamamoto, Y Kojima, O Hayaishi. Studies on monooxygenases: I. General properties of crystalline L-lysine monooxygenase. *Journal of Biological Chemistry*. **1969**, 244(11):2935-41.
61. O Revelles, M Espinosa-Urgel, S Molin, JL Ramos. The *davDT* operon of *Pseudomonas putida*, involved in lysine catabolism, is induced in response to the pathway intermediate δ -aminovaleric acid. *Journal of Bacteriology*. **2004**, 186(11):3439-46.
62. M Muhyaddin, PJ Roberts, GN Woodruff. Presynaptic γ -aminobutyric acid receptors in the rat anococcygeus muscle and their antagonism by 5-aminovaleric acid. *Br J Pharmacol*. **1982**, 77(1):163-8.
63. M Auteri, MG Zizzo, R Serio. GABA and GABA receptors in the gastrointestinal tract: from motility to inflammation. *Pharmacol Res*. **2015**, 93:11-21.
64. CM Adams, BE Eckenroth, EE Putnam, S Doublié, A Shen. Structural and functional analysis of the CspB protease required for *Clostridium* spore germination. *PLOS Pathogens*. **2013**, 9(2):e1003165.
65. JJ Trombetta, D Gennert, D Lu, R Satija, AK Shalek, A Regev. Preparation of single-cell RNA-Seq libraries for next generation sequencing. *Current protocols in molecular biology*. **2014**, 107:4.22.1-4.22.17.
66. H Li, R Durbin. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. **2010**, 26(5):589-95.

67. Y Wei, B Li, D Prakash, JG Ferry, SJ Elliott, J Stubbe. A ferredoxin disulfide reductase delivers electrons to the *Methanosarcina barkeri* class III ribonucleotide reductase. *Biochemistry*. **2015**, 54(47):7019-28.
68. E Gasteiger, A Gattiker, C Hoogland, I Ivanyi, RD Appel, A Bairoch. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*. **2003**, 31(13):3784-8.