

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Discovery and Applications of Radical Amino Acid Halogenases

Permalink

<https://escholarship.org/uc/item/04c7h5mz>

Author

Neugebauer, Monica Elena

Publication Date

2020

Peer reviewed|Thesis/dissertation

Discovery and Applications of Radical Amino Acid Halogenases

by

Monica E Neugebauer

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy in
Chemical Engineering
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:

Professor Michelle C. Y. Chang, Co-chair

Professor Wenjun Zhang, Co-chair

Professor Judith P. Klinman

Professor Jay D. Keasling

Spring 2020

Discovery and Applications of Radical Amino Acid Halogenases

© 2020

by Monica E Neugebauer

Abstract

Discovery and Applications of Radical Amino Acid Halogenases

by

Monica E Neugebauer

Doctor of Philosophy in Chemical Engineering

University of California, Berkeley

Professor Michelle C. Y. Chang, Co-chair

Professor Wenjun Zhang, Co-chair

Through evolution, Nature has developed enzymes that perform challenging reactions. By exploring biosynthetic pathways in bacteria, we can discover and repurpose these enzymes for synthetic applications. Here, we have performed metabolomics and enzyme characterization to elucidate the biosynthesis of β -ethynylserine (β es), an alkyne-containing amino acid made by *Streptomyces cattleya*. In the process, we discovered a novel $\text{Fe}^{\text{II}}/\alpha$ -ketoglutarate ($\text{Fe}^{\text{II}}/\alpha\text{KG}$)-dependent halogenase, BesD, with unprecedented reactivity toward amino acids. We solved the crystal structure of BesD, discovered homologs that enable the formation of nine new chlorinated amino acids, and demonstrated that the newly discovered enzymes also perform bromination and azidation. Next, we demonstrated that halogenated amino acids can be further transformed enzymatically into chlorinated heterocycles, diamines, keto-acids, and peptides. Finally, we investigated how halogenases achieve regioselectivity, substrate selectivity, and catalytic selectivity (halogenation over the closely related hydroxylation reaction). These radical amino acid halogenases set the stage for the discovery of novel natural products that are formed via halogenated intermediates.

Table of Contents

<i>Table of Contents</i>	i
<i>List of Figures, Schemes, and Tables</i>	iii
<i>List of Abbreviations</i>	ix
<i>Acknowledgments</i>	xi

Chapter 1: Introduction

1.1 <i>Rare functional groups in Nature</i>	2
1.2 <i>Enzymatic halogenation strategies</i>	3
1.3 <i>Substrate scope of Fe^{II}/αKG-dependent halogenases</i>	4
1.4 <i>Mechanistic investigation of radical halogenases</i>	6
1.5 <i>Thesis motivation and organization</i>	9
1.6 <i>References</i>	

Chapter 2: Discovery of a pathway for β-ethynylserine biosynthesis

2.1 <i>Introduction</i>	17
2.2 <i>Materials and methods</i>	18
2.3 <i>Results and discussion</i>	29
2.4 <i>Conclusions</i>	68
2.5 <i>References</i>	69

Chapter 3: Discovery and applications of radical amino acid halogenases

3.1 <i>Introduction</i>	73
3.2 <i>Materials and methods</i>	75
3.3 <i>Results and discussion</i>	87
3.4 <i>Conclusions</i>	131
3.5 <i>References</i>	132

Chapter 4: Mechanistic investigation of halogenation vs. hydroxylation in the BesD family

<i>4.1 Introduction</i>	137
<i>4.2 Materials and methods</i>	139
<i>4.3 Results and discussion</i>	146
<i>4.4 Conclusions</i>	180
<i>4.5 References</i>	181

Chapter 5: Structural basis of regioselectivity and substrate selectivity in radical amino acid halogenases

<i>5.1 Introduction</i>	185
<i>5.2 Materials and methods</i>	186
<i>5.3 Results and discussion</i>	192
<i>5.4 Conclusions</i>	216
<i>5.5 References</i>	217

Appendices

<i>Appendix 1: Plasmids and oligonucleotides</i>	220
<i>Appendix 2: Supplementary figures</i>	239
<i>Appendix 3: Python scripts for data analysis</i>	249
<i>Appendix 4: Strain list for cloned constructs</i>	253

List of Figures, Schemes, and Tables

Chapter 1

Figure 1.1	Natural products with rare functional groups	2
Figure 1.2	Enzymatic strategies for halogenation	3
Figure 1.3	Substrates of reported radical halogenases	5
Figure 1.4	Comparison of radical halogenases and hydroxylases	6
Figure 1.5	Structures of radical halogenases	7
Figure 1.6	Stopped-flow studies of radical halogenases	7
Figure 1.7	Mechanistic insights into SyrB2	8
Figure 1.8	Measuring active site geometry of halogenases	8

Chapter 2

Figure 2.1	Natural products of acetylenic origin	17
Table 2.1	Genes in β es biosynthetic gene cluster	29
Figure 2.2	Fatty acid desaturase knockouts	30
Figure 2.3	Pra production in <i>S. catenulae</i>	31
Figure 2.4	Comparative genomics to identify alkyne cluster	32
Figure 2.5	Genomic context of β es/Pra biosynthetic clusters	33
Figure 2.6	Genomic context of Pra biosynthetic clusters	34
Figure 2.7	Alkyne production in <i>Streptomyces</i> spp.	35
Figure 2.8	Alkyne production phenotypes in pathway knockouts	36
Figure 2.9	Biosynthetic pathway for production of β es from lysine	38
Figure 2.10	Extraction ion chromatograms of pathway intermediates	39
Figure 2.11	Sequence alignment of the halogenase BesD	40
Figure 2.12	LC/MS characterization of BesD halogenase	41
Figure 2.13	Characterization of BesD G139D mutant	42
Figure 2.14	LC/MS analysis of 4-Cl-lysine decomposition	43
Figure 2.15	4-Cl-lysine decomposition pathway	45
Figure 2.16	NMR characterization of 4-Cl-lysine	46
Figure 2.17	NMR characterization of lactam products	47

<i>Figure 2.18</i>	<i>BesC homology model</i>	48
<i>Figure 2.19</i>	<i>Overall reaction of BesC</i>	48
<i>Figure 2.20</i>	<i>LC/MS analysis of BesC</i>	49
<i>Figure 2.21</i>	<i>Mass spectra of BesC product</i>	49
<i>Figure 2.22</i>	<i>BesC produces allylglycine from lysine</i>	50
<i>Figure 2.23</i>	<i>Detection of BesC ammonia co-product</i>	52
<i>Figure 2.24</i>	<i>Detection of BesC formaldehyde co-product</i>	53
<i>Figure 2.25</i>	<i>Ammonia and formaldehyde quantification</i>	54
<i>Table 2.2</i>	<i>Crystallography table for BesC apo structure</i>	55
<i>Figure 2.26</i>	<i>Apo crystal structure of BesC</i>	56
<i>Figure 2.27</i>	<i>In vitro characterization of BesB</i>	57
<i>Figure 2.28</i>	<i>LC/MS analysis of BesB in D₂O</i>	58
<i>Figure 2.29</i>	<i>Proposed mechanism of BesB</i>	59
<i>Figure 2.30</i>	<i>LC/MS data for BesA and BesE knockouts</i>	61
<i>Figure 2.31</i>	<i>Metabolomics of γ-Glu-Pra and γ-Glu-βes</i>	62
<i>Figure 2.32</i>	<i>QQQ fragmentation of γ-Glu-Pra and γ-Glu-βes</i>	63
<i>Figure 2.33</i>	<i>LC/MS characterization of γ-Glu-Pra and γ-Glu-βes</i>	63
<i>Figure 2.34</i>	<i>Assay of BesA activity on amino acids</i>	64
<i>Figure 2.35</i>	<i>Kinetic analysis of BesA</i>	65
<i>Figure 2.36</i>	<i>Reconstitution of the BesA and BesE activity</i>	66
<i>Figure 2.37</i>	<i>In vitro reconstitution of βes biosynthesis</i>	67
<i>Scheme 2.1</i>	<i>Diverse functional groups in the βes pathway</i>	68

Chapter 3

<i>Figure 3.1</i>	<i>Products of BesD, WelO5, and SyrB2</i>	74
<i>Figure 3.2</i>	<i>Sequence comparison of BesD, WelO5, and SyrB2</i>	88
<i>Table 3.1</i>	<i>Data collection and refinement of BesD</i>	89
<i>Figure 3.3</i>	<i>Crystal structure of BesD</i>	90
<i>Figure 3.4</i>	<i>Comparison of halogenases and hydroxylases</i>	91
<i>Figure 3.5</i>	<i>Proposed mechanism of BesD</i>	92
<i>Figure 3.6</i>	<i>LC/MS analysis of BesD active site mutants</i>	93
<i>Figure 3.7</i>	<i>O₂ binding comparison in BesD and WelO5</i>	95

Figure 3.8	Conservation of lysine-binding residues in BesD	96
Figure 3.9	Phylogenetic tree of BesD halogenase homologs	97
Figure 3.10	Sequence similarity network of BesD homologs	98
Figure 3.11	Genomic context of HalA-HalE	99
Figure 3.12	SDS-PAGE of purified halogenases	100
Figure 3.13	Summary of LC/MS analysis of HalA-HalE	101
Table 3.2	Summary of tested halogenases	102
Table 3.3	Substrate profiles of assayed halogenases	103
Table 3.4	Percent identity matrix of amino acid halogenases	104
Figure 3.14	LC/MS analysis of regioselective lysine halogenation	105
Figure 3.15	NMR analysis of 5-Cl-lysine methyl ester	106
Figure 3.16	LC/MS analysis of 4,4-dichlorolysine	108
Figure 3.17	NMR analysis of 4,4-dichlorolysine methyl ester	109
Figure 3.18	LC/MS analysis of 5,5-dichlorolysine	110
Figure 3.19	NMR analysis of 5,5-dichlorolysine methyl ester	111
Figure 3.20	LC/MS analysis of 4-Cl-ornithine	112
Figure 3.21	Coupled reaction for monitoring succinate formation	112
Figure 3.22	Ornithine halogenase kinetics	113
Figure 3.23	NMR analysis of 4-Cl-ornithine	114
Figure 3.24	LC/MS analysis of aliphatic amino acid halogenases	115
Figure 3.25	Kinetic analysis of aliphatic amino acid halogenase HalE	116
Figure 3.26	NMR analysis of 4-Cl-leucine	117
Figure 3.27	NMR analysis of 4-Cl-isoleucine	118
Figure 3.28	NMR analysis of 5,5-dichloroleucine	119
Figure 3.29	Sequence alignment of HalA-HalE	120
Figure 3.30	Variable C-termini of amino acid halogenases	121
Figure 3.31	Bromination and azidation catalyzed by HalB	122
Figure 3.32	SDS-PAGE of OCD, LDC, IlvE, and RapL	123
Figure 3.33	In vitro formation of chlorinated heterocycles by RapL	124
Figure 3.34	In vitro formation of chlorinated heterocycles by OCD	125
Figure 3.35	In vitro formation of chlorinated diamines by LDC	126
Figure 3.36	In vitro formation of chlorinated α -keto-acids by IlvE	127

<i>Figure 3.37</i>	<i>Chlorinated peptide formation by IVTT</i>	129
<i>Figure 3.38</i>	<i>Stapled peptide analysis by IVTT and LCMS</i>	129
<i>Figure 3.39</i>	<i>Summary of results for chlorinated peptides by IVTT</i>	130
<i>Figure 3.40</i>	<i>Applications of radical amino acid halogenases</i>	131

Chapter 4

<i>Figure 4.1</i>	<i>Reaction outcomes in halogenases vs. hydroxylases</i>	138
<i>Figure 4.2</i>	<i>Phylogenetic tree of halogenases and hydroxylases</i>	147
<i>Figure 4.3</i>	<i>Sequence alignment of Hal and Hydrox</i>	148
<i>Figure 4.4</i>	<i>LC/MS analysis of Hal and Hydrox</i>	149
<i>Figure 4.5</i>	<i>Kinetic analysis of Hal and Hydrox</i>	150
<i>Figure 4.6</i>	<i>NMR characterization of 4-OH-lysine</i>	151
<i>Figure 4.7</i>	<i>LC/MS characterization of Hydrox D144G</i>	152
<i>Table 4.1</i>	<i>Crystallography table for Hydrox</i>	154
<i>Figure 4.8</i>	<i>Crystal structure of Hydrox</i>	155
<i>Figure 4.9</i>	<i>Conformational change in Hydrox</i>	156
<i>Figure 4.10</i>	<i>Overlay of Hydrox and BesD crystal structures</i>	157
<i>Figure 4.11</i>	<i>αKG binding overlay in Hydrox and BesD</i>	158
<i>Figure 4.12</i>	<i>DNA shuffle library design</i>	159
<i>Figure 4.13</i>	<i>Strategy for shuffle library screening</i>	160
<i>Figure 4.14</i>	<i>CuAAC reaction of Pra with azidocoumarin probe</i>	160
<i>Figure 4.15</i>	<i>Comparison of CuAAC in cell culture media</i>	161
<i>Figure 4.16</i>	<i>CuAAC screen comparison of Hal and Hydrox expression</i>	162
<i>Figure 4.17</i>	<i>CuAAC reaction Pra with CalFluor488-azide</i>	164
<i>Figure 4.18</i>	<i>CuAAC using azidocoumarin vs. CalFluor488-azide</i>	164
<i>Figure 4.19</i>	<i>Data from shuffle library screening</i>	165
<i>Figure 4.20</i>	<i>Heat map of sequence conservation in halogenase hits</i>	166
<i>Figure 4.21</i>	<i>Scatter plot of conservation in halogenase hits</i>	166
<i>Figure 4.22</i>	<i>Mapping critical residues onto the Hydrox structure</i>	167
<i>Figure 4.23</i>	<i>Heat map of conservation in non-halogenating variants</i>	168
<i>Figure 4.24</i>	<i>Scatter plot of conservation in non-halogenating variants</i>	168
<i>Table 4.2</i>	<i>Conservation of residues 151 and 225 in the screen</i>	169

<i>Figure 4.25</i>	<i>SDS-PAGE of purified Hal and Hydrox variants</i>	170
<i>Figure 4.26</i>	<i>LC/MS analysis of engineered Hydrox variants</i>	170
<i>Figure 4.27</i>	<i>Design of a chimera based on the shuffle library</i>	171
<i>Table 4.3</i>	<i>Description of residue grouping in the chimera</i>	174
<i>Figure 4.28</i>	<i>SDS-PAGE of purified Hydrox and Chimera variants</i>	175
<i>Figure 4.29</i>	<i>LC/MS analysis of Hydrox and Chimera variants</i>	175
<i>Figure 4.30</i>	<i>Proposed role of residue 225 on halogenation outcome</i>	176
<i>Figure 4.31</i>	<i>Spectroscopic techniques for probing halogenase mechanism</i>	177
<i>Table 4.4</i>	<i>Crystallography tables for BesD bound to vanadium</i>	178
<i>Figure 4.32</i>	<i>Crystal structure of BesD bound to vanadium</i>	179

Chapter 5

<i>Figure 5.1</i>	<i>Reactions of BesD, HalB, and HalD</i>	185
<i>Figure 5.2</i>	<i>Sequence comparison of BesD, HalB, and HalD</i>	193
<i>Table 5.1</i>	<i>Crystallography table for HalB with lysine</i>	194
<i>Figure 5.3</i>	<i>Crystal structure of HalB bound to lysine</i>	195
<i>Figure 5.4</i>	<i>Overall structures of BesD and HalB</i>	196
<i>Figure 5.5</i>	<i>Structural alignment of BesD and HalB</i>	196
<i>Figure 5.6</i>	<i>Stopped-flow absorption assay for BesD and HalB</i>	198
<i>Figure 5.7</i>	<i>Strategy for screening HalB library variants</i>	199
<i>Figure 5.8</i>	<i>Results of CuAAC screening</i>	200
<i>Figure 5.9</i>	<i>Kinetics of HalB and HalB M246 variants</i>	202
<i>Figure 5.10</i>	<i>Design of chimeric variant of BesD and HalB</i>	204
<i>Figure 5.11</i>	<i>Coupled reactions for succinate detection</i>	205
<i>Figure 5.12</i>	<i>Succinate formation in BesD, HalB, and the chimeric enzymes</i>	205
<i>Table 5.2</i>	<i>Crystallography table for HalD bound to ornithine</i>	207
<i>Figure 5.13</i>	<i>Crystal structure of HalD</i>	208
<i>Figure 5.14</i>	<i>Overall fold comparison of BesD, HalB, and HalD</i>	208
<i>Figure 5.15</i>	<i>Active site comparison of HalD and BesD</i>	209
<i>Figure 5.16</i>	<i>Active site comparison of HalD and HalB</i>	210
<i>Table 5.3</i>	<i>Crystallography table for HalB conformational change</i>	212
<i>Figure 5.17</i>	<i>Crystal structure of HalB conformational change</i>	213

<i>Table 5.4</i>	<i>Crystallography table for apo BesD</i>	214
<i>Figure 5.18</i>	<i>Conformational change in HalB, BesD, and Hydrox</i>	215

Appendix I

<i>Table A1.1</i>	<i>Strains, plasmids, oligonucleotides, and sequences used in Chapter 2</i>	221
<i>Table A1.2</i>	<i>Strains, plasmids, oligonucleotides, and sequences used in Chapter 3</i>	224
<i>Table A1.3</i>	<i>Strains, plasmids, oligonucleotides, and sequences used in Chapter 4</i>	231
<i>Table A1.4</i>	<i>Strains, plasmids, oligonucleotides, and sequences used in Chapter 5</i>	236

Appendix II

<i>Figure A2.1</i>	<i>Phylogenetic tree of BesD homologs</i>	240
<i>Figure A2.2</i>	<i>Shuffle library sequencing results: halogenating variants</i>	245
<i>Figure A2.3</i>	<i>Shuffle library sequencing results: non-halogenating variants</i>	247

Appendix III

<i>Figure A3.1</i>	<i>Python script for annotation of BesD homologs</i>	250
<i>Figure A3.2</i>	<i>Python script for shuffle library analysis</i>	251

Appendix IV

<i>Figure A4.1</i>	<i>Strain list for constructs in Chapter 2</i>	254
<i>Figure A4.2</i>	<i>Strain list for constructs in Chapter 3</i>	254
<i>Figure A4.3</i>	<i>Strain list for constructs in Chapter 4</i>	255
<i>Figure A4.4</i>	<i>Strain list for untested constructs related to Chapter 4</i>	256
<i>Figure A4.5</i>	<i>Strain list for constructs in Chapter 5</i>	256

List of Abbreviations

α KG	α -ketoglutarate
ATP	adenosine-5' -triphosphate
ACP	acyl-carrier protein
ARS	aminoacyl-tRNA synthetase
β es	β -ethynylserine
CuAAC	copper catalyzed azide alkyne cycloaddition
bp	base pairs
d	day
Da	daltons
DTT	DL-dithiothreitol
EIC	extracted ion chromatogram
h	hour
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His	histidine
HPLC	high-pressure liquid chromatography
Glu-Pra	γ -glutamyl-propargylglycine
Glu-Bes	γ -glutamyl- β -ethynylserine
Ggt	γ -glutamyl-transpeptidase
kDa	kilodaltons
LB	lysogeny broth
LC	liquid chromatography
LC/MS	liquid chromatography mass spectrometry
LC/MS-MS	liquid chromatography tandem mass spectrometry
min	minute
MWCO	molecular weight cutoff
MRM	multiple reaction monitoring
Met	methionine
NADH	nicotinamide adenine dinucleotide
Ni-NTA	nickel nitrilotriacetic acid
Ppant	Phosphopantetheine
Pra	propargylglycine
QQQ	triple quadrupole tandem mass spectrometer
QTOF	quadrupole-time of flight mass spectrometer

SCS	succinyl coenzyme-A synthetase
SDS	sodium dodecyl sulfate
SF-abs	stopped-flow absorption
TB	terrific broth
TCEP	tris(2-carboxyethyl)phosphine

Acknowledgments

I have been lucky to have incredible research mentors during my undergraduate and graduate studies. First, I would like to thank my PhD advisor, Prof. Michelle Chang for her endless support and mentorship. Michelle has always encouraged me to follow the science that I have found interesting and to push myself to learn new techniques. As a result, I have had a diverse training experience and a lot of fun while in graduate school. She has given me the confidence to give public talks, to write, and to pursue a career in science. This work would not have been possible without her guidance. I would like to thank Prof. Alice Ting for giving me my first research experience in chemical biology as an undergraduate. That formative research experience led me to pursue a PhD. I am also grateful to my thesis committee members Wenjun Zhang, Judith Klinman, and Jay Keasling for their scientific advice over the years.

I would like to acknowledge past and present members of the Chang Lab for creating a supportive and creative environment in which to do science. I would like to thank Dr. Stephanie Jones and Dr. Ben Thuronyi for their mentorship in the lab and for their advice in the early stages of project selection. Dr. Jon McMurry, for sharing his bioinformatics expertise and for his “fun science facts of the day.” Dr. Mike Blaisse, Dr. Joe Gallagher, Dr. Matt Davis, Dr. Omer Ad, Dr. Jase Gehring, Dr. Chia-I Lin, Kersh Theva, Jason Fang, Ed Koleski, and Dr. Hongjun Dong, for scientific discussion and support. Dr. Vivian Yu, for being a great friend and scientist. I am so grateful for all the times she saved me by sharing her snacks in lab. Sasilada Sirirungruang, who joined the Chang Lab at the same time as me and has been there for the ups and downs of the last six years. Dr. Ningkun Wang, for getting me started on crystallography and for being the best desk-mate I could ask for. I am especially grateful for my fellow chemical engineer, Jorge Marchand. He is an incredibly creative and generous scientist. I don’t know where I would be if he had not allowed me to join him on the alkyne biosynthesis project that he started. I would like to thank the brilliant new graduate students who have joined the halogenase and alkyne amino acid projects: Eli Kissman has been great to collaborate with during my final year. Max Sosa and Doug Millar continue to inspire me with their ideas and enthusiasm. Finally, I would like to thank Kiera Sumida and Nikko Sambold, undergraduates who have contributed so much to the halogenase projects. I am excited to follow their scientific careers in the future. I am also grateful to many scientists who helped me along the way. George Meigs, James Holton, Christine Gee, Elizabeth Wittenborn, Ben Horst, and Ming-Yuan Su all took time to help me with aspects of protein crystallography, and for that I am grateful.

I would like to thank my family, who have sacrificed so much to give me the opportunities that have led me to where I am today. I am eternally grateful for your support. I would also like to thank Vinati and Sam for being such amazing friends for nearly a decade. Thank you for all the encouraging messages, reunion trips, and endless support.

Lastly, I would like to thank my husband, Jeff, who provided constant love and support through the good times and the tough times. I am so lucky to be married to such an incredible person who also happens to be a tremendous scientist.

Chapter 1: *Introduction*

1.1. Rare functional groups in Nature

Strategies for building chemical diversity in synthetic chemistry differ from strategies found in Nature. While synthetic strategies take advantage of a tremendous diversity of functional groups in order to build complexity, biological systems instead perform chemistry using a relatively limited set of building blocks with fewer accessible functional groups. Despite access to fewer building blocks, biological systems generate incredibly complex molecules with exquisite selectivity through the use of enzymes. Although both synthetic and biological approaches have greatly advanced our ability to make compounds, the chemical space accessible through the two strategies is relatively orthogonal. As such, the discovery of novel enzymes that install synthetically useful functional groups site-selectively into complex molecules can serve to bridge these two strategies for generating chemical diversity and allow access to novel structures and classes of compounds [1–3]. Widely used biological building blocks containing useful synthetic functional groups can be interfaced with cellular machinery to form a variety of biomolecules with synthetic handles, thereby merging the structural spaces found in Nature and synthetic chemistry.

While the majority of compounds found in Nature contain only common biological functional groups, natural products containing ‘rare’ functional groups can be found among the secondary metabolites of microbes, plants, and fungi (*Figure 1.1*). These include the alkene [4], alkyne [5], phosphonate [6], N-nitroso [7], cyano [8], diazo [9], and halogen (X = F, Cl, Br, I) [10] functional groups. The enzymes responsible for these transformations provide an avenue for installing ‘rare’ functionality into molecules of interest. Notably, many of these functional groups enable synthetic transformations on biomolecules. For example, the alkyne and alkene functional groups facilitate the azide-alkyne cycloaddition [11,12] and tetrazine ligation [13] reactions, respectively. Another rare functional group with synthetic utility is the halogen [14]. The introduction of halogens into a functional group-dense scaffold is especially useful as it serves as a reactive handle for the selective formation of new chemical bonds as well as a substituent for tuning the activity of the final product. As a result, both synthetic and enzymatic approaches to the formation of carbon-halogen bonds are currently intense areas of interest.

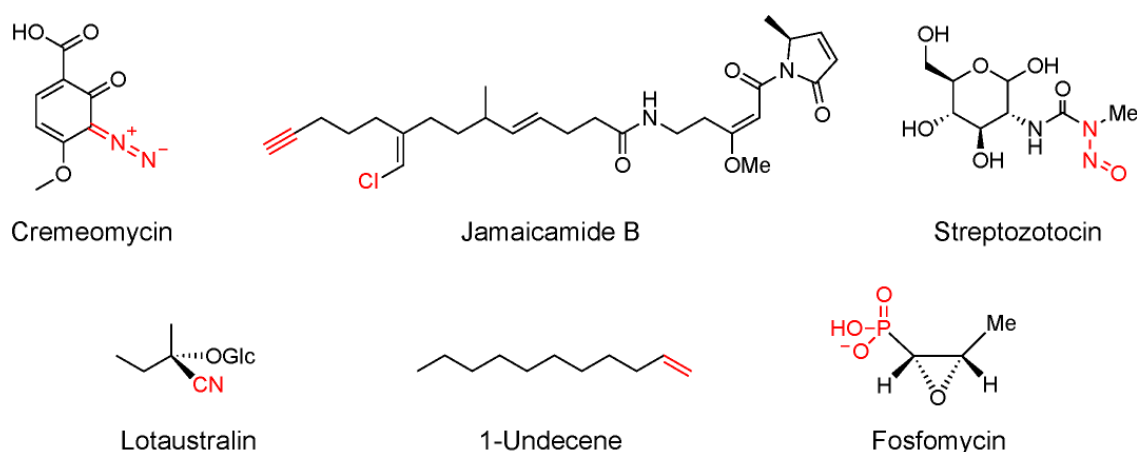


Figure 1.1. Examples of natural products containing ‘rare’ functional groups.

1.2. Enzymatic halogenation strategies

Introduction. Over 4,000 halogenated natural products have been discovered to date. Organochlorines and organobromines comprise the vast majority of these findings, while only ~200 iodinated natural products and a ~30 fluorinated compounds have been identified thus far [10,15,16]. Although the utility of halogens for synthetic chemistry is well-accepted, the role of halogens in natural products continues to be elucidated. In the antibiotic vancomycin, the presence of a halogen is critical for tuning the molecule's bioactivity. In salinosporamide A, the halogen instead serves as a leaving group to facilitate covalent inhibition of a target enzyme. More recently, examples have emerged of biosynthetic pathways that leverage the reactivity of halogens in order to perform chemistry. In the biosynthesis of coronamic acid, chlorine serves as a leaving group to form the cyclopropyl ring in the final product [17], while in cylindrocyclophane biosynthesis, chlorine facilitates aromatic ring alkylation [18]. Given the diversity of organohalogens in Nature, a number of different strategies for enzymatic halogenation have been discovered. These strategies can be divided into three classes depending on the oxidation state of the halogen intermediate: electrophilic halogenation (X^+), nucleophilic halogenation (X^-), and radical halogenation ($X^\cdot$) (Figure 1.2).

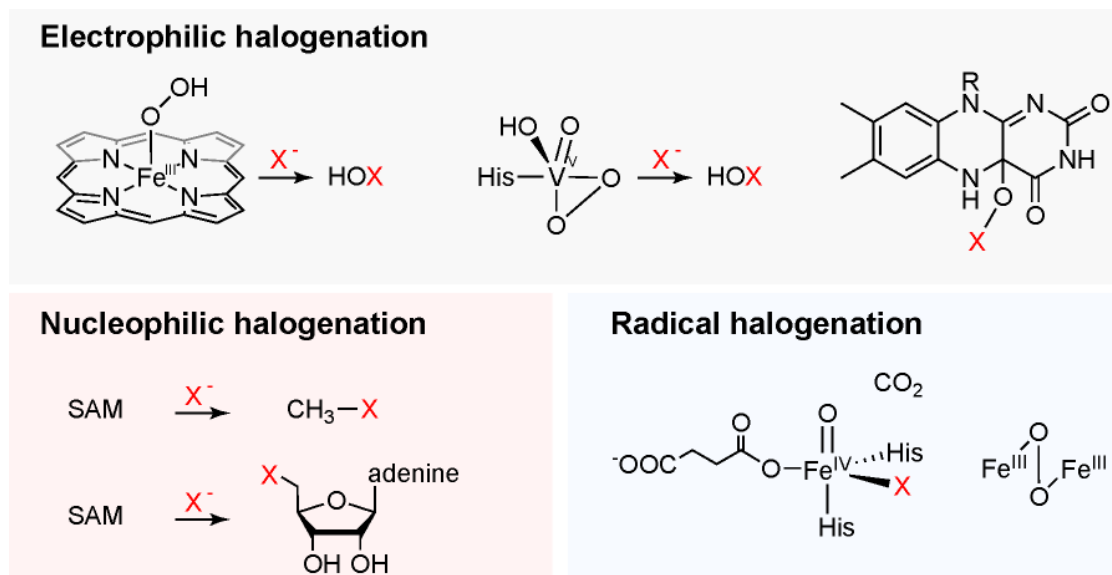


Figure 1.2. Enzymatic strategies for halogenation. Electrophilic halogenases utilize heme, vanadium, or flavin cofactors. Nucleophilic halogenases require S-adenosyl-methionine (SAM) as a substrate. Radical halogenation can be catalyzed by mononuclear Fe/ α KG-dependent enzymes or by putative binuclear Fe-dependent enzymes.

Electrophilic halogenation. Electrophilic halogenases oxidize a halide (X^-) to generate an X^+ equivalent (in the form of hypohalite, XO^-) that is activated for nucleophilic attack by electron-rich substrates. Depending on the enzymatic cofactor, either oxygen (O_2) or hydrogen peroxide (H_2O_2) or can serve as the oxidant. Flavin-dependent halogenases use oxygen to generate the X^+ equivalent, which is then transferred to an active-site lysine residue to form a haloamine species. This lysine residue ultimately directs X^+ to the target substrate, enabling regioselective halogenation [19]. Heme-dependent and vanadium-dependent haloperoxidases use hydrogen peroxide to generate hypohalite. The enzymes can then direct halogenation of electron rich substrates [20], or hypohalite can diffuse from the active site to enable promiscuous halogenation of electron-rich substrates [21]. While electrophilic halogenases have been found to activate

chloride, bromide, and iodide [22], no electrophilic fluorinases have been identified, likely due to the redox potential and high electronegativity of the fluoride ion [23].

Nucleophilic halogenation. Nucleophilic halogenases activate the halide anion (X^-) for nucleophilic attack on the substrate. The enzyme FIA, a fluorinase from *Streptomyces cattleya*, catalyzes nucleophilic attack of fluoride on S-adenosyl-methionine (SAM), with methionine serving as the leaving group. Notably, this is the only reported strategy for enzymatic fluorination [24,25], although nucleophilic chlorinases have been reported as well [26]. Nucleophilic attack on SAM can also result in the methyl halide product [27]. Likely due to the requirement for a strong leaving group, SAM is the only known substrate for nucleophilic halogenation.

Radical halogenation. Radical halogenases catalyze halogenation of unactivated $C_{sp^3}-H$ bonds by generating a halogen radical ($X^\bullet$) that can recombine with a substrate radical. The majority of radical halogenases that have been discovered to date are Fe^{II}/α -ketoglutarate ($Fe^{II}/\alpha KG$)-dependent enzymes in the Cupin superfamily [10]. In this class of enzymes, the catalytic iron is bound by two active site histidine residues, the αKG cofactor, and the halogen to be transferred to the substrate [28]. Upon oxygen binding, the enzyme catalyzes the decarboxylation of the αKG cofactor to succinate, thereby generating a Fe^{IV} -oxo intermediate [29]. This ferryl intermediate abstracts a hydrogen atom from the substrate to generate a substrate radical, which subsequently rebounds with the Fe-bound halogen to yield the halogenated product. When provided with bromide, azide, or nitrite in place of chloride, $Fe^{II}/\alpha KG$ -dependent halogenases have been shown to catalyze bromination, azidation, and nitration [30]. The ability to install a range of functional groups onto unactivated substrates makes this class of enzymes a promising target for engineering. More recently, a new class of radical halogenases, with very low homology to the $Fe^{II}/\alpha KG$ -dependent family, was identified. CylC, a putative binuclear Fe-enzyme, installs chlorine onto an unactivated substrate during cylindrocyclophane biosynthesis [18]. The mechanism of CylC is still under investigation.

1.3. Substrate scope of $Fe^{II}/\alpha KG$ -dependent halogenases

Introduction. $Fe^{II}/\alpha KG$ -dependent halogenases are members of the large and highly diverse Cupin superfamily. Enzymes in the Cupin superfamily catalyze a wide range of reactions, including halogenation, hydroxylation, and stereoinversion on a range of substrates, including small molecules, proteins, and nucleotides [31–35]. Although the Cupin superfamily contains over 360,000 sequences, only 11 radical halogenases had been discovered in the decade between the first reported halogenase in 2005 and the beginning of this dissertation project in 2015. The difficulty in identifying additional halogenases arises from the variability in the Cupin superfamily. Many of the halogenases discovered to date share very low sequence identity and group into different protein families within the superfamily; thus, a simple homology-based bioinformatic search using a known halogenase as a query cannot readily reveal new halogenases with different substrates [36]. The substrates of known radical halogenases can be divided into four groups: carrier protein-dependent halogenases [37], late stage indole-alkaloid halogenases [38], amino acid halogenases (this work), and the recently discovered nucleotide halogenases [39] (*Figure 1.3*).

Carrier protein-dependent halogenases. Nearly all $Fe^{II}/\alpha KG$ -dependent radical halogenases discovered to date only act on carrier-protein bound substrates. SyrB2, the first radical halogenase to be reported, was discovered in 2005 in the process of elucidating the biosynthesis of syringomycin E [37]. SyrB2 chlorinates threonine when it is tethered to the carrier protein SyrB1

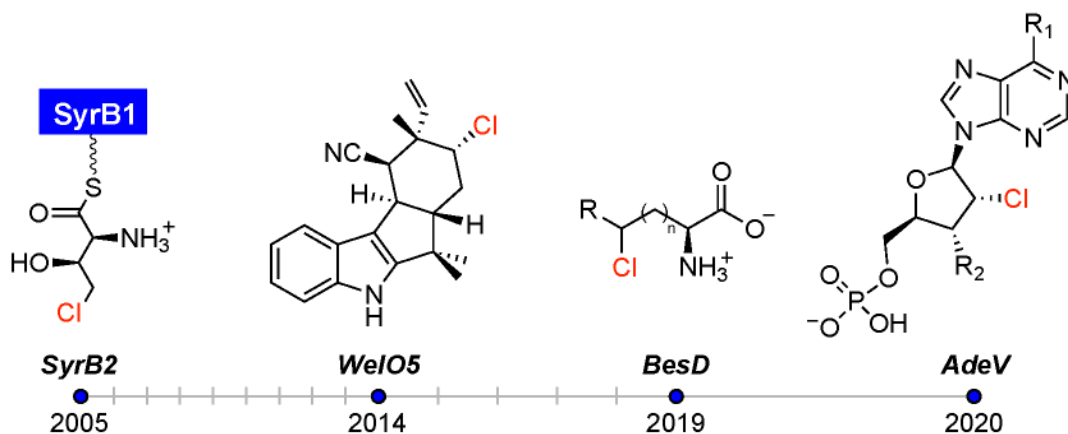


Figure 1.3. Substrates of reported radical amino acid halogenases. The BesD family was discovered during the course of this thesis project.

via a phosphopantetheine arm (aminoacyl-*S*-enzyme Thr-*S*-SyrB1). However, no activity is observed on the freestanding amino acid. Following the discovery of SyrB2, eight other radical halogenases were reported to act on amino-acyl or fatty-acyl substrates [17,40–46]. These halogenases are tailoring enzymes within polyketide and non-ribosomal peptide assembly lines. For nearly a decade, no halogenases were reported to chlorinate an untethered substrate.

Indole alkaloid halogenases. The first halogenase reported to act directly on a free-standing substrate was reported in 2014, nearly 10 years after the discovery of SyrB2. WelO5 is a cyanobacterial indole alkaloid halogenase that chlorinates the welwitindolinone compound 12-*epi*-Fisherindole U [38,47]. WelO5 has extremely low sequence identity to SyrB2 (<15%) and groups into a different protein family (Pfam) within the Cupin superfamily [36]. WelO5 demonstrated nearly 5-10-fold more turnovers before inactivation when compared to SyrB2. Thus, WelO5 is a promising candidate for late-stage modification of compounds. A homolog of WelO5, AmbO5, chlorinates the ambiguine class of indole alkaloids, and was found to be promiscuous in substrate scope [47]. The indole alkaloid halogenases WelO5 and AmbO5 possess a substrate binding lid that confers substrate selectivity. This domain of the protein is the target of promising engineering efforts to develop these enzymes as late-stage synthetic catalysts [48].

Amino acid halogenases. BesD, discovered by our research group, is the first radical halogenase that acts directly on a free-standing amino acid substrate without the requirement for a carrier protein. BesD chlorinates the amino acid lysine to form 4-Cl-lysine as the first step in the biosynthesis of the acetylenic amino acid, β -ethynylserine (β es) [49]. In work described in this dissertation, we additionally discovered several BesD homologs that together enable the synthesis of nine chlorinated amino acid compounds. Due to the simplicity of their amino acid substrates, halogenases in the BesD family hold promise for integrating chlorinated compounds into biosynthetic pathways.

Nucleotide halogenases. Shortly before the time of this writing, the novel halogenase AdeV was reported to act on nucleotide substrates [39]. AdeV installs chlorine at the 2' position of 2'-deoxyadenosine monophosphate in the biosynthesis of adechlorin. The enzyme displays some degree of promiscuity by also chlorinating 2',3'-dideoxyadenosine-5'-monophosphate and 2'-deoxyadenosine 5' monophosphate. The discovery of AdeV provides promising prospects for enzymatic access to nucleotide analogs with potential anticancer and antiviral applications [50].

1.4. Mechanistic investigation of radical halogenases

Introduction. The most well-studied members of the $\text{Fe}^{\text{II}}/\alpha\text{KG}$ -dependent class of enzymes are the hydroxylases, which have been extensively characterized by biochemical, crystallographic, and spectroscopic techniques [51–60]. Hydroxylases possess an HXD/ EX_nH motif and coordinate Fe^{II} with two side chain histidines, one side chain carboxylate, and αKG . Upon oxygen binding, the enzyme decarboxylates αKG to succinate to generate a Fe^{IV} -oxo intermediate, which abstracts a hydrogen atom from the substrate [61]. The resulting substrate radical then rebounds with the Fe^{III} -OH group, yielding a hydroxylated product. $\text{Fe}^{\text{II}}/\alpha\text{KG}$ -dependent halogenases proceed through a similar mechanism. However, instead of the HXD/E motif found in hydroxylases, halogenases possess an HXG/A motif, where the smaller Ala or Gly residue allows room for the chloride to coordinate directly to Fe during the catalytic cycle (*Figure 1.4*). Following hydrogen atom abstraction, the substrate radical selectively rebounds with Fe^{III} -Cl, rather than Fe^{III} -OH, yielding a chlorinated product [62,63]. Since the initial discovery of SyrB2 as an $\text{Fe}^{\text{II}}/\alpha\text{KG}$ -dependent enzyme, several experimental techniques have contributed to the mechanistic understanding of radical halogenation.

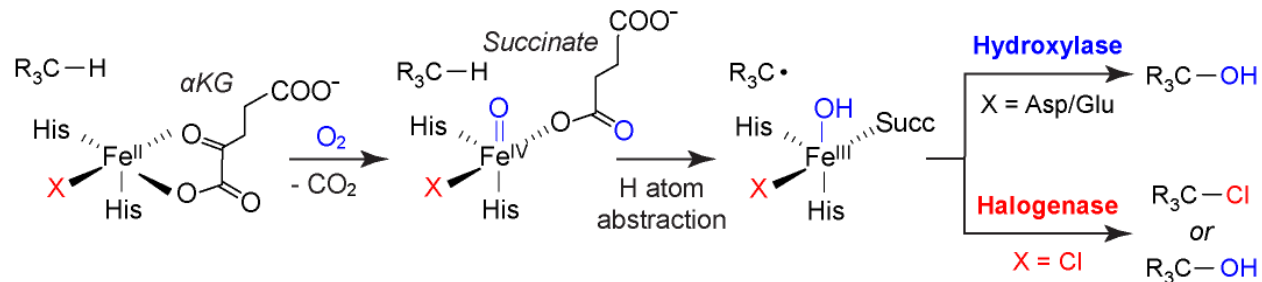


Figure 1.4. Comparison of the active sites of radical halogenases and hydroxylases. While hydroxylases possess an HXD/E motif, halogenases instead possess an HXG/A motif to allow room for chloride to coordinate directly to Fe.

Structural insights into radical halogenase mechanisms. By analogy to the mechanism of radical hydroxylases, Walsh and coworkers proposed that SyrB2 proceeds through a Fe^{IV} -oxo intermediate for hydrogen atom abstraction from the substrate, followed by rebound of an Fe^{III} -Cl intermediate [37]. Critical evidence in support of this mechanism came from the crystal structure of SyrB2, which confirmed that the halogen was indeed directly coordinated to the catalytic Fe [28]. However, due to the challenges associated with co-crystallizing SyrB2 with the carrier protein SyrB1, no substrate-bound crystal structure has been obtained (*Figure 1.5A*). Although several groups pursued substrate modelling and docking studies [64,65], the lack of a co-crystal structure of SyrB2 presented challenges for investigating the mechanism of halogenation.

The discovery of WelO5, which chlorinates a free-standing substrate, enabled the first crystallographic insights into substrate placement relative to the chloride ligand in the active site [66]. Surprisingly, in the anaerobic structure of WelO5, the carbon in the substrate that is targeted for hydrogen atom abstraction is far from the chloride ligand (5.5 Å), while the putative O_2 binding site suggests that the -OH ligand would be much closer (< 4.2 Å) in the absence of ligand rearrangement. The authors therefore proposed that the oxo ligand may be directed away from the substrate by a second-sphere residue (Ser189), thereby enabling competitive halogenation (*Figure 1.5B*). While crystallographic studies of halogenase reactant complexes have provided valuable insights, other methods are required to probe the short-lived intermediates in the catalytic cycle.

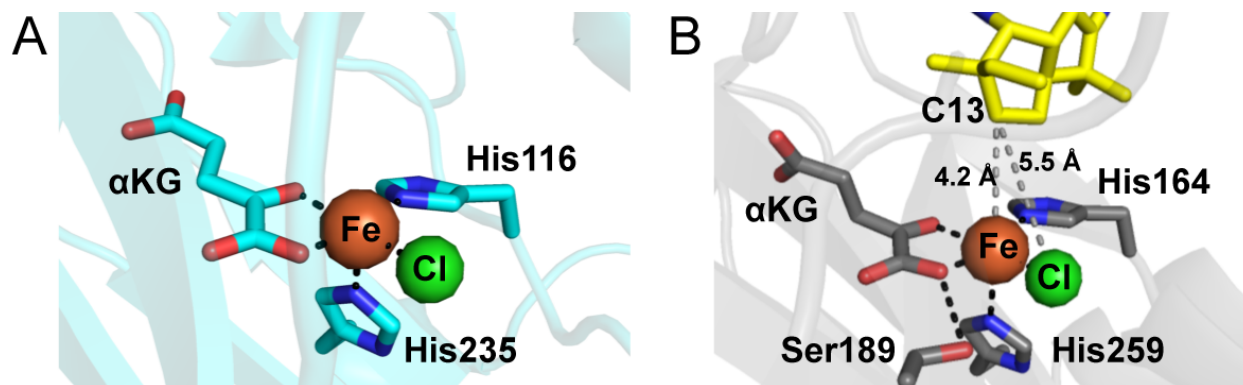


Figure 1.5. Structures of radical halogenases. (A) Structure of SyrB2 (PDB 2FCT) without the substrate bound [28]. (B) Structure of WelO5 (PDB 5IQ7) with substrate bound [66].

Stopped-flow absorption studies. Stopped-flow absorption (SF-abs) experiments can be used to monitor the formation and decay of the Fe^{IV} -oxo species, which absorbs at 320 nm [67]. SF-abs experiments with SyrB2 confirmed that halogenases proceed through a hydrogen atom abstraction mechanism (HAA) and showed that SyrB2 positions its native substrate (Thr) for halogenation by orienting the substrate away from the Fe^{IV} -oxo species, thereby sacrificing the rate of hydrogen atom abstraction in order to increase reaction fidelity [68,69]. Non-native substrates that were expected to be closer to Fe in the active site (norvaline) demonstrated faster rates of hydrogen atom abstraction, but increased hydroxylation during the rebound step. SF-abs experiments suggest that substrate positioning is a critical factor in controlling reaction partitioning to favor halogenation (Figure 1.6).

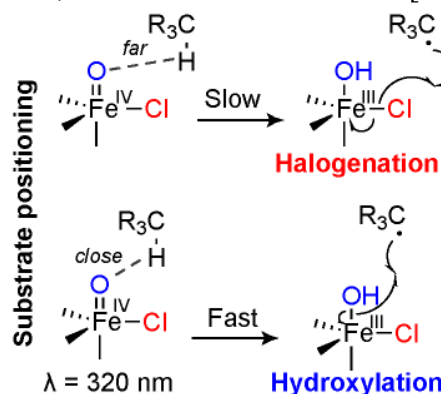


Figure 1.6. SF-abs correlation of HAA rate and reaction outcome.

Spectroscopic characterization of the ferryl intermediate. Freeze quench (FQ) experiments were performed to trap and characterize the ferryl intermediate in halogenases by various spectroscopic methods [70]. Mössbauer spectroscopy of CytC3 and SyrB2 led to the identification of the halogenase ferryl intermediate and supported the hypothesis that halogenases and hydroxylases proceed through similar mechanisms [63]. The Mössbauer spectroscopy also revealed that CytC3 and SyrB2 accumulated two ferryl intermediates that are in equilibrium. These were later proposed to be hydrogen bonded and non-bonded species based on density functional theory (DFT) calculations [71]. The ferryl intermediate was further characterized by nuclear resonance vibrational spectroscopy (NRVS) [71]. DFT calculations to simulate the NRVS spectra of candidate structures revealed that the SyrB2 ferryl intermediate has a trigonal bipyramidal geometry. Calculations also showed that halogenases proceed through a π pathway for HAA, with the substrate C–H bond approximately perpendicular to the Fe^{IV} -oxo bond (Figure 1.7) [71–73]. The π pathway leads to an intermediate in which halogen rebound is competitive with hydroxylation.

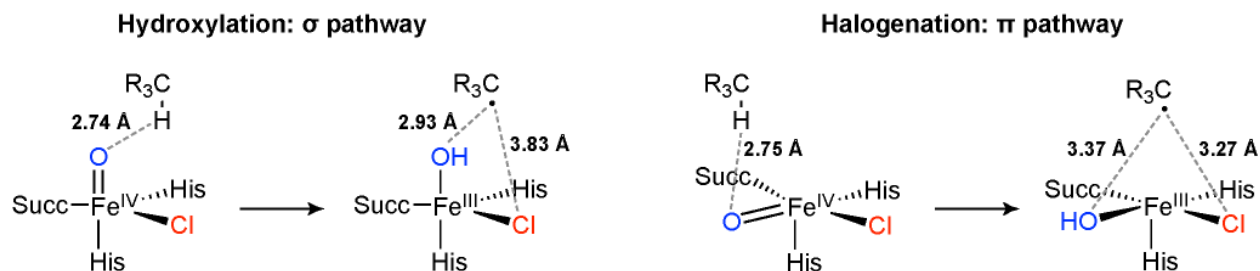


Figure 1.7. Insights into the mechanism of halogenation selectivity in SyrB2. Spectroscopic studies of ferryl intermediate (isolated by freeze-quenching) combined with DFT calculations suggest that hydrogen atom abstraction proceeds through a π pathway in halogenases. Adapted from ref. [71].

Characterization of stable mimics of halogenase intermediates. The transient nature of the reactive intermediates complicates their characterization. Therefore, to complement the spectroscopic studies described above, new strategies for characterizing stable mimics of key catalytic steps have recently been developed. EPR spectroscopy was used to measure distances between the SyrB2[Fe-NO]⁷ complex (a mimic of the [Fe-O₂] complex) and a deuterium on the native (Thr) and non-native substrates [69]. These experiments corroborated the results from SF-

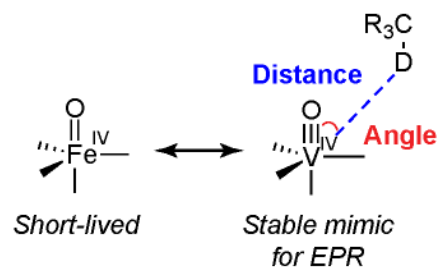


Figure 1.8. Active site geometry measured with stable mimics of intermediates.

abs experiments on SyrB2 which showed that substrate positioning correlates to the reaction outcome. Geometric information can also be obtained by substitution of the Fe^{IV}-oxo with vanadyl (V^{IV}O)²⁺, a stable structural mimic of the unstable Fe^{IV}-oxo intermediate. Both EPR and X-ray crystallography can be used to determine the distances and angles between a deuterated substrate and the vanadyl probe [74] (Figure 1.8). Thus far, vanadyl has been employed to study the arginine hydroxylase VioC and may prove useful for studying radical halogenases in the future. One major challenge to employing these spectroscopic techniques is the requirement for a selectively deuterated substrate. While deuterated amino acids can be obtained and ligated onto the phosphopantetheine arm of SyrB1, selectively deuterated analogs of the indole alkaloid substrate of WelO5 are challenging to obtain.

Experimental challenges. Crystallographic and spectroscopic techniques have enabled significant insights into the mechanism of radical halogenation. However, the SyrB2 and WelO5 systems each present technical challenges toward gaining a complete understanding of halogenation. Ideally, both crystallography and spectroscopy would be performed on the same halogenase system. Although spectroscopic techniques have been very successful in SyrB2, the requirement for a carrier-protein-tethered substrate has prevented substrate co-crystallization. In contrast, crystallographic studies of WelO5 have been successful, but the challenge of obtaining deuterated substrate analogs has limited spectroscopic characterization. Thus, the discovery of new halogenases that are amenable to both characterization strategies could greatly enhance our understanding of radical halogenation.

Halogenase engineering efforts. To date, very few radical halogenases have been discovered in Nature [10]. In contrast, many hydroxylases in the Fe^{II}/αKG-dependent have been identified and found to act on a wide range of substrates [75]. Thus, one strategy for expanding the scope of enzymatic halogenation is to engineer the broader class of hydroxylases to instead perform halogenation. Towards this goal, attempts have been made to enable halogenation by engineering

the HXD/E motif of hydroxylases into the HXG/A required for halogenation. However, this has proven difficult, as these efforts have largely led to no observable halogenation [53,76]. In one promising case, modest halogenation yields were achieved by mutagenesis of the enzyme SadA [77]. However, the majority of the product was still hydroxylated. These efforts indicate that additional factors other than the HXD/E vs HXG/A motifs must contribute to selective rebound of $\text{Fe}^{\text{III}}\text{-Cl}$ rather than $\text{Fe}^{\text{III}}\text{-OH}$.

1.5. Thesis motivation and organization

The discovery and characterization of new radical halogenases may provide insights into strategies for expanding the enzymatic toolkit for C–H halogenation. This goal, along with the background introduced in this chapter, motivates the work in this thesis. In this context, we present studies that expand the scope and understanding of radical halogenases. Chapter 2 describes our discovery of BesD, a novel lysine halogenase that catalyzes the first step in the biosynthesis of an acetylenic amino acid. In Chapter 3 we report the structural characterization of BesD and the bioinformatic identification of several additional members of the new amino acid halogenase family. Chapter 4 describes our investigation of reaction selectivity by comparison of BesD to a closely related hydroxylase homolog. In Chapter 5, we explore the structural basis of regioselectivity and substrate selectivity within the BesD family members. Taken together, this work expands the number of halogenated compounds that can be made enzymatically, contributes to the mechanistic understanding of halogenation, and demonstrates that hydroxylases can be engineered to perform halogenation.

1.6. References

1. Koeller, K. M. & Wong, C.-H. Enzymes for chemical synthesis. *Nature* **409**, 232–240 (2001) doi:10.1038/35051706.
2. Kan, S. B. J., Huang, X., Gumulya, Y., Chen, K. & Arnold, F. H. Genetically programmed chiral organoborane synthesis. *Nature* **552**, 132 (2017) doi:10.1038/nature24996.
3. Savile, C. K., Janey, J. M., Mundorff, E. C., Moore, J. C., Tam, S., Jarvis, W. R., Colbeck, J. C., Krebber, A., Fleitz, F. J., Brands, J., Devine, P. N., Huisman, G. W. & Hughes, G. J. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **329**, 305–309 (2010) doi:10.1126/science.1188934.
4. Rui, Z., Li, X., Zhu, X., Liu, J., Domigan, B., Barr, I., Cate, J. H. D. & Zhang, W. Microbial biosynthesis of medium-chain 1-alkenes by a nonheme iron oxidase. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 18237–42 (2014) doi:10.1073/pnas.1419701112.
5. Edwards, D. J., Marquez, B. L., Nogle, L. M., McPhail, K., Goeger, D. E., Roberts, M. A. & Gerwick, W. H. Structure and Biosynthesis of the Jamaicamides, New Mixed Polyketide-Peptide Neurotoxins from the Marine Cyanobacterium *Lyngbya majuscula*. *Chem. Biol.* **11**, 817–833 (2004) doi:10.1016/j.chembiol.2004.03.030.
6. Wang, C., Chang, W., Guo, Y., Huang, H., Peck, C., Pandelia, E., Lin, G. & H, C. Evidence that the Fosfomycin- Producing Epoxidase , HppE , Is a Non – Heme-Iron Peroxidase. *Science* **342**, 991–995 (2013) doi:10.1126/science.1240373.
7. Ng, T. L., Rohac, R., Mitchell, A. J., Boal, A. K. & Balskus, E. P. An N-nitrosating metalloenzyme constructs the pharmacophore of streptozotocin. *Nature* **566**, 94–99 (2019) doi:10.1038/s41586-019-0894-z.
8. Jensen, N. B., Zagrobelny, M., Hjernø, K., Olsen, C. E., Houghton-Larsen, J., Borch, J., Møller, B. L. & Bak, S. Convergent evolution in biosynthesis of cyanogenic defence compounds in plants and insects. *Nat. Commun.* **2**, 273 (2011) doi:10.1038/ncomms1271.
9. Sugai, Y., Katsuyama, Y. & Ohnishi, Y. A nitrous acid biosynthetic pathway for diazo group formation in bacteria. *Nat. Chem. Biol.* **12**, 73–75 (2016) doi:10.1038/nchembio.1991.
10. Agarwal, V., Miles, Z. D., Winter, J. M., Eustáquio, A. S., El Gamal, A. A. & Moore, B. S. Enzymatic halogenation and dehalogenation reactions: pervasive and mechanistically diverse. *Chem. Rev.* **117**, 5619–5674 (2017) doi:10.1021/acs.chemrev.6b00571.
11. Liang, L. & Astruc, D. The copper(I)-catalyzed alkyne-azide cycloaddition (CuAAC) “click” reaction and its applications. An overview. *Coord. Chem. Rev.* **255**, 2933–2945 (2011) doi:10.1016/j.ccr.2011.06.028.
12. Jewett, J. C. & Bertozzi, C. R. Cu-free click cycloaddition reactions in chemical biology. *Chem. Soc. Rev.* **39**, 1272 (2010) doi:10.1039/b901970g.
13. Blackman, M. L., Royzen, M. & Fox, J. M. Tetrazine Ligation: Fast Bioconjugation Based on Inverse-Electron-Demand Diels–Alder Reactivity. *J. Am. Chem. Soc.* **130**, 13518–13519 (2008) doi:10.1021/ja8053805.
14. Anslyn, E. V. & Dougherty, D. A. *Modern Physical Organic Chemistry*. (University Science, 2006). doi:10.1021/ed083p387.
15. O’Hagan, D. & B. Harper, D. Fluorine-containing natural products. *J. Fluor. Chem.* **100**, 127–133 (1999) doi:10.1016/S0022-1139(99)00201-8.

16. Chan, K. K. J. & O'Hagan, D. The Rare Fluorinated Natural Products and Biotechnological Prospects for Fluorine Enzymology. in *Methods in enzymology* vol. 516 219–235 (2012). doi:10.1016/B978-0-12-394291-3.00003-4.
17. Vaillancourt, F. H., Yeh, E., Vosburg, D. A., O'Connor, S. E. & Walsh, C. T. Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis. *Nature* **436**, 1191–1194 (2005) doi:10.1038/nature03797.
18. Nakamura, H., Schultz, E. E. & Balskus, E. P. A new strategy for aromatic ring alkylation in cylindrocyclophane biosynthesis. *Nat. Chem. Biol.* **13**, 916–921 (2017) doi:10.1038/nchembio.2421.
19. Dong, C., Flecks, S., Unversucht, S., Haupt, C., Pée, K.-H. van & Naismith, J. H. The structure of tryptophan 7-halogenase (PrnA) suggests a mechanism for regioselective chlorination. *Science* **309**, 2216 (2005) doi:10.1126/SCIENCE.1116510.
20. Winter, J. M. & Moore, B. S. Exploring the Chemistry and Biology of Vanadium-dependent Haloperoxidases. *J. Biol. Chem.* **284**, 18577–18581 (2009) doi:10.1074/jbc.R109.001602.
21. Hofrichter, M. & Ullrich, R. Heme-thiolate haloperoxidases: versatile biocatalysts with biotechnological and environmental significance. *Appl. Microbiol. Biotechnol.* **71**, 276–288 (2006) doi:10.1007/s00253-006-0417-3.
22. Gkotsi, D. S., Ludewig, H., Sharma, S. V., Connolly, J. A., Dhaliwal, J., Wang, Y., Unsworth, W. P., Taylor, R. J. K., McLachlan, M. M. W., Shanahan, S., Naismith, J. H. & Goss, R. J. M. A marine viral halogenase that iodates diverse substrates. *Nat. Chem.* **11**, 1091–1097 (2019) doi:10.1038/s41557-019-0349-z.
23. Vaillancourt, F. H., Yeh, E., Vosburg, D. A., Garneau-Tsodikova, S. & Christopher T. Walsh. Nature's Inventory of Halogenation Catalysts: Oxidative Strategies Predominate. **106**, (2006) doi:10.1021/CR050313I.
24. Xiaofeng Zhu, David A. Robinson, Andrew R. McEwan, David O'Hagan, & Naismith, J. H. Mechanism of Enzymatic Fluorination in *Streptomyces cattleya*. (2007) doi:10.1021/JA0731569.
25. Deng, H., Ma, L., Bandaranayaka, N., Qin, Z., Mann, G., Kyeremeh, K., Yu, Y., Shepherd, T., Naismith, J. H. & O'Hagan, D. Identification of fluorinases from *Streptomyces* sp MA37, *Nocardia brasiliensis*, and *Actinoplanes* sp N902-109 by genome mining. *ChemBioChem* **15**, 364–368 (2014) doi:10.1002/cbic.201300732.
26. Eustáquio, A. S., Pojer, F., Noel, J. P. & Moore, B. S. Discovery and characterization of a marine bacterial SAM-dependent chlorinase. *Nat. Chem. Biol.* **4**, 69–74 (2008) doi:10.1038/nchembio.2007.56.
27. Wuosmaa, A. M. & Hager, L. P. Methyl chloride transferase: a carbocation route for biosynthesis of halometabolites. *Science* **249**, 160–2 (1990) doi:10.1126/science.2371563.
28. Blasiak, L. C., Vaillancourt, F. H., Walsh, C. T. & Drennan, C. L. Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature* **440**, 368–371 (2006) doi:10.1038/nature04544.
29. Matthews, M. L., Krest, C. M., Barr, E. W., Vaillancourt, F. H., Walsh, C. T., Green, M. T., Krebs, C., Bollinger, J. M. & Jr. Substrate-triggered formation and remarkable stability of the C-H bond-cleaving chloroferryl intermediate in the aliphatic halogenase, SyrB2. *Biochemistry* **48**, 4331–43 (2009) doi:10.1021/bi900109z.
30. Matthews, M. L., Chang, W., Layne, A. P., Miles, L. a, Krebs, C. & Bollinger, J. M. Direct nitration and azidation of aliphatic carbons by an iron-dependent halogenase. *Nat. Chem. Biol.* **10**, 209–215 (2014) doi:10.1038/nchembio.1438.

31. Bollinger, J. M., Chang, W., Matthews, M. L., Martinie, R. J., Boal, A. K. & Krebs, C. Mechanisms of 2-oxoglutarate-dependent oxygenases: the hydroxylation paradigm and beyond. in *2-Oxoglutarate-Dependent Oxygenases* 95–122 (Royal Society of Chemistry, London, 2015). doi:10.1039/9781782621959.
32. Dunwell, J. M., Purvis, A. & Khuri, S. Cupins: the most functionally diverse protein superfamily? *Phytochemistry* **65**, 7–17 (2004) doi:10.1016/J.PHYTOCHEM.2003.08.016.
33. Chang, W., Guo, Y., Wang, C., Butch, S. E., Rosenzweig, A. C., Boal, A. K., Krebs, C. & Bollinger, J. M. Mechanism of the C5 stereoinversion reaction in the biosynthesis of carbapenem antibiotics. *Science* **343**, 1140–1144 (2014) doi:10.1126/science.1248000.
34. Masson, N., Singleton, R. S., Sekirnik, R., Trudgian, D. C., Ambrose, L. J., Miranda, M. X., Tian, Y.-M., Kessler, B. M., Schofield, C. J. & Ratcliffe, P. J. The FIH hydroxylase is a cellular peroxide sensor that modulates HIF transcriptional activity. *EMBO Rep.* **13**, 251–7 (2012) doi:10.1038/embor.2012.9.
35. Fu, Y., Dai, Q., Dr., Zhang, W., Ren, J., Dr., Pan, T., Prof., He, C. & Prof. The AlkB Domain of Mammalian ABH8 Catalyzes Hydroxylation of 5-Methoxycarbonylmethyluridine at the Wobble Position of tRNA. *Angew. Chem. Int. Ed. Engl.* **49**, 8885 (2010) doi:10.1002/ANIE.201001242.
36. Pandurangan, A. P., Stahlhacke, J., Oates, M. E., Smithers, B. & Gough, J. The SUPERFAMILY 2.0 database: a significant proteome update and a new webserver. *Nucleic Acids Res.* **47**, D490–D494 (2019) doi:10.1093/nar/gky1130.
37. Vaillancourt, F. H., Yin, J. & Walsh, C. T. SyrB2 in syringomycin E biosynthesis is a nonheme Fe(II) α -ketoglutarate- and O₂-dependent halogenase. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 10111–10116 (2005) doi:10.1073/pnas.0504412102.
38. Hillwig, M. L. & Liu, X. A new family of iron-dependent halogenases acts on freestanding substrates. *Nat. Chem. Biol.* **10**, 6–10 (2014) doi:10.1038/nchembio.1625.
39. Zhang, Y., Zhao, C., Yan, S., Li, Q., Zhu, H., Zhong, Z., Ye, Y. & Deng, Z. An iron(II)- and α -ketoglutarate-dependent halogenase acts on nucleotide substrates. *Angew. Chem. Int. Ed. Engl.* anie.201914994 (2020) doi:10.1002/anie.201914994.
40. Galonić, D. P., Vaillancourt, F. H. & Walsh, C. T. Halogenation of unactivated carbon centers in natural product biosynthesis: trichlorination of leucine during barbamide biosynthesis. *J. Am. Chem. Soc.* **128**, 3900–1 (2006) doi:10.1021/ja060151n.
41. Ueki, M., Galonić, D. P., Vaillancourt, F. H., Garneau-Tsodikova, S., Yeh, E., Vosburg, D. A., Schroeder, F. C., Osada, H. & Walsh, C. T. Enzymatic generation of the antimetabolite γ,γ -dichloroaminobutyrate by NRPS and mononuclear iron halogenase action in a streptomycete. *Chem. Biol.* **13**, 1183–1191 (2006) doi:10.1016/j.chembiol.2006.09.012.
42. Neumann, C. S. & Walsh, C. T. Biosynthesis of (–)-(1 S, 2 R)-Allocoronamic Acyl Thioester by an Fe^{II}-Dependent Halogenase and a Cyclopropane-Forming Flavoprotein. *J. Am. Chem. Soc.* **130**, 14022–14023 (2008) doi:10.1021/ja8064667.
43. Jiang, W., Heemstra, J. R., Forseth, R. R., Neumann, C. S., Manaviazar, S., Schroeder, F. C., Hale, K. J. & Walsh, C. T. Biosynthetic Chlorination of the Piperazate Residue in Kutzneride Biosynthesis by KthP. *Biochemistry* **50**, 6063–6072 (2011) doi:10.1021/bi200656k.
44. Khare, D., Wang, B., Gu, L., Razelun, J., Sherman, D. H., Gerwick, W. H., Håkansson, K. & Smith, J. L. Conformational switch triggered by α -ketoglutarate in a halogenase of curacin A biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 14099–104 (2010) doi:10.1073/pnas.1006738107.
45. Flatt, P. M., O'Connell, S. J., McPhail, K. L., Zeller, G., Willis, C. L., Sherman, D. H. & Gerwick, W.

- H. Characterization of the initial enzymatic steps of barbamide biosynthesis. *J. Nat. Prod.* **69**, 938–944 (2006) doi:10.1021/np050523q.
46. Pratter, S. M., Ivkovic, J., Birner-Gruenberger, R., Breinbauer, R., Zangger, K. & Straganz, G. D. More than just a Halogenase: Modification of Fatty Acyl Moieties by a Trifunctional Metal Enzyme. *ChemBioChem* **15**, 567–574 (2014) doi:10.1002/cbic.201300345.
 47. Hillwig, M. L., Zhu, Q., Ittiarnornkul, K. & Liu, X. Discovery of a promiscuous non-heme iron halogenase in ambiguine alkaloid biogenesis: implication for an evolvable enzyme family for late-stage halogenation of aliphatic carbons in small molecules. *Angew. Chemie Int. Ed.* **55**, 5780–5784 (2016) doi:10.1002/anie.201601447.
 48. Hayashi, T., Ligibel, M., Sager, E., Voss, M., Hunziker, J., Schroer, K., Snajdrova, R. & Buller, R. Evolved Aliphatic Halogenases Enable Regiocomplementary C–H Functionalization of a Pharmaceutically Relevant Compound. *Angew. Chemie Int. Ed.* **58**, 18535–18539 (2019) doi:10.1002/anie.201907245.
 49. Marchand, J. A., Neugebauer, M. E., Ing, M. C., Lin, C.-I., Pelton, J. G. & Chang, M. C. Y. Discovery of a pathway for terminal-alkyne amino acid biosynthesis. *Nature* **567**, 420–424 (2019) doi:10.1038/s41586-019-1020-y.
 50. Jordan, P. C., Stevens, S. K. & Deval, J. Nucleosides for the treatment of respiratory RNA virus infections. *Antivir. Chem. Chemother.* **26**, 204020661876448 (2018) doi:10.1177/2040206618764483.
 51. McCusker, K. P. & Klinman, J. P. Modular behavior of TauD provides insight into the origin of specificity in alpha-ketoglutarate-dependent nonheme iron oxygenases. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19791–19795 (2009) doi:10.1073/pnas.0910660106.
 52. Hausinger, R. P. Critical Reviews in Biochemistry and Molecular Biology Fe(II)/ α -Ketoglutarate-Dependent Hydroxylases and Related Enzymes Fe(II)/ α -Ketoglutarate-Dependent Hydroxylases and Related Enzymes. *Crit. Rev. Biochem. Mol. Biol.* **39**, 21–68 (2004) doi:10.1080/10409230490440541.
 53. Gorres, K. L., Pua, K. H. & Raines, R. T. Stringency of the 2-His–1-Asp active-site motif in prolyl 4-hydroxylase. *PLoS One* **4**, e7635 (2009) doi:10.1371/journal.pone.0007635.
 54. Iyer, S. R., Chaplin, V. D., Knapp, M. J. & Solomon, E. I. O₂ Activation by Nonheme Fe^{II} α -Ketoglutarate-Dependent Enzyme Variants: Elucidating the Role of the Facial Triad Carboxylate in FIH. *J. Am. Chem. Soc.* **140**, 11777–11783 (2018) doi:10.1021/jacs.8b07277.
 55. Price, J. C., Barr, E. W., Hoffart, L. M., Krebs, C. & Bollinger, J. M. Kinetic dissection of the catalytic mechanism of taurine: α -ketoglutarate dioxygenase (TauD) from *Escherichia coli*. *Biochemistry* **44**, 8138–8147 (2005) doi:10.1021/bi050227c.
 56. McCusker, K. P. & Klinman, J. P. An Active-Site Phenylalanine Directs Substrate Binding and C–H Cleavage in the α -Ketoglutarate-Dependent Dioxygenase TauD. *J. Am. Chem. Soc.* **132**, 5114–5120 (2010) doi:10.1021/ja909416z.
 57. Elkins, J. M., Ryle, M. J., Clifton, I. J., Dunning Hotopp, J. C., Lloyd, J. S., Burzlaff, N. I., Baldwin, J. E., Hausinger, R. P. & Roach, P. L. X-ray Crystal Structure of *Escherichia coli* Taurine/ α -Ketoglutarate Dioxygenase Complexed to Ferrous Iron and Substrates. *Biochemistry* **41**, 5185–5192 (2002) doi:10.1021/bi016014e.
 58. Price, J. C., Barr, E. W., Tirupati, B., Bollinger, J. M. & Krebs, C. The First Direct Characterization of a High-Valent Iron Intermediate in the Reaction of an α -Ketoglutarate-Dependent Dioxygenase: A High-Spin Fe(IV) Complex in Taurine/ α -Ketoglutarate Dioxygenase (TauD) from *Escherichia coli*. *Biochemistry* **42**, 7497–7508 (2003) doi:10.1021/bi030011f.

59. Hoffart, L. M., Barr, E. W., Guyer, R. B., Bollinger, J. M. & Krebs, C. Direct spectroscopic detection of a C-H-cleaving high-spin Fe(IV) complex in a prolyl-4-hydroxylase. *Proc. Natl. Acad. Sci.* **103**, 14738–14743 (2006) doi:10.1073/pnas.0604005103.
60. Riggs-Gelasco, P. J., Price, J. C., Guyer, R. B., Brehm, J. H., Barr, E. W., Bollinger, J. M. & Krebs, C. EXAFS Spectroscopic Evidence for an Fe=O Unit in the Fe(IV) Intermediate Observed during Oxygen Activation by Taurine: α -Ketoglutarate Dioxygenase. *J. Am. Chem. Soc.* **126**, 8108–8109 (2004) doi:10.1021/ja048255q.
61. Solomon, E. I., Decker, A. & Lehnert, N. Non-heme iron enzymes: contrasts to heme catalysis. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3589–94 (2003) doi:10.1073/pnas.0336792100.
62. Puri, M., Biswas, A. N., Fan, R., Guo, Y. & Que, L. Modeling non-heme iron halogenases: high-spin oxoiron(IV)–halide complexes that halogenate C–H bonds. *J. Am. Chem. Soc.* **138**, 2484–2487 (2016) doi:10.1021/jacs.5b11511.
63. Galonić, D. P., Barr, E. W., Walsh, C. T., Bollinger, J. M. & Krebs, C. Two interconverting Fe(IV) intermediates in aliphatic chlorination by the halogenase CytC3. *Nat. Chem. Biol.* **3**, 113–116 (2007) doi:10.1038/nchembio856.
64. Kulik, H. J. & Drennan, C. L. Substrate placement influences reactivity in non-heme Fe(II) halogenases and hydroxylases. *J. Biol. Chem.* **288**, 11233–41 (2013) doi:10.1074/jbc.M112.415570.
65. Kulik, H. J., Blasiak, L. C., Marzari, N. & Drennan, C. L. First-principles study of non-heme Fe(II) halogenase SyrB2 reactivity. *J. Am. Chem. Soc.* **131**, 14426–33 (2009) doi:10.1021/ja905206k.
66. Mitchell, A. J., Zhu, Q., Maggiolo, A. O., Ananth, N. R., Hillwig, M. L., Liu, X. & Boal, A. K. Structural basis for halogenation by iron- and 2-oxo-glutarate-dependent enzyme WelO5. *Nat. Chem. Biol.* **12**, 636–640 (2016) doi:10.1038/nchembio.2112.
67. Bollinger, J. M. & Krebs, C. Stalking intermediates in oxygen activation by iron enzymes: Motivation and method. *J. Inorg. Biochem.* **100**, 586–605 (2006) doi:10.1016/j.jinorgbio.2006.01.022.
68. Matthews, M. L., Neumann, C. S., Miles, L. a, Grove, T. L., Booker, S. J., Krebs, C., Walsh, C. T. & Bollinger, J. M. Substrate positioning controls the partition between halogenation and hydroxylation in the aliphatic halogenase, SyrB2. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 17723–17728 (2009) doi:10.1073/pnas.0909649106.
69. Martinie, R. J., Livada, J., Chang, W., Green, M. T., Krebs, C., Bollinger, J. M. & Silakov, A. Experimental correlation of substrate position with reaction outcome in the aliphatic halogenase, SyrB2. *J. Am. Chem. Soc.* **137**, 6912–6919 (2015) doi:10.1021/jacs.5b03370.
70. Krebs, C. & Martin Bollinger, J. Freeze-quench ^{57}Fe -Mössbauer spectroscopy: trapping reactive intermediates. *Photosynth. Res.* **102**, 295–304 (2009) doi:10.1007/s11120-009-9406-6.
71. Wong, S. D., Srnc, M., Matthews, M. L., Liu, L. V, Kwak, Y., Park, K., Bell, C. B., Alp, E. E., Zhao, J., Yoda, Y., Kitao, S., Seto, M., Krebs, C., Bollinger, J. M. & Solomon, E. I. Elucidation of the Fe(IV)=O intermediate in the catalytic cycle of the halogenase SyrB2. *Nature* **499**, 320–323 (2013) doi:10.1038/nature12304.
72. Srnc, M. & Solomon, E. I. Frontier molecular orbital contributions to chlorination versus hydroxylation selectivity in the non-heme iron halogenase SyrB2. *J. Am. Chem. Soc.* **139**, 2396–2407 (2017) doi:10.1021/jacs.6b11995.
73. Srnc, M., Wong, S. D., Matthews, M. L., Krebs, C., Bollinger, J. M. & Solomon, E. I. Electronic Structure of the Ferryl Intermediate in the α -Ketoglutarate Dependent Non-Heme Iron Halogenase SyrB2: Contributions to H Atom Abstraction Reactivity. *J. Am. Chem. Soc.* **138**, 5110–5122 (2016)

doi:10.1021/jacs.6b01151.

74. Mitchell, A. J., Dunham, N. P., Martinie, R. J., Bergman, J. A., Pollock, C. J., Hu, K., Allen, B. D., Chang, W.-C., Silakov, A., Bollinger, J. M., Krebs, C. & Boal, A. K. Visualizing the Reaction Cycle in an Iron(II)- and 2-(Oxo)-glutarate-Dependent Hydroxylase. *J. Am. Chem. Soc.* **139**, 13830–13836 (2017) doi:10.1021/jacs.7b07374.
75. Islam, M. S., Leissing, T. M., Chowdhury, R., Hopkinson, R. J. & Schofield, C. J. 2-Oxoglutarate-Dependent Oxygenases. *Annu. Rev. Biochem.* **87**, (2018) doi:10.1146/annurev-biochem-061516-044724.
76. Grzyska, P. K., Müller, T. A., Campbell, M. G. & Hausinger, R. P. Metal ligand substitution and evidence for quinone formation in taurine/ α -ketoglutarate dioxygenase. *J. Inorg. Biochem.* **101**, 797–808 (2007) doi:10.1016/j.jinorgbio.2007.01.011.
77. Mitchell, A. J., Dunham, N. P., Bergman, J. A., Wang, B., Zhu, Q., Chang, W., Liu, X. & Boal, A. K. Structure-guided reprogramming of a hydroxylase to halogenate its small molecule substrate. *Biochemistry* **56**, 441–444 (2017) doi:10.1021/acs.biochem.6b01173.

Chapter 2: *Discovery of a pathway for β -ethynylserine biosynthesis*

Portions of this work were published in the following scientific journal:

Marchand, J. A., Neugebauer, M. E., Ing, M. C., Lin, C. I., Pelton, J. G., Chang, M. C. Y. Discovery of a pathway for terminal alkyne amino acid biosynthesis. *Nature* **2019**, 567, 420-424.

Portions of this work were performed in collaboration with the following persons:

Comparative metabolomics and experiments on BesA/B were performed in collaboration with Dr. Jorge Marchand. Dr. Jorge Marchand performed comparative genomics and knockout studies. Dr. Jeff Pelton performed NMR experiments.

2.1. Introduction

While acetylenic natural products form a relatively large class of compounds (~2,000) [1–3], very few of these contain a terminal alkyne, which is required for the Cu-catalyzed azide-alkyne cycloaddition (CuAAC) reaction (Figure 2.1A). Interestingly, a set of terminal alkyne containing amino acids had been reported to be produced by a select group of fungi and bacteria (Figure 2.1B) [4–9]. Of the 11 discovered terminal-alkyne containing amino acids, three (Bes, Pra, and ethynylglycine) have been reported to be produced by streptomycetes. Notably, terminal-alkyne amino acids, such as Pra, have previously been used for both residue- [10–12] and site-specific [13–18] labeling of proteins and proteomes. However, this previous work relied on terminal-alkyne amino acids which must be chemically synthesized and then exogenously fed to the organism. As such, the identification of the biosynthetic pathway for these amino acids could potentially provide a general route to encode the *in situ* production of terminal alkyne-containing proteins directly from glucose.

Early isotope-labeling experiments showed that most terminal-alkyne compounds originate from acetate-derived fatty acid or polyketide biosynthetic pathways and were postulated to utilize a fatty acid desaturase-like enzyme to form the triple bond [1]. More recently, fatty acid desaturases were also identified to participate in the formation of the terminal alkyne of the polyketide natural products, Jamaicaamide B and Carmabin A [19,20] (Figure 2.1C). The most well characterized acetylenase from these biosynthetic pathways is JamB, a membrane-bound fatty acid acetylenase which desaturates a C6-acyl-ACP. Homologues of JamB have also been discovered to accept a wider range of acyl chain lengths [21]. However, the few terminal-alkyne amino acids described to date are structurally different from products originating from fatty acids and could therefore arise through alternative biosynthetic mechanisms [9]. We thus focused on elucidating the genes for production of β -ethynylserine (β es) in a genetically-tractable soil bacterium, *Streptomyces cattleya*.

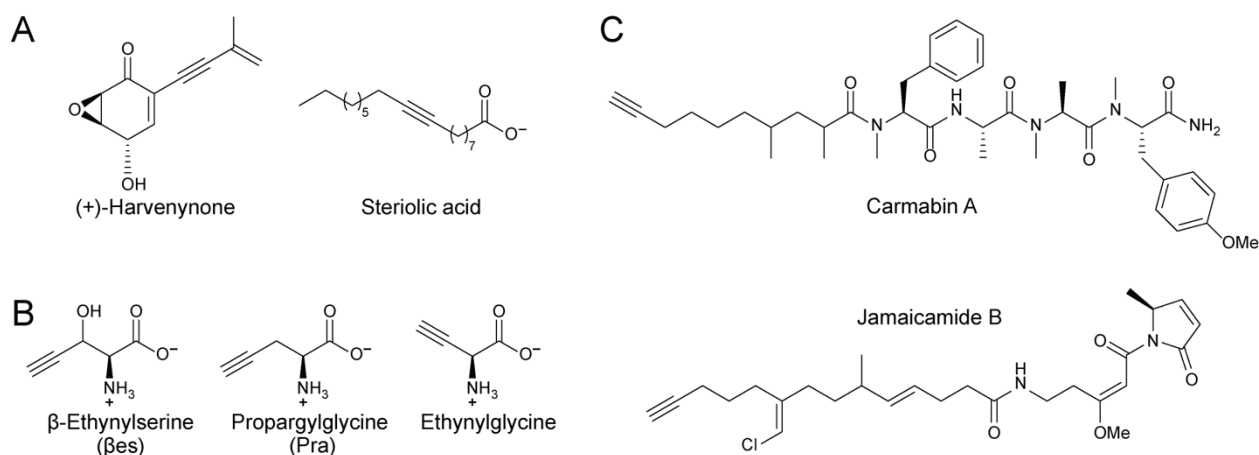


Figure 2.1. Natural products of acetylenic origin. (A) Natural products with internal alkynes. (B) Terminal-alkyne amino acids from streptomycetes. (C) Terminal alkyne natural products with fatty acid biosynthetic origins.

2.2. Materials and methods

Commercial materials. Luria-Bertani (LB) Broth Miller, LB Agar Miller, Terrific Broth (TB), yeast extract, malt extract, and glycerol were purchased from EMD Biosciences (Darmstadt, Germany). Carbenicillin (Cb), kanamycin (Km), chloramphenicol (Cm), isopropyl- β -D-thiogalactopyranoside (IPTG), phenylmethanesulfonyl fluoride (PMSF), tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl), sodium chloride, dithiothreitol (DTT), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), magnesium chloride hexahydrate, acetonitrile, ethylene diamine tetraacetic acid disodium dihydrate (EDTA), and dextrose were purchased from Fisher Scientific (Pittsburgh, PA). Phosphoenolpyruvate (PEP), adenosine triphosphate sodium salt (ATP), adenosine monophosphate sodium salt (AMP), nicotinamide adenine dinucleotide reduced form dipotassium salt (NADH), pyruvate kinase, lactate dehydrogenase, isocitrate dehydrogenase, lysozyme, magnesium sulfate heptahydrate, poly(ethyleneimine) solution (PEI), ammonium iron (II) sulfate hexahydrate, β -mercaptoethanol (β ME), sodium phosphate dibasic heptahydrate, D-mannitol, apramycin sulfate salt (Am), N,N,N',N'-tetramethyl-ethane-1,2-diamine (TEMED), betaine, dimethylsulfoxide (DMSO), pyridoxal-5'-phosphate (PLP), acetonitrile (LC/MS-grade), ammonium formate (LC/MS-grade), 4-amino-3-penten-2-one (Fluoral-P), L-allylglycine (allylglycine), copper sulfate (CuSO_4), Tris(3-hydroxypropyltriazolylmethyl)amine (THPTA), maltose, L-propargylglycine (Pra), and L-lysine, L-glutamate, γ -glutamyl-p-nitroaniline, γ -glutamyl-transpeptidase from equine kidney (Ggt), and glutamate dehydrogenase were purchased from Sigma-Aldrich (St. Louis, MO). 4-Chloro-lysine was purchased from AKos GmbH (Steinen, Germany). (2S)-2-amino-4-chloropent-4-enoic acid (4-Cl-allylglycine) was purchased from Enamine, Ltd (Kiev, Ukraine). (2S)-2-amino-4-bromopent-4-enoic acid (4-Br-allylglycine) and Hygromycin B was purchased from Santa Cruz Biotechnology (Santa Cruz, California). 3-Azido-7-hydroxycoumarin was purchased from AK Scientific, Inc (Union City, CA). Formaldehyde (16% (w/v), methanol-free) and PageRuler Plus Prestained Protein Ladder was purchased from ThermoFisherScientific (Waltham, MA). Succinyl-CoA synthetase was purchased from Megazyme International (Bray, Ireland). Tamra-azide, 5-isomer was purchased from Lumiprobe (Hunt Valley, Maryland). Bacto™ Agar was purchased from BD (Sparks, Maryland) Soy flour was purchased from Berkeley Bowl (Berkeley, CA). Deuterium oxide, L-lysine $\cdot$ 2HCl ($^{13}\text{C}_6$, 99%; $^{15}\text{N}_2$, 99%), L-lysine $\cdot$ 2HCl (4,4,5,5- D_4 , 96-98%), D-dextrose (1- ^{13}C 98%), and L-glutamate (^{15}N , 99%) were purchased from Cambridge Isotope Laboratories (Andover, MA). SuperCos 1 Cosmid Vector Kit was purchased from Agilent (Santa Clara, CA). Formic acid was purchased from Acros Organics (Morris Plains, NJ). Restriction enzymes, T4 DNA ligase, Antarctic phosphatase, Phusion DNA polymerase, T5 exonuclease, and Taq DNA ligase were purchased from New England Biolabs (Ipswich, MA). Deoxynucleotides (dNTPs), were purchased from Invitrogen (Carlsbad, CA). Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA), resuspended at a stock concentration of 100 μM in 10 mM Tris-HCl, pH 8.5, and stored at either 4 $^\circ\text{C}$ for immediate usage or -20 $^\circ\text{C}$ for longer term usage. DNA purification kits and Ni-NTA agarose were purchased from Qiagen (Valencia, CA). SoluLyse reagent was purchased from Genlantis (San Diego, CA). Complete EDTA-free protease inhibitor was purchased from Roche Applied Science (Penzberg, Germany). Strep-Tactin Superflow resin was purchased from IBA GmbH (Göttingen, Germany). PD-10 desalting columns were purchased from GE Healthcare Life, (Pittsburg, PA). Amicon Ultra 3,000 MWCO and 30,000 MWCO centrifugal concentrators and 5,000 MWCO regenerated cellulose ultrafiltration membranes were purchased from Millipore (Billerica, MA). Acrylamide/bis-acrylamide (30%, 37.5:1), electrophoresis grade sodium dodecyl sulfate (SDS), and ammonium persulfate were

purchased from Bio-Rad Laboratories (Hercules, CA). TEV protease was purchased from the University of California, Berkeley Macro Lab (Berkeley, CA).

Bacterial strains. *Streptomyces cattleya* NRRL 8057 (ATCC 35852) was purchased from the American Tissue Type Collection (Manassas, VA). *Streptomyces lavenduligriseus* (NRRL B 3173), *Streptomyces catenulae* (NRRL B2342), *Streptomyces achromogenes* (NRRL B-2120), *Streptomyces* sp. (NRRL S-1448), and *Streptomyces* sp. (NRRL S-31) were provided by Agricultural Research Service Culture Collection: Bacterial Foodborne Pathogens and Mycology Research Unit (National Center for Agricultural Utilization Research, Peoria, IL). *E. coli* DH10B-T1^R was used for plasmid construction. *E. coli* BL21 Star (DE3) was used for heterologous protein production of BesA, BesB, BesC, BesD, BesE. *E. coli* GM272 harboring the non-transmissible, oriT-mobilizing plasmid pUZ8002 was used for conjugative plasmid transfer into *S. cattleya*.

Comparative genomics and clustering analysis. Protein genome sequences and associated protein lists for 27 *Streptomyces* spp. were downloaded from NCBI database (NIH). Three species of Streptomyces have been reported to produce terminal-alkyne amino acids, *S. cattleya*, *S. catenulae*, and *Streptomyces* sp. HLR-599A. However, *S. sp. HLR-599A* is held in a private collection and its genome has not been sequenced; therefore it was not included in this study. An all-by-all protein BLAST was performed between *S. cattleya* (chromosome: NC_015876, plasmid: NC_015875) and *S. catenulae* (NZ_JODY000000000) as the in-group and *S. cattleya* + *S. hygroscopicus*, *S. rimosus*, *S. griseus*, *S. albulus*, *S. filamentosus*, *S. viridochromogenes*, *S. lavendulae*, *S. chartreusis*, *S. ipomoneae*, *S. svinceus*, *S. pristinaespiralis*, *S. bingchengensis*, *S. ghanaensis*, *S. coelicolor*, *S. clavuligerus*, *S. ambofaciens*, *S. grieseoplanus*, *S. scabiei*, *S. avermenitis*, *S. canus*, *S. antibioticus*, *S. roseochromogenus*, *S. cyaneogriseus*, *S. prunicolor*, *S. lydicus* as the out-group. Strict adjacency criteria, within 2 CDSs of best homologue, was used to determine if proteins were clustered in both organisms. Proteins were clustered and visualized using Cytoscape. Proteins that had a lower E-value in the In-group were marked ‘unique’, while proteins with lower E-value in the Out-group were marked ‘common’. Of the initial ~650 clusters of proteins, greater than 3 proteins in size, only 4 were unique between *S. cattleya* and *S. catenulae*. Pipeline and tutorial for comparative clustering analysis is available online as a php script on Github (<https://github.com/jamarchand/io>).

Cell culture of *Streptomyces* spp. for metabolite analysis. Single colonies of *S. catenulae*, *S. sp. NRRL-1448*, *S. sp. NRRL-S31*, *S. achromogenes*, *S. lavendulae*, *S. griseus*, *S. cattleya*, *S. cattleya* Δ SCAT_0184::Am^R, *S. cattleya* Δ SCAT_4823::Am^R, *S. cattleya* Δ SCAT_p1525::Am^R, *S. cattleya* Δ besA::Am^R, *S. cattleya* Δ besB::Am^R, *S. cattleya* Δ besC::Am^R, *S. cattleya* Δ besD::Am^R, *S. cattleya* Δ besE::Am^R, *S. cattleya* Δ besB::Am^R attB^{ΦC31}::ermEp^{*}-besB, *S. cattleya* Δ besC::Am^R attB^{ΦC31}::ermEp^{*}-besC, and *S. cattleya* Δ besD::Am^R attB^{ΦC31}::ermEp^{*} besD were picked from soy-mannitol plates and inoculated into GYM pH 7 (25 mL). After 3 d of growth, cultures were sub-cultured into GYM pH 7 (25 mL), and grown for 7-14 d.

Derivatization and characterization of β -ethynylserine and propargylglycine production in cell culture. Culture supernatants were sampled by removing 1 mL of culture broth and 20 centrifuging at 9,800 $\times$ g for 10 min at room temperature. 100 μ L of supernatant was added to a solution containing 3-azido-7-hydroxy-coumarin (20 μ M), THPTA (100 μ M), and sodium ascorbate (10 mM) with and without CuSO₄ (50 μ M) in 400 μ L of PBS (pH 7). After incubating for 10 min, aliquots were analyzed using an Agilent 1290 UPLC on a Poroshell 120 SB-Aq column (2.7 μ m, 2.1 $\times$ 50 mm; Agilent) using a linear gradient from 5% to 95% acetonitrile over 5 min at a flow rate of 0.6 mL/min with 0.1% (v/v) formic acid, pH 3.5 as the mobile phase. Mass spectra

were acquired using an Agilent 6530 QTOF with the following source and acquisition parameters: Gas temperature = 325 °C; drying gas = 10 l/min; nebulizer = 45 psig; capillary voltage = 3500 V; fragmentor 150 V; skimmer 65 V; oct 1 RF vpp = 750 V; acquisition rate = 3 spectra/s; acquisition time = 333.3 ms/spectrum.

Construction of cosmids for fatty acid desaturase gene disruption in *S. cattleya*. *S. cattleya* NRRL 8057 cosmid library was constructed using the SuperCos 1 Cosmid Vector Kit (Stratagene). Genes *SCAT_p1525*, *SCAT_0184*, *SCAT_2136*, and *SCAT_4823* on cosmids were disrupted by lambda-red mediated recombination with an Am^R/OriT cassette. The Am^R/OriT cassette was amplified from pIJ773 using primers sets LR-SCAT_0184-F/R, LR-SCAT_2136-F/R, LR-SCAT_4823-F/R, and LR-SCAT_p1525-F/R (*Appendix A1.1*). These cassettes were used to disrupt the fatty acid desaturases in their respective cosmids using the REDIRECT method [22]. Apramycin resistant colonies were picked and disruption of *SCAT_0184*, *SCAT_2136*, *SCAT_4823*, *SCAT_p1525* on each cosmid was confirmed by sequencing.

Construction of plasmids for *bes* cluster gene disruption and complementation in *S. cattleya*. Primers for amplifying homology arms for gene disruption were designed to produce amplicons of either 2 kb or 4 kb of homology upstream of start codon and downstream of stop codon for gene of interest. For gene complementation, primers were designed to clone native *S. cattleya besB*, *besC*, or *besD* genes into pSET152 (Hyg^R) under an ermEp* promoter. Standard molecular biology techniques were used to carry out plasmid construction using *E. coli* DH10B-T1^R as the cloning host. All PCR amplifications were carried out with Phusion polymerase using the oligonucleotides listed in *Appendix A1.1*. For gene disruption, pIJ773-BesA, pIJ773-BesB, pIJ773-BesC, pIJ773-BesD, and pIJ773-BesE were constructed by amplification of 2 kb regions from *S. cattleya* genomic DNA using the respective primers in the K773 primer set where K773-BesX-UP-F/R corresponds to the primers used to generate upstream fragment of gene *besX*, while K773-BesX-DOWN-F/R corresponds to the downstream fragment (*Appendix A1.1*). For gene complementation, pSET152-BesB, pSET152-BesC, and pSET152-BesD were constructed using primer sets pSET152-BesX-F/R. Following plasmid construction, all cloned inserts were sequenced at Quintara Biosciences (Berkeley, CA).

Conjugative transfer of plasmids and cosmids from *E. coli* to *S. cattleya*. Plasmids and cosmids for gene disruption or complementation were transferred into *S. cattleya* by conjugation with *E. coli* GM272 harboring the *oriT*-mobilizing plasmid pUZ8002 using the literature protocol³⁴. Gene disruption was verified by PCR amplification of the target locus and neighboring loci using primer set ‘dck’, followed by sequencing to validate correct disruption, generating *S. cattleya* Δ *SCAT_0184::Am^R*, *S. cattleya* Δ *SCAT_4823::Am^R*, *S. cattleya* *SCAT_p1525::Am^R*, *S. cattleya* Δ *besA::Am^R*, *S. cattleya* Δ *besB::Am^R*, *S. cattleya* Δ *besC::Am^R*, *S. cattleya* Δ *besD::Am^R*, and *S. cattleya* Δ *besE::Am^R* (*Appendix A1.1*). Complementation of *S. cattleya* knockout strains was carried out by introducing an expression cassette for the appropriate *bes* gene driven by the ermEp* promoter into the phage attachment site. *S. cattleya* Δ *besB::Am^R*, *S. cattleya* Δ *besC::Am^R*, and *S. cattleya* Δ *besD::Am^R* were transformed respectively with pSET152-BesB, pSET152-BesC, pSET152-BesD to generate *S. cattleya* Δ *besB::Am^R attB^{ΦC31::ermEp*-besB}*, *S. cattleya* Δ *besC::Am^R attB^{ΦC31::ermEp*-besC}*, and *S. cattleya* Δ *besD::Am^R attB^{ΦC31::ermEp*-besD}*. Complementation was verified by PCR using primers designated to amplify the wild-type gene (*Appendix A1.1*).

Extraction of polar metabolites from lysates and supernatants of *Streptomyces* spp. cultures. An aliquot of culture broth (1 mL) was removed and centrifuged at 9,800 $\times$ g for 10 min at room temperature to pellet cells. The supernatant was then diluted 1:1 in methanol with 1% (v/v) formic acid, centrifuged at 9,800 $\times$ g for 10 min to remove precipitate for LC/MS analysis. Culture pellets were then collected after removal of the supernatant and resuspended by vigorous vortexing in a 50% (v/v) solution of methanol (100 μ L of methanol/100 mg of wet cell pellet). These solutions were subjected to five freeze-thaw cycles from -80°C (30 min) to -20°C (30 min) with vortexing in between. Solutions were then centrifuged at 9,800 $\times$ g for 10 min to remove remaining cell debris before LC/MS analysis. Comparative metabolic experiments of *S. cattleya* wildtype, $\Delta besA$, $\Delta besB$, $\Delta besC$, $\Delta besD$, and $\Delta besE$ strains were carried out on cell pellet extracts.

General procedure for high resolution HPLC/MS analysis of polar metabolites. Samples containing polar metabolites were analyzed using an Agilent 1290 UPLC on a SeQuant ZICpHILIC (5 μ m, 2.1 $\times$ 100 mm; EMD-Millipore) using the following buffers: Buffer A (90% acetonitrile, 10% water, 10 mM ammonium formate) and Buffer B (90% water, 10% acetonitrile, 10 mM ammonium formate). A linear gradient from 95% to 60% Buffer A over 17 min followed by a linear gradient from 60% to 33% Buffer A over 8 min was then applied at a flow rate of 0.2 mL/min. Mass spectra were acquired on either positive or negative ionization modes using an Agilent 6530 QTOF (Agilent) with the following source and acquisition parameters: Gas temperature = 325 °C; drying gas = 10 l/min; nebulizer = 45 psig; capillary voltage = 3500 V; fragmentor 150 V; skimmer 65 V; oct 1 RF vpp = 750 V; acquisition rate = 3 spectra/s; acquisition time = 333.3 ms/spectrum.

Comparative metabolite profiling of *S. cattleya* wildtype, $\Delta besA$, $\Delta besB$, $\Delta besC$, $\Delta besD$, and $\Delta besE$ strains. Raw mass spectra data was converted to mzXML format using MZconvert. Peak detection was processed on XCMS platform [23] from the Nomura Lab. Data-collection parameters were set to default HPLC Orbitrap values (centwave feature detection, 0.5 minimum fraction of samples in one group to be a valid group, P-value thresholds for significant feature = 0.05, isotopic ppm error = 5, m/z absolute error = 0.015) except the following: ppm = 20, signal/noise threshold = 2, mz wid = 0.025, and with loess non-linear retention time alignment (default settings). List of m/z values, retention times, peak intensities, and peak areas were exported to Excel for processing.

Construction of plasmids for protein expression. Standard molecular biology techniques were used to carry out plasmid construction using *E. coli* DH10B-T1^R as the cloning host. All PCR amplifications were carried out with Phusion polymerase using the oligonucleotides listed in Appendix A1.1. Following plasmid construction, all cloned inserts were sequenced at Quintara Biosciences (Berkeley, CA) or the Barker Hall Sequencing Facility at UC Berkeley (Berkeley, CA).

Expression of His₆MBP-, His₁₀-, and Strep-tagged proteins. *E. coli* BL21 Star (DE3) was transformed with appropriate protein expression plasmid. An overnight TB culture of the freshly transformed cells was used to inoculate TB (1 L) containing the appropriate antibiotics (50 μ g/mL kanamycin) in a 2.8 L-baffled shake flask to OD₆₀₀ = 0.05. The cultures were grown at 37°C at 200 rpm to OD₆₀₀ = 0.6 to 0.8 at which point cultures were cooled on ice for 20 min, followed by induction of protein expression with IPTG (0.2 mM) and overnight growth at 16°C. For His₁₀-BesE, His₆MBP-BesC, His₁₀-BesD, Strep-BesD, Strep-BesD G139D the medium was supplemented with Fe(SO₄) · 7H₂O (420 μ M) at the time of induction. Cell pellets were harvested by centrifugation at 9,800 $\times$ g for 7 min at 4°C and stored at -80°C.

Expression testing of His₆MBP-tagged BesB variants for solubility. Expression testing of His₆MBP-BesB was performed by the University of California, Berkeley Macro Lab (Berkeley, CA). Vector pSV272.1-BesB was transformed into *E. coli* Rosetta2(DE3) pLysS, *E. coli* BL21(DE3) CodonPlus RIL, *E. coli* BL21 Star (DE3), *E. coli* SoluBL21, and *E. coli* Shuffle T7 Express. Transformants were grown overnight in autoinducing Magic Media (Invitrogen), in a 96-well block at 37°C at 200 rpm. Cultures were then moved to 16°C and allowed to continue growing for an additional 40 h. Cells were lysed using BugBuster (EMD Millipore) and proteins purified using Ni-NTA resin (Qiagen). BL21 Star (DE3) was selected as the production host from these studies based on yield of soluble protein and the lack of need for a secondary antibiotic.

Purification of His₆MBP-BesB. Frozen cell pellets were thawed and resuspended at 5 mL/g of cell paste in lysis buffer (50 mM HEPES, 300 mM NaCl, 20 mM β ME, 10% (v/v) glycerol, pH 7.5) supplemented with phenylmethane sulfonyl fluoride (PMSF, 0.5 mM). The cell paste was homogenized then lysed by passage through a French Pressure cell (Thermo Scientific; Waltham, MA) at 14,000 psi. The lysate was then centrifuged at $13,500 \times g$ for 20 min at 4°C to separate the soluble and insoluble fractions. DNA was precipitated in the soluble fraction with 0.15% (w/v) polyethyleneimine and stirring at 4°C for 30 min. The precipitated DNA was then removed by centrifugation at $13,500 \times g$ for 20 min at 4°C. The remaining soluble lysate was diluted three-fold then passed over amylose column (1 mL resin / 20 g of cell paste) for isolation by gravity flow. Column was washed with wash buffer (50 mM HEPES, 300 mM NaCl, 20 mM β ME, 10% (v/v) glycerol, pH 7.5) for 20 column volumes and eluted with elution buffer (50 mM HEPES, 300 mM NaCl, 20 mM β ME, 10% (v/v) glycerol, 10 mM maltose, pH 7.5). Fractions containing the target protein were pooled by $A_{280 \text{ nm}}$ and concentrated using an Amicon Ultra spin concentrator (100 kDa MWCO, Millipore). Protein was then exchanged into storage buffer (50 mM HEPES, 100 mM NaCl, 20 μ M PLP, 20% (v/v) glycerol, 1 mM DTT, pH 7.5) using a PD-10 desalting column. Based on concerns over solubility, His₆MBP-BesB was used directly for *in vitro* assays. Final protein concentration before storage (1.2 mg/mL) was estimated using the $\epsilon_{280 \text{ nm}}$ calculated by ExPASy ProtParam for His₆MBP-BesB and measured by nanodrop ($\epsilon_{280 \text{ nm}} = 112,440 \text{ M}^{-1} \text{ cm}^{-1}$). Protein was aliquoted, flash-frozen in liquid nitrogen, and stored at -80°C.

Purification of His₁₀-BesA, His₁₀-BesD, His₁₀-BesE, and His₆MBP-BesC. Frozen cell pellets were thawed and resuspended at 5 mL/g of cell paste in lysis buffer (50 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, 20 mM β ME, 10% (v/v) glycerol, pH 7.5) supplemented with PMSF (0.5 mM). The cell paste was homogenized then lysed by passage through a French Pressure cell (Thermo Scientific; Waltham, MA) at 9,000 psi. The lysate was then centrifuged at $13,500 \times g$ for 20 min at 4°C to separate the soluble and insoluble fractions. DNA was precipitated in the soluble fraction with 0.15% (w/v) polyethyleneimine and stirring at 4°C for 30 min. The precipitated DNA was then removed by centrifugation at $13,500 \times g$ for 20 min at 4°C. The soluble lysate was incubated with Ni-NTA (0.5 mL resin/g of cell paste) for 45 min at 4°C, then resuspended and loaded onto a column by gravity flow. The column was washed with wash buffer (50 mM sodium phosphate, 300 mM NaCl, 20 mM imidazole, 20 mM β ME, 10% (v/v) glycerol, pH 7.5) for 15-20 column volumes. The column was then eluted with elution buffer (50 mM sodium phosphate, 300 mM NaCl, 300 mM imidazole, 20 mM β ME, 10% (v/v) glycerol, pH 7.5).

For His₆MBP-BesC and His₁₀-BesD, the protein was dialyzed three times following elution from Ni-NTA column for 1.5 h against dialysis buffer (50 mM HEPES, 100 mM NaCl, 1 mM DTT, pH 7.5) to remove imidazole. After the third round of dialysis, protein was incubated with

TEV protease (1 mg protease/25mg protein) and dialyzed overnight into dialysis buffer. Cleaved and dialyzed protein was passed through a Ni-NTA column (2 mL) to remove TEV and His₆MBP. The eluent was then diluted to a final salt concentration of 20 mM NaCl using buffer A (50 mM HEPES, 20% (v/v) glycerol, 1 mM DTT, pH 7.5) and loaded onto a 5 mL HiTrap-Q column for ion exchange. The protein was eluted using a gradient from 0-100% buffer A to buffer B (50 mM HEPES, 1 M NaCl, 20% (v/v) glycerol, 1 mM DTT, pH 7.5) over 40 min. Fractions containing the target protein were pooled by A_{280 nm} and concentrated using an Amicon Ultra spin concentrator (10 kDa MWCO, Millipore). Protein was then exchanged into storage buffer (50 mM HEPES, 100 mM sodium chloride, 20% (v/v) glycerol, 1 mM DTT, pH 7.5) using PD-10 desalting columns. Final protein concentrations before storage were estimated using the $\epsilon_{280\text{ nm}}$ calculated by ExPASy ProtParam and measured by nanodrop, and are as follows: His₁₀-BesA: 5.6 mg/mL ($\epsilon_{280\text{ nm}} = 40,255\text{ M}^{-1}\text{ cm}^{-1}$), His₁₀-BesE: 2.9 mg/mL ($\epsilon_{280\text{ nm}} = 48,930\text{ M}^{-1}\text{ cm}^{-1}$), BesC: 4.8 mg/mL ($\epsilon_{280\text{ nm}} = 34,380\text{ M}^{-1}\text{ cm}^{-1}$), BesD: 1.6 mg/mL ($\epsilon_{280\text{ nm}} = 47,565\text{ M}^{-1}\text{ cm}^{-1}$). All proteins were aliquoted, flash-frozen in liquid nitrogen, and stored at -80°C.

Purification of Strep-BesD and Strep-BesD G139D. Frozen cell pellets were thawed and resuspended at 5 mL/g of cell paste in lysis buffer (100 mM Tris HCl, 150 mM NaCl, 1 mM EDTA, 20% (v/v) glycerol, 2.5 mM DTT, pH 8.0) supplemented with PMSF (0.5 mM). The cell paste was homogenized then lysed by passage through a French Pressure cell (Thermo Scientific; Waltham, MA) at 9,000 psi. The lysate was then centrifuged at $13,500 \times g$ for 20 min at 4°C to separate the soluble and insoluble fractions. DNA was precipitated in the soluble fraction with 0.15% (w/v) polyethyleneimine and stirring at 4°C for 30 min. The precipitated DNA was then removed by centrifugation at $13,500 \times g$ for 20 min at 4°C. Avidin was then directly added to the soluble lysate (20 µg/L *E. coli* cell culture OD₆₀₀ = 1). Lysate was then passed over Strep-Tactin Superflow resin (0.5 mL resin /g cell pellet) and washed with 10 column volumes of wash buffer (100 mM Tris HCl, 150 mM NaCl, 1 mM EDTA, 20% (v/v) glycerol, 2.5 mM DTT, pH 8.0). Protein was then eluted with 6 column volumes of elution buffer (100 mM Tris HCl, 150 mM NaCl, 1 mM EDTA, 10 mM desthiobiotin, 20% (v/v) glycerol, 2.5 mM DTT, pH 8.0). Fractions containing the target protein were pooled by A_{280 nm} and concentrated using an Amicon Ultra spin concentrator (10 kDa MWCO, Millipore). Protein was then exchanged into storage buffer (50 mM HEPES, 100 mM sodium chloride, 20% (v/v) glycerol, 2.5 mM DTT, pH 7.5) using PD-10 desalting columns. Final protein concentrations before storage were estimated using the $\epsilon_{280\text{ nm}}$ calculated by ExPASy ProtParam and measured by nanodrop and are as follows: Strep-BesD: 8.9 mg/mL ($\epsilon_{280\text{ nm}} = 47,565\text{ M}^{-1}\text{ cm}^{-1}$), Strep-BesD G139D: 7.3 mg/mL ($\epsilon_{280\text{ nm}} = 47,565\text{ M}^{-1}\text{ cm}^{-1}$). All proteins were aliquoted, flash-frozen in liquid nitrogen, and stored at -80°C.

Preparation and purification of 4-Cl-lysine. 4-Cl-lysine was prepared in a 200 µL reaction containing BesD (70 µM), L-lysine · HCl (0.35 mM), αKG (5 mM), (NH₄)₂Fe(SO₄)₂·6H₂O (1 mM), and sodium ascorbate (2.5 mM) in 50 mM HEPES (pH 7.5). Chloride is provided by the enzyme storage buffer (NaCl) and the lysine hydrochloride substrate. Samples were allowed to react for 1 h before quenching in 1.5 volumes of methanol with 1% (v/v) formic acid. 4-Cl-lysine was purified using an Agilent 1200 HPLC on a SeQuant ZIC-HILIC Semi-Preparative column (5 µm, 200 Å, 150 × 21.1 mm, EMD Millipore) using the following buffers: Buffer A (90% acetonitrile, 10% water, 10 mM ammonium formate) and Buffer B (90% water, 10% acetonitrile, 10 mM ammonium formate). A gradient was applied as follows: linear gradient from 100% to 40% Buffer A over 1 h, followed by a hold of 40% Buffer A for 18 min, then to 30% Buffer A over 5 min, followed by a linear gradient to 75% A over 4 min at a flow rate of 5 ml/min. The fractions

containing the 4-Cl-lysine were determined by LC-QTOF MS screening, pooled, and dried by rotary evaporation followed by Speed-vac when the volume reached <1.5 mL.

Preparation and purification of [$^{13}\text{C}_6,^{15}\text{N}_2$]-4-Cl-lysine methyl ester. [$^{13}\text{C}_6,^{15}\text{N}_2$]-4-Cl-lysine methyl ester was prepared in a 200 μL reaction containing BesD (70 μM), [$^{13}\text{C}_6,^{15}\text{N}_2$]-L-lysine $\cdot$ 2HCl (0.35 mM), αKG (5 mM), $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ (1 mM), sodium ascorbate (2.5 mM) in 50 mM HEPES (pH 7.5). Chloride is provided by the enzyme storage buffer (NaCl) and the lysine hydrochloride substrate. The reaction was quenched after 20 min by transferring into 3 mL of 3 M HCl in methanol followed by a 10 min incubation at 50°C. Samples were then cooled at room temperature for 3 h before neutralizing by dropwise addition of 10 M NaOH. Precipitate was removed by centrifugation at $10,000 \times g$ for 10 min. The [$^{13}\text{C}_6,^{15}\text{N}_2$]-4-Cl-lysine methyl ester was purified using an Agilent 1200 HPLC on a SeQuant ZIC-HILIC Semi-Preparative column (5 μm , 200 $\text{\AA}$, 150×21.1 mm, EMD Millipore) using the following buffers: Buffer A (90% acetonitrile, 10% water, 10 mM ammonium formate) and Buffer B (90% water, 10% acetonitrile, 10 mM ammonium formate). A gradient was applied as follows: linear gradient from 100% to 40% Buffer A over 1 h, followed by a hold of 40% Buffer A for 18 min, then to 30% Buffer A over 5 min, followed by a linear gradient to 75% A over 4 min at a flow rate of 5 ml/min. The fractions containing the 4-Cl-lysine methyl ester were determined by LC-QTOF MS screening, pooled, and dried by rotary evaporation followed by Speed-Vac when the volume reached <1.5 mL. Samples were resuspended in a mixture of 60% D_3 -acetonitrile:40% D_2O (v/v) or D_2O for LC/QTOF-MS and NMR analysis, respectively. ESI-MS $[\text{M}+\text{H}]^+$: calculated for $\text{C}^{13}\text{C}_6\text{H}_{16}^{35}\text{Cl}^{15}\text{N}_2\text{O}$ m/z , 203.1037, found m/z 203.1037. $\text{C}^{13}\text{C}_6\text{H}_{16}^{37}\text{Cl}^{15}\text{N}_2\text{O}$ m/z , 205.1007, found m/z 205.1008.

NMR characterization of [$^{13}\text{C}_6,^{15}\text{N}_2$]-4-Cl-lysine methyl ester. All experiments were recorded on a Bruker Avance II spectrometer operating at 900 MHz and at 298 K. The instrument was equipped with a CP TXI cryoprobe. A constant-time ^1H - ^{13}C HSQC experiment (Bruker pulse sequence hsqcctetgpsi) was recorded with carrier frequencies set to 5.48 ppm (^1H) and 40 ppm (^{13}C) and the spectral widths were set to 16 ppm (^1H) and 80 ppm (^{13}C). The recycle delay was set to 1.5 s. A 2D ^1H - ^{13}C version of the HCCH-COSY experiment (Bruker pulse sequence hcchcogp3d2) was recorded with carrier frequencies of 5.48 ppm (^1H) and 39 ppm (^{13}C) with spectral widths of 14 ppm (^1H) and 75 ppm (^{13}C). The recycle delay was set to 1.8 s. To filter out buffer signals, the 1D ^1H spectrum was obtained from the first block of a 2D ^1H - ^{13}C HSQC experiment (pulse sequence hsqcctgpsi2.2) with adiabatic decoupling, using 16 K complex points in the direct dimension and only one point in the indirect ^{13}C dimension. The carrier frequencies were set to 4.7 ppm (^1H) and 40 ppm (^{13}C) and the ^1H spectral width was set to 16 ppm. The FID was recorded for 1.13 sec for each of 128 scans. After zero filling, the final resolution was 0.22 Hz/pt. A 1D ^{13}C spectrum was obtained by direct ^{13}C detection (pulse sequence zgpg30) using Waltz16 decoupling of protons. A total of 32 K complex points were recorded with the carrier set to 102 ppm (^{13}C) and 5 ppm (^1H) and with a ^{13}C spectral width of 200 ppm. The FID was recorded for 0.72 sec for each of 1509 scans. After zero filling, the final resolution was 0.34 Hz/pt. Chemical shifts (ppm) were obtained from 1D-NMR spectra. Please note that the methyl ester is formed by quenching with unlabeled methanol and does not appear using these data collection parameters. ^1H NMR (900 MHz, 100% D_2O): δ 4.25 (H_γ , t, J = 11.5 Hz), 4.16 (H_α , m), 3.31 ($\text{H}_{\epsilon 2}$, m), 3.28 ($\text{H}_{\epsilon 1}$, m), 2.39 ($\text{H}_{\beta 1}$, m), 2.23 ($\text{H}_{\delta 1}$, m), 2.18 ($\text{H}_{\beta 2}$, m), 2.13 ($\text{H}_{\delta 2}$, m). ^{13}C NMR (900 MHz, 100% D_2O): δ 58.4 (C_γ , t, J = 37 Hz), 53.5 (C_α , dd, $J_{\text{C}_\alpha \text{C}_\beta}$ = 35 Hz, $J_{\text{C}_\alpha \text{CO}}$ = 57 Hz), 39.5 (C_ϵ , t, J = 4.5 Hz), 42.2 (C_β , m), 37.7 (C_δ , d, J = 2 Hz).

Preparation and purification of ϵ -lactam cyclization product of 4-chlorolysine (2b). An enzymatic reaction (250 μ L) containing BesD (18 μ M), [$^{13}\text{C}_6,^{15}\text{N}_2$]-L-lysine $\cdot$ 2HCl (5 mM), α KG (10 mM), $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ (1 mM), and sodium ascorbate (2.5 mM) in 50 mM HEPES (pH 7.5) was run for 16 h at room temperature. Chloride is provided by the enzyme storage buffer (NaCl) and the lysine hydrochloride substrate. The reaction was then quenched by transfer into 750 μ L of methanol with 1% (v/v) formic acid. The precipitate was removed by centrifugation at $10,000 \times g$ for 10 min. [$^{13}\text{C}_6,^{15}\text{N}_2$]-4-hydroxy-2-amino- ϵ -lactam was purified using an Agilent 1200 HPLC on a SeQuant ZIC-HILIC Semi-Preparative column (5 μ m, 200 $\text{\AA}$, 150×21.1 mm, EMD Millipore) using the following buffers: Buffer A (90 % acetonitrile, 10% water, 10 mM ammonium formate) and Buffer B (90% water, 10% acetonitrile, 10 mM ammonium formate). A linear gradient from 90% to 10% Buffer A was applied over 1.5 h at a flow rate of 5 mL/min. The peak determined to be the ϵ -lactam by LC/MS was collected manually, lyophilized, and resuspended in 90% d^6 -DMSO:10% D_2O (v/v) for NMR analysis on a Bruker 900 MHz instrument.

NMR characterization of ϵ -lactam cyclization product of 4-chlorolysine (2b). All experiments were recorded on a Bruker Avance II spectrometer operating at 900 MHz and at 298 K. The instrument was equipped with a CP TXI cryoprobe. A constant-time ^1H - ^{13}C HSQC spectrum (Bruker pulse sequence hsqcctetgppsp) was recorded with a ^{13}C evolution period of 13.6 ms, so that signals from ^{13}C attached to two ^{13}C neighbors appeared opposite in phase to those with one ^{13}C neighbor. The carrier frequencies were set to 4.7 ppm (^1H) and 40 ppm (^{13}C) and the spectral widths were set to 16 ppm (^1H) and 80 ppm (^{13}C). The recycle delay was set to 1.5 s. A 2D ^1H - ^{13}C version of the HCCH-COSY experiment (Bruker pulse sequence hcchcogp3d2) was recorded with carrier frequencies of 5.5 ppm (^1H) and 39 ppm (^{13}C) and the spectral widths were set to 14 ppm (^1H) and 75 ppm (^{13}C). The recycle delay was set to 1.8 s. A 2D ^1H - ^{15}N HSQC experiment (Bruker pulse sequence hsqcetf3gpsi) was recorded with carrier frequencies of 4.65 ppm (^1H) and 70 ppm (^{15}N) with spectral widths of 16 ppm (^1H) and 140 ppm (^{15}N). The recycle delay was set to 1.0 s. A 2D ^1H - ^{13}C version of the HNCO experiment (Bruker pulse sequence hncogp3d) was recorded with carrier frequencies of 4.65 ppm (^1H), 119.5 ppm (^{15}N) and 176 ppm (^{13}C) and the spectral widths were set to 16 ppm (^1H) and 22 ppm (^{13}CO). The recycle delay was 1.0 s. Finally, a 2D ^1H - ^{13}C version of the HNCA experiment (pulse sequence hncacbgp3d with ^{13}C - ^{13}C evolution delay set to 20 μ s, effectively making it an HNCA experiment) was recorded with carrier frequencies of 4.65 ppm (^1H), 119.5 ppm (^{15}N), and 40 ppm (^{13}C) and spectral widths were set to 14 ppm (^1H) and 75 ppm (^{13}C). The recycle delay was set to 1.0 s. Chemical shifts (ppm) were obtained from 2D-NMR spectra at a resolution too low to observe coupling constants. ^1H NMR (900 MHz, 90% d^6 -DMSO:10% D_2O): δ 4.11 (H_α), 8.1 (CONH), 3.90 (H_γ), 3.20 ($\text{H}_{\epsilon 1}$), 3.10 ($\text{H}_{\epsilon 2}$), 2.13 ($\text{H}_{\beta 1}$), 2.04 ($\text{H}_{\delta 1}$), 1.69 ($\text{H}_{\beta 2}$), 1.35 ($\text{H}_{\delta 2}$). ^{13}C NMR (900 MHz, 90% d^6 -DMSO:10% D_2O): δ 172.0 ($\text{C}=\text{O}$), 69.4 (C_γ), 48.2 (C_α), 36.6 (C_ϵ), 36.2 (C_β), 35.9 (C_δ). ^{15}N NMR (900 MHz, 90% d^6 -DMSO:10% D_2O): δ 116.3 (CONH).

Detection of ammonia as a coproduct of BesC catalysis. Ammonia was detected using a glutamate dehydrogenase-coupled assay [24]. Reactions (100 μ L) contained BesD (10 μ M), L-lysine $\cdot$ HCl or [$^{13}\text{C}_6,^{15}\text{N}_2$]-L-lysine $\cdot$ 2HCl (10 mM), sodium α KG (5 mM), sodium ascorbate (1 mM), $\text{Fe}(\text{II})\text{SO}_4 \cdot 7\text{H}_2\text{O}$ (0.5 mM), glutamate dehydrogenase (3 U/mL), and NADPH (0.3 mM) in 100 mM HEPES (pH 7.5). Chloride is provided by the enzyme storage buffer (NaCl) and the lysine hydrochloride substrate. Standard curves of 4-Cl-allylglycine, allylglycine, and ^{15}N -glutamate were prepared in the reaction buffer lacking BesD and BesC. Reactions were initiated by addition of BesC (10 μ M) and incubated at room temperature for 2 h before quenching in 1.5 volumes of

methanol with 1% (v/v) formic acid containing 5 μ M Phe as an internal standard before centrifuging at $9,800 \times g$ for 10 min. Samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above. For spectrophotometric analysis of NADH consumption, reactions were prepared as described above except that glutamate dehydrogenase and NADPH were omitted. After 1 h incubation at room temperature, glutamate dehydrogenase and NADPH were added to 3 U/mL and 0.3 mM, respectively, and the A_{340} was monitored on a SpectraMax M2 Microplate Reader (Molecular Devices).

Detection of formaldehyde as a coproduct of BesC catalysis. Formaldehyde was detected using the Fluoral-P reagent [25]. Reactions (100 μ L) contained BesD (5 μ M), L-lysine $\cdot$ HCl or $^{13}\text{C}_6$, $^{15}\text{N}_2$ - L-lysine $\cdot$ 2HCl (5 mM), sodium α KG (5 mM), sodium ascorbate (1 mM), $\text{Fe(II)SO}_4 \cdot 7\text{H}_2\text{O}$ (0.5 mM), and 4-amino-3-penten-2-one (Fluoral-P) (20 mM) in 100 mM HEPES (pH 7.5). Reactions were initiated by addition of BesC (15 μ M). Chloride is provided by the enzyme storage buffer (NaCl) and the lysine hydrochloride substrate. Standard curves of 4-Cl-allylglycine, allylglycine, and formaldehyde were prepared in a reaction mixture omitting BesD and BesC. After incubating for 16 h at room temperature, the reaction product of formaldehyde and Fluoral-P (3,5-diacetyl-1,4-dihydro-2,6-lutidine) was quantified by fluorescence using a BioTek SynergyMx plate reader ($\lambda_{\text{ex}} = 410$ nm, $\lambda_{\text{em}} = 500$ nm) by comparison to a formaldehyde standard curve (0 - 200 μ M). Samples were then quenched in 1.5 volumes of methanol with 1% (v/v) formic acid containing 5 μ M Phe as an internal standard and analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above to quantify the total alkene product formed.

Steady-state kinetic characterization of BesA. Reactions (100 μ L) contained varying concentrations of L-Pra (0.4-5 mM) or L-cysteine (1.25-20 mM), with L-glutamate (10 mM), ATP (2.5 mM), MgCl_2 (5 mM), phosphoenol-pyruvate (PEP, 10 mM), NADH (0.3 mM), lactate dehydrogenase (LDH, 10 U/mL), and pyruvate kinase (PK, 10 U/mL) in 100 mM HEPES (pH 7.5). Reactions were initiated by addition of BesA (6.4 μ M final concentration). Initial rates were measured by monitoring λ_{340} using a SpectraMax M2 Microplate Reader (Molecular Devices) at room temperature. Kinetic parameters (k_{cat} , K_{M}) were determined by fitting the data using Microcal Origin to the equation: $v_0 = v_{\text{max}} [\text{S}] / (K_{\text{M}} + [\text{S}])$, where v is the initial rate and $[\text{S}]$ is the substrate concentration. Error bars on graphs represent mean $\pm$ s.d. ($n = 3$). Error in $k_{\text{cat}}/K_{\text{M}}$ is calculated by propagation of error from the individual kinetic parameters.

In vitro assay for BesA. Reactions (50 μ L) contained L-Pra (2 mM), L-glutamate (5 mM), ATP (2.5 mM), MgCl_2 (5 mM) in 100 mM MOPS buffer (pH 7). Reactions were initiated by addition of BesA (6.4 μ M) and incubated at room temperature for 1 h before quenching in 1.5 volumes of methanol with 1% (v/v) formic acid. Reactions with BesA were compared with the production of γ -Glu-Pra from commercial γ -glutamyl-transpeptidase (Ggt). Ggt reactions (50 μ L) contained L-Pra (2 mM), γ -glutamyl-p-nitroaniline (5 mM) in 100 mM MOPS buffer (pH 7). Reactions were initiated by addition of Ggt (1 U) and allowed to proceed for 1 h before quenching in 1.5 volumes of methanol with 1% (v/v) formic acid. Samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above.

In vitro assay for BesE. BesE was assayed by coupling to either BesA or Ggt in order to produce Glu-Pra in situ. With BesA coupling, reactions (50 μ L) contained BesA (6.4 μ M), L-Pra (2 mM), L-glutamate (5 mM), ATP (2.5 mM), MgCl_2 (5 mM), sodium α KG (5 mM), $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ (1 mM), and sodium ascorbate (1 mM) in 100 mM MOPS buffer (pH 7). With Ggt coupling, BesA and L-glutamate were replaced with Ggt (1 U) and γ -glutamyl-

pnitroaniline (5 mM). Reactions were initiated by addition of BesE (4.4 μ M) and incubated at room temperature for 1 h before quenching in 1.5 volumes of methanol with 1% (v/v) formic acid. Samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above.

***In vitro* reconstitution of Pra biosynthesis from L-lysine with BesB, BesC, BesD.** Reactions (50 μ L) contained BesD (2.9 μ M), BesC (15.7 μ M), BesB (2.4 μ M), sodium α KG (5 mM), sodium ascorbate (5 mM), and $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ (1 mM), and PLP (10 μ M) in 100 mM HEPES (pH 7.5). Chloride is introduced from the enzyme storage buffer (NaCl) and lysine hydrochloride substrate. Reactions were initiated by addition of L-lysine $\cdot$ HCl (4 mM) and incubated at room temperature for 1 h before quenching in 1.5 volumes of methanol with 1% (v/v) formic acid. Samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above.

***In vitro* reconstitution of Glu- β es biosynthesis with BesA, BesB, BesC, BesD and BesE.** Reactions (50 μ L) contained BesD (2.9 μ M), BesC (15.7 μ M), BesB (2.4 μ M), BesA (6.4 μ M), BesE (4.4 μ M), sodium α KG (5 mM), sodium ascorbate (5 mM), and $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ (1 mM), PLP (10 μ M), L-glutamate (5 mM), ATP (2.5 mM), and MgCl_2 (5 mM) in 100 mM HEPES (pH 7.5). Chloride is introduced from the enzyme storage buffer (NaCl) and lysine hydrochloride substrate. Reactions were initiated by addition of L-lysine $\cdot$ HCl (4 mM) and allowed to proceed for 16 h before being quenched in 1.5 volumes of methanol with 1% (v/v) formic acid. Samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis.

Succinate formation by BesE. Reactions (100 μ L) contained ATP (2.5 mM), MgCl_2 (5 mM), phosphoenol-pyruvate (PEP, 1 mM), NADH (0.3 mM), lactate dehydrogenase (LDH, 10 U/mL), pyruvate kinase (PK, 10 U/mL), succinyl CoA synthetase (SCS, 3.2 U/ml), coenzyme A (1 mM), sodium α KG (1 mM), $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ (0.2 mM), and sodium ascorbate (2 mM) in 100 mM MOPS buffer (pH 7). Reactions were initiated by addition of BesD (2.9 μ M) in the presence and absence of Pra. Initial rates of NADH consumption were measured by monitoring A_{340} using a SpectraMax M2 Microplate Reader (Molecular Devices) at room temperature.

Quantification of Fe stoichiometry in BesC. Protein in storage buffer (50 mM HEPES, 100 mM sodium chloride, 20% (v/v) glycerol, 2.5 mM DTT, pH 7.5) was degassed in a glass vial with a rubber septum by sparging with Argon for 15 minutes. After transferring to an anaerobic chamber, BesC was incubated with 5 eq. $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ for 30 min. Then protein was desalted using ZebaSpin desalting columns (7 kD MWCO). Samples were removed from the glovebox and diluted with H_2O to a volume of 50 μ L. Following the addition of 100 μ L of 2M HCl, samples were incubated at 100°C for 30 min. 200 μ L ferrozine (10 mM), 150 μ L saturated ammonium acetate, and 40 μ L sodium ascorbate (14.9 mg/mL) were added to each sample. Precipitated protein was removed by centrifugation at 13,000 $\times$ g for 10 min, and A_{562} of the supernatants was quantified using a plate reader. In parallel, a ferrozine standard curve was generated for Fe (0-180 μ M).

BesC crystallography and structure solution. Apo crystals of BesC from *S. cattleya* were obtained by the hanging drop vapor diffusion method by combining equal volumes of protein solution (10 mg/mL BesC in 50 mM HEPES pH 7.5 and 1 mM DTT) and reservoir solution (calcium acetate (0.6 mM) and PEG 3350 (25% (v/v))). Small hexagonal plate crystals grew in 1 day. Crystals were transferred to an Eppendorf tube containing 100 μ L of seed buffer (calcium

acetate (0.6 mM) and PEG 3350 (25% (v/v))) at 4°C and vortexed for 30 seconds with 10 × 1 mM diameter zirconia/silica beads to produce a micro-seed solution. The seed solution was stored at -80°C for future use. SeMet crystals were prepared by micro-seeding equal volumes of protein solution (10 mg/mL BesC in 50 mM HEPES pH 7.5 and 1 mM TCEP) and reservoir solution (calcium acetate (0.6 mM) and PEG 3350 (25% (v/v))). Small hexagonal plate crystals grew in 1 day. Crystals were flash-frozen in liquid nitrogen and data were collected at Beamline 8.3.1 at the Advanced Light Source (Lawrence Berkeley National Laboratory). For phasing purposes, an energy scan was performed at the Se edge and then data were collected at the following two wavelengths with inverse beam: 0.979620 Å (12656.631 eV, halfway between the peak and the inflection point) and 0.953680 Å (13000.613 eV, high energy remote). Data were processed with XDS[26] and scaled and merged with Aimless[27] within the CCP4 suite[28]. The structure was phased using the experimental phasing module within Phenix. Buccaneer[29] was used to build an initial model of the structure, which was refined iteratively in COOT[30] and Phenix[31]. The structure was analyzed in Pymol, which was also used to create figures.

2.3. Results and discussion

Discovery of β es biosynthetic gene cluster. The reported terminal-alkyne amino acids that have been described are structurally different from products that originate from fatty acids. We thus focused on identifying the genes for production of β -ethynylserine (β es) in a genetically tractable soil bacterium, *S. cattleya*. We initially targeted all four fatty acid desaturases found in *S. cattleya* for gene disruption. However, knockout of the three non-essential fatty acid desaturases (SCAT_4823, SCAT_p1525 and SCAT_0184) did not disrupt β es production, suggesting that β es is not produced by a canonical desaturation (Figure 2.2). We then turned to comparative genomics to identify uniquely shared gene clusters among *Streptomyces* spp. that produce terminal-alkyne amino acids. Of the two other known producers, only *Streptomyces catenulae* had an available genome sequence and was therefore validated for terminal alkyne production (Figure 2.3). When we cultured *S. cattleya*, we detected Pra in cell-pellet metabolite extracts and β es in the supernatant. Since both *S. catenulae* and *S. cattleya* produce Pra, we hypothesized that the genes responsible for Pra biosynthesis would be shared between the two organisms. By comparison of the two genomes, we were able to reduce the roughly 650 total gene clusters conserved between them to just four unique clusters not found in the other 26 control *Streptomyces* spp. genomes (Figure 2.4). Among these four, we focused on one particular cluster that contained gene functions predicted to be associated with amino acid metabolism.

The putative β es biosynthetic cluster features six proteins (Table 2.1). BesA possesses a highly conserved ATP-grasp domain and shares homology with carboxylate-amine ligases. BesB is homologous to the pyridoxal-5'-phosphate (PLP)-dependent cystathionine- β -lyase/cystathionine- γ -synthase family. BesC displays homology to enzymes in the HemeO non-haem iron oxidase superfamily, and both BesD and BesE are predicted to be non-haem Fe/ α -ketoglutarate (α KG)-dependent oxidases. Finally, BesF is a putative EamA-like transporter, homologues of which have been implicated in amino acid efflux.

	Gene	aa	Homologue	%ID/%S	Proposed function
←	besF	297	EamA family transporter [<i>Streptomyces xylophagus</i>]	69/78	Amino acid transporter
→	besA	511	Carboxylate-amine ligase [<i>Cylindrospermopsis raciborskii</i>]	32/50	Amino acid ligase
←	besB	465	Cystathionine beta-lyase [<i>Opitutaceae</i> bacterium TSB47]	42/54	Gamma-lyase
←	besC	252	Pyrroloquinoline-quinone synthase [<i>Marinobacter pelagius</i>]	44/65	Oxidase
←	besD	268	ArpA protein [<i>Pseudomonas azotoformans</i>]	56/71	Halogenase
→	besE	257	Proline hydroxylase [<i>Streptomyces misionensis</i>]	54/66	Hydroxylase

Table 2.1. Genes in β es biosynthetic gene cluster with nearest homologue, percent identity and similarity to nearest homologue (%ID/%S), and proposed function.

Validation of putative β es cluster. By searching the non-redundant protein database for additional instances of these genes we identified five additional *Streptomyces* spp. that had not been reported to produce terminal alkyne amino acids yet were also found to possess the *besABCD* genes (Figure 2.5, 2.6, Appendix A1.3). Three of these new strains were found to produce either Pra or β es in cell culture (Figure 2.7). Based on these results and the comparison of gene clusters, we propose that BesABCD are responsible for Pra biosynthesis and form the core of the cluster, while BesE is related to the conversion of Pra to β es by hydroxylation at the β -position.

Gene	Homologue	%ID/%S	aa
SCAT_0184	Fatty acid desaturase [<i>Streptomyces uncialis</i>]	50/64	346
SCAT_2136	Fatty acid desaturase [<i>Streptomyces megasporus</i>]	83/91	332
SCAT_4823	Acyl-CoA desaturase [<i>Streptomyces</i> sp. NRRL S-813]	98/99	372
SCAT_p1525	Δ Fatty acid desaturase [<i>Streptomyces</i> sp. IMTB 2501]	86/90	361

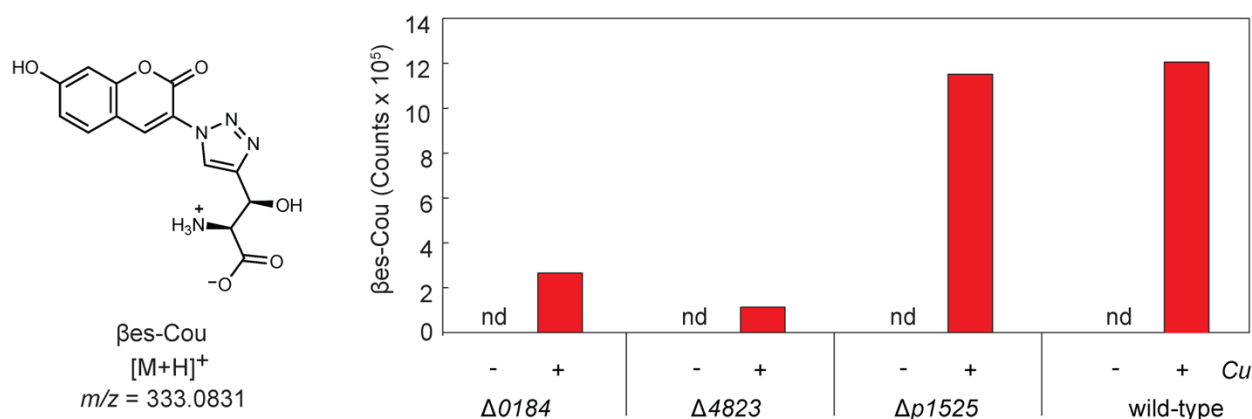


Figure 2.2. (A) A total of four fatty acid desaturases were identified in the genome of *S. cattleya* and targeted for deletion. Percent identity (%ID) and percent similarity (%S) compared to the closest sequenced homologue with the host in brackets. The number of amino acids (aa) in the protein is also tabulated. SCAT_0184 is adjacent to type I polyketide synthase and may be part of a biosynthetic cluster. SCAT_p1525 may also have a metabolic role, although its genomic context is not highly informative. SCAT_2136 and SCAT_4823 are clustered with proteins thought to be involved in primary metabolism and rRNA transcription. We were unable to identify any deletion strains of SCAT_2136 suggesting SCAT_2136 may be an essential fatty acid desaturase. Knockouts of SCAT_0184, SCAT_4823, and SCAT_p1525 were found to still produce β es by CuAAC derivatization with 3-azido-7-hydroxycoumarin followed by LC/QTOF monitoring the β es-coumarin CuAAC product (β es-Cou). The reaction where Cu(I) is omitted provides a negative control since Cu(I) is required to for the cycloaddition to occur.

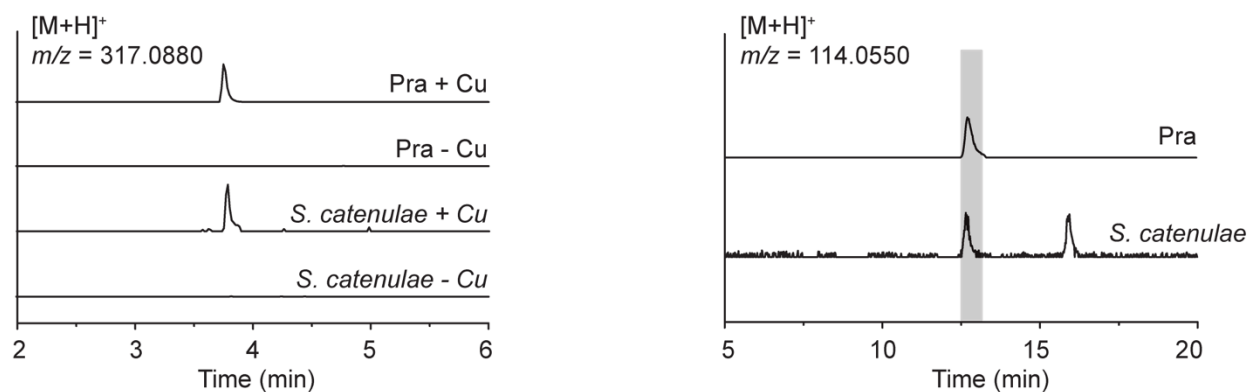


Figure 2.3. Extracted ion chromatograms of the culture supernatant of *S. catenulae* with and without CuAAC-derivatization monitoring either the Pra-coumarin adduct (Pra-Cou; $m/z = 317.0880$) or Pra ($m/z = 114.0550$) compared to a synthetic Pra standard using LC/QTOF. Chromatograms shown are representative of at least 3 experimental replicates.

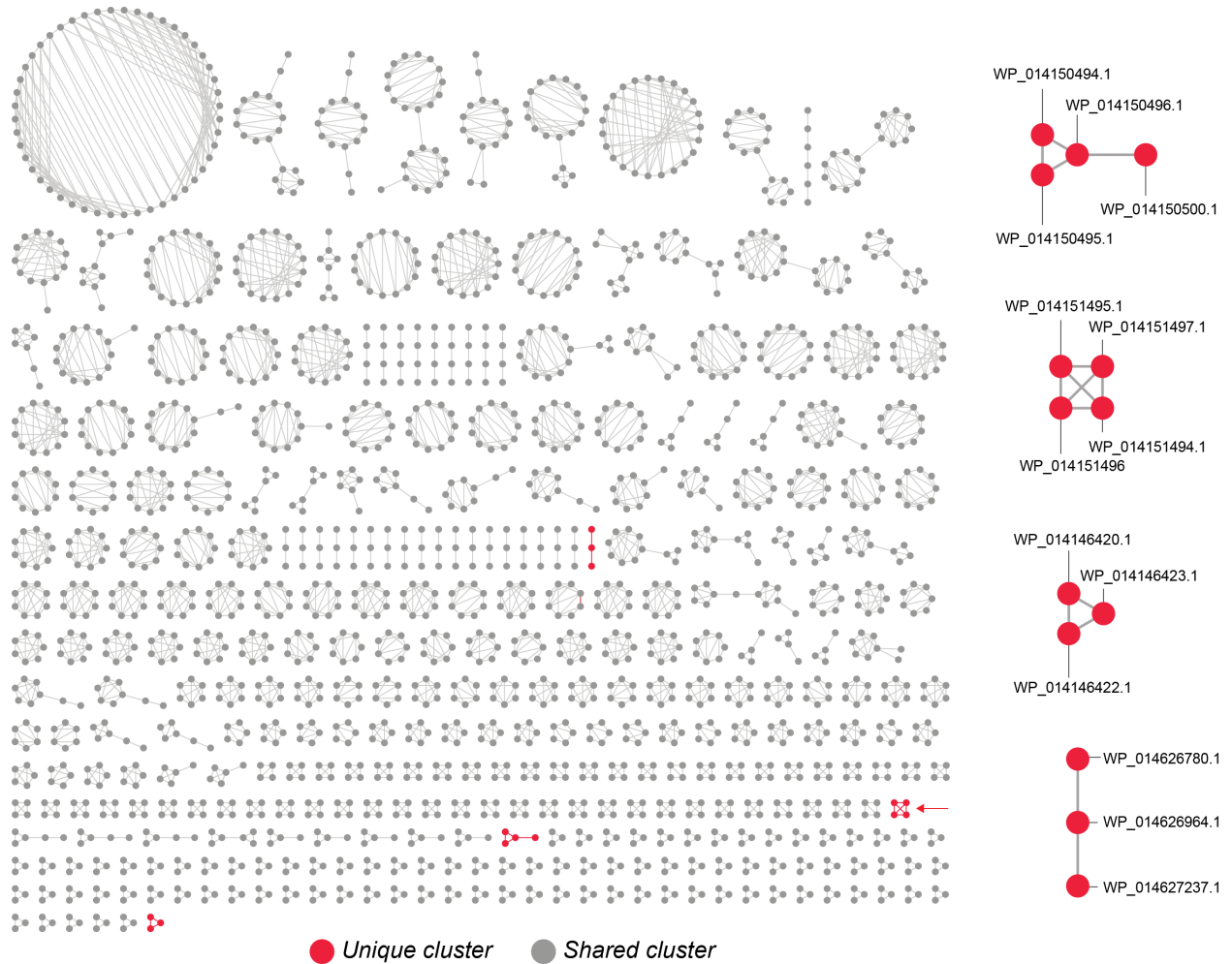
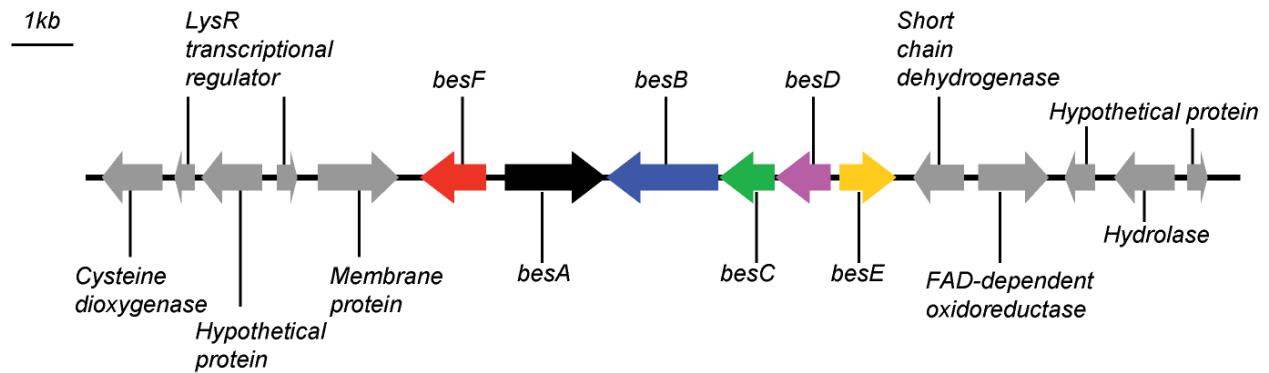


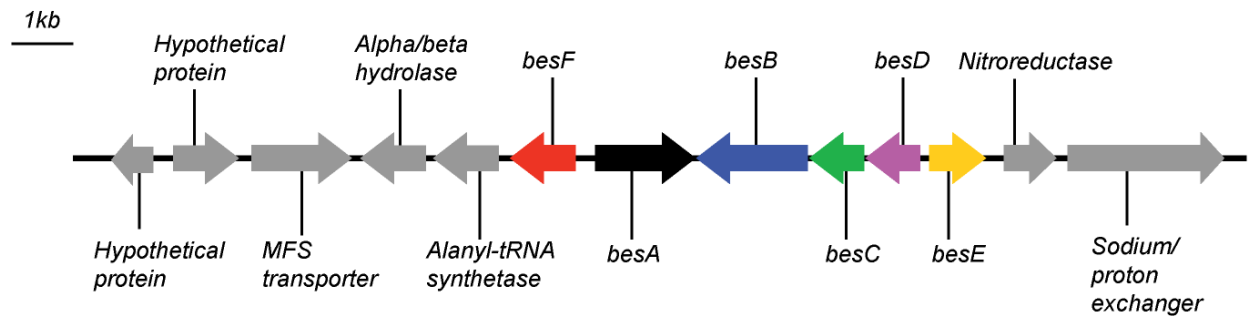
Figure 2.4. We generated a list of approximately 650 gene clusters (> 2 genes) conserved between both species. Clusters were generated using Cytoscape [32]. Nodes are individual genes while edges connect genes that are within 2 protein coding regions from one another. Gene clusters are comprised of connected nodes. Connected grey nodes are gene clusters also found in any the other 26 *Streptomyces spp.* used in the analysis. Connected red nodes are gene clusters found to be unique between *S. cattleya* and *S. catenulae*. We then curated this list by removing gene clusters that are more closely related between *S. cattleya* and 26 other *Streptomyces spp.*, narrowing down the number of conserved unique clusters to just four. One of these unique clusters contains member nodes WP_014151494.1, WP_014151495.1, WP_014151496.1, WP_014151497.1, which composes the putative β es cluster (indicated with a red arrow).

βes/Pra clusters (*bes*ABCDEF)

S. cattleya



S. sp. NRRL S-1448



S. alboverticillatus

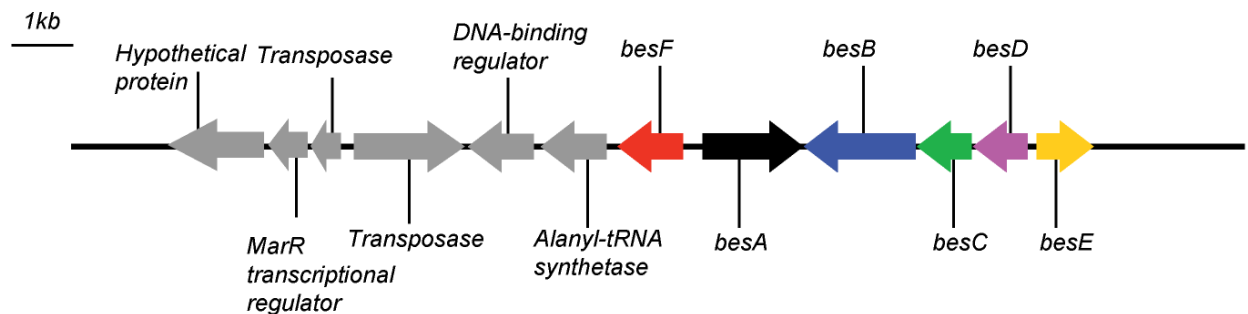
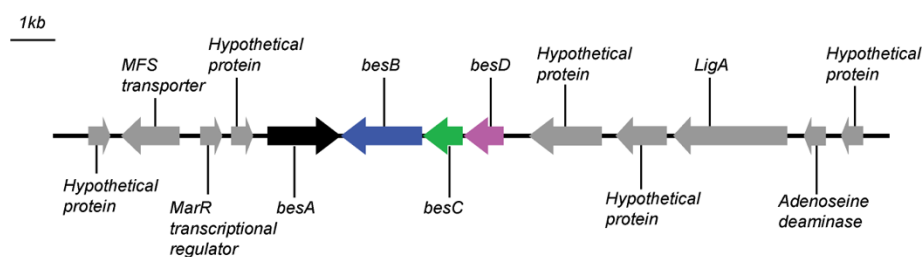


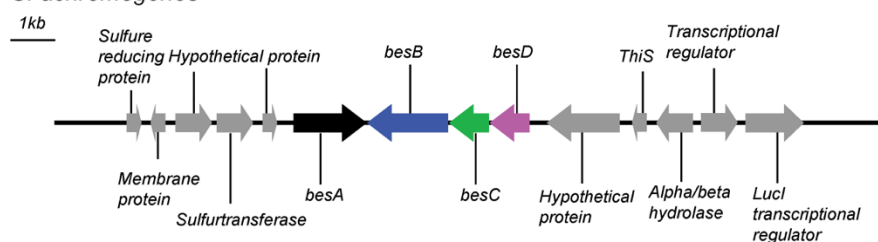
Figure 2.5. Genome clusters and genomic context of βes/Pra biosynthetic clusters (colored) in *S. cattleya*, *S. sp. NRRL S-1448*, *S. alboverticillatus*. The βes/Pra core cluster contains *besE* and *besF* in addition to *besABCD*, whereas the Pra core cluster contain only *besABCD*. The *S. alboverticillatus besF* is predicted based on partial sequencing of the gene from the incomplete genome assembly.

Pra clusters (*besABCD*)

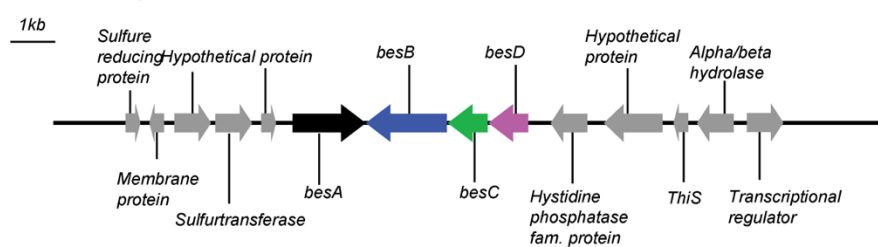
S. catenulae



S. achromogenes



S. lavenduligriseus



S. sp. NRRL S-31

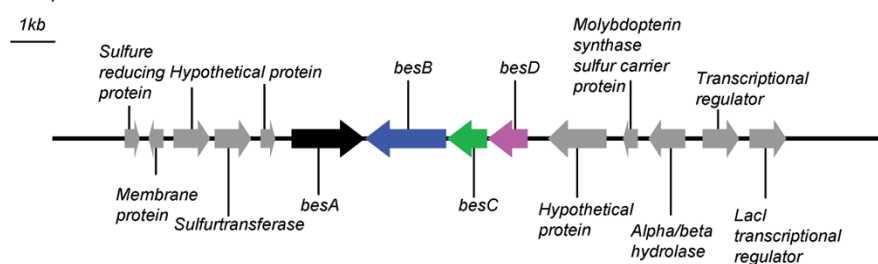


Figure 2.6. Genome clusters and genomic context of Pra biosynthetic clusters (colored) in *S. catenulae*, *S. achromogenes*, *S. lavenduligriseus*, and *S. sp. NRRL S-31*.

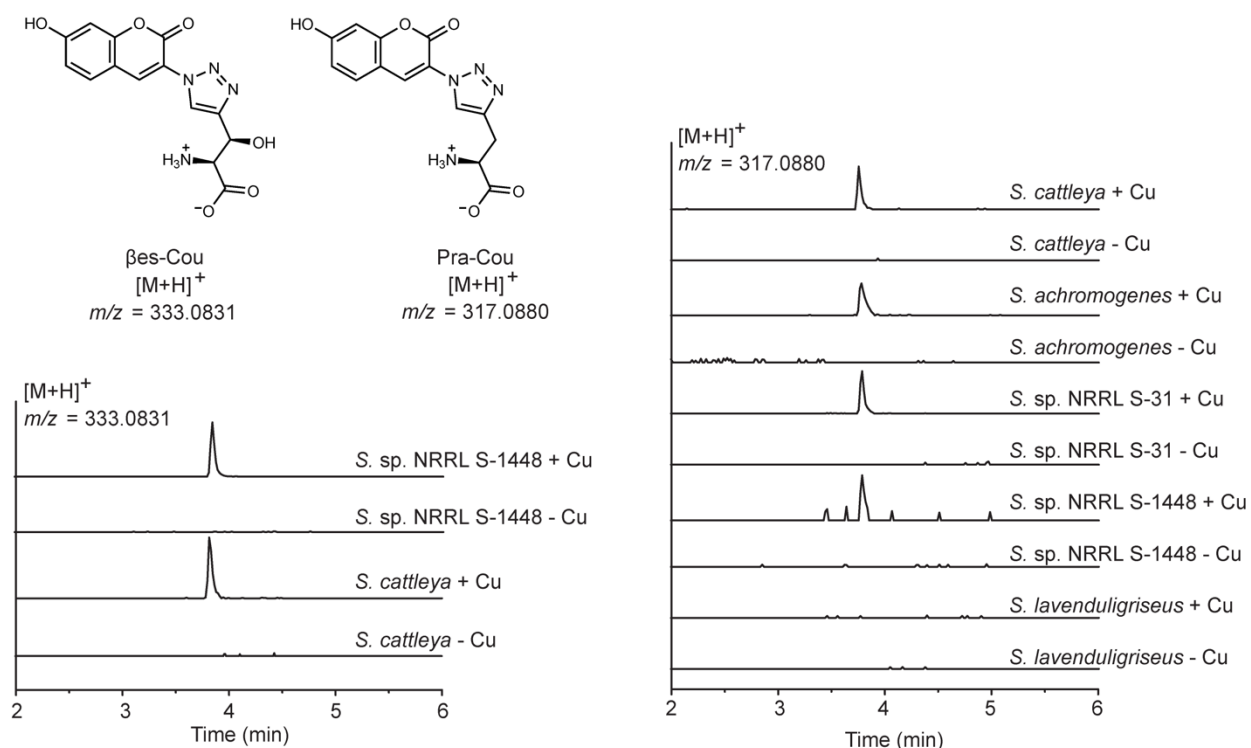


Figure 2.7. Characterization of *Streptomyces* spp. containing the βes/Pra and Pra biosynthetic clusters for their ability to produce terminal alkyne amino acids. The presence of βes and Pra was analyzed by derivatization with 3-azido-7-OH-coumarin by CuAAC and monitoring either the βes-coumarin adduct (βes-Cou; $m/z = 333.0831$) or Pra-coumarin adduct (Pra-Cou; $m/z = 317.0880$) in the presence and absence of Cu(I) using LC/QTOF. The reaction where Cu(I) is omitted provides a negative control since it is required for the cycloaddition to occur. Extracted ion chromatograms of *S. cattleya* and *S. sp. NRRL S-1448* supernatants showing the production of the βes (βes-Cou; $m/z = 333.0831$) (left). Extracted ion chromatograms of *S. cattleya*, *S. achromogenes*, *S. sp. NRRL S-31*, *S. sp. NRRL S-1448*, and *S. lavenduligriseus* supernatants showing the production of Pra (Pra-Cou; $m/z = 317.0880$; right). *S. lavenduligriseus* was cultured but not found to produce Pra. Attempts to amplify genes in the bes cluster failed, suggesting that this particular strain may have lost the bes genes. Chromatograms shown are representative of at least 3 experimental replicates.

The biosynthetic genes *besA–besE* were then targeted for deletion from the genome to generate the *S. cattleya* $\Delta besA$, $\Delta besB$, $\Delta besC$, $\Delta besD$, and $\Delta besE$ knockout strains. Amino acid analysis of the supernatant and intracellular metabolites of these knockouts showed that the $\Delta besB$, $\Delta besC$, and $\Delta besD$ strains no longer produced detectable amounts of Pra or β es (Figure 2.8). Complementation studies further support the assignment of terminal-alkyne amino acid production to the β es gene cluster (Figure 2.8). In comparison, the $\Delta besA$ strain showed lowered yields of both products, and the $\Delta besE$ strain still produced Pra but not β es. These results support the hypothesis that *besB*, *besC* and *besD* are crucial for formation of Pra, and suggest that BesE is the hydroxylase that converts Pra to β es.

Identification of pathway intermediates through comparative metabolomics. To identify pathway intermediates, we performed an untargeted comparative metabolomics experiment on the intracellular metabolite extracts of *S. cattleya* wild-type, $\Delta besA$, $\Delta besB$, $\Delta besC$, $\Delta besD$, and $\Delta besE$ strains using high resolution liquid chromatography with mass spectrometry (LC–MS). The resulting mass spectra were processed using XCMS [23] for feature detection to generate an initial set of about 1,000 lead compounds with high statistical significance ($P < 0.05$) and high fold change (>2) between wild-type and knockout strains. By focusing on the metabolites that accounted for the highest degree of variance, we were able to identify a set of putative intermediates and precursors for β es biosynthesis (Figure 2.9A). On the basis of these findings, we hypothesized that β es biosynthesis is unexpectedly initiated by chlorination of lysine (1) at the C_γ position by the BesD halogenase (Figure 2.9B, Figure 2.10). The resulting 4-Cl-lysine (2) is then subject to an unusual oxidative cleavage by the BesC oxidase to form 4-Cl-allylglycine (3). The PLP-dependent enzyme BesB can next catalyse γ -elimination of chloride from this intermediate followed by isomerization to form the first terminal-alkyne product, Pra (4). We propose that Pra is ligated to glutamate by BesA to form a γ -glutamyl-dipeptide (5), which serves as the substrate for the BesE hydroxylase to produce the corresponding β es dipeptide (6) in the last step. As no other conserved gene was found in these clusters, it is possible that a non-pathway-specific peptidyl-transglutaminase or amide hydrolase is responsible for the release of free Pra and β es.

BesD is a lysine halogenase. The sequence of BesD contains the characteristic HXG motif that is required for halide coordination to the active-site iron in the Fe/ α KG-dependent halogenase family, rather than the HXD/E motif of the related family of hydroxylases (Figure 2.11) [33]. Indeed, when lysine was incubated with purified BesD, the major product of the reaction exhibited a characteristic Cl isotopic pattern and could be assigned as Cl-lysine (Figure 2.12). To further support the assignment of BesD as a halogenase, we mutated G139, from the HXG motif, to aspartate. This abolished the formation of Cl-lysine and resulted solely in the formation of hydroxylysine (Figure 2.13).

Although LC/MS enabled the assignment of the product Cl-lysine, the regioselectivity remained unknown. We therefore attempted to purify Cl-lysine for characterization by NMR. However, this proved challenging due to the inherent instability of the compound, which decomposed to several additional products that we detected by LC/MS (Figure 2.14). We were able to verify that the degradation products were formed non-enzymatically by making Cl-lysine using purified BesD, filtering the enzyme out of the reaction using size-exclusion spin filters, and tracking the products by LC/MS thereafter. We observed a decrease in Cl-lysine over time, while the three degradation products increased. The rate of degradation of Cl-lysine was slightly higher

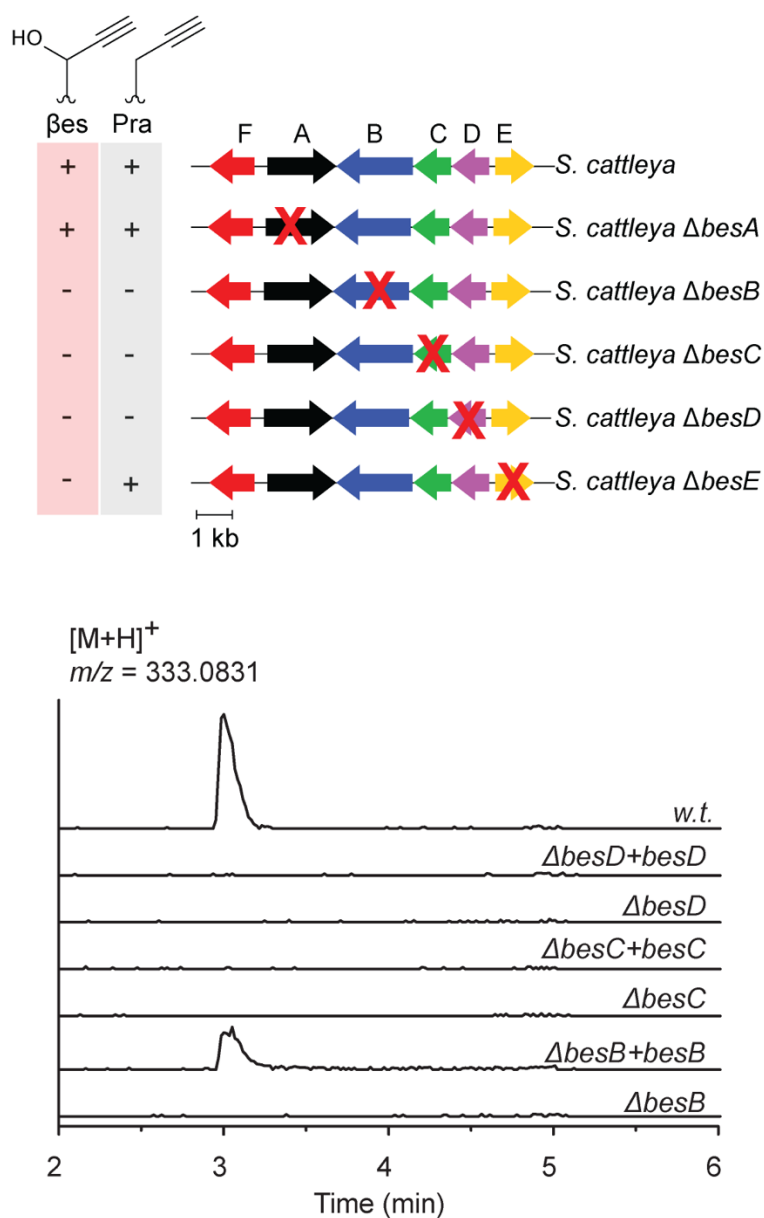


Figure 2.8. (Top) Terminal alkyne production phenotypes of knockouts of each gene in the β es biosynthetic gene cluster from *S. cattleya*. (Bottom) *S. cattleya* knockout strains were complemented with the corresponding wild-type gene expressed from the ErmEp* promoter by insertion into the phage attachment site in the genome. Knockout and complementation strains were cultured for 7 d. The supernatants were derivatized with 3-azido-7-OH-coumarin by CuAAC and analyzed for the β es-coumarin adduct (β es-Cou; $m/z = 333.0831$) by LC/QTOF. Only the *besB* complementation strain appears to recover β es production, possibly due to polar effects related to ordering of the genes in the putative operon (*besD-besC-besB*).

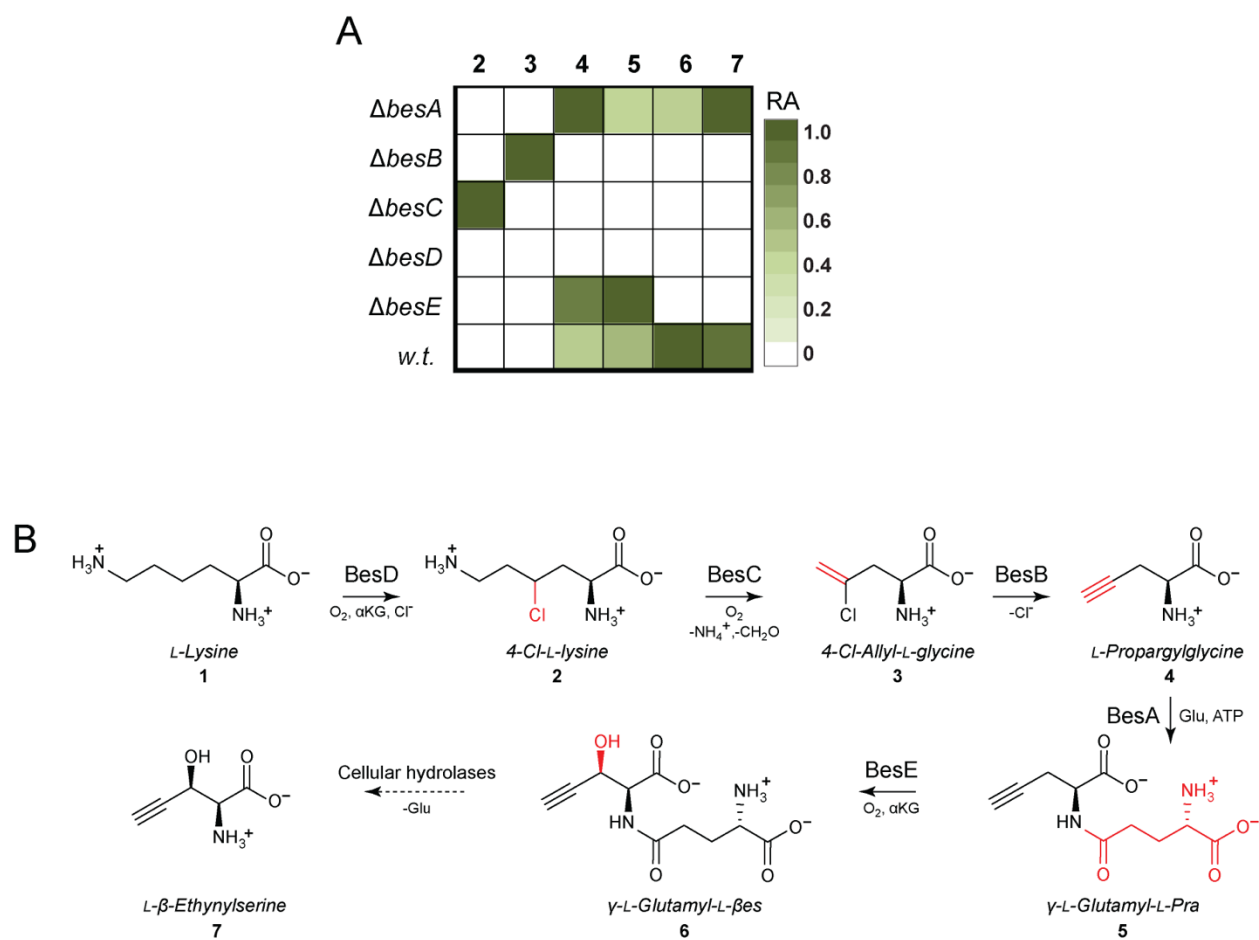


Figure 2.9. Biosynthetic pathway for production of β es from lysine. (A) Relative abundance of pathway amino acids found in *S. cattleya* knockouts strains, by average integrated peak area of each compound (n = 5) and normalized by relative abundance (RA) within each species (green, more abundant; white, less abundant). (B) Putative pathway for biosynthesis of β -ethynylserine (β es) from lysine.

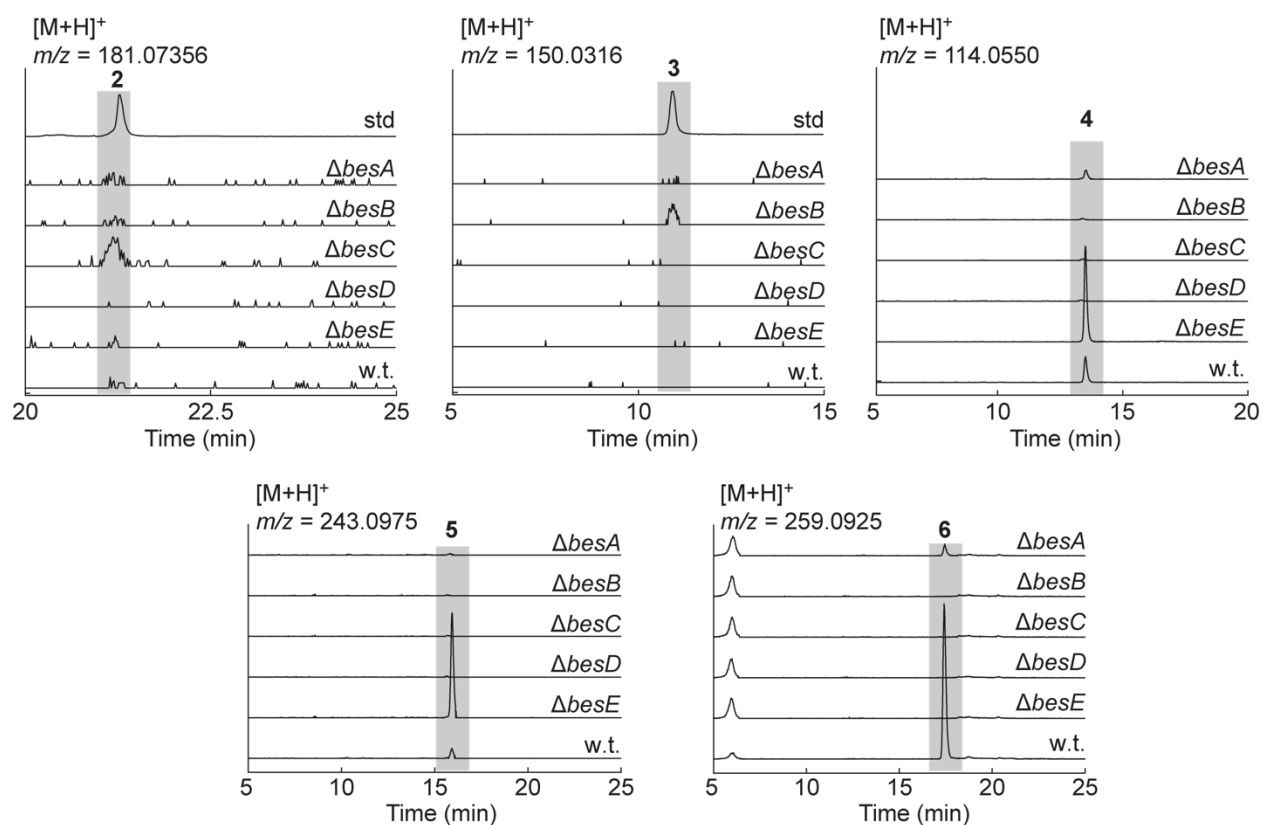


Figure 2.10. Extracted ion chromatograms of pathway intermediates identified using comparative metabolomics from *S. cattleya* pellet extracts for 4-Cl-lysine (**2**, $m/z = 181.0738$), 4-Cl-allylglycine (**3**, $m/z = 150.0316$), propargylglycine (**4**, $m/z = 114.0550$), γ -Glu-Pra (**5**, $m/z = 243.0975$), and γ -Glu-Bes (**6**, $m/z = 259.0925$). Chromatograms shown are representative of at least 5 independent biological replicates.

Halogenase							H X A/G	
D	W	H	Q	A	D	T	SyrB2	
A	V	H	V	G	N	D	WeIO5	
G	W	H	W	G	D	F	BesD	

Hydroxylase							H X D/E	
N	W	H	T	D	V	T	TauD	
K	A	H	Y	D	Y	F	3ITQ	
F	I	H	Y	D	D	I	BesE	

Figure 2.11. BesD is a Fe/ α KG-dependent halogenase with a conserved HXG halogenation motif.

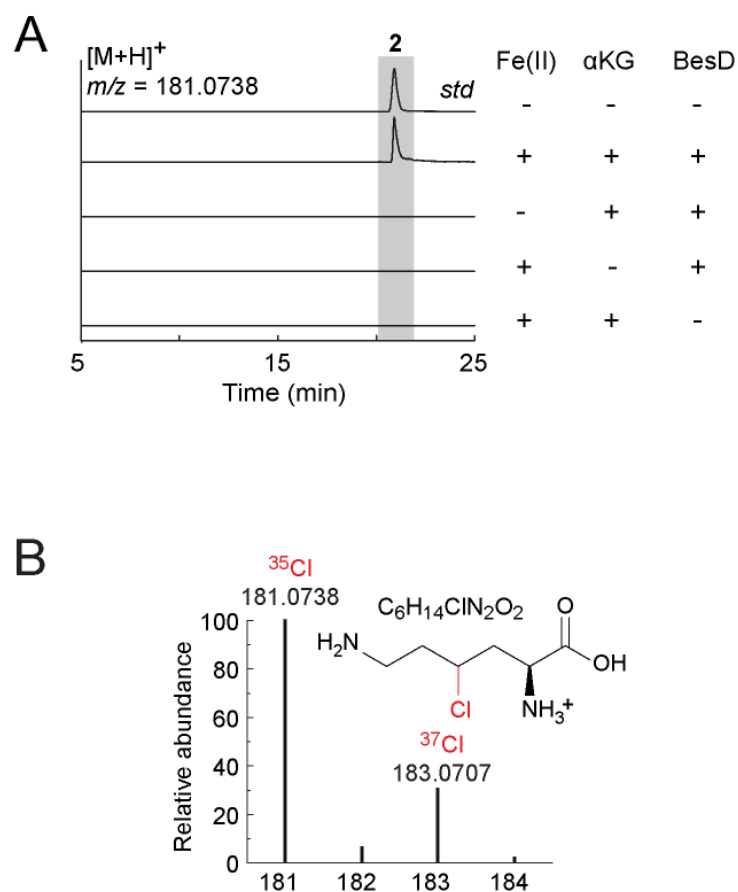


Figure 2.12. (A) BesD activity requires lysine, Cl^- , Fe(II) and αKG (4-Cl-lysine, **2**; m/z = 181.0738). Chromatogram shown are representative of at least 3 independent experimental replicates. (B) The mass spectra of the product formed when BesD is incubated with lysine and Cl^- displays the characteristic chlorine isotope pattern. Spectra shown is representative of at least 3 experimental replicates.

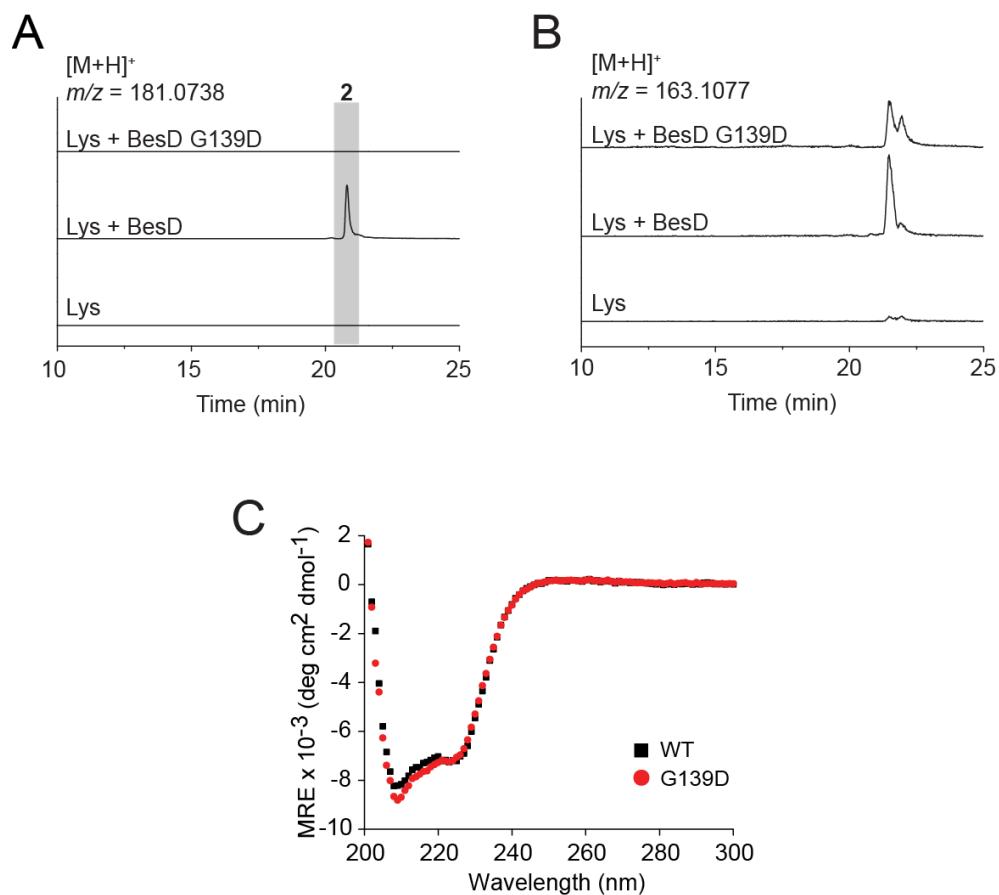


Figure 2.13. (A) LC/QTOF traces monitoring production of 4-Cl-lysine (**2**, $m/z = 181.0738$) in reactions with no enzyme, BesD, or the BesD G139D mutant. Chromatogram shown is representative of 3 experimental replicates. (B) LC/QTOF traces monitoring production of OH-lysine ($m/z = 163.0738$). A low level of 4-OH-lysine is expected from direct hydroxylation of lysine by wild-type BesD, as has been commonly observed in other members of the α KG/Fe-dependent halogenase family. (C) Circular dichroism (CD) spectra of wild-type BesD compared to the G139D mutant suggests that BesD G139D retains a similar fold or secondary structure content (MRE, molar residue ellipticity). Spectra shown is representative of 3 independent replicates.

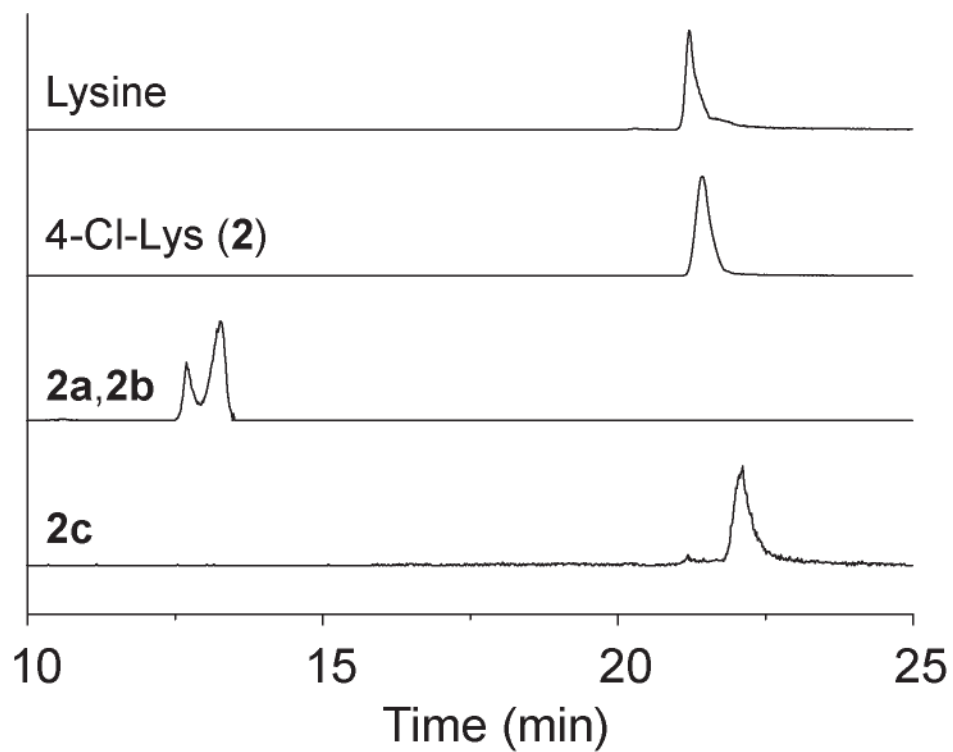


Figure 2.14. Extracted ion chromatograms of the decomposition products observed upon halogenation of lysine by BesD by LC/QTOF. (Lysine, $m/z = 147.1128$; 4-Cl-lysine (**2**), $m/z = 181.0738$; ϵ -lactam (**2a**)/ γ -lactone (**2b**), $m/z = 145.0972$; 4-OH-lysine (**2c**), $m/z = 163.1077$). Chromatogram shown is representative of at least 3 experimental replicates.

at higher pH (pH 8) than neutral (pH 7) or slightly acidic (pH 6) (*Figure 2.15A*). Two degradation products had identical mass ($[M+H]^+$ $m/z = 145.0975$), corresponding to a loss of chloride and a hydrogen, while the third degradation product had a mass corresponding to OH-lysine ($[M+H]^+$ $m/z = 163.1077$).

Degradation through an intramolecular carboxylate-mediated cyclization of amino acids has previously been reported [34]. Thus, we propose that degradation of Cl-lysine occurs through an intramolecular nucleophilic attack on the C_γ , leading to formation of a 5-membered lactone (*Figure 2.15B*). The lactone can then either hydrolyze to form 4-OH-lysine or undergo an ϵ -amine mediated ring-opening to form a stable 7-membered lactam. The lactam can also undergo hydrolysis to form 4-OH-lysine. Based on this proposed degradation mechanism, we performed methyl-esterification to stabilize 4-Cl-lysine for NMR characterization.

To improve the signal of NMR, we used $^{13}C/^{15}N$ -labeled lysine as substrate for chlorination. A modified constant time HSQC ($^{13}C/^1H$) was used to determine the identity of the C_α and C_ϵ , which contain only one adjacent ^{13}C . HCCH-COSY was then used to determine connectivity of the identified carbons. Based on NMR analysis of the methyl-esterified product, the chlorination was determined to occur at the γ -position (*Figure 2.16*). Subsequent experiments were performed on the $^{13}C/^{15}N$ lactam degradation product to confirm the degradation pathway. In addition to a constant time HSQC and HCCH-COSY, we also performed $^{15}N/^1H$ HSQC, $^{13}C/^1H$ HNCO, and $^{13}C/^1H$ HNCA (*Figure 2.17*).

BesC is an Fe-dependent oxidase. In addition to 4-Cl-lysine, we identified a second chlorinated metabolite from comparative metabolomics studies, which had a mass consistent with a loss of $-CH_5N$ from 4-Cl-lysine. This compound accumulated in the $\Delta besB$ strain but was not observed in the $\Delta besC$ strain (*Figure 2.9*), suggesting that BesC, a putative non-heme dinuclear Fe enzyme (*Figure 2.18*), is the most likely candidate for its production. We thus propose that BesC catalyses an unusual oxidative C–C bond-cleaving reaction to form a five-carbon terminal-alkene amino acid, 4-Cl-allylglycine (**3**) with eventual release of the methylamine fragment as ammonia and formaldehyde (*Figure 2.19*). To further characterize BesC, we tested the ability of the purified enzyme to produce 4-Cl-allylglycine (**3**) *in vitro* from purified 4-Cl-lysine or a coupled reaction with BesD and lysine (*Figure 2.20*). The BesC product was confirmed to be **3** by comparison to a commercial standard and exhibited the characteristic Cl isotope pattern (*Figure 2.21*).

While investigating BesC, we explored off-pathway reactions that could be occurring *in vivo*. We tested whether BesC could oxidize lysine to form allylglycine (*Figure 2.22*). When we tested the enzyme *in vitro*, we discovered that BesC does make allylglycine *in vitro*. However, competition experiments revealed that Cl-lysine is the preferred substrate for BesC; the presence of Cl-lysine in the reactions decreases allylglycine production despite an excess of lysine substrate. It remains to be determined whether this is due to preferential binding of Cl-lysine over lysine, or to the difference in bond dissociation energy (BDE) between the target hydrogen atom in the two substrates.

To further characterize BesC, we sought to identify the co-products that are released upon C–C bond cleavage. We proposed that the terminal N_ϵ and C_ϵ could be lost as ammonia and formaldehyde, respectively. To investigate the fate of the N_ϵ , we developed an assay to capture the resulting ammonia co-product. Glutamate dehydrogenase (GDH) is an enzyme that catalyzes the reductive amination of α KG to glutamate using ammonia as the nitrogen source [24]. By using an isotopically-labelled lysine substrate ($^{15}N_2, ^{13}C_6$ -lysine), we could directly detect ^{15}N

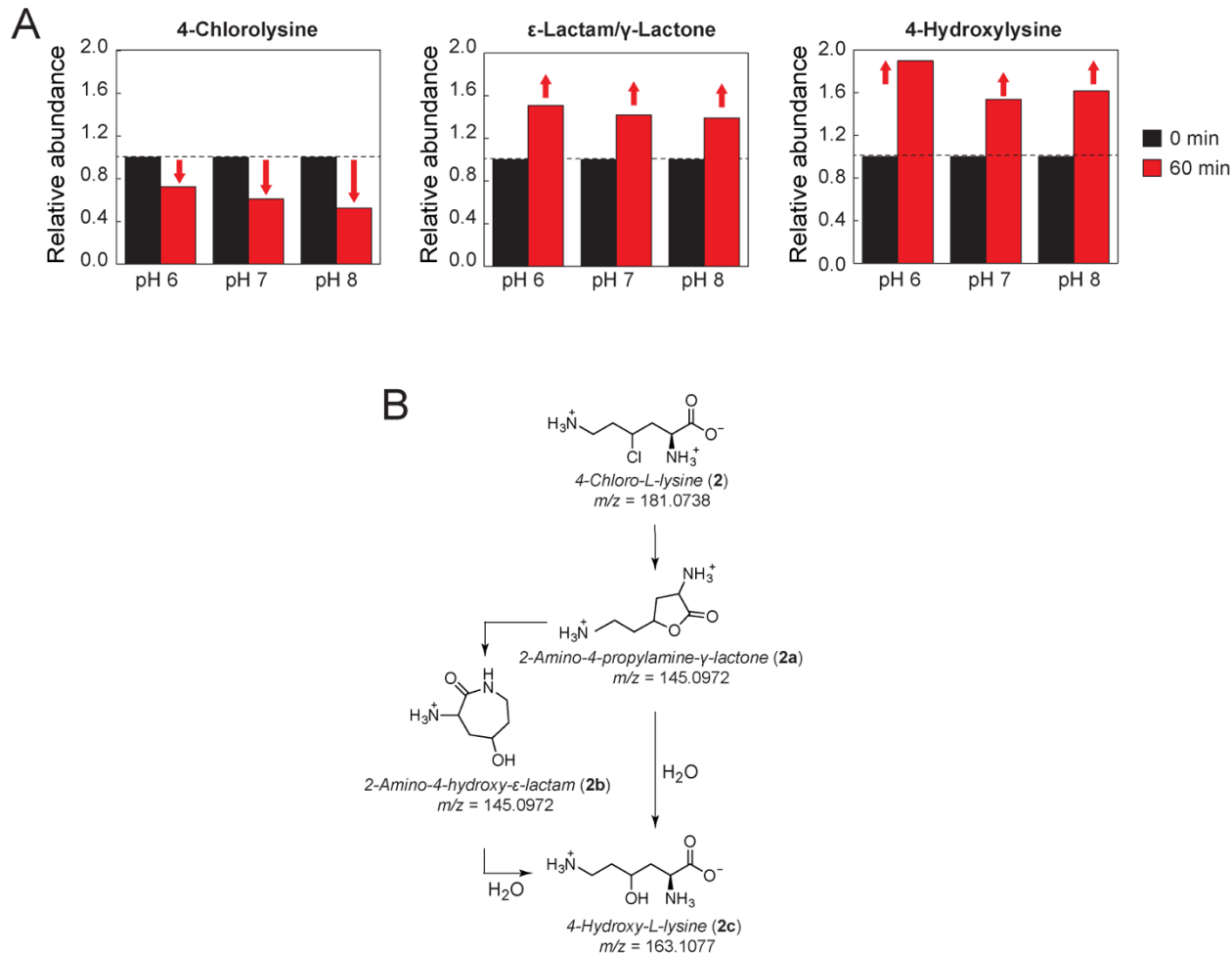


Figure 2.15. (A) Integrated EICs for each of the observed products at pH 6, 7, and 8 following removal of BesD by filtration (Pall Nanosep filter, 10 kDa MWCO). Data are plotted relative to lysine, which is taken to be constant following enzyme removal. After 1 h, Cl-lysine decreases, while the lactam, lactone, and OH lysine products increase in abundance. (B) 4-Cl-lysine (**2**) is proposed to undergo an intramolecular nucleophilic attack forming 2-amino-4-propylamine- γ -lactone (**2a**) with an estimated $t_{1/2} = 1$ h. The γ -lactone is also unstable can form either the 2-amino-4-hydroxy- ϵ -lactam (**2b**) or 4-OH-lysine (**2c**). 4-OH-lysine could also arise from direct hydroxylation of lysine by BesD, as a low level of hydroxylation is commonly observed in other members of the Fe/ α KG-dependent halogenase family.

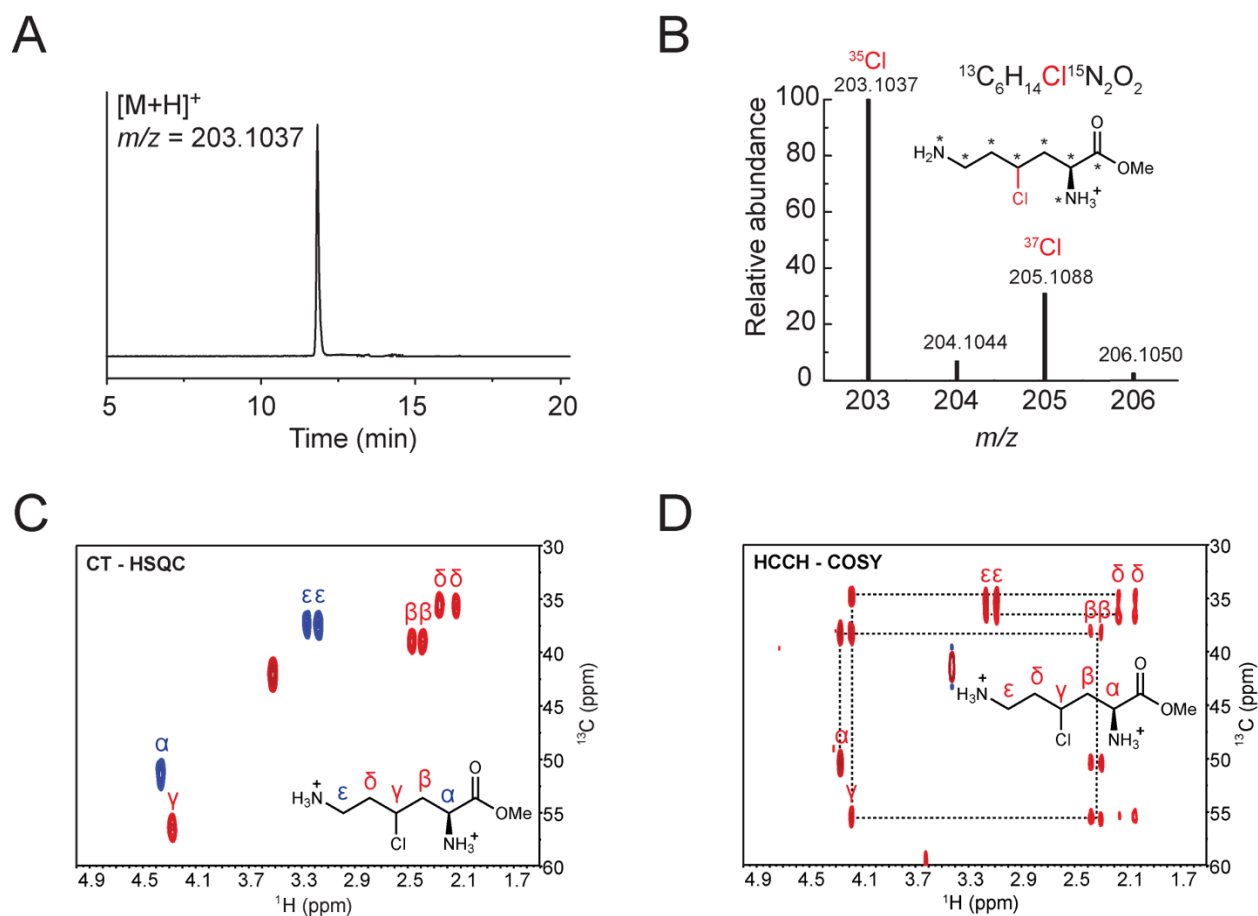


Figure 2.16. (A) BesD was incubated with fully ^{15}N - and ^{13}C -labeled lysine, Fe, and αKG for 20 min before quenching with methanolic HCl (3M) to yield the $[\text{}^{15}\text{N}_2, \text{}^{13}\text{C}_6]\text{-4-Cl-lysine}$ methyl ester product. Chloride was provided as NaCl in the enzyme storage buffer. We found that derivatization was necessary in order to avoid intramolecular lactam formation. The derivatized BesD product was then isolated by HPLC and characterized. Extracted ion chromatograms of the $[\text{}^{15}\text{N}_2, \text{}^{13}\text{C}_6]\text{-4-Cl-lysine}$ ($m/z = 203.1037$) produced by BesD. Chromatogram shown is representative of 3 independent experimental replicates. (B) Mass spectra show the characteristic Cl isotope pattern for $[\text{}^{15}\text{N}_2, \text{}^{13}\text{C}_6]\text{-4-Cl-lysine}$ methyl ester produced by BesD. (C) 2D ^1H - ^{13}C CT-HSQC was used to confirm assignment of C_α and C_ϵ of the $[\text{}^{15}\text{N}_2, \text{}^{13}\text{C}_6]\text{-4-Cl-lysine}$ methyl ester produced by BesD. In this experiment carbons with two neighbors (red) have opposite phase of carbons with one or three neighbors (blue). (D) 2D ^1H - ^{13}C HCCH COSY was used to assign connectivity of carbons of the $[\text{}^{15}\text{N}_2, \text{}^{13}\text{C}_6]\text{-4-Cl-lysine}$ methyl ester produced by BesD.

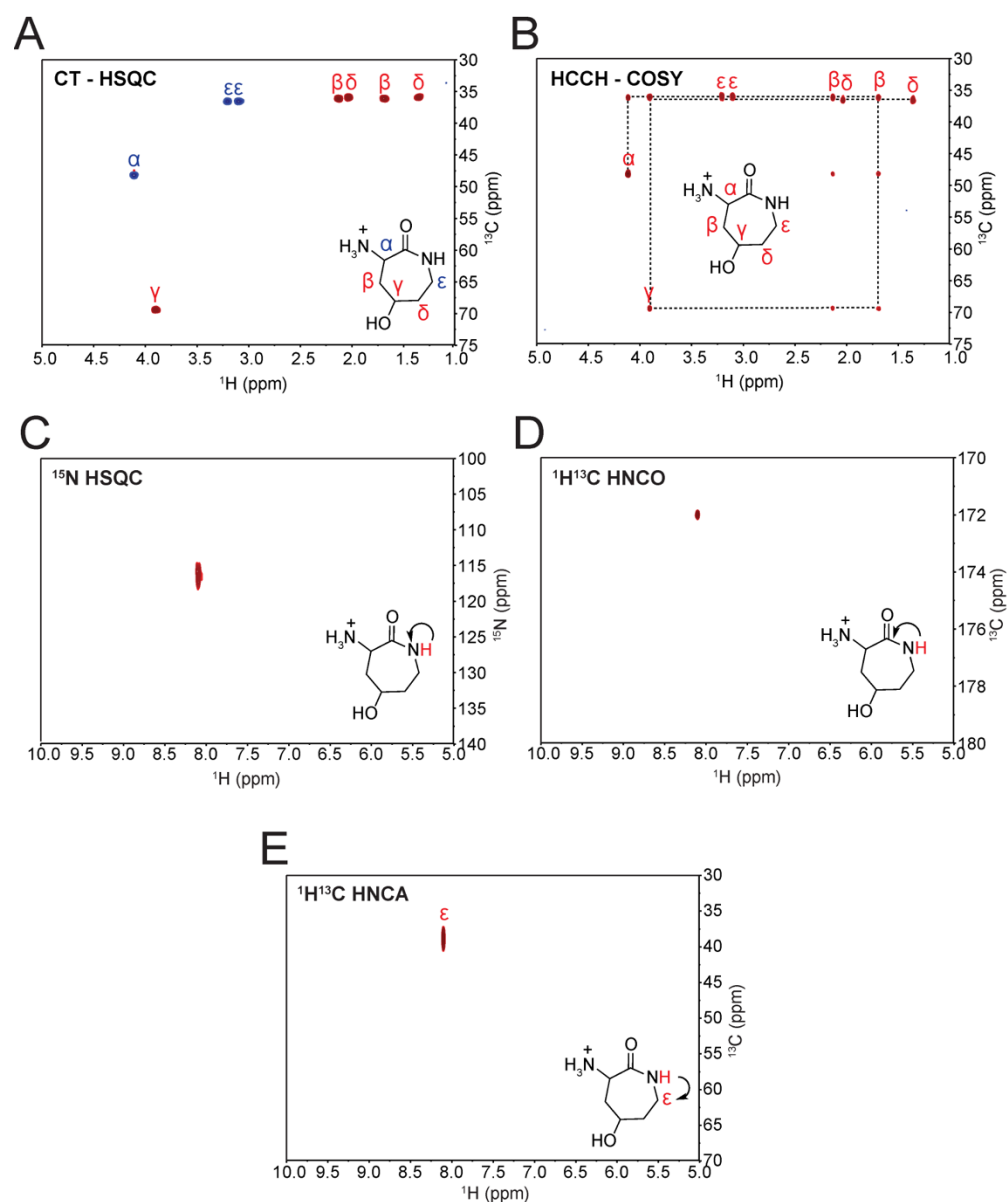


Figure 2.17. (A) For obtaining the lactam structure, BesD was incubated with fully ^{15}N and ^{13}C -labeled L-lysine, Fe, and αKG for 16 h before quenching with 1% (v/v) formic acid in methanol to yield the $[\text{N}_2, \text{C}_6]$ -lactam as the major product. Chloride was provided as NaCl in the enzyme storage buffer. NMR analysis of the ϵ -lactam was carried out after HPLC purification. 2D ^1H - ^{13}C CT-HSQC with a 13.6 ms ^{13}C evolution period was used to confirm assignment of C_α and C_ϵ . In this experiment carbons with two neighbors (red) have opposite phase of carbons with one or three neighbors (blue). (B) ^1H - ^{13}C HCCH COSY was used to assign backbone connectivity of carbons for the lactam. (C) The ^1H - ^{15}N HSQC spectrum confirms the presence of the amide in the lactam structure, with reported chemical shift typical of amides (δ 6 - 9 ppm). (D) The 2D ^1H - ^{13}C HNCO spectrum confirms that the epsilon N is adjacent to a carbonyl species. (E) The 2D ^1H - ^{13}C HNCA spectrum confirms that the epsilon N is coupled to C_ϵ , forming a cyclic structure.

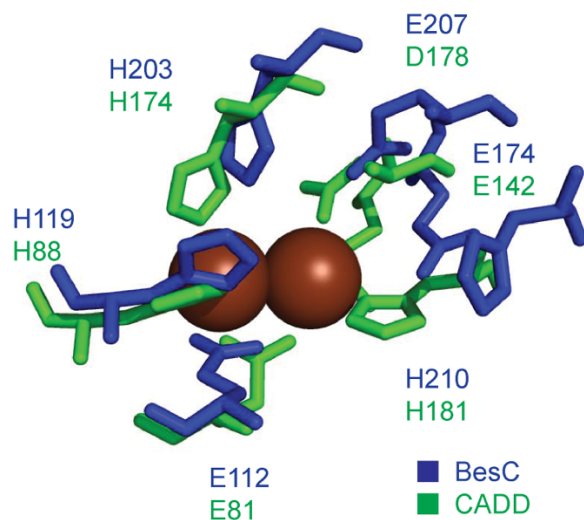


Figure 2.18. BesC shows homology to non-heme diiron proteins such as the *Chlamydia* protein Associated with Death Domains (CADD, 34% sequence similarity) [35] and stearoyl-ACP desaturase (26% sequence similarity) [36]. Overlay of the predicted homology model of BesC (blue) with the CADD crystal structure (green) reveals six conserved putative Fe binding residues in the active site. Putative iron ligands shown in brown. Homology model was generated using Phyre 2 [37].

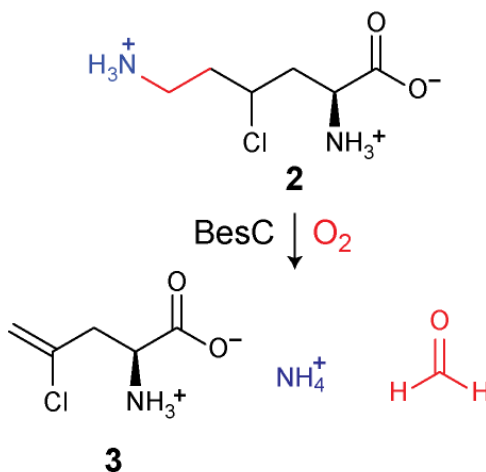


Figure 2.19. Reaction catalyzed by BesC, forming ammonia and formaldehyde as co-products from oxidative cleavage between C δ and C ϵ .

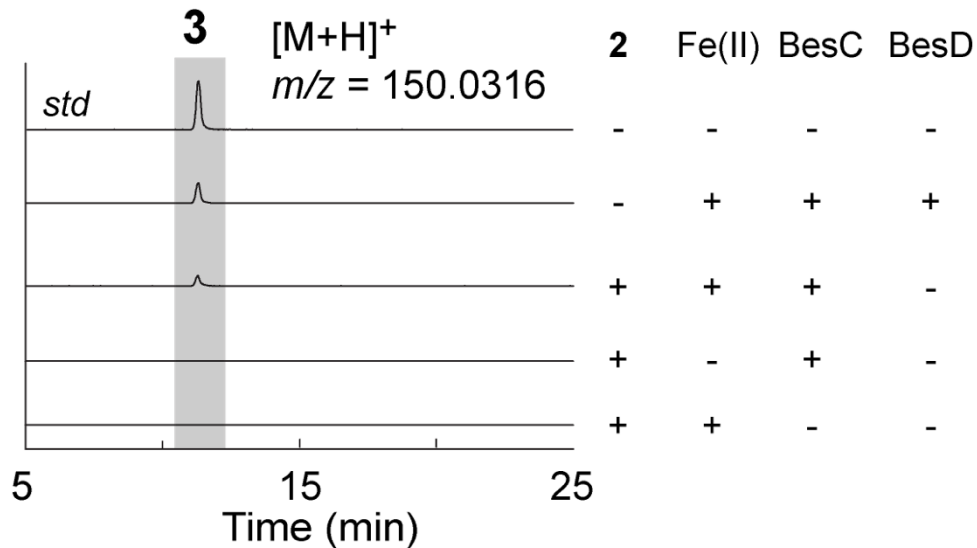


Figure 2.20. BesC catalyzes formation of 4-Cl-allylglycine from 4-Cl-lysine and requires Fe(II) as a cofactor (4-Cl-allylglycine, **3**; $m/z = 150.0316$). Chromatograms shown are representative of at least 3 experimental replicates.

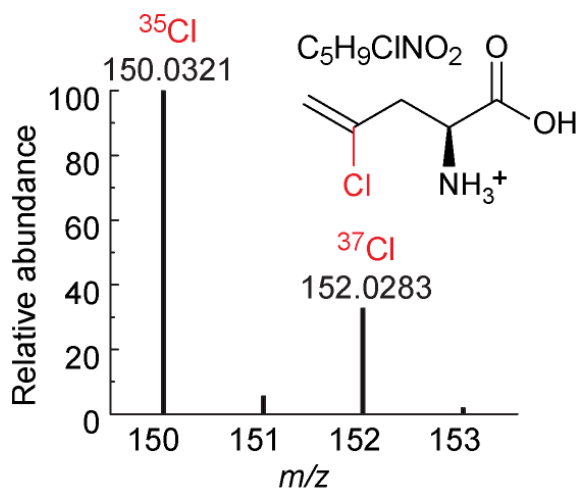


Figure 2.21. Mass spectra of product formed when BesC is incubated with 4-Cl-lysine. Spectra shown is representative of at least 3 experimental replicates.

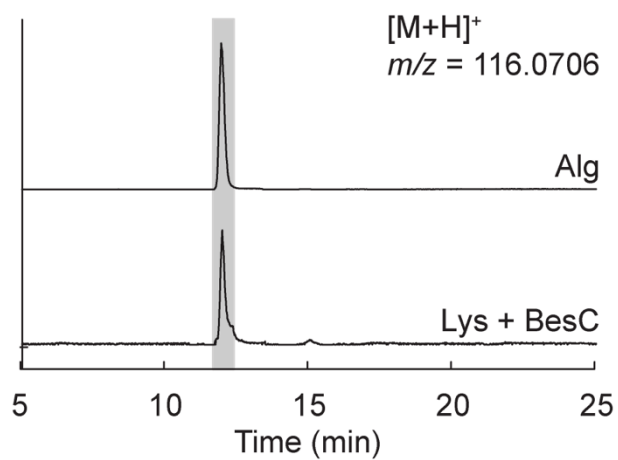
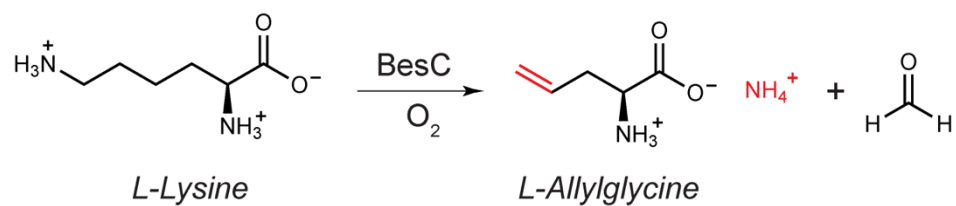


Figure 2.22. BesC is also competent to react with lysine directly to produce allylglycine *in vitro*. LC/QTOF traces show allylglycine produced by BesC compared to a synthetic standard (Alg, $m/z = 116.0706$). Chromatogram shown is representative of 3 experimental replicates.

incorporation into glutamate as N_α by LC/MS. We reacted lysine with BesD, BesC, and GDH. LC/MS analysis revealed a clear mass shift of glutamate from $m/z = 148.0603$ to 149.0578 when ^{15}N lysine was used (Figure 2.23). This experiment confirms that ^{15}N is incorporated into glutamate and that ammonia is a co-product of the BesC reaction.

We used a similar strategy to determine the fate of C_ϵ ; In these experiments, the putative formaldehyde co-product reacts with two molecules of Fluoral-P to form 3,5-diacetyl-1,4-dihydro-2,6-lutidine (DDL), releasing ammonia in the process (Figure 2.24A). In a coupled reaction with BesD, BesC, and $^{15}\text{N}_2, ^{13}\text{C}_6$ -lysine as a substrate, we observed a mass shift of DDL from $m/z = 194.1176$ to 195.1209 (Figure 2.24). The experiments confirm that C_ϵ is released as formaldehyde in the BesC reaction. Both release of ammonia and formaldehyde were found to be stoichiometric with terminal-alkene amino acid formation (Figure 2.25). Taken together, the data show that BesC catalyzes alkene formation with release of formaldehyde and ammonia as the final co-products of the reaction.

Based on homology models, BesC is a putative binuclear iron enzyme. We therefore attempted to determine the stoichiometry of iron using the ferrozine assay [38]. BesC was incubated anaerobically with 5 eq. of Fe. Bovine serum albumin (BSA) was used as a negative control. After 30 min of incubation, samples were desalted to remove unbound Fe. Protein was precipitated using hydrochloric acid, and Fe content was quantified by the ferrozine assay (A_{562}). Two replicates were performed for BesC; one sample had a Fe:BesC stoichiometry of $\sim 3:1$, and the other had a stoichiometry of $0.8:1$. The BSA control had only ~ 0.25 eq. of Fe. Due to the inconsistent results, the stoichiometry of Fe loading in BesC remains unclear at this time.

We next performed X-ray protein crystallography of BesC in an effort to identify the Fe- and lysine-binding active site residues. The crystals diffracted to $1.7 \text{ \AA}$. However, BesC has less than 20% sequence identity to solved structures in the Protein Data Bank (PDB). Since this is too low for molecular replacement techniques, the structure of BesC was phased experimentally by crystallizing apo SeMet-labeled protein and collecting data at the Se edge (Table 2.2). BesC crystallized with four molecules per asymmetric unit in space group P21 with 43% twin fraction. It displays the seven-helix bundle that is characteristic of diiron enzymes from the heme-oxygenase-like family (Figure 2.26). However, one key helix near the active site had poor electron density and could not be modelled into the structure. Since this missing helix contains three of the six Fe-binding residues, the lack of density could be due to the absence of Fe in the crystals. Recent structural studies of SznF from *S. achromogenes* and UndA from *Pseudomonas aeruginosa*, two other enzymes the heme-oxygenase family, reported difficulty binding the second equivalent of Fe [39,40], resulting in weak electron density in the homologous helical region. Future efforts on this front include solving the Fe and lysine-bound structures of BesC.

We next turned our attention to formation of the terminal-alkyne functional group of Pra. On the basis of the accumulation of **3** upon deletion of the *besB* gene (Figure 2.9), we assigned the PLP-dependent BesB enzyme as the acetylenase that catalyses the γ -elimination of chloride from **3** to achieve substrate desaturation. *In vitro* assays with purified BesB show that Pra can be produced either directly from a 4-Cl-allylglycine standard or by a coupled assay with BesC and BesD, and that the presence of a γ -halogen as a leaving group is required (Figure 2.27). As an initial probe of the BesB mechanism, we used high-resolution mass spectroscopy to measure deuterium exchange along the backbone of the substrate. Using 4-Cl-allylglycine (**3**) as a substrate in D_2O resulted in formation of $[\text{M}+2\text{D}]^+$ Pra, suggesting that at least two deprotonation–

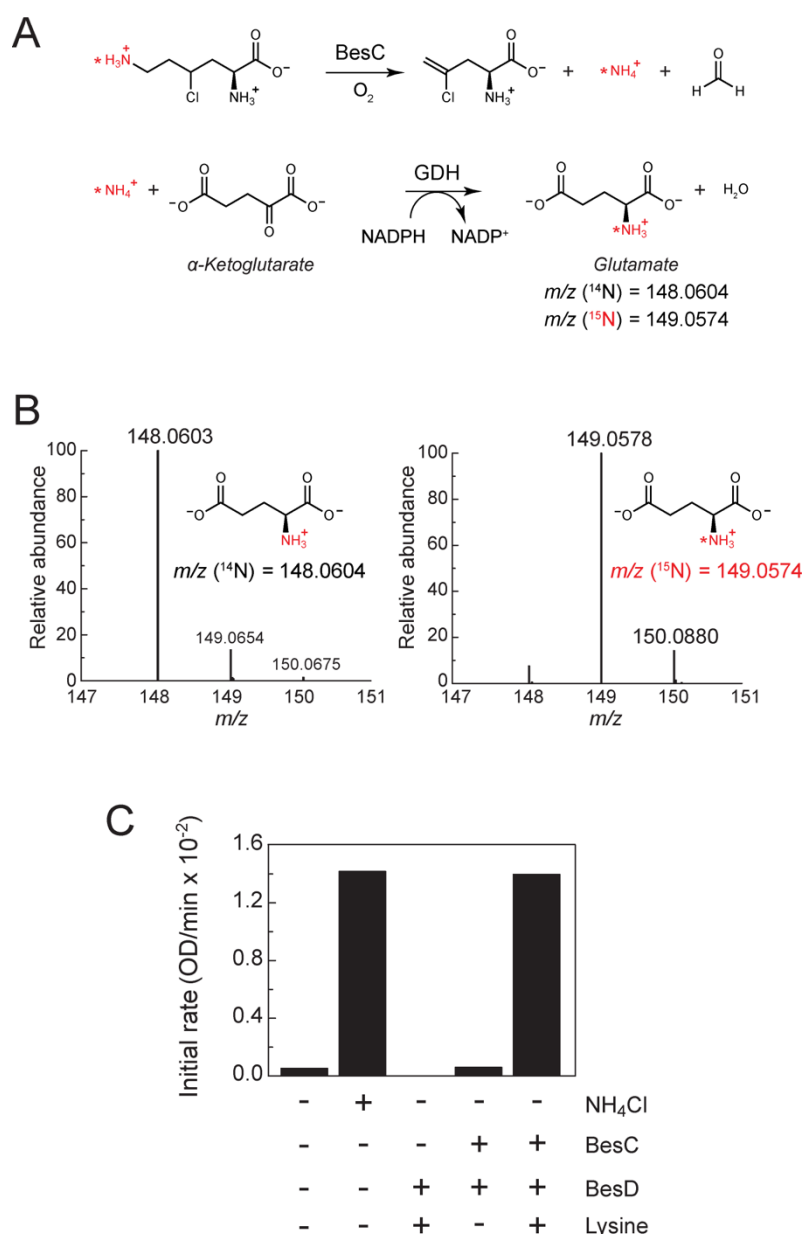


Figure 2.23. (A) The BesC reaction is run as a coupled reaction with BesD starting from lysine, α KG, and chloride. To detect the ammonia co-product, we employed glutamate dehydrogenase (GDH), which produces glutamate from ammonia and α KG with concurrent oxidation of NADPH. (B) The release of ammonia from 4-Cl-lysine is monitored through incorporation of the nitrogen from the ^{15}N -labeled lysine substrate into glutamate using LC/QTOF. Positive ionization mass spectrum showing formation of ^{14}N -L-glutamate or ^{15}N -glutamate when unlabeled lysine (left) or $^{15}\text{N}_2, ^{13}\text{C}_6$ -lysine (right) is used as the substrate, respectively. Spectra shown are representative of 3 experimental replicates. (C) Consumption of NADPH was observed spectrophotometrically. Reactions contained BesC, BesD, lysine, α KG, sodium ascorbate, Fe(II), and chloride as indicated. After 1 h, NADPH and glutamate dehydrogenase were added, and NADPH consumption was monitored by A_{340} .

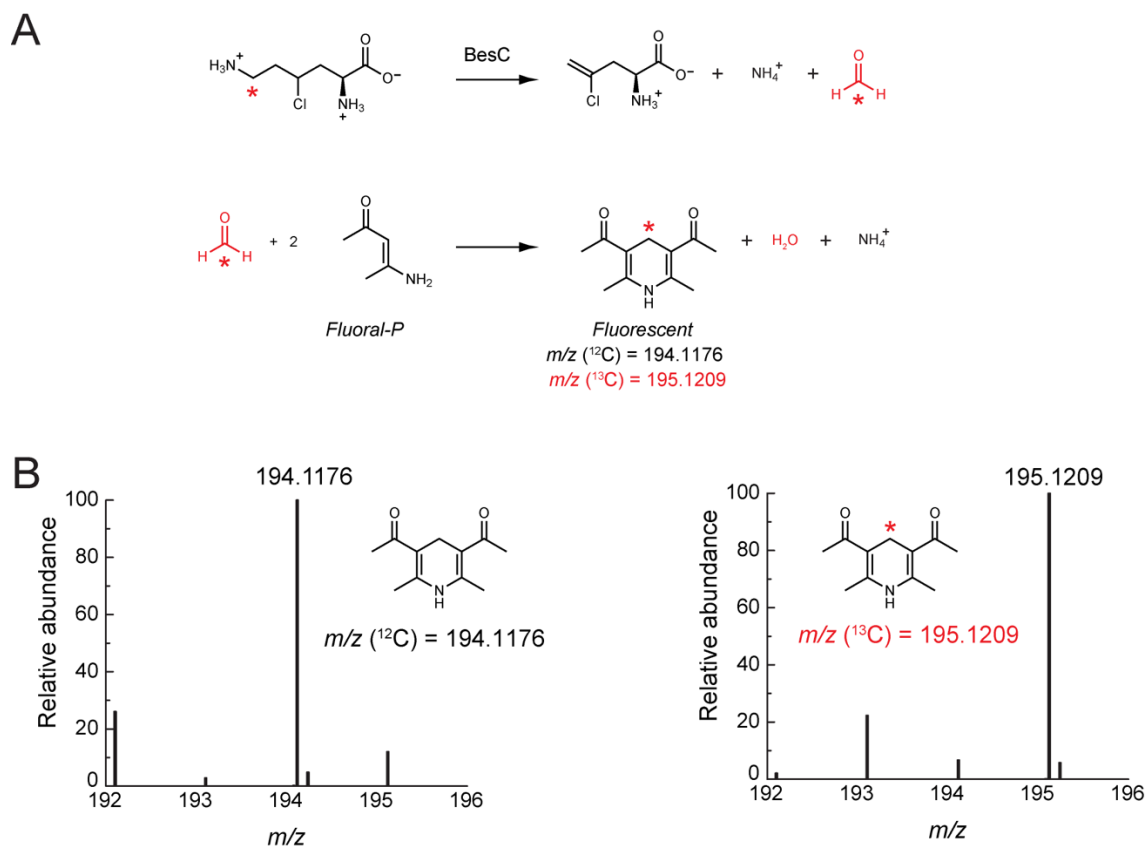


Figure 2.24. Detection of formaldehyde as a co-product of the BesC reaction. The BesC reaction is run as a coupled reaction with BesD starting from lysine, α KG, and chloride. (A) To derivatize the putative formaldehyde co-product, we performed the enzymatic reaction in the presence of Fluoral-P. The release of formaldehyde from 4-Cl-lysine was monitored through incorporation of the carbon from ^{13}C -labeled lysine substrate into 3,5-diacetyl-1,4-dihydro-2,6-lutidine (DDL). (B) Positive ionization mass spectrum showing formation of ^{12}C -DDL or ^{13}C -DDL when unlabeled lysine (left) or [$^{15}\text{N}_2$, $^{13}\text{C}_6$]-lysine (right) is used as the substrate, respectively. Spectra shown are representative of 3 experimental replicates.

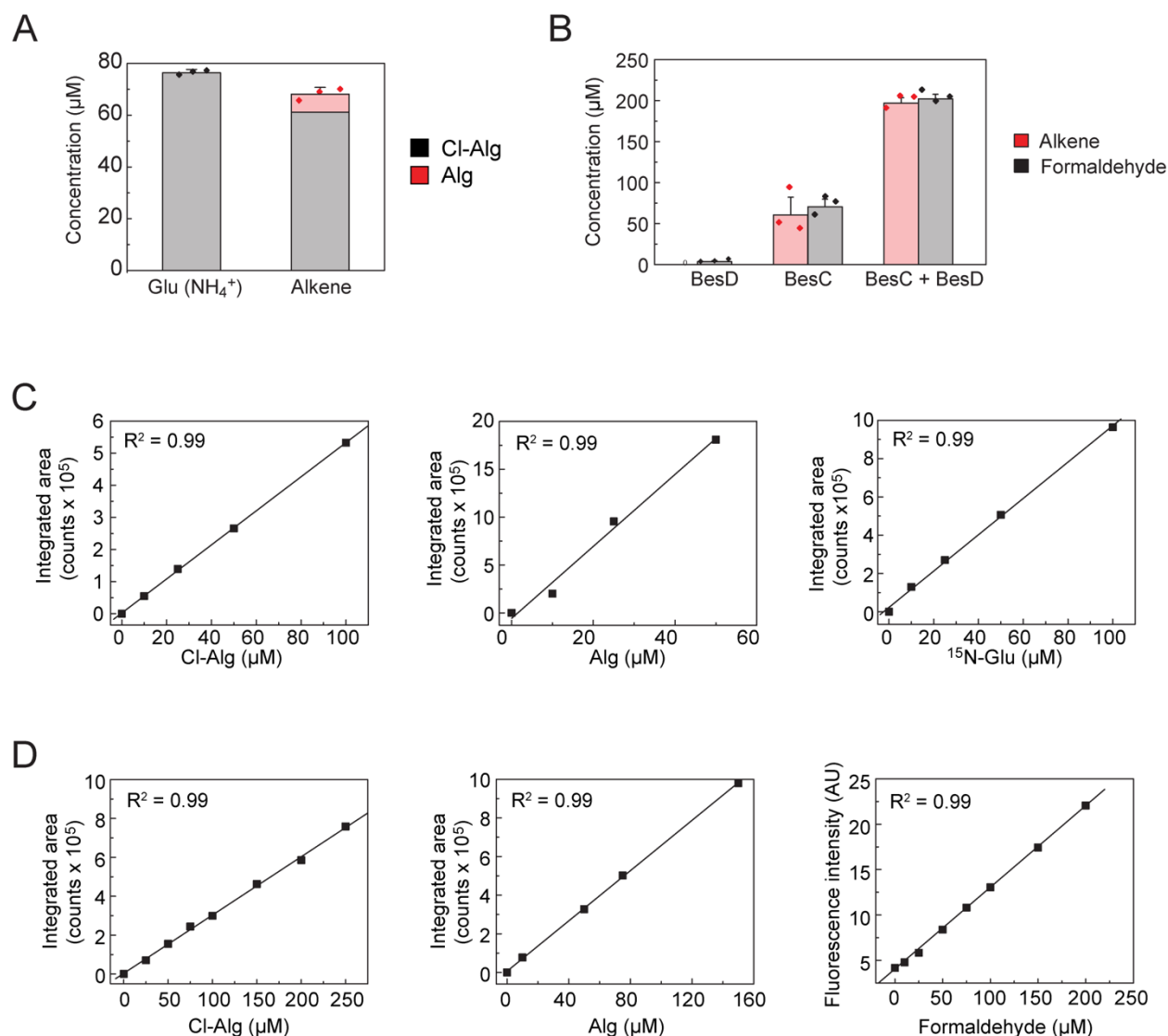


Figure 2.25. (A) Quantification by LC/QTOF showing stoichiometric ammonia release measured as glutamate (Glu) compared total alkene product (4-Cl-allylglycine, CI-Alg; allylglycine, Alg) in a coupled assay of BesD and BesC with lysine using glutamate dehydrogenase. Data are mean $\pm$ s.d. ($n = 3$ independent experiments). (B) Quantification of formaldehyde produced from Fluoral-P reaction with 5 mM lysine and purified BesD, BesC, or both BesC and BesD. The resulting product was quantified by fluorescence and compared to alkene product formation measured by LC/QTOF and found to be stoichiometric. Data are mean $\pm$ s.d. ($n = 3$ independent experimental replicates). (C) Standard curves for ammonia quantification were constructed by integrating extracted ion counts for the relevant species by LC/QTOF. R^2 value shown is determined by ordinary least squares. (D) Standard curves for formaldehyde quantification were generated by integration of the extracted ion chromatogram from LC/TOF (CI-Alg, Alg) or by fluorescence quantification (formaldehyde). R^2 value shown was determined by ordinary least squares.

Table 2.2. Data collection and refinement parameters for BesC from *S. cattleya*.

	BesC SeMet $\lambda = 0.953680 \text{ \AA}$	BesC SeMet $\lambda = 0.979620 \text{ \AA}$
Data collection		
Space group	P 1 21 1	P 1 21 1
Cell dimensions		
$a, b, c \text{ (\AA)}$	60.917, 68.977, 126.748	60.917, 68.977, 126.748
$\alpha, \beta, \gamma \text{ (}^\circ\text{)}$	90, 89.993, 90	90, 89.993, 90
Resolution ($\text{\AA}$)	63.37-1.7 (1.761-1.7) *	63.37-1.7 (1.761-1.7) *
R_{sym} or R_{merge}	0.0798 (0.5588)	0.08463 (0.5791)
$I / \sigma I$	14.18 (2.94)	14.58 (2.69)
Completeness (%)	96.36 (85.59)	96.36 (85.59)
Redundancy	8.2 (7.3)	8.0 (6.3)
$CC_{1/2}$	0.999 (0.929)	0.998 (0.887)
Refinement		
Resolution ($\text{\AA}$)	63.37-1.7	
No. reflections	941530 (81185)	
$R_{\text{work}} / R_{\text{free}}$	0.2062/0.2533	
No. atoms	8082	
Protein	7036	
Ligand/ion	22	
Water	1024	
B -factors		
Protein	21.11	
Ligand/ion	27.62	
Water	27.05	
R.m.s. deviations		
Bond lengths ($\text{\AA}$)	0.004	
Bond angles ($^\circ$)	0.89	

*Data were collected on one crystal at two different wavelengths. *Values in parentheses are for highest-resolution shell.

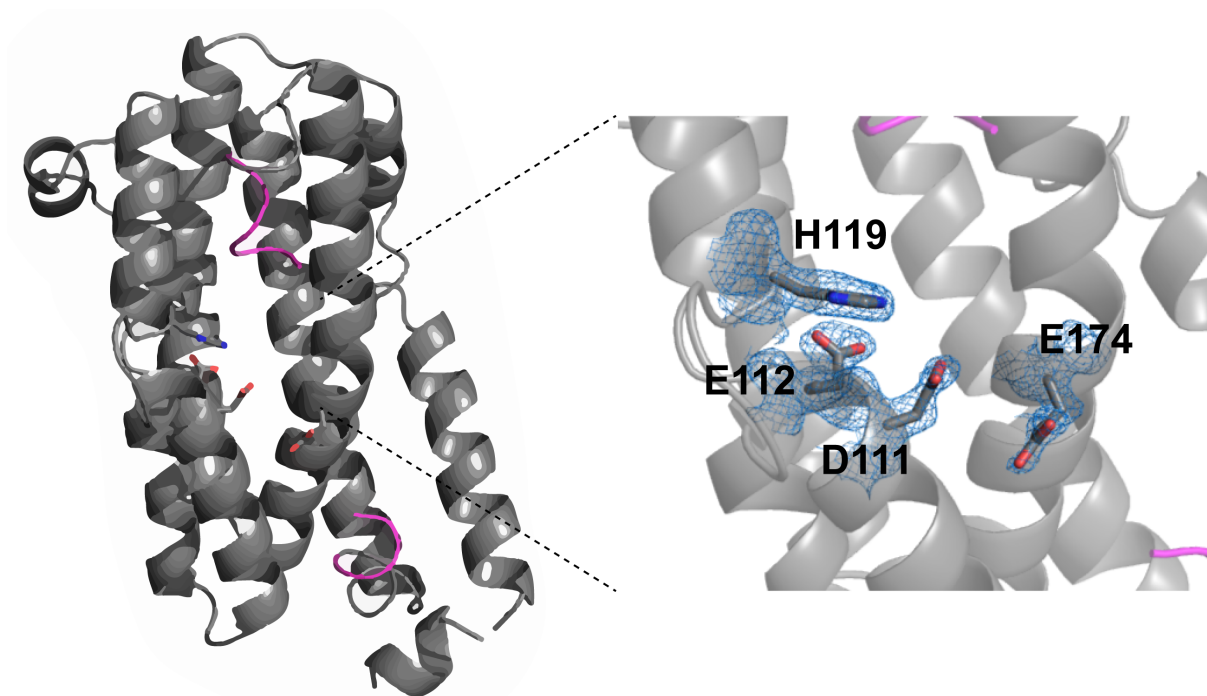


Figure 2.26. Crystal structure of apo BesC from *S. cattleya*. (Left) The overall fold of BesC shows the seven-helix bundle that is characteristic of diiron enzymes in the heme-oxygenase-like family. (Right) $2F_o - F_c$ map of key residues in the BesC active site (blue mesh, contoured at 1σ). The active site contains electron density for the putative Fe-binding residues Glu112, His119, and Glu174, in addition to Asp111. The helix shown in magenta, which contains the putative Fe-binding residues His203, Glu207, and His210, cannot be fully modeled due to lack of electron density.

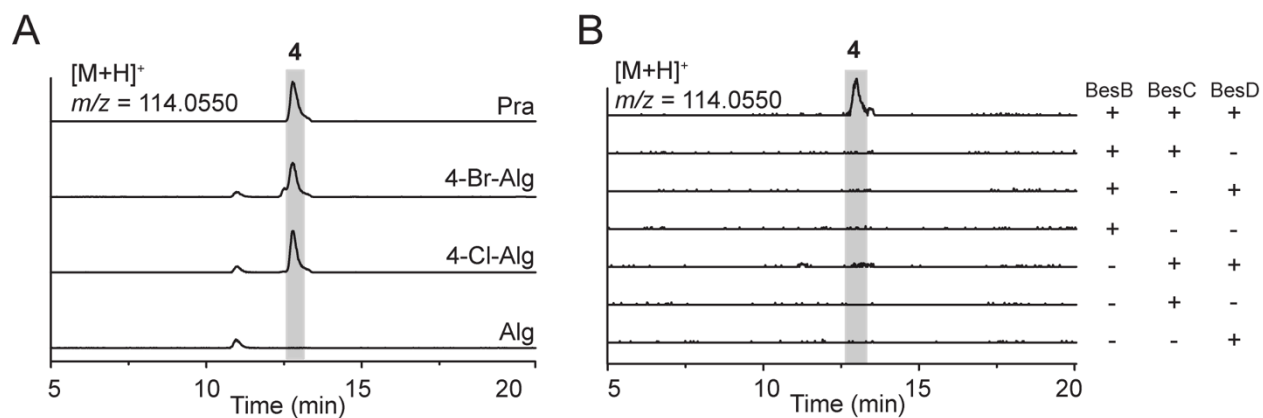


Figure 2.27. *In vitro* characterization of BesB. (A) LC/QTOF analysis of *in vitro* reactions containing BesB, PLP, and substrate. When the substrate is 4-Cl-Alg or 4-Br-Alg, propargylglycine (**4**, *m/z* = 114.0550) is observed. However, no propargylglycine is observed if allylglycine is used as a substrate. Chromatograms shown are representative of at 3 independent experimental replicates. (B) *In vitro* reconstitution of propargylglycine (**4**, *m/z* = 114.0550) formation from lysine with purified BesB, BesC, and BesD and Fe(II), α KG, chloride, PLP, and ascorbate monitored by LC/QTOF. Chromatograms shown are representative of at 3 independent experimental replicates.

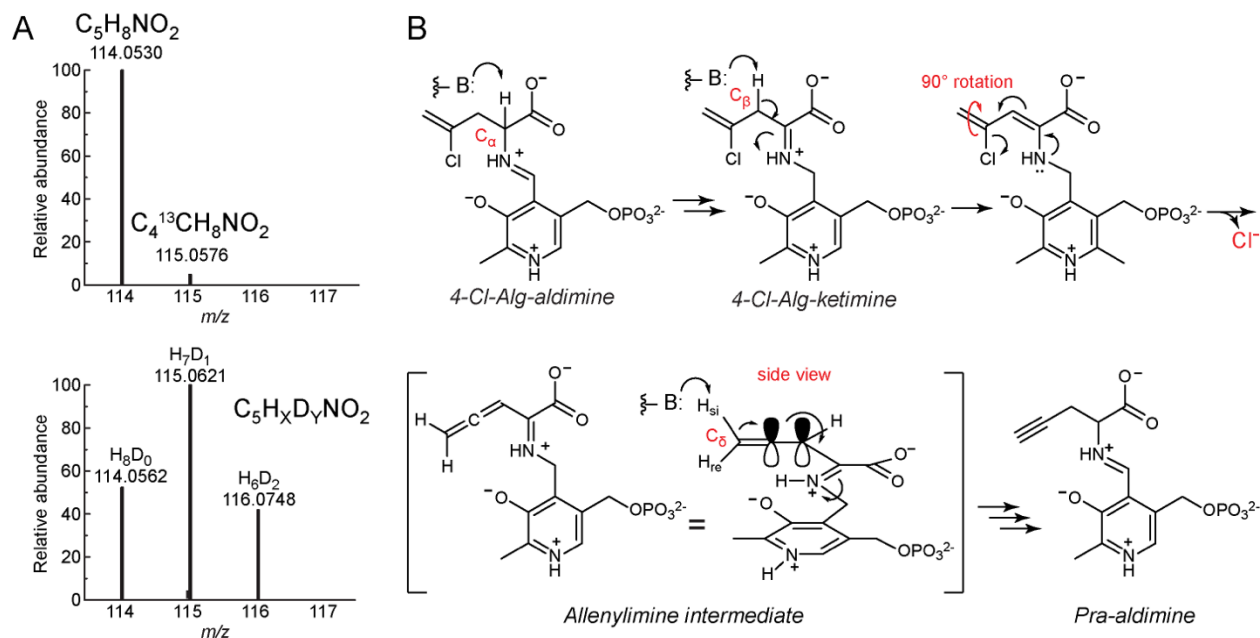


Figure 2.28. (A) Mass spectra showing deuterium incorporation at two positions of Pra (**4**) when BesB reaction is carried out in D₂O compared to H₂O. (top) Extracted ions for the BesB reaction run in H₂O. (Expected m/z : C₅H₈NO₂ 114.0550 [100%], C₄¹³CH₈NO₂ 115.0580 [5.9%]; Observed m/z : 114.0530 [100%], 115.0576 [4.8%]). (bottom) Extracted ions for the BesB reaction run in D₂O (Observed m/z : 114.0562 [52.3%], 115.0558 [3.2%, not labeled for clarity], 115.0621 [100%], 116.0748 [41.8%]). Spectra shown is representative of 3 experimental replicates. (B) Putative mechanism for BesB-catalyzed formation of Pra (**4**) from 4-Cl-allylglycine (**3**). Formation of a covalent external aldimine adduct between PLP and 4-Cl-allylglycine followed by C_α deprotonation and protonation of the quinonoid intermediate (not shown) is proposed to lead to formation of the initial ketimine intermediate. C_β deprotonation results in elimination of chloride upon 90° bond rotation to form an allenylimine intermediate. Subsequent C_δ deprotonation could then initiate isomerization of the allene to form the terminal alkyne. Following deprotonation of the PLP-bound alkyne to regenerate the quinonoid form (not shown), reprotonation of C_α reforms the external aldimine, from which Pra can be released as the free amino acid (not shown).

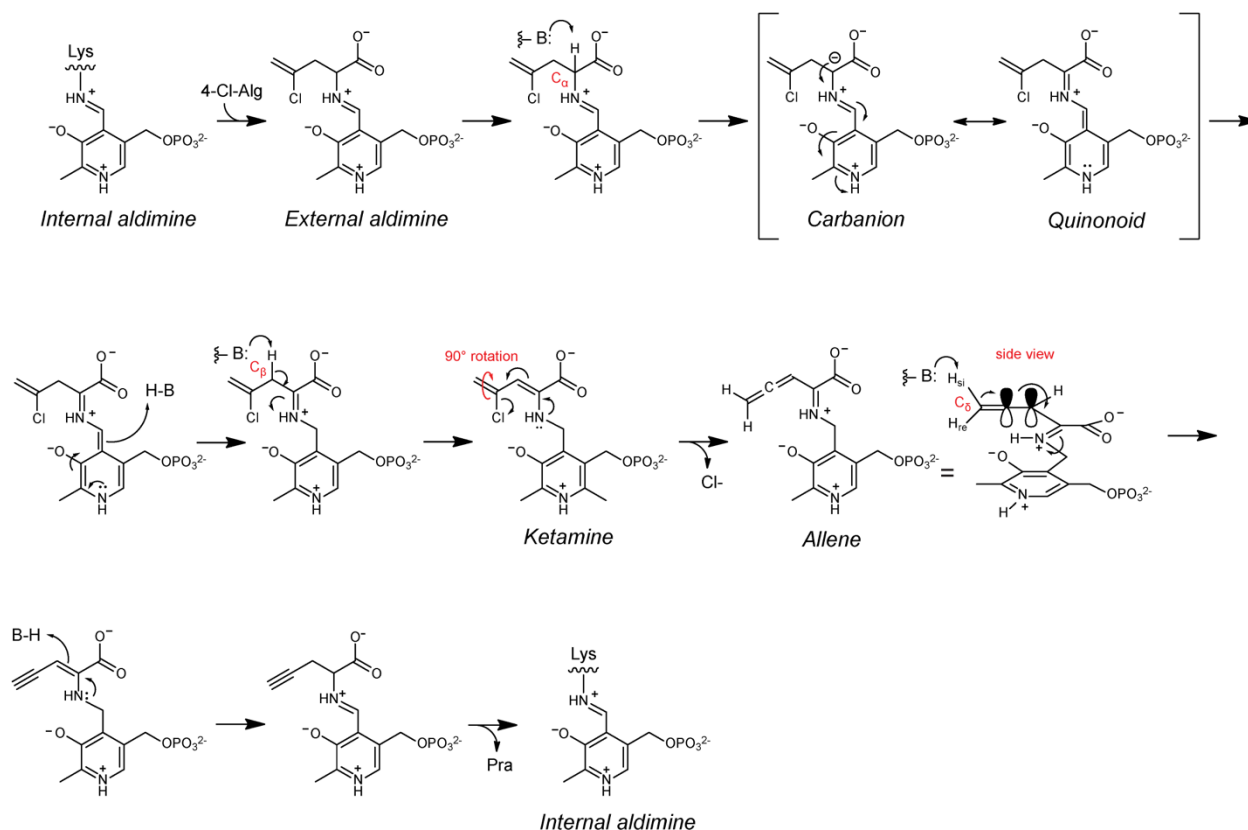


Figure 2.29. A mechanistic model for BesB is proposed in analogy to cystathionine γ -synthase and related PLP mechanisms that form allene intermediates. In the first step, the external 4-Cl-allylglycinealdimine is formed. Deprotonation at the C_α -position yields the 4-Cl-allylglycine-carbanion or quinonoid or intermediate. Deprotonation at the C_β -position produces the 4-Cl-allylglycine-ketamine intermediate. Elimination of chloride can then occur upon a 90° rotation to break planarity/aromatization to for a proposed terminal-allene intermediate. Based on work using Pra as a mechanistic inhibitor of PLP enzymes, isomerization between the terminal alkyne and allene can occur. At this time, we suggest that the reverse can occur to convert the PLP-bound allene to the terminal alkyne, yielding a PLP-bound Pra intermediate. The C_δ deprotonation to initiate this step is facilitated by overlap of the C_δ -H bond with the conjugated C_β - C_γ and PMP-imine π -bonds. Subsequent steps then allow for regeneration of the external aldimine with Pra, which is then released from the cofactor.

protonation events occur during the course of BesB catalysis (*Figure 2.28*). This model is consistent with a canonical cystathionine β -lyase/cystathionine γ -synthase mechanism in which C_{α} -deprotonation leads to eventual formation of a 4-Cl-allylglycine-ketimine intermediate (*Figure 2.29*) [41]. A second deprotonation on the amino acid at C_{β} would initiate the elimination of chloride to form an allene intermediate, which could subsequently isomerize to the terminal alkyne before eventual conversion back to aldimine and release from PLP. This proposed mechanism follows from previous studies of Pra as a mechanism-based inhibitor of PLP enzymes, based on its ability to isomerize to the corresponding allene [42,43].

BesA is a Pra- γ -glutamyl ligase. From the terminal-alkyne amino acid profile of the different streptomycete producers along with the $\Delta besE$ knockout strain (*Figure 2.30*), it seemed reasonable to believe that hydroxylation of Pra by BesE yields β es (**7**). However, no β es was observed when purified BesE was incubated with Pra in the presence of Fe, ascorbate, and α KG. We therefore looked for an alternative substrate for BesE, which we were able to identify as a glutamate-Pra dipeptide (**5**) through comparative metabolomics of *S. cattleya* $\Delta besE$ and wild-type strains (*Figure 2.31*). This compound was also present at a low concentration in the $\Delta besA$ knockout. Upon fragmentation, characteristic fragments consistent with amide bond cleavage were detected (*Figure 2.32*). Comparison of the products of purified BesA and a commercial γ -glutamyl transpeptidase further support assignment of this product as γ -Glu-Pra (*Figure 2.33*). Additional biochemical characterization indicates that BesA is selective for Pra over norvaline, allylglycine, and the standard proteinogenic amino acids (*Figure 2.34*, *Figure 2.35*). Notably, *besA* is conserved among these Pra biosynthetic gene clusters, suggesting that Glu-Pra formation has a physiological function, perhaps to reduce side reactions of Pra [42,43] or to facilitate transport.

BesE is a Glu-Pra hydroxylase. After discovering that Glu-Pra is an intermediate for β es biosynthesis, we then incubated Glu-Pra with BesE, Fe, ascorbate, α KG. Reactions were analysed by LC/MS, and production of γ -glutamyl- β es (Glu-Bes, **6**) was observed. These results show that Glu-Pra, and not Pra, is the substrate for BesE (*Figure 2.36*). The resulting hydroxylation gives rise to the question of how BesE can create a radical so close to a terminal alkyne without rearrangement. Further work on characterizing BesE should probe this question, as well as investigate the active site residues involved in terminal alkyne recognition and specificity.

Reconstitution of β es pathway *in vitro*. Elucidation of the final committed step allowed us to fully reconstitute the biosynthesis of Glu- β es (**6**) *in vitro* (*Figure 2.37*). In a single-pot reaction with lysine, purified protein (BesA, BesB, BesC, BesD, and BesE), and necessary co-factors (α KG, Fe, NaAsc, PLP, ATP, Glu) we were able to successfully make Glu- β es *in vitro*. While our target product was free β es, rather than the dipeptide, we believe that other enzymes found in ubiquitously in cells can carry out Glu- β es hydrolysis. One such candidate enzyme is γ -glutamyl transferase (Ggt). The mechanism by which *S. cattleya* and other Glu- β es/Glu-Pra producing streptomycetes hydrolyze these dipeptides is currently unknown.

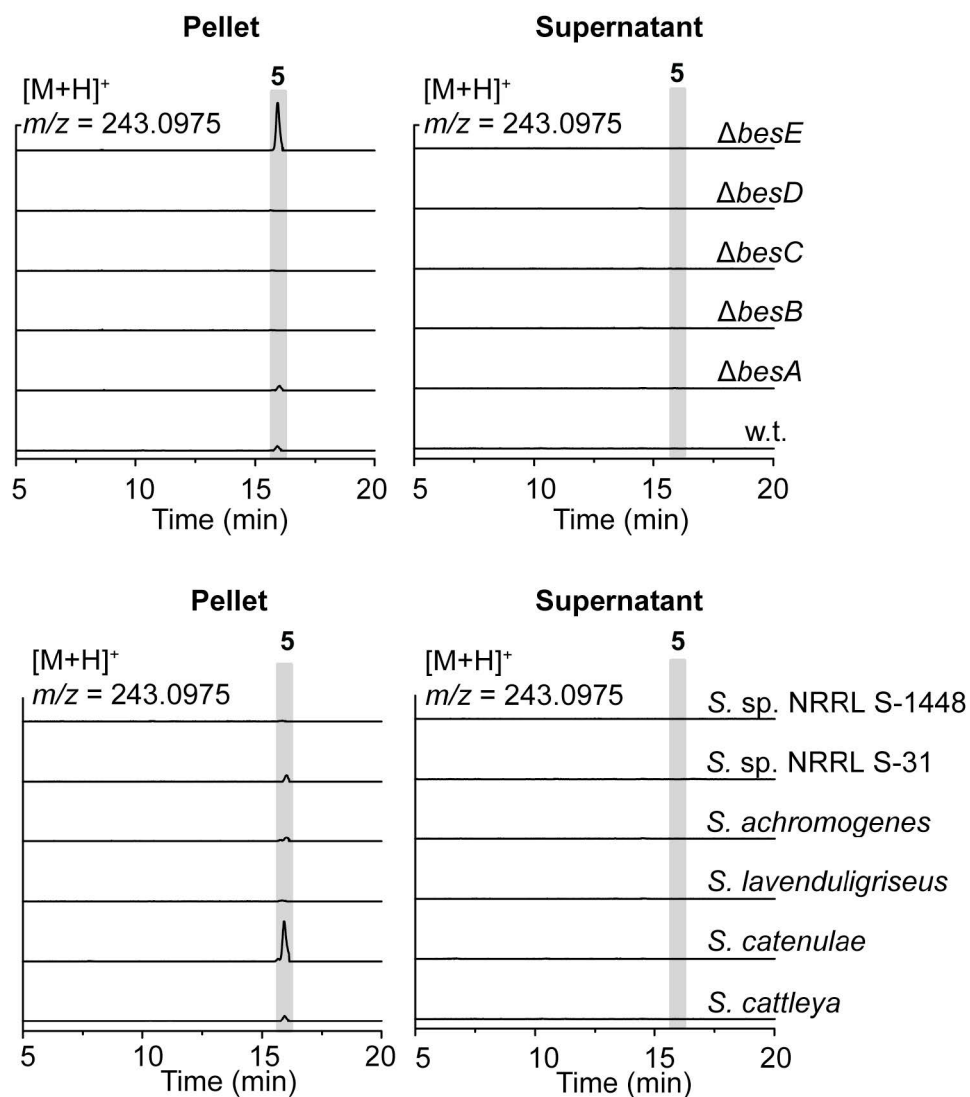


Figure 2.30. LC/QTOF traces for γ -Glu-Pra (**5**, $m/z = 243.0975$) from metabolites extracted from cell pellets and supernatants of *S. cattleya* w.t. and *bes* knockout strains as well as *S. sp.* NRRL S-1448, *S. sp.* NRRL S-31, *S. achromogenes*, *S. lavenduligriseus*, and *S. catenulae*. γ -Glu-Pra was detected in the cell pellets of *S. cattleya* w.t., the $\Delta besA$ and $\Delta besE$ knockout strains, *S. catenulae*, *S. achromogenes*, *S. sp.* NRRL S-31, and *S. sp.* NRRL S-1448. Lack of detection in the culture supernatants suggests that it is not a secreted metabolite. Chromatograms shown are representative of at least 3 independent experimental replicates

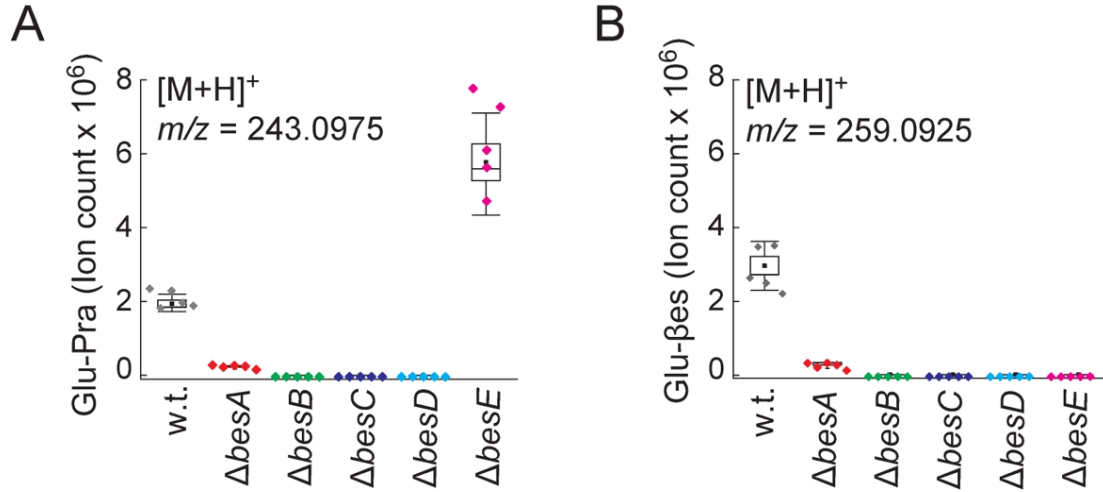


Figure 2.31. (A) Relative abundance of the γ -Glu-Pra dipeptide in *S. cattleya* wildtype and pathway knockout strains (n = 5 biologically independent experiments; dot = mean; line = median; box edges = 25th/75th percentile; error bars = 95th percentile). (B) Relative abundance of the γ -Glu- β es dipeptide in *S. cattleya* wildtype and pathway knockout strains (n = 5 biologically independent experiments; dot = mean; line = median; box edges = 25th/75th percentile; error bars = 95th percentile).

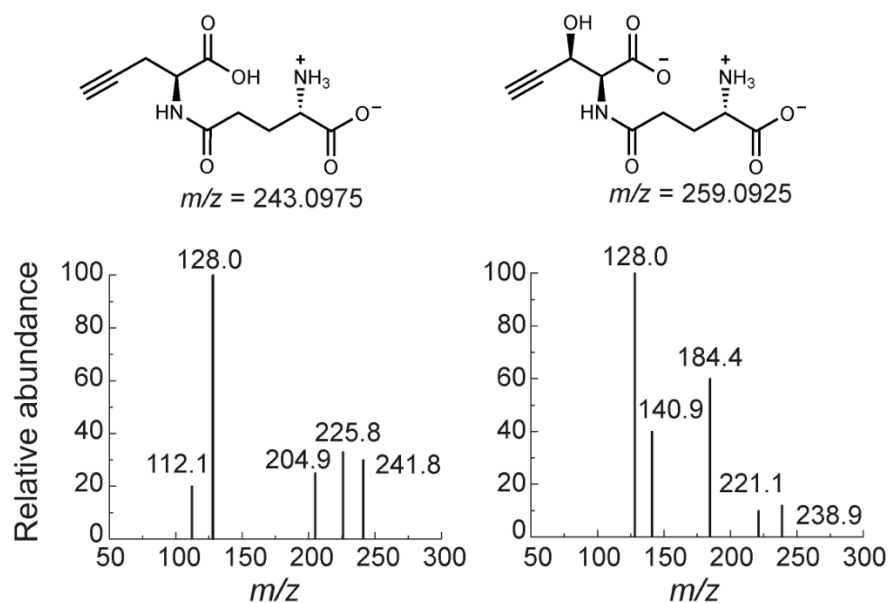


Figure 2.32. Fragmentation of γ -Glu-Pra (left) and γ -Glu- β es (right) in negative ionization mode by LC/QQQ shows fragments characteristic of γ -glutamyl-dipeptides ($m/z = 128.0$). Spectra shown are representative of at least 3 independent experimental replicates.

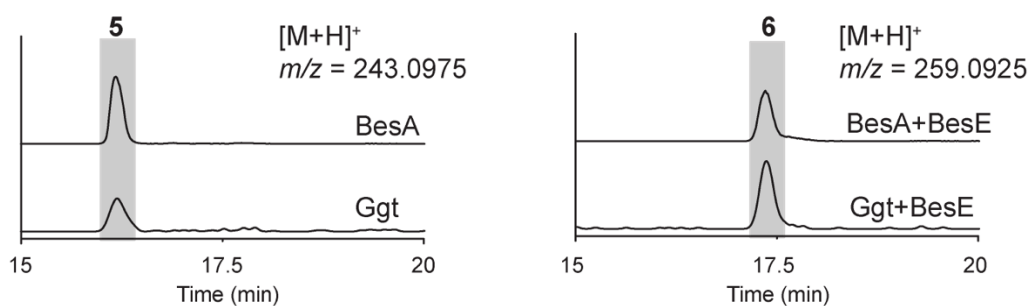


Figure 2.33. LC/MS traces for enzymatically synthesized γ -Glu-Pra (**5**, $m/z = 243.0975$) by either Ggt or BesA and LC/QTOF traces for enzymatically synthesized γ -Glu- β es (**6**, $m/z = 259.0925$) by either Ggt + BesE or BesA + BesE. Chromatograms shown are representative of at least 3 independent experimental replicates.

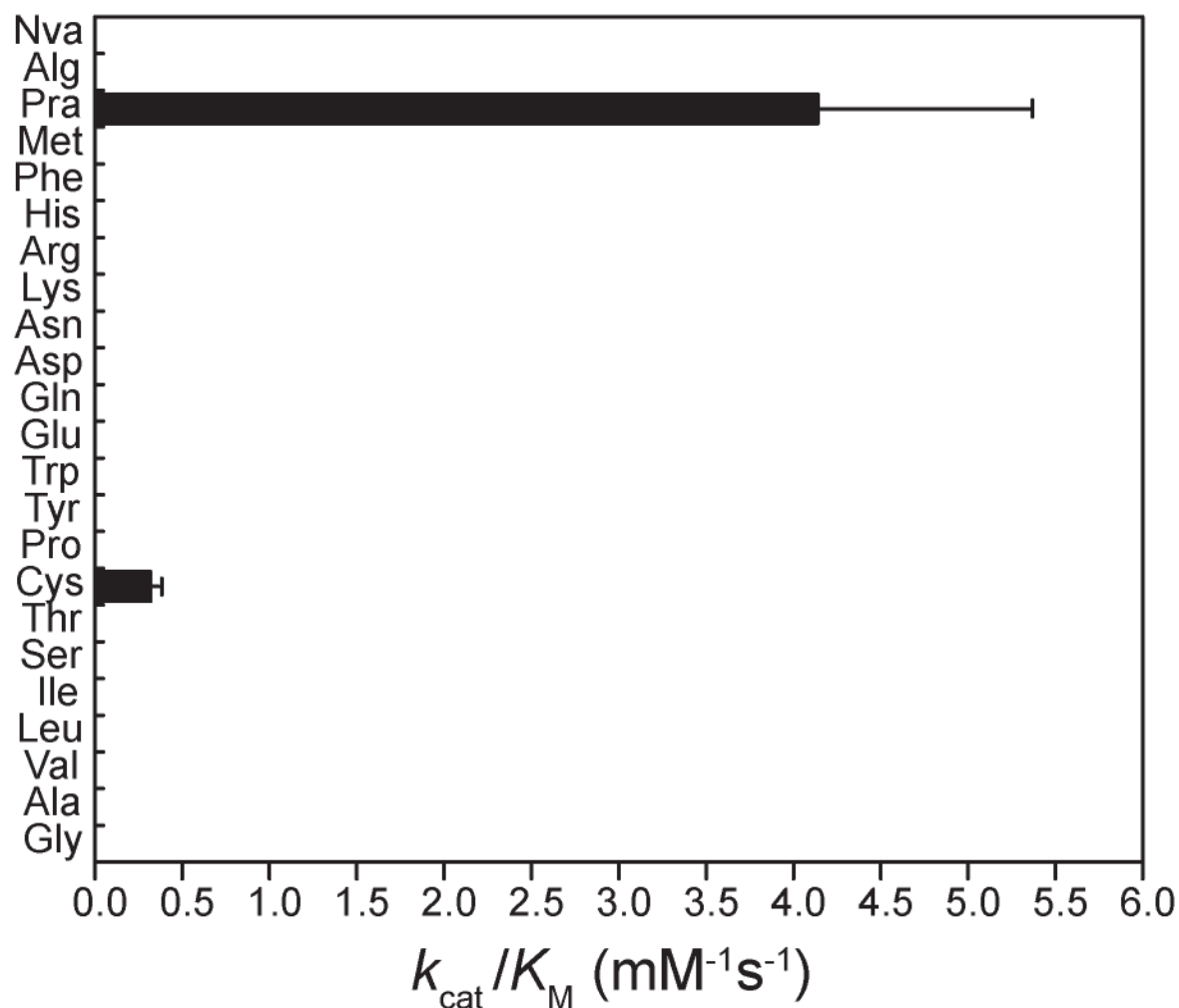


Figure 2.34. Steady-state kinetic analysis for BesA reactions between glutamate and various amino acid partners (Nva, norvaline; Alg, allylglycine). Each bar represents the $k_{\text{cat}}/K_{\text{M}}$ value for the corresponding amino acid calculated from the individual kinetic terms (k_{cat} and K_{M}) obtained from nonlinear curve fitting of data to the Michaelis-Menten equation. Data shown are mean $\pm$ standard error of the mean ($n = 3$ experimental replicates) obtained by propagation. Only reactions between glutamate and propargylglycine or cysteine showed any detectable activity.

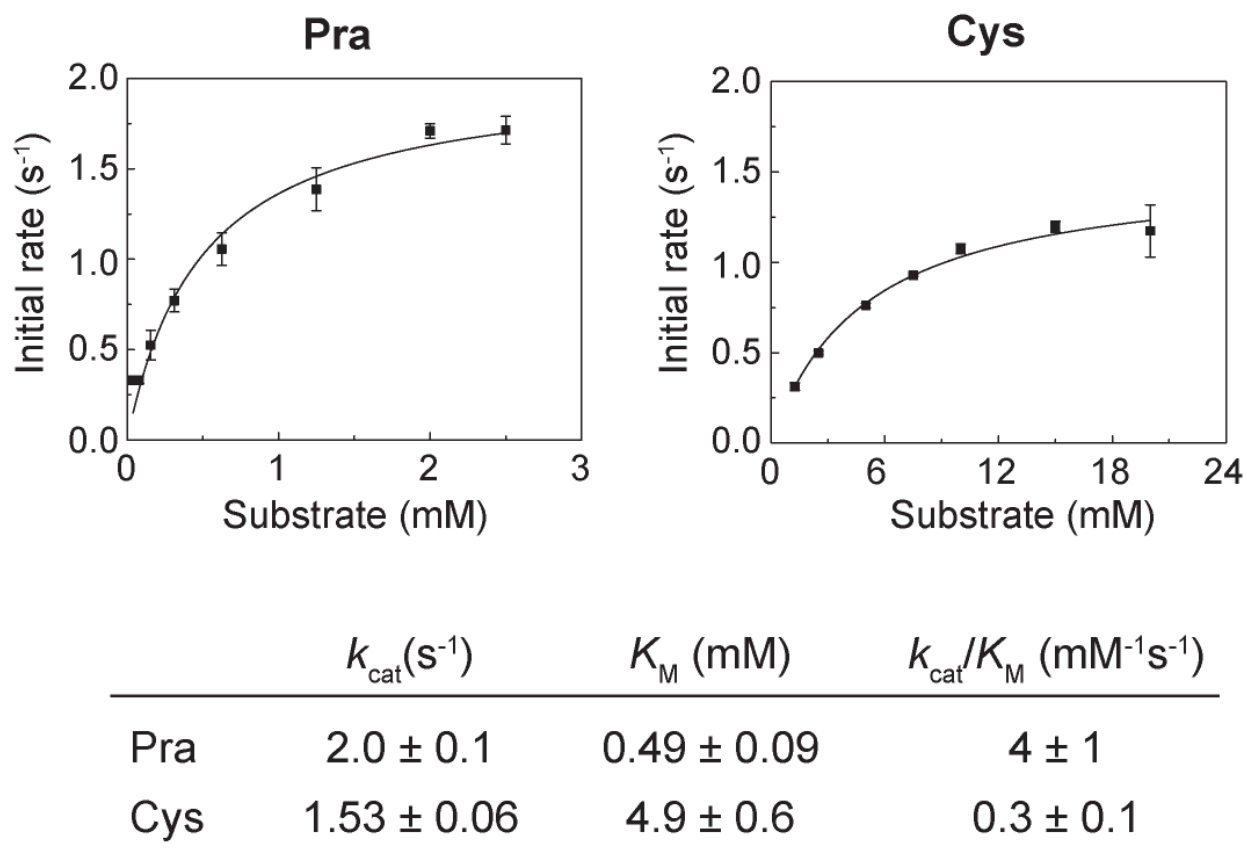


Figure 2.35. Steady state kinetic analysis of BesA with 10 mM glutamate and propargylglycine or cysteine as substrates. Data are mean $\pm$ s.d. ($n = 3$ experimental replicates). Table contains k_{cat} , K_{M} , and $k_{\text{cat}}/K_{\text{M}}$ calculated by non-linear curve fitting to the Michaelis-Menten equation. Data are mean $\pm$ s.e. Error in $k_{\text{cat}}/K_{\text{M}}$ is obtained by propagation from the individual kinetic terms.

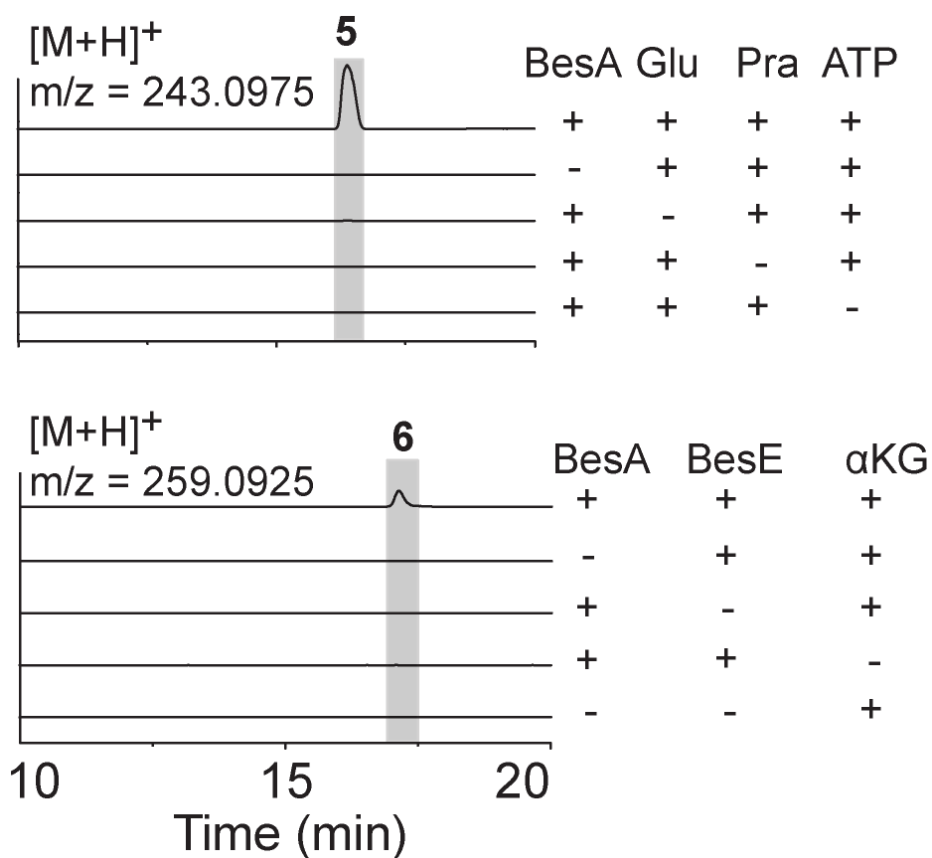


Figure 2.36. (Top) Extracted ion chromatograms for γ -Glu-Pra in reactions containing a combination of BesA, BesE, and cofactors (γ -Glu-Pra, **5**; m/z = 243.0975). (Bottom) Extracted ion chromatograms for γ -Glu- β es in reactions containing a combination of BesA, BesE, and cofactors (γ -Glu- β es, **6**; m/z = 259.0925). Chromatograms shown are representative of at least 3 independent experimental replicates.

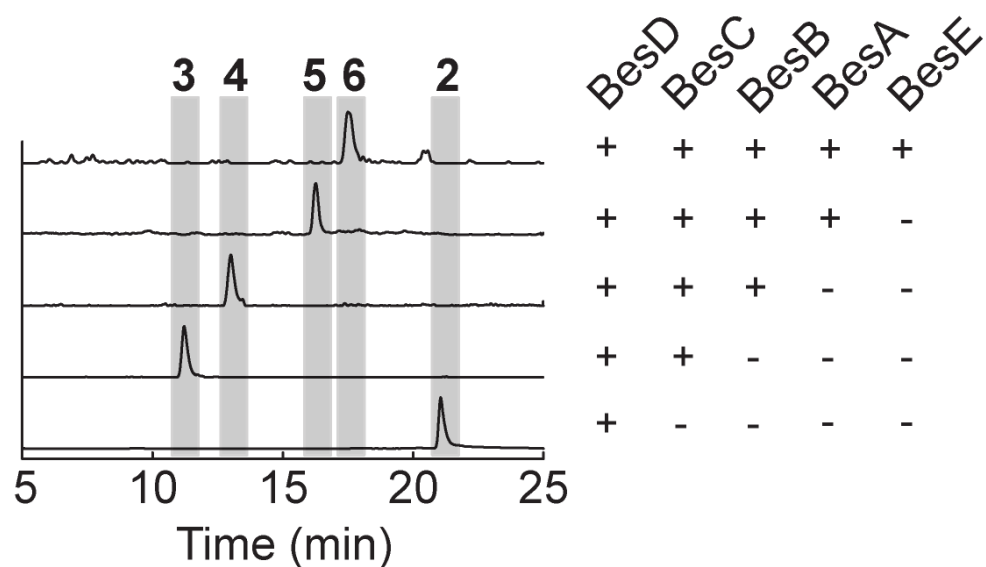
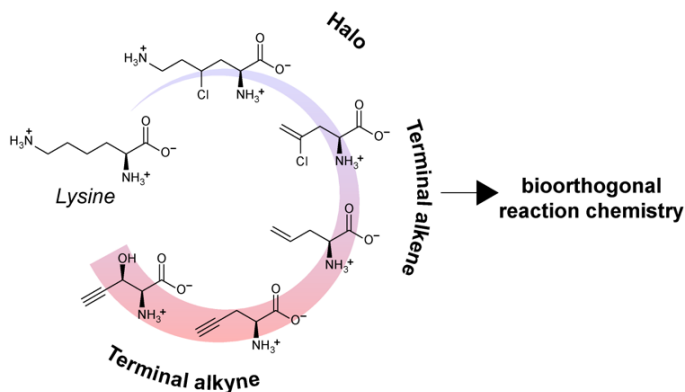


Figure 2.37. *In vitro* reconstitution of β es biosynthesis. Extracted ion chromatogram for each pathway intermediate in the presence of corresponding enzymes required to catalyze each step (**2**, $m/z = 181.0738$; **3**, $m/z = 150.0316$; **4**, $m/z = 114.550$; **5**, $m/z = 243.0975$; **6**, $m/z = 259.0925$). Peaks are normalized for each trace to the most abundant ion.

2.4. Conclusion

Though first discovered in the 1970s and 1980s, the biosynthetic origin of terminal-alkyne amino acids remained unknown. This work is the first to uncover how the terminal alkyne amino acids β es and Pra can be synthesized enzymatically from the primary metabolite lysine. The biosynthetic pathway includes several unusual chemical transformations and contains several interesting enzymes including an amino acid halogenase (BesD), a diiron-dependent enzyme that oxidatively cleaves a C-C bond (BesC), and a novel acetylenase (BesB). Together, the newly discovered enzymes enable the production of halo, terminal-alkene, and terminal-alkyne amino acids.



Scheme 2.1. Unnatural amino acids made in this study. The diverse functional groups may enable applications in bioorthogonal chemistry.

The subsequent introduction of these functional groups into peptides and proteins either site- or residue-specifically opens new opportunities to use a variety of downstream bioorthogonal chemistries, including alkyne-mediated CuAAC click reactions, olefin metathesis or polymerization, and halogen-dependent reactions for manipulating the structure, function, and physical properties of products.

2.5. References

1. Minto, R. E. & Blacklock, B. J. Biosynthesis and function of polyacetylenes and allied natural products. *Prog. Lipid Res.* **47**, 233–306 (2008) doi:10.1016/j.plipres.2008.02.002.
2. Kuklev, D. V., Domb, A. J. & Dembitsky, V. M. Bioactive acetylenic metabolites. *Phytomedicine* **20**, 1145–1159 (2013) doi:10.1016/j.phymed.2013.06.009.
3. Scrimgeour, C. M. Natural acetylenic and olefinic compounds, excluding marine natural products. *Aliphatic Relat. Nat. Prod. Chem.* **2**, (1979) doi:10.1039/9781847555564-00001.
4. Potgieter, H. C., Vermeulen, N. M. J., Potgieter, D. J. J. & Strauss, H. F. A toxic amino acid, 2(S)3(R)-2-amino-3-hydroxypent-4-ynoic acid from the fungus *Sclerotium rolfsii*. *Phytochemistry* **16**, 1757–1759 (1977) doi:10.1016/0031-9422(71)85083-5.
5. Sanada, M., Miyano, T. & Iwadare, S. Ethynylserine, an antimetabolite of L-threonine, from *Streptomyces cattleya*. *J. Antibiot. (Tokyo)*. **39**, 304–305 (1986) doi:10.7164/antibiotics.39.304.
6. Scannell, J. P., Pruess, D. L., Demny, T. C., Weiss, F., Williams, T. & Stempel, A. Antimetabolites Produced by Microorganisms II. *J. Antibiot. (Tokyo)*. **24**, 239–244 (1971) doi:10.7164/antibiotics.24.239.
7. Zhou, Z.-Y., Shi, G.-Q., Fontaine, R., Wei, K., Feng, T., Wang, F., Wang, G.-Q., Qu, Y., Li, Z.-H., Dong, Z.-J., Zhu, H.-J., Yang, Z.-L., Zeng, G. & Liu, J.-K. Evidence for the Natural Toxins from the Mushroom *Trogia venenata* as a Cause of Sudden Unexpected Death in Yunnan Province, China. *Angew. Chemie Int. Ed.* **51**, 2368–2370 (2012) doi:10.1002/anie.201106502.
8. Kuroda, Y., Okuhara, M., Goto, T., Iguchi, E., Kohsaka, M., Aoki, H. & Imanaka, H. FR-900130, a novel amino acid antibiotic. I. Discovery, taxonomy, isolation, and properties. *J. Antibiot. (Tokyo)*. **33**, 125–131 (1980) doi:10.7164/antibiotics.33.125.
9. Hatanaka, S.-I. Amino Acids from Mushrooms. in 1–140 (Springer, Vienna, 1992). doi:10.1007/978-3-7091-9150-7_1.
10. Simone, A. De, Acevedo-Rocha, C. G., Hoesl, M. G. & Budisa, N. Towards reassignment of the methionine codon AUG to two different noncanonical amino acids in bacterial translation. *Croat. Chem. Acta* **89**, (2016) doi:10.5562/cca2915.
11. Dieterich, D. C., Lee, J. J., Link, A. J., Graumann, J., Tirrell, D. A. & Schuman, E. M. Labeling, detection and identification of newly synthesized proteomes with bioorthogonal non-canonical amino-acid tagging. *Nat. Protoc.* **2**, 532–540 (2007) doi:10.1038/nprot.2007.52.
12. Truong, F., Yoo, T. H., Lampo, T. J. & Tirrell, D. A. Two-strain, cell-selective protein labeling in mixed bacterial cultures. *J. Am. Chem. Soc.* **134**, 8551–6 (2012) doi:10.1021/ja3004667.
13. Lang, K. & Chin, J. W. Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chem. Rev.* **114**, 4764–4806 (2014) doi:10.1021/cr400355w.
14. Uttamapinant, C., Howe, J. D., Lang, K., Beránek, V., Davis, L., Mahesh, M., Barry, N. P. & Chin, J. W. Genetic Code Expansion Enables Live-Cell and Super-Resolution Imaging of Site-Specifically Labeled Cellular Proteins. *J. Am. Chem. Soc.* **137**, 4602–4605 (2015) doi:10.1021/ja512838z.
15. de Graaf, A. J., Kooijman, M., Hennink, W. E. & Mastrobattista, E. Nonnatural Amino Acids for Site-Specific Protein Conjugation. *Bioconjug. Chem.* **20**, 1281–1295 (2009) doi:10.1021/bc800294a.
16. Sakamoto, K., Hayashi, A., Sakamoto, A., Kiga, D., Nakayama, H., Soma, A., Kobayashi, T., Kitabatake, M., Takio, K., Saito, K., Shirouzu, M., Hirao, I. & Yokoyama, S. Site-specific incorporation of an unnatural amino acid into proteins in mammalian cells. *Nucleic Acids Res.* **30**, 4692–4699 (2002) doi:10.1093/nar/gkf589.

17. Santoro, S. W., Wang, L., Herberich, B., King, D. S. & Schultz, P. G. An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat. Biotechnol.* **20**, 1044–1048 (2002) doi:10.1038/nbt742.
18. Bindman, N. A. & van der Donk, W. A. A General Method for Fluorescent Labeling of the N-Termini of Lanthipeptides and Its Application to Visualize their Cellular Localization. *J. Am. Chem. Soc.* **135**, 10362–10371 (2013) doi:10.1021/ja4010706.
19. Zhu, X., Liu, J. & Zhang, W. De novo biosynthesis of terminal alkyne-labeled natural products. *Nat. Chem. Biol.* **11**, 115–20 (2015) doi:10.1038/nchembio.1718.
20. Edwards, D. J., Marquez, B. L., Nogle, L. M., McPhail, K., Goeger, D. E., Roberts, M. A. & Gerwick, W. H. Structure and Biosynthesis of the Jamaicamides, New Mixed Polyketide-Peptide Neurotoxins from the Marine Cyanobacterium *Lyngbya majuscula*. *Chem. Biol.* **11**, 817–833 (2004) doi:10.1016/j.chembiol.2004.03.030.
21. Zhu, X., Su, M., Manickam, K. & Zhang, W. Bacterial Genome Mining of Enzymatic Tools for Alkyne Biosynthesis. *ACS Chem. Biol.* **10**, 2785–2793 (2015) doi:10.1021/acscchembio.5b00641.
22. Gust, B., Challis, G. L., Fowler, K., Kieser, T. & Chater, K. F. PCR-targeted *Streptomyces* gene replacement identifies a protein domain needed for biosynthesis of the sesquiterpene soil odor geosmin. *Proc. Natl. Acad. Sci.* **100**, 1541–1546 (2003) doi:10.1073/pnas.0337542100.
23. Tautenhahn, R., Patti, G. J., Rinehart, D. & Siuzdak, G. XCMS Online: A Web-Based Platform to Process Untargeted Metabolomic Data. *Anal. Chem.* **84**, 5035–5039 (2012) doi:10.1021/ac300698c.
24. Fisher, H. F. L-Glutamate dehydrogenase from bovine liver. *Methods Enzymol.* **113**, 16–27 (1985) doi:10.1016/s0076-6879(85)13006-5.
25. Compton, B. J. & Purdy, W. C. Fluoral-P, a member of a selective family of reagents for aldehydes. *Anal. Chim. Acta* **119**, 349–357 (1980) doi:10.1016/S0003-2670(01)93636-0.
26. Kabsch, W. XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125–32 (2010) doi:10.1107/S0907444909047337.
27. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D. Biol. Crystallogr.* **69**, 1204–14 (2013) doi:10.1107/S0907444913000061.
28. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D. Biol. Crystallogr.* **67**, 235–42 (2011) doi:10.1107/S0907444910045749.
29. Cowtan, K. The Buccaneer software for automated model building. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 1002–1011 (2006) doi:10.1107/S0907444906022116.
30. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of *Coot*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010) doi:10.1107/S0907444910007493.
31. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–21 (2010) doi:10.1107/S0907444909052925.
32. Lopes, C. T., Franz, M., Kazi, F., Donaldson, S. L., Morris, Q. & Bader, G. D. Cytoscape Web: an interactive web-based network browser. *Bioinformatics* **26**, 2347–2348 (2010) doi:10.1093/bioinformatics/btq430.
33. Blasiak, L. C., Vaillancourt, F. H., Walsh, C. T. & Drennan, C. L. Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature* **440**, 368–371 (2006) doi:10.1038/nature04544.

34. Merkx, R., de Bruin, G., Kruithof, A., van den Bergh, T., Snip, E., Lutz, M., El Oualid, F. & Ovaa, H. Scalable synthesis of γ -thiolysine starting from lysine and a side by side comparison with δ -thiolysine in non-enzymatic ubiquitination. *Chem. Sci.* **4**, 4494 (2013) doi:10.1039/c3sc51599k.
35. Schwarzenbacher, R., Stenner-Liewen, F., Liewen, H., Robinson, H., Yuan, H., Bossy-Wetzel, E., Reed, J. C. & Liddington, R. C. Structure of the Chlamydia protein CADD reveals a redox enzyme that modulates host cell apoptosis. *J. Biol. Chem.* **279**, 29320–29324 (2004) doi:10.1074/jbc.M401268200.
36. Lindqvist, Y., Huang, W., Schneider, G. & Shanklin, J. Crystal structure of delta9 stearyl-acyl carrier protein desaturase from castor seed and its relationship to other di-iron proteins. *EMBO J.* **15**, 4081–92 (1996) doi:10.1002/j.1460-2075.1996.tb00783.x.
37. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845–858 (2015) doi:10.1038/nprot.2015.053.
38. Davis, M. D., Kaufman, S. & Milstien, S. A modified ferrozine method for the measurement of enzyme-bound iron. *J. Biochem. Biophys. Methods* **13**, 39–45 (1986) doi:10.1016/0165-022X(86)90006-0.
39. Rui, Z., Li, X., Zhu, X., Liu, J., Domigan, B., Barr, I., Cate, J. H. D. & Zhang, W. Microbial biosynthesis of medium-chain 1-alkenes by a nonheme iron oxidase. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 18237–42 (2014) doi:10.1073/pnas.1419701112.
40. Ng, T. L., Rohac, R., Mitchell, A. J., Boal, A. K. & Balskus, E. P. An N-nitrosating metalloenzyme constructs the pharmacophore of streptozotocin. *Nature* **566**, 94–99 (2019) doi:10.1038/s41586-019-0894-z.
41. Brzovic, P., Holbrook, E. L., Greene, R. C. & Dunn, M. F. Reaction mechanism of Escherichia coli cystathionine γ -synthase: direct evidence for a pyridoxamine derivative of vinylglyoxylate as a key intermediate in pyridoxal phosphate dependent γ -elimination and γ -replacement reactions. *Biochemistry* **29**, 442–451 (1990) doi:10.1021/bi00454a020.
42. Sun, Q., Collins, R., Huang, S., Holmberg-Schiavone, L., Anand, G. S., Tan, C.-H., van-den-Berg, S., Deng, L.-W., Moore, P. K., Karlberg, T. & Sivaraman, J. Structural Basis for the Inhibition Mechanism of Human Cystathionine γ -Lyase, an Enzyme Responsible for the Production of H₂S. *J. Biol. Chem.* **284**, 3076–3085 (2009) doi:10.1074/jbc.M805459200.
43. Marcotte, P. & Walsh, C. Vinylglycine and propargylglycine: complementary suicide substrates for L-amino acid oxidase and D-amino acid oxidase. *Biochemistry* **15**, 3070–3076 (1976) doi:10.1021/bi00659a021.

Chapter 3: *Discovery and applications of radical amino acid halogenases*

Portions of this work were published in the following scientific journal:

Neugebauer, M. E., Sumida, K. H., Pelton, J. G., McMurry, J. L., Marchand, J. A., Chang, M. C. Y. A family of radical halogenases for the engineering of amino-acid-based products. *Nat. Chem. Biol.* **2019**, 15, 1009-1016.

Portions of this work were performed in collaboration with the following persons:

Dr. Jeff Pelton performed NMR experiments. Halogenase biochemistry was performed in collaboration with Kiera Sumida.

3.1. Introduction

The development of approaches to bridge synthetic and cellular chemistry can enable access to novel structures and classes of compounds by combining the exquisite selectivity of enzymes with breadth of functional groups used for synthetic transformations [1–3]. In particular, the introduction of halogens (X = F, Cl, Br, I) into a functional group-dense scaffold is especially useful, serving both to tune bioactivity [4,5] and act as a reactive handle for the formation of new chemical bonds for diversification and modification of late-stage synthetic intermediates [6,7]. While the role of halogens in biosynthetic transformations is continuing to be elucidated [8–11], their value in a variety of synthetic transformations and the design of synthetic routes is well-accepted [12]. As a result, both synthetic and enzymatic approaches to the formation of carbon-halogen bonds are currently areas of major interest [6–8,13–15].

The selective modification of unactivated sp³ C–H bonds is particularly difficult, as it is challenging to identify catalysts that are sufficiently powerful to activate these bonds while maintaining regio- and stereoselectivity [16]. Remarkably, Nature has evolved a set of non-heme Fe^{II}/ α -ketoglutarate (Fe^{II}/ α KG)-dependent enzymes that can achieve this task, using a high-valent metal-oxo intermediate to generate a substrate radical that can rebound with the bound halide ligand [17–19]. However, only a handful of these radical halogenases have been characterized and are notable for their complex substrates (*Figure 3.1*). The SyrB2 family has been found to maintain a strict requirement for carrier-protein tethered substrates [20], whereas the WelO5 family halogenates late-stage indole alkaloid natural products [21]. Since these types of intermediates are not readily modified using downstream enzymatic pathways, the discovery of new enzymes that act on simple and modular building blocks would greatly expand the biosynthetic potential of radical halogenation.

In the process of elucidating the biosynthesis of β es, we discovered a novel halogenase, BesD, that chlorinates the free amino acid lysine. BesD is the first α KG-dependent radical halogenase reported to chlorinate an amino acid without the requirement for a carrier protein [9]. In this chapter, we explore the structural basis for direct halogenation of a free amino acid by BesD. In addition, we use bioinformatics to explore the sequence space around BesD to identify additional halogenases that chlorinate different polar and nonpolar amino acids. Taking advantage of the central role of amino acid building blocks in the cellular production of metabolites, natural products, and macromolecules [22–24], we demonstrate that the halogenated amino acids produced by the BesD family can be enzymatically converted into a range of compound classes such as nitrogen heterocycles, amines, ketoacids, and peptides. Taken together, this work expands the utility of radical halogenases and highlights the promise of the BesD halogenase family in the biological production of diverse halogenated compounds.

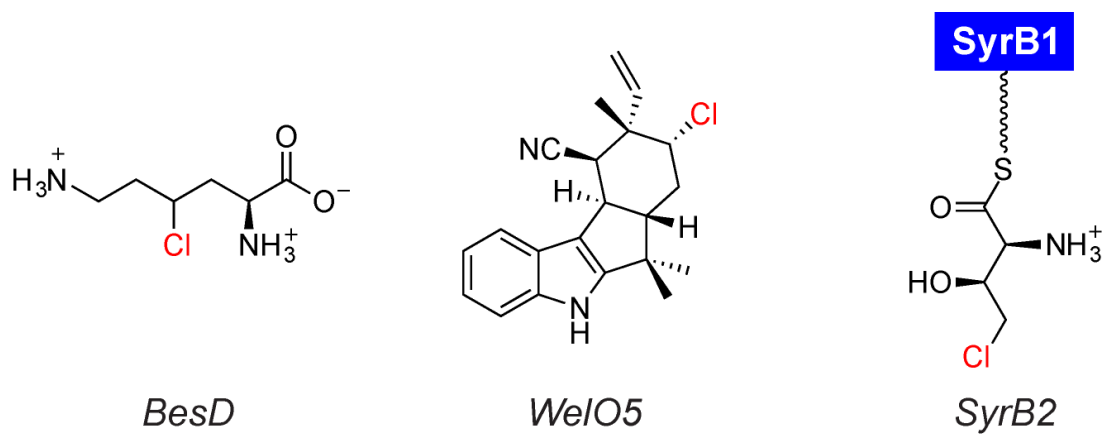


Figure 3.1. Products of the Fe^{II}/αKG-dependent halogenases BesD, WelO5, and SyrB2.

3.2. Materials and methods

Commercial materials. Luria-Bertani (LB) Broth Miller, LB Agar Miller, Terrific Broth (TB), and glycerol were purchased from EMD Biosciences (Darmstadt, Germany). Carbenicillin (Cb), kanamycin (Km), chloramphenicol (Cm), isopropyl- β -D-thiogalactopyranoside (IPTG), sodium chloride, dithiothreitol (DTT), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), magnesium chloride hexahydrate, acetonitrile, ethylene diamine tetraacetic acid disodium dihydrate (EDTA), hydrochloric acid, 3kDa MWCO dialysis tubing, ammonium bicarbonate, sodium acetate, chloroform, and sodium hydroxide were purchased from Fisher Scientific (Pittsburgh, PA). Phosphoenolpyruvate (PEP), adenosine triphosphate sodium salt (ATP), nicotinamide adenine dinucleotide reduced form dipotassium salt (NADH), pyruvate kinase, lactate dehydrogenase, lysozyme, poly(ethyleneimine) solution (PEI), ammonium iron (II) sulfate hexahydrate, α -ketoglutaric acid sodium salt, β -mercaptoethanol (β ME), N,N,N',N'-tetramethyl-ethane-1,2-diamine (TEMED), betaine, dimethylsulfoxide (DMSO), pyridoxal-5'-phosphate (PLP), sodium L-ascorbate, acetonitrile (LC/MS-grade), ammonium formate (LC/MS-grade), Deuterium oxide, L-lysine for crystallography, 20 canonical amino acids, seleno-L-methionine, L-leucine ($^{13}\text{C}_6$, 98%), L-isoleucine ($^{13}\text{C}_6$, 98%; ^{15}N , 98%), DL-norleucine, 2-(N-Morpholino)ethanesulfonic acid hydrate, 4-morpholineethanesulfonic acid (MES), Pall Nanosep centrifugal Omega membrane (3,000 MWCO), sodium bromide, and sodium azide were purchased from Sigma-Aldrich (St. Louis, MO). PageRuler Plus Prestained Protein Ladder was purchased from ThermoFisherScientific (Waltham, MA). Succinyl-CoA synthetase was purchased from Megazyme International (Bray, Ireland). Methanol (D_4 , 99.8%), chloroform (D , 99.8%), L-lysine $\cdot 2\text{HCl}$ ($^{13}\text{C}_6$, 99%; $^{15}\text{N}_2$, 99%) and L-ornithine ($^{13}\text{C}_5$, 99%; $^{15}\text{N}_2$, 99%) were purchased from Cambridge Isotope Laboratories (Andover, MA). Formic acid was purchased from Acros Organics (Morris Plains, NJ). Restriction enzymes, T4 DNA ligase, Antarctic phosphatase, Phusion DNA polymerase, Q5 DNA polymerase, T5 exonuclease, Taq DNA ligase, and the PURExpress *in vitro* transcription/translation kit $\Delta(\text{aa}, \text{tRNA})$ were purchased from New England Biolabs (Ipswich, MA). Deoxynucleotides (dNTPs), were purchased from Invitrogen (Carlsbad, CA). Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA), resuspended at a stock concentration of 100 μM in water and stored at either 4 $^\circ\text{C}$ for immediate usage or -20 $^\circ\text{C}$ for long-term storage. DNA purification kits and Ni-NTA agarose were purchased from Qiagen (Valencia, CA). Zirconia/silica beads were purchased from BioSpec Products (Bartlesville, OK). Complete EDTA-free protease inhibitor was purchased from Roche Applied Science (Penzberg, Germany). PD-10 desalting columns and Superdex 75 16/60 pg column were purchased from GE Healthcare (Pittsburg, PA). Amicon Ultra 10,000 MWCO centrifugal concentrators and Milli-Q Gradient water purification system were purchased from Millipore (Billerica, MA). Acrylamide/bis-acrylamide (30%, 37.5:1), electrophoresis grade sodium dodecyl sulfate (SDS), and ammonium persulfate were purchased from Bio-Rad Laboratories (Hercules, CA). Ultrayield baffled flasks were purchased from Thompson Instrument Company (Oceanside, CA).

Bioinformatic selection of BesD halogenase homologs for screening. To select halogenase homologs for functional screening, the sequence of BesD (WP_014151497) was used as a query for BLAST against the non-redundant protein database with a cutoff value of e^{-5} , yielding 261 hits. Blast hits were aligned using MUSCLE [25] and visualized in AliView [26], and the sequence position of the HXD/E or HXG/A motif was identified manually. A Bioython script was used to parse the list of alignments to segregate the halogenases (HXG/A) for storage in a

separate file, yielding 105 sequences which were retrieved using BatchEntrez (NCBI). From the list of halogenases, redundant sequences of 98% identity were removed using CD-HIT [27]. The remaining halogenase sequences with homology to BesD were input into the EFI-EST web tool (<https://efi.igb.illinois.edu/efi-est>) [28], using Option C to read FASTA files with headers. After initial processing, sequences below 150 amino acids in length were excluded from the sequence similarity network. Using Cytoscape Version 3.6.1, the network was adjusted by deleting edges with low alignments scores until the halogenases identified for β es biosynthesis [9] were separated into an isofunctional cluster, which occurred at an alignment score value of 88 (which corresponds to a sequence identity of ~56%). The EFI-GNT web tool (<https://efi.igb.illinois.edu/efi-gnt>) [28] was used to identify adjacent genes that were shared among 50% of the halogenase sequences in each cluster, revealing genes associated with amino acid metabolism.

Phylogenetic analysis of BesD homologs. The list of halogenases for phylogenetic analysis was obtained as described above. Halogenase sequences were aligned using MUSCLE. The MEGAX interface was then used to analyze the alignment by Maximum-Likelihood based on the JTT matrix-based model [29], and uncertainty in the topology of the resulting tree was evaluated with 500 bootstrap replicates.

Bioinformatic analysis of active site conservation among halogenases vs hydroxylases. Separate lists of halogenases (HXG/A motif) and hydroxylases (HXD/E motif) were obtained as described above. Redundant sequences with at least 90% identity to other sequences under consideration were removed using CD-HIT, leaving 48 halogenases and 97 hydroxylases. Next, the halogenase and hydroxylase lists were separately analyzed with LOGOS (<http://weblogo.threeplusone.com/create.cgi>) [30] to generate a graphical representation of sequence conservation within each data set.

Bioinformatic analysis of the variability of the C-termini of BesD halogenase homologs. As above, a list of BesD halogenase homologs within a cutoff value of e^{-5} was obtained. From the list of 105 halogenases, redundant sequences of 90% identity were removed using CD-HIT, leaving 48 sequences. These remaining halogenases were input into LOGOS to generate a graphical representation of sequence conservation throughout the halogenases.

Bacterial strains. *E. coli* DH10B-T1^R was used for plasmid construction. BL21(DE3)–Star harboring the pRARE2 plasmid was used for heterologous protein production of all halogenases. *Streptomyces cattleya* NRRL 8057 (ATCC 35852) was obtained from the American Tissue Type Collection (Manassas, VA). The following strains were purchased from DSMZ: *Pseudomonas orientalis* (DSMZ 17489), *Pseudomonas trivialis* (DSMZ 14937), *Pseudomonas* sp. *Root562* (DSMZ 102504), and *Pseudomonas fulva* (DSMZ 17717). *Pseudomonas putida* KT2440 was a gift from the Keasling laboratory (UC Berkeley).

Construction of plasmids for protein expression. Gibson assembly was used to carry out plasmid construction using *E. coli* DH10B-T1^R as the cloning host. PCR amplifications were carried out with Phusion polymerase or Q5 polymerase (New England Biolabs, Ipswich MA) using the oligonucleotides listed in *Appendix A1.2*. The intermediate cloning plasmid, pET16-His-PrescissionCutSite-IMPDH, was constructed by amplification of IMPDH from *E. coli* gDNA followed by insertion into NdeI/BamHI-digested pET16b. The halogenases WP_014151497 (*S. cattleya* BesD), WP_057723975 (*P. orientalis* HalA), WP_057008702 (*P. trivialis* HalD), WP_056854138 (*P. sp. Root562* HalE), WP_028687259 (*P. fulva* HalE), as well as lysine decarboxylase WP_001020973 (LDC, *E. coli*), aminotransferase 1104250A (IlvE, *E. coli*) and

ornithine cyclodeaminase WP_010954390 (OCD, *P. putida* KT2440) were amplified from genomic DNA of the corresponding organism and inserted into NdeI/BamHI-digested pET16-His-PrescissionCutSite-IMPDH. The halogenases WP_016975823 (*P. fluorescens* HalA), WP_044582715 (*S. iranensis* HalB), WP_046063366 (*P. kilonensis* HalD), WP_047703910 (*P. mediterranea* HalH) were obtained as GeneBlocks and amplified with primers from *Appendix A1.2* before Gibson insertion into NdeI/BamHI-digested His-PrescissionCutSite-IMPDH. The remaining halogenases in *Appendix A1.2* and the lysine cyclodeaminase RapL were obtained as GeneBlocks containing the overhangs required for direct Gibson insertion into NdeI/BamHI-digested pET16-His-Prescission-IMPDH. To clone mutants of *S. lavenduligriseus* BesD (SlBesD), SlBesD was amplified in two pieces from the pET16-His-SlBesD plasmid. The first piece was amplified with Sl-BesD-F and the reverse primers indicated in *Appendix A1.2*. The second piece was amplified with the forward primers indicated in *Appendix A1.2* and either Sl-BesD-F (for R74A, E120A, H134A, D140A) or pET16-anneal-R (for N219A, T221A, W238A, W239A). Then both pieces were inserted into NdeI/BamHI-digested pET16-His-PrescissionCutSite-IMPDH to yield the mutant constructs. pSV272.1-His₆-MBP-PfRoot562 HalE was constructed by genomic DNA amplification with the primers indicated in *Appendix A1.2* followed by insertion into SfoI/EcoRI-digested pSV272.1. IVTT-PepC and IVTT-PepA were cloned by vector-amplification of the DHFR plasmid from the PURExpress IVTT kit (E640S, NEB) using the PepC-F/R or PepA-F/R primer sets, respectively, followed by Gibson assembly. Following plasmid construction, all cloned inserts were sequenced at Quintara Biosciences (San Francisco, CA) or the Barker Hall Sequencing Facility at UC Berkeley (Berkeley, CA).

Expression of His₁₀-tagged and His₁₀-MBP-tagged proteins. *E. coli* BL21 Star (DE3) harboring the pRARE2 plasmid was transformed with the appropriate protein expression plasmid. An overnight TB culture of the freshly transformed cells was used to inoculate TB (1 L) containing the appropriate antibiotics (50 µg/mL carbenicillin or kanamycin with 50 µg/mL chloramphenicol) in a 2.8 L-baffled shake flask to OD₆₀₀ = 0.05. The cultures were grown at 37°C at 200 rpm to OD₆₀₀ = 0.6 to 0.8 at which point cultures were cooled on ice for 20 min, followed by induction of protein expression with IPTG (0.25 mM) and overnight growth at 16°C. Cell pellets were harvested by centrifugation at 9,800 × g for 7 min at 4°C and stored at -80°C.

Purification of His₁₀-tagged and His₁₀-MBP-tagged proteins. Frozen cell pellets were thawed and resuspended at 5 mL/g of cell paste in lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM imidazole, 20 mM βME, 20% (v/v) glycerol, pH 7.5) supplemented with EDTA-free Protease Inhibitor Cocktail (Roche). The cell paste was homogenized and then lysed by passage through a French Pressure cell (Thermo Scientific; Waltham, MA) at 9,000 psi or by sonication with a Misonix Sonicator 3000 (power = 7.5, 5 s on, 25 s off, 2 min total process time, 1/2" tip). The lysate was then centrifuged at 13,500 × g for 20 min at 4°C to separate the soluble and insoluble fractions. DNA was precipitated in the soluble fraction with 0.15% (w/v) polyethyleneimine and stirring at 4°C for 30 min. The precipitated DNA was then removed by centrifugation at 13,500 × g for 20 min at 4°C. The soluble lysate was incubated with Ni-NTA (0.5 mL resin/g of cell paste) for 45 min at 4°C, then resuspended and loaded onto a column by gravity flow. The column was washed with wash buffer (50 mM HEPES, 300 mM NaCl, 20 mM imidazole, 20 mM βME, 20% (v/v) glycerol, pH 7.5) for 15-20 column vol. The column was then eluted with elution buffer (50 mM HEPES, 300 mM NaCl, 300 mM imidazole, 20 mM βME, 20% (v/v) glycerol, pH 7.5). Fractions containing the target protein were pooled by A_{280 nm} and concentrated using an Amicon Ultra spin concentrator (10 kDa MWCO, Millipore). Protein was then exchanged into storage buffer (50 mM

HEPES, 100 mM sodium chloride, 20% (v/v) glycerol, 1 mM DTT, pH 7.5) using PD-10 desalting columns. Note that for the PLP-dependent enzymes IlvE and LDC, 20 μ M PLP was present in all the purification and storage buffers. Final protein concentrations before storage were estimated using the $\epsilon_{280\text{ nm}}$ calculated by ExPASy ProtParam [31] and measured by nanodrop. They are as follows: *S. cattleya* BesD: 2.21 mg/mL ($\epsilon_{280\text{ nm}} = 47,565\text{ M}^{-1}\text{ cm}^{-1}$), *P. fluorescens* HalA: 4.19 mg/mL ($\epsilon_{280\text{ nm}} = 47,565\text{ M}^{-1}\text{ cm}^{-1}$), *P. orientalis* HalA: 3.11 mg/mL ($\epsilon_{280\text{ nm}} = 50,545\text{ M}^{-1}\text{ cm}^{-1}$), *S. wuyuanensis* HalB: 10.67 mg/mL ($\epsilon_{280\text{ nm}} = 45,045\text{ M}^{-1}\text{ cm}^{-1}$), *S. toyocaensis* HalB: 12.84 mg/mL ($\epsilon_{280\text{ nm}} = 43,555\text{ M}^{-1}\text{ cm}^{-1}$), *S. viridosporus* HalB: 8.36 mg/mL ($\epsilon_{280\text{ nm}} = 43,555\text{ M}^{-1}\text{ cm}^{-1}$), *A. awajinensis* HalB: 2.82 mg/mL ($\epsilon_{280\text{ nm}} = 46,075\text{ M}^{-1}\text{ cm}^{-1}$), *S. griseus* HalB: 5.74 mg/mL ($\epsilon_{280\text{ nm}} = 42,065\text{ M}^{-1}\text{ cm}^{-1}$), *S. afghaniensis* HalB: 4.03 mg/mL ($\epsilon_{280\text{ nm}} = 42,065\text{ M}^{-1}\text{ cm}^{-1}$), *P. iranensis* HalB: 2.37 mg/mL ($\epsilon_{280\text{ nm}} = 45,045\text{ M}^{-1}\text{ cm}^{-1}$), *S. prunicolor* HalB: 1.24 mg/mL ($\epsilon_{280\text{ nm}} = 41,035\text{ M}^{-1}\text{ cm}^{-1}$), *P. anisa* HalC: 1.25 mg/mL ($\epsilon_{280\text{ nm}} = 53,525\text{ M}^{-1}\text{ cm}^{-1}$), *P. kilonensis* HalD: 4.06 mg/mL ($\epsilon_{280\text{ nm}} = 60,515\text{ M}^{-1}\text{ cm}^{-1}$), *P. sp. SHC52* HalD: 4.93 mg/mL ($\epsilon_{280\text{ nm}} = 60,515\text{ M}^{-1}\text{ cm}^{-1}$), *P. trivialis* HalD: 3.52 mg/mL ($\epsilon_{280\text{ nm}} = 60,515\text{ M}^{-1}\text{ cm}^{-1}$), *P. sp. Root562* HalE: 0.73 mg/mL ($\epsilon_{280\text{ nm}} = 62,005\text{ M}^{-1}\text{ cm}^{-1}$), *P. fulva* HalE: 1.30 mg/mL ($\epsilon_{280\text{ nm}} = 57,410\text{ M}^{-1}\text{ cm}^{-1}$), *S. pristinaespiralis* HalF: 2.23 mg/mL ($\epsilon_{280\text{ nm}} = 32,430\text{ M}^{-1}\text{ cm}^{-1}$), *M. pelagius* HalG: 8.92 mg/mL ($\epsilon_{280\text{ nm}} = 49,515\text{ M}^{-1}\text{ cm}^{-1}$), *P. corrugata* HalH: 6.46 mg/mL ($\epsilon_{280\text{ nm}} = 46,660\text{ M}^{-1}\text{ cm}^{-1}$), *P. mediteranea* HalH: 6.68 mg/mL ($\epsilon_{280\text{ nm}} = 46,410\text{ M}^{-1}\text{ cm}^{-1}$), *S. lavenduligriseus* BesD (SlBesD): 6.48 mg/mL, SlBesD N219A: 3.02 mg/mL, SlBesD T221A: 3.52 mg/mL, SlBesD R74A: 6.65 mg/mL, SlBesD D140A: 2.73 mg/mL, SlBesD E120A: 4.51 mg/mL, SlBesD H134A: 3.67 mg/mL ($\epsilon_{280\text{ nm}} = 50,545\text{ M}^{-1}\text{ cm}^{-1}$); SlBesD W238A: 6.80 mg/mL, SlBesD W239A: 7.28 mg/mL ($\epsilon_{280\text{ nm}} = 45,045\text{ M}^{-1}\text{ cm}^{-1}$), RapL 0.54 mg/mL ($\epsilon_{280\text{ nm}} = 3,230\text{ M}^{-1}\text{ cm}^{-1}$), OCD, 2.13 mg/mL ($\epsilon_{280\text{ nm}} = 20,065\text{ M}^{-1}\text{ cm}^{-1}$), IlvE: 13.7 mg/mL ($\epsilon_{280\text{ nm}} = 49,515\text{ M}^{-1}\text{ cm}^{-1}$), LDC: 4.01 mg/mL ($\epsilon_{280\text{ nm}} = 106,605\text{ M}^{-1}\text{ cm}^{-1}$). All proteins were aliquoted, flash-frozen in liquid nitrogen, and stored at -80°C .

Preparation of proteins for crystallization. For the BesD halogenase from *S. cattleya*, the protein was dialyzed following elution from Ni-NTA column $3 \times 1:50$ against dialysis buffer (50 mM HEPES, 100 mM NaCl, 1 mM DTT, pH 7.5) for 1.5 h to remove imidazole. After the third round of dialysis, protein was incubated with Prescission protease (1 mg protease/50 mg protein) and dialyzed $3 \times 1:50$ overnight into dialysis buffer. Cleaved and dialyzed protein was passed through Ni-NTA (2 mL) to remove Prescission and the His₁₀ tag. The eluent was diluted to a final salt concentration of 20 mM NaCl using Buffer A (50 mM HEPES, 20% (v/v) glycerol, 1 mM DTT, 1 mM EDTA, pH 7.5) and loaded onto a 5 mL HiTrap-Q column for ion exchange with the AKTA Purifier FPLC system (GE Healthcare). The protein was eluted using a gradient from 0-100% buffer A to buffer B (50 mM HEPES, 1 M NaCl, 20% (v/v) glycerol, 1 mM DTT, 1 mM EDTA, pH 7.5) over 40 min. The protein sample was concentrated to 2 mL and loaded onto a Superdex 75 16/60 pg (GE Healthcare) column equilibrated with SEC Buffer (20 mM HEPES, 100 mM NaCl, 1 mM DTT, pH 7.5). The protein eluent was concentrated to 15 mg/mL and glycerol was added to a final concentration of 5% (v/v) before flash freezing in liquid nitrogen and storage at -80°C .

Crystallization and structure determination. Crystals of BesD from *S. cattleya* were obtained by the hanging drop vapor diffusion method by combining equal volumes of protein solution (10 mg/mL BesD, lysine (1.5 mM), α KG (3 mM, pH 7)) and reservoir solution (MES pH 6.5 (100 mM), sodium chloride (0.6 mM) containing 20% (v/v) PEG 4000). Thick rectangular plate crystals grew in 4 days. Crystals were transferred to an Eppendorf tube containing 100 μ L of seed buffer (MES pH 6.5 (100 mM), sodium chloride (0.6 mM) containing 40% (v/v) PEG 4000) at 4°C and vortexed for 30 seconds with $10 \times 1\text{ mm}$ diameter zirconia/silica beads to produce a micro-

seed solution. The seed solution was stored at -80°C for future use. Fe-bound BesD crystals were prepared in a Coy Anaerobic chamber by micro-seeding equal volumes of protein solution (10 mg/mL BesD, lysine (1.5 mM), α KG (3 mM, pH 7)) and reservoir solution (MES pH 6.5 (100 mM), sodium chloride (0.6 mM), containing 20% (v/v) PEG 4000). Thick rectangular rods grew within 2 weeks. Crystals were soaked anaerobically in a solution containing $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ (1 mM), α KG (1 mM), sodium ascorbate (1 mM), lysine (1 mM), and glycerol (20% (v/v)) for 30 min before flash freezing in liquid nitrogen. Data were collected at Beamline 8.3.1 at the Advanced Light Source (Lawrence Berkeley National Laboratory) at a wavelength of 1.11 Å.

Data were processed with XDS [32] and scaled and merged with Aimless [33] in within the CCP4 suite [34]. The anomalous signal from the Fe atoms was sufficient for obtaining initial phases using the CRANK2 pipeline [35]. Buccaneer [36] was used to build an initial model of the structure, which was refined iteratively in COOT [37] and Phenix [38]. Ligands were added to the model using COOT and refined in Phenix. The overall *B*-factors of the protein are higher than that of the ligands due to partially disordered regions of the protein. However, the *B*-factors of the ligands are similar to that of the surrounding protein residues in the active site. Omit maps for the ligand complex including Fe, Cl, α KG, lysine, His137, and His 204 were generated using the Phenix Map function. The Fe, Cl, α KG, lysine, His137, and His 204 were selected for removal from refinement prior to calculation of the maps. No map is visible around those ligands at the -3 σ contour level. The structure was analyzed in Pymol, which was also used to create figures.

General procedure for high resolution HPLC/MS analysis of polar metabolites with HILIC. Samples containing polar metabolites were analyzed using an Agilent 1290 UPLC on a SeQuant ZIC-pHILIC (5 μ m, 2.1 $\times$ 100 mm; EMD-Millipore) using the following buffers: Buffer A (90% acetonitrile, 10% water, 10 mM ammonium formate) and Buffer B (90% water, 10% acetonitrile, 10 mM ammonium formate). A linear gradient from 95% to 60% Buffer A over 17 min followed by a linear gradient from 60% to 33% Buffer A over 8 min was then applied at a flow rate of 0.2 mL/min. Mass spectra were acquired in positive ionization mode using an Agilent 6530 QTOF (Agilent).

***In vitro* screening of BesD homolog activity.** Reactions (50 μ L) contained the following amino acids: L-alanine, L-arginine, L-asparagine, L-aspartate, L-cysteine, L-glutamine, L-glutamate, L-glycine, L-histidine, L-isoleucine, L-leucine, L-lysine, L-methionine, L-phenylalanine, L-proline, L-serine, L-threonine, L-tryptophan, L-valine, L-ornithine, and DL-norleucine (0.5 mM each), sodium α KG (5 mM), sodium ascorbate (1 mM), $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ (1 mM), and sodium chloride (5 mM) in 100 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of purified halogenase variants (10 μ M final concentration) and allowed to proceed for 1 h at room temperature before quenching in 2 vol of methanol with 1% (v/v) formic acid. Samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above. Following identification of substrates from the amino acid pool, the reactions were repeated with individual amino acids lysine, ornithine, leucine, isoleucine, and norleucine tested separately (0.5 mM each). In the case of the aliphatic halogenases which chlorinate leucine, isoleucine, and norleucine, the enzymes were re-assayed against each of the three amino acids separately (0.5 mM each) to confirm activity, since the substrates and products of these isomers have the same exact mass.

***In vitro* halogenation assay of BesD mutants.** Reactions (50 μ L) contained L-lysine $\cdot$ HCl (1 mM), sodium α KG (5 mM), sodium ascorbate (1 mM), $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ (1 mM), and

sodium chloride (5 mM) in 100 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of purified halogenase mutants or WT BesD from *S. lavanduligriseus* (20 μ M final concentration) and allowed to proceed for 20 min at room temperature before quenching in 2 vol of methanol with 1% (v/v) formic acid. Samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above.

Determination of kinetic parameters for PkHalD and PfHalA on lysine and ornithine. Reactions (100 μ L) contained ATP (2.5 mM), MgCl₂ (5 mM) phosphoenol-pyruvate (PEP, 1 mM), NADH (0.3 mM) lactate dehydrogenase (LDH, 10 U/mL), pyruvate kinase (PK, 10 U/mL) succinyl-CoA synthetase (SCS, 3.2 U/mL), coenzyme A (1 mM), sodium α KG (1 mM), (NH₄)₂Fe(SO₄)₂ · 6H₂O (0.2 mM), sodium chloride (10 mM), and sodium ascorbate (2 mM) in 100 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of PkHalD (5 μ M) or PfHalA (5 μ M) in the presence of varying concentrations of L-lysine · HCl (0 - 16 mM) or L-ornithine · HCl (0 - 2 mM). Initial rates of NADH consumption were measured by monitoring A₃₄₀ using a SynergyMx Microplate Reader (BioTek) at room temperature. k_{cat} and K_M were determined by fitting to initial rate data with Origin (OriginLab, Northampton, MA) using the equation:

$$v_0 = \frac{k_{cat}[S]}{K_M + [S]}$$

where v_0 is the initial rate and $[S]$ is the substrate concentration. Analogous experiments were performed to assess the activity of PrHalE on leucine, isoleucine, norleucine, and lysine.

Bromination and azidation assays. Reactions (50 μ L) contained L-lysine (1 mM), sodium α KG (5 mM), sodium ascorbate (5 mM), and (NH₄)₂Fe(SO₄)₂ · 6H₂O (1 mM) in 50 mM sodium acetate buffer (pH 7, adjusted with acetic acid). The anion was provided as either sodium bromide (100 mM) or sodium azide (0.5 mM). Note that halogenases used for these assays were desalted into sodium acetate (25 mM, pH 7) to prevent competition from the native chloride ligand found in the enzyme storage buffer. Reactions were initiated by addition of SwHalB (30 μ M final concentration) and allowed to proceed for 1 h at room temperature before quenching in 2 vol of Methanol with 1% (v/v) formic acid. Samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above.

Lysine cyclodeaminase reactions. Reactions (25 μ L) contained L-lysine · HCl (1 mM), sodium α KG (5 mM), sodium ascorbate (100 μ M), (NH₄)₂Fe(SO₄)₂ · 6H₂O (50 μ M), and sodium chloride (1 mM) in 50 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of SwHalB (15 μ M final concentration) and allowed to proceed for 2 h at room temperature to form chlorinated lysine substrates for the lysine cyclodeaminase (RapL). Cyclodeamination was initiated by addition of 1 volume (25 μ L) of cyclodeaminase solution (NAD⁺ (1.8 mM) and RapL (14 μ M) in 50 mM HEPES buffer pH 7.5) and allowed to proceed for 2 h at room temperature before quenching in 2 vol of Methanol with 1% (v/v) formic acid. Samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above.

Ornithine cyclodeaminase reactions. Reactions (25 μ L) contained L-ornithine · HCl (1 mM), sodium α KG (5 mM), sodium ascorbate (100 μ M), (NH₄)₂Fe(SO₄)₂ · 6H₂O (50 μ M), and sodium chloride (1 mM) in 50 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of PkHalD (15 μ M final concentration) and allowed to proceed for 2 h at room temperature to form chlorinated ornithine substrate for the ornithine cyclodeaminase (OCD). Cyclodeamination was

initiated by addition of 1 volume (25 μ L) of cyclodeaminase solution (NAD^+ (1.8 mM) and OCD (20 μ M) in 50 mM HEPES buffer pH 7.5) and allowed to proceed for 2 h at room temperature before quenching in 2 vol of Methanol with 1% (v/v) formic acid. Samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above.

Lysine decarboxylase reactions. Reactions (50 μ L) contained L-lysine $\cdot$ HCl (1 mM), sodium α KG (5 mM), sodium ascorbate (200 μ M), $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ (100 μ M), and sodium chloride (1 mM) in 50 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of SwHalB (30 μ M final concentration) and allowed to proceed for 2 h at room temperature to form chlorinated lysine substrates for the lysine decarboxylase (LDC). Decarboxylation was initiated by addition of 15 μ L of lysine decarboxylase (LDC, 48 μ M) and allowed to proceed for 2 h at room temperature before quenching in 2 vol of methanol with 1% (v/v) formic acid. Samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above.

Aliphatic amino acid transaminase reactions. Reactions (50 μ L) contained DL-norleucine (1 mM), sodium α KG (5 mM), sodium ascorbate (200 μ M), $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ (100 μ M), and sodium chloride (1 mM) in 50 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of SrHalE (10 μ M final concentration) and allowed to proceed for 1.5 h at room temperature to form chlorinated norleucine substrates for the aliphatic amino acid aminotransferase (IlvE). Transamination was initiated by addition of 15 μ L of IlvE (370 μ M) and allowed to proceed for 1 h at room temperature before quenching in 2 vol of Methanol with 1% (v/v) formic acid. Samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF (negative ion mode) using a Phenomenex (Torrence, CA) Rezex-ROA Organic Acid H^+ column (150 $\times$ 4.6 mM) and Carbo- H^+ Security Guard cartridge. In the isocratic method, the sample was eluted over the course of 10 min with 0.5% (v/v) formic acid.

***In vitro* transcription/translation (IVTT) of a chlorinated peptide.** The IVTT reactions were performed using the PURExpress Δ (aa, tRNA) kit (NEB), which separates the amino acids and tRNA from the other reaction components. Initial halogenation reactions (4.2 μ L) contained the 20 canonical amino acids (1 mM), sodium α KG (1 mM), sodium ascorbate (167 μ M), $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ (167 μ M), and 0.5 mM NaCl in 33 mM HEPES buffer pH 7.5. Halogenation was initiated by addition of PfHalA or SwHalB (83 μ M final concentration) and allowed to proceed for 40 min at room temperature. IVTT was then initiated by addition of the remaining PURExpress kit components along with the plasmid encoding PepC (7 ng/ μ L final concentration), resulting in a $\sim$ 3-fold dilution of the initial halogenation reaction components. The halogenases were omitted for control reactions as noted in the data. The final reactions (14 μ L) were moved to 37°C for 4 h. Following dilution with 25 μ L of 10 mM magnesium acetate, protein components were removed from the reaction by passage through a Pall Nanosep 10 kDa spin column. The flow-through was analyzed using an Agilent 1290 UPLC on a Poroshell 120 SB-Aq column (2.7 μ m, 2.1 $\times$ 50 mm; Agilent) using a linear gradient from 0 to 100% acetonitrile over 5 min at a flow rate of 0.6 mL/min with 0.1% (v/v) formic acid as the mobile phase. Mass spectra were acquired using an Agilent 6530 QTOF.

General method for HPLC purification of chlorinated amino acids for NMR analysis. Methyl-esterified chlorinated amino acids were purified using an Agilent 1200 HPLC on a SeQuant ZIC-HILIC Semi-Preparative column (5 μ m, 200 Å, 150 $\times$ 21.1 mm, EMD Millipore)

using the following buffers: Buffer A (90% acetonitrile, 10% water, 10 mM ammonium formate) and Buffer B (90% water, 10% acetonitrile, 10 mM ammonium formate). A gradient was applied as follows: linear gradient from 100% to 40% Buffer A over 1 h, followed by a hold of 40% Buffer A for 18 min, then to 30% Buffer A over 5 min, followed by a linear gradient to 75% A over 4 min at a flow rate of 5 mL/min. The fractions containing the desired chlorinated compounds were determined by LC-QTOF MS screening, pooled, and dried by rotary evaporation followed by speed-vacuum when the volume reached <1.5 mL.

General methods for NMR data collection. All experiments were recorded on a Bruker Avance II spectrometer operating at 900 MHz and at 298 K. The instrument was equipped with a CP TXI cryoprobe and was controlled using Topspin (version 3.2) software. Data were processed using Topspin (version 3.2) by zero-filling once in each dimension, followed by apodization, Fourier transformation, and phasing, and were referenced to 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). Spectra were analyzed using Mnova software (Mestrelab Research, Escondido, CA, USA).

Preparation, purification, and NMR analysis of [$^{13}\text{C}_6$, $^{15}\text{N}_2$]-5-Cl-lysine methyl ester. [$^{13}\text{C}_6$, $^{15}\text{N}_2$]-5-Cl-lysine methyl ester was prepared in a 200 μL reaction containing HalB from *S. iranensis* (SiHalB, 70 μM), [$^{13}\text{C}_6$, $^{15}\text{N}_2$]-L-lysine $\cdot$ 2HCl (0.35 mM), αKG (5 mM), $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ (1 mM), sodium ascorbate (2.5 mM) in 50 mM HEPES (pH 7.5). Chloride is provided by the enzyme storage buffer (NaCl) and the lysine hydrochloride substrate. The reaction was quenched after 60 min by transferring into 3 mL of 3 M HCl in methanol followed by incubation at 50°C for 3 h. Samples were then cooled at room temperature for 1 h before neutralizing by dropwise addition of 10 M NaOH. Precipitate was removed by centrifugation at $10,000 \times g$ for 10 min. The [$^{13}\text{C}_6$, $^{15}\text{N}_2$]-5-Cl-lysine methyl ester was purified using an Agilent 1200 HPLC on a SeQuant ZIC-HILIC Semi-Preparative column as described above. Samples were resuspended in 60% D_3 -acetonitrile:40% D_2O for NMR analysis on a Bruker 900 MHz instrument and for LC/QTOF-MS analysis. High resolution ESI-MS $[\text{M}+\text{H}]^+$: calculated for $\text{C}^{13}\text{C}_6\text{H}_{16}^{35}\text{Cl}^{15}\text{N}_2\text{O}_2^+$ m/z , 203.1037, found m/z 203.1034. $\text{C}^{13}\text{C}_6\text{H}_{16}^{37}\text{Cl}^{15}\text{N}_2\text{O}_2^+$ m/z , 205.1007, found m/z 205.1006.

A constant-time ^1H - ^{13}C HSQC experiment (Bruker pulse sequence hsqcctetgsp) was recorded with carrier frequencies set to 5.48 ppm (^1H) and 40 ppm (^{13}C) and the spectral widths set to 16 ppm (^1H) and 80 ppm (^{13}C). The carboxyl chemical shift was set to 176 ppm, and the ^1H - ^{13}C coupling constant was set to 145 Hz. The recycle delay was set to 1.5 sec. The constant-time evolution period was set to 13.6 ms, to distinguish ^1H - ^{13}C pairs with one or three carbon neighbors from those with two carbon neighbors. A 2D ^1H - ^{13}C version of the 3D HCCH-COSY experiment (Bruker pulse sequence hcchcogp3d2) was used to identify neighboring ^1H and ^{13}C resonances. The carrier frequencies were set to 5.48 ppm (^1H) and 39 ppm (^{13}C) with spectral widths of 14 ppm (^1H) and 75 ppm (^{13}C). The recycle delay was set to 1.8 sec. Please note that the methyl ester is formed by quenching with unlabeled methanol and does not appear using these data collection parameters. ^1H NMR (900 MHz, 60% D_3 -acetonitrile:40% D_2O): δ 4.30 (H_δ), 4.19 (H_α), 3.47 ($\text{H}_{\epsilon 1}$), 3.24 ($\text{H}_{\epsilon 2}$), 2.25 ($\text{H}_{\beta 1}$), 2.16 ($\text{H}_{\beta 2}$), 2.11 ($\text{H}_{\gamma 1}$), 1.86 ($\text{H}_{\gamma 2}$). ^{13}C NMR (900 MHz, 60% D_3 -acetonitrile:40% D_2O): δ 58.91 (C_δ), 53.11 (C_α), 46.10 (C_ϵ), 31.22 (C_γ), 27.31 (C_β).

Preparation, purification, and NMR analysis of [$^{13}\text{C}_6$, $^{15}\text{N}_2$]-5,5-dichlorolysine methyl ester. [$^{13}\text{C}_6$, $^{15}\text{N}_2$]-5,5-dichlorolysine methyl ester was prepared in a 400 μL reaction containing HalB from *S. wuyuanensis* (SwHalB, 70 μM), [$^{13}\text{C}_6$, $^{15}\text{N}_2$]-L-lysine $\cdot$ 2HCl (0.35 mM), αKG (5 mM), $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ (1 mM), sodium ascorbate (2.5 mM) in 50 mM HEPES (pH 7.5). Chloride is provided by the enzyme storage buffer (NaCl) and the lysine hydrochloride substrate.

The reaction was quenched after 60 min by transferring into 3 mL of 3 M HCl in methanol followed by incubation at 50°C for 3 h. Samples were then cooled at room temperature for 1 h before neutralizing by dropwise addition of 10 M NaOH. Precipitate was removed by centrifugation at $10,000 \times g$ for 10 min. The $[^{13}\text{C}_6, ^{15}\text{N}_2]$ -5,5-dichlorolysine methyl ester was purified using an Agilent 1200 HPLC on a SeQuant ZIC-HILIC Semi-Preparative column as described above. Samples were resuspended in D_2O for NMR analysis on a Bruker 900 MHz instrument and for LC/QTOF-MS analysis. High resolution ESI-MS $[\text{M}+\text{H}]^+$: calculated for $\text{C}^{13}\text{C}_6\text{H}_{15}^{35}\text{Cl}_2^{15}\text{N}_2\text{O}_2^+$ m/z , 237.0647, found m/z 237.0647. $\text{C}^{13}\text{C}_6\text{H}_{15}^{35}\text{Cl}^{37}\text{Cl}^{15}\text{N}_2\text{O}_2^+$ m/z , 239.0618, found m/z 239.0618. $\text{C}^{13}\text{C}_6\text{H}_{15}^{37}\text{Cl}_2^{15}\text{N}_2\text{O}_2^+$ m/z , 241.0588, found m/z 241.0587.

A constant-time ^1H - ^{13}C HSQC spectrum was recorded with carrier frequencies set to 4.7 (^1H) and 40 (^{13}C) ppm, and spectral widths set to 16 ppm (^1H) and 80 ppm (^{13}C). The recycle delay was set to 1.5 sec. A HCCH-COSY spectrum was recorded with carrier frequencies set to 4.7 ppm (^1H) and 39 ppm (^{13}C) and spectral widths of 14 ppm (^1H) and 75 ppm (^{13}C). HCCH-COSY spectra depend on neighboring ^1H - ^{13}C pairs to resolve resonances. The experiment fails when a carbon neighbor has no attached protons, as was the case for the δ position of 5,5-dichlorolysine. To confirm the resonance of C_δ , a 2D ^1H - ^{13}C long-range HCCH experiment (Bruker pulse sequence `hcchctgplr`) was used. The carrier frequencies were set to 4.7 (^1H) and 50 ppm (^{13}C), and the spectral widths were set to 14 ppm (^1H) and 140 ppm (^{13}C). The recycle delay was 2.0 sec. Please note that the methyl ester is formed by quenching with unlabeled methanol and does not appear using these data collection parameters. ^1H NMR (900 MHz, D_2O): δ 4.32 (H_α), 3.82 (H_ϵ), 2.60 ($\text{H}_{\gamma 1}$), 2.50 ($\text{H}_{\gamma 2}$), 2.50 ($\text{H}_{\beta 1}$), 2.44 ($\text{H}_{\beta 2}$). ^{13}C NMR (900 MHz, D_2O): δ 88.84 (C_δ), 53.98 (C_ϵ), 53.33 (C_α), 41.66 (C_γ), 26.56 (C_β).

Preparation, purification, and NMR analysis of $[^{13}\text{C}_6, ^{15}\text{N}_2]$ -4,4-dichlorolysine methyl ester. $[^{13}\text{C}_6, ^{15}\text{N}_2]$ - 4,4-dichlorolysine methyl ester was prepared in a 400 μL reaction containing HalC from *L. anisa* (LaHalC, 70 μM), $[^{13}\text{C}_6, ^{15}\text{N}_2]$ - L-lysine $\cdot$ 2HCl (1 mM), αKG (5 mM), $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ (1 mM), sodium ascorbate (2.5 mM) in 50 mM HEPES (pH 7.5). Chloride is provided by the enzyme storage buffer (NaCl) and the lysine hydrochloride substrate. The reaction was quenched after 4 h by transferring into 3 mL of 3 M HCl in methanol followed by incubation at 50°C for 3 h. Samples were then cooled at room temperature for 1 h before neutralizing by dropwise addition of 10 M NaOH. Precipitate was removed by centrifugation at $10,000 \times g$ for 10 min. The $[^{13}\text{C}_6, ^{15}\text{N}_2]$ -4,4-dichlorolysine methyl ester was purified using an Agilent 1200 HPLC on a SeQuant ZIC-HILIC Semi-Preparative column as described above. Samples were resuspended in D-methanol for NMR analysis on a Bruker 900 MHz instrument and for LC/QTOF-MS analysis. High resolution ESI-MS $[\text{M}+\text{H}]^+$: calculated for $\text{C}^{13}\text{C}_6\text{H}_{15}^{35}\text{Cl}_2^{15}\text{N}_2\text{O}_2^+$ m/z , 237.0647, found m/z 237.0651. $\text{C}^{13}\text{C}_6\text{H}_{15}^{35}\text{Cl}^{37}\text{Cl}^{15}\text{N}_2\text{O}_2^+$ m/z , 239.0618, found m/z 239.0621. $\text{C}^{13}\text{C}_6\text{H}_{15}^{37}\text{Cl}_2^{15}\text{N}_2\text{O}_2^+$ m/z , 241.0588, found m/z 241.0591.

A constant-time HSQC experiment was recorded with carrier frequencies of 4.7 ppm (^1H) and 40 ppm (^{13}C), and with spectral widths of 16 ppm (^1H) and 80 ppm (^{13}C). The recycle delay was set to 2 seconds. A HCCH-COSY experiment was recorded with carrier frequencies of 4.7 ppm (^1H) and 39 ppm (^{13}C), and spectral widths of 14 ppm (^1H) and 75 ppm (^{13}C). Because C_γ lacked protons, a 1,1 Adequate experiment was used to assign the shift of the C_γ carbon via correlation between H_δ and C_γ . The experiment was recorded with carrier frequencies of 4.67 ppm (^1H) and 105 ppm (^{13}C), and with spectral widths of 16 ppm (^1H) and 200 ppm (^{13}C). The recycle delay was set to 1.5 sec. A 2D ^1H - ^{13}C plane of a 3D HCACO experiment (Bruker pulse sequence `hcacogp3d`) was used to unambiguously distinguish the H_α resonance from the other signals. The

experiment was recorded with carrier frequencies set to 4.7 ppm (^1H) and 50 ppm (^{13}C), and spectral widths set to 16 ppm (^1H) and 60 ppm (^{13}C). The recycle delay was set to 1.5 sec. ^1H NMR (900 MHz, methanol- D): δ 3.86 (H_α), 3.33 (H_ϵ), 2.96 ($\text{H}_{\beta 1}$), 2.79 ($\text{H}_{\delta 1}$), 2.68 ($\text{H}_{\delta 2}$), 2.51 ($\text{H}_{\beta 2}$). ^{13}C NMR (900 MHz, methanol- D): δ 82.72 (C_γ), 53.12 (C_α), 52.83 (C_β), 46.01 (C_δ), 37.32 (C_ϵ).

Preparation, purification, and NMR analysis of [$^{13}\text{C}_5$, $^{15}\text{N}_2$]-4-Cl-ornithine methyl ester. [$^{13}\text{C}_5$, $^{15}\text{N}_2$]-4-Cl-ornithine methyl ester was prepared in a 400 μL reaction containing HalD from *P. kilonensis* (PkJalD, 70 μM), [$^{13}\text{C}_5$, $^{15}\text{N}_2$]-L-ornithine $\cdot$ HCl (0.5 mM), αKG (5 mM), $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ (1 mM), sodium ascorbate (2.5 mM) in 50 mM HEPES (pH 7.5)). Chloride is provided by the enzyme storage buffer (NaCl) and the ornithine hydrochloride substrate. The reaction was quenched after 60 min by transferring into 3 mL of 3 M HCl in methanol followed by incubation at 50°C for 3 h. Samples were then cooled at room temperature for 1 h before neutralizing by dropwise addition of 10 M NaOH. Precipitate was removed by centrifugation at $10,000 \times g$ for 10 min. The [$^{13}\text{C}_5$, $^{15}\text{N}_2$]-4-Cl-ornithine methyl ester was purified using an Agilent 1200 HPLC on a SeQuant ZIC-HILIC Semi-Preparative column as described above. Samples were resuspended in D_2O for NMR analysis on a Bruker 900 MHz instrument and for LC/QTOF-MS analysis. High resolution ESI-MS [$\text{M}+\text{H}$] $^+$: calculated for $\text{C}^{13}\text{C}_5\text{H}_{14}^{35}\text{Cl}^{15}\text{N}_2\text{O}_2^+$ m/z , 188.0847, found m/z 188.0847. $\text{C}^{13}\text{C}_5\text{H}_{14}^{37}\text{Cl}^{15}\text{N}_2\text{O}_2^+$ m/z , 190.0817, found m/z 190.0817.

A constant-time HSQC experiment was recorded with carrier frequencies of 4.7 ppm (^1H) and 40 ppm (^{13}C) and spectral widths of 16 ppm (^1H) and 80 ppm (^{13}C). The recycle delay was set to 1.5 sec. A HCCH-COSY spectrum was recorded with carrier frequencies set to 4.7 ppm (^1H) and 39 ppm (^{13}C), and spectral widths of 14 ppm (^1H) and 80 ppm (^{13}C). The recycle delay was set to 2 sec. Please note that the methyl ester is formed by quenching with unlabeled methanol and does not appear using these data collection parameters. ^1H NMR (900 MHz, D_2O): δ 4.69 (H_γ), 4.24 (H_α), 3.77 ($\text{H}_{\delta 1}$), 3.50 ($\text{H}_{\delta 2}$), 2.55 ($\text{H}_{\beta 1}$), 2.29 ($\text{H}_{\beta 2}$). ^{13}C NMR (900 MHz, D_2O): δ 51.5 (C_γ), 48.8 (C_δ), 46.1 (C_α), 32.3 (C_β).

Preparation, purification, and NMR analysis of [$^{13}\text{C}_6$]-4-Cl-leucine methyl ester. [$^{13}\text{C}_6$]-4-Cl-leucine methyl ester was prepared in a 400 μL reaction containing MBP-*P. Root562* HalE (80 μM), [$^{13}\text{C}_6$] L-leucine (1 mM), αKG (2 mM), $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ (0.25 mM), sodium ascorbate (2 mM), sodium chloride (5 mM), and 40% glycerol (v/v) in 30 mM sodium acetate (pH 6.0). The reaction was quenched after 90 min by passage through a 10 kDa MWCO Pall Nanosep spin filter to remove protein. The flow-through was dried using a speed vacuum, resuspended in 200 μL of 3 M HCl in methanol, and incubated at 50°C for 3 h. The methyl-esterified samples were then dried using a speed vacuum and resuspended in 2 mL of 90% acetonitrile:10% water. Precipitate was removed by centrifugation at $10,000 \times g$ for 10 min. The [$^{13}\text{C}_6$]-4-Cl-leucine methyl ester was purified using an Agilent 1200 HPLC on a SeQuant ZIC-HILIC Semi-Preparative column as described above. Fractions containing the compound of interest were dried by speed vacuum, resuspended with 100 μL of saturated ammonium bicarbonate in D_2O , and extracted 3 times with 100 μL chloroform (D, 99.8%) for NMR analysis on a Bruker 900 MHz instrument and for LC/QTOF-MS analysis. High resolution ESI-MS [$\text{M}+\text{H}$] $^+$: calculated for $\text{C}^{13}\text{C}_6\text{H}_{15}^{35}\text{ClNO}_2^+$ m/z , 186.0987, found m/z 186.0988. $\text{C}^{13}\text{C}_6\text{H}_{15}^{37}\text{ClNO}_2^+$ m/z , 188.0958, found m/z 188.0958.

A constant-time HSQC was recorded with carrier frequencies of 4.7 ppm (^1H) and 50 ppm (^{13}C), and with spectral widths of 14 ppm (^1H) and 100 ppm (^{13}C). The recycle delay was set to 2 s. A HCCH-COSY spectrum was recorded with carrier frequencies of 4.7 ppm (^1H) and 50 ppm (^{13}C), with spectral widths of 14 ppm (^1H) and 100 ppm (^{13}C). The recycle delay was set to 1.8 s. To confirm the resonance of C_γ , a 2D ^1H - ^{13}C long-range HCCH experiment that was recorded with

carrier frequencies of 5 ppm (^1H) and 40 ppm (^{13}C), and with spectral widths of 14 ppm (^1H) and 120 ppm (^{13}C). The recycle delay was set to 2 s. Please note that the methyl ester is formed by quenching with unlabeled methanol and does not appear using these data collection parameters. ^1H NMR (900 MHz, chloroform- D): δ 3.75 (H_α), 2.37 ($\text{H}_{\beta 1}$), 1.91 ($\text{H}_{\beta 2}$), 1.69 ($\text{H}_{\delta 1}$), 1.64 ($\text{H}_{\delta 2}$). ^{13}C NMR (900 MHz, chloroform- D): δ 69.47 (C_γ), 52.39 (C_α), 50.32 (C_β), 33.14 ($\text{C}_{\delta 2}$), 32.97 ($\text{C}_{\delta 1}$).

Preparation, purification, and NMR analysis of [$^{13}\text{C}_6$, ^{15}N]-4-Cl-isoleucine methyl ester. [$^{13}\text{C}_6$, ^{15}N]-4-Cl-isoleucine methyl ester was prepared in a 400 μL reaction containing *P. Root562* HalE (80 μM), [$^{13}\text{C}_6$, ^{15}N] L-isoleucine (1 mM), αKG (2 mM), $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ (0.25 mM), sodium ascorbate (2 mM), sodium chloride (5 mM), and 40% glycerol (v/v) in 30 mM sodium acetate (pH 6.0). The reaction was quenched after 90 min by passage through a 10 kDa MWCO Pall Nanosep spin filter to remove protein. The flow-through was dried using a speed vacuum, resuspended in 200 μL of 3 M HCl in methanol, and incubated at 50°C for 3 h. Samples were then dried using a speed vacuum and resuspended in 2 mL of 90% acetonitrile and 10% water. Precipitate was removed by centrifugation at $10,000 \times g$ for 10 min. The [$^{13}\text{C}_6$, ^{15}N]-4-Cl-isoleucine methyl ester was purified using an Agilent 1200 HPLC on a SeQuant ZIC-HILIC Semi-Preparative column as described above. Fractions containing the compound of interest were dried by speed vacuum, resuspended with 100 μL of saturated ammonium bicarbonate in D_2O , and extracted 3 times with 100 μL chloroform (D, 99.8%) for NMR analysis on a Bruker 900 MHz instrument and for LC/QTOF-MS analysis. High resolution ESI-MS $[\text{M}+\text{H}]^+$: calculated for $\text{C}^{13}_6\text{H}_{15}^{35}\text{Cl}^{15}\text{NO}_2^+$ m/z , 187.0957, found m/z 187.0959. $\text{C}^{13}_6\text{H}_{15}^{37}\text{Cl}^{15}\text{NO}_2^+$ m/z , 189.0928, found m/z 189.0927.

A constant-time HSQC experiment was recorded with carrier frequencies set to 4.7 ppm (^1H) and 50 ppm (^{13}C), and with spectral widths set to 14 ppm (^1H) and 100 ppm (^{13}C). The recycle delay was set to 2 s. A 2D HCCH-COSY spectrum was recorded with carrier frequencies of 4.7 ppm (^1H) and 50 ppm (^{13}C), and spectral widths of 14 ppm (^1H) and 100 ppm (^{13}C). The recycle delay was set to 2 s. Please note that the methyl ester is formed by quenching with unlabeled methanol and does not appear using these data collection parameters. [$^{13}\text{C}_6$, ^{15}N]-4-Cl-isoleucine methyl ester, ^1H NMR (900 MHz, chloroform- D): δ 4.39 ($\text{H}_{\gamma 2}$), 3.55 (H_α), 2.14 (H_β), 1.49 (H_δ), 0.99 ($\text{H}_{\gamma 1}$). ^{13}C NMR (900 MHz, chloroform- D): δ 58.91 ($\text{C}_{\gamma 2}$), 56.98 (C_α), 45.39 (C_β), 20.97 (C_δ), 12.18 ($\text{C}_{\gamma 1}$). [$^{13}\text{C}_6$, ^{15}N]-isoleucine methyl ester, ^1H NMR (900 MHz, chloroform- D): δ 3.37 (H_α), 1.75 (H_β), 1.46 ($\text{H}_{\gamma 2}$), 1.22 ($\text{H}_{\gamma 2}$), 0.95 ($\text{H}_{\gamma 1}$), 0.92 (H_δ). ^{13}C NMR (900 MHz, chloroform- D): δ 58.68 (C_α), 38.72 (C_β), 24.34 ($\text{C}_{\gamma 2}$), 15.53 ($\text{C}_{\gamma 1}$), 11.44 (C_δ).

Preparation, purification, and NMR analysis of 5,5-dichloronorleucine methyl ester (unlabeled). 5,5-dichloronorleucine methyl ester was prepared in a 500 μL reaction containing MBP-*P. Root562* HalE (20 μM), DL-norleucine (1 mM), αKG (2 mM), $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ (0.5 mM), sodium ascorbate (2 mM) in 50 mM NaAc (pH 6.0). After 16 h, protein was removed by passage through a 10 kDa MWCO Pall Nanosep spin filter. After dilution of the flow-through to 1.5 mL with acetonitrile, 5,5-dichloronorleucine was purified using an Agilent 1200 HPLC on a SeQuant ZIC-HILIC Semi-Preparative column as described above. Fractions containing 5,5-dichloronorleucine were dried using a speed vacuum, resuspended in 200 μL of 3 M HCl in methanol, and incubated at 50°C for 3 h to methyl esterify the amino acid. Samples were again vacuumed to dryness. The sample was then resuspended with 100 μL of saturated ammonium bicarbonate in D_2O and extracted 3 times with 100 μL chloroform (D, 99.8%) for NMR analysis on a Bruker 900 MHz instrument and for LC/QTOF-MS analysis. High resolution ESI-MS

[M+H]⁺: calculated for C₇H₁₄³⁵Cl₂NO₂⁺ *m/z*, 214.0396, found *m/z* 214.0398. C₇H₁₄³⁵Cl³⁷ClNO₂⁺ *m/z*, 216.0367, found *m/z* 216.0369. C₇H₁₄³⁷Cl₂NO₂⁺ *m/z*, 218.0337, found *m/z* 218.0341.

An edited ¹H-¹³C HSQC experiment (Bruker pulse sequence hsqcedetgpsisp2.2), in which carbons with one or three protons are opposite in phase to those with two protons, was recorded with carrier frequencies set to 5 ppm (¹H) and 50 ppm (¹³C), and with spectral widths of 11 ppm (¹H) and 100 ppm (¹³C). The recycle delay was 1.5 s. A ¹H-¹³C HMBC experiment (Bruker pulse sequence hmbcgpplndqf) was used to correlate protons with carbon spins two and three bonds removed via long-range proton-carbon couplings. The experiment was optimized for a long-range proton-carbon coupling of 10 Hz. Carrier frequencies were set to 5 ppm (¹H) and 50 ppm (¹³C), with spectral widths of 11 ppm (¹H) and 100 ppm (¹³C). The recycle delay was set to 1.5 s. Gradient COSY and TOCSY experiments were recorded with Bruker pulse sequences cosygpqf and dipsi2etgpsi, respectively. For each experiment, the carrier frequency was set to 5 ppm, the spectral widths were set to 11 ppm in both dimensions, and the recycle delay was set to 1.5 s. The mixing time in the TOCSY experiment was set to 100 ms. ¹H NMR (900 MHz, chloroform-D): δ 3.51 (H_α), 2.41 (H_{γ1}), 2.29 (H_{γ2}), 2.19 (H_ε), 2.16 (H_{β1}), 1.93 (H_{β2}). ¹³C NMR (900 MHz, chloroform-D): δ 89.9 (C_δ), 53.3 (C_α), 45.8 (C_γ), 32.6 (C_ε), 30.8 (C_β).

3.2. Results and discussion

Structural characterization of BesD. The Fe^{II}/ α KG-dependent halogenases are part of the large and highly diverse Cupin superfamily, which contains members that catalyze a host of different reactions, including hydroxylation, halogenation, olefin epoxidation, and stereoinversion on a broad range of substrates [39]. As such, the prediction and discovery of new activities using bioinformatic approaches has been limited by the sequence variability of family members. In this case, BesD has very low identity to the carrier protein-dependent halogenase SyrB2 (7% ID) and the indole alkaloid halogenase WelO5 (11% ID) and thus displays higher sequence similarity to predicted but uncharacterized Fe^{II}/ α KG-dependent hydroxylases ($\leq 46\%$ ID) than to other known halogenases (*Figure 3.2*). A query of the superfamily database reveals that BesD groups in a different protein family within the Cupin Superfamily compared to previously reported Fe^{II}/ α KG-dependent halogenases [40]. To gain a better mechanistic understanding of BesD, we solved the X-ray crystal structure of BesD with lysine, Fe, chlorine, and α KG bound in the active site (*Table 3.1, Figure 3.3*).

BesD was crystallized anaerobically with lysine and α KG and soaked with Fe^{II}. The X-ray structure was solved at 1.95 Å resolution by Fe-SAD phasing. The structure contains four BesD monomers in the asymmetric unit, all of which possess the β -sandwich topology characteristic of Fe/ α KG enzymes [18]. Each monomer contains a single Fe coordinated by His137, His 204, chloride, and α KG in a distorted square pyramidal geometry in the active site, which contains the HXG/A motif that is characteristic of halogenases. The binding site for lysine in BesD is largely polar, with a network of hydrogen bonds for binding to the carboxylate (Arg74, His134, Trp238), α -amine (Asp140), and ϵ -amine (Glu120, Asn219, and Thr221) of the substrate (*Figure 3.3*). In addition, Trp239 on the C-terminus of the protein stacks over the aliphatic side chain of lysine, closing over the substrate at the entrance of the active site. The structure also reveals a water network between Thr221 and the chloride ligand.

The crystal structure highlights differences in substrate binding between BesD and the carrier protein-dependent amino acid halogenases. In BesD, the positioning of the lysine substrate appears to be strongly determined by a two-point interaction between the free carboxylate of lysine and the guanidinium group of an active-site arginine (Arg74). In contrast in SyrB2, the carboxylate of threonine substrate is ligated to a phosphopantetheine (PPant) arm, which both tethers threonine to the carrier protein (SyrB1) through a thioester linkage and masks the carboxylate. It has been proposed that the Thr-PPant arm enters the active site through a tunnel and that interactions with threonine, the PPant arm, as well as SyrB1 are likely important for positioning the substrate in the active site for catalysis [41,42]. Another notable difference is that BesD appears to have a covering lid that holds lysine in the active site for catalysis. This substrate binding mode is more reminiscent of the indole alkaloid halogenase, WelO5, in which the substrate is held into place by a helix that closes over the active site [19].

Catalytic selectivity in BesD. Understanding how radical halogenases catalyze halogenation selectively over the closely related hydroxylation reaction remains an active area of investigation. Fe^{II}/ α KG-dependent halogenases are closely related both evolutionarily and mechanistically to the larger and more well-studied class of Fe^{II}/ α KG-dependent hydroxylases [43,44]. Upon activation with O₂ and α KG, both hydroxylases and halogenases generate a high-valent Fe^{IV}-oxo species which can abstract an H atom from the substrate [45]. For hydroxylases, rebound of the substrate radical with the resulting hydroxo group is the only possible pathway as the coordination site

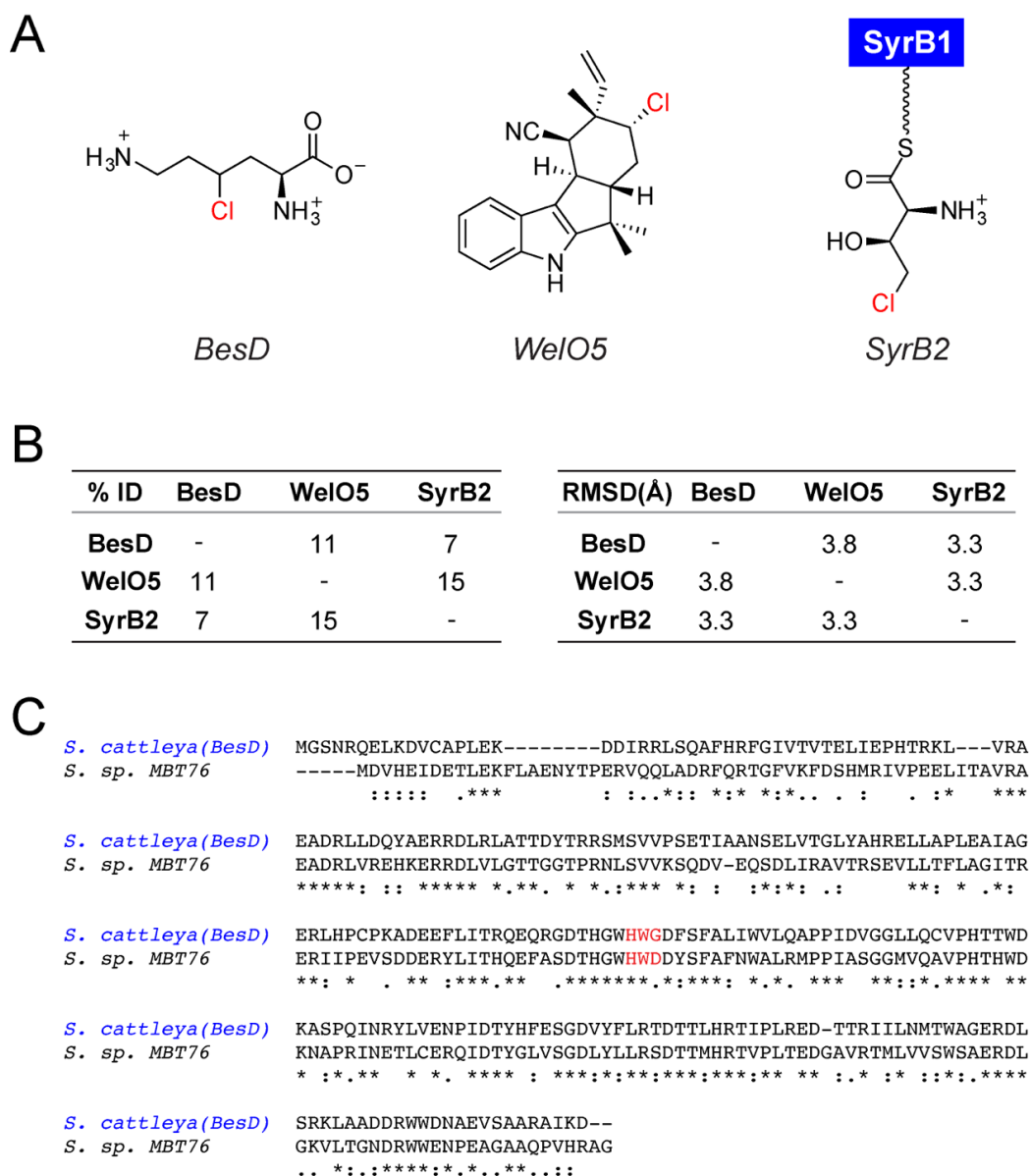


Figure 3.2. Structures of the substrates of *BesD*, *WelO5*, and *SyrB2*. (B) Sequence comparison (left) and RMSD (right) of *BesD* (PDB 6NIE), *WelO5* (PDB 5IQT), and *SyrB2* (PDB 2FCU). RMSD calculations were performed using the DALI server [46]. (C) An alignment of *BesD* to its nearest hydroxylase homolog demonstrates that *BesD* is closer in sequence similarity (46% ID) to its nearest putative hydroxylase neighbor (WP_107105619, *S. sp. MBT76*) with the characteristic HXD motif, shown in red, than to the halogenases *SyrB2* and *WelO5*. Alignment and sequence comparisons were performed with MUSCLE [25].

Table 3.1. Data collection and refinement parameters for BesD from *S. cattleya*.

	BesD Fe-bound (6NIE)
Data collection	
Space group	P1
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	61.3578, 68.9165, 79.7021
α , β , γ (°)	115.363, 91.5673, 112.951
Resolution (Å)	70.03–1.95 (2.02–1.95) *
<i>R</i> _{sym} or <i>R</i> _{merge}	0.158 (1.583)
<i>I</i> / σI	10.70 (1.79)
Completeness (%)	97.37 (95.7)
Redundancy	13.6 (11.5)
CC _{1/2}	0.998 (0.78)
Refinement	
Resolution (Å)	70.03–1.95
No. reflections	74970 (7338)
<i>R</i> _{work} / <i>R</i> _{free}	0.180/0.217
No. atoms	8788
Protein	7892
Ligand/ion	49
Water	847
<i>B</i> -factors	
Protein	42.46
Ligand/ion	33.26
Fe	29
Cl	26
Lysine	30
α KG	35
Water	43.66
R.m.s. deviations	
Bond lengths (Å)	0.01
Bond angles (°)	1.29

*Data were collected on one crystal. *Values in parentheses are for highest-resolution shell.

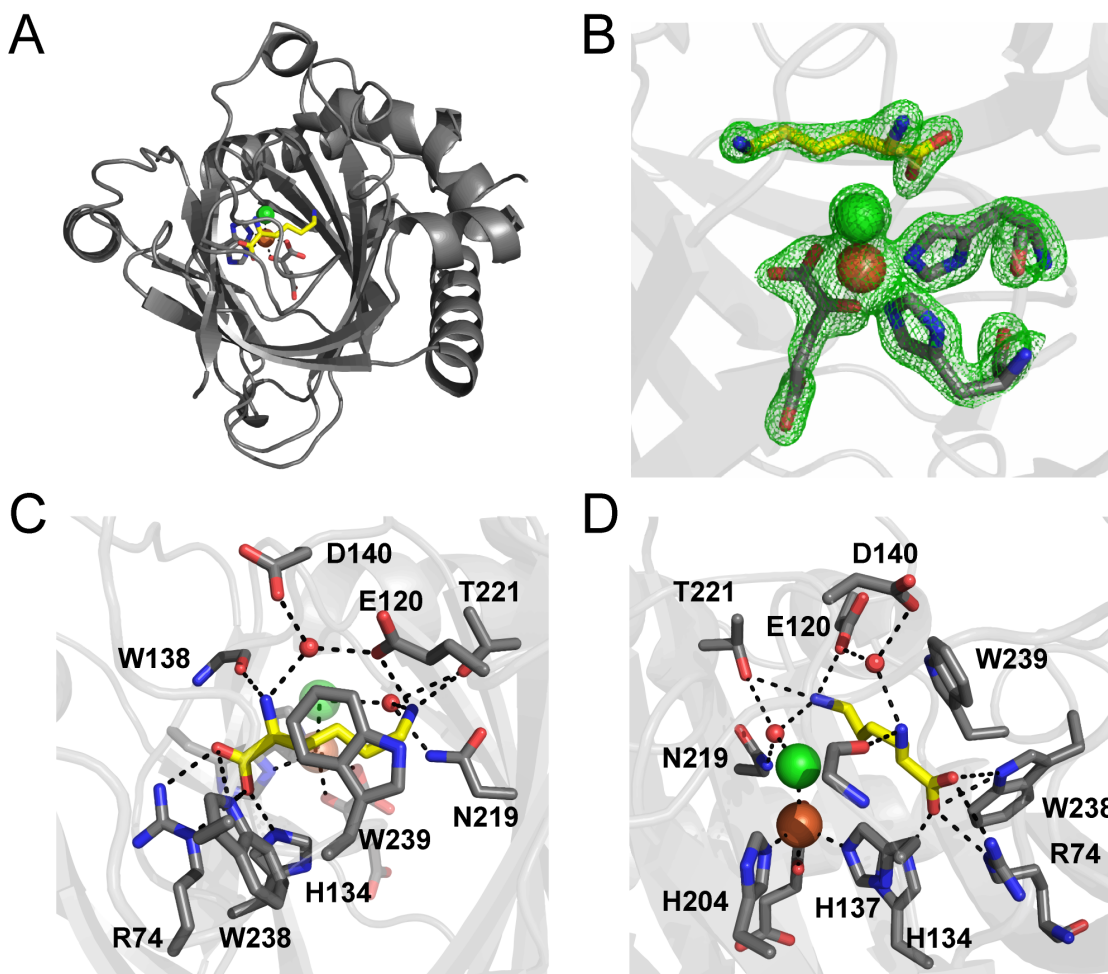


Figure 3.3. (A) Overall fold of substrate-bound BesD. (B) $F_o - F_c$ omit map for the lysine (yellow) and Fe-bound complex with two histidines, Cl (green), Fe (brown), and α KG. Omit map was calculated by removing ligands before refinement and is contoured at $\pm 3\sigma$. (c, d) Binding site of the substrate lysine. Hydrogen bonding interactions coordinate the carboxylate and two amines of lysine. Trp239 stacks in front of the aliphatic sidechain of lysine to cover the substrate in the active site.

occupied by a halide in halogenases is filled by the Asp or Glu ligand from the HXD/E motif. In halogenases, however, rebound with either the halogen or the hydroxyl group is possible, leading to potential reaction partitioning between halogenation and hydroxylation, respectively (*Figure 3.4*) [47]. In the case of BesD, mutation of the active site glycine of the HXG motif to an aspartate eliminates halogenation and results in only hydroxylation [9]. Indeed, it has been shown that halogenases can carry out a low level of off-pathway hydroxylation, especially when challenged with non-native substrates [48]. Efforts to engineer halogenases from hydroxylases have also been difficult, suggesting that second-sphere interactions are important in controlling reaction partitioning [49].

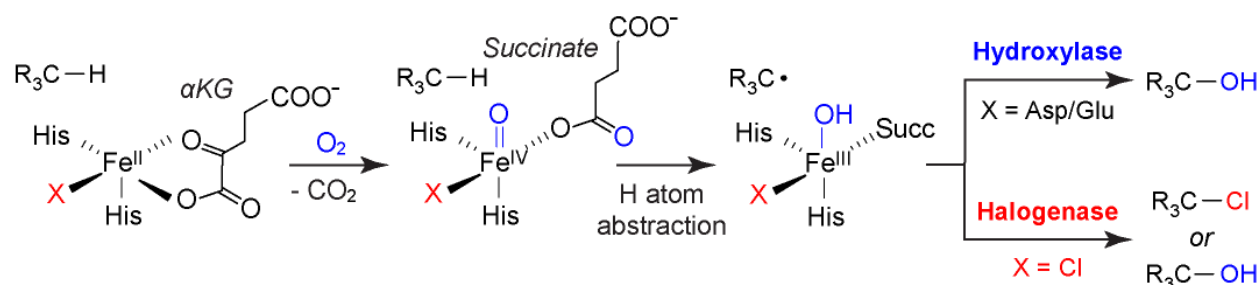


Figure 3.4. Radical halogenases are related in reaction mechanism to radical hydroxylases. In halogenases, Asp or Glu is instead replaced with a Gly or Ala residue, allowing room for direct coordination of chloride to the Fe. In this case, rebound of either the -Cl or the -OH radical is possible, leading to halogenated or hydroxylated products, respectively.

Combining insights from our structure with prior studies of $\text{Fe}^{\text{II}}/\alpha\text{KG}$ -dependent enzymes [17], we can propose a mechanism for halogenation by BesD. Halogenases must orient the Fe^{IV} -oxo intermediate such that it is suitably positioned to perform hydrogen atom abstraction while disfavoring rebound of the resulting Fe^{III} -OH intermediate in favor of Fe^{III} -Cl rebound. Our structure captured under anaerobic conditions most likely represents a snapshot of the active site prior to O_2 binding. By analogy to other $\text{Fe}^{\text{II}}/\alpha\text{KG}$ -dependent enzymes in the literature [18,19], the putative vacant site for oxygen binding appears to be *trans* to His137, on the opposite side of the His204-Cl- αKG plane from the substrate carbon that is targeted for hydrogen atom abstraction (*Figure 3.5*). If the oxo moiety remains *trans* to His 137 upon formation of the Fe^{IV} -oxo intermediate, and if the substrate position does not change, the distance between the oxo moiety and site of hydrogen atom abstraction (C_4 of lysine) would be ~ 5.0 Å on the basis of the Fe^{II} - $\text{C}_4(\text{lysine})$ distance (4.8 Å) and estimated Fe^{IV} -oxo bond length (1.6 Å) [45]. While the precise structures of the reaction intermediates are unknown, this distance is longer than that calculated from density-functional theory studies on SyrB2 [45]. It is also possible that conformational change of the substrate or a shift [19,50] of the oxo ligand toward the position *trans* to His204 before, during, or after decarboxylation of αKG may be needed to facilitate hydrogen atom abstraction to bring the substrate and Fe-center geometries within the required range [45]. Finally, rebound of a suitably positioned halogen would yield the chlorinated product.

In order to further investigate catalytic selectivity in BesD, we performed alanine scan mutagenesis of active site residues in BesD. Protein variants were expressed, purified, and incubated with lysine before product analysis by LC/MS (*Figure 3.6*). Most alanine variants

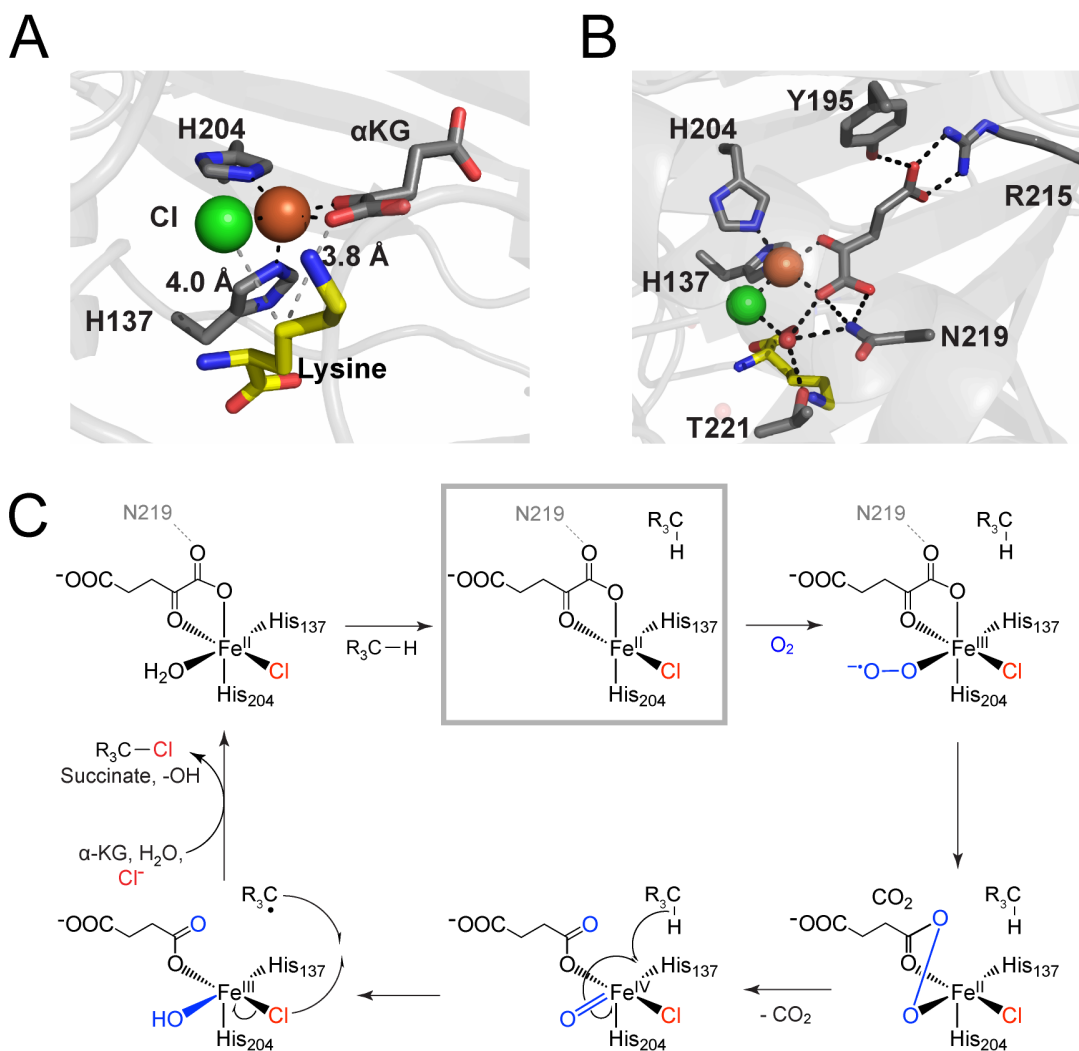


Figure 3.5. (A) The ligand complex displays a vacant site for oxygen binding that is distal to the substrate. The target γ -carbon of lysine is 4.0 Å from Cl and 3.8 Å from α KG (indicated with grey dotted lines). (B) Top view of Fe site showing α KG contacts and hydrogen bonding between the C1-carboxylate of α KG and Asn219. (C) Proposed mechanism of halogenation by BesD. The Fe is coordinated by His137, His204, chloride, and α KG in a distorted square pyramidal geometry (grey box). Based on the crystal structure in this study, the putative vacant site for oxygen binding appears to be *trans* to His137, on the opposite side of the His204-Cl- α KG plane from the substrate carbon that is targeted for hydrogen atom abstraction. In this class of enzymes, binding of the substrate followed by O₂ leads to decarboxylation of α KG via attack of the distal oxygen on C2 of α KG. It is possible that conformational change of the substrate or shift of the oxo toward the position *trans* to His204 before, during, or after decarboxylation of α KG may be needed to facilitate hydrogen atom abstraction. Finally, rebound of the suitably positioned chloride would yield the halogenated product. Note that the precise positions of the oxo species in the proposed intermediates are unknown.

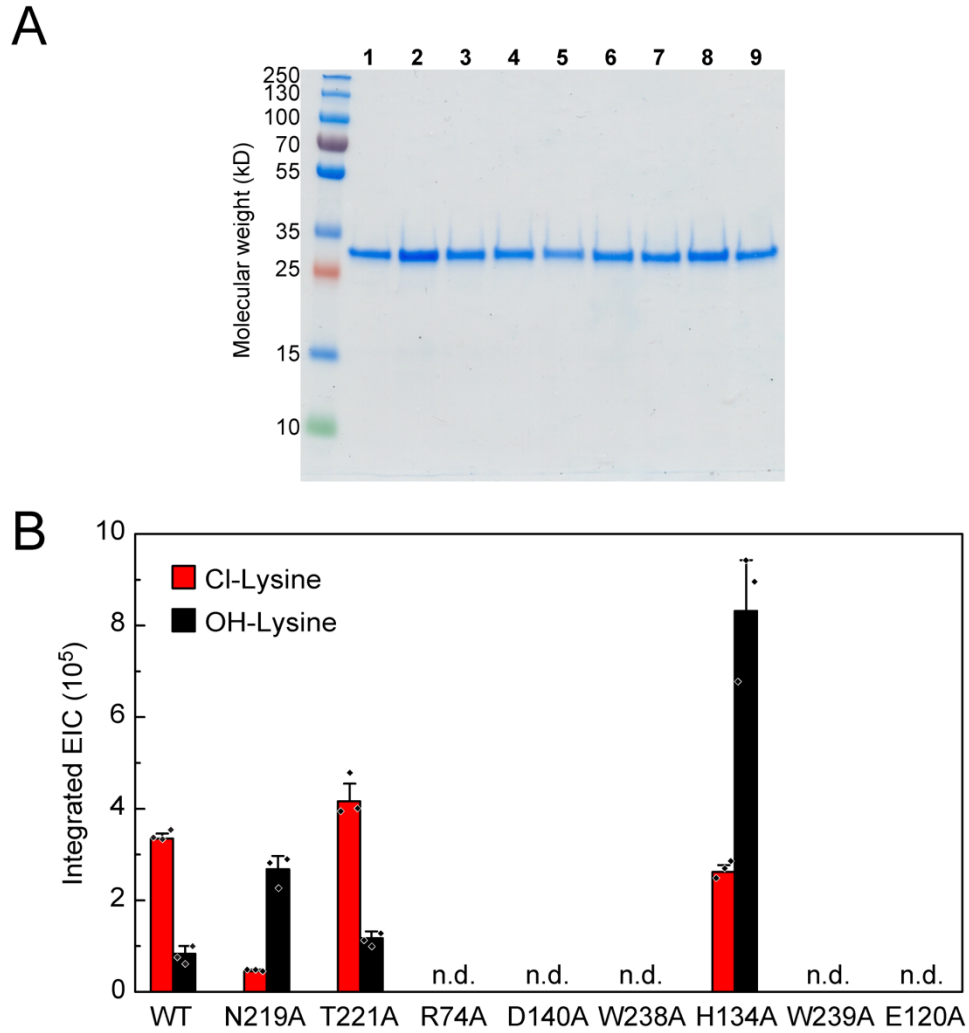


Figure 3.6. (A) SDS-PAGE of purified WT BesD and alanine mutants from *Streptomyces lavanduligriseus*. Lane 1, WT; 2, N219A; 3, T221A; 4, R74A; 5, D140A; 6, W238A; 7, H134A; 8, W239A; 9, E120A. (B) LCMS analysis of chlorolysine and hydroxyllysine production by BesD and mutants. Samples were analyzed by LCMS/QTOF and extracted ion counts for Cl-lysine ($m/z = 181.0738$) and hydroxyllysine ($m/z = 163.1077$) were integrated for each sample. Data are mean $\pm$ sd ($n = 3$ technical replicates).

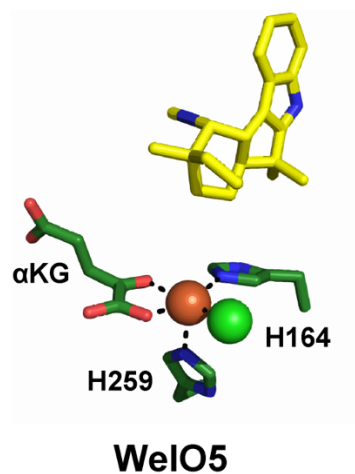
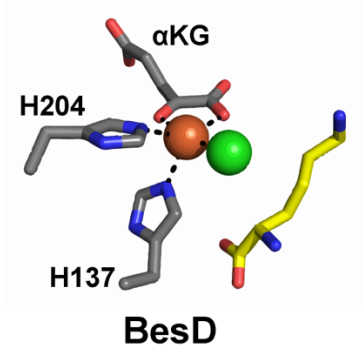
abolished all activity (both halogenation and hydroxylation). However, three variants yielded interesting profiles. Strikingly, H134A increased overall enzyme activity and abolished enzyme selectivity, resulting in decreased halogenation:hydroxylation ratios. The N219A variant also abolished selectivity for halogenation, albeit with reduced activity compared to WT BesD. Interestingly, the T221A mutation had no effect on activity or halogenation selectivity despite the conservation of a polar amino acid at position 221 in halogenases.

Analysis of the BesD crystal structure in conjunction with sequence conservation of these residues suggests that His134 and Asn219 play important roles in the second-sphere interactions that control reaction partitioning. His134 is hydrogen-bonded to the carboxylate of lysine in the BesD structure, plays an important role in orienting lysine relative to the iron complex, and is also highly conserved. The loss in halogenation selectivity upon mutation of His134 is consistent with studies that implicate the precise positioning of the substrate within the active site to favor chloride rebound over hydroxide rebound [48,51]. At 2.9 Å away from α KG, Asn219 may serve to orient α KG through hydrogen bonding such that the vacant putative O₂ binding site is located far from the substrate. Additionally, Asn219 could be involved in orienting the Fe^{III}-OH intermediate to differentially favor chloride rebound, analogous to the proposed role of Ser189 in WelO5, if the oxo does shift towards the substrate (*Figure 3.7*) [19]. Upon examining the sequence conservation of key active site residues in predicted halogenase and hydroxylase sequences with homology to BesD, we find that Asn219 is highly conserved in the halogenase homologs while greater variability is displayed by hydroxylase homologs (*Figure 3.8*). This conservation of Asn219 further supports the significance of second-sphere interactions in the BesD amino acid halogenase family.

Discovery of new amino acid halogenases. To explore the sequence space around BesD, we performed a sequence-based homology search (BLAST E-value of e^{-5}) to identify other amino acid halogenase candidates. The related hydroxylases were filtered from the data set based on the presence of the characteristic HXD/E motif, as halogenases are known to possess an HXG/A motif instead to allow for halide coordination to Fe during catalysis [18]. To maximize sampling of functional diversity, we generated a Sequence Similarity Network (SSN) [28] to group the homologous enzymes into eight new putative isofunctional groups, which we termed Clusters A-H (*Figure 3.9-3.10*). Genes adjacent to putative halogenase coding sequences were analyzed with the Enzyme Function Initiative's Genome Neighborhood Tools in order to identify conserved genomic contexts associated with these enzymes [28]. This analysis revealed numerous genes encoding proteins predicted to be involved in amino acid metabolism, such as amino acid transporters, LysR family transcriptional regulators, serine hydroxymethyl transferases, and ATP-Grasp dependent carboxy-amine ligases (*Figure 3.11*). The conservation of genes encoding proteins associated with amino acid metabolism suggested that the BesD homologs might also be amino acid halogenases. We cloned, heterologously expressed, and purified representative halogenase homologs from Clusters A-H and tested the proteins for activity against a panel of amino acids (*Figure 3.12*). The products were analyzed by LC/MS, revealing newly accessible chlorinated amino acid products and identifying substrates for enzymes within the clusters (*Figure 3.13, Table 3.2-3.4*).

Among the characterized halogenases, we discovered members with different regioselectivities. BesD and enzymes from Cluster A halogenate lysine at the γ carbon to produce 4-Cl-lysine. However, HalB from *Streptomyces iranensis* (SiHalB) yielded a product with the same

A



B

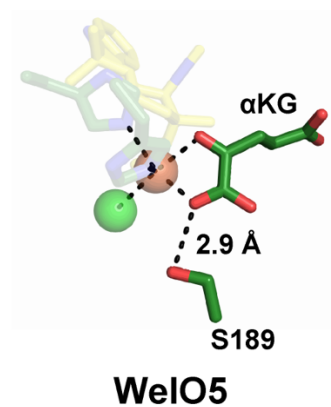
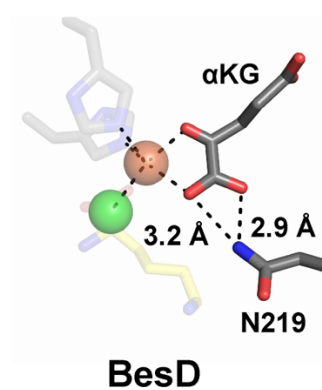


Figure 3.7. O₂ binding site for BesD vs. WelO5. (A) Comparison of BesD (grey) and WelO5 (dark green) active site suggests that in the case of BesD, the putative O₂ binding site is on the opposite side of the α KG-C1-H204 plane from the substrate (yellow). However, in WelO5, the putative O₂ binding site is on the same side as the substrate (yellow). (B) Asn219 in BesD provides hydrogen bonding contacts to the C1 carboxylate of α KG and possibly serves an analogous mechanistic role to Ser189 in WelO5.

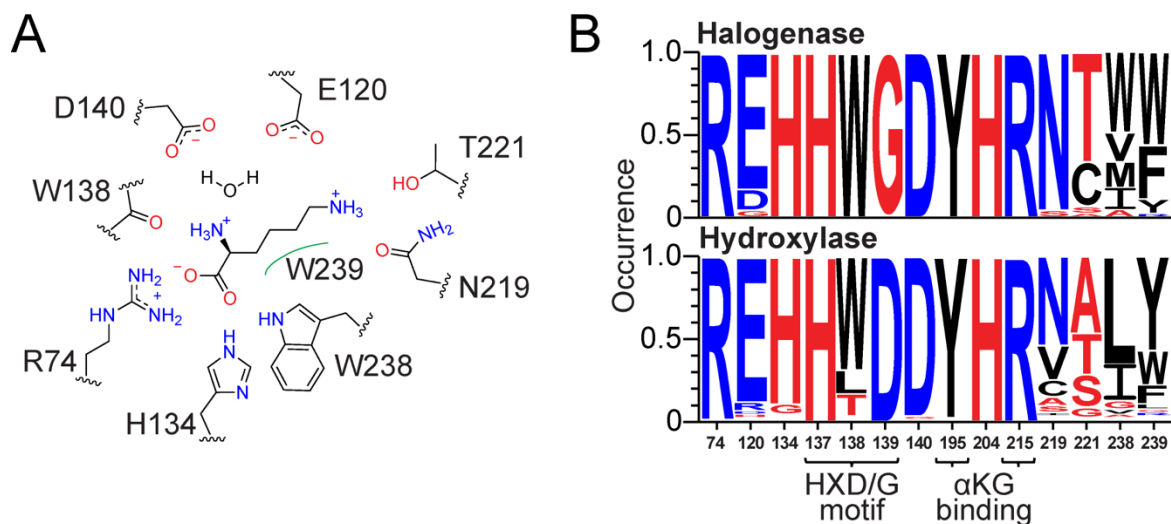


Figure 3.8. Alanine scan of active site residues. (A) Lysine-binding residues of BesD. (B) LOGOS plot of sequence conservation of active site residues within halogenase (top) and hydroxylase (bottom) homologs of BesD. Homologs of BesD (BLAST E-value of e^{-5}) were sorted based on the presence of the HXG (halogenase) or HXD (hydroxylase) motifs.

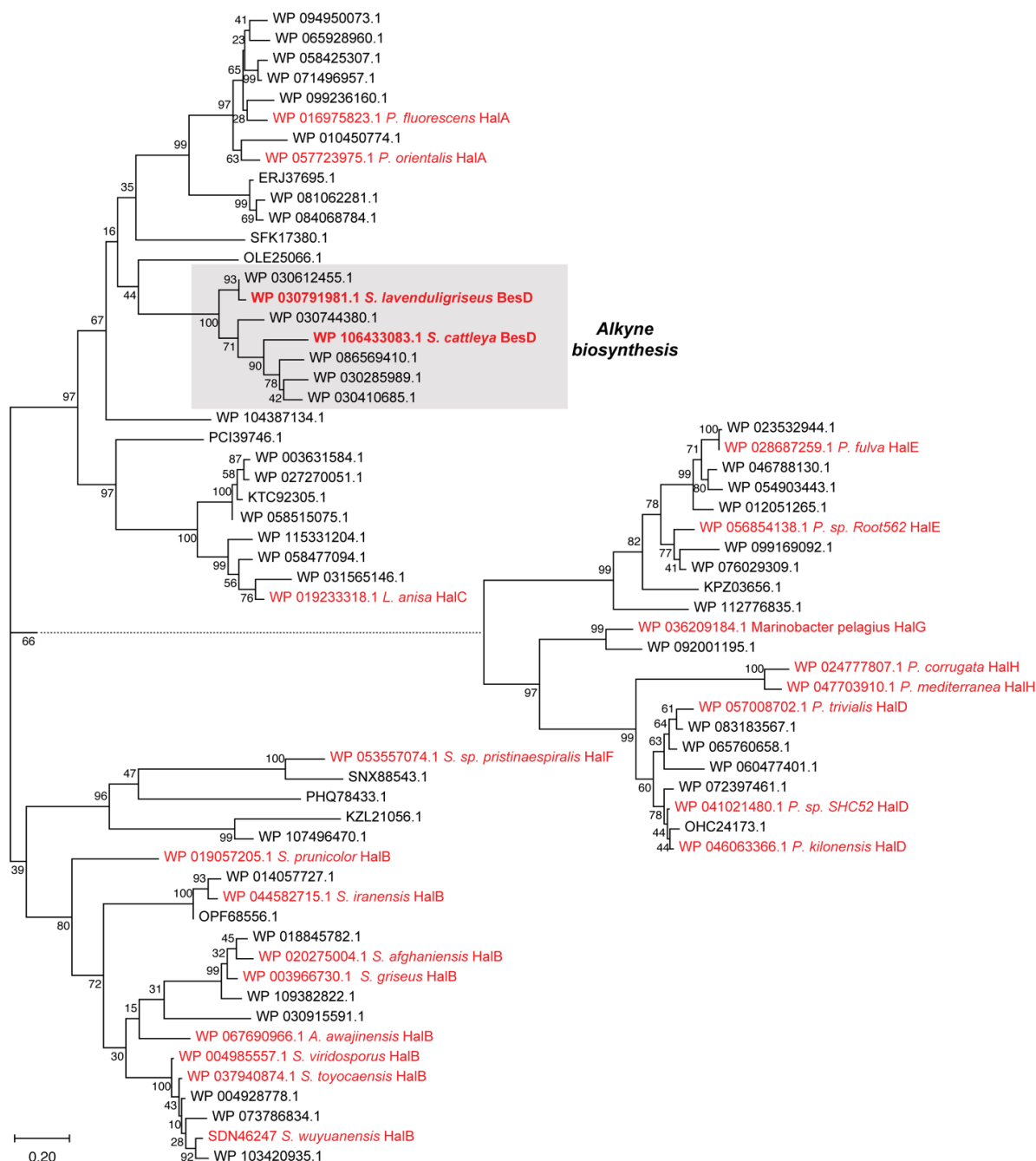


Figure 3.9. Maximum-likelihood phylogeny of BesD homologs. Phylogenetic tree of BesD homologs from the NCBI Non-redundant protein database assembled in MEGA [52] following alignment in MUSCLE [25]. The protein clade containing the alkyne biosynthetic genes [9] is indicated with brackets. Proteins tested in this work are indicated in red. Bootstrap values are indicated at branch points. Note that the tree does not contain sequences of hydroxylases, which were filtered out based on the presence of the HXD/E motif. Note also that the middle branch is extended to fit the page, and the dotted line is not to scale.

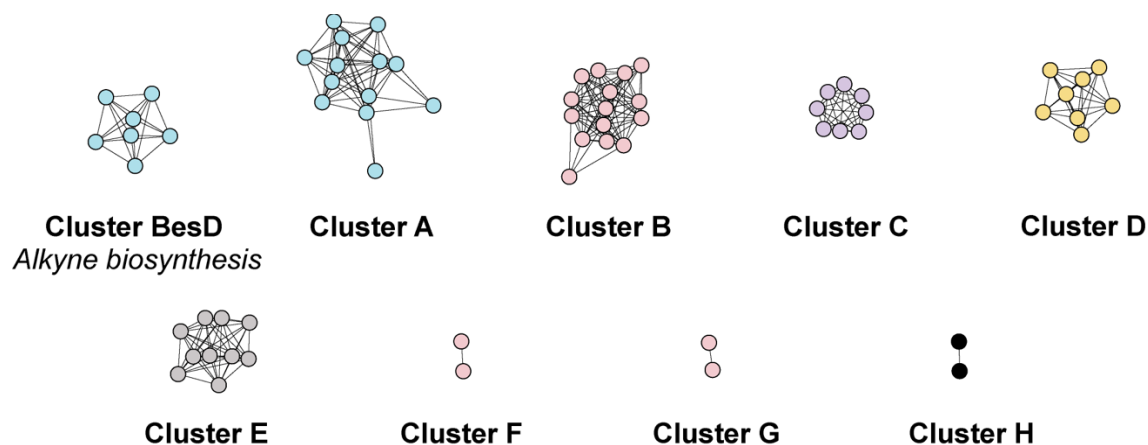


Figure 3.10. Sequence similarity network of BesD homologs generated using the Enzyme Function Initiative's Sequence Similarity Network tools. Homologs from NCBI's non-redundant protein database were identified using BLAST with an E-value cutoff of e^{-5} . Using Cytoscape Version 3.6.1, the network was adjusted by deleting edges with low alignments scores until the halogenases identified for β -ethynylserine (alkyne) biosynthesis [9] were separated into an isofunctional cluster, which occurred at an alignment score value of 88 (which corresponds to a sequence identity of ~56%).

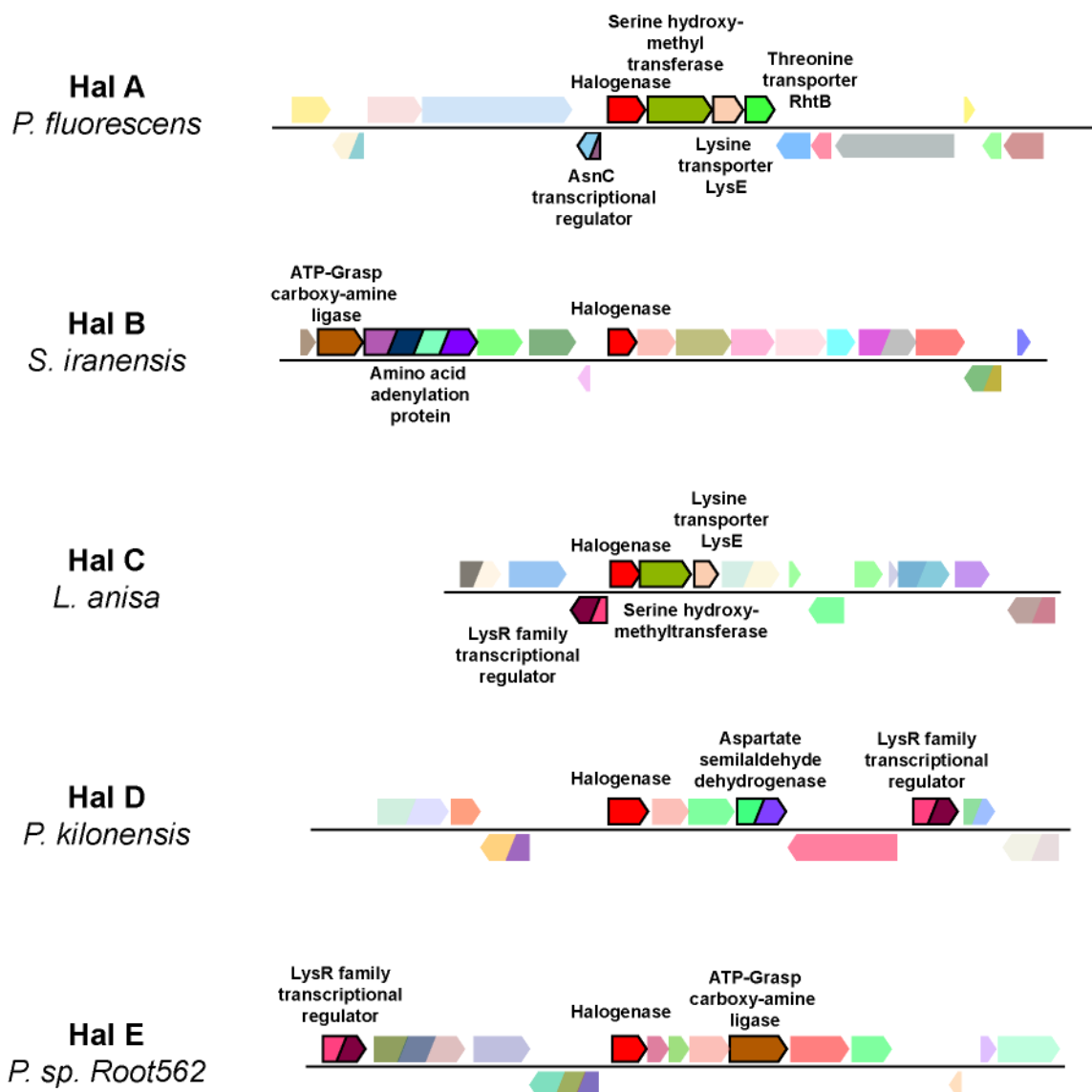


Figure 3.11. The Enzyme Function Initiative's Sequence Similarity and Genome Neighborhood Tools were used to extract the genomic context of the amino acid halogenase clusters. Adjacent genes include amino acid adenylation proteins, ATP-dependent carboxy-amine ligases, LysR family transcriptional regulators, Aspartate semialdehyde dehydrogenases, AsnC family transcriptional regulators, Serine hydroxymethyl transferases, Lysine transporter LysE, and Threonine transporter RhtB, genes associated with amino acid metabolism.

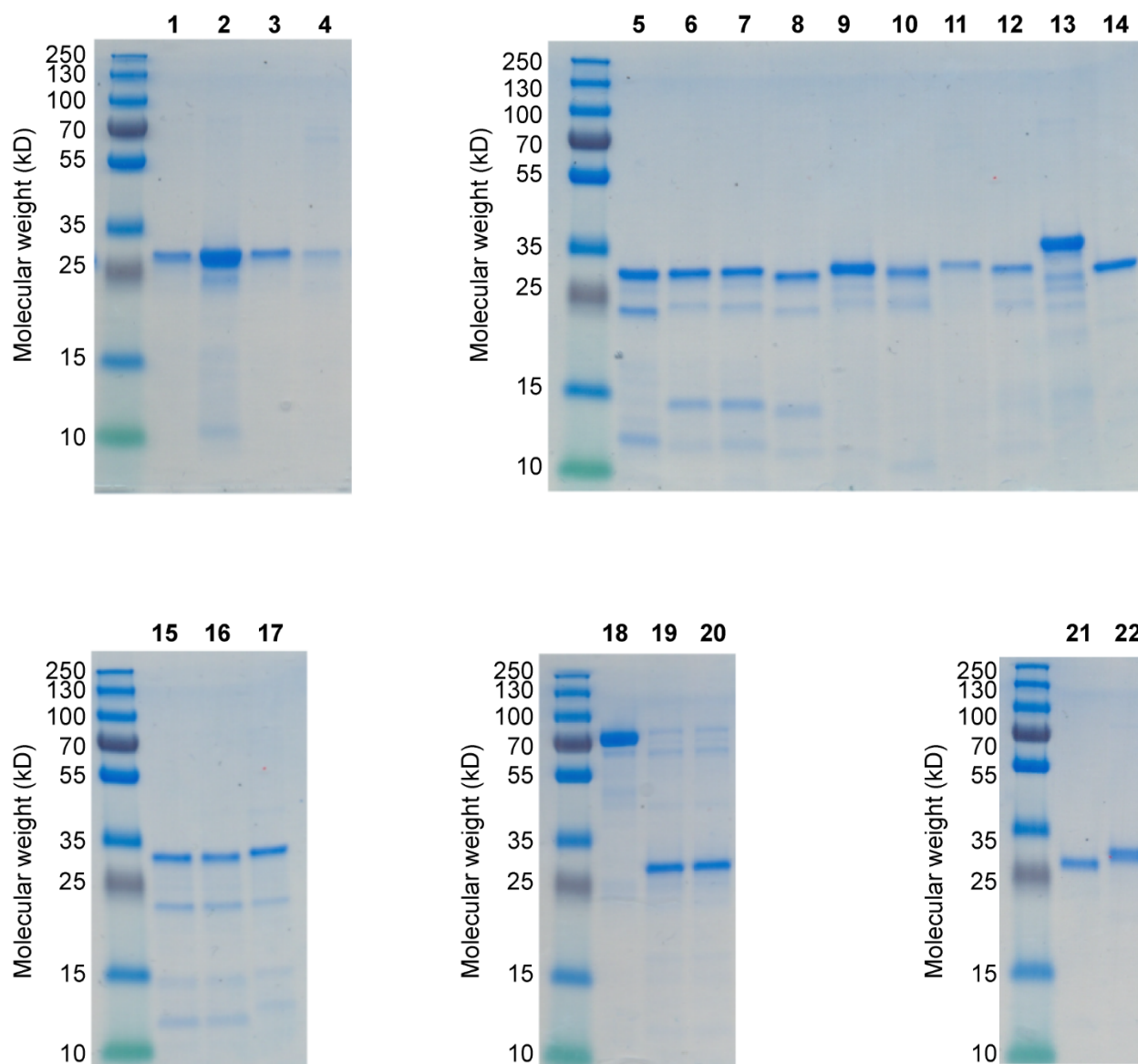


Figure 3.12. SDS-PAGE gels of purified halogenases used in this study. Lane 1, *S. cattleya* BesD; 2, *P. fluorescens* HalA; 3, *P. orientalis* HalA; 4, *L. anisa* HalC; 5, *S. wuyuanensis* HalB; 6, *S. toyocaensis* HalB; 7, *S. viridosporus* HalB; 8, *A. awajinensis* HalB; 9, *S. griesus* HalB; 10, *S. afghaniensis* HalB; 11, *S. iranensis* HalB; 12, *S. prunicolor* HalB; 13, *S. sp. pristinaespiralis* HalF; 14, *M. pelagius* HalG; 15, *P. kilonensis* HalD; 16, *P. sp. SHC52* HalD; 17, *P. trivialis* HalD; 18, MBP-Root sp. 562 HalE; 19, Root sp. 562 HalE; 20, *P. fulva* HalE; 21, *P. corrugata* Hal H; 22, *P. mediterranea* HalH. Protein purification was performed one time.

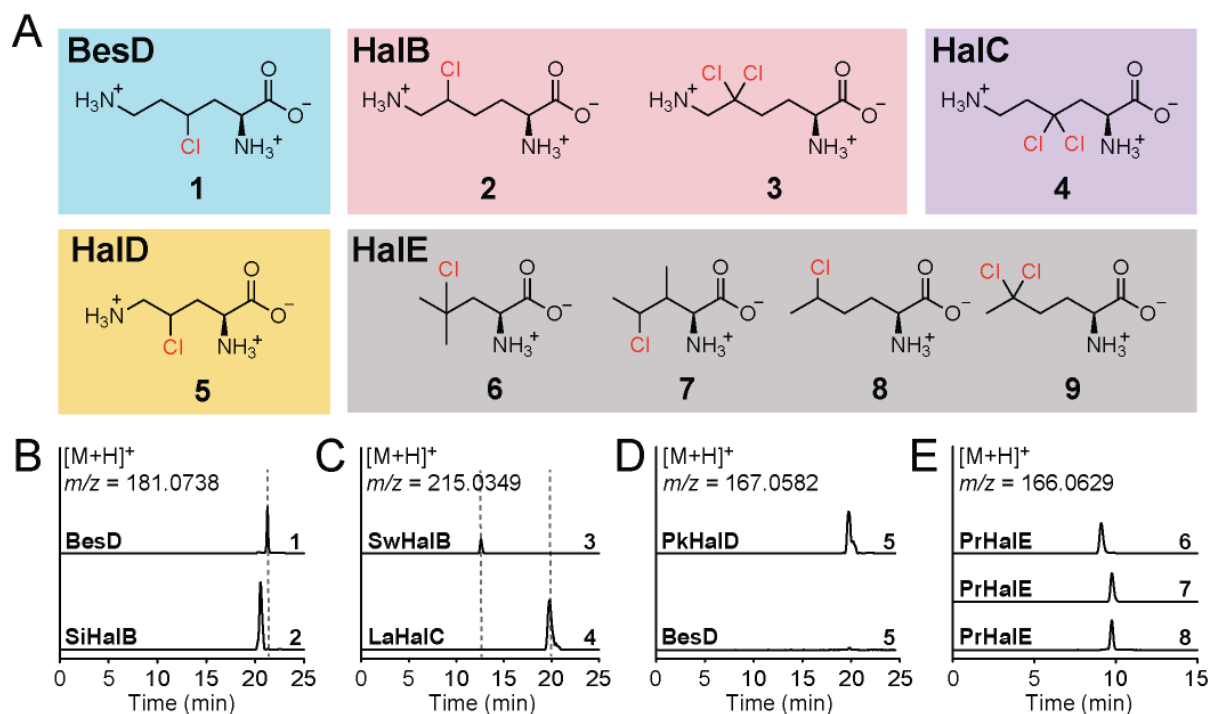


Figure 3.13. Amino acid halogenase diversity. (A) BesD homologs and products of enzymatic amino acid halogenation. (B) Regioselective mono-halogenation is observed with 4-Cl-lysine (**1**, $m/z = 181.0738$) or 5-Cl-lysine (**2**, $m/z = 181.0738$) produced by BesD or SiHalB, respectively. (C) Regioselective di-halogenation is also observed with 5,5-dichloro-lysine (**3**, $m/z = 215.0349$) or 4,4-dichloro-lysine (**4**, $m/z = 215.0349$) produced by SwHalB or LaHalC, respectively. (D) Ornithine is a substrate for both PkHalD and BesD (4-Cl-ornithine, **5**, $m/z = 167.0582$). (E) The non-polar amino acids, leucine, isoleucine, and norleucine are substrates for PrHalE (4-Cl-leucine, **6**, $m/z = 166.0629$; 4-Cl-isoleucine, **7**, $m/z = 166.0629$; 4-Cl-norleucine, **8**, $m/z = 166.0629$). Peak intensities are normalized in panel E. Assays include Fe^{II} , ascorbate, α KG, NaCl, in addition to the amino acid substrate. All extracted ion chromatograms are representative of at least 3 experimental replicates.

Table 3.2. Summary of halogenases in this study.

#	Organism	Accession	Cluster	Product
1	<i>Streptomyces cattleya</i>	WP_014151497	BesD	4-Cl-Lysine
2	<i>Streptomyces lavenduligriseus</i>	WP_030791981	BesD	4-Cl-Lysine
3	<i>Pseudomonas fluorescens</i>	WP_016975823	A	4-Cl-Lysine
4	<i>Pseudomonas orientalis</i>	WP_057723975	A	4-Cl-Lysine
5	<i>Streptomyces wuyuanensis</i>	SDN46247	B	5-Cl-Lysine, 5,5-dichloroysine
6	<i>Streptomyces toyocaensis</i>	WP_037940874	B	5-Cl-Lysine, 5,5-dichloroysine
7	<i>Streptomyces viridosporus</i>	WP_004985557	B	5-Cl-Lysine, 5,5-dichloroysine
8	<i>Actinoplanes awajinensis</i>	WP_067690966	B	5-Cl-Lysine, 5,5-dichloroysine
9	<i>Streptomyces griseus</i>	WP_003966730	B	5-Cl-Lysine, 5,5-dichloroysine
10	<i>Streptomyces afghaniensis</i>	WP_020275004	B	5-Cl-Lysine, 5,5-dichloroysine
11	<i>Streptomyces iranensis</i>	WP_044582715	B	5-Cl-Lysine, 5,5-dichloroysine
12	<i>Streptomyces prunicolor</i>	WP_019057205	B	5-Cl-Lysine, 5,5-dichloroysine
13	<i>Legionella anisa</i>	WP_019233318	C	4-Cl-Lysine, 4,4-dichloroysine
14	<i>Pseudomonas kilonensis</i>	WP_046063366	D	4-Cl-Ornithine
15	<i>Pseudomonas SHC52</i>	WP_041021480	D	4-Cl-Ornithine
16	<i>Pseudomonas trivialis</i>	WP_057008702	D	4-Cl-Ornithine
17	<i>Pseudomonas</i> sp. Root562	WP_056854138	E	4-Cl-Leucine, 4-Cl-Isoleucine, 5,5-dichloronorleucine
18	<i>Pseudomonas fulva</i> (multisp)	WP_028687259	E	4-Cl-Leucine, 4-Cl-Isoleucine, 5,5-dichloronorleucine
19	<i>Streptomyces</i> sp. pristinaespiralis	WP_053557074	F	5-Cl-Lysine, 5,5-dichloroysine
20	<i>Marinobacter pelagius</i>	WP_036209184	G	5-Cl-Lysine, 5,5-dichloroysine
21	<i>Pseudomonas corrugata</i>	WP_024777807	H	None detected
22	<i>Pseudomonas mediterranea</i>	WP_047703910	H	None detected

Table 3.3. Substrate profiles of tested halogenases from reaction screening. Reactions (50 μ L) contained the following amino acids: L-alanine, L-arginine, L-asparagine, L-aspartate, L-cysteine, L-glutamine, L-glutamate, L-glycine, L-histidine, L-isoleucine, L-leucine, L-lysine, L-methionine, L-phenylalanine, L-proline, L-serine, L-threonine, L-tryptophan, L-valine, L-ornithine (O), and DL-norleucine (NL) (0.5 mM each), sodium α KG (5 mM), sodium ascorbate (1 mM), $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ (1 mM), and sodium chloride (5 mM) in 100 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of purified halogenase variants (10 μ M final concentration) and allowed to proceed for 1 h at room temperature before quenching in 2 vol of methanol with 1% (v/v) formic acid. Samples were then analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above. Activity detected on substrates is shown in green. Please note that when SDN46247 was tested on individual amino acids, chlorinated product was detected with I, L, and NL, likely because no K was present to compete.

	A	R	N	D	C	E	Q	G	H	I	L	K	M	F	P	S	T	W	V	NL	O
WP_014151497																					
WP_030791981																					
WP_016975823																					
WP_057723975																					
SDN46247																					
WP_037940874																					
WP_004985557																					
WP_067690966																					
WP_003966730																					
WP_020275004																					
WP_044582715																					
WP_019057205																					
WP_019233318																					
WP_046063366																					
WP_041021480																					
WP_057008702																					
WP_056854138																					
WP_028687259																					
WP_053557074																					
WP_036209184																					
WP_024777807																					
WP_047703910																					

Table 3.4. Percent identity matrix of tested halogenases. An all-by-all sequence alignment was performed using Clustal Omega [53]. The sequence identity (%) matrix output is displayed for halogenases A-H. The number representing each halogenase corresponds to the accession number as shown in *Table 3.2*. Enzymes within a cluster have a percent identity of >~56%.

	19	12	11	9	10	8	5	6	7	13	3	4	1	2	17	18	20	21	22	14	15	16
HalF (19)	100	36	37	38	40	36	36	37	37	32	35	34	34	36	31	30	35	28	27	34	33	34
HalB (12)	36	100	52	54	55	54	58	59	58	35	47	47	40	40	42	41	46	29	30	39	38	39
HalB (11)	37	52	100	58	59	58	57	58	59	39	43	44	43	41	40	40	45	33	36	41	41	41
HalB (9)	38	54	58	100	90	66	63	63	64	40	44	44	40	41	38	39	42	32	33	38	38	38
HalB (10)	40	55	59	90	100	63	62	62	63	41	45	46	41	41	38	39	44	32	33	39	39	38
HalB (8)	36	54	58	66	63	100	69	70	72	41	43	42	41	40	42	40	44	34	36	42	42	41
HalB (5)	36	58	57	63	62	69	100	92	91	39	42	42	40	41	43	42	45	33	35	41	40	40
HalB (6)	37	59	58	63	62	70	92	100	96	39	43	42	41	43	41	41	45	35	36	41	40	40
HalB (7)	37	58	59	64	63	72	91	96	100	39	42	42	42	43	41	43	46	36	37	42	41	41
HalC (13)	32	35	39	40	41	41	39	39	39	100	50	54	50	50	38	39	37	28	31	35	37	35
HalA (3)	35	47	43	44	45	43	42	43	42	50	100	83	53	53	41	39	42	31	31	40	41	40
HalA (4)	34	47	44	44	46	42	42	42	42	54	83	100	53	52	40	38	41	33	31	40	41	40
BesD (1)	34	40	43	40	41	41	40	41	42	50	53	53	100	76	38	38	37	29	31	38	38	37
BesD (2)	36	40	41	41	41	40	41	43	43	50	53	52	76	100	37	37	37	28	31	38	38	37
HalE (17)	31	42	40	38	38	42	43	41	41	38	41	40	38	37	100	79	44	29	32	40	41	41
HalE (18)	30	41	40	39	39	40	42	41	43	39	39	38	38	37	79	100	44	32	32	41	41	41
HalG (20)	35	46	45	42	44	44	45	45	46	37	42	41	37	37	44	44	100	36	39	51	53	51
HalH (21)	28	29	33	32	32	34	33	35	36	28	31	33	29	28	29	32	36	100	89	55	56	54
HalH (22)	27	30	36	33	33	36	35	36	37	31	31	31	31	31	32	32	39	89	100	58	59	58
HalD (14)	34	39	41	38	39	42	41	41	42	35	40	40	38	38	40	41	51	55	58	100	97	86
HalD (15)	33	38	41	38	39	42	40	40	41	37	41	41	38	38	41	41	53	56	59	97	100	86
HalD (16)	34	39	41	38	38	41	40	40	41	35	40	40	37	37	41	41	51	54	58	86	86	100

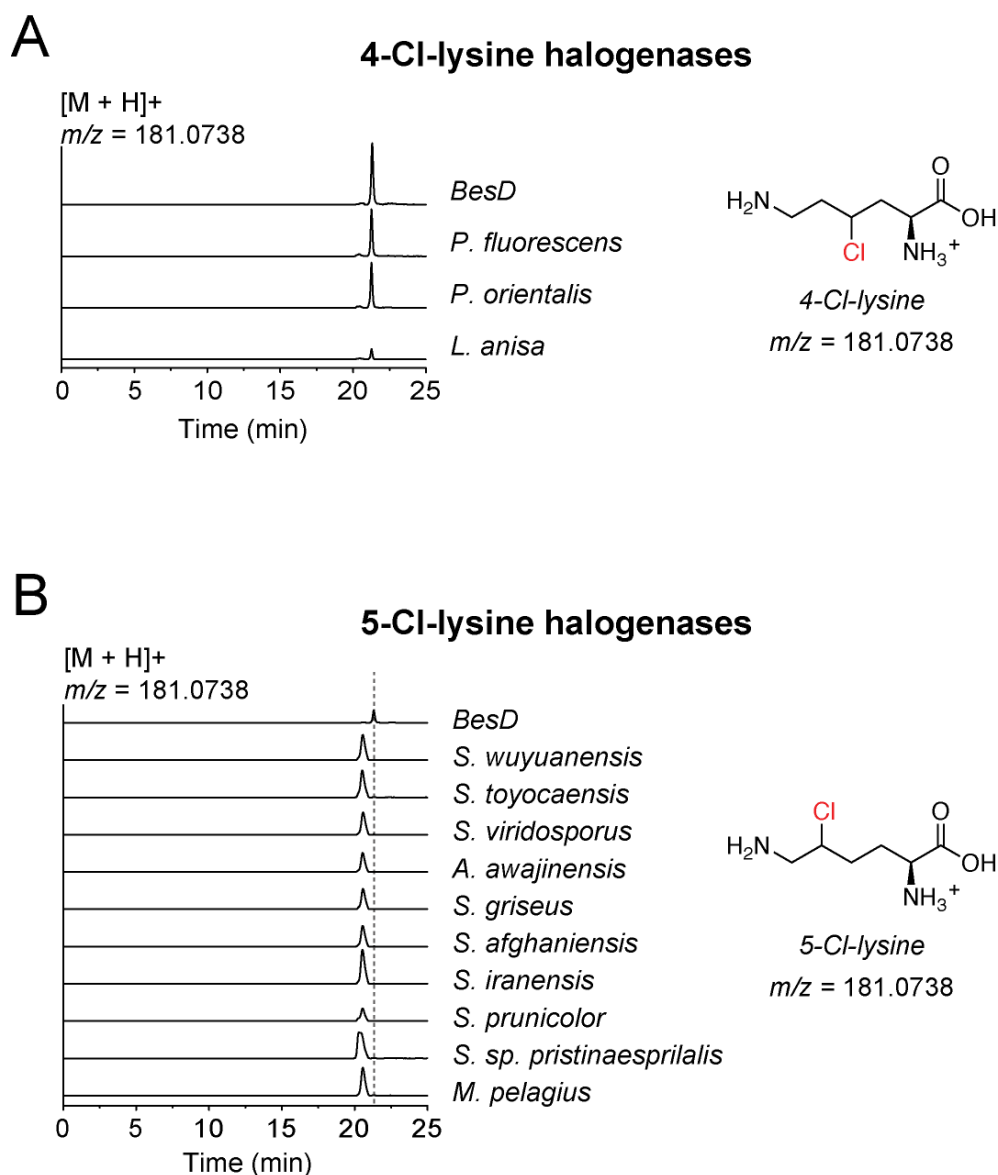


Figure 3.14. LC/MS analysis of regioselective formation of 4-Cl-lysine and 5-Cl-lysine by halogenase homologs. (A) Extracted ion chromatograms of the products observed upon halogenation of lysine to 4-Cl-lysine ($m/z = 181.0738$) by homologs of BesD. Reactions contained L-lysine, α KG, sodium ascorbate, Fe^{II} , and chloride. HalA (from *P. fluorescens* and *P. orientalis*) and HalC (from *L. anisa*) yield mono-chlorinated products with the same exact mass and retention time as 4-Cl-lysine from BesD. (B) Extracted ion chromatograms of the products observed upon halogenation of lysine to 5-Cl-lysine by HalB homologs of BesD. Reactions contained L-lysine, α KG, sodium ascorbate, Fe^{II} , and chloride. HalB homologs (from *S. wuyuanensis*, *S. toyocaensis*, *S. viridosporus*, *A. awajinensis*, *S. griseus*, *S. afghaniensis*, *S. iranensis*, *S. prunicolor*), HalF (from *S. sp. pristinaespiralis*), and HalG (from *M. Pelagius*) yield mono-chlorinated products with the same exact mass as 4-Cl-lysine from BesD ($m/z = 181.0738$), but with an earlier retention time. Note that a small peak corresponding to the mass and retention time of 5-Cl-lysine is observed in the 4-Cl-lysine halogenases, although 4-Cl-lysine is the dominant product. All extracted ion chromatograms are representative of at least 3 experimental replicates.

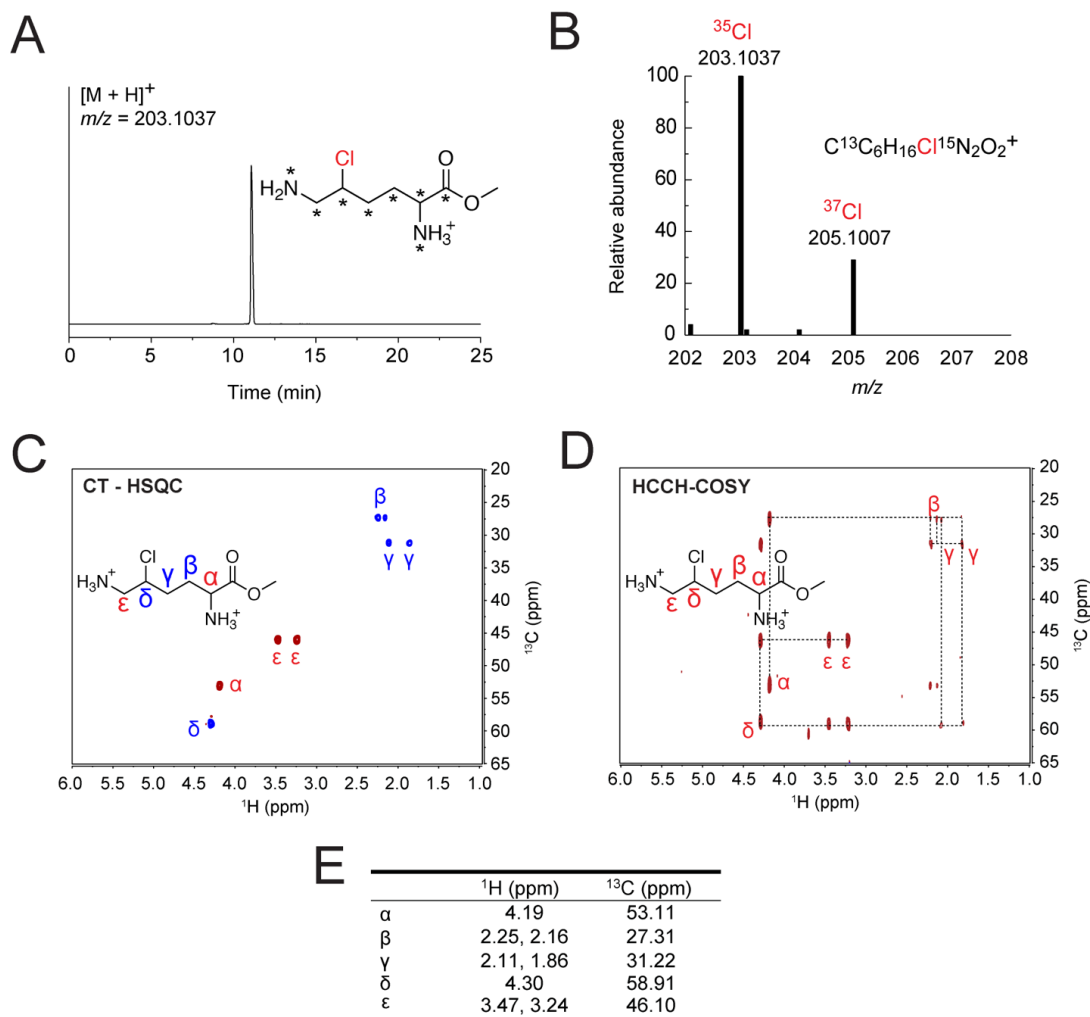


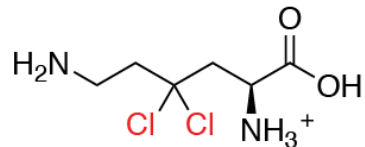
Figure 3.15. NMR analysis of 5-Cl-lysine methyl ester. (A) HalB from *S. iranensis* (SiHalB) was incubated with fully ^{15}N - and ^{13}C -labeled L-lysine, Fe, NaCl, αKG , and ascorbate for 60 min before quenching with methanolic HCl (3M) to yield the $[\text{N}_2^{15}, \text{C}_6^{13}]$ -5-chlorolysine methyl ester product. The derivatized HalB product was then isolated by HPLC and characterized. Extracted ion chromatograms of the $[\text{N}_2^{15}, \text{C}_6^{13}]$ -5-chlorolysine ($m/z = 203.1037$) produced by SiHalB. Purification of the compound was performed once. (B) Mass spectra show the characteristic Cl isotope pattern for $[\text{N}_2^{15}, \text{C}_6^{13}]$ -5-chlorolysine produced by SiHalB. (C) 2D ^1H - ^{13}C CT-HSQC was used to confirm assignment of C_α and C_ϵ of the $[\text{N}_2^{15}, \text{C}_6^{13}]$ -5-chlorolysine methyl ester produced by BesD. In this experiment carbons with two neighbors (blue) have opposite phase of carbons with one or three neighbors (red). (D) 2D ^1H - ^{13}C HCCH COSY was used to assign connectivity of carbons of the $[\text{N}_2^{15}, \text{C}_6^{13}]$ -5-chlorolysine methyl ester produced by SiHalB. (E) Chemical shifts determined from the 2D-NMR data.

exact mass and characteristic chlorinated isotope pattern as 4-Cl-lysine, but with a different retention time (*Figure 3.14*). Through NMR analysis of the methyl esterified product, we identified the product of SiHalB as 5-Cl-lysine (*Figure 3.15*). Though 44% identical, SiHalB and BesD have distinct regioselectivities, highlighting the capability of $\text{Fe}^{\text{II}}/\alpha\text{KG}$ -dependent enzymes to perform regioselective halogenation under mild conditions. We also found halogenases that perform both mono- and di-chlorination of lysine. When lysine is incubated with BesD, only monochlorinated 4-Cl-lysine is observed. However, in addition to monochlorination of lysine, *Legionella anisa* HalC (LaHalC) could also carry out dichlorination to produce 4,4-dichlorolysine (*Figure 3.16-3.17*) while HalB from *Streptomyces wuyuanensis* (SwHalB) yielded 5,5-dichlorolysine (*Figure 3.18-3.19*). Dichlorination has been observed in radical carrier protein-dependent halogenases as well [13].

In addition to differing regioselectivities, other BesD family members were identified that could act on other amino acid substrates. Enzymes from Cluster D exhibit chain-length preference and halogenate the 5-carbon substrate ornithine preferentially over the 6-carbon substrate lysine. In particular, the *Pseudomonas kilonensis* HalD (PkHalD) was found to prefer ornithine ($k_{\text{cat}}/K_{\text{M}} = 330 \pm 70 \text{ mM}^{-1}\text{min}^{-1}$) as a substrate by 25-fold over lysine ($k_{\text{cat}}/K_{\text{M}} = 13.2 \pm 3.6 \text{ mM}^{-1}\text{min}^{-1}$) (*Figure 3.20-3.23*). Finally, Cluster E was found to contain aliphatic amino acid halogenases. When we assayed *P. sp. Root562* HalE (PrHalE), we observed no activity on the amino acids lysine and ornithine. However, when we incubated PrHalE with the aliphatic amino acids leucine, isoleucine, and norleucine, we observed the expected mass and isotopic patterns for chlorinated products of those amino acids (*Figure 3.24-3.25*). We additionally observed a dichlorinated species when norleucine was used as a substrate. Further characterization by NMR indicated that PrHalE chlorinates leucine and isoleucine (CH_2) at the γ -position and norleucine at the δ -position (*Figure 3.26-3.28*).

Remarkably, halogenases from Clusters A-H have evolved to accommodate different amino acids, with regioselectivity for specific sites on those substrates, while maintaining fidelity for halogenation over the competing side reaction of hydroxylation. In order to further explore the structural basis for these differences, Halogenase A-E sequences were aligned (*Figure 3.29*). As expected, the carboxylate and amine binding residues Arg74, His134, and Asp140 are highly conserved, regardless of substrate preference. However, it was surprising to note that the ϵ -amine binding residues of BesD (Glu120, Asn219, and Thr221) were also highly conserved even for halogenases that act on non-polar amino acids. The greatest disparity in sequence occurs at W238 and W239, which are located on the C-terminal end of the protein sequence. Sequence alignment reveals that the C-terminal regions of Halogenases A-E are highly variable, differing by as many as 10 amino acids in length and resulting in unreliable alignments of that region. Structural analysis of BesD shows that the C-terminus closes over the substrate in the active site of the protein, occluding it from solvent (*Figure 3.30*). The C-terminus of BesD additionally interacts with a corresponding internal loop near the active site (residues 117-119), which also varies among the halogenase clusters. Combined with the structure, the sequence analysis suggests that the C-terminus and the corresponding loop could be important determinants of substrate selectivity.

Biosynthetic applications of amino acid halogenases. Amino acids are key building blocks for metabolism, serving as the source for many metabolites and natural products, as well as ribosomally- and non-ribosomally synthesized peptides and proteins. Therefore, the discovery of the BesD family of halogenases raises the possibility that halogens and other functional groups could be incorporated into these classes of compounds through engineering downstream



4,4-dichlorolysine

$m/z = 215.0349$

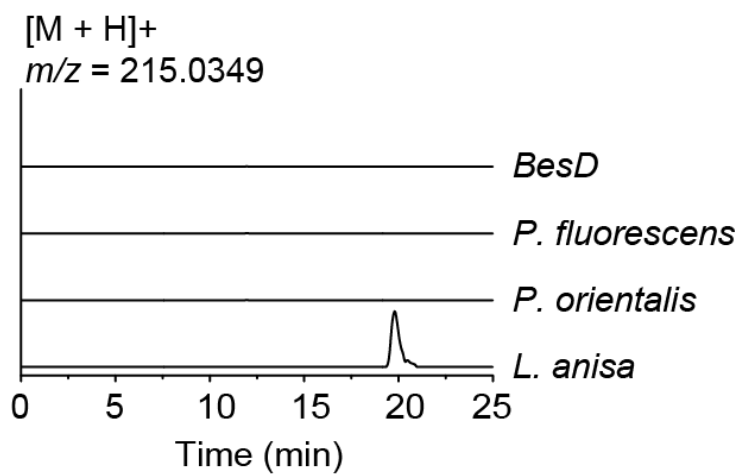


Figure 3.16. LC/MS analysis of dichlorination of lysine to 4,4-dichlorolysine by HalC from *L. anisa*. Extracted ion chromatograms of the products observed upon halogenation of lysine to 4,4-dichlorolysine ($m/z = 215.0349$) by HalC from *L. anisa* (LaHalC). Reactions contained L-lysine, α KG, sodium ascorbate, Fe^{II}, and chloride. BesD and homologs from Cluster A (*P. fluorescens* and *P. orientalis*) do not yield dichlorolysine, and no product with the expected mass is observed. In contrast, LaHalC yields 4,4-dichlorolysine ($m/z = 215.0349$). All extracted ion chromatograms are representative of at least 3 experimental replicates.

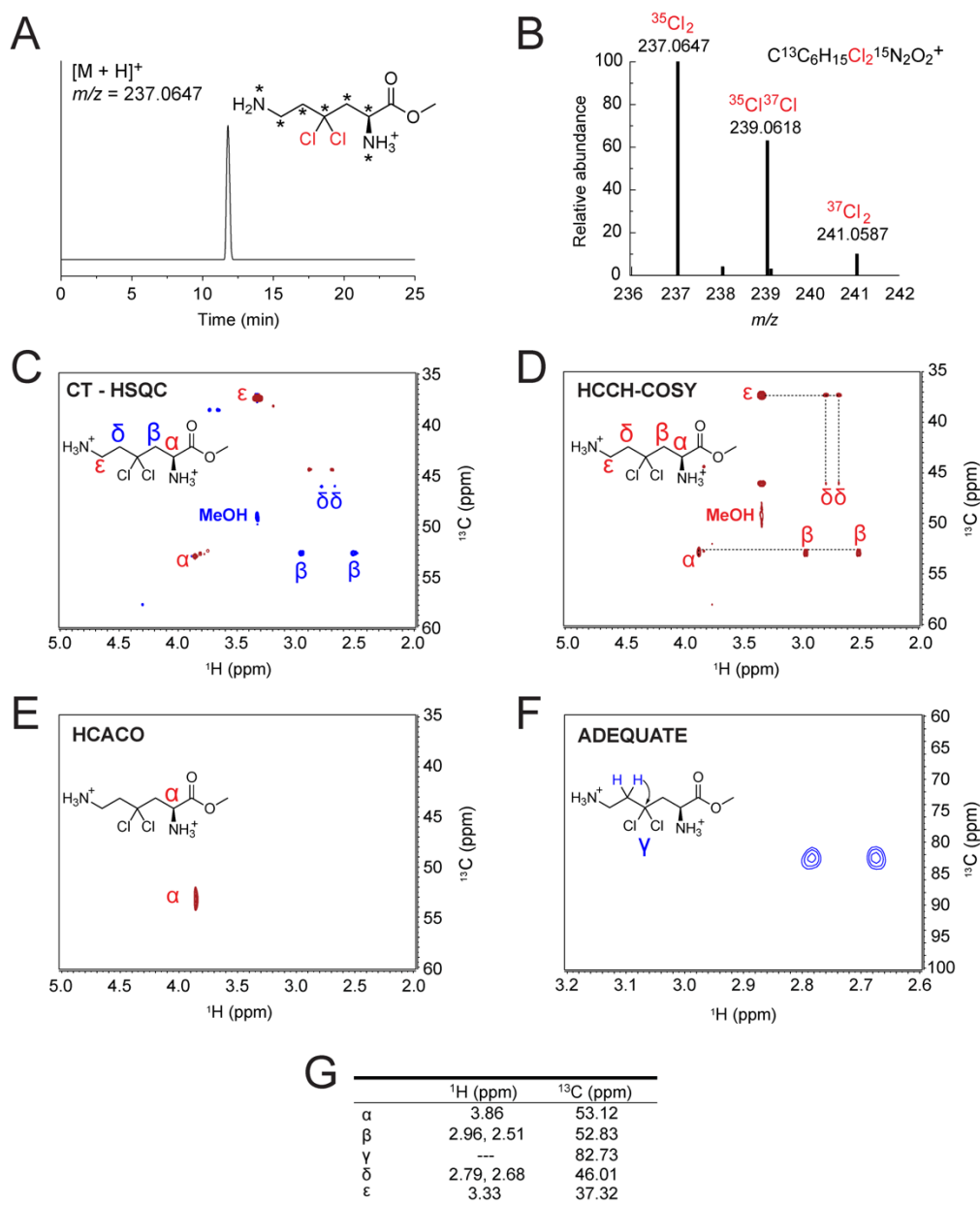
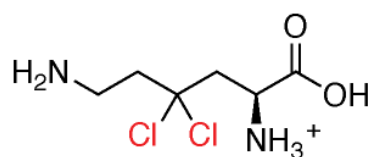
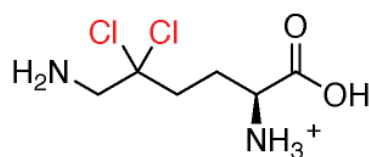


Figure 3.17. NMR analysis of 4,4-dichlorolysine methyl ester. (A) HalC from *L. anisa* was incubated with fully ^{15}N - and ^{13}C -labeled L-lysine, Fe, NaCl, αKG , and ascorbate for 4 h before quenching with methanolic HCl (3M) to yield the $[\text{}^{15}\text{N}_2, \text{}^{13}\text{C}_6]$ -4,4-dichlorolysine methyl ester product. The derivatized HalC product was then isolated by HPLC and characterized. Extracted ion chromatograms of the $[\text{}^{15}\text{N}_2, \text{}^{13}\text{C}_6]$ -4,4-dichlorolysine methyl ester ($m/z = 237.0647$). Purification of the compound was performed once. (B) Mass spectra show the characteristic Cl isotope pattern for $[\text{}^{15}\text{N}_2, \text{}^{13}\text{C}_6]$ -4,4-dichlorolysine methyl ester. (C) 2D ^1H - ^{13}C CT-HSQC was used to confirm assignment of C_α and C_ϵ of the $[\text{}^{15}\text{N}_2, \text{}^{13}\text{C}_6]$ -4,4-dichlorolysine methyl ester produced by HalC. In this experiment carbons with two neighbors (blue) have opposite phase of carbons with one or three neighbors (red). (D) 2D ^1H - ^{13}C HCCH COSY was used to assign connectivity of carbons of the $[\text{}^{15}\text{N}_2, \text{}^{13}\text{C}_6]$ -4,4-dichlorolysine methyl ester produced by BesD. (E) HCACO to confirm assignment of α proton. (F) 1,1 ADEQUATE was used to obtain the shift of C_7 by correlation with H_δ . (G) Chemical shifts determined from the 2D-NMR data.



4,4-dichlorolysine

$m/z = 215.0349$



5,5-dichlorolysine

$m/z = 215.0349$

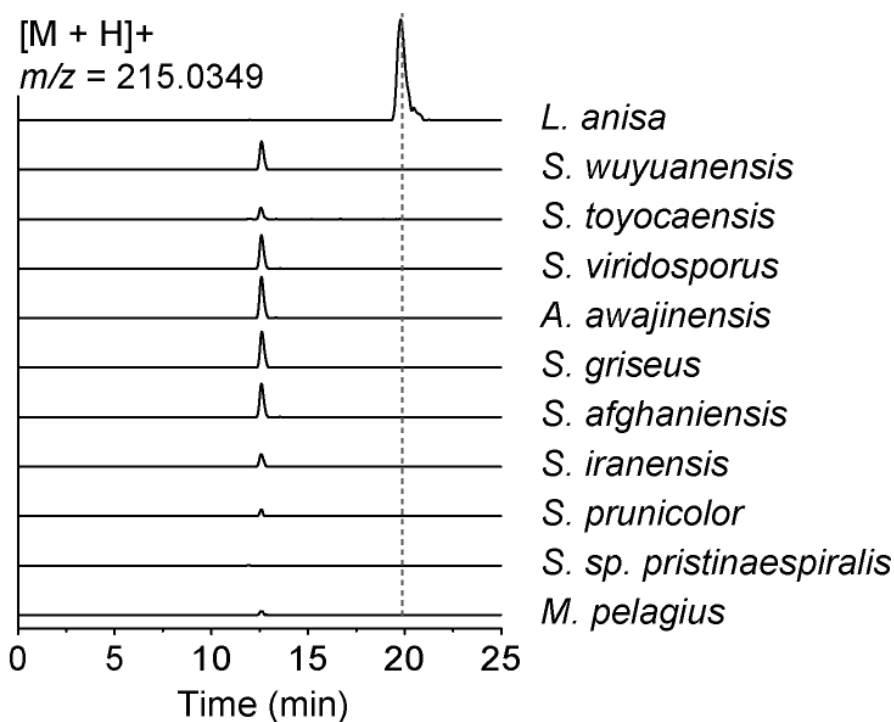


Figure 3.18. LC/MS comparison of 4,4-dichlorolysine vs. 5,5-dichlorolysine products. Extracted ion chromatograms of the products observed upon halogenation of lysine to 4,4-dichlorolysine ($m/z = 215.0349$) by HalC from *L. anisa* and halogenation of lysine to 5,5-dichlorolysine ($m/z = 215.0349$) by halogenase homologs from Cluster B (from *S. wuyuanensis*, *S. toyocaensis*, *S. viridosporus*, *A. awajinensis*, *S. griseus*, *S. afghaniensis*, *S. iranensis*, *S. prunicolor*), Cluster F (from *S. sp. pristinaespiralis*), and Cluster G (from *M. pelagius*). Reactions contained L-lysine, α KG, sodium ascorbate, Fe^{II}, and chloride. The structures of the observed products were confirmed by NMR. All extracted ion chromatograms are representative of at least 3 experimental replicates.

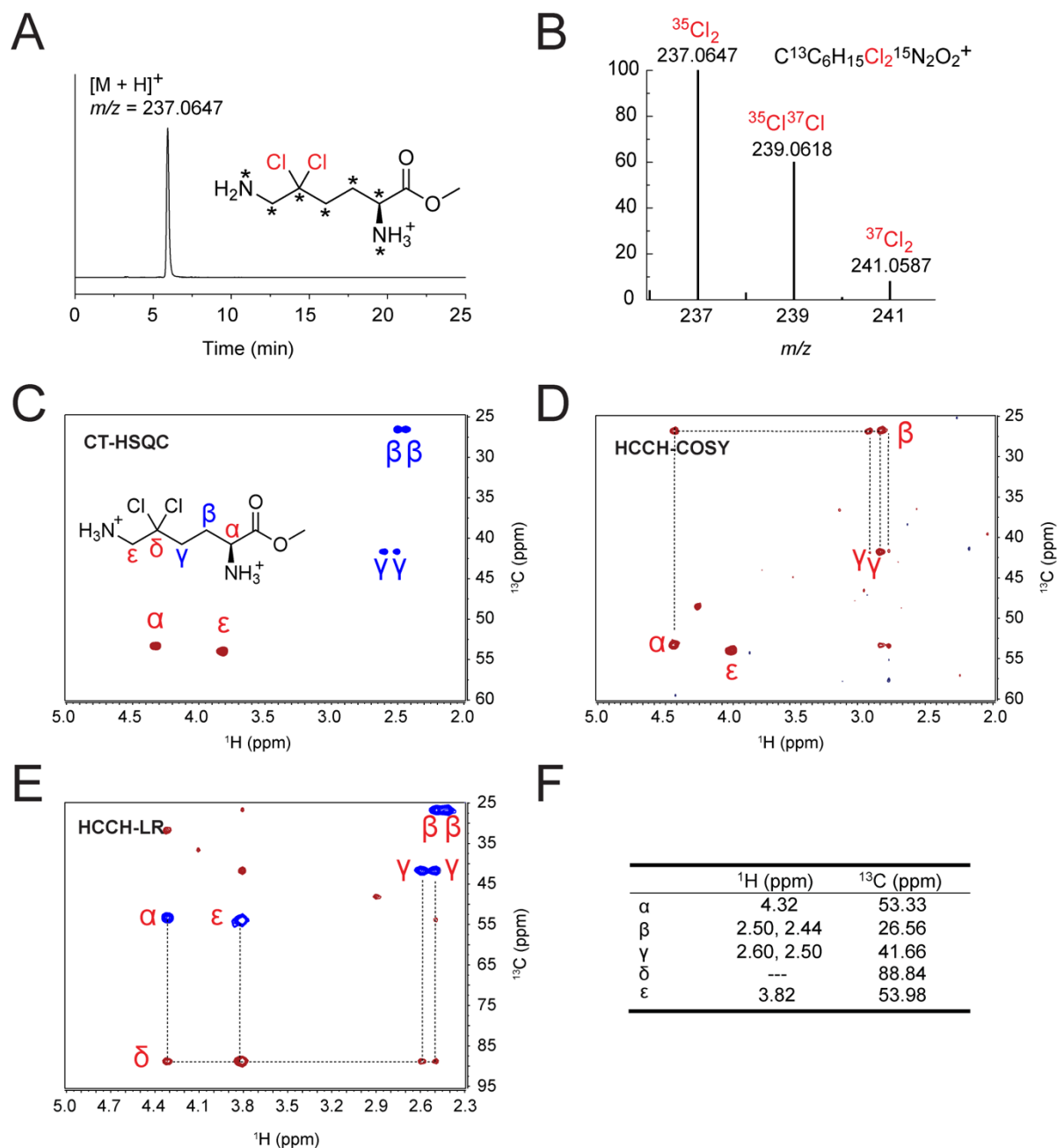


Figure 3.19. NMR analysis of 5,5-dichlorolysine methyl ester. (A) HalB from *S. wuyuanensis* (SwHalB) was incubated with fully ^{15}N - and ^{13}C -labeled L-lysine, Fe, NaCl, and αKG for 120 min before quenching with methanolic HCl (3M) to yield the $[\text{N}_2^{15}, \text{C}_6^{13}]$ -5,5-dichlorolysine methyl ester product. The derivatized SwHalB product was then isolated by HPLC and characterized. Purification of the compound was performed once. (B) Extracted ion chromatograms of the $[\text{N}_2^{15}, \text{C}_6^{13}]$ -5,5-dichlorolysine methyl ester ($m/z = 237.0647$). Mass spectra show the characteristic Cl isotope pattern for $[\text{N}_2^{15}, \text{C}_6^{13}]$ -5,5-dichlorolysine methyl ester. (C) 2D ^1H - ^{13}C CT-HSQC was used to confirm assignment of C_α and C_ϵ of the $[\text{N}_2^{15}, \text{C}_6^{13}]$ -5,5-dichlorolysine methyl ester produced by SwHalB. In this experiment carbons with two neighbors (blue) have opposite phase of carbons with one or three neighbors (red). (D) 2D ^1H - ^{13}C HCCH COSY was used to assign connectivity of carbons of the $[\text{N}_2^{15}, \text{C}_6^{13}]$ -5,5-dichlorolysine methyl ester. (E) A 2D ^1H - ^{13}C long-range HCCH experiment was used to obtain the shift of C_δ by correlation to H_α , H_γ , and H_ϵ . (F) Chemical shifts determined from the 2D-NMR data.

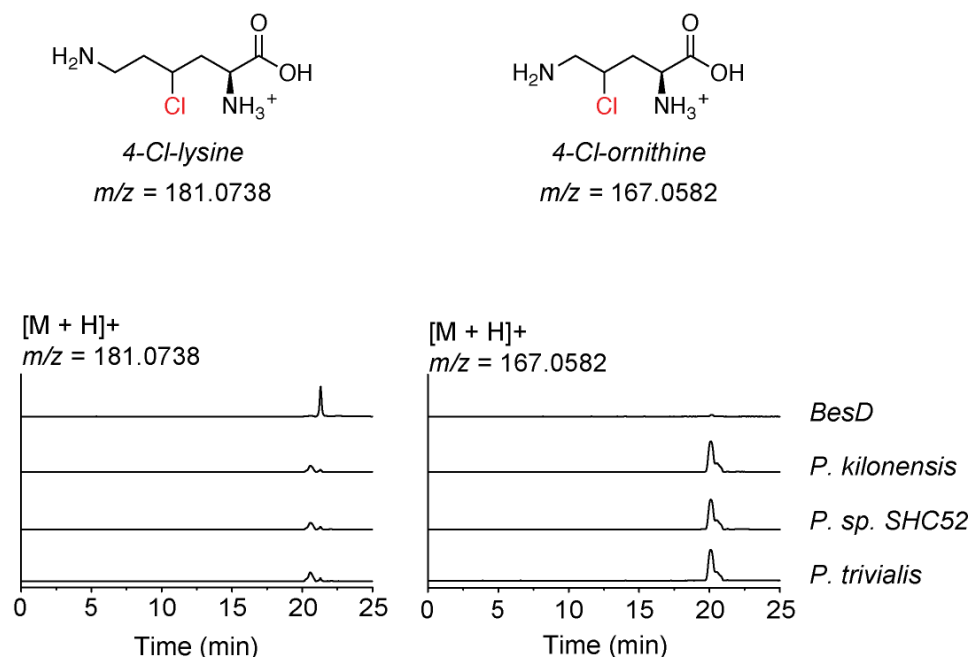


Figure 3.20. LC/MS analysis of HalD chlorination of ornithine. Extracted ion chromatograms of the products observed upon incubation of halogenase homologs with Fe, ascorbate, NaCl, α KG, lysine and ornithine. Halogenase homologs from *P. kilonensis* (PkHalD), *P. sp. SHC52* (PsHalD), and *P. trivialis* (PtHalD) yield a lower extracted ion count by LC/MS analysis than the halogenase from *S. cattleya* (BesD). Instead, these halogenases yield more 4-Cl-ornithine ($m/z = 167.0582$) than BesD from *S. cattleya*. All extracted ion chromatograms are representative of at least 3 experimental replicates.

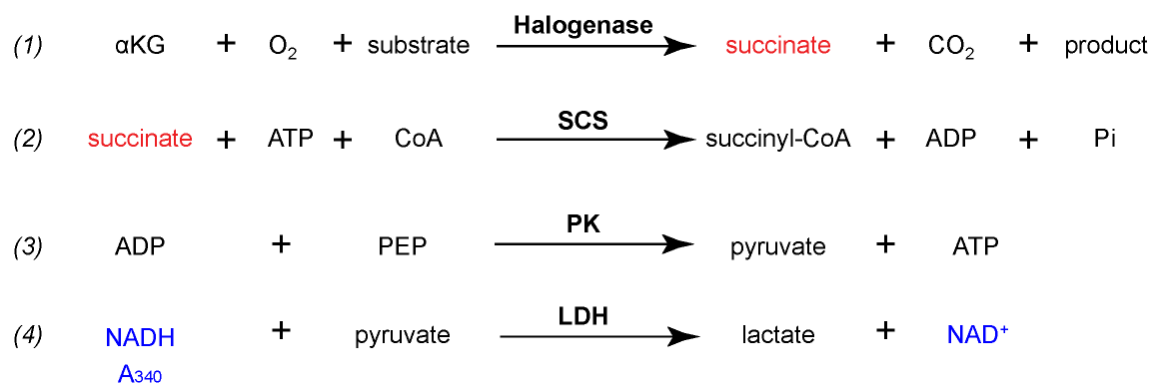


Figure 3.21. Schematic for the coupled reaction for monitoring halogenase activity through NADH oxidation. The rate of succinate formation by HalA from *P. fluorescens* (PfHalA) and HalD from *P. kilonensis* (PkHalD) was monitored by the change in A_{340} using an NADH-coupled assay with succinyl-CoA synthetase (SCS), pyruvate kinase (PK) and lactate dehydrogenase.

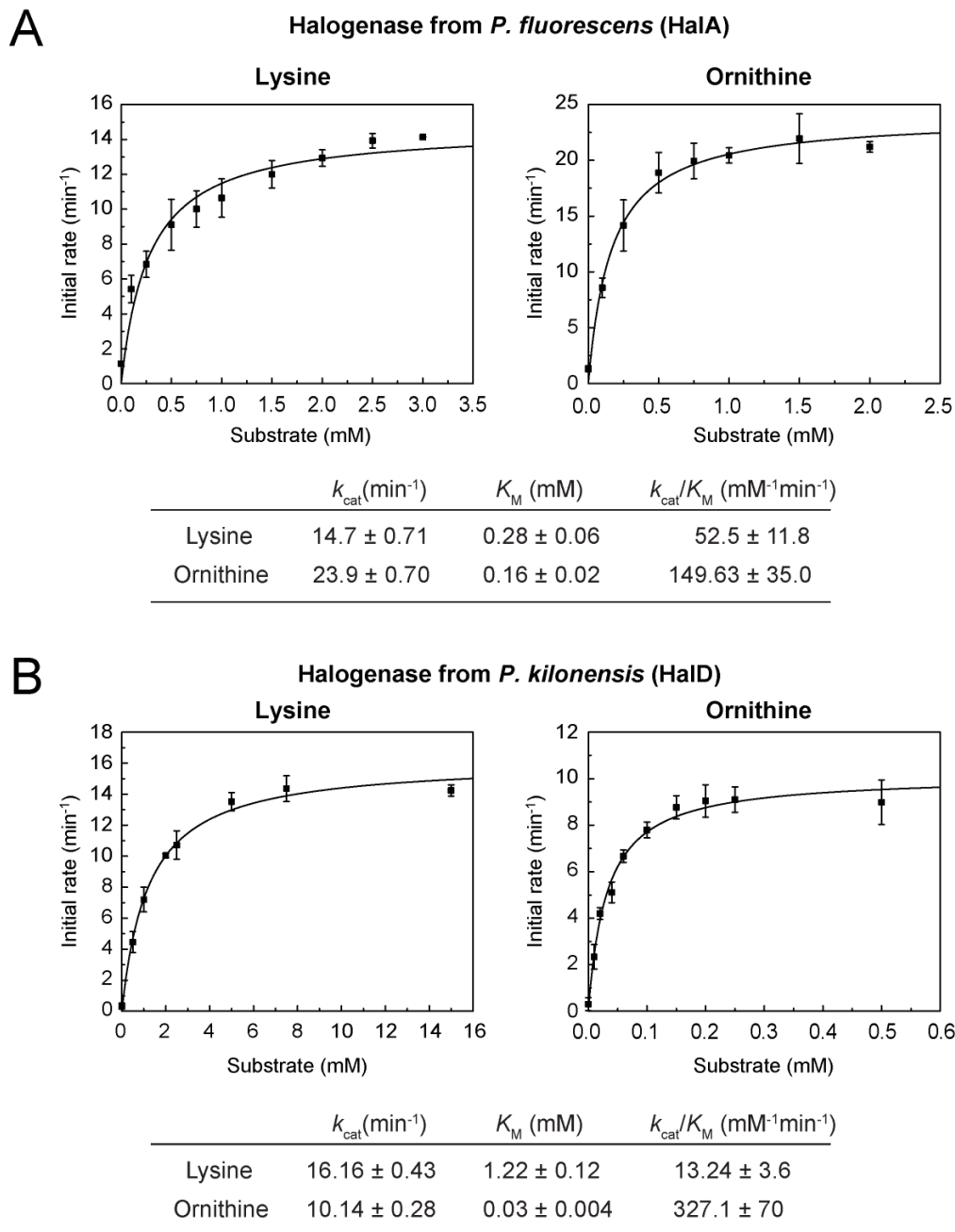


Figure 3.22. Ornithine halogenase kinetics. (A) Steady-state kinetic analysis for 4-Cl-lysine halogenase PfHalA with lysine or ornithine as substrates. Data are mean ± sd (n = 3 technical replicates). Table contains k_{cat} , K_M , and k_{cat}/K_M calculated by non-linear curve fitting to the Michaelis-Menten equation. Data in the table (kinetic parameters) are mean ± s.e. Error in k_{cat}/K_M is obtained by propagation from the individual kinetic terms. (B) Steady-state kinetic analysis for 4-Cl-ornithine halogenase (PkHalD) with lysine or ornithine as substrates. Data are mean ± sd (n = 3 technical replicates). Table contains k_{cat} , K_M , and k_{cat}/K_M calculated by non-linear curve fitting to the Michaelis-Menten equation. Data in the table (kinetic parameters) are mean ± s.e. Error in k_{cat}/K_M is obtained by propagation from the individual kinetic terms.

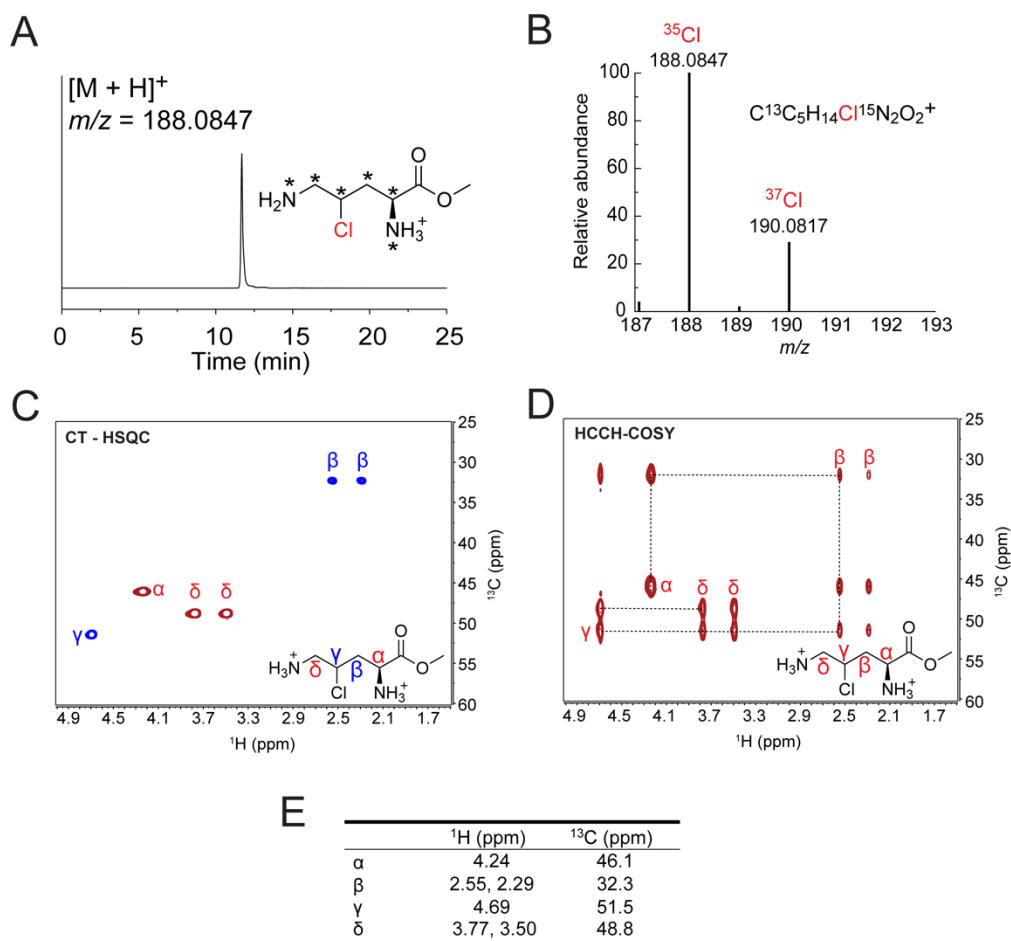


Figure 3.23. NMR characterization of 4-Cl-ornithine. (A) HalD from *P. kilonensis* (PkHalD) was incubated with fully ^{15}N - and ^{13}C -labeled L-ornithine, Fe, NaCl, α KG, and ascorbate for 60 min before quenching with methanolic HCl (3M) to yield the $[^{15}N_2, ^{13}C_5]$ -4-Cl-ornithine methyl ester product. The derivatized PkHalD product was then isolated by HPLC and characterized. (A) Extracted ion chromatograms of the $[^{15}N_2, ^{13}C_5]$ -4-Cl-ornithine methyl ester ($m/z = 188.0847$) produced by PkHalD. Purification of the compound was performed once. (B) Mass spectra show the characteristic Cl isotope pattern for $[^{15}N_2, ^{13}C_5]$ -4-Cl-ornithine. (C) 2D 1H - ^{13}C CT-HSQC was used to confirm assignment of C_α and C_δ of the $[^{15}N_2, ^{13}C_5]$ -4-Cl-ornithine methyl ester produced by PkHalD. In this experiment carbons with two neighbors (blue) have opposite phase of carbons with one or three neighbors (red). (D) 2D 1H - ^{13}C HCCH COSY was used to assign connectivity of carbons of the $[^{15}N_2, ^{13}C_5]$ -4-chloroornithine methyl ester produced by PkHalD. (E) Chemical shifts determined from the 2D-NMR data.

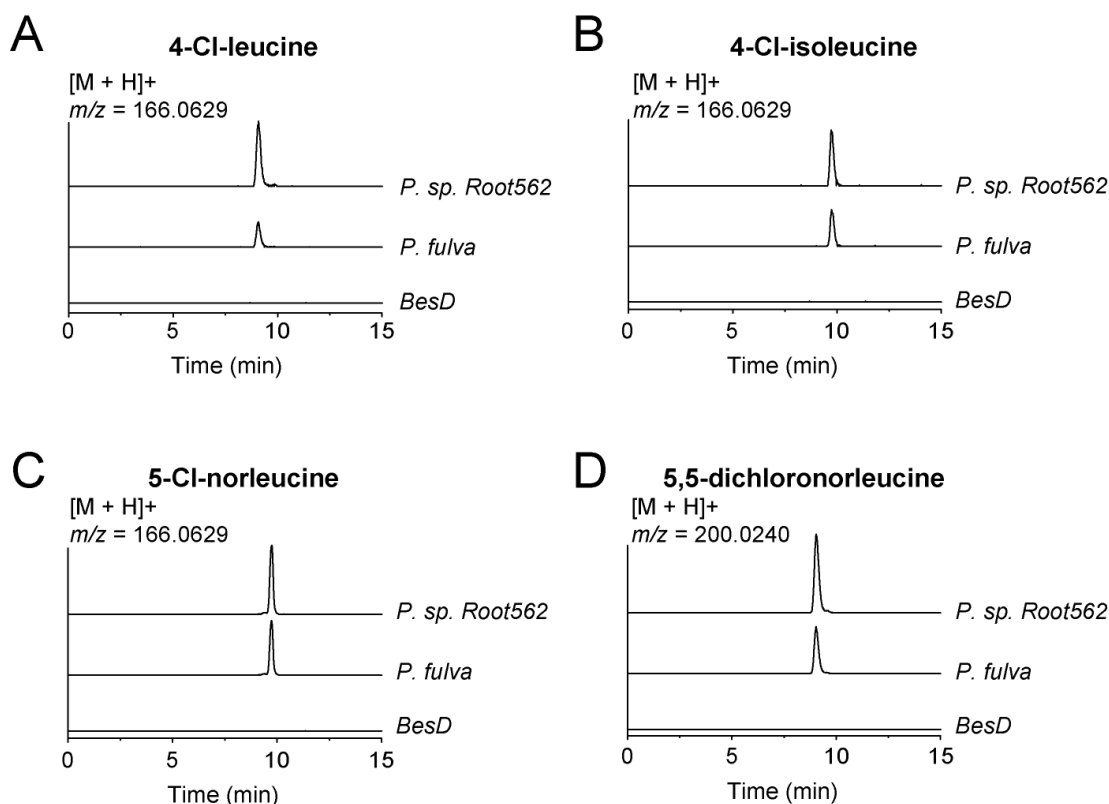
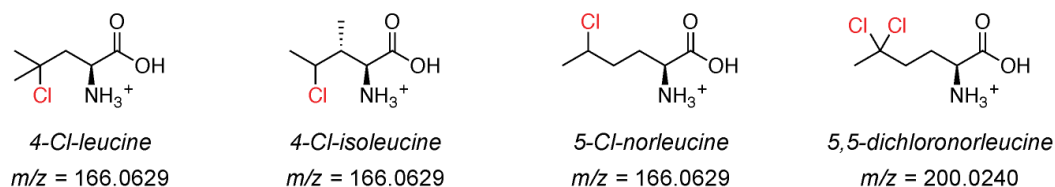


Figure 3.24. LC/MS analysis of products of aliphatic amino acid halogenases. (A) Extracted ion chromatograms of the 4-Cl-leucine ($m/z = 166.0629$) product observed upon incubation of *P. sp. Root562* HalE (PrHalE), *P. fulva* HalE (PfHalE), and *S. cattleya* BesD (ScBesD) with Fe, ascorbate, NaCl, α KG, and leucine for 90 min. (B) Extracted ion chromatograms of the 4-Cl-isoleucine ($m/z = 166.0629$) product observed upon incubation of *P. sp. Root562* HalE (PrHalE), *P. fulva* HalE (PfHalE), and *S. cattleya* BesD (ScBesD) with Fe, ascorbate, NaCl, α KG, and isoleucine for 90 min. (C, D) Extracted ion chromatograms of the 5-Cl-norleucine ($m/z = 166.0629$) and 5,5-dichloronorleucine ($m/z = 200.0240$) products observed upon incubation of *P. sp. Root562* HalE (PrHalE), *P. fulva* HalE (PfHalE), and *S. cattleya* BesD (ScBesD) with Fe, ascorbate, NaCl, α KG, and DL-norleucine for 90 min. All extracted ion chromatograms are representative of at least 3 experimental replicates.

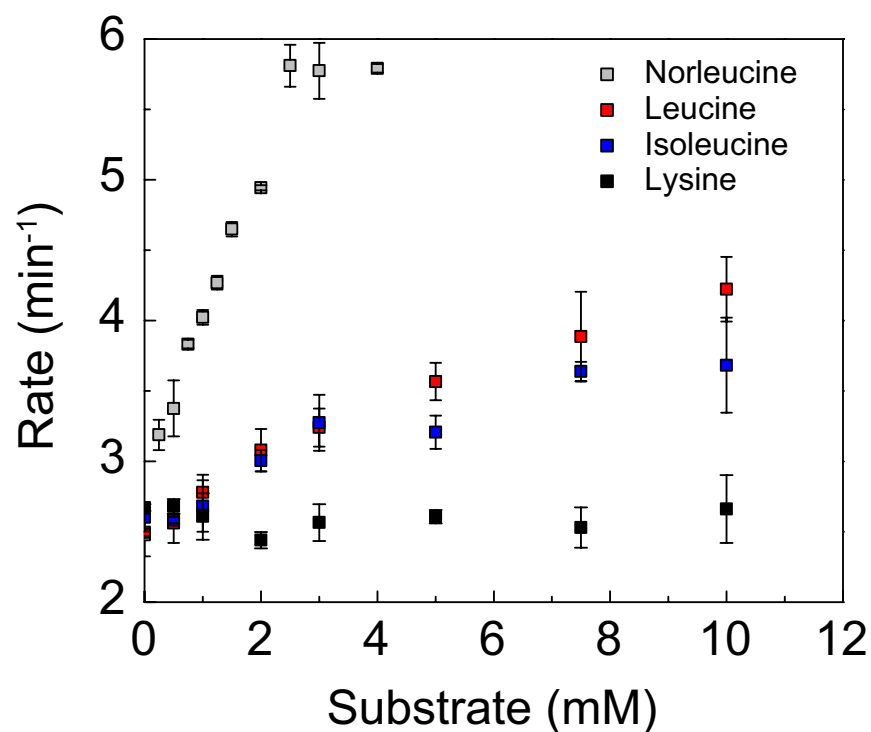


Figure 3.25. Kinetic analysis of aliphatic amino acids halogenase, PrHalE. Steady-state kinetic analysis for aliphatic amino acid halogenase PrHalE with norleucine, leucine, isoleucine, and lysine as substrates. A coupled reaction for monitoring halogenase activity through NADH oxidation was performed as in *Figure 3.22*. The rate of succinate formation by HalE from *P. sp. Root562* was monitored by the change in A_{340} using an NADH-coupled assay with succinyl-CoA synthetase (SCS), pyruvate kinase (PK) and lactate dehydrogenase. Data are mean $\pm$ sd ($n = 3$ technical replicates). Note that the enzyme has a relatively high rate of succinate formation ($\sim 2.5 \text{ min}^{-1}$) even in the absence of substrate when compared to PfHalA and PkHalD. This indicates a high rate of turnover of αKG to succinate without halogenation of the primary substrate, or “uncoupling,” which has been reported in $\text{Fe}^{\text{II}}/\alpha\text{KG}$ -dependent enzymes under certain conditions [54]. The rate of the enzyme increases with increasing concentration of norleucine, leucine, and isoleucine. However, the rate does not increase even when 10 mM lysine is added to the reactions.

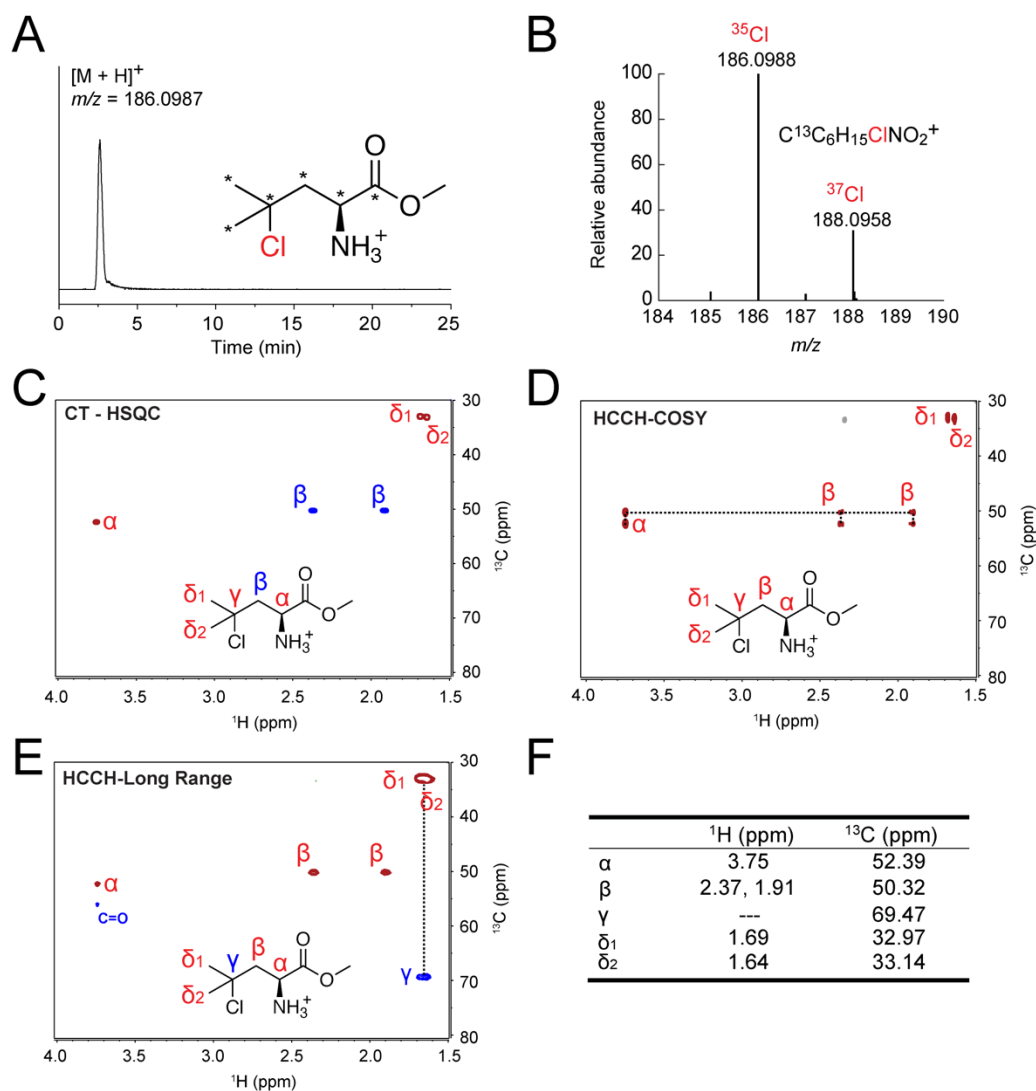


Figure 3.26. NMR analysis of 4-Cl-leucine. (A) HalE from *P. sp.* Root562 (PrHalE) was incubated with fully ^{13}C -labeled L-leucine, Fe, NaCl, and αKG for 90 min before quenching with methanolic HCl (3M) to yield the $[\text{C}_6]$ -4-chloroleucine methyl ester product. The derivatized product was then isolated by HPLC and characterized. Extracted ion chromatograms of the $[\text{C}_6]$ -4-Cl-leucine methyl ester ($m/z = 186.0987$) produced by PrHalE. Purification of the compound was performed once. (B) Mass spectra show the characteristic Cl isotope pattern for $[\text{C}_6]$ -4-Cl-leucine produced by PrHalE. (C) 2D ^1H - ^{13}C CT-HSQC was used to confirm assignment of C_β of the $[\text{C}_6]$ -4-Cl-leucine methyl ester. In this experiment carbons with two neighbors (blue) have opposite phase of carbons with one or three neighbors (red). (D) 2D ^1H - ^{13}C HCCH COSY was used to assign connectivity of carbons of the $[\text{C}_6]$ -4-Cl-leucine methyl ester produced by PrHalE. (E) A 2D ^1H - ^{13}C long-range HCCH was used to obtain the shift of C_γ using the adjacent H_δ proton shifts. (F) Chemical shifts determined from the 2D-NMR data.

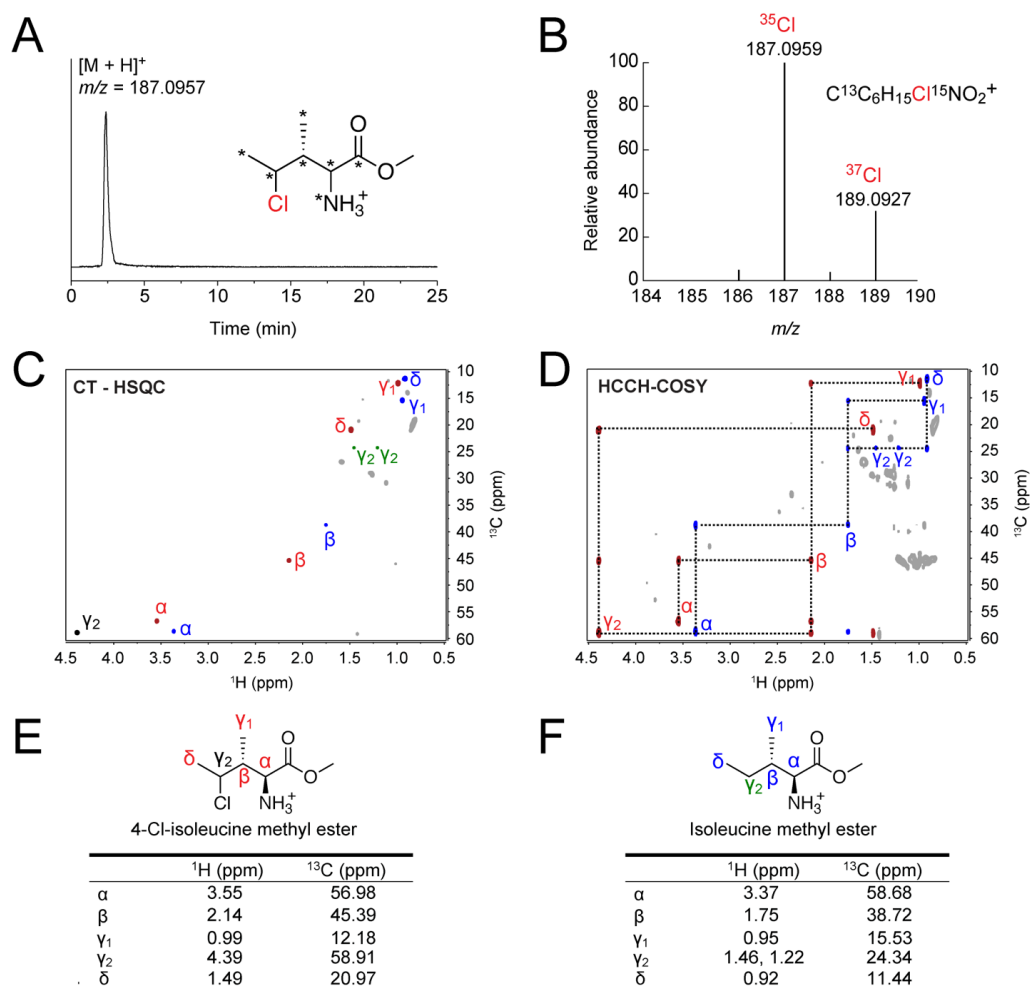


Figure 3.27. NMR analysis of 4-Cl-isoleucine. (A) HalE from *P. sp.* Root562 was incubated with fully ¹⁵N- and ¹³C-labeled L-isoleucine, Fe, NaCl, and αKG for 90 min before quenching with methanolic HCl (3M) to yield the [¹⁵N, ¹³C₆]-4-Cl-isoleucine methyl ester product. The derivatized product was then isolated by HPLC and characterized. Extracted ion chromatograms of the [¹⁵N, ¹³C₆]-4-Cl-isoleucine methyl ester (*m/z* = 187.0957) produced by HalE. Purification of the compound was performed once. (B) Mass spectra show the characteristic Cl isotope pattern for [¹⁵N, ¹³C₆]-4-Cl-isoleucine produced by HalE. Note that methyl esterified [¹⁵N, ¹³C₆]-isoleucine (the substrate) co-elutes with [¹⁵N, ¹³C₆]-4-Cl-isoleucine in the purification. (C) 2D ¹H-¹³C CT-HSQC was used to assign C_{γ2} of the [¹⁵N, ¹³C₆]-4-Cl-isoleucine and [¹⁵N, ¹³C₆]-isoleucine methyl esters. In this experiment carbons with two neighbors (black/green) have opposite phase of carbons with one or three neighbors (red/blue). (D) 2D ¹H-¹³C HCCH COSY was used to assign connectivity of carbons of the [¹⁵N, ¹³C₆]-4-Cl-isoleucine methyl ester produced by HalE and the residual [¹⁵N, ¹³C₆]-isoleucine methyl ester. (E, F) Structures and shifts obtained from the 2D-NMR data.

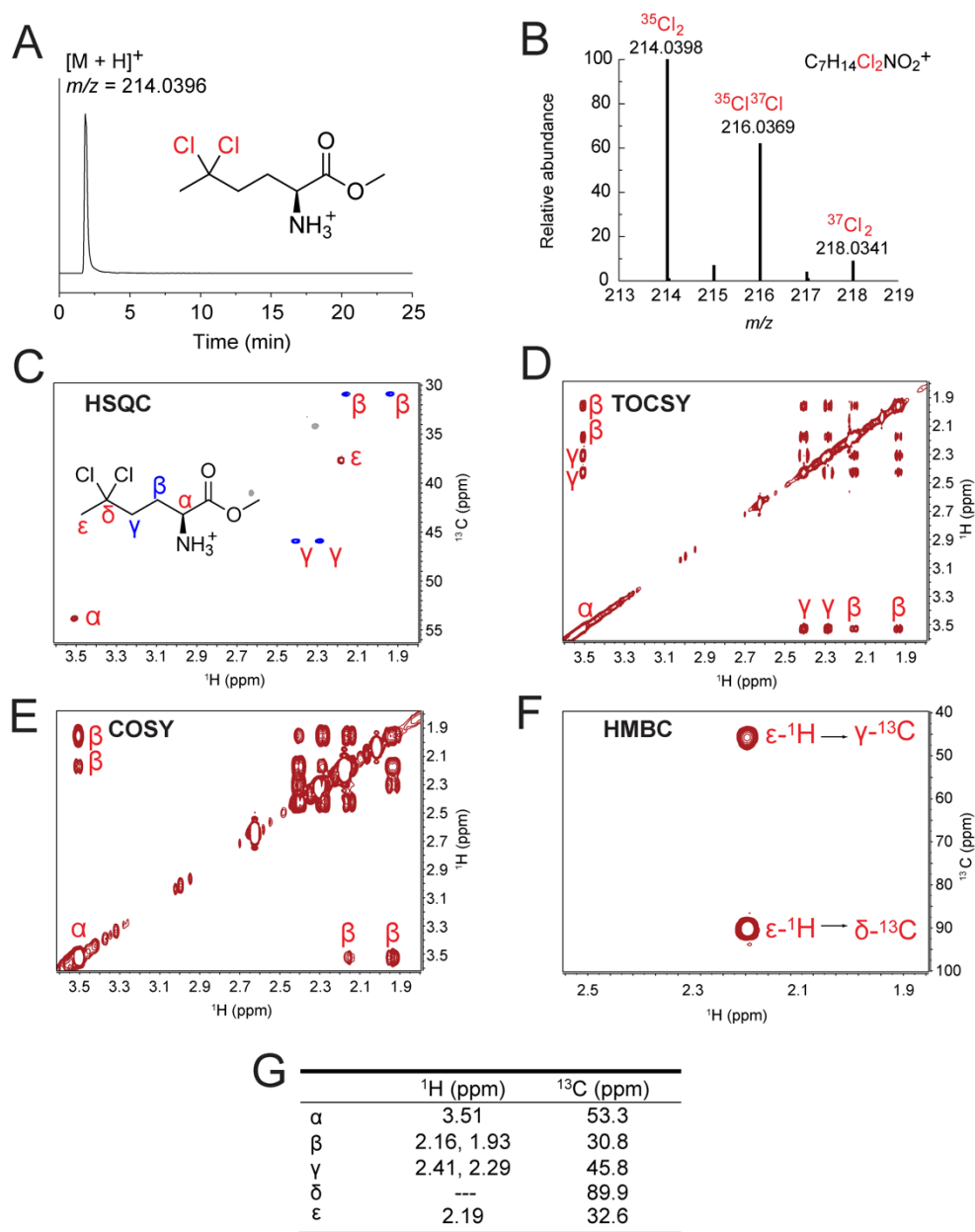


Figure 3.28. NMR analysis of 5,5-dichloronorleucine. HalE from *P. sp.* Root562 was incubated with DL-norleucine, Fe, NaCl, α KG, and ascorbate for 16 h to produce 5,5-dichloronorleucine before quenching the reaction and precipitating the protein with acetonitrile. The 5,5-dichloronorleucine product was isolated by HPLC, derivatized with 3M HCl in methanol to yield the 5,5-dichloronorleucine methyl ester and extracted into chloroform- D for LC/MS and NMR analysis. (A) Extracted ion chromatograms of the 5,5-dichloronorleucine methyl ester ($m/z = 214.0396$) produced by HalE. Purification of the compound was performed once. (B) Mass spectra show the characteristic Cl isotope pattern for 5,5-dichloronorleucine methyl ester. (C) 2D ^1H - ^{13}C CT-HSQC was used to confirm assignment of C_α and C_ϵ of the 5,5-dichloronorleucine methyl ester. In this experiment carbons with two neighbors (blue) have opposite phase of carbons with one or three neighbors (red). (D) 2D-TOCSY and (E) 2D ^1H - ^{13}C HCCH COSY were used to assign connectivity of carbons of the 5,5-dichloronorleucine methyl ester produced by HalE. (F) HMBC was used to obtain the chemical shift of the δ carbon, which is dichlorinated and lacks protons. (G) Chemical shifts obtained from the 2D-NMR data.

	1	10	20	30	40
<i>S. cattleya</i> (BesD)	MGS-----NRQEL---KDV	CAPLEKDDIRRLSQA	FHRFGIVTVTE--LIE	PHTRKLV	
<i>P. fluorescens</i>	MNY-----VLDEARLQ	EHHATNFESSVFALR	HEFARNGFIKVRN--	IVDDDLREKI	
<i>S. iranensis</i>	MTTNTPTTGVRTLDQDA	FDRHLKEFSENAPILE	YSQAFRRREGYVKLTG--	LVSADLFAEV	
<i>S. wuyuanensis</i>	MTENSITAA--NVEELIA	KNIAERFADDHEVLGL	SQHFRREGYVKLPG--	LVSPFVDAV	
<i>L. anisa</i>	MNT-----LNDTIS--	SYSNAFTA	KKIRKLREEFSRHGIV	KIPD--IVSEQ	LKQNI
<i>P. kilonensis</i>	MSEQTSSLVIEVMEQQL	AKHFQAILQDENRMKQ	IRNEFRDGYFNFKNFS	SFLPKRILENV	
<i>P. sp. Root562</i>	MGVV-----SADQAH	LEILNSNPYETRDRFF	WEQHRFARDGYLKINH--	LIDPLVRDAI	
	*	:	.	* * * ...	:: :
	50	60	70	80	90
<i>S. cattleya</i> (BesD)	RAEADRLLDQYAERRDL	RLLATTDYTR	RSMSVVPSETIAANSEL	VTGLYAHRELLAPLEAI	
<i>P. fluorescens</i>	TREVNSLIDRQLERDL	LHLATTDNTP	RYMSVVRSEFIAENSTL	INTLSKSKGLETLSQI	
<i>S. iranensis</i>	VSEVGDLLDQHAQRDI	ELKETGNSP	RRMHTVSAADIARDSRL	ISAIYDSASIRDTLGRI	
<i>S. wuyuanensis</i>	AAETHQLIDTHQKRIDI	RKETGDSF	RYMSTVGQKAIATDGL	IPAVYESTALKGFLSRL	
<i>L. anisa</i>	HNEIYGLLENYAERRDL	LLKTTDNTF	RYLSVVRSELIEEQSKM	IREVRES	DALLQFLEKI
<i>P. kilonensis</i>	HAEVHALLDEYSVRRDV	TPSTGNTY	RKMYNVNQPEIAEGGT	FIPALYQSES	LRKFLGNI
<i>P. sp. Root562</i>	MVDVKYLMAHHAKRRD	FLIESTGNTP	RYLSNVSRDNIEEFGR	YIPLAYKSQY	LSLIGNI
	:	*:	* * . :	* . :	:
	110	120	130	140	150
<i>S. cattleya</i> (BesD)	AGERLHPCPKADEEFL	ITRQEQRGDT	HGWHWGD	FSFALI	WVLQAP-PIDV
<i>P. fluorescens</i>	AGTQLIASVSKDEEY	LITKQERKGD	HGWHWGD	YSFALI	WIIETP-SIAK
<i>S. iranensis</i>	AQADVRCPCWEGE	KYVIIRQDQPGD	HGWHWGD	FSYTLI	WIIQAP-GPEV
<i>S. wuyuanensis</i>	AKKEVMGCPWDEE	KYIITRQHQKGD	HGWHWGD	FSFTVI	WLEAP-SLEY
<i>L. anisa</i>	TQEKLLTHVKEDE	EKFVITKQEFSGD	HGWHWGD	YSFALI	WIVEMP-PVDC
<i>P. kilonensis</i>	AGDDLASC-WEQE	QYLVTKLSHPGD	HGWHWGD	YPYTM	WIIIEAPDP
<i>P. sp. Root562</i>	VKEEVIPTPWKWD	NYIINSQHKSGD	HGWHWGD	YPYTI	WIVEAP-SIECG
	.	:	*****:	*****:	*
	170	180	190	200	210
<i>S. cattleya</i> (BesD)	TWDKASPQINRYLVEN	PIDTYHFESGDVYFL	RDTTLHRTIPL	REDTTRIILN	MTWAGER
<i>P. fluorescens</i>	SWDKSNPRIHELLCSN	PIATYGFVTGDIYFL	RDTTLHRTIPL	NEDATRIILN	MTWAAEK
<i>S. iranensis</i>	NWDKQNPVRVHEYLQ	NHPRTYANATGDLYFL	RSDTTLHRTIPL	SRPATRIILN	TCWASTA
<i>S. wuyuanensis</i>	DWNKDDPRVEDYLQ	KHPIRSYGHAKGDL	YLLRSDTTLHRTV	PLNADRTRIIL	NTCWASRA
<i>L. anisa</i>	SWDKEAPDLYQYFC	DNPIYTHSFVPGDV	YLLKADTTLHRTV	PLRYDATRIM	LNMTWGSIR
<i>P. kilonensis</i>	EWDKQNPQIWQYIL	NNPIKSYHHLKGDV	YFLKSDTTLHVV	PIQGETTRIIL	NTCWASAH
<i>P. sp. Root562</i>	NWDKKNPRVENLLV	ENSIDSYFHTG	DVYLLKADTTLH	RVKPLTKDVQRI	ILNTTWERAR
	:	* :	: . : *	::	*****: *
	230	240			
<i>S. cattleya</i> (BesD)	DLSRKLAAD--DR	WWDNAEVSAA	RAIKD		
<i>P. fluorescens</i>	DLSRN-LHGN-DR	WWEHQHVEAA	KSILT-		
<i>S. iranensis</i>	DVERPTTHETMDA	MFSS-----			
<i>S. wuyuanensis</i>	DQQKATTHETMNA	MF-----			
<i>L. anisa</i>	QDNPFSLVD--DR	WWDNLNAQAAL	HVHN		
<i>P. kilonensis</i>	DRRTDVAHESIEV	IWDTKARTTEA	---		
<i>P. sp. Root562</i>	DRDRHVEHET--F	VFRD-----			
	*	:			

Figure 3.29. Sequence alignment of representative halogenases A-E. Multiple sequence alignment of BesD and representative homologs from this study. Included are *S. cattleya* BesD (4-Cl-lysine halogenase, BesD), *P. fluorescens* HalA (4-Cl-lysine), *S. iranensis* HalB (5-Cl-lysine and 5,5-dichlorolysine), *S. wuyuanensis* HalB (5-Cl-lysine and 5,5-dichlorolysine), *L. anisa* HalC (4-Cl-lysine and 4,4-dichlorolysine), *P. kilonensis* HalD (4-Cl-ornithine), *P. sp. Root562* HalE (4-Cl-leucine, 4-Cl-isoleucine, and 5,5-dichloronorleucine). The alignment was performed with MUSCLE [2]. Residues investigated in this study are colored in red.

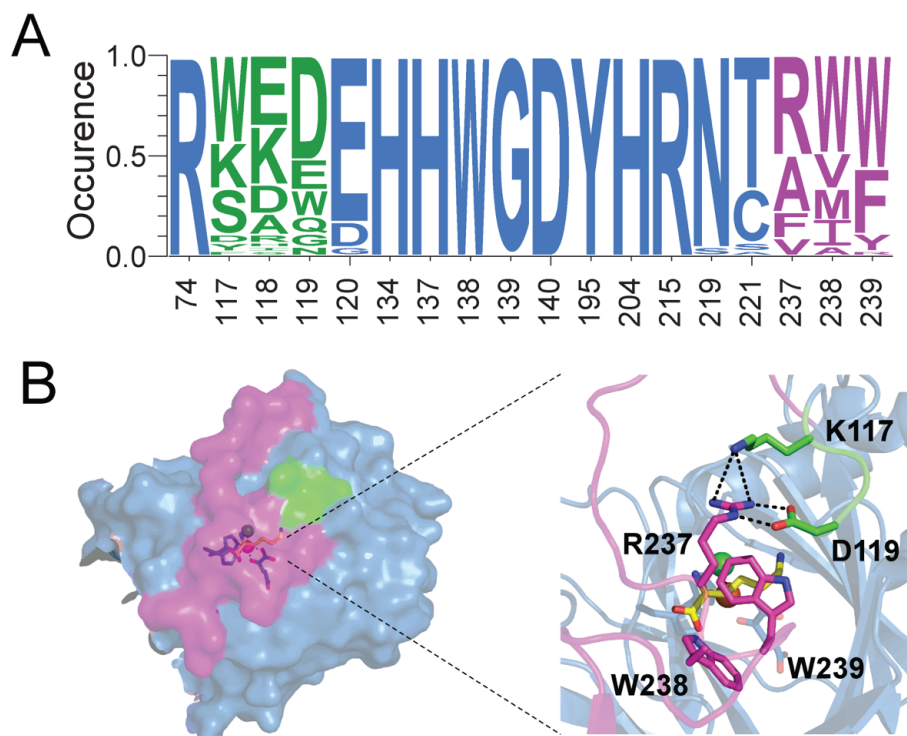


Figure 3.30. Variable C-terminus of halogenases. Of all the substrate-contacting residues W238 and W239 are the most variable among halogenase homologs of BesD. (A) W238 and W239 are located on the C-terminus of the protein (magenta), which closes over the active site of BesD. (B) The C-termini of BesD homologs is highly variable, differing by as much as 10 amino acids in length. R237 on the C-terminus of BesD contacts K117 and D119 of the internal loop of BesD (green), which is also variable among the halogenases. The variable C terminus and internal loop may be key factors in determining halogenase substrate specificity.

biosynthetic transformations. We first examined the ability to expand the repertoire of the amino acid halogenases to carry out bromination and azidation by replacement of the Cl^- ligand with Br^- and N_3^- , respectively [55]. When NaCl was replaced with NaBr and NaN_3 , SwHalB catalyzed the corresponding bromination and azidation reactions as observed by LC/MS (Figure 3.31). Thus, a variety of functional handles can be installed onto amino acid substrates to diversify amino acid building blocks using SwHalB simply by providing an alternate ligand for the Fe complex. In particular, bromine and azide functional groups serve as handles for substitution reactions [56] and as precursors to amines [57] and nitrogen containing heterocycles [58], respectively.

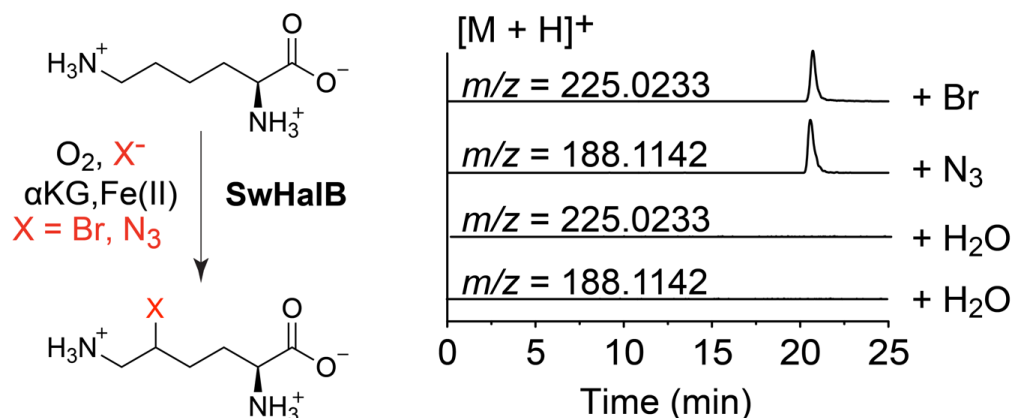


Figure 3.31. Bromination and azidation can be carried out with BesD halogenases. Bromolysine ($m/z = 225.0233$) and azidolysine ($m/z = 188.1142$) are produced by SwHalB using NaBr or NaN_3 , respectively. Extracted ion chromatograms are representative of at least 3 experimental replicates.

Amino acids also participate in a broad range of metabolic pathways and can be directly transformed into a variety of new compound classes. For example, heterocycles are key structural components in a number of bioactive molecules and drugs [59] and can be made by oxidative cyclization by lysine cyclodeaminase (RapL) [60] and ornithine cyclodeaminase (OCD) [61] to form pipecolate and proline, respectively. To test whether the cyclodeaminase enzymes were sufficiently promiscuous to generate chlorinated heterocycles, we cloned, expressed, and purified RapL and OCD (Figure 3.32). Coupled reactions with lysine or ornithine halogenases and RapL or OCD, respectively, yielded chlorinated compounds with the predicted masses and isotope patterns of chloropipecolate, dichloropipecolate, and chloroproline (Figure 3.33-3.34). Thus, the newly discovered amino acid halogenases can be further enzymatically elaborated to form 5- and 6- membered chlorinated heterocycles.

Due to the utility of diamines as precursors for polymers such as Nylon [62], we next sought to access chlorinated diamines through coupled enzymatic reactions. Lysine decarboxylase (LDC) catalyzes the PLP-dependent decarboxylation of lysine to 1,5-pentanediamine [63]. LDC from *E. coli* was heterologously expressed and purified to evaluate its tolerance of a chlorinated substrate. To generate a chlorinated diamine, lysine decarboxylase was added to a reaction following the enzymatic synthesis of 5,5-dichlorolysine by SwHalB. A mass consistent with a dichlorinated diamine was observed by LCMS (Figure 3.35). We additionally accessed chlorinated α -keto acids through IlvE, an aliphatic amino acid amino transferase. IlvE from *E. coli* has been shown to catalyze transamination of leucine, isoleucine, and norleucine into their corresponding α -keto acids [64]. Incubation of enzymatically synthesized chloronorleucine with IlvE yielded products with the predicted mass and isotope pattern of chlorinated α -ketoacids as observed by LC/MS (Figure

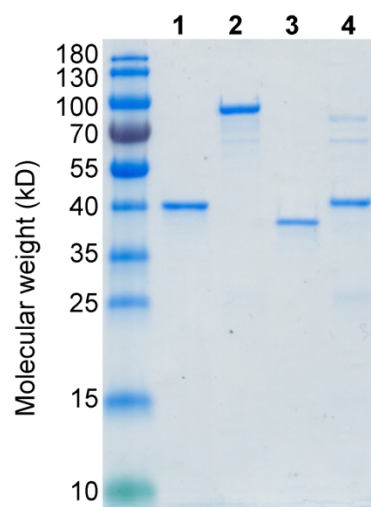


Figure 3.32. SDS-PAGE of amino acid metabolism proteins ornithine cyclodeaminase (OCD), lysine decarboxylase (LDC), aliphatic amino acid aminotransferase (IlvE), and lysine cyclodeaminase (RapL). Lane **1**, OCD; **2**, LDC; **3**, IlvE; **4**, RapL. Protein purification was performed one time.

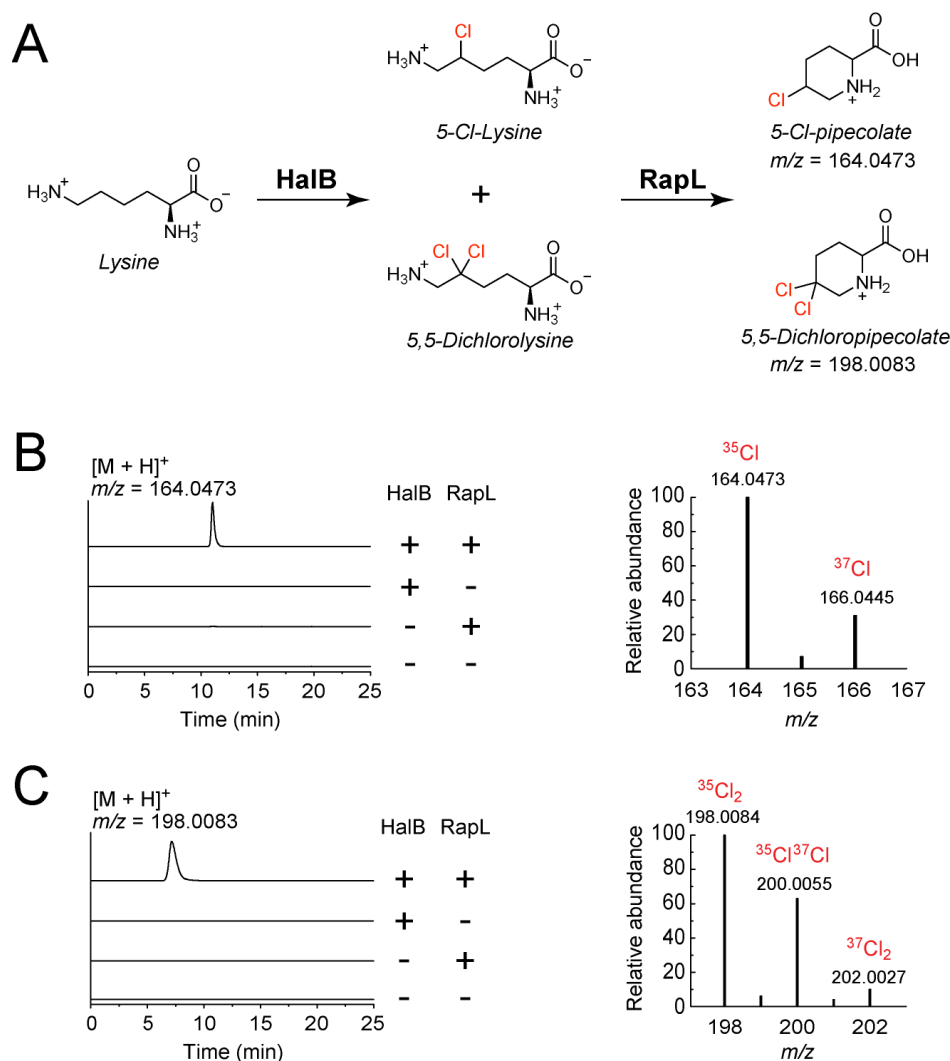


Figure 3.33. *In vitro* formation of chlorinated heterocycles by lysine cyclodeaminase (RapL). (A) Reaction scheme for the production of 5-Cl-pipicolate and 5,5-dichloropipicolate by HalB from *S. wuyuanensis* (SwHalB) and RapL from *S. hygrosopicus*. SwHalB was incubated with lysine, Fe, NaCl, α KG, and ascorbate for 90 min to form 5-Cl-lysine and 5,5-dichlorolysine before addition of RapL and NAD⁺ to the reaction mixture. Expected products are drawn to represent the [M+H]⁺ mass detected in positive scan mode on the QTOF. After 60 min, reactions were quenched by addition of 2 vol of 1% formic acid in methanol. (B) Extracted ion chromatograms and mass spectra for the formation of 5-Cl-pipicolate ($m/z = 164.0473$) upon addition of RapL to 5-chlorolysine formed by SwHalB. (C) Extracted ion chromatograms and mass spectra for the formation of 5,5-dichloropipicolate ($m/z = 198.0083$) upon addition of RapL to 5,5-dichlorolysine formed by SwHalB. All data are representative of $n = 3$ experimental replicates.

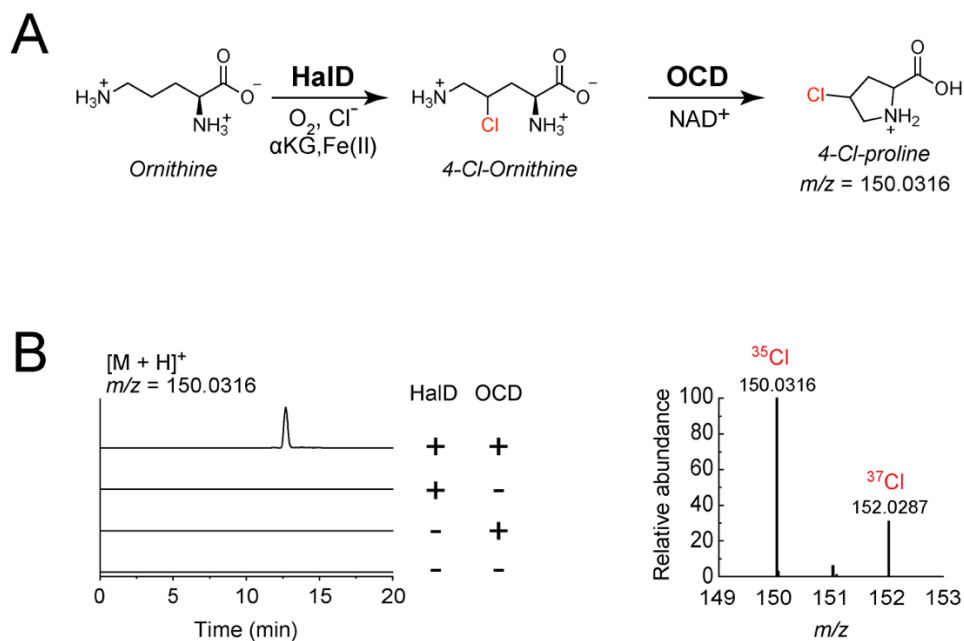


Figure 3.34. *In vitro* formation of chlorinated heterocycles by ornithine cyclodeaminase (OCD). (A) Reaction scheme for the production of 4-Cl-proline from ornithine via a coupled reaction with purified HalD from *P. kilonensis* (PkHalD) and OCD from *E. coli*. PkHalD was incubated with ornithine, Fe, NaCl, α KG, and ascorbate for 90 min to form 4-Cl-ornithine before addition of OCD and NAD^+ to the reaction mixture. After 60 min, reactions were quenched by addition of 2 vol of 1% formic acid in methanol. Expected products are drawn to represent the $[\text{M}+\text{H}]^+$ mass detected in positive scan mode on the QTOF. (B) Extracted ion chromatograms and mass spectra for the formation of 4-Cl-proline ($m/z = 150.0316$) upon addition of OCD to 4-Cl-ornithine formed by PkHalD. All data are representative of $n = 3$ experimental replicates.

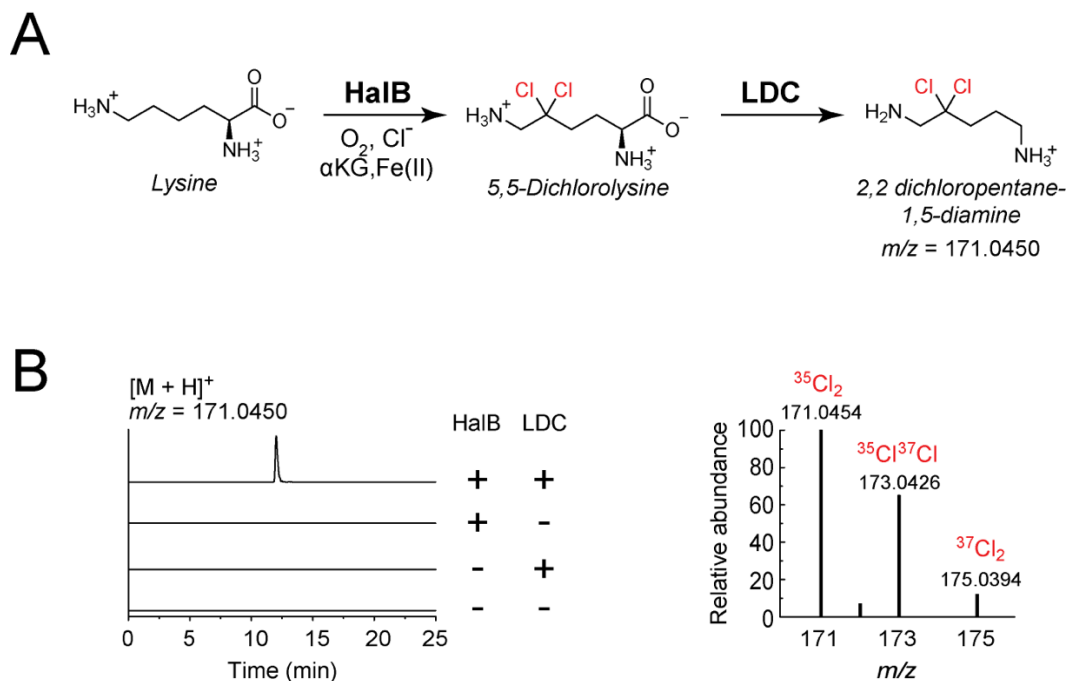


Figure 3.35. *In vitro* chlorinated diamine formation by lysine decarboxylase (LDC). (A) Reaction scheme for the production of 2,2 dichloropentane-1,5-diamine from lysine via a coupled reaction with HalB purified from *S. wuyuanensis* (SwHalB) and LDC from *E. coli*. SwHalB was incubated with lysine, Fe, NaCl, α KG, and ascorbate for 90 min to form 5,5-dichlorolysine before addition of LDC to the reaction mixture. After 60 min, reactions were quenched by addition of 2 vol of 1% formic acid in methanol. Expected products are drawn to represent the $[M+H]^+$ mass detected in positive scan mode on the QTOF. (B) Extracted ion chromatograms for the formation of 2,2 dichloropentane-1,5-diamine ($m/z = 171.0450$) upon addition of LDC to 5,5-dichlorolysine formed by SwHalB. Mass spectra show the characteristic chlorinated isotope pattern for 2,2 dichloropentane-1,5-diamine. All data are representative of $n = 3$ experimental replicates.

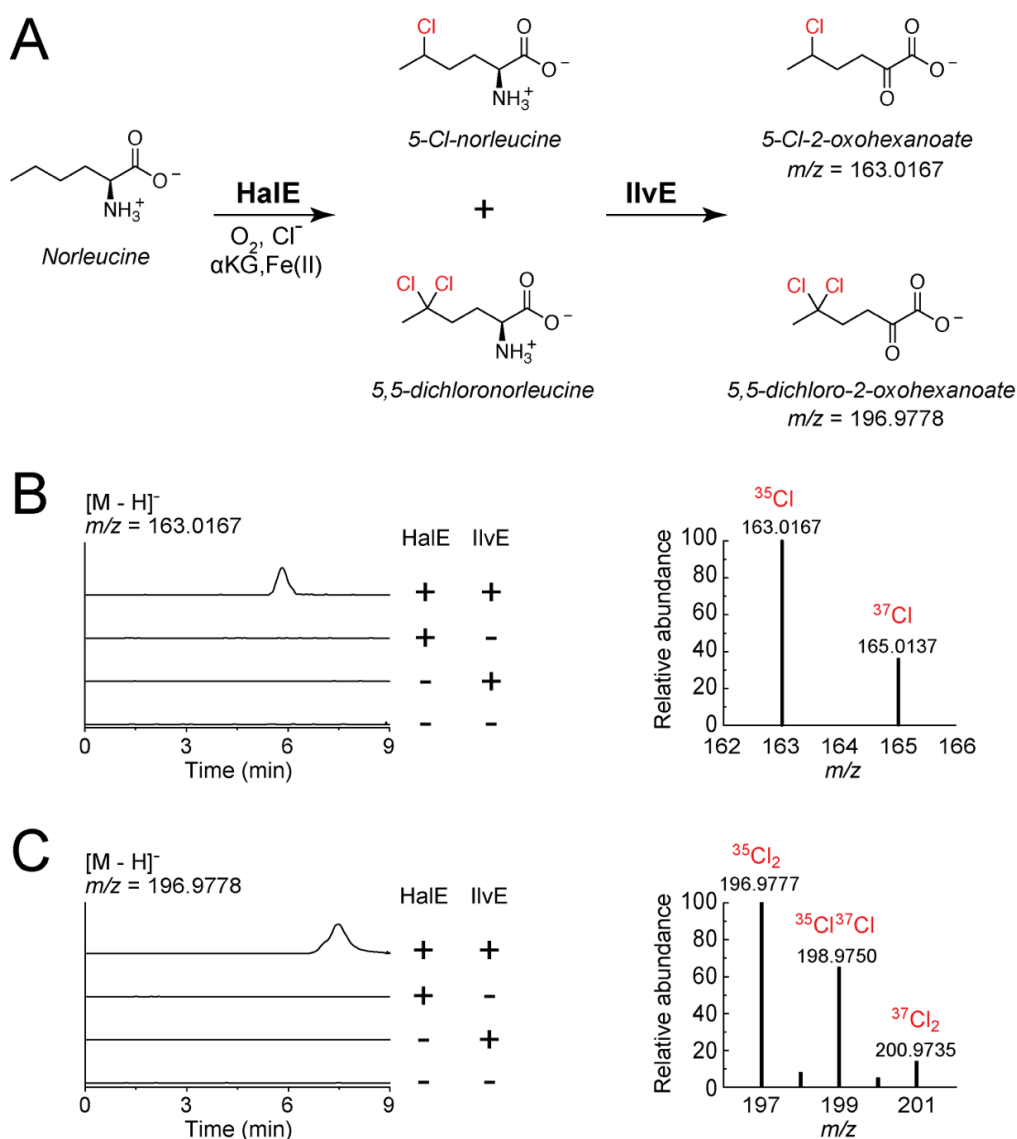


Figure 3.36. *In vitro* α -keto acid formation by the aliphatic amino acid aminotransferase, IlvE. (A) Reaction scheme for the production of the chlorinated α -keto acids from norleucine via a coupled reaction with HaIE purified from *P. sp.* Root562 (PrHaIE) and IlvE from *E. coli*. PrHaIE was incubated with DL-norleucine, Fe, NaCl, α KG, and ascorbate for 90 min before addition of IlvE to the reaction mixture. After 60 min, reactions were quenched by addition of 2 vol of 1% formic acid in methanol. Expected products are drawn to represent the [M-H]⁻ mass detected in negative scan mode on the QTOF. (B) Extracted ion chromatograms for 5-Cl-2-oxohexanoate ($m/z = 163.0167$) product formation by PrHaIE and IlvE. (C) 5,5-dichloro-2-oxohexanoate ($m/z = 196.9778$) product formation by PrHaIE and IlvE. All data are representative of $n = 3$ experimental replicates.

3.36). Finally, we investigated whether chlorolysine synthesized by BesD homologs could be incorporated into peptides with the *in vitro* transcription/translation (IVTT) system [65]. Following incubation of amino acids with PfHalA or SwHalB, transcription and translational machinery from the PURExpress IVTT kit, along with a plasmid encoding a 9-amino acid peptide (METRSKNML, PepC) were added to the reaction mixture. When the amino acids were pre-incubated with either halogenase, a product corresponding to the exact mass of the chlorinated peptide was detected by LC/MS (*Figure 3.37*). Therefore, the halogenases from the BesD family can be coupled with IVTT systems to generate chlorinated peptides from 4-Cl-lysine and 5-Cl-lysine.

The METRSKNML peptide (PepC) sequence has been shown to adopt an alpha helical structure and encodes glutamate and lysine at the *i* and *i*+4 positions of the helix [66]. We therefore wondered whether the presence of chlorolysine in the peptide might facilitate peptide stapling through nucleophilic attack by the encoded glutamate (*Figure 3.38*). We found that when 5-Cl-lysine was incorporated into PepC, a mass consistent with a stapled peptide was observed by LC/MS. In contrast, when 4-Cl-lysine was incorporated, the stapled peptide mass was not observed. This could be due to differences in accessibility of the chlorine in 5-Cl-lysine compared to 4-Cl-lysine; glutamate may not be able to access chlorine in 4-Cl-lysine within the constraints of a helical peptide.

Finally, we compared results for incorporating Cl-lysine into two different peptides (*Figure 3.39*). PepC (METRSKNML) has an internal lysine. We also designed PepA (MSERSSNMK), which has a lysine residue at the C-terminus. We found that the 4-Cl-lysine peptide could not be detected when lysine was encoded at the terminal position in PepA. This could be due to the fact that the carboxylate of lysine is free to undergo nucleophilic attack on chlorine and cyclize in a manner similar to what is observed in free 4-Cl-lysine. In contrast, the carboxylate from the internal 4-Cl-lysine in PepC is protected from cyclization. Therefore, 4-Cl-lysine can be stabilized through incorporation into peptides. Taken together, this work raises the possibility that chlorinated, brominated, and azidated amino acids may be incorporated into peptides through coupling of BesD halogenases with IVTT.

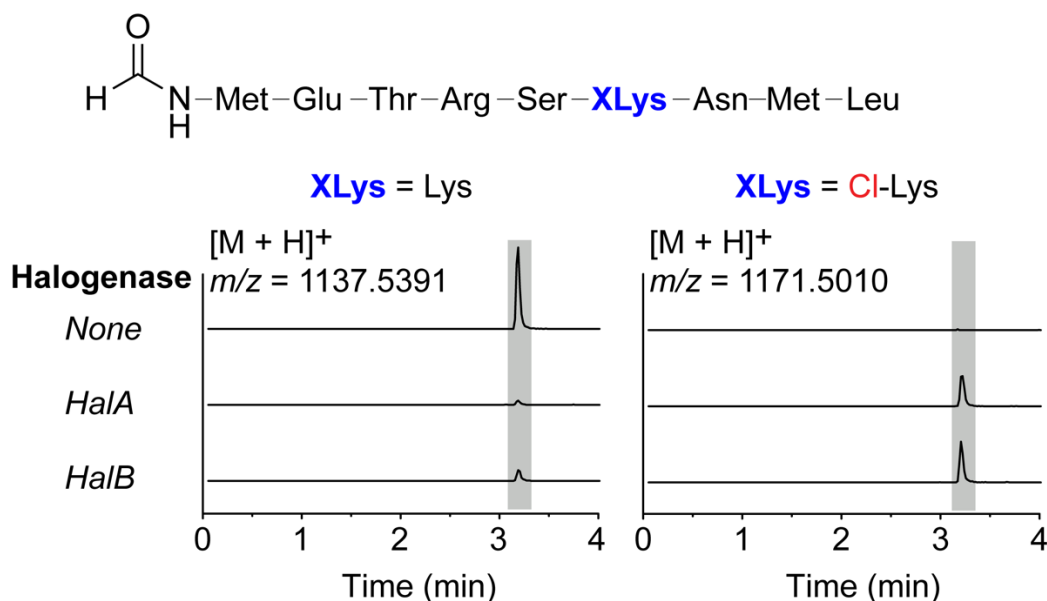


Figure 3.37. PfHalA or SwHalB could be used to generate Cl-lysine *in situ* followed by *in vitro* transcription/translation (IVTT) initiated by addition of PURExpress kit components along with a plasmid encoding the METRSKNML peptide. Extracted ion chromatogram shows the production of peptides containing lysine ($m/z = 1137.5391$) or Cl-lysine ($m/z = 1171.5010$) and are representative of 3 experimental replicates.

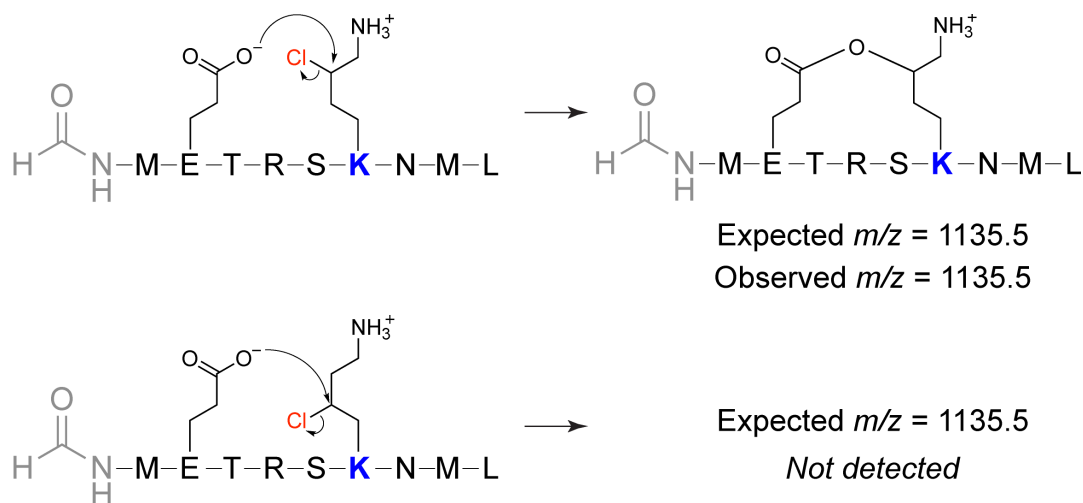
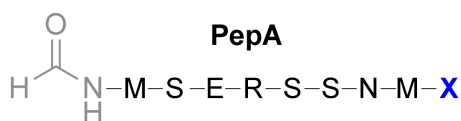
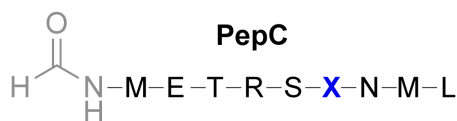


Figure 3.38. Incorporation of 5-Cl-lysine (generated by HalB) into the METRSKNML peptide leads to the production of a mass consistent with the i+4 stapled peptide product. HalA or HalB was used to generate Cl-lysine *in situ* followed by *in vitro* transcription/translation (IVTT) initiated by addition of PURExpress kit components along with a plasmid encoding the METRSKNML peptide. Extracted ion chromatograms showed the production of a compound consistent with the mass expected from nucleophilic attack on chlorolysine from 5-Cl-lysine ($m/z = 1135.5$), but not from 4-Cl-lysine.



X	[M + H] ⁺
Lys	1097.5
OH-Lys	1113.5
Cl-Lys	1131.4
(2Cl)-Lys	1165.4



X	[M + H] ⁺
Lys	1137.5
OH-Lys	1153.5
Cl-Lys	1171.5
(2Cl)-Lys	1205.5

Condition	Observed <i>m/z</i>	Difference	Modification
No plasmid	None	NA	NA
PepA + Lys	1097.5	0.0	***
PepA + 5-OH-Lys	1113.5	16.0	+OH
PepA + HalA (4-Cl-Lys)	1095.5, 1113.5	-2.0, 16.0	cyclo, +OH
PepA + HalB (5-Cl-Lys)	1131.4, 1165.4	34.0, 16.0	+Cl, +2Cl
PepA + ¹³ C ₆ ¹⁵ N ₂ -5-Cl-Lys	1139.5	42.0	+Cl+ isotope
PepC + Lys	1137.5	0.00	***
PepC + 5-OH-Lys	1153.5	15.99	+OH
PepC + HalA (4-Cl-Lys)	1171.5, 1153.5	34.0, 16.0	+Cl, +OH
PepC + HalB (5-Cl-Lys)	1171.5, 1205.5, 1135.5	34.0, 16.0, -2.0	+Cl, +2Cl, cyclo
PepC + ¹³ C ₆ ¹⁵ N ₂ -5-Cl-Lys	1179.5	42.0	+Cl + isotope

Figure 3.39. Summary of results from experiments to incorporate lysine derivatives into peptides. PfHalA or SwHalB could be used to generate Cl-lysine *in situ* followed by *in vitro* transcription/translation (IVTT) initiated by addition of PURExpress kit components along with a plasmid encoding MSERSSNMK (PepA) or METRSKNML (PepC). The difference between the mass of the modified peptide and the unmodified lysine peptide is reported in the table. As controls, purified 5-OH-lysine (Sigma Aldrich) or ¹³C₆¹⁵N₂-5-Cl-lysine (made by HalB) were added to the IVTT reactions.

3.4. Conclusion

BesD directly chlorinates the free amino acid lysine. The simple substrate of BesD makes it a promising system for investigating the mechanism of Fe^{II}/αKG-dependent halogenases. Toward this goal, we have identified several new enzymes within this family and solved the crystal structure of BesD. Characterization of the radical halogenases in the BesD family has shown that these enzymes accept both polar and non-polar amino acids, demonstrate regioselective halogenation of C_{sp3}-H bonds, as well as carry out mono- and di-chlorination. Structural studies of BesD further provide valuable insight into its substrate selectivity and the potential mechanisms by which reaction partitioning between halogenation and hydroxylation is achieved.

Radical amino acid halogenases also have the potential to enable the production of diverse compounds containing useful functional handles. In addition to showing that bromination and azidation are catalyzed by the BesD family, we have demonstrated that downstream enzymatic pathways can be constructed to utilize halogenase-modified intermediates to produce different classes of compounds, including nitrogen heterocycles, diamines, keto acids, and ribosomally-synthesized peptides. Altogether, these results highlight the potential to integrate this family of halogenases directly into engineered biological reaction pathways and expand our ability to access a wide variety of new compounds (*Figure 3.40*).

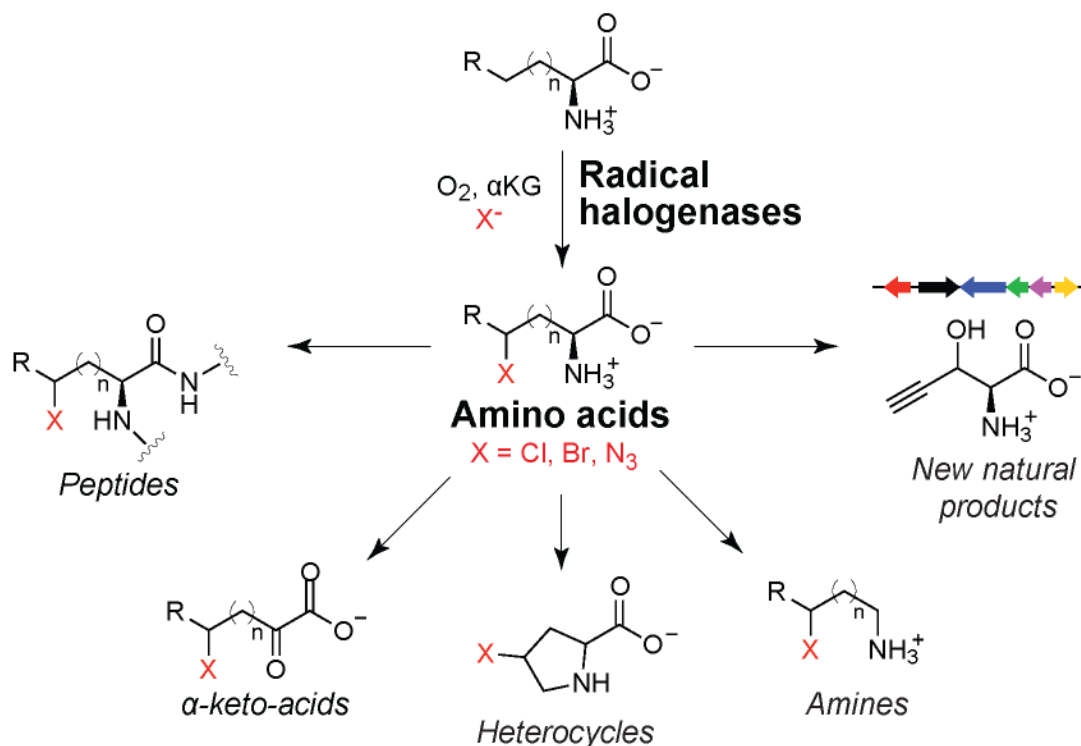


Figure 3.40. Radical halogenases can be used to diversify amino acids, which can then be incorporated into downstream pathways to make a range of products with valuable functional groups.

3.5. References

1. Koeller, K. M. & Wong, C.-H. Enzymes for chemical synthesis. *Nature* **409**, 232–240 (2001) doi:10.1038/35051706.
2. Kan, S. B. J., Huang, X., Gumulya, Y., Chen, K. & Arnold, F. H. Genetically programmed chiral organoborane synthesis. *Nature* **552**, 132 (2017) doi:10.1038/nature24996.
3. Savile, C. K., Janey, J. M., Mundorff, E. C., Moore, J. C., Tam, S., Jarvis, W. R., Colbeck, J. C., Krebber, A., Fleitz, F. J., Brands, J., Devine, P. N., Huisman, G. W. & Hughes, G. J. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **329**, 305–309 (2010) doi:10.1126/science.1188934.
4. Harris, C., Kannan Rajamoorthi, Kopecka, H. & Harris, T. The Role of the Chlorine Substituent in the Antibiotic Vancomycin: Preparation and Characterization of Mono- and Didechlorovancomycin. *J. Am. Chem. Soc.* **107**, 6652–6658 (1985) doi:10.1021/ja00309a038.
5. Groll, M., Huber, R. & Potts, B. C. M. Crystal structures of Salinosporamide A (NPI-0052) and B (NPI-0047) in complex with the 20S proteasome reveal important consequences of beta-lactone ring opening and a mechanism for irreversible binding. *J. Am. Chem. Soc.* **128**, 5136–5141 (2006) doi:10.1021/ja058320b.
6. Latham, J., Brandenburger, E., Shepherd, S. A., Menon, B. R. K. & Micklefield, J. Development of Halogenase Enzymes for Use in Synthesis. *Chem. Rev.* **118**, 232–269 (2018) doi:10.1021/acs.chemrev.7b00032.
7. Gkotsi, D. S., Dhaliwal, J., McLachlan, M. M., Mulholland, K. R. & Goss, R. J. Halogenases: powerful tools for biocatalysis (mechanisms, applications, and scope). *Curr. Opin. Chem. Biol.* **43**, 119–126 (2018) doi:10.1016/j.cbpa.2018.01.002.
8. Neumann, C. S., Fujimori, D. G. & Walsh, C. T. Halogenation strategies in natural product biosynthesis. *Chem. Biol.* **15**, 99–109 (2008) doi:10.1016/j.chembiol.2008.01.006.
9. Marchand, J. A., Neugebauer, M. E., Ing, M. C., Lin, C.-I., Pelton, J. G. & Chang, M. C. Y. Discovery of a pathway for terminal-alkyne amino acid biosynthesis. *Nature* **567**, 420–424 (2019) doi:10.1038/s41586-019-1020-y.
10. Vaillancourt, F. H., Yeh, E., Vosburg, D. A., O'Connor, S. E. & Walsh, C. T. Cryptic chlorination by a non-haem iron enzyme during cyclopropyl amino acid biosynthesis. *Nature* **436**, 1191–1194 (2005) doi:10.1038/nature03797.
11. Nakamura, H., Schultz, E. E. & Balskus, E. P. A new strategy for aromatic ring alkylation in cylindrocyclophane biosynthesis. *Nat. Chem. Biol.* **13**, 916–921 (2017) doi:10.1038/nchembio.2421.
12. Anslyn, E. V. & Dougherty, D. A. *Modern Physical Organic Chemistry*. (University Science, 2006). doi:10.1021/ed083p387.
13. Agarwal, V., Miles, Z. D., Winter, J. M., Eustáquio, A. S., El Gamal, A. A. & Moore, B. S. Enzymatic halogenation and dehalogenation reactions: pervasive and mechanistically diverse. *Chem. Rev.* **117**, 5619–5674 (2017) doi:10.1021/acs.chemrev.6b00571.
14. Liang, T., Neumann, C. N. & Ritter, T. Introduction of Fluorine and Fluorine-Containing Functional Groups. *Angew. Chemie Int. Ed.* **52**, 8214–8264 (2013) doi:10.1002/anie.201206566.
15. Petrone, D. A., Ye, J. & Lautens, M. Modern Transition-Metal-Catalyzed Carbon-Halogen Bond Formation. *Chem. Rev.* **116**, 8003–104 (2016) doi:10.1021/acs.chemrev.6b00089.
16. Shilov, A. E. & Shul'pin, G. B. Activation of C–H Bonds by Metal Complexes. *Chem. Rev.* **97**, 2879–2932 (1997) doi:10.1021/CR9411886.
17. Bollinger, J. M., Chang, W., Matthews, M. L., Martinie, R. J., Boal, A. K. & Krebs, C. Mechanisms of 2-oxoglutarate-dependent oxygenases: the hydroxylation paradigm and beyond. in *2-Oxoglutarate-Dependent Oxygenases* 95–122 (Royal Society of Chemistry, London, 2015). doi:10.1039/9781782621959.

18. Blasiak, L. C., Vaillancourt, F. H., Walsh, C. T. & Drennan, C. L. Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature* **440**, 368–371 (2006) doi:10.1038/nature04544.
19. Mitchell, A. J., Zhu, Q., Maggiolo, A. O., Ananth, N. R., Hillwig, M. L., Liu, X. & Boal, A. K. Structural basis for halogenation by iron- and 2-oxo-glutarate-dependent enzyme WelO5. *Nat. Chem. Biol.* **12**, 636–640 (2016) doi:10.1038/nchembio.2112.
20. Vaillancourt, F. H., Yin, J. & Walsh, C. T. SyrB2 in syringomycin E biosynthesis is a nonheme Fe(II) alpha-ketoglutarate- and O₂-dependent halogenase. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 10111–10116 (2005) doi:10.1073/pnas.0504412102.
21. Hillwig, M. L. & Liu, X. A new family of iron-dependent halogenases acts on freestanding substrates. *Nat. Chem. Biol.* **10**, 6–10 (2014) doi:10.1038/nchembio.1625.
22. Ortega, M. A. & van der Donk, W. A. New insights into the biosynthetic logic of ribosomally synthesized and post-translationally modified peptide natural products. *Cell Chem. Biol.* **23**, 31–44 (2016) doi:10.1016/j.chembiol.2015.11.012.
23. Rungtaphan, W., Qu, X. & O'Connor, S. E. Integrating carbon-halogen bond formation into medicinal plant metabolism. *Nature* **468**, 461–464 (2010) doi:10.1038/nature09524.
24. Challis, G. L., Ravel, J. & Townsend, C. A. Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chem. Biol.* **7**, 211–224 (2000) doi:10.1016/s1074-5521(00)00091-0.
25. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004) doi:10.1186/1471-2105-5-113.
26. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–8 (2014) doi:10.1093/bioinformatics/btu531.
27. Huang, Y., Niu, B., Gao, Y., Fu, L. & Li, W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **26**, 680–2 (2010) doi:10.1093/bioinformatics/btq003.
28. Gerlt, J. A. Genomic enzymology: web tools for leveraging protein family sequence-function space and genome context to discover novel functions. *Biochemistry* **56**, 4293–4308 (2017) doi:10.1021/acs.biochem.7b00614.
29. Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275–82 (1992) doi:10.1093/bioinformatics/8.3.275.
30. Crooks, G. E., Hon, G., Chandonia, J.-M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–90 (2004) doi:10.1101/gr.849004.
31. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D. & Bairoch, A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31**, 3784–8 (2003) doi:10.1093/nar/gkg563.
32. Kabsch, W. XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125–32 (2010) doi:10.1107/S0907444909047337.
33. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D. Biol. Crystallogr.* **69**, 1204–14 (2013) doi:10.1107/S0907444913000061.
34. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D. Biol. Crystallogr.* **67**, 235–42 (2011) doi:10.1107/S0907444910045749.
35. Skubák, P. & Pannu, N. S. Automatic protein structure solution from weak X-ray data. *Nat. Commun.* **4**, 2777 (2013) doi:10.1038/ncomms3777.

36. Cowtan, K. The Buccaneer software for automated model building. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 1002–1011 (2006) doi:10.1107/S0907444906022116.
37. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of *Coot*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010) doi:10.1107/S0907444910007493.
38. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–21 (2010) doi:10.1107/S0907444909052925.
39. Dunwell, J. M., Purvis, A. & Khuri, S. Cupins: the most functionally diverse protein superfamily? *Phytochemistry* **65**, 7–17 (2004) doi:10.1016/J.PHYTOCHEM.2003.08.016.
40. Pandurangan, A. P., Stahlhacke, J., Oates, M. E., Smithers, B. & Gough, J. The SUPERFAMILY 2.0 database: a significant proteome update and a new webserver. *Nucleic Acids Res.* **47**, D490–D494 (2019) doi:10.1093/nar/gky1130.
41. Kulik, H. J. & Drennan, C. L. Substrate placement influences reactivity in non-heme Fe(II) halogenases and hydroxylases. *J. Biol. Chem.* **288**, 11233–41 (2013) doi:10.1074/jbc.M112.415570.
42. Matthews, M. L., Krest, C. M., Barr, E. W., Vaillancourt, F. H., Walsh, C. T., Green, M. T., Krebs, C. & Bollinger, J. M. Substrate-Triggered Formation and Remarkable Stability of the C–H Bond-Cleaving Chloroferryl Intermediate in the Aliphatic Halogenase, SyrB2. *Biochemistry* **48**, 4331–4343 (2009) doi:10.1021/bi900109z.
43. Puri, M., Biswas, A. N., Fan, R., Guo, Y. & Que, L. Modeling non-heme iron halogenases: high-spin oxoiron(IV)–halide complexes that halogenate C–H bonds. *J. Am. Chem. Soc.* **138**, 2484–2487 (2016) doi:10.1021/jacs.5b11511.
44. Galonić, D. P., Barr, E. W., Walsh, C. T., Bollinger, J. M. & Krebs, C. Two interconverting Fe(IV) intermediates in aliphatic chlorination by the halogenase CytC3. *Nat. Chem. Biol.* **3**, 113–116 (2007) doi:10.1038/nchembio856.
45. Wong, S. D., Srnc, M., Matthews, M. L., Liu, L. V, Kwak, Y., Park, K., Bell, C. B., Alp, E. E., Zhao, J., Yoda, Y., Kitao, S., Seto, M., Krebs, C., Bollinger, J. M. & Solomon, E. I. Elucidation of the Fe(IV)=O intermediate in the catalytic cycle of the halogenase SyrB2. *Nature* **499**, 320–323 (2013) doi:10.1038/nature12304.
46. Holm, L. & Laakso, L. M. Dali server update. *Nucleic Acids Res.* **44**, W351-5 (2016) doi:10.1093/nar/gkw357.
47. Srnc, M. & Solomon, E. I. Frontier molecular orbital contributions to chlorination versus hydroxylation selectivity in the non-heme iron halogenase SyrB2. *J. Am. Chem. Soc.* **139**, 2396–2407 (2017) doi:10.1021/jacs.6b11995.
48. Matthews, M. L., Neumann, C. S., Miles, L. a, Grove, T. L., Booker, S. J., Krebs, C., Walsh, C. T. & Bollinger, J. M. Substrate positioning controls the partition between halogenation and hydroxylation in the aliphatic halogenase, SyrB2. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 17723–17728 (2009) doi:10.1073/pnas.0909649106.
49. Mitchell, A. J., Dunham, N. P., Bergman, J. A., Wang, B., Zhu, Q., Chang, W., Liu, X. & Boal, A. K. Structure-guided reprogramming of a hydroxylase to halogenate its small molecule substrate. *Biochemistry* **56**, 441–444 (2017) doi:10.1021/acs.biochem.6b01173.
50. Zhang, Z., Ren, J. shan, Harlos, K., McKinnon, C. H., Clifton, I. J. & Schofield, C. J. Crystal structure of a clavamate synthase-Fe(II)-2-oxoglutarate-substrate-NO complex: evidence for metal centered rearrangements. *FEBS Lett.* **517**, 7–12 (2002) doi:10.1016/S0014-5793(02)02520-6.
51. Martinie, R. J., Livada, J., Chang, W., Green, M. T., Krebs, C., Bollinger, J. M. & Silakov, A. Experimental correlation of substrate position with reaction outcome in the aliphatic halogenase,

- SyrB2. *J. Am. Chem. Soc.* **137**, 6912–6919 (2015) doi:10.1021/jacs.5b03370.
52. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018) doi:10.1093/molbev/msy096.
 53. Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D. & Higgins, D. G. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011) doi:10.1038/msb.2011.75.
 54. McCusker, K. P. & Klinman, J. P. Modular behavior of TauD provides insight into the origin of specificity in alpha-ketoglutarate-dependent nonheme iron oxygenases. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 19791–19795 (2009) doi:10.1073/pnas.0910660106.
 55. Matthews, M. L., Chang, W., Layne, A. P., Miles, L. a, Krebs, C. & Bollinger, J. M. Direct nitration and azidation of aliphatic carbons by an iron-dependent halogenase. *Nat. Chem. Biol.* **10**, 209–215 (2014) doi:10.1038/nchembio.1438.
 56. Fu, G. C. Transition-metal catalysis of nucleophilic substitution reactions: a radical alternative to SN1 and SN2 processes. *ACS Cent. Sci.* **3**, 692–700 (2017) doi:10.1021/acscentsci.7b00212.
 57. Nyffeler, P. T., Liang, C.-H., Koeller, K. M. & Wong, C.-H. The chemistry of amine-azide interconversion: catalytic diazotransfer and regioselective azide reduction. *J. Am. Chem. Soc.* **124**, 10773–10778 (2002) doi:10.1021/ja0264605.
 58. Sletten, E. M. & Bertozzi, C. R. Bioorthogonal chemistry: fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed. Engl.* **48**, 6974–6998 (2009) doi:10.1002/anie.200900942.
 59. Roughley, S. D. & Jordan, A. M. The medicinal chemist's toolbox: an analysis of reactions used in the pursuit of drug candidates. *J. Med. Chem.* **54**, 3451–3479 (2011) doi:10.1021/jm200187y.
 60. Gatto, G. J., Boyne, M. T., Kelleher, N. L. & Walsh, C. T. Biosynthesis of pipecolic acid by RapL, a lysine cyclodeaminase encoded in the rapamycin gene cluster. *J. Am. Chem. Soc.* **128**, 3838–3847 (2006) doi:10.1021/ja0587603.
 61. Goodman, J. L., Wang, S., Alam, S., Ruzicka, F. J., Frey, P. A. & Wedekind, J. E. Ornithine cyclodeaminase: structure, mechanism of action, and implications for the mu-crystallin family. *Biochemistry* **43**, 13883–13891 (2004) doi:10.1021/bi048207i.
 62. Wendisch, V. F., Mindt, M. & Pérez-García, F. Biotechnological production of mono- and diamines using bacteria: recent progress, applications, and perspectives. *Appl. Microbiol. Biotechnol.* **102**, 3583–3594 (2018) doi:10.1007/s00253-018-8890-z.
 63. Takatsuka, Y., Yamaguchi, Y., Ono, M. & Kamio, Y. Gene cloning and molecular characterization of lysine decarboxylase from *Selenomonas ruminantium* delineate its evolutionary relationship to ornithine decarboxylases from eukaryotes. *J. Bacteriol.* **182**, 6732–6741 (2000) doi:10.1128/jb.182.23.6732-6741.2000.
 64. Amorim Franco, T. M., Hegde, S. & Blanchard, J. S. Chemical mechanism of the branched-chain aminotransferase IlvE from *Mycobacterium tuberculosis*. *Biochemistry* **55**, 6295–6303 (2016) doi:10.1021/acs.biochem.6b00928.
 65. Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K. & Ueda, T. Cell-free translation reconstituted with purified components. *Nat. Biotechnol.* **19**, 751–755 (2001) doi:10.1038/90802.
 66. Spokoiny, A. M., Zou, Y., Ling, J. J., Yu, H., Lin, Y.-S. & Pentelute, B. L. A perfluoroaryl-cysteine S(N)Ar chemistry approach to unprotected peptide stapling. *J. Am. Chem. Soc.* **135**, 5946–9 (2013) doi:10.1021/ja400119t.

Chapter 4: *Mechanistic investigation of halogenation vs. hydroxylation in the BesD family*

Portions of this work were performed in collaboration with the following persons:

Dr. Jeff Pelton performed NMR experiments. Hydroxylase crystallography was performed with Elijah Kissman. Fluorogenic screening of the library was developed in collaboration with Dr. Jorge Marchand and Doug Millar. Halogenase biochemistry was performed in collaboration with Nicholas Sambold.

4.1. Introduction

Halogenation of small molecules is a useful chemical transformation. As a result, new enzymatic and synthetic strategies for selective halogenation are highly desirable. In particular, selective halogenation of unactivated C–H bonds is difficult to achieve. Nature has evolved a class of Fe^{II}/ α KG dependent enzymes that are able to perform halogenation of unactivated C–H bonds, making them potentially useful biocatalysts for this challenging chemical transformation.

Halogenases are part of the diverse superfamily of Fe^{II}/ α -ketoglutarate-dependent enzymes, which perform a wide range of transformations, including hydroxylation of unactivated C–H bonds. Fe^{II}/ α KG-dependent hydroxylases and halogenases proceed through similar mechanisms (*Figure 4.1*). Hydroxylases possess an HXD/EX_nH motif and coordinate Fe^{II} with two side chain histidines, one side chain carboxylate, and the cofactor α KG. Upon oxygen binding, the enzyme decarboxylates α KG to succinate to generate a Fe^{IV}-oxo intermediate, which abstracts a hydrogen atom from the substrate. The resulting substrate radical then rebounds with the Fe^{III}-OH group, yielding a hydroxylated product. In contrast, halogenases possess an HXG/A motif, where the smaller Ala or Gly residue allows room for the chloride to coordinate directly to Fe during the catalytic cycle. Following hydrogen atom abstraction, the substrate radical selectively rebounds with Fe^{III}-Cl, rather than Fe^{III}-OH, yielding a chlorinated product [1,2].

To date, very few radical halogenases have been discovered in Nature [3]. In contrast, many Fe^{II}/ α KG-dependent hydroxylases have been identified and found to act on a wide range of substrates [4]. Thus, one strategy for expanding the scope of enzymatic halogenation is to engineer the broader class of hydroxylases to instead perform halogenation. Towards this goal, attempts have been made to enable halogenation by engineering the HXD/E motif of hydroxylases into the HXG/A required for halogenation. This has proven difficult, as these efforts have largely led to no observable halogenation [5,6]. In one promising case, modest halogenation yields were achieved by mutagenesis of the enzyme SadA [7]. However, the majority of the product was still hydroxylated. These efforts indicate that additional factors other than the HXG/A vs HXD/E motifs must contribute to selective rebound of Fe^{III}-Cl rather than Fe^{III}-OH. Because the few halogenases that have been reported to date have carrier-protein-bound substrates [8–10] or complex free-standing substrates [11–13], it has been challenging to study the basis of halogenation in this class of enzymes. We recently discovered BesD, a radical halogenase that chlorinates the free amino acid lysine directly, without the requirement for a carrier protein [14]. The simple substrate of BesD makes it a convenient system for investigating halogenation selectivity.

In this chapter, we used bioinformatics to identify a lysine halogenase and hydroxylase pair within the BesD family. The enzymes are 71% identical and modify the same substrate with identical regioselectivities but possess divergent reaction outcomes (halogenation vs. hydroxylation). The high sequence identity makes this a promising system for studying reaction partitioning between halogenation and hydroxylation. We characterized the enzymes biochemically and solved the crystal structure of the novel hydroxylase. Next, we used DNA-shuffling of the two genes to generate a library of chimeric proteins and screened the library for chlorination activity to identify second-sphere residues that are critical for halogenation. Finally, based on insights from the screen and structure, we successfully engineered the hydroxylase to perform halogenation.

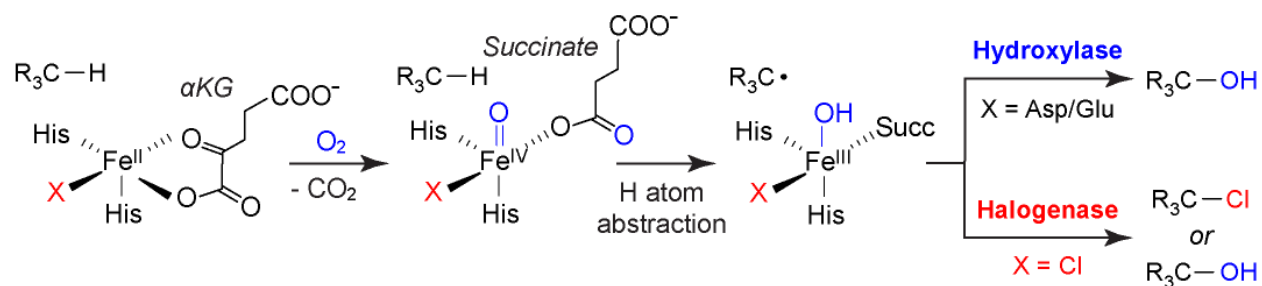


Figure 4.1. Divergent reaction outcomes in halogenases and hydroxylases. Radical halogenases are related in reaction mechanism to radical hydroxylases. In halogenases, Asp or Glu is instead replaced with a Gly or Ala residue, allowing room for direct coordination of chloride to the Fe. In this case, rebound of either the -Cl or the -OH radical is possible, leading to halogenated or hydroxylated products, respectively.

4.2. Materials and methods

Commercial materials. Luria-Bertani (LB) Broth Miller, LB Agar Miller, Terrific Broth (TB), and glycerol were purchased from EMD Biosciences (Darmstadt, Germany). Carbenicillin (Cb), kanamycin (Km), chloramphenicol (Cm), isopropyl- β -D-thiogalactopyranoside (IPTG), sodium chloride, dithiothreitol (DTT), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), magnesium chloride hexahydrate, acetonitrile, ethylene diamine tetraacetic acid disodium dihydrate (EDTA), hydrochloric acid, 3kDa MWCO dialysis tubing, ammonium bicarbonate, sodium acetate, and sodium hydroxide were purchased from Fisher Scientific (Pittsburgh, PA). Phosphoenolpyruvate (PEP), adenosine triphosphate sodium salt (ATP), nicotinamide adenine dinucleotide reduced form dipotassium salt (NADH), pyruvate kinase, lactate dehydrogenase, lysozyme, poly(ethyleneimine) solution (PEI), ammonium iron (II) sulfate hexahydrate, α -ketoglutaric acid sodium salt, β -mercaptoethanol (β ME), N,N,N',N'-tetramethylethane-1,2-diamine (TEMED), betaine, dimethylsulfoxide (DMSO), sodium L-ascorbate, acetonitrile (LC/MS-grade), ammonium formate (LC/MS-grade), Deuterium oxide, L-lysine for crystallography, 2-(N-Morpholino)ethanesulfonic acid hydrate, 4-morpholineethanesulfonic acid (MES), Pall Nanosep centrifugal Omega membrane (3,000 MWCO), AeraSeal films, and copper sulfate (CuSO_4) were purchased from Sigma-Aldrich (St. Louis, MO). PageRuler Plus Prestained Protein Ladder was purchased from ThermoFisherScientific (Waltham, MA). Succinyl-CoA synthetase was purchased from Megazyme International (Bray, Ireland). L-lysine $\cdot$ 2HCl ($^{13}\text{C}_6$, 99%; $^{15}\text{N}_2$, 99%) was purchased from Cambridge Isotope Laboratories (Andover, MA). Formic acid was purchased from Acros Organics (Morris Plains, NJ). Restriction enzymes, T4 DNA ligase, Phusion DNA polymerase, Q5 DNA polymerase, T5 exonuclease, and Taq DNA ligase, were purchased from New England Biolabs (Ipswich, MA). Deoxynucleotides (dNTPs), were purchased from Invitrogen (Carlsbad, CA). Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA), resuspended at a stock concentration of 100 μM in water and stored at either 4 $^\circ\text{C}$ for immediate usage or -20 $^\circ\text{C}$ for long-term storage. DNA purification kits and Ni-NTA agarose were purchased from Qiagen (Valencia, CA). Zirconia/silica beads were purchased from BioSpec Products (Bartlesville, OK). Complete EDTA-free protease inhibitor was purchased from Roche Applied Science (Penzberg, Germany). PD-10 desalting columns and Superdex 75 16/60 pg column were purchased from GE Healthcare (Pittsburg, PA). Amicon Ultra 10,000 MWCO centrifugal concentrators and Milli-Q Gradient water purification system were purchased from Millipore (Billerica, MA). Acrylamide/bis-acrylamide (30%, 37.5:1), electrophoresis grade sodium dodecyl sulfate (SDS), and ammonium persulfate were purchased from Bio-Rad Laboratories (Hercules, CA). Ultrayield baffled flasks were purchased from Thompson Instrument Company (Oceanside, CA). CalFluor 488 azide and 2-(4-((bis((1-(tert-butyl)-1H-1,2,3-triazol-4-yl)methyl)amino)methyl)-1H-1,2,3-triazolyl) acetic acid (BTAA) was purchased from Click Chemistry Tools (Scottsdale, AZ). 96-well plates with 2 ml volume and v-bottom were purchased from Corning (Corning, NY). 3-Azido-7-hydroxycoumarin was purchased from AK Scientific, Inc. (Union City, CA).

Bioinformatic identification of a homologous halogenase and hydroxylase pair. To identify halogenase and hydroxylase homologs, the sequence of BesD (WP_014151497) was used as a query for BLAST against the non-redundant protein database with a cutoff value of e^{-5} , yielding 349 hits. From the list, redundant sequences of 98% identity were removed using CD-HIT [15], leaving 256 sequences. Sequences were aligned using MUSCLE [16] and visualized in AliView [17], and the sequence position of the HXD/E or HXG/A motif was identified manually.

A Bioython script was used to parse the list of alignments to annotate the halogenases (HXG/A) and hydroxylases (HXD/E) based on the key Fe-binding motif (*Appendix A3.1*). The MEGAX interface was then used to analyze the alignment by Maximum-Likelihood based on the JTT matrix-based model [18], and uncertainty in the topology of the resulting tree was evaluated with 100 bootstrap replicates (*Appendix A2.1*).

Bacterial strains. *E. coli* DH10B-T1^R was used for plasmid construction. BL21(DE3)–Star harboring the pRARE2 plasmid was used for heterologous protein production of all halogenase and hydroxylase variants.

Construction of plasmids. Gibson assembly [19] was used to carry out plasmid construction using *E. coli* DH10B-T1^R as the cloning host. PCR amplifications were carried out with Phusion polymerase (New England Biolabs, Ipswich MA) using the oligonucleotides listed in *Appendix A1.3*. Following plasmid construction, all cloned inserts were sequenced at Quintara Biosciences (San Francisco, CA) or the Barker Hall Sequencing Facility at UC Berkeley (Berkeley, CA). The halogenase (Hal, WP_122981682) and hydroxylase (WP_107105619) DNA sequences were initially optimized for expression in *E. coli*. Although the amino acid sequence of Hal is 71% identical to Hydrox, the nucleotide sequences of the *E. coli*-optimized genes were only ~65% identical. Hal was therefore manually optimized (Hal_optimized) to maximize nucleotide sequence identity with the hydroxylase to improve DNA shuffling. For example, if Hydrox encoded for Gly with codons GGG and Hal encoded for Gly at the same position with GGA, the codons in Hal were changed to match that of Hydrox. After optimization, the two sequences shared ~85% identity at nucleotide level while maintaining their original amino acid sequences, which shared 71% identity. The nucleotide sequences are listed in *Appendix 1.3*.

pET16-His₁₀-Hal, pET16-His₁₀-Hal_optimized, pET16b-His₁₀-Hydrox and pET16b-His₁₀-Hydrox point mutants. The intermediate cloning plasmid, pET16-His-PrescissionCutSite-IMPDH, was constructed by amplification of IMPDH from *E. coli* gDNA followed by insertion into NdeI/BamHI-digested pET16b. The halogenase WP_122981682 (Hal and Hal_optimized) and hydroxylase WP_107105619 (Hydrox), were obtained as GeneBlocks (*Appendix A1.3*, IDT DNA, Coralville, IA) containing the appropriate overhangs for Gibson insertion into NdeI/BamHI-digested His-PrescissionCutSite-IMPDH. To clone pET16-His₁₀-Hydrox D144G, Hydrox was amplified in two pieces from the pET16-His₁₀-Hydrox plasmid. The first piece was amplified with HalHydrox-F and the Hydrox D144G-R primer indicated in *Appendix A1.3*. The second piece was amplified with the Hydrox D144G-F primer indicated in *Appendix A1.3* and pET16-anneal-R. Then both pieces were inserted into NdeI/HindIII-digested pET16-His-PrescissionCutSite-IMPDH to yield the final constructs. pET16-His₁₀-Hydrox D144G N151I was cloned in a similar manner from the pET16-His₁₀-Hydrox D144G plasmid using the HalHydrox-F/Hydrox D114G N151I-R (piece 1) and Hydrox D114G N151I-F/pET16-anneal-R (piece 2) primer pairs. pET16-His₁₀-Hydrox D144G V225N was cloned from the pET16-His₁₀-Hydrox D144G plasmid using the HalHydrox-F/Hydrox D114G V225N-R (piece 1) and Hydrox D114G V225N-F/pET16-anneal-R (piece 2) primer pairs. The triple mutant pET16-His₁₀-Hydrox D144G N151I V225N was cloned analogously from Hydrox D144G N151I using the HalHydrox-F/Hydrox V225N-R (piece 1) and Hydrox V225N-F/pET16-anneal-R (piece 2) primer pairs.

pET16-His₁₀-Chimera. The three pieces for Gibson insertion (ChimeraPiece1, ChimeraPiece2, and ChimeraPiece3) were amplified from pET16-His₁₀-Hydroxylase D144G with the primer pairs HalHydrox-F/ChP1-R, ChP2-F/R, and ChP3-F/pET16-anneal-R, respectively. The three pieces were joined using overlap extension PCR [20] and then inserted into NdeI/HindIII-digested

pET16-His-PrescissionCutSite-IMPDH using Gibson to yield the final pET16-His₁₀-Chimera construct. To generate additional mutations on the Chimera scaffold (Chimera L150F, Chimera ST218-219GAV, Chimera I222M, Chimera M225V T226S, Chimera A228S, Chimera K230E E234G D236V), the Chimera was amplified in two pieces from the pET16-His-Chimera plasmid. The first piece was amplified with HalHydrox-F and the reverse primers indicated in *Appendix A1.3*. The second piece was amplified with the forward primers indicated in *Appendix A1.3* and pET16-anneal-R. Then both pieces were inserted into NdeI/HindIII-digested pET16-His-PrescissionCutSite-IMPDH to yield the mutant constructs.

Hydrox D144G N151I V225N point mutants. To generate additional mutations on the Hydrox D144G N151I V225N (HydroxTripleMutant, HydroxTM), constructs were amplified in two pieces from the pET16-His-Hydrox D144G N151I V225N plasmid. The first piece was amplified with HalHydrox-F and the reverse primers indicated in *Appendix A1.3*. The second piece was amplified with the forward primers indicated in *Appendix A1.3* and pET16-anneal-R. Then both pieces were inserted into NdeI/HindIII-digested pET16-His-PrescissionCutSite-IMPDH to yield the mutant constructs. Variants cloned by this strategy include HydroxTM F150L, HydroxTM GAV218/219 to ST, HydroxTM M223I, HydroxTM V226M S227T, HydroxTM S229A, HydroxTM E231K G235E V237D.

pBesCB. A fragment containing *besC*-T7p-MBP-*besB* was PCR amplified from pPra [14] using primers pBesCB-F/R and cloned into pACYC-DUET using Gibson assembly and NcoI/KpnI cut sites.

Additional Hal and Hydrox variants that were cloned but not assayed in this work are described in *Appendix A1.3D-F*. All of these constructs were cloned into NdeI/HindIII-digested pET16-His-PrescissionCutSite-IMPDH using Gibson assembly.

Construction of a shuffle library of Hal and Hydrox D144G. Hal_{opt} and Hydrox D144G were PCR-amplified from pET16-His-Hal_{optimized} and pET16-His-Hydrox D144G, respectively using HalHydrox-F and HalHydrox-R. Following PCR purification, 1000 ng of each of the PCR products was mixed in equimolar ratios and digested on ice with 1 U DNaseI (New England Biolabs, Ipswich MA) for 1 min 15 s. The 50 μ L reactions were quenched by adding 20 μ L of 100 mM EDTA. Digested DNA was analyzed on a 1% agarose gel. Fragments centered around 200 bp were isolated and purified by gel purification (Qiagen). 200 ng of purified fragments were re-assembled and extended using Phusion polymerase with the following program: 98 °C for 30 s (denature), 42 °C for 45 s (annealing), 72 °C for 1 min (extension). Fully reassembled products (~900 bp) were gel purified and re-amplified using Phusion polymerase and the HalHydrox-F/R primer pair. The shuffled products were inserted into NdeI/BamHI-digested pET16-His-PrescissionCutSite-IMPDH using Gibson assembly to yield the mutant constructs.

Click screen media optimization with 3-azido-7-hydroxycoumarin. For media optimization, we tested the CuAAC reaction in LB, TB, 2X YT, autoinducing media, M9 minimal media, and M9 with additives (magnesium sulfate (2 mM), calcium chloride (0.1 mM), glucose (4% (w/v)), thiamine (1 mg/L), 0.1 mM each of the 20 canonical amino acids). Media samples (50 μ L) containing Pra were pipetted into a 96-well black plate with a clear bottom (Corning, NY). To initiate the reaction, 50 μ L of a 2X CuAAC reaction solution in 1X PBS was added to each well to give the final reaction solution (3-azido-7-hydroxycoumarin (0.5 mM), Pra (0 - 1 mM), copper (II) sulfate (0.5 mM), THPTA (1 mM), and sodium ascorbate (5 mM)). Reactions were performed

for 30 min in the dark before analyzing fluorescence of the Pra-coumarin product (Pra-Cou, $\lambda_{\text{ex}} = 404$ nm, $\lambda_{\text{em}} = 477$ nm) using a SynergyMx Microplate Reader (BioTek) at room temperature.

Click screen with CalFluor488 probe. 50 μ L M9 media containing Pra was transferred to a fresh 96-well plate and mixed 1:1 with CuAAC solution to yield the final 100 μ L reaction mixture (CalFluor 488 azide (1 μ M), CuSO₄ (0.5 mM), BTAA (0.1 mM), and sodium ascorbate (5 mM)). Following incubation in the dark for 15 min, the fluorescence was measured using a SynergyMx Microplate Reader (BioTek) at room temperature ($\lambda_{\text{ex}} = 485$ nm, $\lambda_{\text{em}} = 528$ nm).

Screening the shuffle library for alkyne formation. Chemically-competent *E. coli* BL21 cells harboring the pBesCB plasmid (Cm^R) were transformed with the shuffle library plasmids (Amp^R) and grown on LB plates with the appropriate antibiotic. Colonies were inoculated into individual wells of a 96-well plate containing 200 μ L LB with the appropriate antibiotic and grown overnight. Dense cultures were diluted 20-fold into 96-well plates (2 ml v-bottom, Corning, NY) containing 1 ml of LB, covered with Aeraseal Film (Sigma) and grown to mid-log phase (OD = 0.6-0.8) before chilling the plates on ice for 30 min. Cells were pelleted by centrifugation at 4,000 $\times$ g for 20 min and then resuspended in 1 ml of M9 media (sodium phosphate (6.6 g/L), potassium phosphate (3 g/L), sodium chloride (0.5 g/L), ammonium chloride (1 g/L), magnesium sulfate (2 mM), calcium chloride (0.1 mM), glucose (4% (w/v)), thiamine (1 mg/L)) supplemented with 0.5 mM Lysine, 0.25 mM IPTG, and the appropriate antibiotics. Cells were grown for 2 d at 16°C with shaking at 225 rpm. Note that at the end of the 2 d, some of the cells had settled in the bottom of the plate, despite shaking during growth. The remaining cells were then pelleted by centrifugation. The supernatant (50 μ L) was transferred to a fresh 96-well black plate with a clear bottom (Corning). 50 μ L of freshly-prepared click solution (CalFluor 488 azide (1 μ M), CuSO₄ (0.5 mM), BTAA (0.1 mM), sodium ascorbate (5 mM)) was then added to the cell supernatant and incubated for 15 min in the dark. The fluorescence of the samples was measured using a SynergyMx Microplate Reader (BioTek) at room temperature ($\lambda_{\text{ex}} = 485$ nm, $\lambda_{\text{em}} = 528$ nm).

Cells from the 30 fluorescent wells were grown overnight, minipreped (Qiagen) and sequenced with the Prescission-seq-F primer (*Appendix A1.3*), which anneals to the shuffle library vector but not to the pBesCB vector which was also present in the cells. In addition, a total of 46 wells of non-halogenating cells were sequenced. 33 of these gave reads corresponding to the expected shuffle library, while 13 yielded sequences of empty vectors or severely truncated genes that were unsuccessful cloning artifacts of from the shuffle library generation procedure.

Expression of His₁₀-tagged proteins. *E. coli* BL21 Star (DE3) harboring the pRARE2 plasmid was transformed with the appropriate protein expression plasmid. An overnight TB culture of the freshly transformed cells was used to inoculate TB (1 L) containing the appropriate antibiotics (50 μ g/mL carbenicillin with 50 μ g/mL chloramphenicol) in a 2.8 L-baffled shake flask to OD₆₀₀ = 0.05. The cultures were grown at 37°C at 200 rpm to OD₆₀₀ = 0.6 to 0.8 at which point cultures were cooled on ice for 20 min, followed by induction of protein expression with IPTG (0.25 mM) and overnight growth at 16°C. Cell pellets were harvested by centrifugation at 9,800 $\times$ g for 7 min at 4°C and stored at -80°C.

Purification of His₁₀-tagged proteins. Frozen cell pellets were thawed and resuspended at 5 mL/g of cell paste in lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM imidazole, 20 mM β ME, 20% (v/v) glycerol, pH 7.5) supplemented with EDTA-free Protease Inhibitor Cocktail (Roche). The cell paste was homogenized and then lysed by sonication with a Qsonica Q700 sonicator (Amplitude = 45, 5 s on, 25 s off, 2.5 min total process time, 1/2" tip). The lysate was then

centrifuged at $13,500 \times g$ for 20 min at 4°C to separate the soluble and insoluble fractions. DNA was precipitated in the soluble fraction with 0.15% (w/v) polyethyleneimine and stirring at 4°C for 30 min. The precipitated DNA was then removed by centrifugation at $13,500 \times g$ for 20 min at 4°C . The soluble lysate was incubated with Ni-NTA (0.5 mL resin/g of cell paste) for 45 min at 4°C , then resuspended and loaded onto a column by gravity flow. The column was washed with wash buffer (50 mM HEPES, 300 mM NaCl, 20 mM imidazole, 20 mM β ME, 20% (v/v) glycerol, pH 7.5) for 15-20 column vol. The column was then eluted with elution buffer (50 mM HEPES, 300 mM NaCl, 300 mM imidazole, 20 mM β ME, 20% (v/v) glycerol, pH 7.5). Fractions containing the target protein were pooled by $A_{280\text{ nm}}$ and concentrated using an Amicon Ultra spin concentrator (10 kDa MWCO, Millipore). Protein was then exchanged into storage buffer (50 mM HEPES, 100 mM sodium chloride, 20% (v/v) glycerol, 1 mM DTT, pH 7.5) using PD-10 desalting columns. Final protein concentrations before storage were estimated using the $\epsilon_{280\text{ nm}}$ calculated by ExPASy ProtParam [21] and measured by nanodrop. They are as follows: *S. sp.* MBT76 Hydrox: 6.75 mg/mL ($\epsilon_{280\text{ nm}} = 45,950\text{ M}^{-1}\text{cm}^{-1}$), *A. teichiomyceticus* Hal: 7.49 mg/mL ($\epsilon_{280\text{ nm}} = 44,460\text{ M}^{-1}\text{cm}^{-1}$), Hydrox D144G: 6.36 mg/mL ($\epsilon_{280\text{ nm}} = 45,950\text{ M}^{-1}\text{cm}^{-1}$), Hydrox D144G V225N: 5.06 mg/mL ($\epsilon_{280\text{ nm}} = 45,950\text{ M}^{-1}\text{cm}^{-1}$), Hydrox D144G V225N N151I: 11.85 mg/mL ($\epsilon_{280\text{ nm}} = 45,950\text{ M}^{-1}\text{cm}^{-1}$), Hydrox D144G N151I: 10.08 mg/mL ($\epsilon_{280\text{ nm}} = 45,950\text{ M}^{-1}\text{cm}^{-1}$), Chimera L150F: 12.55 mg/mL ($\epsilon_{280\text{ nm}} = 45,950\text{ M}^{-1}\text{cm}^{-1}$), Chimera ST218/219 to GAV: 12.14 mg/mL ($\epsilon_{280\text{ nm}} = 45,950\text{ M}^{-1}\text{cm}^{-1}$), Chimera I222M: 11.27 mg/mL ($\epsilon_{280\text{ nm}} = 45,950\text{ M}^{-1}\text{cm}^{-1}$), Chimera M225V T226S: 14.57 mg/mL ($\epsilon_{280\text{ nm}} = 45,950\text{ M}^{-1}\text{cm}^{-1}$), Chimera A228S: 16.48 mg/mL ($\epsilon_{280\text{ nm}} = 45,950\text{ M}^{-1}\text{cm}^{-1}$), Chimera K230E E234G D236V: 19.55 mg/mL ($\epsilon_{280\text{ nm}} = 45,950\text{ M}^{-1}\text{cm}^{-1}$), Hydrox F150L: 9.86 mg/mL ($\epsilon_{280\text{ nm}} = 45,950\text{ M}^{-1}\text{cm}^{-1}$), Hydrox GAV218/219 to ST: 12.33 mg/mL ($\epsilon_{280\text{ nm}} = 45,950\text{ M}^{-1}\text{cm}^{-1}$), Hydrox M223I: 12.6 mg/mL ($\epsilon_{280\text{ nm}} = 45,950\text{ M}^{-1}\text{cm}^{-1}$), Hydrox V226M S227T: 15.74 mg/mL ($\epsilon_{280\text{ nm}} = 45,950\text{ M}^{-1}\text{cm}^{-1}$), Hydrox S229A: 10.95 mg/mL ($\epsilon_{280\text{ nm}} = 45,950\text{ M}^{-1}\text{cm}^{-1}$), Hydrox E231K G235E V237D: 13.73 mg/mL ($\epsilon_{280\text{ nm}} = 45,950\text{ M}^{-1}\text{cm}^{-1}$). All proteins were aliquoted, flash-frozen in liquid nitrogen, and stored at -80°C .

Preparation of proteins for crystallization. Following elution from Ni-NTA column, Hydrox protein was incubated with Prescission protease (1 mg protease/50 mg protein) and dialyzed 1:50 overnight into dialysis buffer to remove imidazole. Cleaved and dialyzed protein was passed through Ni-NTA (2 mL) to remove Prescission and the His₁₀ tag. The eluent was the diluted to a final salt concentration of 20 mM NaCl using Buffer A (50 mM HEPES, 20% (v/v) glycerol, 1 mM DTT, 1 mM EDTA, pH 7.5) and loaded onto a 5 mL HiTrap-Q column for ion exchange with the AKTA Purifier FPLC system (GE Healthcare). The protein was eluted using a gradient from 0-100% buffer A to buffer B (50 mM HEPES, 1 M NaCl, 20% (v/v) glycerol, 1 mM DTT, 1 mM EDTA, pH 7.5) over 40 min. The protein sample was concentrated to 2 mL and loaded onto a Superdex 75 16/60 pg (GE Healthcare) column equilibrated with SEC Buffer (20 mM HEPES, 100 mM NaCl, 1 mM DTT, pH 7.5). The protein eluent was concentrated to 15 mg/mL and glycerol was added to a final concentration of 5% (v/v) before flash freezing in liquid nitrogen and storage at -80°C .

Crystallization and structure determination. Crystals of Hydrox were obtained by the hanging drop vapor diffusion method by combining equal volumes of protein solution (3 mg/mL Hydrox, lysine (3 mM), α KG (3 mM, pH 7)) and reservoir solution (HEPES pH 6.8 (100 mM), sodium iodide (200 mM) containing 22% (w/v) PEG 3350). Densely-packed thin rods and plates grew in 1 day. Crystals were transferred to an Eppendorf tube containing 250 μL of reservoir solution and vortexed for 30 seconds with $10 \times 1\text{ mm}$ diameter zirconia/silica beads to produce a

micro-seed solution. The seed solution was stored at -80°C for future use. Fe-bound Hydrox crystals were prepared in a Coy Anaerobic chamber by micro-seeding equal volumes of protein solution (3 mg/mL Hydrox, lysine (3 mM), αKG (3 mM, pH 7)) and reservoir solution (HEPES pH 7.0 (100 mM), sodium nitrate (200 mM) containing 28% (w/v) PEG 3350). Thin rectangular rods grew within 3 days. Crystals were soaked anaerobically in a solution containing $\text{Fe}^{\text{II}}\text{Cl}_2 \cdot 4\text{H}_2\text{O}$ (167 μM) for 3 min before flash freezing in liquid nitrogen. Crystals of BesD from *S. cattleya* were obtained aerobically by the hanging drop vapor diffusion method by combining equal volumes of protein solution (10 mg/mL BesD, lysine (1.5 mM), succinate (3 mM, pH 7)), vanadium sulfate (0.5 mM) and reservoir solution (MES pH 6.5 (100 mM), sodium chloride (0.6 mM) containing 20% (v/v) PEG 4000). Rectangular, gel-like crystals grew in 4 days. Crystals were flash frozen in liquid nitrogen and data were collected at Beamline 8.3.1 at the Advanced Light Source (Lawrence Berkeley National Laboratory) at a wavelength of 1.11 Å.

Data were processed with XDS [22] and scaled and merged with Aimless [23] in within the CCP4 suite [24]. The data were phased with molecular replacement using the structure of BesD (6NIE) as a search model. The structure was refined iteratively in COOT [25] and Phenix [26]. Ligands were added to the model using COOT and refined in Phenix. Omit maps for the ligand complex including Fe, αKG , lysine, and Fe-bound histidines were generated using the Phenix Map function. The Fe, αKG , lysine, and Fe-bound histidines were selected for removal from refinement prior to calculation of the maps. Structural alignments were performed using the Chimera software [27]. The structure was analyzed in Pymol, which was also used to create figures.

General procedure for high resolution HPLC/MS analysis of polar metabolites with HILIC. Samples containing polar metabolites were analyzed using an Agilent 1290 UPLC on a SeQuant ZIC-pHILIC (5 μm , 2.1×100 mm; EMD-Millipore) using the following buffers: Buffer A (90% acetonitrile, 10% water, 10 mM ammonium formate) and Buffer B (90% water, 10% acetonitrile, 10 mM ammonium formate). A linear gradient from 95% to 60% Buffer A over 17 min followed by a linear gradient from 60% to 33% Buffer A over 8 min was then applied at a flow rate of 0.2 mL/min. Mass spectra were acquired in positive ionization mode using a 6530 QTOF (Agilent). A standard curve was not built due to difficulties with quantifying the unstable 4-Cl-lysine product [14]. Therefore, the reported values are not normalized for concentration of the product.

In vitro LC/MS assays of Hal and Hydrox variants. Reactions (50 μL) contained L-lysine $\cdot$ HCl (3 mM), sodium αKG (5 mM), sodium ascorbate (5 mM), $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ (1 mM), and sodium chloride (5 mM) in 100 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of purified Hal and Hydrox variants (20 μM final concentration) and allowed to proceed for 25 min at room temperature before quenching in 2 vol of methanol with 1% (v/v) formic acid. Following centrifugation at $18,000 \times g$ for 20 min (at 4°C) to remove precipitated protein, samples were analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF using the protocol for polar metabolite analysis above.

Determination of kinetic parameters for Hal and Hydrox on lysine. Reactions (100 μL) contained ATP (2.5 mM), MgCl_2 (5 mM) phosphoenol-pyruvate (PEP, 1 mM), NADH (0.3 mM) lactate dehydrogenase (LDH, 10 U/mL), pyruvate kinase (PK, 10 U/mL) succinyl-CoA synthetase (SCS, 3.2 U/mL), coenzyme A (1 mM), sodium αKG (1 mM), $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ (0.2 mM), sodium chloride (10 mM), and sodium ascorbate (2 mM) in 100 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of Hal (2.5 μM) or Hydrox (2.5 μM) in the presence of varying

concentrations of L-lysine · HCl (0 - 5 mM). Initial rates of NADH consumption were measured by monitoring A₃₄₀ using a SynergyMx Microplate Reader (BioTek) at room temperature. k_{cat} and K_M were determined by fitting to initial rate data with Origin (OriginLab, Northampton, MA) using the equation:

$$v_0 = \frac{k_{\text{cat}}[S]}{K_M + [S]}$$

where v_0 is the initial rate and $[S]$ is the substrate concentration.

HPLC purification of hydroxylysine for NMR analysis. Hydroxylysine was purified using an Agilent 1200 HPLC on a SeQuant ZIC-HILIC Semi-Preparative column (5 μm , 200 $\text{\AA}$, 150 $\times$ 21.1 mm, EMD Millipore) using the following buffers: Buffer A (90% acetonitrile, 10% water, 10 mM ammonium formate) and Buffer B (90% water, 10% acetonitrile, 10 mM ammonium formate). A gradient was applied as follows: linear gradient from 100% to 40% Buffer A over 1 h, followed by a hold of 40% Buffer A for 18 min, then to 30% Buffer A over 5 min, followed by a linear gradient to 75% A over 4 min at a flow rate of 5 mL/min. The fractions containing the desired compound were determined by LC/QTOF MS screening, pooled, and dried by speed-vacuum.

NMR data collection. All experiments were recorded on a Bruker Avance II spectrometer operating at 900 MHz and at 298 K. The instrument was equipped with a CP TXI cryoprobe and was controlled using Topspin (version 3.2) software. Data were processed using Topspin (version 3.2) by zero-filling once in each dimension, followed by apodization, Fourier transformation, and phasing, and were referenced to 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS). Spectra were analyzed using Mnova software (Mestrelab Research, Escondido, CA, USA).

Preparation, purification, and NMR analysis of [¹³C₆,¹⁵N₂]-4-OH-lysine. [¹³C₆,¹⁵N₂]-4-OH-lysine methyl ester was prepared in a 600 μL reaction containing Hydrox (60 μM), [¹³C₆,¹⁵N₂]-L-lysine · 2HCl (1 mM), αKG (5 mM), $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ (1 mM), sodium ascorbate (5 mM) in 50 mM HEPES (pH 7.5). The reaction was quenched after 16 h by addition of 1.5 mL of acetonitrile with 1% formic acid. Precipitate was removed by centrifugation at 15,000 $\times g$ for 10 min. [¹³C₆,¹⁵N₂]-4-OH-lysine was purified using an Agilent 1200 HPLC on a SeQuant ZIC-HILIC Semi-Preparative column as described above. Samples were dried by speed vacuum and then resuspended in D₂O for NMR analysis on a Bruker 900 MHz instrument and for LC/QTOF-MS analysis. High resolution ESI-MS $[\text{M}+\text{H}]^+$: calculated for ¹³C₆H₁₆¹⁵N₂O₃⁺ m/z , 171.1219, found m/z 171.1219.

A constant-time ¹H-¹³C HSQC experiment (Bruker pulse sequence hsqcctetgsp) was recorded with carrier frequencies set to 5.48 ppm (¹H) and 40 ppm (¹³C)) and the spectral widths set to 16 ppm (¹H) and 80 ppm (¹³C). The carboxyl chemical shift was set to 176 ppm, and the ¹H-¹³C coupling constant was set to 145 Hz. The recycle delay was set to 1.5 sec. The constant-time evolution period was set to 13.6 ms, to distinguish ¹H-¹³C pairs with one or three carbon neighbors from those with two carbon neighbors. A 2D ¹H-¹³C version of the 3D HCCH-COSY experiment (Bruker pulse sequence hcchcogp3d2) was used to identify neighboring ¹H and ¹³C resonances. The carrier frequencies were set to 5.48 ppm (¹H) and 39 ppm (¹³C) with spectral widths of 14 ppm (¹H) and 75 ppm (¹³C). The recycle delay was set to 1.8 sec. ¹H NMR (900 MHz, D₂O): δ 3.85 (H _{α}), 3.85 (H _{γ}), 3.03 (H _{ϵ}), 1.96 (H _{β}), 1.81 (H _{δ_1}), 1.76 (H _{δ_2}). ¹³C NMR (900 MHz, D₂O): δ 66.48 (C _{γ}), 52.78 (C _{α}), 36.67 (C _{ϵ}), 36.14 (C _{β}), 33.98 (C _{δ}).

4.3. Results and discussion

Discovery of a closely related halogenase and hydroxylase pair. The enzyme BesD is the first radical halogenase reported to chlorinate a free-standing amino acid directly without the requirement for a carrier protein-tethered substrate. By searching for BesD homologs possessing the HXG motif, we discovered additional enzymes that chlorinate a range of different amino acids [28]. We then extended this analysis to include sequences with an HXD/E motif in an effort to identify new enzymes that hydroxylate, rather than chlorinate, amino acids. We performed a BLAST search with an E-value of 10^{-5} and assembled the hits on a phylogenetic tree (Figure 4.2). The sequences on the tree were annotated as halogenases or hydroxylases based on the presence of an HXD or an HXG motif. For the most part, the hydroxylases and halogenases grouped separately on the tree, indicating that halogenation evolved first, followed by diversification of the enzymes for various amino acid substrates. However, there was one exception; the hydroxylase from *Streptomyces* sp. MBT76 (Hydrox, WP_107105619) is located within the clade of halogenases that chlorinate lysine at C₄. Strikingly, Hydrox is 71% identical in amino acid sequence to the putative halogenase from *Actinoplanes teichomyceticus* (Hal, WP_122981682). Due to the high percent identity between the two enzymes, we hypothesized that this system could facilitate the identification of key sequence differences responsible for their divergent reaction outcomes (Figure 4.3).

To confirm the predicted activity of the enzyme pair, we cloned, expressed, and purified the proteins from *E. coli*. By incubating purified protein with the substrate lysine and the cofactors Fe, α KG, and chloride and analyzing the products by LC/MS, we confirmed that Hal and Hydrox are indeed a lysine halogenase and hydroxylase, respectively (Figure 4.4). Next, we analyzed the kinetic parameters of Hal and Hydrox. We performed steady-state kinetic analysis using an assay that couples NADH oxidation to succinate formation by the enzyme of interest (Figure 4.5) [29]. We found that Hydrox possesses a two-fold higher k_{cat} (22.5 ± 0.7 vs. $10.5 \pm 0.4 \text{ min}^{-1}$), as well as a higher K_{M} (0.34 ± 0.05 vs. $0.12 \pm 0.03 \text{ mM}$) compared to Hal. However, both enzymes have a K_{M} value within the estimated intracellular concentration of lysine in bacteria [30]. While the LC/MS and kinetic analyses enable the assignment lysine as a substrate, it does not reveal information about the regioselectivity of the two enzymes. We therefore used HSQC and COSY NMR to confirm that Hydrox modifies C₄, which is the same position that is chlorinated by Hal (Figure 4.6). Thus, the enzymes diverge only in reaction outcome, and not in substrate- or regioselectivity, making this a convenient system for studying halogenation and hydroxylation.

Probing the role of the HXD/G motif on reaction outcome. In other hydroxylase systems, attempts have been made to engineer halogenase activity by mutating the HXD motif to HXG. This change would presumably allow room for chloride to coordinate directly to the Fe in the active site. In the hydroxylases TauD and Prolyl-4-hydroxylase, this change alone did not lead to detectable levels of halogenation [5,6]. Encouragingly, in the hydroxylase SadA, this mutation led to the formation of some halogenated product, although hydroxylation was still the dominant outcome [7]. We therefore tested whether the lysine hydroxylase that we identified could be engineered to perform halogenation by mutating the HXD motif to the HXG found in the halogenases (Hydrox D144G). Consistent with previously reported attempts, the mutagenesis failed to lead to efficient halogenation (Figure 4.7). Instead, we found Hydrox D144G has diminished hydroxylation activity, possibly due to impaired oxygen activation by the reactant complex, as has been found in FIH hydroxylase mutants [31]. Since mutagenesis of the facial triad

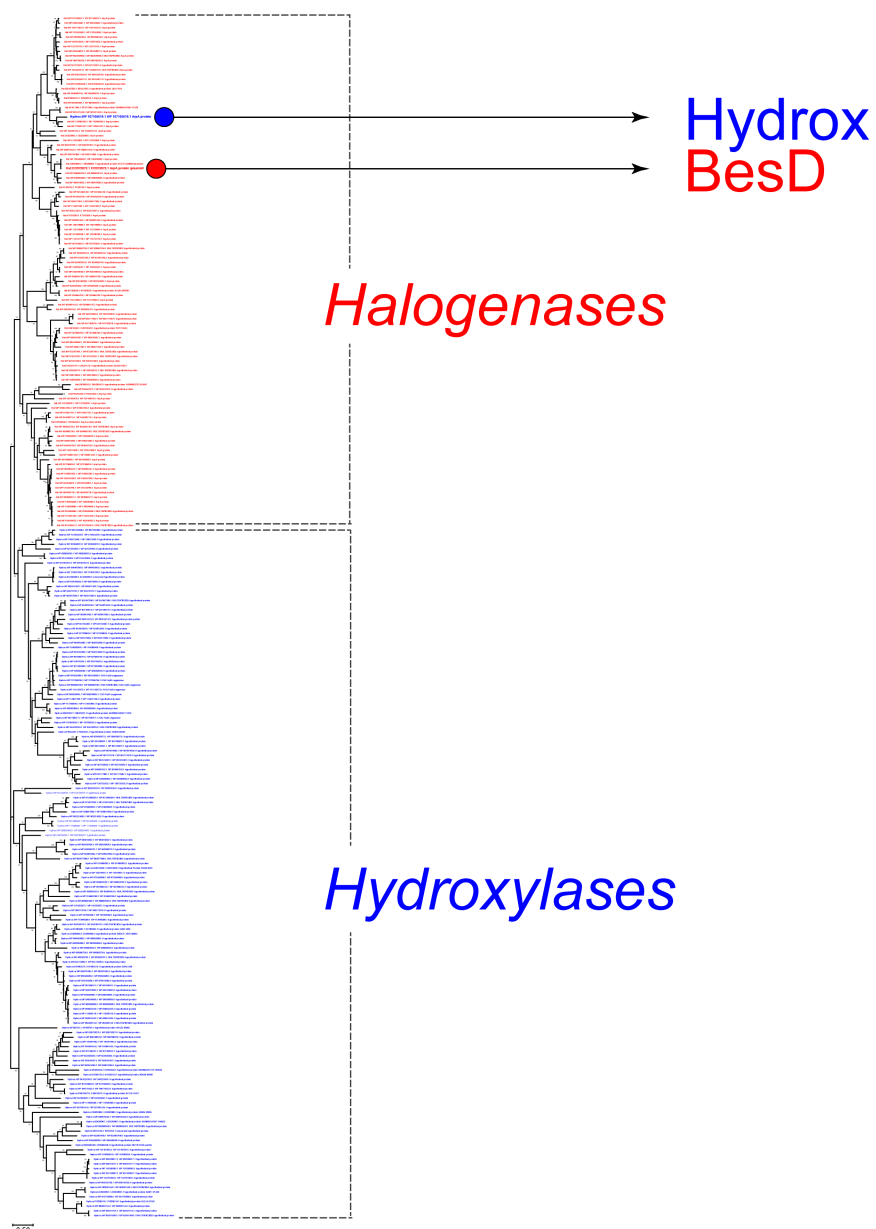


Figure 4.2. Phylogenetic analysis of a hydroxylase with homology to BesD. Phylogenetic tree of BesD halogenase (red) and hydroxylase (blue) homologs from the NCBI non-redundant protein database assembled in MEGA [32] following sequence alignment in MUSCLE [16]. The assignments are based on the presence of the HXG/A (halogenase) or HXD/E (hydroxylase) sequence motifs. BesD from *S. cattleya* and the hydroxylase from *S. sp.* MBT76 (Hydrox, WP_107105619) are annotated on the tree.

A

% ID	BesD	Hydrox	Hal
BesD	-	44.6	51.0
Hydrox	44.6	-	71.7
Hal	51.0	71.7	-

B

```

S. cattleya (BesD)      MGSNRQELKD----VCAPLEKDDIRRLSQAFHRFGIV---TVTELIEPHTRKLVRAEAD
S. sp. MBT76 (Hydrox)  --MDVHEIDETLEKFLAENYTPERVQQLADRFQRTGFVKFDSHMRIVPEELITAVRAEAD
A. teichomyceticus (Hal) --MNVEQIDENLAKFLAERYTPESVAQLADRFHRFGFVKFDAANRLVPDELQTAVREECD
                        :  :::      :      :  : .*: : *: * : * :      :  :      . ** * . *

S. cattleya (BesD)      RLLDQYAERRDLRLATTDYTRRSMSVVPSETIAANSELVTGLYAHRELLAPLEAIAGERL
S. sp. MBT76 (Hydrox)  RLVREHKERRDLVLGTTGGTPRNLSVVKSQDV-EQSDLIRAVTRSEVLLTFLAGITRERI
A. teichomyceticus (Hal) LLIEQHKERRNLLSTTGNTPRMSVVKSEEI-EKSELISLTSRSEVLLGFLAGITREEI
                        * : : : * : * : * . * * : * : * : * : : * : * : * : * : * :

S. cattleya (BesD)      HPCPKADEEFLITRQEQRGDTHGWHWGDFS FALI WVLQAPPIDVGGLLQCVPHTTWDKAS
S. sp. MBT76 (Hydrox)  IPEVSDDERYLITHQEFASDTHGWHWGDYSFAFNWALRMPPIASGGMVQAVPHTHWDKNA
A. teichomyceticus (Hal) IPEVSSDERYLITHQEFKSDTHGWHWGDYSFALI WALRMPPIEHGGMLQAVPHTHWDKSN
                        * . ** : * : * . * . * : * : * : * : * : * : * : * : * : * :

S. cattleya (BesD)      PQINRYLVENPIDTYHFESGDVYFLRTDTTLHRTIPLRED-TTRIILNMTWAGERDLSRK
S. sp. MBT76 (Hydrox)  PRINETLCERQIDTYGLVSGDLYLLRSDTTMHRTVPLTEDGAVRTMLVVSWSAERDLGKV
A. teichomyceticus (Hal) PRINQTLCEINTHGLESGLYLLRTDTTLHRTVPLSED-STRTILNMTWAAKRDLEKD
                        * . * * * * . * : : : * : * : * : * : * : * : * : * : * : * :

S. cattleya (BesD)      LAADDRWWDNAEVSAAAIKD--
S. sp. MBT76 (Hydrox)  LTGNDRWWENPEAGAAQPVHRAG
A. teichomyceticus (Hal) LVGNDRWWENPEAEAAARAVDNA-
                        * . : * : * : * . * . * :

```

Figure 4.3. Sequence alignment of BesD, Hal, and Hydrox. (A) Sequence identity matrix for the halogenases from *S. cattleya* (BesD) and *A. teichomyceticus* (Hal, WP_122981682), and the hydroxylase from *S. sp. MBT76* (Hydrox, WP_107105619). (B) Sequence alignment of BesD, Hal, and Hydrox highlighting the HXG and HXD motifs. Alignments were performed in MUSCLE [16].

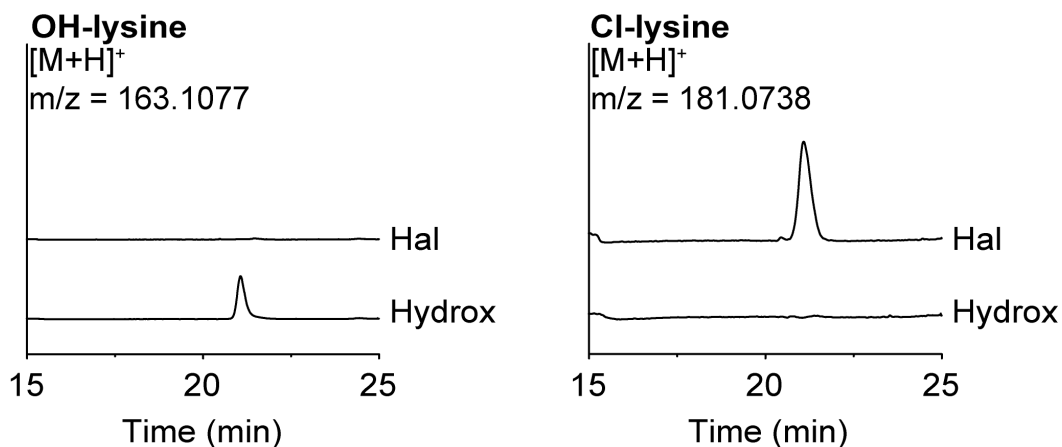
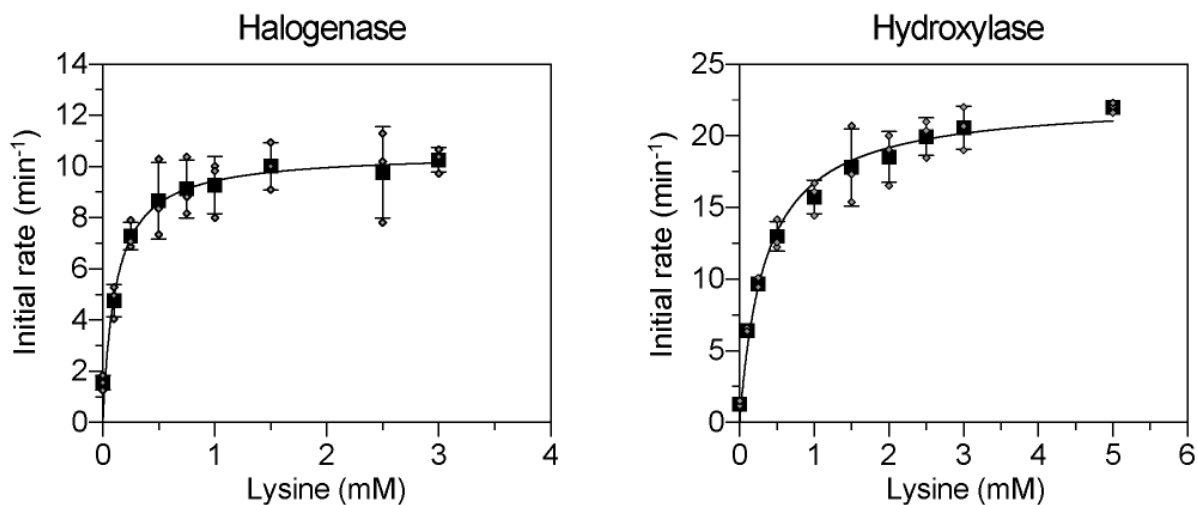


Figure 4.4. LC/MS analysis of Hydrox and Hal. Reactions (50 μ L) contained L-lysine $\cdot$ HCl (3 mM), sodium α KG (5 mM), sodium ascorbate (5 mM), $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ (1 mM), and sodium chloride (5 mM) in 100 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of purified Hal or Hydrox (20 μ M final concentration) and allowed to proceed for 30 min at room temperature before quenching in 2 vol of methanol with 1% (v/v) formic acid. Following centrifugation at $18,000 \times g$ for 20 min ($^\circ\text{C}$) to remove precipitated protein, samples were analyzed by LC/MS on an Agilent 1290 UPLC/QQQ. Extracted ions chromatograms are shown for chlorolysine ($m/z = 181.0738$) and hydroxylysine ($m/z = 163.1077$) for each sample. The data are representative of 3 experimental replicates.



	k_{cat} (min ⁻¹)	K_{M} (mM)	$k_{\text{cat}}/K_{\text{M}}$ (mM ⁻¹ min ⁻¹)
Hal	10.5 ± 0.4	0.12 ± 0.03	91 ± 15
Hydrox	22.5 ± 0.5	0.34 ± 0.05	66 ± 13

Figure 4.5. Steady-state kinetic analysis of Hal and Hydrox with the amino acid lysine as a substrate. Data are mean ± sd (n = 3 technical replicates). Table contains k_{cat} , K_{M} , and $k_{\text{cat}}/K_{\text{M}}$ calculated by non-linear curve fitting to the Michaelis-Menten equation. Data in the table (kinetic parameters) are mean ± s.e. Error in $k_{\text{cat}}/K_{\text{M}}$ is obtained by propagation from the individual kinetic terms.

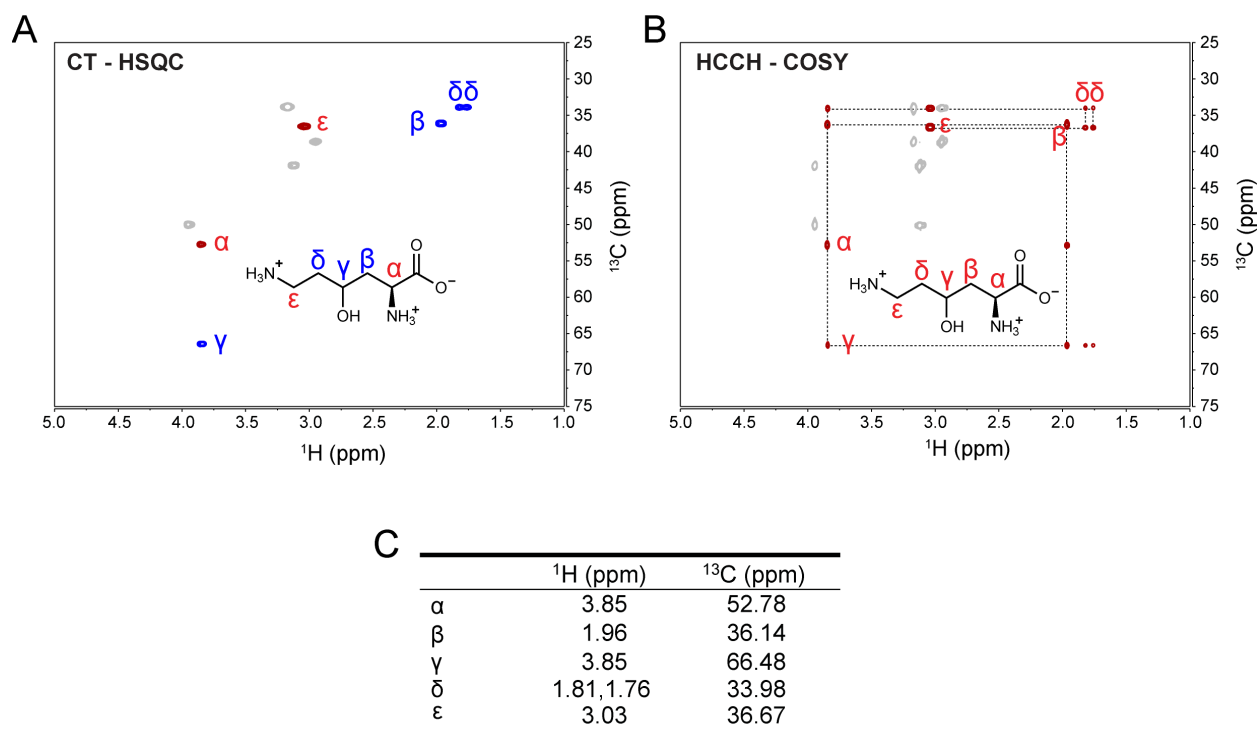


Figure 4.6. NMR analysis of 4-OH-lysine. Hydrox from *S. sp* MBT76 was incubated with fully ^{15}N - and ^{13}C -labeled lysine (1 mM), Fe (0.25 mM), αKG (5 mM), and sodium ascorbate (1 mM) for 16 h before addition of acetonitrile to precipitate the protein. Following centrifugation and removal of precipitated protein, the Hydrox product was isolated by HPLC and screened by LC/MS. Purification of the compound was performed once. (A) 2D ^1H - ^{13}C CT-HSQC was used to confirm assignment of C_α and C_ϵ of $[\text{}^{15}\text{N}_2, \text{}^{13}\text{C}_6]$ -4-hydroxylysine produced by Hydrox. In this experiment carbons with two neighbors (blue) have opposite phase of carbons with one or three neighbors (red). (B) 2D ^1H - ^{13}C HCCH COSY was used to assign connectivity of carbons of $[\text{}^{15}\text{N}_2, \text{}^{13}\text{C}_6]$ -4-hydroxylysine produced by Hydrox. (C) Chemical shifts determined from the 2D-NMR data.

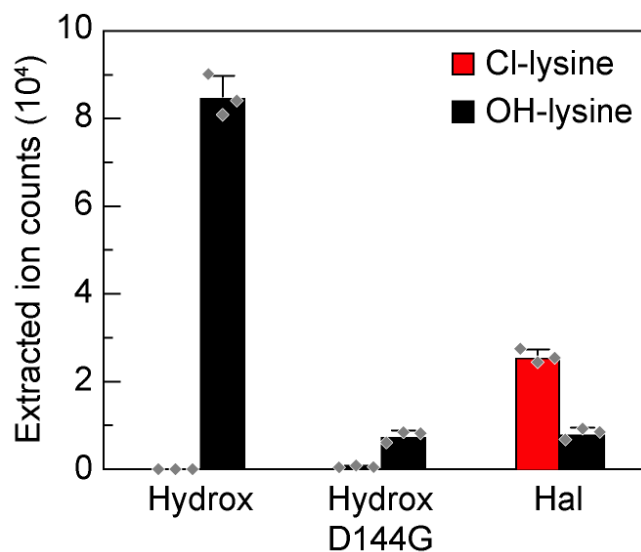


Figure 4.7 Mutagenesis of the HXD motif of Hydrox to HXG does not lead to efficient halogenation. Reactions (50 μ L) contained L-lysine $\cdot$ HCl (3 mM), sodium α KG (5 mM), sodium ascorbate (5 mM), $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ (1 mM), and sodium chloride (5 mM) in 100 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of purified Hal or Hydrox variants (20 μ M final concentration) and allowed to proceed for 25 min at room temperature before quenching in 2 vol of methanol with 1% (v/v) formic acid. Following centrifugation at $18,000 \times g$ for 20 min ($^\circ\text{C}$) to remove precipitated protein, samples were analyzed by LC/MS on an Agilent 1290 UPLC-6530 QTOF. Integrated extracted ions counts are shown for chlorolysine ($m/z = 181.0738$) and hydroxylysine ($m/z = 163.1077$) for each sample. Data are average and standard deviation of technical replicates ($n = 3$).

in Hydrox did not enable halogenation, we next sought to investigate the role of other residues in determining the reaction outcome.

Crystal structure of Hydrox. In order to identify differences in the second-sphere coordination environment between lysine halogenases and hydroxylases, we solved an anaerobic X-ray crystal structure of Hydrox to 2.4 Å resolution (*Table 4.1, Figure 4.8*). The protein was co-crystallized with lysine and α KG in a coy anaerobic chamber, and then soaked anaerobically with $\text{Fe}^{\text{II}}\text{Cl}_2 \cdot 4\text{H}_2\text{O}$ to reconstitute the reactant complex. The protein crystalized with 4 copies per asymmetric unit. Chains A, B, and C have density for Fe, α KG, and lysine in the active site, while chain D lacks density for the cofactors and substrate. Note that the enzyme in chain D is in an alternate conformation; the C-terminal covering lid, which closes over the substrate in chains A-C is disordered, and the domain containing Arg80, which binds the carboxylate of the lysine substrate, is projected away from the active site and adopts an alternate secondary structure (*Figure 4.9*).

A comparison of Hydrox and the halogenase BesD (6NIE) reveals many similarities between the two structures. In the following analysis, residue numbering is based on the Hydrox sequence. The overall fold of the two enzymes is highly conserved, with an RMSD of 1.7 Å (*Figure 4.10*). The enzymes also share a highly conserved lysine binding site; In Hydrox, the carboxylate-binding residues (Arg80, His139, Trp244) and α -amine-binding residues (Asp145) are identical to those found in the halogenase. In addition, Trp245, which closes over lysine in the active site, is also conserved. The most notable difference lies at the ϵ -amine-coordinating residues (Glu125, Val225, Ser227). Asn225, which coordinates the ϵ -amine via a water in BesD, is replaced by Val225 in Hydrox. In addition, Ser227 is instead a Thr in BesD. Based on the structures, the overall fold and lysine substrate binding modes of the two enzymes appears to be highly conserved.

However, the orientation of α KG differs between the structure of Hydrox and BesD (*Figure 4.11*). In both enzymes, the C₁ carboxylate of α KG is trans to His209. However, in BesD, the complex displays a distorted geometry due to hydrogen bonding between Asn225 and the C₁ carboxylate of α KG. In Hydrox, the C₁ carboxylate of α KG is instead directly trans to His209. The difference in the angle of rotation between the α KG in BesD and Hydrox is $\sim 20^\circ$. This is likely because the Asn225 found in BesD is replaced by a valine residue in Hydrox, which cannot participate in hydrogen bonding. The crystal structure reveals that second sphere residues can alter the configuration of α KG in the respective reactant complexes of the halogenases and hydroxylases. We next wondered whether these second sphere residues might contribute to the observed reaction outcome in these enzymes. To gain further insight, we developed an independent screening-based strategy to identify key residues required for halogenation.

Designing a platform to screen enzyme variants. We envisioned that important sequence components could be identified by constructing a DNA-shuffled library of Hydrox and Hal chimeras and screening the variants to determine whether they perform halogenation or hydroxylation. We elected to use Hydrox D144G as a template, rather than wild type Hydrox, to minimize library members that are incapable of halogenation (since Gly144 is essential for chloride binding and halogenation). To maximize the efficiency of DNA shuffling, we codon-optimized Hal to achieve 85% sequence identity at the nucleotide level when compared to the hydroxylase [33]. Following PCR-amplification of Hal and Hydrox D144G, equimolar amounts of the amplified DNA were digested with DNaseI and reassembled using PCR, yielding a library of chimeric proteins composed of the two parent sequences (*Figure 4.12*).

Table 4.1. Data collection and refinement parameters for Hydrox with Lysine, α KG, and Fe bound in the active site.

Hydrox with Lys, α KG, and Fe	
Data collection	
Space group	P12 ₁ 1
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	72.626, 93.53, 91.339
α , β , γ (°)	90, 107.564, 90
Resolution (Å)	87.08–2.4 (2.486–2.4)*
<i>R</i> _{sym} or <i>R</i> _{merge}	0.2843 (2.156)
<i>I</i> / σ <i>I</i>	5.62 (1.10)
Completeness (%)	99.4 (98.78)
Redundancy	6.9 (6.9)
CC _{1/2}	0.986 (0.389)
Refinement	
Resolution (Å)	87.08–2.4
No. reflections	45400 (4454)
<i>R</i> _{work} / <i>R</i> _{free}	0.192/0.230
No. atoms	8399
Protein	7913
Ligand/ion	70
Water	416
<i>B</i> -factors	46.12
Protein	46.17
Ligand/ion	40.71
Fe	40
Lysine	45
α KG	34
Water	46.03
R.m.s. deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.93

*Values are reported for the highest-resolution shell.

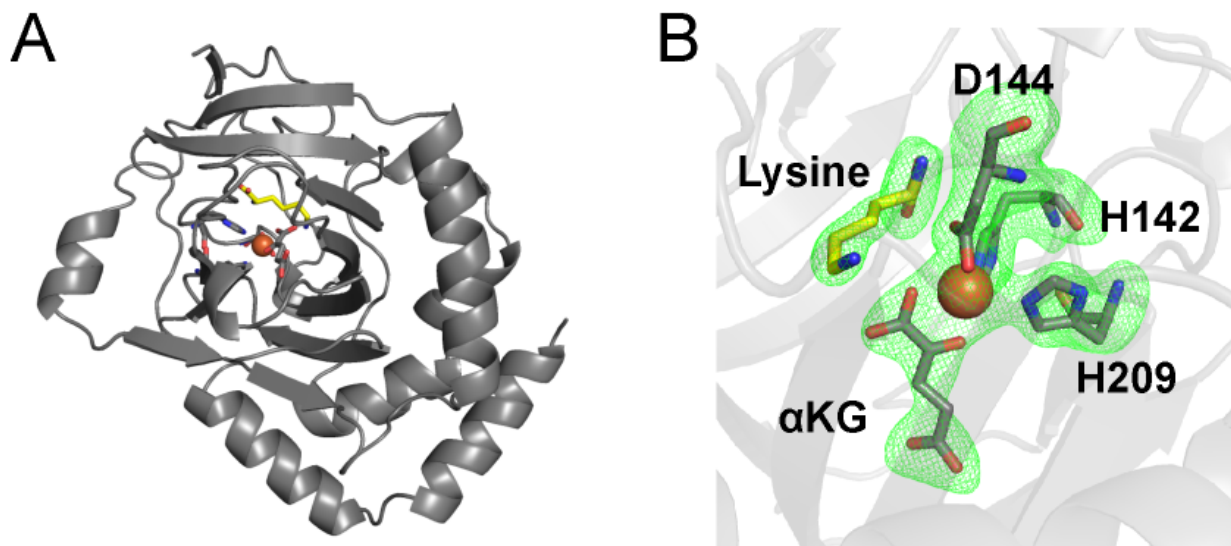


Figure 4.8. Crystal structure of Hydrox at 2.4 Å resolution. (A) Overall fold of Hydrox. (B) $F_o - F_c$ omit map (contoured at 2σ) for lysine (yellow), α KG (grey), Fe (brown), and the facial triad residues (His 142, His 209, and Asp 144).

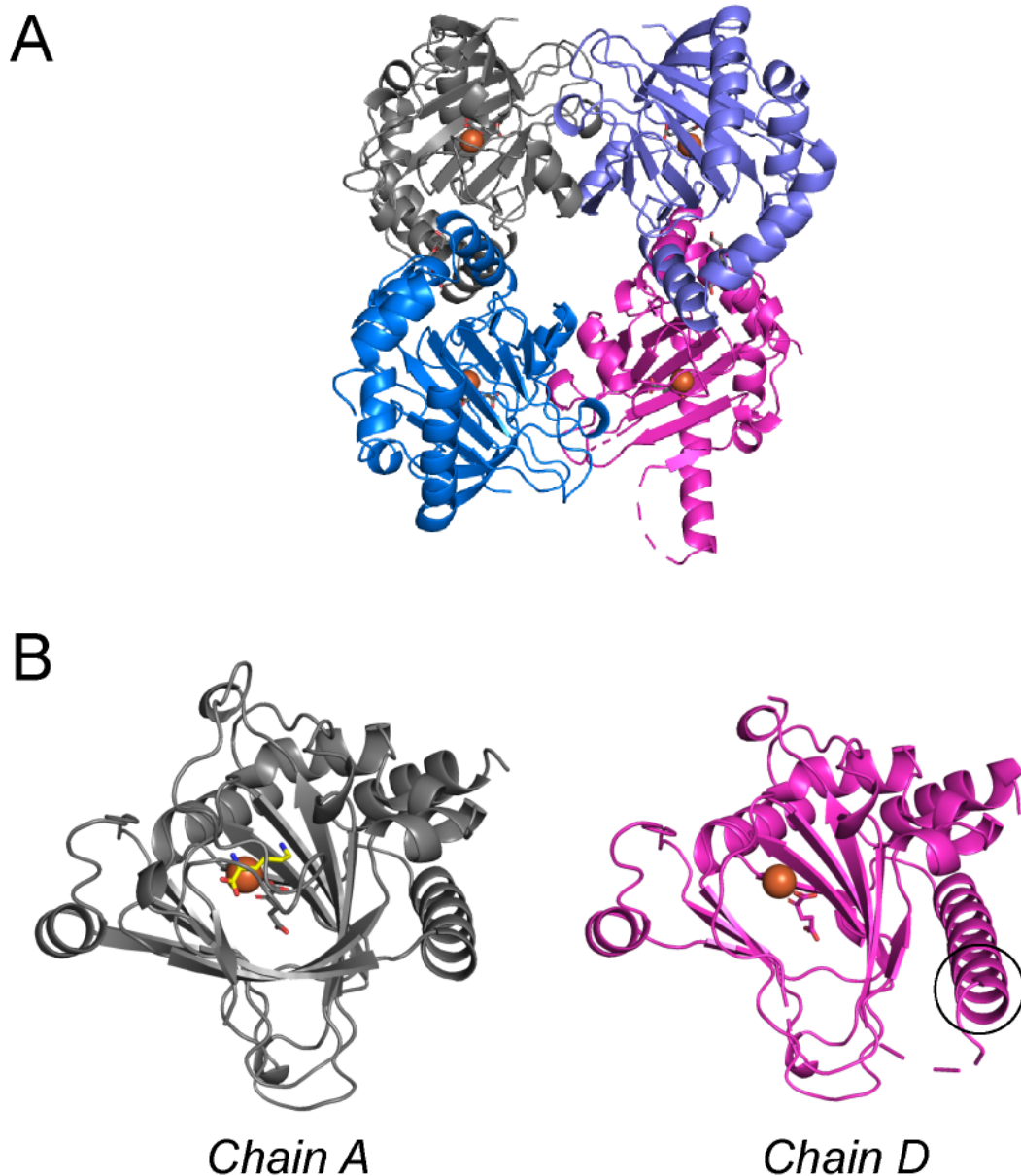


Figure 4.9. Hydrox has 4 chains per asymmetric unit. (A) Depiction of the 4 chains per asymmetric unit. (B) Chain D (magenta) is in a different conformation than the other three chains and lacks density for the substrate lysine in the active site. Chain A (grey) is shown as a representative of chains A-C. The marked helix in chain D (black circle) is extended when compared to chain A. In chain A, those residues instead form antiparallel beta strands that close over the active site and interact with the substrate carboxylate via Arg80. In addition, Chain D lacks density for the substrate lysine and for the C-terminal substrate covering lid that closes over the active site.

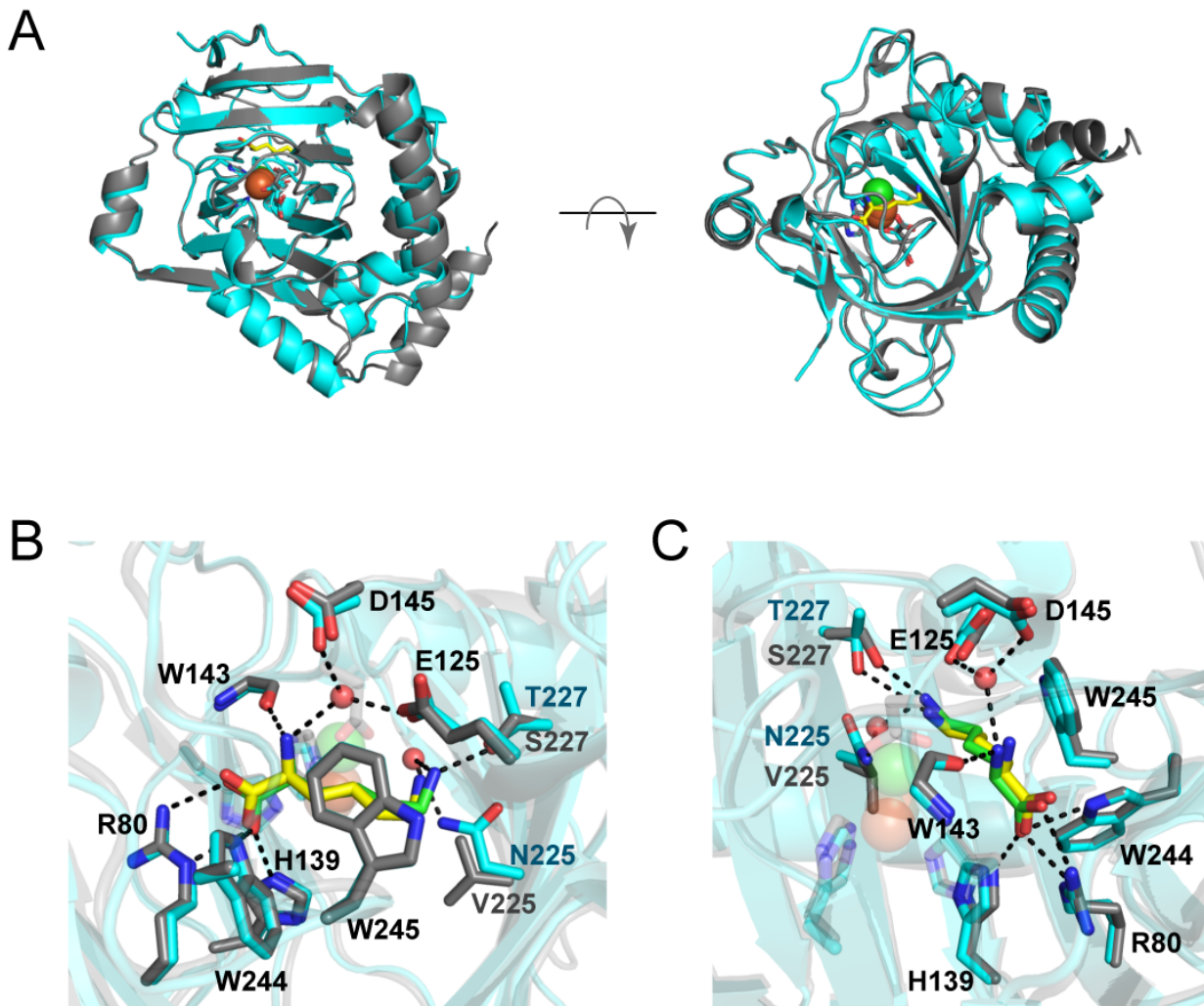


Figure 4.10. Comparison of the Hydrox structure (grey) to the structure of BesD (cyan, PDB 6NIE). (A) An overlay of Hydrox (grey) and BesD (cyan) was performed using Chimera [27]. (B) The lysine binding sites of Hydrox and BesD. Lysine is shown in yellow for Hydrox and in green for BesD. (C) Alternate view of the lysine binding site of Hydrox (grey) and BesD (cyan). The residues in the binding sites are conserved with the exception of positions 225 and 227.

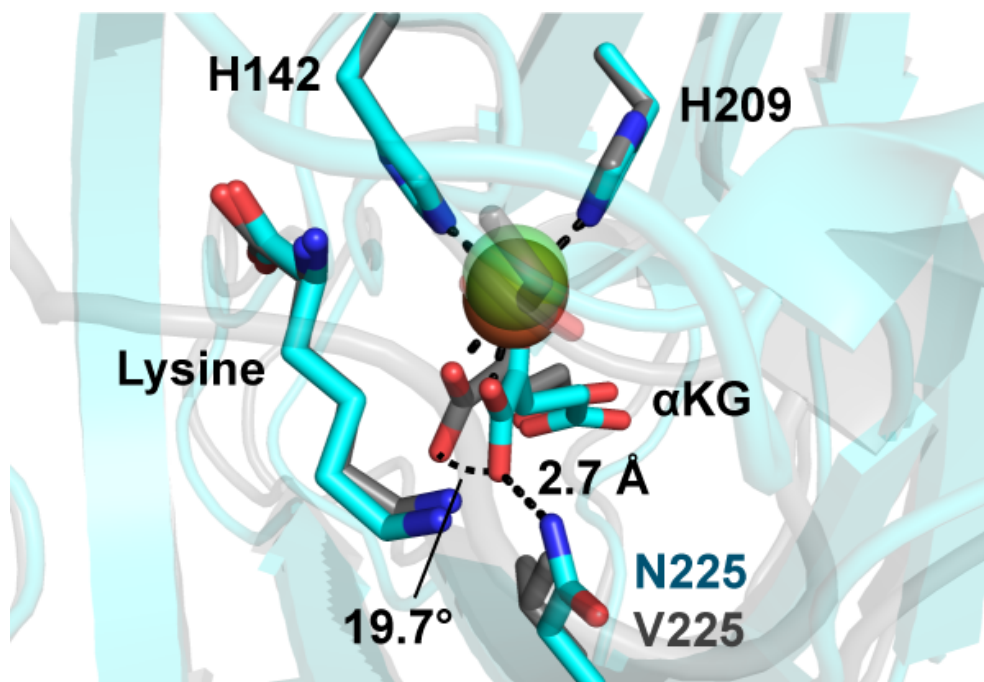


Figure 4.11. Comparison of the α KG binding configuration in Hydrox (grey) and BesD (cyan, PDB 6NIE). In both cases, the O_2 binding site appears to be trans to His142. However, the geometry is distorted in the halogenase due to interaction of N225 and α KG. The resulting α KG (Hydrox) -iron- α KG (BesD) angle is $\sim 20^\circ$. The two structures were overlaid using the match function in Chimera [27] with His142, His209, and Fe as the match targets.

Design and optimization of CuAAC-based screening. To probe the reaction outcome of the shuffle library variants, we developed a high-throughput screen based on the role of 4-Cl-lysine as an intermediate in the alkyne amino-acid biosynthetic pathway. We previously discovered the halogenase BesD in the process of elucidating the biosynthesis of β -ethynylserine (β es), an alkyne-containing amino acid [14]. In this pathway, BesD produces 4-Cl-lysine, which is then converted to 4-Cl-allylglycine by BesC, and then to propargylglycine (Pra) by BesB. Thus, in two enzymatic steps, 4-Cl-lysine can be converted into Pra, which contains an alkyne functional group that can be leveraged for copper-catalyzed azide-alkyne cycloaddition (CuAAC) chemistry with a fluorogenic probe [34]. We envisioned that cells containing the two downstream enzymes, BesC and BesB, could be transformed with a library of Hal/Hydrox mutants and screened for fluorescence. Variants that yielded halogenases would result in a fluorescent signal. However, variants that performed only hydroxylation would yield no fluorescence, as 4-OH-lysine is not a substrate for BesC (Figure 4.13). We initiated assay development with 3-azido-7-hydroxycoumarin as the fluorogenic probe (Figure 4.14) [35].

We optimized screening conditions in order to achieve the sensitivity required to distinguish halogenation from hydroxylation. We first sought a cell culture medium that would facilitate robust cell growth while enabling the CuAAC reaction to proceed efficiently. It has been shown that for CuAAC-based screens, the culture media can greatly influence the screening outcome [34]. We therefore compared the CuAAC reaction in LB, TB, 2 $\times$ YT, autoinducing, M9 media, and M9 with additives (magnesium sulfate (2 mM), calcium chloride (0.1 mM), glucose (4% (w/v)), thiamine (1 mg/L), 0.1 mM each of the 20 canonical amino acids). The reaction of 3-azido-7-hydroxycoumarin with Pra in the presence of copper (II) sulfate, THPTA, and sodium ascorbate gave the greatest fluorescent signal in the presence M9 without additives, while LB media performed second-best of the attempted conditions (Figure 4.15AB). By systematically including M9 additives one at a time, we determined that the amino acid mixture was responsible for the signal suppression (Figure 4.15C). We therefore chose to proceed with LB and M9 media without amino acid supplementation for subsequent experiments.

To test whether the fluorogenic screen could detect Pra generated by cells expressing the halogenase and the downstream enzymes BesB and BesC, we generated chemically competent *E. coli* harboring BesB and BesC on a pACYC vector under a T7 promoter. We then transformed these cells with either Hal or Hydrox, grew the cells in a 96-well plate format, and induced protein expression in either M9 or LB media supplemented with 0.5 mM lysine. After 2 days, the cell supernatant was collected and subjected to CuAAC with 3-azido-7-hydroxycoumarin ($\lambda_{\text{ex}} = 404$ nm, $\lambda_{\text{em}} = 477$ nm). In M9 media, we achieved a ~6-10-fold change in fluorescence for cells expressing Hal compared to cells transformed with Hydrox (Figure 4.16). These data indicate that

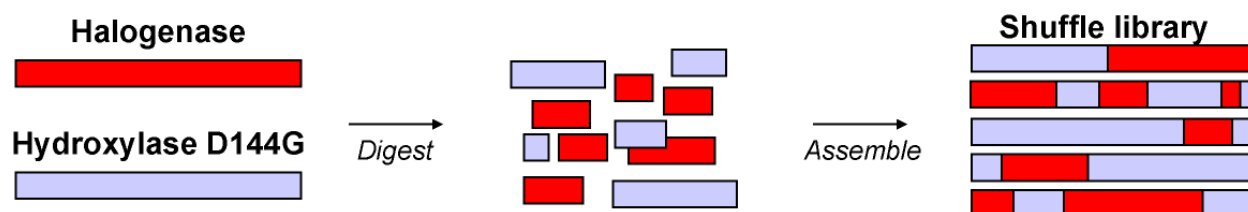


Figure 4.12. Shuffle library design. DNA encoding the wild type halogenase and the hydroxylase D144G mutant were digested with DNaseI and reassembled by PCR.

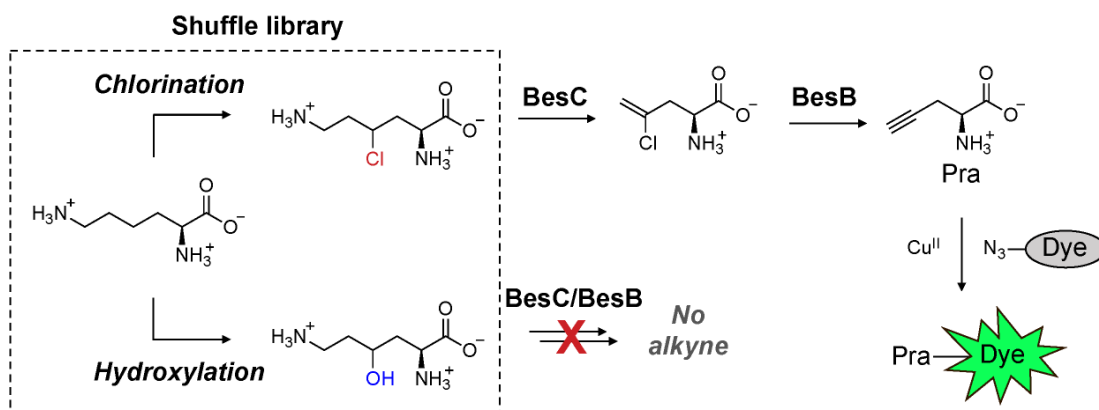


Figure 4.13. Strategy for screening the shuffle library for variants that perform halogenation. In two enzymatic steps, 4-Cl-lysine can be converted into Pra, which contains an alkyne functional group that can be leveraged for CuAAC chemistry with a fluorogenic probe. In contrast, variants that perform only hydroxylation would yield no fluorescent product downstream, as 4-OH-lysine is not a substrate for the enzymes in the alkyne biosynthetic pathway.

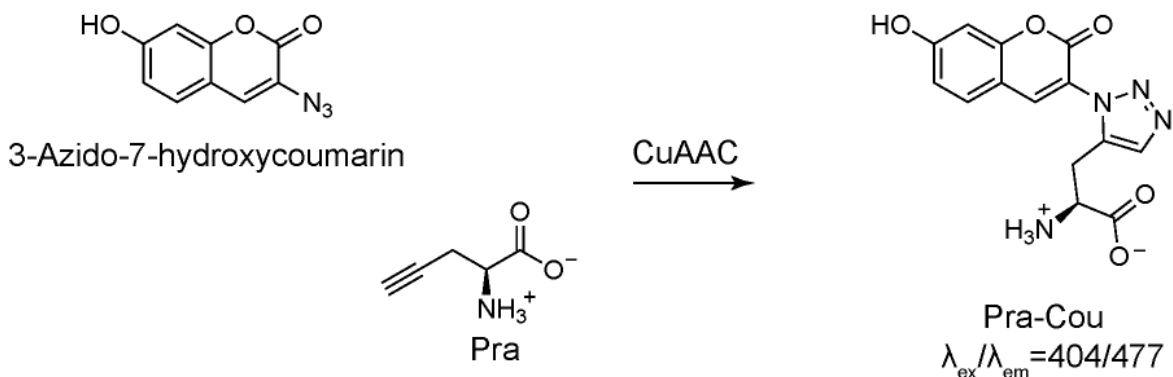


Figure 4.14. Propargylglycine (Pra) produced by cells expressing halogenase variants can react with 3-azido-7-hydroxycoumarin in the presence of copper (II), ascorbate, and the THPTA ligand. The resulting propargylglycine-coumarin product (Pra-Cou) is fluorescent and can be detected using a plate reader.

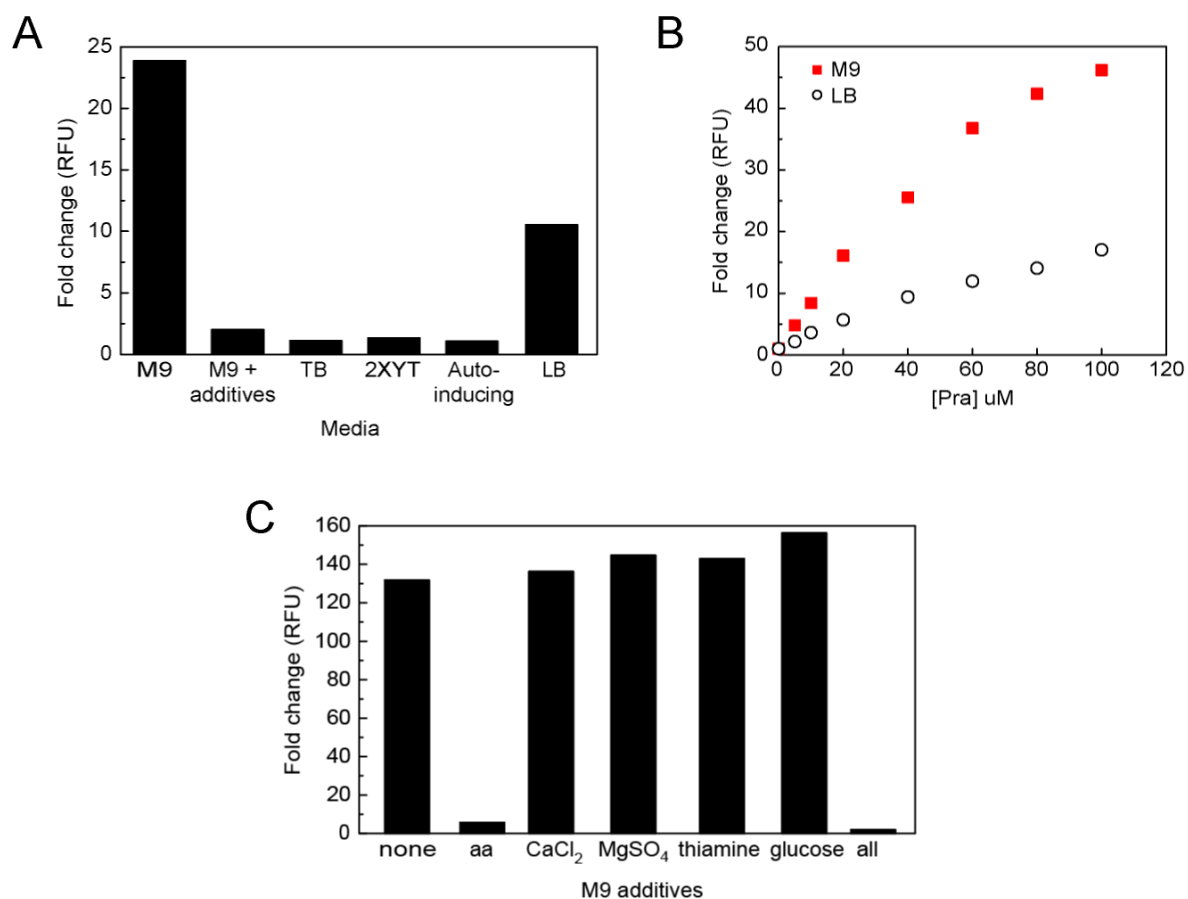


Figure 4.15. Testing the CuAAC reaction in different cell culture media. (A) Comparison of M9, TB, 2XYT, Autoinducing, and LB media. The reactions contained 3-azido-7-hydroxycoumarin (0.5 mM), Pra (1 mM), copper (II) sulfate (0.5 mM), THPTA (1 mM), and sodium ascorbate (5 mM). Reactions (100 μ L) were performed for 30 min before analyzing fluorescence ($\lambda_{\text{ex}} = 404$ nm, $\lambda_{\text{em}} = 477$ nm) using a SynergyMx Microplate Reader (BioTek) at room temperature. (B) Comparison of CuAAC in M9 and LB media. Reaction conditions are as in part (A), but with varying concentrations of Pra (1-100 μ M). (C) CuAAC performed in M9 with various additives (magnesium sulfate (2 mM), calcium chloride (0.1 mM), glucose (4% (w/v)), thiamine (1 mg/L), 0.1 mM each of the 20 canonical amino acids).

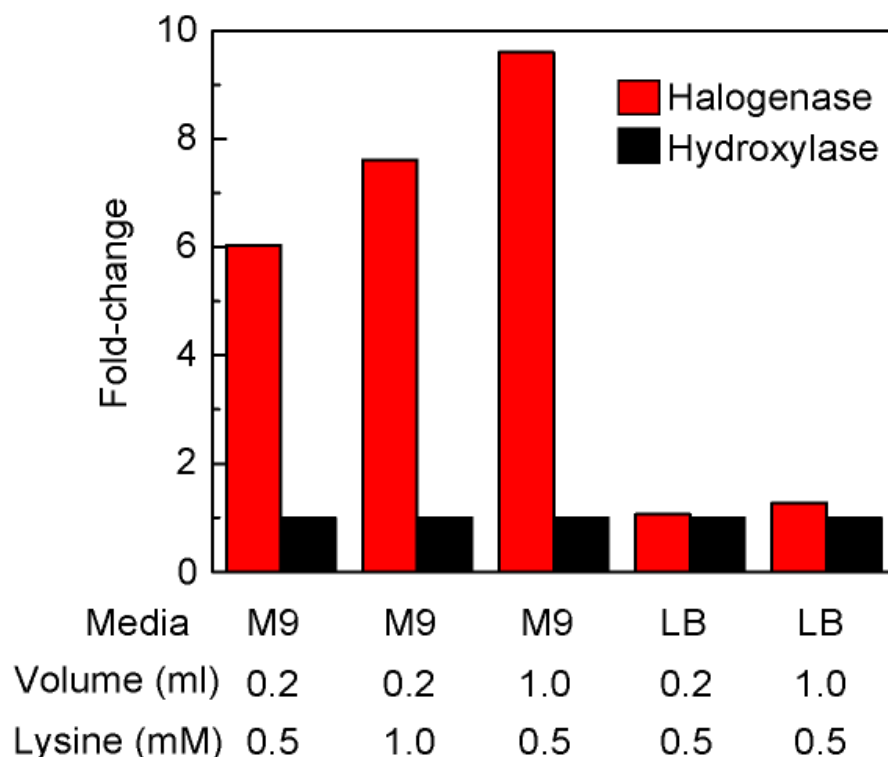


Figure 4.16. The alkyne screening platform enables detection of chlorolysine produced by the halogenase via CuAAC. Cells expressing either the halogenase (red) or hydroxylase (black) along with downstream enzymes BesB and C were grown in 96-well plates 16 deg. C for 48 hours. Samples contained varied media types (M9 or LB), culture volumes (200 μ L or 1 ml), and lysine addition (0.5 or 1.0 mM). After the 48 h growth, cells were pelleted and 50 μ L of supernatant was mixed with 50 μ L of reaction solution to yield the final reaction mixture (100 μ L, 3-azido-7-hydroxycoumarin (0.5 mM), Pra (1 mM), copper (II) sulfate (0.5 mM), THPTA (1 mM), and sodium ascorbate (5 mM)). After 30 min reaction time, fluorescence was measured using a SynergyMx Microplate Reader (BioTek) at room temperature (λ_{ex} = 404 nm, λ_{em} = 477 nm). The data shown above are representative of n = 2 experimental replicates.

the CuAAC screen can be used to distinguish between chlorination and hydroxylation outcomes. The data additionally show that supplementing the cultures with lysine improved the signal for the screen, likely by increasing the substrate pool available to the halogenase.

Finally, we compared two fluorogenic probes for the CuAAC reaction to determine which would provide the greatest sensitivity for Pra detection. The CalFluor probes have been shown to demonstrate high sensitivity for the alkyne functional groups [34,36]. We therefore compared the sensitivity of CalFluor488 to our original probe, 3-azido-7-hydroxycoumarin, across a range of Pra concentrations (*Figure 4.17*). Although 3-azido-7-hydroxycoumarin provided a linear response across a wider range of Pra concentrations, CalFluor488 yielded a greater sensitivity at low Pra concentrations (*Figure 4.18*). We therefore proceeded to use CalFluor488 for our subsequent experiments.

Library screening and analysis. With this assay in hand, we screened 216 variants from the shuffle library and found that 30 of the variants yielded fluorescence upon CuAAC with the CalFluor 488 fluorogenic probe (*Figure 4.19*). We sequenced the 30 hits that gave greater than 3-fold increase in fluorescence in order to determine which residues were conserved in the variants that performed halogenation (*Figure 4.20-4.21, Appendix A2.1*). We found that 5 of the hits contained the sequence of the parent halogenase residue, but the remaining 25 were chimeras. Analysis of the chimeras revealed that a majority of the residues were not critical for halogenation. In particular, the N-terminus through residue ~120 was highly variable in the hits; even variants containing the N-terminal half from Hydrox were still able to perform halogenation. The data suggest that the residues that are critical for halogenation do not lie on the N-terminal half of the protein sequence.

However, 11 residues (and a single amino acid deletion) were conserved in 100% of the halogenase hits, suggesting they may contribute to the observed reaction outcome. These residues were mapped onto the crystal structure of Hydrox in order to gain insight into their roles (*Figure 4.22*). Based on this analysis, we found that the mutations that were conserved in all of the halogenase hits were located along two of the beta strands that comprise the core of the protein. Note that DNA shuffling requires stretches of sequence homology to enable crossovers to occur. Therefore, the contiguous nature of the conserved residues may be partially due to low identity between the halogenase and hydroxylase templates in those regions, as this would prevent shuffling between those stretches.

We next compared the residues of the Hydrox and Hal at the conserved positions. Based on this analysis, two residues stood out: Ile 151 and Asn225 are conserved in all of the halogenase hits and are located near the α KG in the active site. In contrast, the corresponding residues in wild-type Hydrox are Asn151 and Val225. Notably, Hal has a nonpolar residue at 151 and a polar one at 225, while this trend is reversed in the hydroxylases. We previously found that mutation of the equivalent Asn225 in BesD to Ala diminished halogenation selectivity [28]. Thus, the recovery of Asn225 within our hits is consistent with the importance of this residue for halogenation and suggests that the shuffle screen can be used to identify critical sequence components.

To gain further insight into the role of residues 151 and 225, we sequenced a random subset of the non-halogenating variants from our screen that failed to give a detectable fluorescence signal (*Figure 4.23-4.24, Appendix A.2.2*). Note that the lack of fluorescence could either be due to

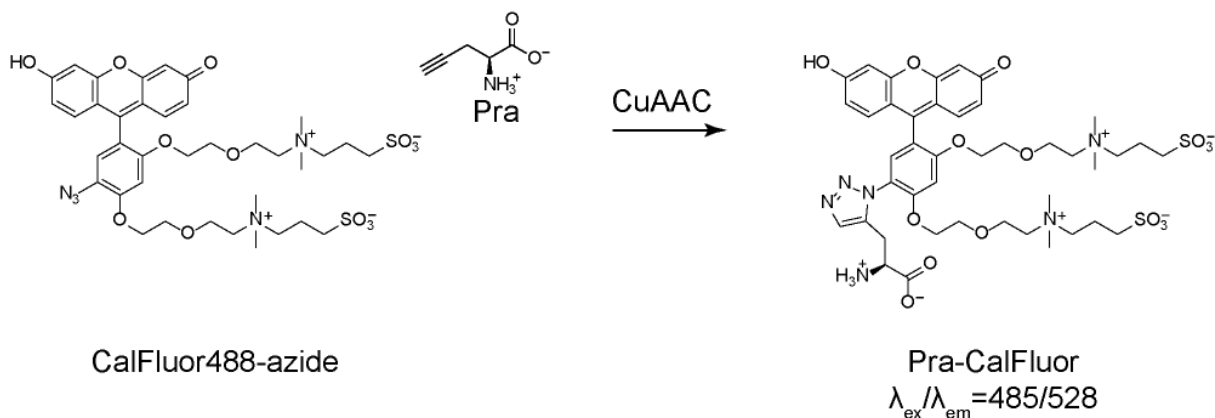


Figure 4.17. Propargylglycine (Pra) produced by cells expressing halogenase variants can react with CalFluor488-azide in the presence of copper (II), ascorbate, and the THPTA ligand. The resulting propargylglycine-CalFluor product (Pra-CalFluor) is fluorescent and can be detected using a plate reader.

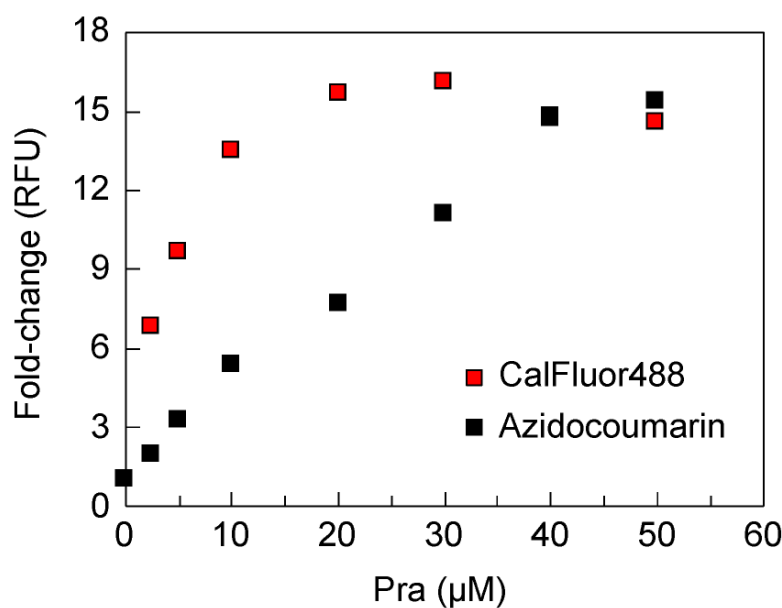


Figure 4.18. Comparison of the sensitivity of 3-azido-7-hydroxycoumarin to CalFluor488 across a range of Pra concentrations. For the CalFluor488 probe, 50 μL M9 media containing Pra was transferred to a fresh 96-well plate and mixed 1:1 with CuAAC solution to yield the final reaction mixture (CalFluor 488 azide (1 μM), CuSO_4 (0.5 mM), BTTA (0.1 mM), and sodium ascorbate (5 mM)). Following incubation in the dark for 15 min, the fluorescence was measured using a SynergyMx Microplate Reader (BioTek) at room temperature ($\lambda_{\text{ex}} = 485 \text{ nm}$, $\lambda_{\text{em}} = 528 \text{ nm}$). For 3-azido-7-hydroxycoumarin, 50 μL of M9 was mixed with 50 μL of reaction solution to yield the final reaction mixture (100 μL , 3-azido-7-hydroxycoumarin (0.2 mM), Pra (1 mM), copper (II) sulfate (0.5 mM), THPTA (1 mM), and sodium ascorbate (5 mM)). After 30 min reaction time, fluorescence was measured using a SynergyMx Microplate Reader (BioTek) at room temperature ($\lambda_{\text{ex}} = 404 \text{ nm}$, $\lambda_{\text{em}} = 477 \text{ nm}$). In both cases, the fold-change is calculated relative to the fluorescence value of the probe with 0 μM Pra added to the sample.

Plate 1

	1	2	3	4	5	6	7	8	9	10	11	12
A	0.96	1	0.96	1.08	0.96	0.99	0.97	0.91	24.49	27.45	23.77	25.98
B	25.5	19.76	0.77	0.8	0.96	0.79	27.76	0.81	0.74	0.91	0.77	0.78
C	1.18	0.87	22.92	0.8	0.87	1.08	0.85	0.87	0.89	0.78	0.73	16.24
D	0.93	0.86	0.92	0.84	0.82	0.88	0.93	0.82	0.74	0.86	0.94	0.94
E	1.03	1.12	0.95	23.19	0.86	0.79	22.35	0.97	0.92	1.34	0.87	20.67
F	0.83	0.94	0.92	0.82	16.02	0.81	1.02	0.93	0.79	0.93	0.75	0.78
G	26.29	0.94	0.83	0.83	9.82	1.1	0.91	20.68	0.93	0.96	0.82	0.9
H	1.45	0.89	0.88	0.8	0.76	0.78	0.82	0.97	0.79	0.88	17.54	1.54

Plate 2

	1	2	3	4	5	6	7	8	9	10	11	12
A	1.02	1.01	0.98	1	0.9	0.98	0.97	1	21.66	25.32	15.67	13.48
B	0.99	1	0.96	10.11	0.91	0.93	1.54	0.85	0.95	0.96	1.03	1.1
C	1.11	23.39	0.93	0.86	1.05	0.97	0.93	0.83	0.94	1.07	0.87	0.89
D	1.04	0.88	0.85	0.96	9.35	0.86	0.9	0.89	0.81	19.68	1.01	24.98
E	0.92	0.93	0.87	0.86	0.84	0.94	0.83	0.84	17.79	20.28	1	1
F	1.03	1.03	1.04	0.93	0.93	0.95	16.44	0.91	0.84	0.88	0.94	0.95
G	0.08	0.1	0.09	0.09	0.08	0.08	0.1	0.1	0.09	0.08	0.1	0.09
H	1.2	0.95	1.1	0.94	13.55	0.87	1.03	0.97	0.87	1.05	10.17	24.01

Plate 3

	1	2	3	4	5	6	7	8	9	10	11	12
A	0.99	1.01	0.98	1.02	0.95	0.96	0.85	0.91	30.39	28.74	28.12	21.04
B	25.46	0.94	1.18	1.05	2.36	0.94	0.9	0.94	28.76	0.99	1.07	0.98
C	0.89	1.05	0.91	0.87	0.82	0.89	0.87	0.87	1.01	0.84	0.98	0.79
D	1.04	1.93	0.94	0.9	1.03	1.05	27.95	0.91	23.07	16.74	9.08	0.94
E	1.14	1.14	0.9	1.01	1	1	0.84	0.82	0.92	0.86	0.93	0.97

Figure 4.19. Results of screening the shuffle library with CuAAC. Cells expressing BesB, BesC, and a member of the shuffle library are grown in M9 media supplemented with 0.5 mM lysine for two days. Controls are present on each plate as follows: Hydrox (A1-4), Hydrox D144G (A5-8), Hal (A9-12). After two days of expression, cells were pelleted by centrifugation, and the supernatant was transferred to a fresh plate and mixed 1:1 with CuAAC solution to yield the final reaction mixture (CalFluor 488 azide (1 μ M), CuSO₄ (0.5 mM), BTTA (0.1 mM), and sodium ascorbate (5 mM)). Following incubation in the dark for 15 min, the fluorescence was measured using a SynergyMx Microplate Reader (BioTek) at room temperature (λ_{ex} = 485 nm, λ_{em} = 528 nm). Values for each well are reported relative to the average of wells A1-4 (Hydrox) on each plate.

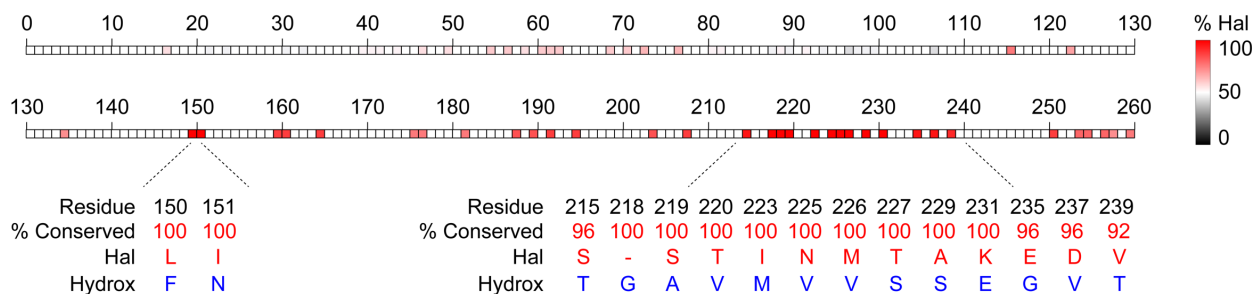


Figure 4.20. Sequence conservation of the Hal parent residues among the sequenced halogenase hits from the shuffle library. The 30 hits from the click screen were sequenced and aligned. Five of the sequences were that of wild type Hal and are excluded from the analysis. The remaining 25 chimeras were analyzed to determine the conservation of the Hal parent residue at each position. If 100% of the halogenase hits possess the Hal residue at a given sequence position, the residue is colored red. If 50% of the hits have the Hal residue, or if a residue is the same in Hal and Hydrox, the position is colored white. Positions at which none of the sequenced hits had the Hal parent residue are colored black. Residues 150, 151, 218-220, 223, 225-227, 229, and 231 are conserved in 100% of the hits.

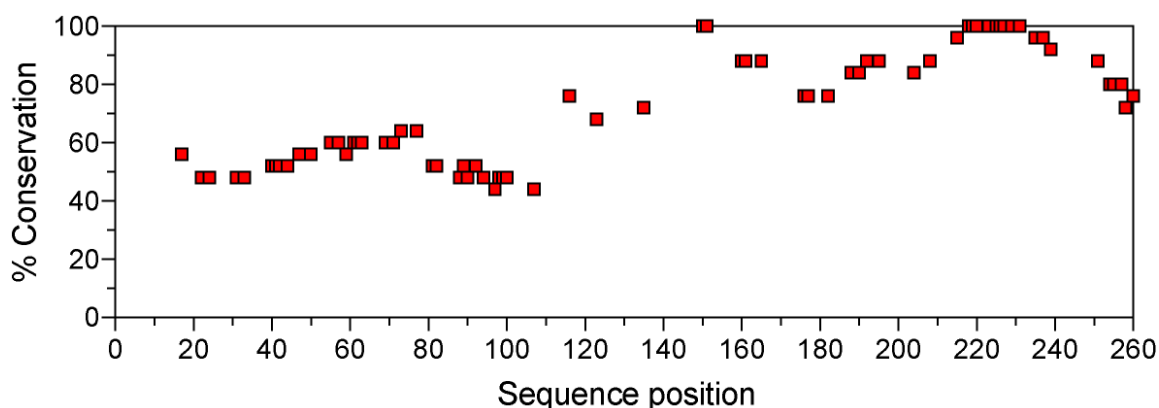


Figure 4.21. Sequence conservation of the Hal parent residues among the sequenced halogenase hits from the shuffle library (alternative representation). The 30 hits from the click screen were sequenced and aligned. Five of the sequences were that of wild type Hal and are excluded from the analysis. The remaining 25 chimeras were analyzed to determine the conservation of the Hal parent residue at each position. Residues 150, 151, 218-220, 223, 225-227, 229, and 231 are conserved in 100% of the hits.

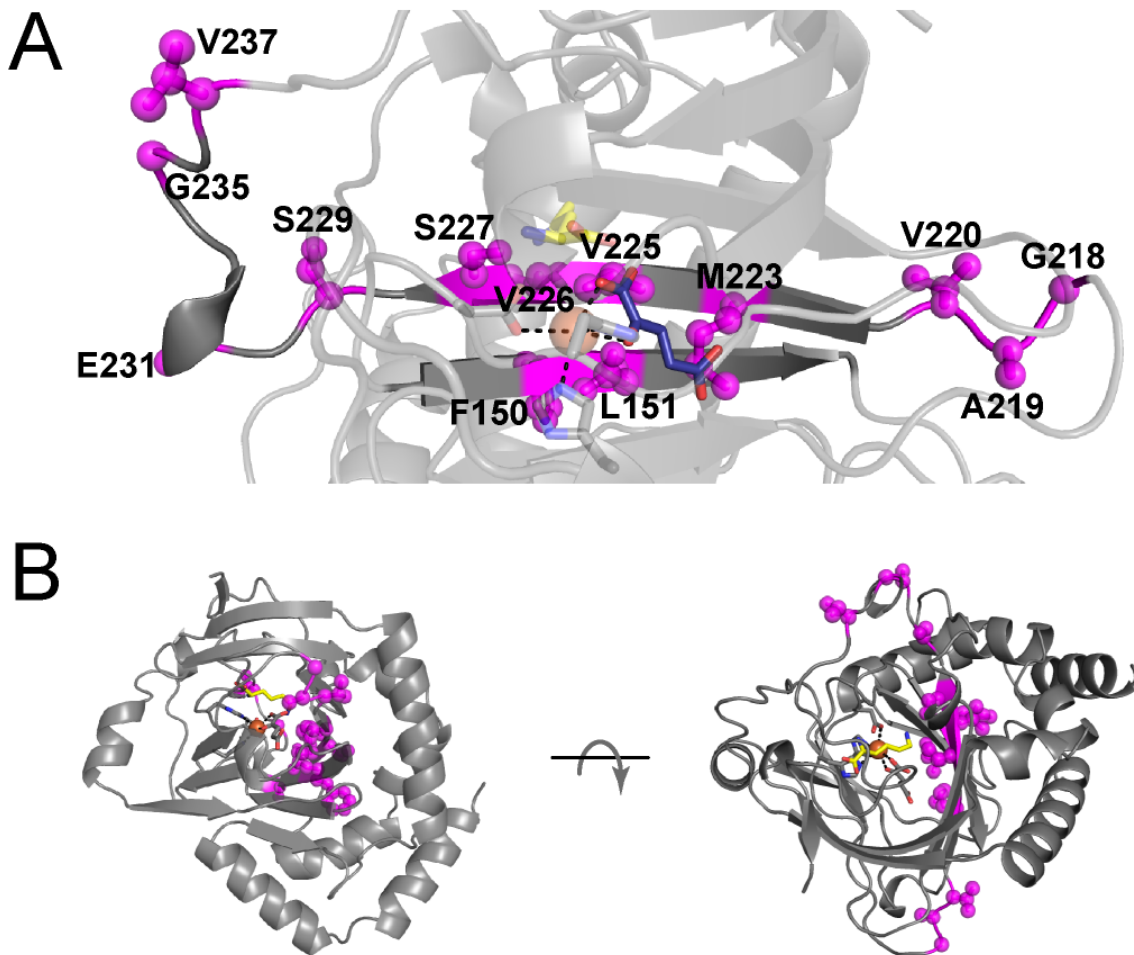


Figure 4.22. Mapping the residues that are conserved in the halogenase library hits onto the structure of Hydrox. (A) The key positions that were conserved in the halogenase shuffle hits are located along two antiparallel beta strands that are adjacent to α KG in the active site. A few of the mutations extend beyond the beta strand to nearby loops. (B) The residues that were conserved in the halogenase hits (magenta) are mapped onto the overall fold of the structure.

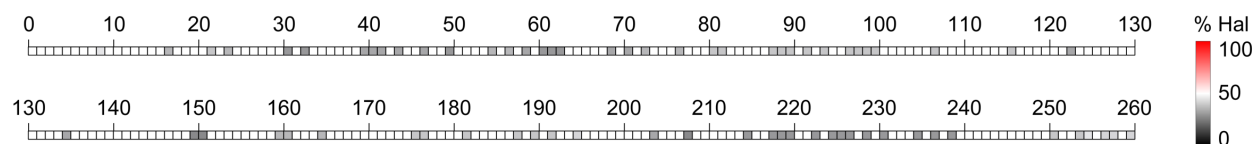


Figure 4.23. Sequence conservation of the Hal parent residues among the randomly sequenced non-halogenating variants from the shuffle library. Forty-six variants were sequenced in total. Thirteen of these yielded poor sequencing reads, or empty vectors. The remaining 33 chimeras were aligned and analyzed to determine the conservation of the Hal parent residue at each position. If 100% of the sequenced constructs possess the Hal residue at a given sequence position, that residue is colored red. If 50% of the hits have the Hal residue, or if a residue is the same in Hal and Hydrox, the position is colored white. Positions at which none of the sequenced hits had the Hal parent residue are colored black. The non-halogenating data set has a low occurrence (<50%) of the Hal parent residue at all sequence positions.

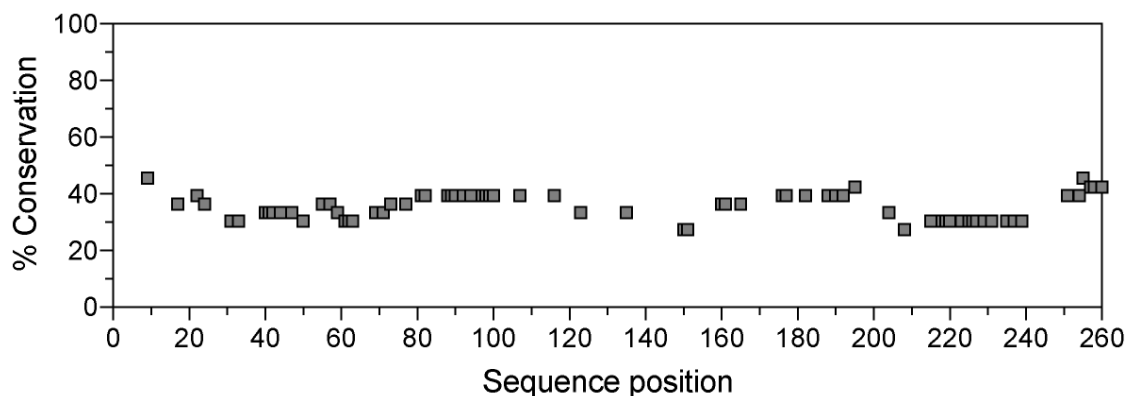


Figure 4.24. Sequence conservation of the Hal parent residues among the randomly sequenced non-halogenating variants from the shuffle library (alternative representation). Forty-six variants were sequenced in total. Thirteen of these yielded poor sequencing reads, or empty vectors. The remaining 33 chimeras were aligned and analyzed to determine the conservation of the Hal parent residue at each position. The non-halogenating data set has a low occurrence (<50%) of the Hal parent residue at all sequence positions.

hydroxylation or due to an inactive enzyme. Out of 46 total non-halogenating sequences, 13 yielded empty vectors, truncated genes, or poor sequencing reads, but 33 of the variants yielded intact genes. We found that these sequences contained various combinations of Asn151, Ile 151, Val225, and Asn225, but none of the sequenced variants had both Ile151 and Asn225 (*Table 4.2*). It is possible that the sensitivity of the screen is insufficient to detect variants with low halogenation activity, or that sequencing even more of the non-halogenating variants would have revealed sequences that possess Ile151 and Asn225 yet fail to yield detectable halogenation signal. However, the data support the hypothesis that both Ile151 and Asn225 are required for halogenation activity.

Table 4.2. Conservation of residues 151 and 225 in the shuffle screen.

Position		Fraction observed	
151	225	Hal	Not hal
Ile	Asn	30/30	0
Asn	Asn	0	14/33
Ile	Val	0	9/33
Asn	Val	0	10/33

Engineering Hydrox to perform halogenation. We next sought to determine whether Hydrox D144G could be engineered to perform halogenation by addition of the residues identified through the shuffle library. We first cloned, expressed, and purified Hydrox D144G V225N and assayed the enzyme by LC/MS (*Figure 4.25- 4.26*). Encouragingly, the double mutant (Hydrox D144G V225N) has diminished hydroxylation activity compared to Hydrox D144G and also displayed a ~2-fold increase in halogenation. Since our screen indicated that only variants with *both* Asn225 and Ile151 were able to give high signal for halogenation, we purified the triple mutant, Hydrox D144G V225N N151I. Encouragingly, the triple mutant yielded a five-fold increase in halogenation. The double mutant Hydrox D144G N151I failed to achieve the same result. In the BesD crystal structure, Asn225 and Ile151 are only 3.4 Å apart. Therefore, the additional Asn151I mutation could possibly have alleviated polar interactions that may disrupt the positioning of Asn225. Thus, the addition of both second sphere-residues Ile151 and Asn225 to Hydrox D144G enabled successful engineering of Hydrox to perform halogenation.

Although Hydrox D144G V225N N151I performed halogenation, there was still a large disparity in chlorination activity when compared to the wild type Hal. This difference in activity highlights the challenges associated with engineering hydroxylases to perform halogenation. We reasoned that further improvements could be achieved by incorporating additional residues found to be important in the shuffle library. Since the screening results showed that the halogenases had conserved residues along two beta strands adjacent to Asn225, we hypothesized that a boost in activity could be achieved by generating a chimera that encompassed these regions. Starting with Hydrox D144G, we designed a Chimera by grafting 13 highly conserved residues into the hydroxylase. The resulting chimera only differs from the hydroxylase by a total of 14 residues (93.5% sequence identity, *Figure 4.27*). Excitingly, we found that the Chimera performed halogenation within error of the wild type Hal. Thus, our screening results led to a successful halogenase engineering by elucidating key residues that could not be identified simply by inspection of the crystal structure.

Finally, we probed the relative contribution that the 14 different residues had toward improving halogenation. Using the crystal structure, we grouped the mutations into six groups based on proximity, solvent exposure, and polarity (*Table 4.3*). Each set was then either added (to Hydrox D144G V225N N151I) or removed (from the chimera). The 12 resulting enzyme variants were

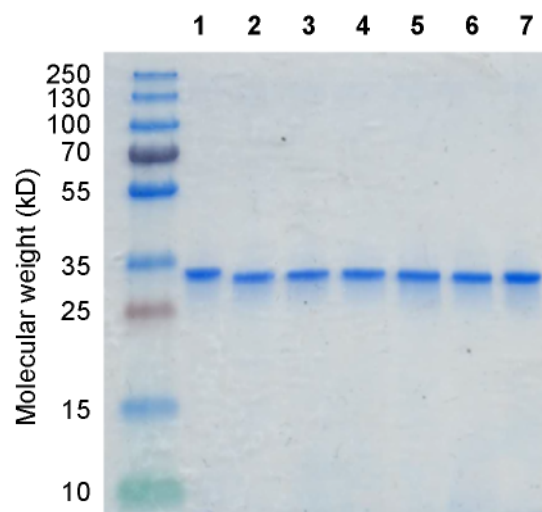


Figure 4.25. SDS-PAGE gel of purified Hal and Hydrox variants. Lane 1, Hal; 2, Hydrox; 3, Hydrox D144G; 4, Hydrox D144G V225N; 5, Hydrox D144G N151I; 6, Hydrox D144G V225N N151I; 7, Chimera.

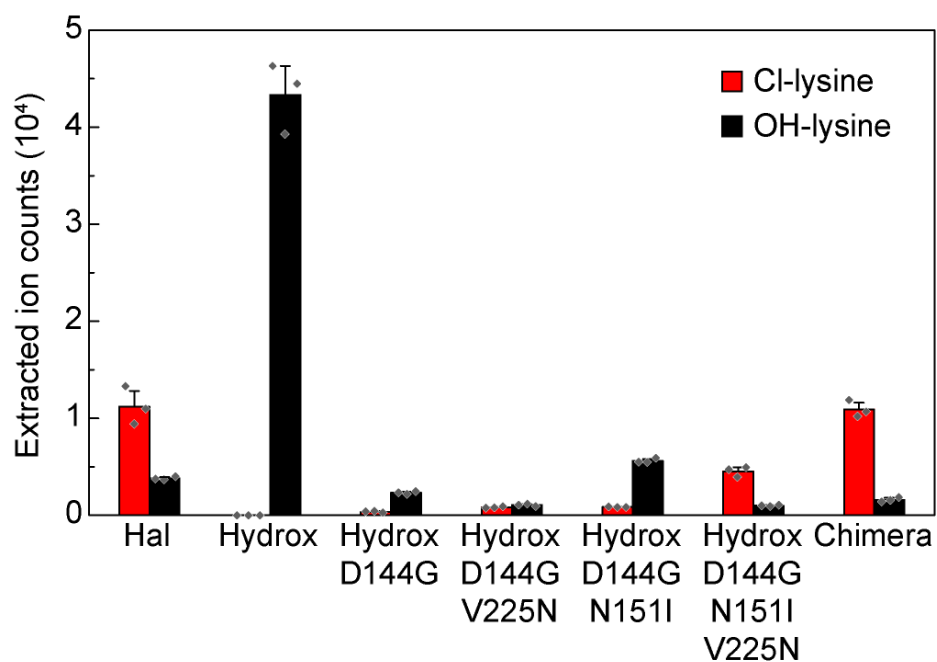


Figure 4.26. LC/MS analysis of engineered Hydrox variants. The mutations were chosen based on residues identified from the shuffle screen and from structures. “Chimera” is an engineered hydroxylase that has a total of 14 residues from the halogenase engineered into it, including D144G, N151I, and V225N.

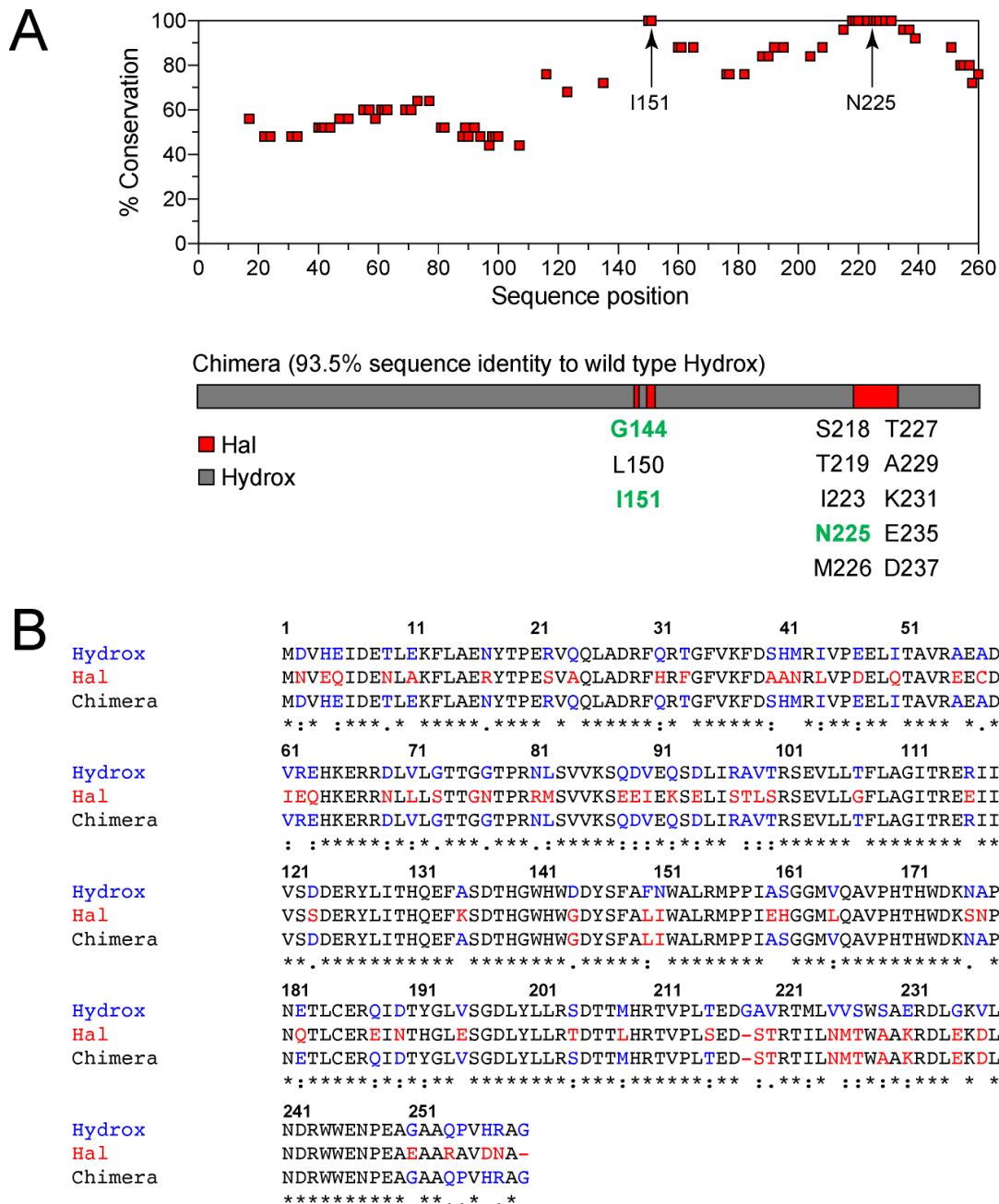


Figure 4.27. Designing a chimera based on the shuffle library. To design an enzyme which captures the key sequence motifs required for halogenation, a chimeric protein was engineered based on the shuffle results. (A) Conservation of halogenase residues that were conserved in library members capable of performing halogenation. Residues that were conserved in 100% of the halogenating hits were grafted into the sequence of the hydroxylase. All the residues that differ between the Chimera and Hydrox are listed below the sequence map. Key primary and secondary sphere residues are highlighted in green. (B) Sequence alignment of Hydrox (blue), Hal (red), and the engineered Chimera. Alignment was performed in MUSCLE [16].

purified and assayed by LC/MS (*Figures 4.28-4.29*). When we added the halogenase groups one at a time to the hydroxylase, we found that none were sufficient to achieve the halogenation levels found in the chimera on their own. Likewise, the systematic removal of the halogenase residues from the Chimera did not result in a full reversal to the lower activity found in the hydroxylase triple mutant (Hydrox D144G V225N N151I). Taken together, the data support the idea that the key residues from the screen likely have an additive effect on halogenation activity. Our previous work on BesD in which we assayed the BesD G144D confirmed that the residue in the facial triad is essential for halogenation [14]. Beyond this, our results suggest that a key factor for halogenation is Asn225, while additional residues along the two core beta sheets aid in the tuning of that residue to maximize activity and selectivity in radical halogenases within the BesD family.

Structural insights into engineered selectivity. While the anaerobic crystal structures of BesD and Hydrox are only snapshots of the enzyme's reactant complex, they provide insight into the possible influence of second sphere residues on key catalytic intermediates. In $\text{Fe}^{\text{II}}/\alpha\text{KG}$ -dependent enzymes, oxygen binds to the Fe^{II} reactant complex, leading to decarboxylation of αKG to succinate. The decarboxylation results in a Fe^{IV} -oxo intermediate, which abstracts a hydrogen atom from the substrate. In halogenases, the resulting substrate radical selectively rebounds with the $\text{Fe}^{\text{III}}\text{-Cl}$, rather than the $\text{Fe}^{\text{III}}\text{-OH}$, yielding a chlorinated product [1,2]. Previous studies on the halogenases SyrB2 and WelO5 suggest that substrate positioning relative to the $\text{Fe}^{\text{III}}\text{-Cl}$ and $\text{Fe}^{\text{III}}\text{-OH}$ is critical for determining whether halogen or hydroxyl rebound is favored. It is proposed that halogenases orient the reactive Fe^{IV} -oxo intermediate such that hydrogen atom abstraction is sluggish but rebound of the resulting $\text{Fe}^{\text{III}}\text{-OH}$ is disfavored. In SyrB2, use of non-native substrates perturbed the positioning of the carbon targeted for modification and resulted in hydroxylation by the enzyme [37]. In WelO5, Ser189 is proposed to orient the oxo away from the substrate, favoring halogen rebound. Mutation of Ser189 to Ala leads to decreased halogenation fidelity [38]. Work by Solomon and co-workers suggests that the determining factor for selective halogenation is the favorable overlap of the substrate radical with the $\text{Fe}^{\text{III}}\text{-Cl}$ orbitals compared to the $\text{Fe}^{\text{III}}\text{-OH}$ [39,40]. Therefore, second-sphere residues that influence the relative overlap of these groups are poised to influence selectivity.

We have identified important second sphere residues by combining crystallography with a functional screening assay. Among the key determining factors of reaction outcome are Asn225 in Hal and Val225 in Hydrox, which could influence relative proximity of C₄ (lysine) to $\text{Fe}^{\text{III}}\text{-Cl}$ (4.0 Å, in Hal) compared to $\text{Fe}^{\text{III}}\text{-OH}$ (*Figure 4.29*). In the crystal structure of BesD, Asn225 hydrogen bonds with αKG in the active site and orients the coordinating oxygen away farther away from C₄ of the substrate (3.8 Å) than in the case of the hydroxylase (3.3 Å). This 0.5 Å difference may serve as a preview of the orientation of subsequent intermediates during the catalytic cycle; once the substrate radical is formed in halogenases, Asn225 could hydrogen bond the $\text{Fe}^{\text{III}}\text{-OH}$ intermediate, decreasing orbital overlap with the substrate radical, and resulting in $\text{Fe}^{\text{III}}\text{-Cl}$ rebound instead. In SyrB2, hydrogen bonding constrains the hydroxyl group such that the $d\pi^*_{\text{Fe-OH}}$ frontier molecular orbital cannot achieve good overlap with the substrate carbon radical [40], which disfavors hydroxylation. In Hydrox, the replacement of Asn225 with Val precludes hydrogen bonding of the $\text{Fe}^{\text{III}}\text{-OH}$ group. Furthermore, DFT calculations on SyrB2 suggest that a Fe-Cl distance that is 0.5 Å closer than Fe-OH (considering ionic radii) results in a lower Gibbs free energy barrier for chlorine rebound, thereby favoring halogenation [40].

Consistent with this proposed role of second-sphere residues, our results suggest that Asn225 serves as a critical switch to enable halogenation. However, the introduction of Asn225 into the secondary sphere of Hydrox D144G is insufficient for halogenation without additional residues that serve to further tune activity. Thus, Asn225 may serve as a critical switch to enable halogenation, while the additional residues identified in the screen that lie along adjacent strands are required to tune the active site for selective halogenation. Indeed, our systematic analysis of the 14 residues in the engineered Chimera revealed that the residues with the greatest impact on halogenation are those immediately adjacent to Asn225 (Ile151, Met226, Thr227), and that the effects of the key residues are additive. The screen enabled unbiased identification of previously unexplored second-sphere residues that work in concert to enable effective chlorination.

Experimental investigation of reactive intermediates in Hal and Hydrox. Future work on these enzyme systems will seek to experimentally probe the role of second-sphere residues on catalytic intermediates. The $\text{Fe}^{\text{III}}\text{-OH}$ intermediate immediately preceding the rebound step is too short-lived to monitor experimentally [41]. However, spectroscopic and crystallographic techniques for characterizing the $\text{Fe}^{\text{IV}}\text{-oxo}$ species have proven useful in other $\text{Fe}^{\text{II}}/\alpha\text{KG}$ -dependent systems [2,42–45]. One such technique is stopped-flow absorption (SF-abs), which can be used to monitor the rise and decay of the $\text{Fe}^{\text{IV}}\text{-oxo}$ species at 320 nm. SF-abs experiments with SyrB2 paired with either native or non-native substrates showed that SyrB2 positions its native substrate for halogenation by orienting the substrate away from the $\text{Fe}^{\text{IV}}\text{-oxo}$ species, thereby sacrificing the rate of $\text{Fe}^{\text{IV}}\text{-oxo}$ hydrogen atom abstraction in order to increase reaction fidelity [37,43]. With substrate-bound crystal structures and successful engineering results, the Hal/Hydrox system provides a unique opportunity to understand how second-sphere residues may alter the reaction outcome. We hope to determine whether Hal exhibits slower $\text{Fe}^{\text{IV}}\text{-oxo}$ decay compared to Hydrox, supporting the hypothesis that second-sphere residues such as Asn225 may orient the oxo away to ensure selectivity at the expense of the rate of H atom abstraction (*Figure 4.31A*). Towards this goal, we have purified Hal, Hydrox, the engineered Chimera, and several Hydrox and Hal point mutants, and successfully demonstrated that these enzymes can be concentrated to the mM concentration range required for these experiments.

In addition to indirect measurements of ferryl orientation using SF-abs, geometric information can also be obtained by substitution of the ferryl with vanadyl ($\text{V}^{\text{IV}}\text{O}$)²⁺. Vanadyl serves as a structural mimic of the unstable $\text{Fe}^{\text{IV}}\text{-oxo}$ intermediate that is generated in αKG -dependent enzymes upon decarboxylation of αKG [44]. In particular, crystal structures of αKG -dependent enzymes bound to the vanadyl intermediate (rather than to the $\text{Fe}^{\text{IV}}\text{-oxo}$, which is too short-lived to crystallize) can reveal information about the positioning of the metal complex relative to the substrate (*Figure 4.31B*) [45]. Towards this goal, we co-crystallized BesD with vanadyl and succinate (*Table 4.4, Figure 4.32*). Although the data reveal density for vanadium between His137 and His 204, density for the chloride and oxo ligands on vanadium is too poor for confident modelling using this data set. Improvements could be made by growing the crystals anaerobically to prevent oxidation of vanadyl to vanadate, or by soaking the crystals with higher concentrations of NaCl prior to freezing. The ability to obtain vanadyl, succinate, chloride, and lysine-bound enzyme will additionally facilitate EPR experiments to measure the distance and angle between a deuterium on the substrate and the vanadyl probe [44]. Taken together, these experiments on Hal and Hydrox can expand our mechanistic understanding of halogenation.

Table 4.3. Grouping the residues that were highly conserved in the shuffle screen based on proximity, solvent exposure, and polarity.

Group	Hydrox	Chimera	Notes
---	D144	G144	Critical
1	F150	L150	Maybe pushing on back helix around residue 176
---	N151	I151	Critical
2	G218	S218	Solvent exposed
2	A219	T219	Solvent exposed
2	V220	---	Solvent exposed
3	M223	I222	Near α KG binding site
---	V225	N224	Critical
4	V226	M225	Close to critical N/V225 residue
4	S227	T226	Very similar residues. Close to N/V225
5	S229	A228	Seems to be interacting with other residues
6	E231	K230	Solvent exposed
6	G235	E234	Solvent exposed
6	V237	D236	Solvent exposed

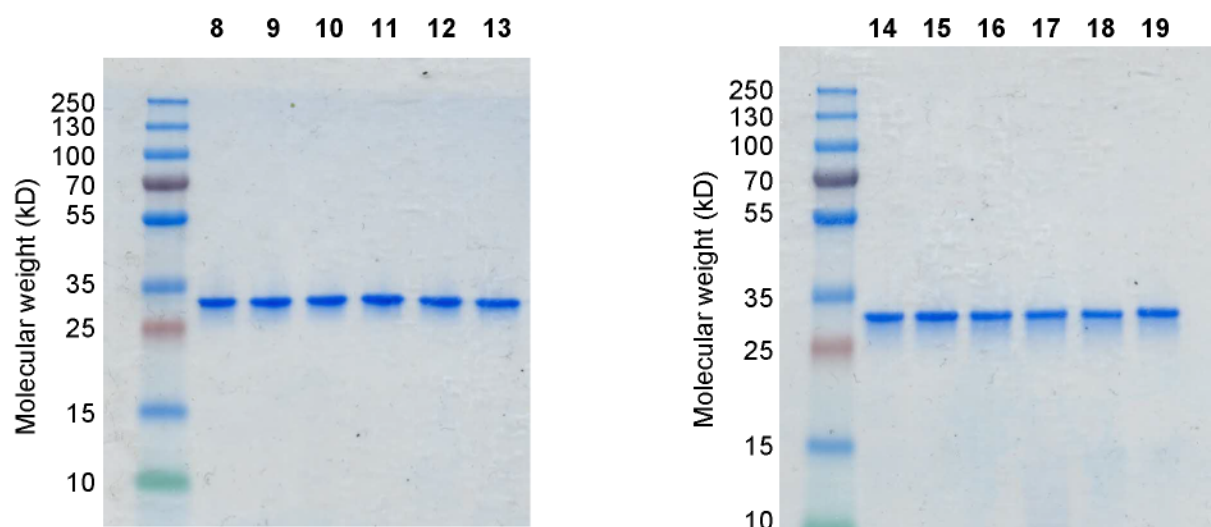


Figure 4.28. SDS-PAGE gel for of purified Hydrox and Chimera variants. Lane **8**, Chimera L150F; **9**, Chimera ST218-219GAV; **10**, Chimera I222M; **11**, Chimera M225V T226S; **12**, Chimera A228S; **13**, Chimera K230E E234G D236V; **14**, HydroxTM F150L; **15**, HydroxTM GAV218/219 to ST; **16**, HydroxTM M223I; **17**, HydroxTM V226M S227T; **18**, HydroxTM S229A; **19**, HydroxTM E231K G235E V237D.

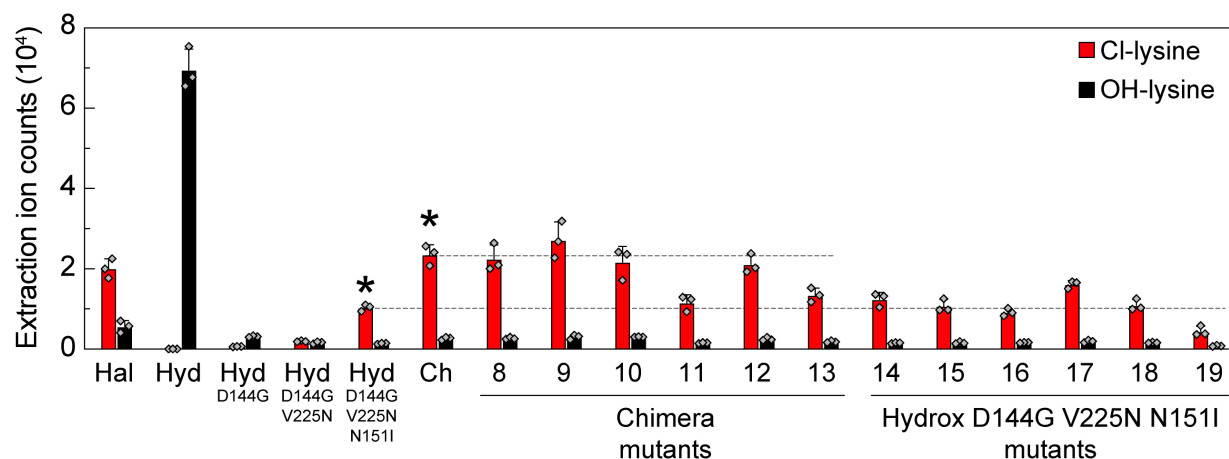
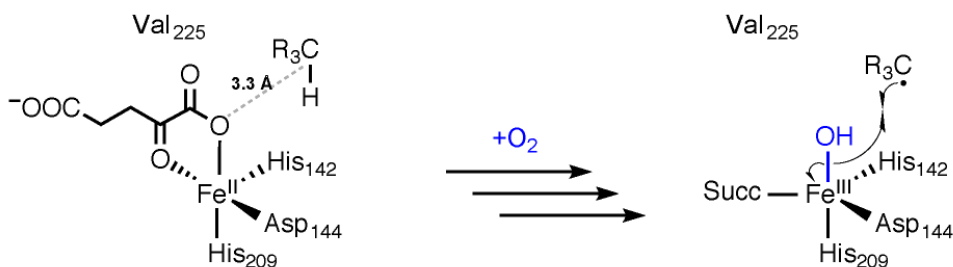


Figure 4.29. LC/MS analysis of engineered Hydrox variants. An asterisk and the corresponding dotted line provide points of comparison for the mutants. The goal of the analysis was to identify Chimera mutants that diminish activity relative to the Chimera (Ch) or Hydrox D144G N151I V225N (HydroxTM) mutants that increase activity relative to the triple mutant (Hydrox D144G N151I V225N). The variants 8-19 correspond to the purified proteins in Figure 4.24: **8**, Chimera L150F; **9**, Chimera ST218-219GAV; **10**, Chimera I222M; **11**, Chimera M225V T226S; **12**, Chimera A228S; **13**, Chimera K230E E234G D236V; **14**, HydroxTM F150L; **15**, HydroxTM GAV218/219 to ST; **16**, HydroxTM M223I; **17**, HydroxTM V226M S227T; **18**, HydroxTM S229A; **19**, HydroxTM E231K G235E V237D.

Hydroxylase



Halogenase

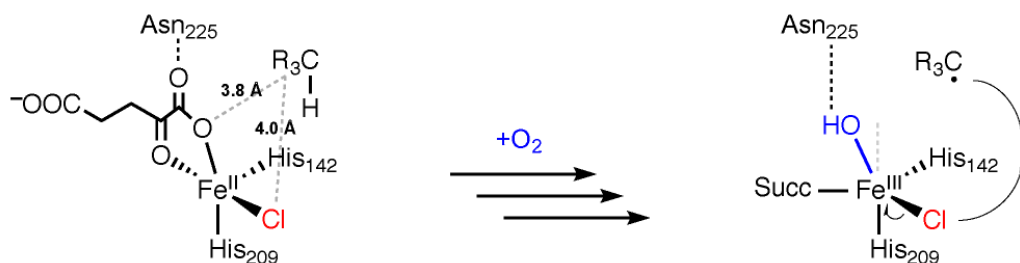


Figure 4.30. The proposed role of residue 225 on halogenation outcome. Residue 225 is Val in the hydroxylase (top), but Asn in the halogenase (bottom). The distance from αKG to the substrate (C₄ of lysine) is 3.3 Å in the hydroxylase. The distance from Cl to C₄ of lysine is 4.0 Å. In the hydroxylase, Asn225 hydrogen bonds with αKG to distort the geometry of the reactant complex, increasing the distance from αKG to the substrate to 3.8 Å. After oxygen activation, decarboxylation of αKG to succinate, and then hydrogen atom abstraction, Asn225 may hydrogen bond with the hydroxyl ligand in the halogenase, thereby favoring halogen rebound instead. Assuming the -OH occupies the same position as the coordinating oxygen in the C₁ carboxylate of αKG, the -Cl ligand would be closer to the substrate radical in the halogenase (considering the larger ionic radius of Cl).

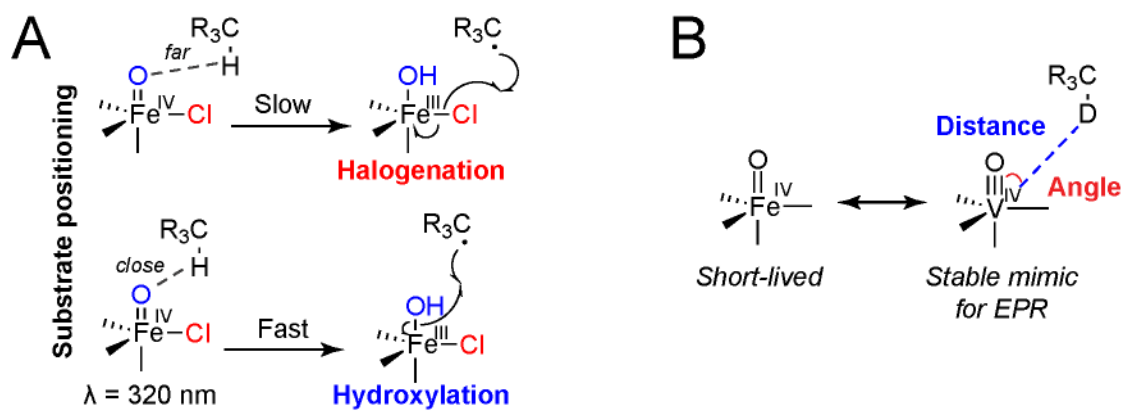


Figure 4.31. Spectroscopic techniques for probing the mechanism of halogenation within the enzymes pair.

Table 4.4. Data collection and refinement parameters for ScBesD with Vanadium and succinate bound in the active site.

ScBesD vanadium and succinate	
Data collection	
Space group	P1
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	61.092, 68.892, 73.448
α , β , γ (°)	111.118, 94.201, 111.151
Resolution (Å)	58.4–2.0 (2.072–2.0)
<i>R</i> _{sym} or <i>R</i> _{merge}	0.1929 (2.968)
<i>I</i> / σI	3.87 (0.32)
Completeness (%)	94.25 (69.46)
Redundancy	3.5 (3.5)
CC _{1/2}	0.994 (0.257)
Refinement	
Resolution (Å)	58.4–2.0
No. reflections	233382 (20859)
<i>R</i> _{work} / <i>R</i> _{free}	0.287/0.320
No. atoms	7649
Protein	7542
Ligand/ion	36
Water	71
<i>B</i> -factors	43.72
Protein	43.79
Ligand/ion	41.54
Water	36.95
R.m.s. deviations	
Bond lengths (Å)	0.004
Bond angles (°)	1.01

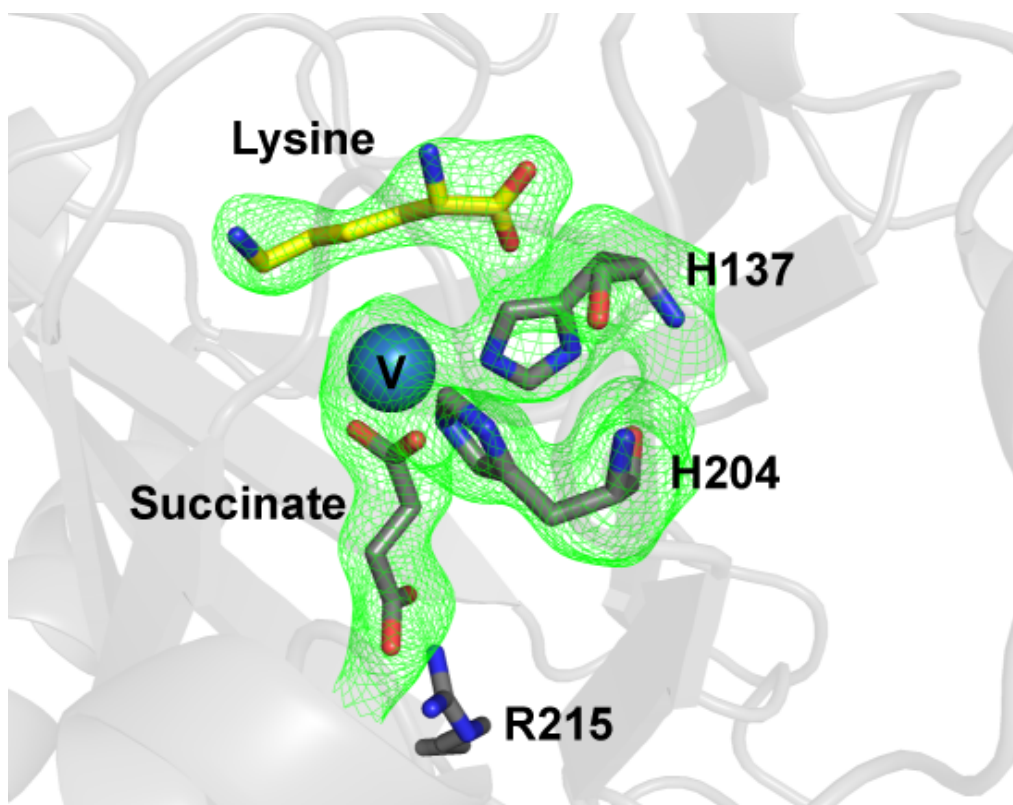


Figure 4.32. Crystal structure of BesD from *S. cattleya* with lysine, vanadium, and succinate bound in the active site. The electron density for the other ligands on vanadium (including chlorine) was not well-defined enough to be modeled in. The $F_o - F_c$ omit map (contoured at 1.5σ) for lysine (yellow), α KG (grey), vanadium (blue), His 137, and His204 is shown in green.

4.4. Conclusion

One strategy for expanding the substrate scope of enzymatic halogenases is to engineer the closely related family of hydroxylases to instead perform halogenation. Thus far, this goal has been difficult to achieve due to second-sphere contributions that cannot be easily identified by inspection. In this chapter, we have combined bioinformatics, structural biology, library generation, and screening to achieve this goal. The identification of a halogenase and hydroxylase pair with high sequence identity aided the identification of important second-sphere residues by narrowing down key sequence differences and by enabling DNA-shuffling of the enzyme pair. By solving the crystal structure of Hydrox for comparison to the halogenase BesD, we have demonstrated that second-sphere residues influence the geometry of ligands on the catalytic Fe and may be critical to halogenation selectivity in the BesD family of enzymes. The crystal structures, along with the results of our library, enabled successful engineering of Hydrox to perform efficient halogenation. Our results suggest that tuning the active site by targeting residues along key beta strands within the cupin barrel may be a promising strategy for engineering hydroxylases to perform halogenation. In conjunction with the crystal structures of Hal and Hydrox, these engineering studies contribute to our mechanistic understanding of halogenation.

4.5. References

1. Puri, M., Biswas, A. N., Fan, R., Guo, Y. & Que, L. Modeling non-heme iron halogenases: high-spin oxoiron(IV)–halide complexes that halogenate C–H bonds. *J. Am. Chem. Soc.* **138**, 2484–2487 (2016) doi:10.1021/jacs.5b11511.
2. Galonić, D. P., Barr, E. W., Walsh, C. T., Bollinger, J. M. & Krebs, C. Two interconverting Fe(IV) intermediates in aliphatic chlorination by the halogenase CytC3. *Nat. Chem. Biol.* **3**, 113–116 (2007) doi:10.1038/nchembio856.
3. Agarwal, V., Miles, Z. D., Winter, J. M., Eustáquio, A. S., El Gamal, A. A. & Moore, B. S. Enzymatic halogenation and dehalogenation reactions: pervasive and mechanistically diverse. *Chem. Rev.* **117**, 5619–5674 (2017) doi:10.1021/acs.chemrev.6b00571.
4. Islam, M. S., Leissing, T. M., Chowdhury, R., Hopkinson, R. J. & Schofield, C. J. 2-Oxoglutarate-Dependent Oxygenases. *Annu. Rev. Biochem.* **87**, (2018) doi:10.1146/annurev-biochem-061516-044724.
5. Gorres, K. L., Pua, K. H. & Raines, R. T. Stringency of the 2-His–1-Asp active-site motif in prolyl 4-hydroxylase. *PLoS One* **4**, e7635 (2009) doi:10.1371/journal.pone.0007635.
6. Grzyska, P. K., Müller, T. A., Campbell, M. G. & Hausinger, R. P. Metal ligand substitution and evidence for quinone formation in taurine/α-ketoglutarate dioxygenase. *J. Inorg. Biochem.* **101**, 797–808 (2007) doi:10.1016/j.jinorgbio.2007.01.011.
7. Mitchell, A. J., Dunham, N. P., Bergman, J. A., Wang, B., Zhu, Q., Chang, W., Liu, X. & Boal, A. K. Structure-guided reprogramming of a hydroxylase to halogenate its small molecule substrate. *Biochemistry* **56**, 441–444 (2017) doi:10.1021/acs.biochem.6b01173.
8. Vaillancourt, F. H., Yin, J. & Walsh, C. T. SyrB2 in syringomycin E biosynthesis is a nonheme Fe(II) α-ketoglutarate- and O₂-dependent halogenase. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 10111–10116 (2005) doi:10.1073/pnas.0504412102.
9. Ueki, M., Galonić, D. P., Vaillancourt, F. H., Garneau-Tsodikova, S., Yeh, E., Vosburg, D. A., Schroeder, F. C., Osada, H. & Walsh, C. T. Enzymatic generation of the antimetabolite γ,γ-dichloroaminobutyrate by NRPS and mononuclear iron halogenase action in a streptomycete. *Chem. Biol.* **13**, 1183–1191 (2006) doi:10.1016/j.chembiol.2006.09.012.
10. Pratter, S. M., Ivkovic, J., Birner-Gruenberger, R., Breinbauer, R., Zangger, K. & Straganz, G. D. More than just a Halogenase: Modification of Fatty Acyl Moieties by a Trifunctional Metal Enzyme. *ChemBioChem* **15**, 567–574 (2014) doi:10.1002/cbic.201300345.
11. Hillwig, M. L. & Liu, X. A new family of iron-dependent halogenases acts on freestanding substrates. *Nat. Chem. Biol.* **10**, 6–10 (2014) doi:10.1038/nchembio.1625.
12. Hillwig, M. L., Zhu, Q., Ittiamornkul, K. & Liu, X. Discovery of a promiscuous non-heme iron halogenase in ambiguine alkaloid biogenesis: implication for an evolvable enzyme family for late-stage halogenation of aliphatic carbons in small molecules. *Angew. Chemie Int. Ed.* **55**, 5780–5784 (2016) doi:10.1002/anie.201601447.
13. Zhang, Y., Zhao, C., Yan, S., Li, Q., Zhu, H., Zhong, Z., Ye, Y. & Deng, Z. An iron(II)- and α-ketoglutarate-dependent halogenase acts on nucleotide substrates. *Angew. Chem. Int. Ed. Engl.* **anie.201914994** (2020) doi:10.1002/anie.201914994.
14. Marchand, J. A., Neugebauer, M. E., Ing, M. C., Lin, C.-I., Pelton, J. G. & Chang, M. C. Y. Discovery of a pathway for terminal-alkyne amino acid biosynthesis. *Nature* **567**, 420–424 (2019) doi:10.1038/s41586-019-1020-y.
15. Huang, Y., Niu, B., Gao, Y., Fu, L. & Li, W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **26**, 680–2 (2010) doi:10.1093/bioinformatics/btq003.

16. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004) doi:10.1186/1471-2105-5-113.
17. Larsson, A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics* **30**, 3276–8 (2014) doi:10.1093/bioinformatics/btu531.
18. Jones, D. T., Taylor, W. R. & Thornton, J. M. The rapid generation of mutation data matrices from protein sequences. *Comput. Appl. Biosci.* **8**, 275–82 (1992) doi:10.1093/bioinformatics/8.3.275.
19. Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A. & Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–5 (2009) doi:10.1038/nmeth.1318.
20. Bryksin, A. & Matsumura, I. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *Biotechniques* **48**, 463–465 (2010) doi:10.2144/000113418.
21. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D. & Bairoch, A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31**, 3784–8 (2003) doi:10.1093/nar/gkg563.
22. Kabsch, W. XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125–32 (2010) doi:10.1107/S0907444909047337.
23. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D. Biol. Crystallogr.* **69**, 1204–14 (2013) doi:10.1107/S0907444913000061.
24. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D. Biol. Crystallogr.* **67**, 235–42 (2011) doi:10.1107/S0907444910045749.
25. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of *Coot*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010) doi:10.1107/S0907444910007493.
26. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–21 (2010) doi:10.1107/S0907444909052925.
27. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. UCSF Chimera: A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004) doi:10.1002/jcc.20084.
28. Neugebauer, M. E., Sumida, K. H., Pelton, J. G., McMurry, J. L., Marchand, J. A. & Chang, M. C. Y. A family of radical halogenases for the engineering of amino-acid-based products. *Nat. Chem. Biol.* **15**, 1009–1016 (2019) doi:10.1038/s41589-019-0355-x.
29. Luo, L., Pappalardi, M. B., Tummino, P. J., Copeland, R. A., Fraser, M. E., Grzyska, P. K. & Hausinger, R. P. An assay for Fe(II)/2-oxoglutarate-dependent dioxygenases by enzyme-coupled detection of succinate formation. *Anal. Biochem.* **353**, 69–74 (2006) doi:10.1016/j.ab.2006.03.033.
30. Guo, A. C., Jewison, T., Wilson, M., Liu, Y., Knox, C., Djoumbou, Y., Lo, P., Mandal, R., Krishnamurthy, R. & Wishart, D. S. ECMDDB: the E. coli Metabolome Database. *Nucleic Acids Res.* **41**, D625–30 (2013) doi:10.1093/nar/gks992.
31. Iyer, S. R., Chaplin, V. D., Knapp, M. J. & Solomon, E. I. O₂ Activation by Nonheme Fe^{II} α -Ketoglutarate-Dependent Enzyme Variants: Elucidating the Role of the Facial Triad Carboxylate in FIH. *J. Am. Chem. Soc.* **140**, 11777–11783 (2018) doi:10.1021/jacs.8b07277.

32. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018) doi:10.1093/molbev/msy096.
33. Zhao, H. Optimization of DNA shuffling for high fidelity recombination. *Nucleic Acids Res.* **25**, 1307–1308 (1997) doi:10.1093/nar/25.6.1307.
34. Zhu, X., Shieh, P., Su, M., Bertozzi, C. R. & Zhang, W. A fluorogenic screening platform enables directed evolution of an alkyne biosynthetic tool. *Chem. Commun.* **52**, 11239–11242 (2016) doi:10.1039/C6CC05990B.
35. Varazo, K., Droumaguet, C. Le, Fullard, K. & Wang, Q. Metal ion detection using a fluorogenic 'click' reaction. *Tetrahedron Lett.* **50**, 7032–7034 (2009) doi:10.1016/J.TETLET.2009.09.166.
36. Shieh, P., Dien, V. T., Beahm, B. J., Castellano, J. M., Wyss-Coray, T. & Bertozzi, C. R. CalFluors: A Universal Motif for Fluorogenic Azide Probes across the Visible Spectrum. *J. Am. Chem. Soc.* **137**, 7145–51 (2015) doi:10.1021/jacs.5b02383.
37. Matthews, M. L., Neumann, C. S., Miles, L. a, Grove, T. L., Booker, S. J., Krebs, C., Walsh, C. T. & Bollinger, J. M. Substrate positioning controls the partition between halogenation and hydroxylation in the aliphatic halogenase, SyrB2. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 17723–17728 (2009) doi:10.1073/pnas.0909649106.
38. Mitchell, A. J., Zhu, Q., Maggiolo, A. O., Ananth, N. R., Hillwig, M. L., Liu, X. & Boal, A. K. Structural basis for halogenation by iron- and 2-oxo-glutarate-dependent enzyme WelO5. *Nat. Chem. Biol.* **12**, 636–640 (2016) doi:10.1038/nchembio.2112.
39. Wong, S. D., Srnc, M., Matthews, M. L., Liu, L. V, Kwak, Y., Park, K., Bell, C. B., Alp, E. E., Zhao, J., Yoda, Y., Kitao, S., Seto, M., Krebs, C., Bollinger, J. M. & Solomon, E. I. Elucidation of the Fe(IV)=O intermediate in the catalytic cycle of the halogenase SyrB2. *Nature* **499**, 320–323 (2013) doi:10.1038/nature12304.
40. Srnc, M. & Solomon, E. I. Frontier molecular orbital contributions to chlorination versus hydroxylation selectivity in the non-heme iron halogenase SyrB2. *J. Am. Chem. Soc.* **139**, 2396–2407 (2017) doi:10.1021/jacs.6b11995.
41. Price, J. C., Barr, E. W., Hoffart, L. M., Krebs, C. & Bollinger, J. M. Kinetic dissection of the catalytic mechanism of taurine:α-ketoglutarate dioxygenase (TauD) from Escherichia coli. *Biochemistry* **44**, 8138–8147 (2005) doi:10.1021/bi050227c.
42. Matthews, M. L., Krest, C. M., Barr, E. W., Vaillancourt, F. H., Walsh, C. T., Green, M. T., Krebs, C., Bollinger, J. M. & Jr. Substrate-triggered formation and remarkable stability of the C-H bond-cleaving chloroferryl intermediate in the aliphatic halogenase, SyrB2. *Biochemistry* **48**, 4331–43 (2009) doi:10.1021/bi900109z.
43. Martinie, R. J., Livada, J., Chang, W., Green, M. T., Krebs, C., Bollinger, J. M. & Silakov, A. Experimental correlation of substrate position with reaction outcome in the aliphatic halogenase, SyrB2. *J. Am. Chem. Soc.* **137**, 6912–6919 (2015) doi:10.1021/jacs.5b03370.
44. Martinie, R. J., Pollock, C. J., Matthews, M. L., Bollinger, J. M., Krebs, C. & Silakov, A. Vanadyl as a Stable Structural Mimic of Reactive Ferryl Intermediates in Mononuclear Nonheme-Iron Enzymes. *Inorg. Chem.* **56**, 13382–13389 (2017) doi:10.1021/acs.inorgchem.7b02113.
45. Mitchell, A. J., Dunham, N. P., Martinie, R. J., Bergman, J. A., Pollock, C. J., Hu, K., Allen, B. D., Chang, W.-C., Silakov, A., Bollinger, J. M., Krebs, C. & Boal, A. K. Visualizing the Reaction Cycle in an Iron(II)- and 2-(Oxo)-glutarate-Dependent Hydroxylase. *J. Am. Chem. Soc.* **139**, 13830–13836 (2017) doi:10.1021/jacs.7b07374.

Chapter 5: *Structural basis of regioselectivity and substrate selectivity in radical amino acid halogenases*

Portions of this work were performed in collaboration with the following persons:

Crystallography experiments were performed in collaboration with Kiera Sumida and Cameron Swenson. Fluorogenic screening of the library was developed in collaboration with Dr. Jorge Marchand and Doug Millar. Fluorogenic screening and halogenase biochemistry were performed in collaboration with Nicholas Sambold and Elijah Kissman. Stopped-flow absorption experiments were performed with collaborators in the Bollinger, Krebs, and Boal labs at Pennsylvania State University.

5.1. Introduction

The diversity of reactions offered by synthetic chemistry enables access to a broad range of valuable products, including pharmaceuticals, polymers, and commodity chemicals. A major advantage of traditional synthetic strategies is versatility; reactions can often be performed on a range of different compounds possessing the required functional groups. Enzymes, on the other hand, offer a complementary strategy. Through evolution in a cellular milieu, enzymes have evolved exquisite selectivity for their target substrates. As a result, enzymatic approaches are particularly promising for transformations where selectivity is difficult to achieve, such as C–H bond activation. Enzymes capable of this chemistry, such as cytochrome P450s, have been evolved for diverse reactions, including C–C-bond formation, amination, silylation, and borylation [1–4]. However, only a handful of enzymes that can chlorinate C_{sp3}–H bonds have been discovered to date, which limits their application to the field of biocatalysis [5–9].

In recent work, we discovered a new family of Fe^{II}/α-ketoglutarate-dependent halogenases (BesD family) that selectively modify amino acid substrates without the requirement for carrier protein-tethered substrates [10,11]. Collectively, this family of enzymes (HalA–E) can produce 4-Cl-lysine (BesD and HalA); 5-Cl-lysine and 5,5-dichlorolysine (HalB); 4,4-dichlorolysine (HalC); 4-Cl-ornithine (HalD); and 4-Cl-leucine, 4-Cl-isoleucine, and 4-Cl-norvaline (HalE). The product scope of these enzymes provides an opportunity to investigate the mechanistic basis of regioselectivity and substrate selectivity in radical halogenases. This information could be used to engineer BesD to accept a broader range of substrates. In particular, since these enzymes act directly on free-standing substrates, they are promising candidates for enzyme engineering.

In this chapter, we perform structural studies of the BesD family members to investigate the basis of selectivity within the family. To understand how BesD and HalB achieve different regioselective outcomes, we solved the structure of HalB and performed comparative structural analyses and mutagenesis. We also solved the crystal structure of the ornithine halogenase (HalD), which reveals the basis of selectivity for a 5-carbon substrate over a 6-carbon substrate. Taken together, the data show that the enzymes are structurally similar except at the C-terminal lid, which closes over the substrate in the active site. To further support the role of the covering lid in determining the reaction outcome, we solved apo crystal structures of HalB, BesD, and a related hydroxylase. In all cases, the apo enzyme adopts an alternate conformation with a disordered substrate covering lid. The data suggest that the lid may be a promising target for future efforts to engineer radical halogenases for chlorination of unactivated C_{sp3}–H bonds across a broad range of substrates.

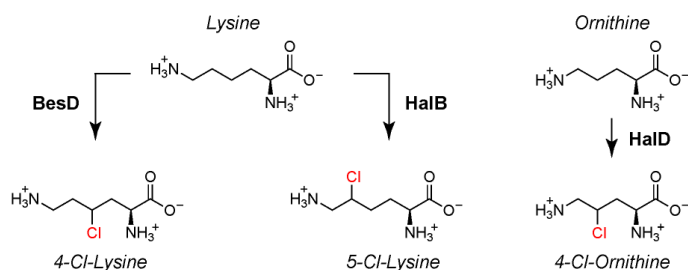


Figure 5.1. Radical amino acid halogenases in this study (BesD, HalB, HalD). BesD and HalB both act on lysine, but with differing regioselectivities. HalD chlorinates ornithine.

5.2. Materials and methods

Commercial materials. Luria-Bertani (LB) Broth Miller, LB Agar Miller, Terrific Broth (TB), and glycerol were purchased from EMD Biosciences (Darmstadt, Germany). Carbenicillin (Cb), kanamycin (Km), chloramphenicol (Cm), isopropyl- β -D-thiogalactopyranoside (IPTG), sodium chloride, dithiothreitol (DTT), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), magnesium chloride hexahydrate, acetonitrile, ethylene diamine tetraacetic acid disodium dihydrate (EDTA), hydrochloric acid, 3kDa MWCO dialysis tubing, ammonium bicarbonate, sodium acetate, chloroform, and sodium hydroxide were purchased from Fisher Scientific (Pittsburgh, PA). Phosphoenolpyruvate (PEP), adenosine triphosphate sodium salt (ATP), nicotinamide adenine dinucleotide reduced form dipotassium salt (NADH), pyruvate kinase, lactate dehydrogenase, lysozyme, poly(ethyleneimine) solution (PEI), ammonium iron (II) sulfate hexahydrate, α -ketoglutaric acid sodium salt, β -mercaptoethanol (β ME), N,N,N',N'-tetramethyl-ethane-1,2-diamine (TEMED), betaine, dimethylsulfoxide (DMSO), sodium L-ascorbate, acetonitrile (LC/MS-grade), ammonium formate (LC/MS-grade), Deuterium oxide, L-lysine for crystallography, 2-(N-Morpholino)ethanesulfonic acid hydrate, 4-morpholineethanesulfonic acid (MES), PEG 3350, PEG 1500, PEG 550 MME, Pall Nanosep centrifugal Omega membrane (3,000 MWCO), AeraSeal films, and copper sulfate (CuSO_4) were purchased from Sigma-Aldrich (St. Louis, MO). PageRuler Plus Prestained Protein Ladder was purchased from ThermoFisherScientific (Waltham, MA). Succinyl-CoA synthetase was purchased from Megazyme International (Bray, Ireland). L-lysine $\cdot$ 2HCl ($^{13}\text{C}_6$, 99%; $^{15}\text{N}_2$, 99%), and 4,4,5,5,-D₄-lysine were purchased from Cambridge Isotope Laboratories (Andover, MA). Formic acid was purchased from Acros Organics (Morris Plains, NJ). Restriction enzymes, T4 DNA ligase, Phusion DNA polymerase, T5 exonuclease, and Taq DNA ligase, were purchased from New England Biolabs (Ipswich, MA). Deoxynucleotides (dNTPs), were purchased from Invitrogen (Carlsbad, CA). Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA), resuspended at a stock concentration of 100 μM in water and stored at either 4 $^\circ\text{C}$ for immediate usage or -20 $^\circ\text{C}$ for long-term storage. DNA purification kits and Ni-NTA agarose were purchased from Qiagen (Valencia, CA). Zirconia/silica beads were purchased from BioSpec Products (Bartlesville, OK). Complete EDTA-free protease inhibitor was purchased from Roche Applied Science (Penzberg, Germany). PD-10 desalting columns and the Superdex 75 16/60 pg column were purchased from GE Healthcare (Pittsburg, PA). Amicon Ultra 10,000 MWCO centrifugal concentrators and Milli-Q Gradient water purification system were purchased from Millipore (Billerica, MA). Acrylamide/bis-acrylamide (30%, 37.5:1), electrophoresis grade sodium dodecyl sulfate (SDS), and ammonium persulfate were purchased from Bio-Rad Laboratories (Hercules, CA). Ultrayield baffled flasks were purchased from Thompson Instrument Company (Oceanside, CA). Tris-hydroxypropyltriazolylmethylamine (THPTA) was purchased from Click Chemistry Tools (Scottsdale, AZ). 96-well plates with 2 ml volume and v-bottom were purchased from Corning (Corning, NY). 3-Azido-7-hydroxycoumarin was purchased from AK Scientific, Inc. (Union City, CA).

Sequence alignment of BesD, HalB, and HalD. Sequences of SlBesD (WP_030791981), ScBesD (WP_014151497), SwHalB (SDN46247), PkHalD (WP_046063366), and PfHalA (WP_016975823) were aligned using MUSCLE [12], which was also used to compute the percent identity matrices.

Bacterial strains. *E. coli* DH10B-T1^R was used for plasmid construction. BL21(DE3)–Star harboring the pRARE2 plasmid was used for heterologous protein production of all halogenase variants.

Construction of plasmids. Gibson assembly [13] was used to carry out plasmid construction using *E. coli* DH10B-T1^R as the cloning host. PCR amplifications were carried out with Phusion polymerase or Q5 polymerase (New England Biolabs, Ipswich MA) using the oligonucleotides listed in *Appendix A1.4*. GeneBlock sequences are listed in *Appendix A1.4* along with primers used to construct plasmids. Following plasmid construction, all cloned inserts were sequenced at Quintara Biosciences (San Francisco, CA) or the Barker Hall Sequencing Facility at UC Berkeley (Berkeley, CA).

Plasmids for protein expression. The intermediate cloning plasmid, pET16-His-PrescissionCutSite-IMPDH, was constructed by amplification of IMPDH from *E. coli* gDNA followed by insertion into NdeI/BamHI-digested pET16b. The halogenase WP_014151497 (*S. cattleya* BesD) was amplified from genomic DNA of the corresponding organism and inserted into NdeI/BamHI-digested pET16-His₁₀-PrescissionCutSite-IMPDH. WP_046063366 (*P. kilonensis* HalD) was obtained as a GeneBlock and amplified with primers as noted (*Appendix A1.2*) before Gibson insertion into NdeI/BamHI-digested His₁₀-PrescissionCutSite-IMPDH. WP_030791981 (*S. lavanduligriseus* BesD) and SDN46247 (*S. wuyuanensis* HalB) were obtained as GeneBlocks containing the overhangs required for direct Gibson insertion into NdeI/BamHI-digested pET16-His-Prescission-IMPDH.

Plasmids for library generation and screening. To clone the M246NNK library of SwHalB, SwHalB was amplified in two pieces from the pET16-His-SwHalB plasmid. The first piece was amplified with SwHalB-F and SwHalB246NNK-R. The second piece was amplified with SwHalB246NNK-F and pET16-anneal-R. Then both pieces were inserted into NdeI/HindIII-digested pET16-His₁₀-PrescissionCutSite-IMPDH to yield the mutant constructs. The F250NNK, M246V F250NNK, M246L F250NNK, and M246I F250NNK libraries were cloned using SwHalB, SwHalB M246V, SwHalB M246L, and SwHalB M246I, respectively, as templates. The first piece was amplified with SwHalB-F and the corresponding reverse primers. The second piece was amplified with pET16-anneal-R and the corresponding forward primers. Both pieces were inserted into NdeI/HindIII-digested pET16-His₁₀-PrescissionCutSite-IMPDH to yield the mutant constructs. To generate pBesCB, a fragment containing besC-T7p-MBP-besB was PCR amplified from pPra [10] using primers pBesCB-F/R and cloned into pACYC-DUET using Gibson ligation and NcoI/KpnI cut sites. Inserts for cloning the chimeric fusion of BesD and HalB (ChimeraDB) were amplified in four pieces from pET16-His-SIBesD (pieces 1-3) or pET16-His-SwHalB (piece 4) using the ChDB primers in *Appendix A1.4*. Pieces 1-3 were annealed using overlap extension PCR [14]. The fused PCR product, along with piece 4, was inserted into NdeI/HindIII-digested pET16-His₁₀-PrescissionCutSite-IMPDH to yield the chimeric construct.

Screening the HalB library for alkyne formation. Chemically-competent *E. coli* BL21 cells harboring the pBesCB plasmid (Cm^R) were transformed with the SwHalB NNK library plasmids (Cb^R) and grown on LB plates with the appropriate antibiotic. Colonies were inoculated into individual wells of a 96-well plate containing 200 μ L LB with the appropriate antibiotic and grown overnight. Dense cultures were diluted 20-fold into 96-well plates (2 ml v-bottom, Corning, NY) containing 1 ml of LB, covered with Aeraseal Film (Sigma) and grown to mid-log phase (OD = 0.6 to 0.8) before chilling the plates on ice for 30 min. Cells were pelleted by centrifugation at 4,000 $\times$ g for 20 min and then resuspended in 1 ml of M9 media (sodium phosphate (6.6 g/L),

potassium phosphate (3 g/L), sodium chloride (0.5 g/L), ammonium chloride (1 g/L), magnesium sulfate (2 mM), calcium chloride (0.1 mM), glucose (4% (w/v)), thiamine (1 mg/L)) supplemented with 0.5 mM lysine, 0.25 mM IPTG, and the appropriate antibiotics. Cells were grown for 2 d at 16°C with shaking at 225 rpm. After 2 d of growth in M9 media, the cultures were supplemented with lysine (0.2 mM), sodium α KG (0.2 mM), and NaCl (0.2 mM) and grown for 2 additional days. Note that at the end of the day 4, some of the cells had settled in the bottom of the plate, despite shaking during growth. The remaining cells were then pelleted by centrifugation for 15 min at $4,000 \times g$. Media samples (50 μ L) containing Pra were pipetted into a 96 well black plate with a clear bottom (Corning, NY). To initiate the reaction, 50 μ L of a $2\times$ CuAAC reaction solution in $1\times$ PBS was added to each well to give the final reaction solution (3-azido-7-hydroxycoumarin (0.5 mM), copper (II) sulfate (0.5 mM), THPTA (1 mM), and sodium ascorbate (5 mM)). Reactions were performed for 30 min in the dark before analyzing fluorescence of the Pra-coumarin product (Pra-Cou, $\lambda_{\text{ex}} = 404$ nm, $\lambda_{\text{em}} = 477$ nm) using a SynergyMx Microplate Reader (BioTek) at room temperature. Cells from the fluorescent wells were grown overnight, minipreped (Qiagen), and sequenced with the Prescission-seq-F primer in *Appendix A1.4*, which anneals to the SwHalB NNK vector but not to the pBesCB vector which was also present in the cells.

Expression of His₁₀-tagged proteins. *E. coli* BL21 Star (DE3) harboring the pRARE2 plasmid was transformed with the appropriate protein expression plasmid. An overnight TB culture of the freshly transformed cells was used to inoculate TB (1 L) containing the appropriate antibiotics (50 μ g/mL carbenicillin or kanamycin with 50 μ g/mL chloramphenicol) in a 2.8 L-baffled shake flask to OD₆₀₀ = 0.05. The cultures were grown at 37°C at 200 rpm to OD₆₀₀ = 0.6 to 0.8 at which point cultures were cooled on ice for 20 min, followed by induction of protein expression with IPTG (0.25 mM) and overnight growth at 16°C. Cell pellets were harvested by centrifugation at $9,800 \times g$ for 7 min at 4°C and stored at -80°C.

Purification of His₁₀-tagged proteins. Frozen cell pellets were thawed and resuspended at 5 mL/g of cell paste in lysis buffer (50 mM HEPES, 300 mM NaCl, 10 mM imidazole, 20 mM β ME, 20% (v/v) glycerol, pH 7.5) supplemented with EDTA-free Protease Inhibitor Cocktail (Roche). The cell paste was homogenized and then lysed by sonication with a Qsonica Q700 sonicator (Amplitude = 45, 5 s on, 25 s off, 2.5 min total process time, 1/2" tip). The lysate was then centrifuged at $13,500 \times g$ for 20 min at 4°C to separate the soluble and insoluble fractions. DNA was precipitated in the soluble fraction with 0.15% (w/v) polyethyleneimine and stirring at 4°C for 30 min. The precipitated DNA was then removed by centrifugation at $13,500 \times g$ for 20 min at 4°C. The soluble lysate was incubated with Ni-NTA (0.5 mL resin/g of cell paste) for 45 min at 4°C, then resuspended and loaded onto a column by gravity flow. The column was washed with wash buffer (50 mM HEPES, 300 mM NaCl, 20 mM imidazole, 20 mM β ME, 20% (v/v) glycerol, pH 7.5) for 15-20 column vol. The column was then eluted with elution buffer (50 mM HEPES, 300 mM NaCl, 300 mM imidazole, 20 mM β ME, 20% (v/v) glycerol, pH 7.5). Fractions containing the target protein were pooled by A_{280 nm} and concentrated using an Amicon Ultra spin concentrator (10 kDa MWCO, Millipore). Protein was then exchanged into storage buffer (50 mM HEPES, 100 mM sodium chloride, 20% (v/v) glycerol, 1 mM DTT, pH 7.5) using PD-10 desalting columns. Final protein concentrations before storage were estimated using the $\epsilon_{280 \text{ nm}}$ calculated by ExPASy ProtParam [15] and measured by nanodrop. They are as follows: SwHalB: 10.67 mg/mL ($\epsilon_{280 \text{ nm}} = 45,045 \text{ M}^{-1} \text{ cm}^{-1}$), SwHalB M246L: 3.93 mg/mL ($\epsilon_{280 \text{ nm}} = 45,045 \text{ M}^{-1} \text{ cm}^{-1}$), SwHalB M246I: 4.35 mg/mL ($\epsilon_{280 \text{ nm}} = 45,045 \text{ M}^{-1} \text{ cm}^{-1}$), SwHalB M246V: 4.35 mg/mL ($\epsilon_{280 \text{ nm}} = 45,045 \text{ M}^{-1} \text{ cm}^{-1}$),

SlBesD: 6.48 mg/mL ($\epsilon_{280\text{ nm}} = 50,545\text{ M}^{-1}\text{cm}^{-1}$), ChimeraDB: 2.71 mg/mL ($\epsilon_{280\text{ nm}} = 44920\text{ M}^{-1}\text{cm}^{-1}$). All proteins were aliquoted, flash-frozen in liquid nitrogen, and stored at -80°C .

Preparation of proteins for crystallization. Following elution from the Ni-NTA column, the proteins were incubated with Prescission protease (1 mg protease/50 mg protein) and dialyzed 1:50 overnight into dialysis buffer to remove imidazole. Cleaved and dialyzed protein was passed through Ni-NTA (2 mL) to remove Prescission and the His₁₀ tag. The eluent was diluted to a final salt concentration of 20 mM NaCl using Buffer A (50 mM HEPES, 20% (v/v) glycerol, 1 mM DTT, 1 mM EDTA, pH 7.5) and loaded onto a 5 mL HiTrap-Q column for ion exchange with the AKTA Purifier FPLC system (GE Healthcare). The protein was eluted using a gradient from 0-100% buffer A to buffer B (50 mM HEPES, 1 M NaCl, 20% (v/v) glycerol, 1 mM DTT, 1 mM EDTA, pH 7.5) over 40 min. In the case of PkHalD, EDTA was omitted from the ion exchange purification procedure. The protein sample was concentrated to 2 mL and loaded onto a Superdex 75 16/60 pg (GE Healthcare) column equilibrated with SEC Buffer (20 mM HEPES, 100 mM NaCl, 1 mM DTT, pH 7.5). The protein eluent was concentrated to 15 mg/mL and glycerol was added to a final concentration of 5% (v/v) before flash freezing in liquid nitrogen and storage at -80°C .

HalB crystallization and data collection (lysine bound). Crystals of HalB from *S. wuyuanensis* (SwHalB) were obtained by the hanging drop vapor diffusion method by combining equal volumes of protein solution (SwHalB (3 mg/mL), lysine (3 mM), α KG (3 mM, pH 7)) and reservoir solution (succinate phosphate lysine (SPL) buffer (100 mM), 20% (w/v) PEG 1500, pH 5.5). The 1 M SPL buffer stock was made as a modification of SPG (succinate phosphate glycine) buffer and contained succinic acid (125.3 mM), sodium dihydrogen phosphate monohydrate (437.7 mM), and lysine (436.9 mM). Crystals grew in three days and were transferred to an Eppendorf tube containing 250 μL of reservoir solution and vortexed for 30 seconds with $10 \times 1\text{ mm}$ diameter zirconia/silica beads to produce a micro-seed solution. The seed solution was stored at -80°C for future use. Subsequent crystals were prepared by micro-seeding equal volumes of protein solution (2 mg/mL SwHalB, lysine (3 mM), α KG (3 mM, pH 7)) and reservoir solution (succinate phosphate lysine (SPL) buffer, pH 5.4 (100 mM), 20% (w/v) PEG 1500). Crystals grew in one day and were flash frozen in liquid nitrogen. Data were collected at Beamline 8.3.1 at the Advanced Light Source (Lawrence Berkeley National Laboratory) at a wavelength of 1.11 Å.

HalD crystallization. Crystals of HalD from *P. kilonensis* (PkHalD) were obtained by the hanging drop vapor diffusion method by combining equal volumes of protein solution (PkHalD (6 mg/mL), ornithine (3 mM), α KG (3 mM, pH 7)) and reservoir solution (MES buffer (100 mM), ~35% (w/v) PEG 550 MME, pH 6.3). Crystals grew within two days and microseeds were generated as described above for SwHalB. Subsequent crystals were obtained by microseeding equal volumes of protein solution (PkHalD (6 mg/mL), ornithine (3 mM), α KG (3 mM, pH 7)) and reservoir solution (MES buffer (100 mM), 27% (w/v) PEG 550 MME, pH 6.2). Crystals grew in one day and were flash frozen in liquid nitrogen. Data were collected at Beamline 8.3.1 at the Advanced Light Source (Lawrence Berkeley National Laboratory) at a wavelength of 1.11 Å.

BesD crystallization data collection. Crystals of BesD from *S. lavanduligriseus* (SlBesD) were obtained by the hanging drop vapor diffusion method by combining equal volumes of protein solution (SlBesD (10 mg/mL), lysine (3 mM), α KG (3 mM, pH 7)) and reservoir solution (magnesium acetate (180 mM), 20% (w/v) PEG 3350). Diamond-shaped crystals grew in two days

and were flash frozen in liquid nitrogen. Data were collected at Beamline 8.3.1 at the Advanced Light Source (Lawrence Berkeley National Laboratory) at a wavelength of 1.11 Å.

HalB crystallization and data collection (glycine bound). SwHalB crystals were obtained by the hanging drop vapor diffusion method by combining equal volumes of protein solution (SwHalB (2.5 mg/mL), lysine (3 mM), αKG (3 mM, pH 7)) and reservoir solution (succinate phosphate glycine (SPG) buffer (100 mM), 18% (w/v) PEG 1500, pH 5.2). The 1 M SPG buffer stock contained succinic acid (125.3 mM), sodium dihydrogen phosphate monohydrate (437.7 mM), and glycine (436.9 mM). Crystals grew in three days and were flash frozen in liquid nitrogen. Data were collected at Beamline 8.3.1 at the Advanced Light Source (Lawrence Berkeley National Laboratory) at a wavelength of 1.11 Å.

Structure solution. Data were processed with XDS [16] and scaled and merged with Aimless [17] in within the CCP4 suite [18]. The data were phased with molecular replacement using the structure of BesD (6NIE) as a search model. The structures were refined iteratively in COOT [19] and Phenix [20]. Ligands were added to the model using COOT and refined in Phenix. Omit maps for the ligand complex including were generated using the Phenix Map function by removing atoms of interest from refinement prior to calculation of the maps. The structure was analyzed in Pymol, which was also used to create figures.

Determination of kinetic parameters for SwHalB variants. Reactions (100 μL) contained ATP (2.5 mM), MgCl₂ (5 mM) phosphoenol-pyruvate (PEP, 1 mM), NADH (0.3 mM) lactate dehydrogenase (LDH, 10 U/mL), pyruvate kinase (PK, 10 U/mL) succinyl-CoA synthetase (SCS, 3.2 U/mL), coenzyme A (1 mM), sodium αKG (1 mM), (NH₄)₂Fe(SO₄)₂ · 6H₂O (0.2 mM), sodium chloride (10 mM), and sodium ascorbate (2 mM) in 100 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of the halogenase variant (2.5 μM) in the presence of varying concentrations of L-lysine · HCl (0 - 1 mM). Initial rates of NADH consumption were measured by monitoring A₃₄₀ using a SynergyMx Microplate Reader (BioTek) at room temperature. *k*_{cat} and *K*_M were determined by fitting to initial rate data with Origin (OriginLab, Northampton, MA) using the equation:

$$v_0 = \frac{k_{\text{cat}}[S]}{K_M + [S]}$$

where *v*₀ is the initial rate and [*S*] is the substrate concentration.

Succinate formation by the chimeric SIBesD and SwHalB variant (ChimeraDB). Reactions (100 μL) contained ATP (2.5 mM), MgCl₂ (5 mM) phosphoenol-pyruvate (PEP, 1 mM), NADH (0.3 mM) lactate dehydrogenase (LDH, 10 U/mL), pyruvate kinase (PK, 10 U/mL) succinyl-CoA synthetase (SCS, 3.2 U/mL), coenzyme A (1 mM), sodium αKG (1 mM), (NH₄)₂Fe(SO₄)₂ · 6H₂O (0.2 mM), sodium chloride (10 mM), and sodium ascorbate (2 mM) in 100 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of the halogenase variant (10 μM) in the presence of 5 mM L-lysine · HCl. Initial rates of NADH consumption were measured by monitoring A₃₄₀ using a SynergyMx Microplate Reader (BioTek) at room temperature.

Preparation of proteins for stopped-flow absorption experiments. Following elution from Ni-NTA column, the proteins were incubated with Prescission protease (1 mg protease/50 mg protein) and dialyzed 1:50 overnight into dialysis buffer to remove imidazole. Cleaved and dialyzed protein was passed through Ni-NTA (2 mL) to remove Prescission and the His₁₀ tag.

After addition of EDTA (1 mM, pH 8), the eluent was then diluted to a final salt concentration of 20 mM NaCl using Buffer A (50 mM HEPES, 1 mM DTT, pH 7.5) and loaded onto a 5 mL HiTrap-Q column for ion exchange with the AKTA Purifier FPLC system (GE Healthcare). The protein was eluted using a gradient from 0-100% buffer A to buffer B (50 mM HEPES, 1 M NaCl, pH 7.5) over 40 min. Fractions containing the target protein were pooled by $A_{280\text{ nm}}$ and concentrated using an Amicon Ultra spin concentrator (10 kDa MWCO, Millipore). Protein was then exchanged into storage buffer (100 mM HEPES, 100 mM sodium chloride, 5% (v/v) glycerol, pH 7.5) using PD-10 desalting columns. The protein eluent was concentrated to 1.5 mM before flash freezing in liquid nitrogen and storage at -80°C.

Stopped-flow absorption experiments. Experiments were performed in an anoxic chamber on an Applied Photophysics Ltd. (Leatherhead, UK) SX-20 instrument. An anoxic solution of halogenase (0.8 mM), Fe^{II} (0.64 mM), sodium αKG (10 mM), and either lysine or 4,4,5,5- D_4 -lysine (6 mM) was mixed at 5° C with an equal volume of air-saturated buffer (100 mM HEPES, 100 mM NaCl), giving a final limiting O_2 concentration of ~ 0.2 mM. Absorbance at 320 nm was monitored over the course of 1000 s. Data were fit in origin using the following equation:

$$A_{320}(t) = [\text{R}]_0 \varepsilon_1 [k_1 / (k_2 - k_1)] (e^{-k_1 t} - e^{-k_2 t})$$

which describes the absorbance from species I in a reaction of $\text{R} \rightarrow \text{I} \rightarrow \text{P}$ assuming R and P do not absorb. The parameters k_1 and k_2 are rate constants for the formation and decay of the ferryl intermediate, I. $[\text{R}]_0$ is the initial concentration of the reactant complex, and ε_1 is the molar absorptivity of species I.

General procedure for high resolution HPLC/MS analysis of polar metabolites with HILIC. Samples containing polar metabolites were analyzed using an Agilent 1290 UPLC on a SeQuant ZIC-pHILIC (5 μm , 2.1×100 mm; EMD-Millipore) using the following buffers: Buffer A (90% acetonitrile, 10% water, 10 mM ammonium formate) and Buffer B (90% water, 10% acetonitrile, 10 mM ammonium formate). A linear gradient from 95% to 60% Buffer A over 17 min followed by a linear gradient from 60% to 33% Buffer A over 8 min was then applied at a flow rate of 0.2 mL/min. Mass spectra were acquired in positive ionization scan mode using a QQQ (Agilent).

In vitro assays of halogenase variants. Reactions (50 μL) contained L-lysine $\cdot$ HCl (3 mM), sodium αKG (5 mM), sodium ascorbate (5 mM), $(\text{NH}_4)_2\text{Fe}(\text{SO}_4)_2 \cdot 6\text{H}_2\text{O}$ (1 mM), and sodium chloride (5 mM) in 100 mM HEPES buffer (pH 7.5). Reactions were initiated by addition of purified SwHalB variants (20 μM final concentration) and allowed to proceed for 25 min at room temperature before quenching in 2 vol of methanol with 1% (v/v) formic acid. Following centrifugation at $18,000 \times g$ for 20 min (4 °C) to remove precipitated protein, samples were analyzed by LC/MS on an Agilent 1290 UPLC-QQQ using the protocol for polar metabolite analysis above.

5.3. Results and discussion

BesD and HalB have different regioselectivities. Within the BesD family, we have identified two enzymes with differing regioselective outcomes. The enzymes BesD and HalA both chlorinate the amino acid lysine at C₄ to yield 4-Cl-lysine [10]. Through bioinformatics, we identified and characterized a homolog, HalB, which acts on the same substrate (lysine), but instead chlorinates C₅ of lysine [11] (*Figure 5.2*). An understanding of the basis for this regioselectivity could enable future efforts to engineer enzyme catalysts for regioselective halogenation of diverse small molecules.

Radical halogenases are related to the larger family of radical hydroxylases. Both enzyme classes decarboxylate α KG to succinate to generate a high-valent Fe^{IV}-oxo species which can abstract a hydrogen atom from the substrate, generating a substrate radical [21,22]. In hydroxylases, rebound of the oxo species yields a hydroxylated product. However, in halogenases, the Asp/Glu ligand of the iron is replaced by Gly/Ala, allowing space for the halide to coordinate directly to the Fe and to serve as the rebounding species, yielding a halogenated product [23]. An active area of research lies in understanding how halogenases favor rebound of the halogen rather than rebound of the hydroxyl group. Biochemical [24], spectroscopic [22,25], and computational [26] studies suggest that the precise positioning of the target C–H relative to the Fe^{IV}-oxo species is an important determining factor for selective halogenation. Remarkably, HalB and BesD have evolved to chlorinate different carbons along the sidechain of lysine while maintaining fidelity for halogenation over hydroxylation [11].

Crystal structure of HalB. To understand the structural basis for the regioselectivity of HalB, we solved the lysine-bound crystal structure of SwHalB to 2.0 Å (*Table 5.1, Figure 5.3*). Density for the lysine ligand is clearly visible in the active site of SwHalB. The data were phased by molecular replacement using the structure of BesD (PDB 6NIE) as a search model. The crystals yielded 8 chains per asymmetric unit. A comparison of SwHalB to BesD reveals that enzyme shares high structural similarity, but that the C-termini of the two enzymes, which serve as lids to close over lysine in the active site, differ in secondary structure. While the C-terminus of BesD is longer and lacks secondary structure, the C-terminus of SwHalB is shorter and α -helical in structure (*Figure 5.4*). An overlay of the crystal structures of SwHalB and BesD reveals that the enzymes have conserved carboxylate and α -amine binding residues. As a result, the lysine substrate is located in a similar position in the active site and is not translationally shifted in SwHalB compared to BesD. However, the two enzymes display a striking difference in the conformation of lysine within the active site. In the case of BesD, the 4th carbon of lysine is oriented toward the Fe complex. In contrast, the 5th carbon of SwHalB is instead presented toward the complex, while the 4th carbon is pointed away (*Figure 5.5*).

Notably, the key residues that contribute to the conformational differences of lysine are located on the C-terminal substrate-binding domains of the enzymes. In BesD, Trp239 stacks against the aliphatic sidechain of lysine in the active site. In SwHalB, this Trp residue is absent. Instead, Met246 protrudes into the active and presents C₅ (lysine) toward the Fe complex, making it accessible for halogenation. The structure suggests that Met246 makes a steric contribution to halogenation of the 5th carbon. In addition, Phe250 in SwHalB is positioned 4.1 Å away from C₄ (lysine), forming a pocket which stabilizes C₄ away from the active site.

A

% ID	BesD	HalA	HalB	HalD
BesD	-	51.8	42.4	38.6
HalA	51.8	-	41.4	38.7
HalB	42.4	41.4	-	40.8
HalD	38.6	38.7	40.8	-

B

```

HalD MSEQTSSLVIEVMEQQLAKHFQAILQDENRMKQIRNEFRRDGYFNFKNFSFLPKRILENV
HalB --MTENSITAANVEELIAKNIAERFADDHEVLGLSQHFRREGYVKLPG--LVSPVFDVAV
HalA -----MNYVLDEARLQ--EHHATNFPES-SVFALRHEFARNGFIKVRN--IVDDDLREKI
BesD -----MVGSNRQEL--KDVCAPLEKD-DIRRLSQAFHRFGIVTVTE--LIEPHTRKLV
      .           :           : .. : : : * * * ...   ::      . :

HalD HAEVHALLDEYSVRRDVTVPSTGNTYRKMYNVNQPEIAEGGTFIPALYQSESLRKFLGNI
HalB AAETHQLIDTHQKRIDIRLKETGDSPRYMSTVGQKAIATDGLIPAVYESTALKGFLSRL
HalA TREVNSLIDRQLERRDLHLATTDNTPRYMSVVRSEFIAENSTLINTLSKSKGLETLSQI
BesD RAEADRLLDQYAERRDLRLATTDYTRRSMVVPSETIAANSELVTGLYAHRELLAPLEAI
      * .  *: *      * *: :  * . : * * * .  ** .. :: :      *   *   :

HalD AGDDLASC-WEQEYQLVTKLSHPGDTHGWHWGDYPYTMIIIEAPEDPAIGGVLQCVPHS
HalB AKEEVMGCPWDEEKYIITRQHQKGDTHGWHWGDYSFTVIWIEAPSLEY-GGMLQICIPHT
HalA AGTQLIASVSKDEEYLITKQERKGDTHGWHWGDYSFALIWIIEFPSIAK-GGMLQCVPHS
BesD AGERLHPCPKADEEFLITRQEQRGDTHGWHWGDYSFALIWVLQAPPIDV-GGMLQCVPHS
      *   :   .   :*:*:*:*.   .  *****:..*:*:*:*:*   **:***:***:

HalD EWDKQNPQIWQYILNNPIKSYHHLKGDVYFLKSDTTLHHVVPPIQQETTRIILNTCWASAH
HalB DWNKDDPRVEDYLQKHPIRSYGHAKGDLYLLRSDTTLHRTVPLNADRTRIILNTCWASRA
HalA SWDKSNPRIHELCSNPIATYGFVTGDIYFLRTDTTLHRTIPLNEDATRIILNMTWAAEK
BesD TWDKASPQINRYLVENPIDTYHFESGDVYFLRTDTTLHRTIPLREDTTRIILNMTWAGER
      *: * . *: :   : . : ** : * . . *: *: *. : ***** .. : : : ***** ** .

HalD DRRTDVAHESIEVIWDTKARTTEEA---
HalB DQQKATTHETMNAMFD-----
HalA DLSRNLHGN--DRWEDQHVEAAKSLT-
BesD DLSRKLAAD--DRWWDNAEVSAAAIKD
      *           :           : :

```

Figure 5.2. Comparison of enzymes that produce 4-Cl-lysine (BesD and HalA), 5-Cl-lysine (HalB), and 4-Cl-ornithine (HalD). (A) Sequence identity matrix for BesD, HalA, HalB, and HalD. (B) Sequence alignment of BesD (*S. cattleya*, WP_106433083), HalA (*P. fluorescens*, WP_016975823), HalB (*S. wuyuanensis*, WP_093660971), and HalD (*P. kilonensis*, WP_046063366), were performed in MUSCLE [12].

Table 5.1. Data collection and refinement parameters for SwHalB.

SwHalB, lysine-bound	
Data collection	
Space group	P12 ₁ 1
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	70.241, 136.079, 116.24
α , β , γ (°)	90, 100.016, 90
Resolution (Å)	87.6–2.0 (2.071–2.0)*
<i>R</i> _{sym} or <i>R</i> _{merge}	0.1454 (1.131)
<i>I</i> / σI	9.11 (1.82)
Completeness (%)	99.69 (99.22)
Redundancy	6.9 (6.6)
CC _{1/2}	0.997 (0.658)
Refinement	
Resolution (Å)	87.6–2.0 (2.071–2.0)
No. reflections	144318 (14282)
<i>R</i> _{work} / <i>R</i> _{free}	0.1562/0.2293
No. atoms	18264
Protein	15848
Ligand/ion	64
Water	2352
<i>B</i> -factors	
Protein	28.59
Ligand/ion	28.33
Water	35.93
R.m.s. deviations	
Bond lengths (Å)	0.01
Bond angles (°)	1.31

*Values in parentheses are for highest-resolution shell.

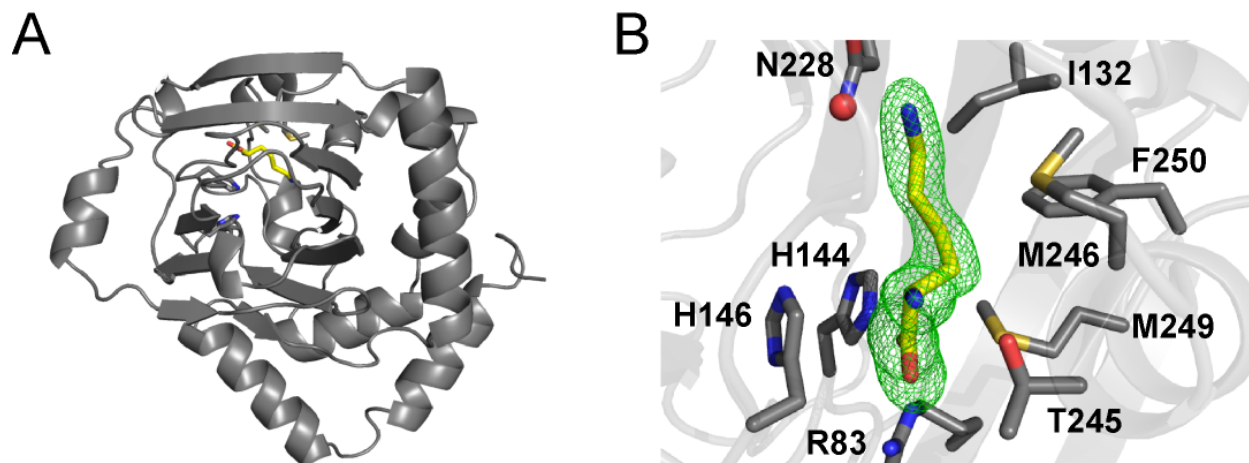


Figure 5.3. Crystal structure of SwHalB bound to lysine (yellow) at 2.0 Å resolution. (A) Overall fold of SwHalB with lysine bound. (B) $F_o - F_c$ omit map (green mesh, contoured at 3 σ) for lysine (yellow). Active site residues contacting the lysine substrate are shown as sticks in grey.

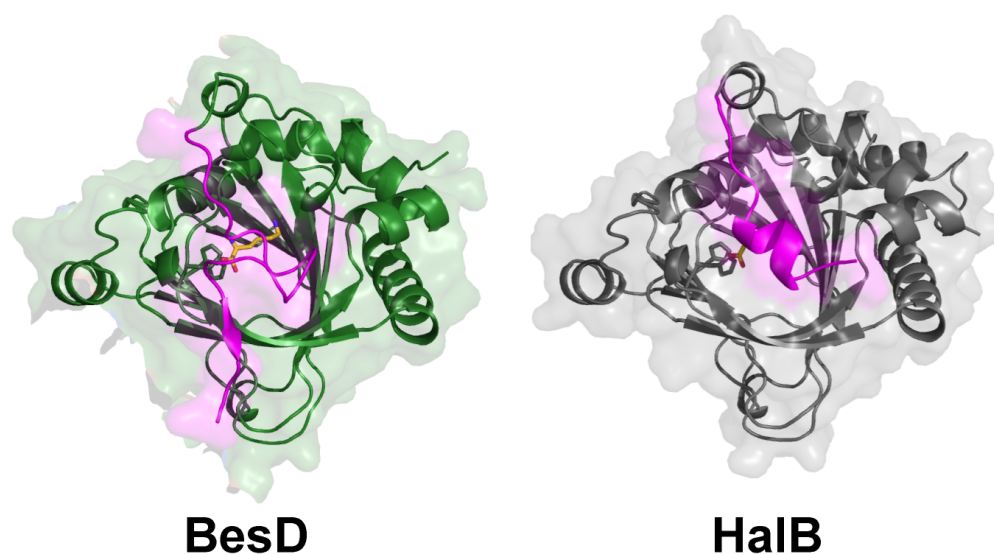


Figure 5.4. Comparison of overall structures of BesD (green, PDB 6NIE) and SwHalB (grey). The enzymes share high structural similarity except at the C-terminal substrate covering lid (magenta).

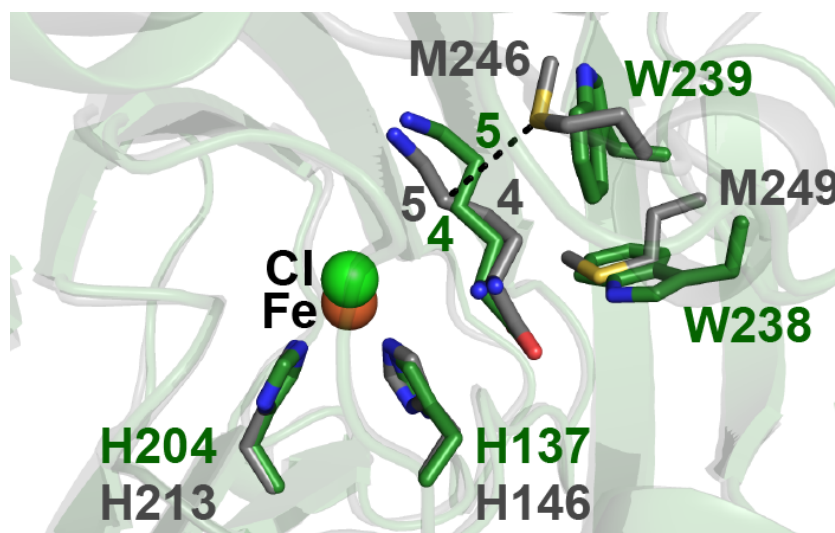


Figure 5.5. Structural alignment of BesD (green) and SwHalB (grey). Although the carboxylate and amines of lysine are in similar positions in both cases, the sidechain samples a different conformation in the two enzymes. In BesD, C₄ is oriented toward chlorine in the active site, while C₅ is oriented away. In contrast, SwHalB orients C₅ toward chlorine in the active site. The substrate conformation is influenced by residues in the substrate covering lids of BesD and SwHalB. In particular, Met246 in SwHalB makes a steric contribution to the reaction outcome by forcing C₅ toward the active site.

SF-Absorption kinetics in BesD and HalB. In radical halogenases, the site from which the hydrogen atom is abstracted determines the position of halogenation [24]. As a result, the rate of ferryl decay can provide information about the relative positioning of the ferryl intermediate and the target site of modification [24]. The structures of BesD and SwHalB suggest that the active site positions lysine such that the target carbons are a comparable distance from the ferryl intermediate; in a structural superposition, C₄ (lysine) in BesD overlays well with C₅ (lysine) of SwHalB. To probe substrate positioning by a complementary strategy, we performed stopped-flow absorption (SF-abs) experiments, monitoring the formation and decay of the ferryl intermediate which absorbs at 320 nm.

Each enzyme was pre-incubated with α KG, lysine, and chloride-containing buffer (HEPES pH 7.5) in an anaerobic chamber. The enzyme solution was then mixed against air-saturated buffer, and the formation and decay of a peak at 320 nm was monitored over time. The experiment was performed with both protiated lysine substrate, as well as 4,4,5,5-D₄-lysine in order to measure a kinetic isotope effect (*Figure 5.6*). The peaks were fit to obtain parameters for the rate of ferryl decay. The rate of hydrogen atom abstraction (corresponding to ferryl decay) was found to be 1.95 s⁻¹ for SlBesD and 1.8 s⁻¹ for SwHalB. Thus, the rate of hydrogen atom abstraction is comparable for both enzymes regardless of regiochemical outcome. With 4,4,5,5-D₄-lysine substrate, these rates slowed to 0.21 s⁻¹ (SlBesD) and 0.04 s⁻¹ (SwHalB). The observed kinetic isotope effects (KIE) are consistent with modification of C₄ (SlBesD) and C₅ (SwHalB). Because both the 4th and 5th carbons of the substrate are labeled, the kinetic isotope effects from SF-abs cannot independently distinguish which of these positions is targeted by each of the enzymes. However, LC-MS and NMR experiments have independently verified the regioselectivities [11]. The SF-abs results suggest that the position of the target site of halogenation is carefully controlled in HalB and BesD.

The KIE for SwHalB (~45) is significantly larger than the KIE in BesD (~10). One possible explanation for the small KIE is that SlBesD may perform off-target HAA (at a position other than on C₄ or C₅) when challenged with a substrate that is deuterated at those positions, as is the case in SyrB2 [24]. To test this hypothesis, we repeated the SF-abs with fully deuterated lysine. However, the KIE did not increase with the fully deuterated substrate compared to 4,4,5,5-D₄-lysine. This suggests that SlBesD is not redirecting HAA to an alternative position. Other possible explanations for the difference in KIE could be related to differences in how SlBesD and SwHalB sample donor-acceptor distances (DAD) for hydrogen atom abstraction. SlBesD may enable both the hydrogen and the deuterium to sample a DAD that facilitates tunneling. This could be further investigated by probing the temperature dependence of the KIE for both systems [27].

Probing the role of Met246 via fluorogenic screening. To further probe the significance of Met246 on regioselectivity, we developed a screen for engineering altered regioselectivity in SwHalB (*Figure 5.7*). The 4-Cl-lysine halogenase, BesD, was discovered in the biosynthetic pathway for Pra; 4-Cl-lysine serves as a pathway intermediate for the downstream enzymes BesC and BesB, which rely on chlorine as a leaving group in order to form the terminal alkyne [10]. If lysine is chlorinated at C₅ instead, BesB and BesC cannot form Pra since 5-Cl-lysine is not a substrate for those enzymes.

To establish the screening platform, we constructed cell lines containing BesC and BesB. The cells were transformed with either BesD, SwHalB or a SwHalB M246 variant, and protein expression was induced in M9 media supplemented with lysine. After two days, additional lysine

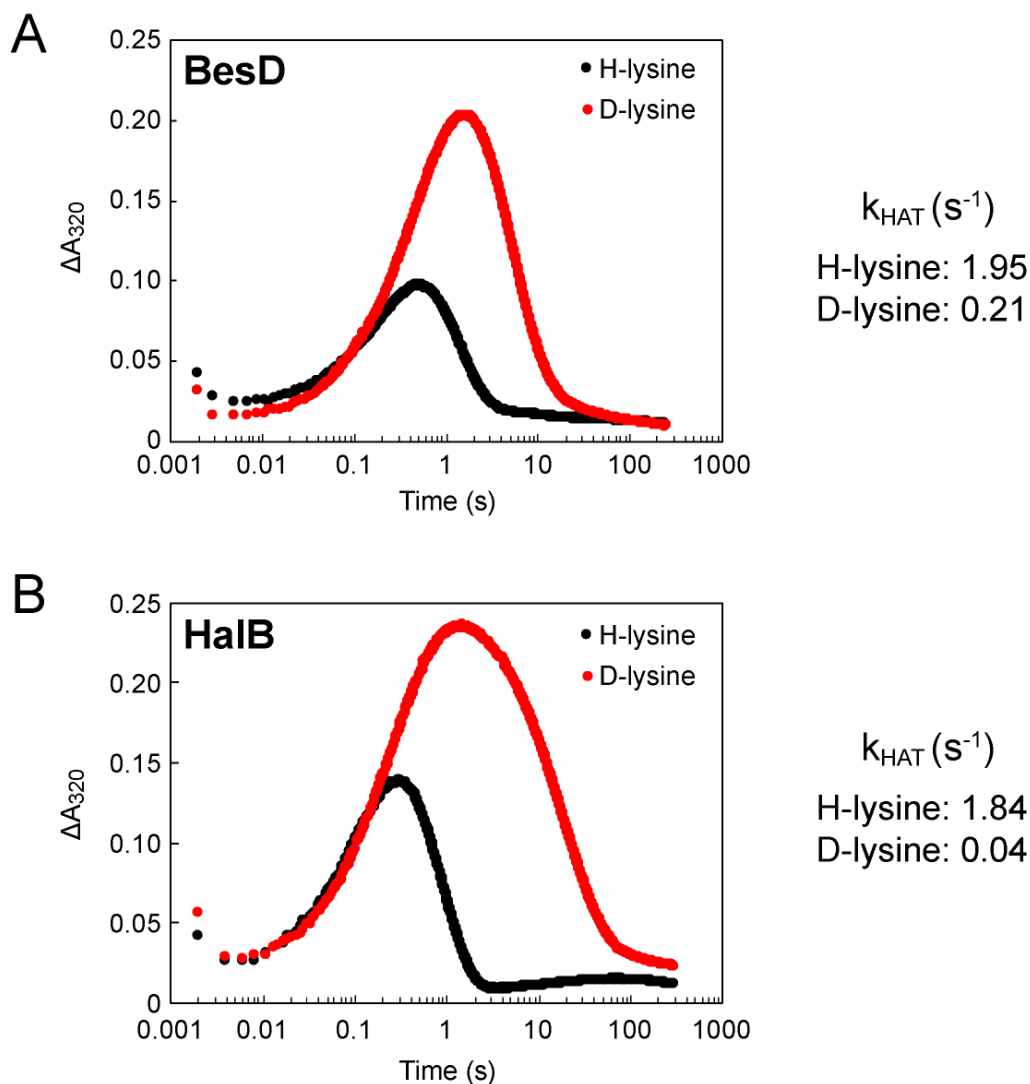


Figure 5.6. Stopped-flow absorbance measures the formation and decay of the ferryl intermediate, which absorbs at 320 nm. (A) Formation and decay of the ferryl intermediate in BesD from *S. lavanduligriseus*. An anoxic solution of SlBesD (0.8 mM), Fe^{II} (0.64 mM), sodium αKG (5 mM), lysine (6 mM, red) or 4,4,5,5- D_4 -lysine (6 mM, black) was mixed at 5° C with an equal volume of air-saturated buffer (100 mM HEPES, 100 mM NaCl), giving a final O_2 concentration of ~0.2 mM. The data were fit according to an equation describing the reaction $\text{R} \rightarrow \text{I} \rightarrow \text{P}$, assuming only I absorbs. The data shown are an average of $n=3$ technical replicates. (B) Formation and decay of the ferryl intermediate in HalB from *S. wuyuanensis*. An anoxic solution of SwHalB (0.8 mM), Fe^{II} (0.64 mM), sodium αKG (5 mM), and lysine (6 mM, red) or 4,4,5,5- D_4 -lysine (6 mM, black) was mixed at 5° C with an equal volume of air-saturated buffer (100 mM HEPES, 100 mM NaCl), giving a final O_2 concentration of ~0.2 mM. The data were fit according to an equation describing the reaction $\text{R} \rightarrow \text{I} \rightarrow \text{P}$, assuming only I absorbs. The data shown are an average of $n=3$ technical replicates.

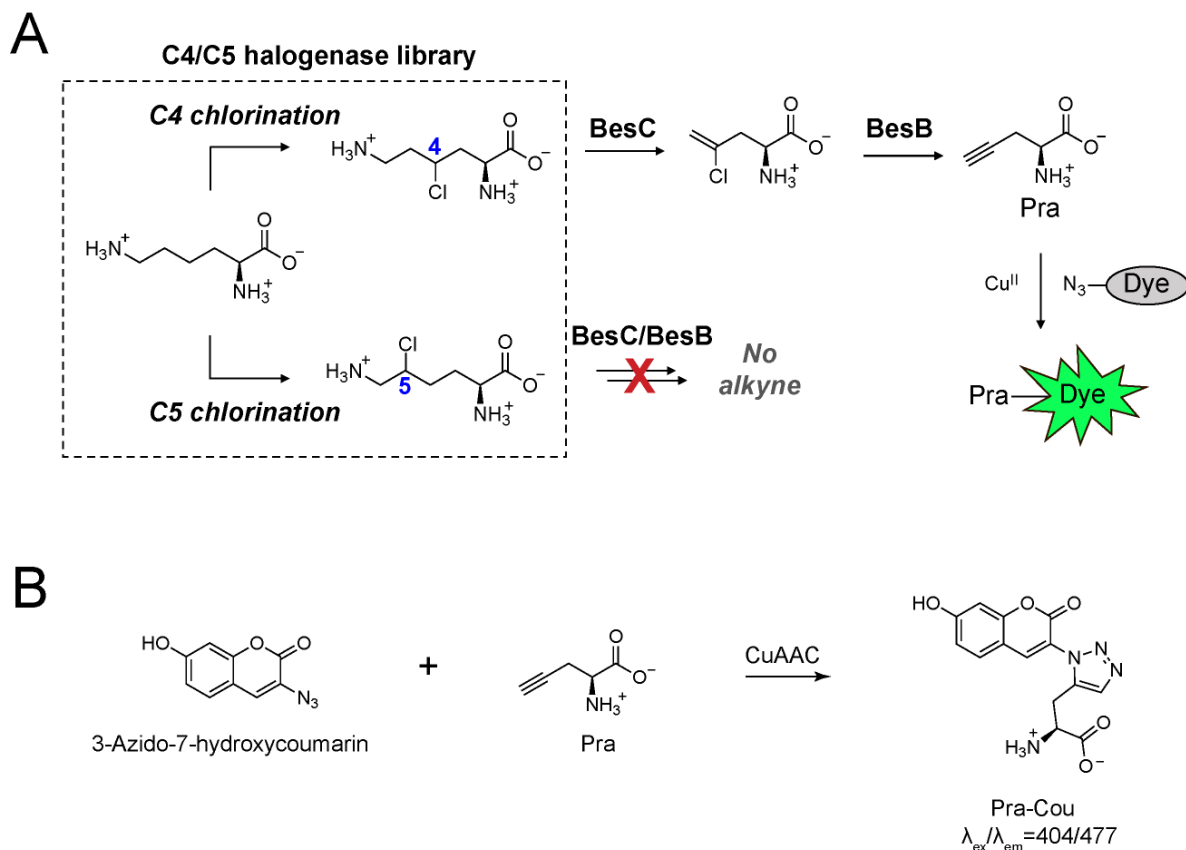


Figure 5.7. Strategy for screening the HalB library for variants that perform halogenation at C4. (A) In two enzymatic steps, 4-Cl-lysine can be converted into Propargylglycine (Pra), which contains an alkyne functional group that can be leveraged for CuAAC chemistry with a fluorogenic probe. In contrast, variants that only halogenase C5 would yield no fluorescent product downstream, as 5-Cl-lysine is not a substrate for the enzymes in the alkyne biosynthetic pathway. (B) Pra produced by cells expressing halogenase variants can react with 3-azido-7-hydroxycoumarin in the presence of copper (II), ascorbate, and the THPTA ligand. The resulting propargylglycine-coumarin product (Pra-Cou) is fluorescent and can be detected using a plate reader ($\lambda_{\text{ex}} = 404 \text{ nm}$, $\lambda_{\text{em}} = 477 \text{ nm}$).

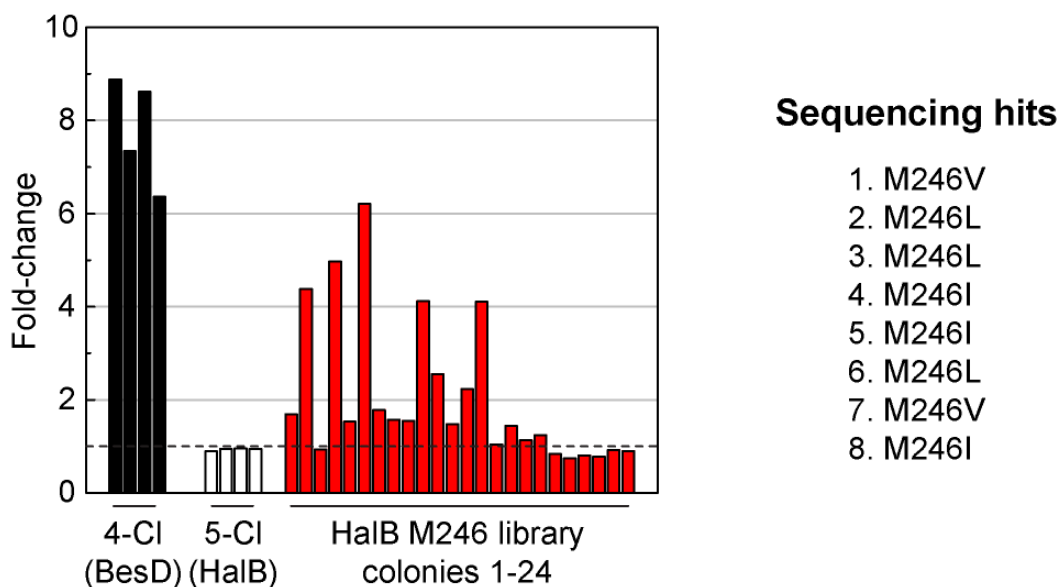


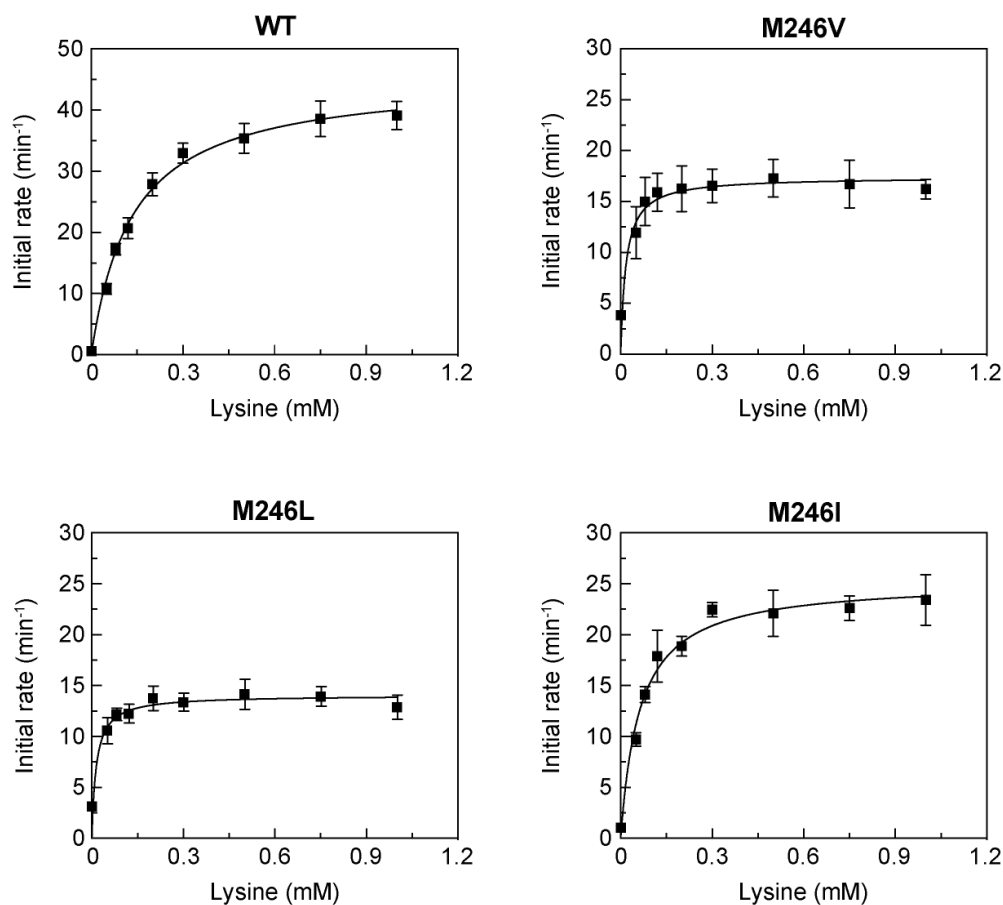
Figure 5.8. Results of a subset of the click screen. Cells expressing BesB, BesC, and a member of the NNK library (red) are grown in M9 media supplemented with 0.5 mM lysine for two days. Control samples were transformed with BesD (black) or SwHalB (white) rather than a library member. After 2 d of growth in M9 media, the cultures were supplemented with additional lysine (0.2 mM), α KG (0.2 mM), and NaCl (0.2 mM) and grown for 2 more days. Cells were then pelleted by centrifugation at 4,000 $\times$ g for 15 min. Media samples (50 μ L) containing Pra were pipetted into a 96 well black plate with a clear bottom. To initiate the reaction, 50 μ L of a 2X CuAAC reaction solution in 1 $\times$ PBS was added to each well to give the final reaction solution (3-azido-7-hydroxycoumarin (0.5 mM), copper (II) sulfate (0.5 mM), THPTA (1 mM), and sodium ascorbate (5 mM)). Reactions were performed for 30 min in the dark before analyzing fluorescence of the Pra-coumarin product (Pra-Cou, λ_{ex} = 404 nm, λ_{em} = 477 nm) using a SynergyMx Microplate Reader (BioTek) at room temperature. Plasmids from cells in fluorescent wells were isolated and sequenced to identify the mutation at position 246.

and α KG were added to the cultures. On the fourth day of expression, the cells were removed by centrifugation, and the supernatants were analyzed by CuAAC with a fluorogenic 3-azido-7-hydroxycoumarin probe. Cells expressing SwHalB gave no fold-change in fluorescence relative to the empty vector control, indicating the absence of alkyne formation. In contrast, cells expressing BesD (4-Cl-lysine halogenase) yielded a 5-8 fold change in fluorescence relative to the negative control (*Figure 5.8*). We then constructed an NNK library at position 246 and transformed the variants into cells containing plasmids for BesC/B expression. Cells were screened for alkyne formation alongside wild type BesD and SwHalB for comparison. The screen identified several wells containing the terminal alkyne product and having a fluorescence increase of 1.5-fold relative to the negative control. Plasmids from alkyne-producing cells were isolated and sequenced, identifying mutations that enabled 4-Cl-lysine formation. Of the 8 sequenced hits, 3 were leucine, 2 were isoleucine, and 3 were valine mutants. The results suggest that replacement of Met246 with a smaller, nonpolar residue may serve to alleviate steric constraints that enforce the proximity of C₅ to the Fe center, thereby enabling C₄ to be accessed for chlorination as well.

We next revisited the SwHalB crystal structure to determine if additional residues may be directly contributing to the observed regioselectivity. The C-terminus of SwHalB (residues 237-251) closes over lysine in the active site and includes residue Phe250, which is 4.1 Å away from the aliphatic sidechain of lysine and may participate in van der Waals contacts with the substrate. To probe the role of Phe250 on regioselectivity, we generated 4 NNK libraries at position 250 while keeping residue 246 constant as Met (wild type), Leu, Ile, or Val. We screened over 100 colonies from each library using the assay described above. However, sequencing of the hits yielded the original Met246 point mutants, with the original Phe at position 250. Therefore, none of the mutations at Phe250 led to an additional improvement beyond the original M246I, M246L, M246V mutations.

Steady-state kinetic characterization of HalB variants. We next investigated whether the mutations that enabled some chlorination of C₄ might also lead to diminished overall activity. To characterize the SwHalB mutants with relaxed regioselectivity (SwHalB M246V, M246L, and M246I), we expressed and purified the proteins from *E. coli* BL21. We performed steady-state kinetic analysis using an assay that couples NADH oxidation to succinate formation by the enzyme of interest (*Figure 5.9*) [28]. We found that all three mutants had a decreased k_{cat} and K_M relative to the wild type enzyme. Radical halogenases are mechanistically similar to radical hydroxylases but selectively direct halogen rebound rather than hydroxyl rebound. Precise control over substrate positioning relative to the two potential rebounding species is thought to be important for halogenation [24,25]. Therefore, it is intriguing that the two halogenases, BesD and HalB, are able to precisely orient lysine within the active site to yield the observed regioselective outcome without compromising selectivity for halogenation. While the continuous kinetic analysis of the SwHalB mutants probes the overall activity, it does not provide information about the resulting product distribution. Therefore, future work should characterize the ratio of 4-Cl-lysine to 5-Cl-lysine in the SwHalB mutants, as well as the ratio of chlorinated:hydroxylated products formed by the mutants. Since 4-Cl-lysine and 5-Cl-lysine have different retention times on a HILIC column, these experiments can be performed by LC/QTOF.

Engineering C-terminal chimeras. Taken together, the structures of BesD and SwHalB, along with the mutagenesis of Met246, suggest that residues along the C-termini of the two enzymes play a role in regioselectivity. We next attempted to determine whether grafting of the C-terminus from SwHalB onto BesD could endow BesD with the ability to produce 5-Cl-lysine. We



	k_{cat} (min ⁻¹)	K_{M} (μM)	$k_{\text{cat}}/K_{\text{M}}$ (μM ⁻¹ min ⁻¹)
WT	45 ± 1.1	130 ± 10	0.34 ± 0.10
M246V	17 ± 0.9	17 ± 8	0.40 ± 0.09
M246L	14 ± 0.7	15 ± 8	0.95 ± 0.09
M246I	25 ± 0.8	63 ± 9	1.01 ± 0.11

Figure 5.9 Kinetics of wild type SwHalB and SwHalB M246 variants on the substrate lysine. Data are mean ± sd (n = 3 technical replicates). Table contains k_{cat} , K_{M} , and $k_{\text{cat}}/K_{\text{M}}$ calculated by non-linear curve fitting to the Michaelis-Menten equation. Data in the table (kinetic parameters) are mean ± s.e. Error in $k_{\text{cat}}/K_{\text{M}}$ is obtained by propagation from the individual kinetic terms.

constructed a chimera of the two enzymes (ChimeraDB), with residues 1-227 from SIBesD followed by residues 237-261 from SwHalB (*Figure 5.10*). In addition, three other mutations (R61I, K116W, L122I) were introduced to enable BesD to accommodate the swapped residues. Note that the BesD variant from *S. lavanduligresius* rather than *S. cattleya* was used due to superior expression levels. Encouragingly, the chimeric variant expressed reasonably well in *E. coli* and could be isolated by Ni-NTA affinity purification at yields comparable to that of SIBesD. When we evaluated the activity of the variant with the NADH-coupled succinate assay, we found that ChimeraDB was able to perform the half reaction and decarboxylate α KG to succinate (*Figure 5.11-5.12*). However, no chlorinated product was observed by LC/MS. At this time, hydroxylation and uncoupling have not been investigated. Thus, this initial attempt to swap the observed regioselectivity by generating a C-terminal chimera was unsuccessful. In future attempts, additional mutations could be made to accommodate the swapped residues, and aberrant hydroxylation activity and uncoupling should also be evaluated.

Crystal structure of ornithine halogenase PkHalD. The enzyme PkHalD selectively chlorinates ornithine ($k_{\text{cat}}/K_M = 330 \pm 70 \text{ mM}^{-1}\text{min}^{-1}$) over lysine ($k_{\text{cat}}/K_M = 13.2 \pm 3.6 \text{ mM}^{-1}\text{min}^{-1}$) [11]. Interestingly, the two substrates only differ in length by one carbon. In PkHalD, the K_M for ornithine is 30 μM , which is in the range of intracellular ornithine concentrations [29]. In contrast, the lysine halogenase HalA has a K_M for ornithine of 160 μM , which is well above the expected concentration of ornithine in the cell [29]. We sought to understand how PkHalD exhibits chain-length preference for the shorter ornithine substrate by performing comparative structural studies.

To understand the structural basis for ornithine selectivity, we co-crystallized PkHalD with ornithine and solved the structure to 1.95 Å (*Figure 5.13*). The data were phased by molecular replacement using the structure of BesD (PDB 6NIE) as a search model. The crystals yielded 4 chains per asymmetric unit and have clear electron density for the substrate ornithine in the active site. In addition, density for the catalytic metal complex, including α KG and chloride ligands, was found between the Fe-binding His residues of the active site. However, since the crystals formed aerobically, the identity of the metal is unlikely to be Fe, as this would lead to catalytic turnover in the presence of oxygen. Instead, the metal is likely a trace component of the crystallization or purification conditions. Ni and Zn are possible candidates, as the protein was purified by Ni-NTA chromatography, and several crystal structures of non-heme Fe enzymes have been solved aerobically with Ni or Zn as Fe substitutes [30–32]. Regardless of its identity, the bound metal appears to be a good structural mimic of Fe based on high correlation with the complex found in the Fe-bound structure of BesD (6NIE).

Comparative analysis of halogenase structures. We next performed a structural comparison of PkHalD (4-Cl-ornithine) to BesD (4-Cl-lysine) and SwHalB (5-Cl-lysine). The cores of all three enzymes share high structural similarity, the carboxylate and α -amine binding residues are conserved, and the substrates bind in a similar location in all three cases. The majority of the structural differences occur at the C-termini. The C-terminus of the PkHalD aligns better with that of SwHalB than BesD. The C-termini of both PkHalD and SwHalB are alpha-helical in structure, while the C-terminus of BesD lacks secondary structure (*Figure 5.14-5.15*). Based on the phylogenetic tree of halogenases in the BesD family, PkHalD and SwHalB are likely more closely related evolutionarily [11]. In addition, the LCMS results suggest that the dominant lysine product produced by PkHalD is 5-Cl-lysine rather than 4-Cl-lysine [11]. Due to the high structural similarity of PkHalD and SwHalB, we chose SwHalB as the primary point of comparison for chain-length specificity.

A 3D ribbon diagram of a protein structure, likely a proteasome subunit, shown in light blue. A magenta-colored ligand, consisting of a long, flexible chain and several clusters of small spheres, is bound to the protein. The protein structure features various alpha-helices and beta-sheets. The magenta ligand is positioned in the center of the protein structure, with one long chain extending upwards and several clusters of spheres located in the middle and lower parts of the structure.

[illegible]

204

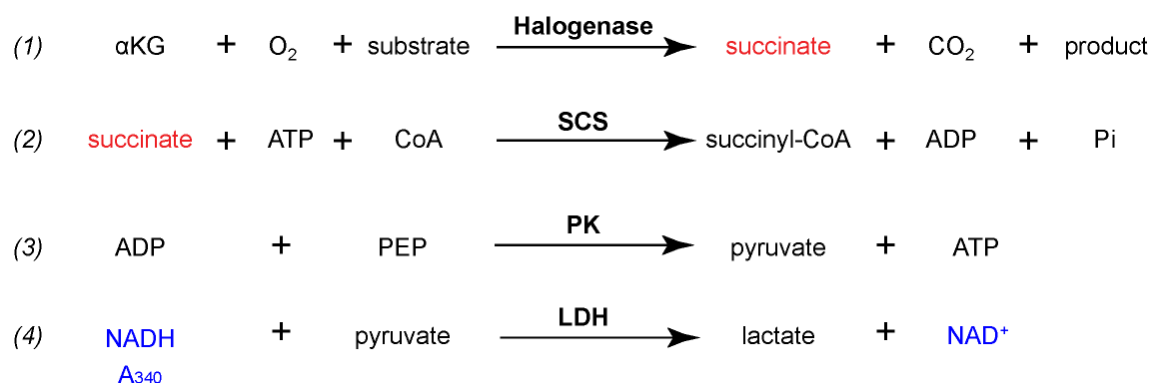


Figure 5.11. Schematic for the coupled reaction for monitoring halogenase activity through NADH oxidation. The rate of succinate formation by halogenases can be monitored by the change in A_{340} using an NADH-coupled assay with succinyl-CoA synthetase (SCS), pyruvate kinase (PK) and lactate dehydrogenase.

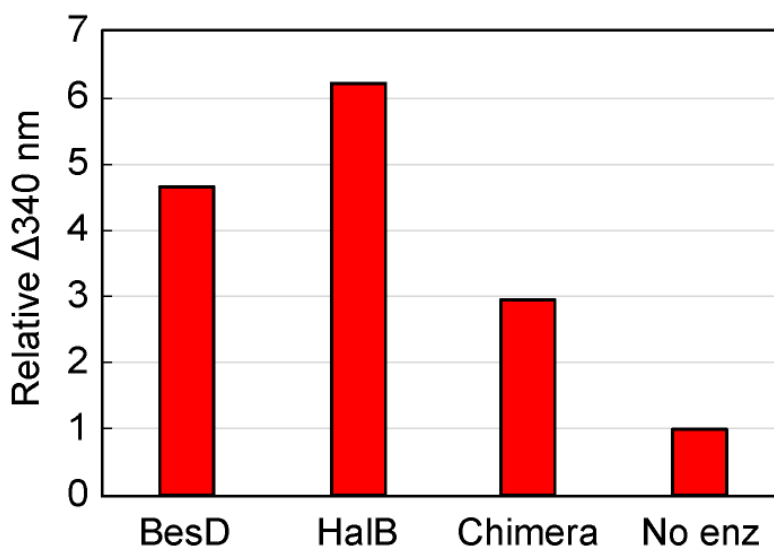


Figure 5.12. Succinate formation in SIBesD, SwHalB, and a chimera of SIBesD and SwHalB (ChimeraDB). Succinate formation is monitored by the change in A_{340} using an NADH-coupled assay as described in Figure 5.11. Reactions were initiated by the addition of SIBesD, SwHalB, or ChimeraDB (10 μM) and A_{340} was monitored using a SynergyMx Microplate Reader (BioTek) at room temperature. The rate of decrease in A_{340} is reported relative to the negative control sample containing no halogenase enzyme.

To probe the structural basis of selectivity, the binding pockets of PkHalD and SwHalB were aligned. In comparing the structures, we noted that the carboxylates and ϵ -amines of the two substrates are in very similar positions in the active site. However, while ornithine is bound in an extended conformation, the side chain of lysine is instead kinked away from the catalytic Fe, making hydrophobic contacts with a binding pocket provided by the C-terminal covering lid. In the lysine halogenase, SwHalB, the substrate binding pocket is composed of Met246, Met249, and Phe250. The pocket allows for the accommodation C₃ and C₄ of lysine, which pucker into the pocket. In the ornithine halogenase PkHalD, the pocket is smaller and is instead composed of Ile250, Ile253, and Trp254 (*Figure 5.16*). Notably, Trp254 found in PkHalD is larger than Phe250 found in BesD; as a result, the kinked sidechain of lysine may not be readily accommodated in the HalD active site. It is important to note that PkHalD does chlorinate lysine in the active site in addition to ornithine; however, the K_M value for lysine is 40 times higher (1.22 ± 0.12 mM vs $0.03 \pm .004$ mM), indicating a strong preference for ornithine [11].

Engineering altered selectivity. The structures of the SwHalB and PkHalD are highly similar. However, three key C-terminal residues tailor the binding pocket to disfavor binding of the larger lysine substrate in PkHalD. In future work, triple mutants of PkHalD should be purified (I250M I253M W254F) and characterized by steady state kinetics in order to determine whether expanding the PkHalD pocket to mimic the pocket found in SwHalB will enable efficient chlorination of lysine. Similarly, SwHalB M246I M249I W250F can be tested to see if making the binding pocket smaller can exclude lysine from binding, while allowing the enzyme to maintain activity on ornithine. The high structural similarity of the substrate lid in SwHalB and PkHalD, compared to BesD, increases the likelihood and activity can be altered within this pair by engineering. Taken together, these crystallography results suggest that engineering of the C-terminal region may enable tuning of substrate selectivity in this family of enzymes.

Conformational change upon substrate binding in the BesD family. In the structures we have solved of BesD family members (ScBesD, SwHalB, PkHalD, Hydrox), the substrate is bound in the active site by a covering lid. In the bound state, the substrate is occluded from solvent, which leads to the question of how the substrates are able to access the enzyme active sites in the first place. In the process of solving the lysine bound SwHalB crystal structure, we solved a structure of the enzyme that provides preliminary insights into this process. One of the SwHalB crystallography datasets revealed two different enzyme conformations depending on whether or not the substrate is bound (*Table 5.3, Figure 5.17*). In chain A, density for the amino acid glycine is present in the enzyme active site, in the position expected for the substrate lysine. The replacement of the lysine with glycine is likely due to the high concentration of glycine in the crystallization conditions (21.8 mM). In contrast, chain B lacks density for glycine in the active site. The lack of substrate in chain B is correlated to structural differences in the substrate binding residues of the enzyme; the C-terminal domain, which closes over the substrate in chain A, is instead disordered in the apo chain B. In addition, the antiparallel beta strands that bind the carboxylate of the substrate via Arg83 are projected away from the active site and instead adopt an extended alpha helical structure. The structure suggests that SwHalB may undergo a conformational change upon substrate binding.

We next sought to determine whether the conformational change may be an artifact of packing in this particular SwHalB crystal, or if it might be relevant for substrate binding across the BesD family. In particular, since the conformational change was observed with glycine in the crystallography conditions, it may not be representative of binding of the native substrate, lysine.

Table 5.2. Data collection and refinement parameters for PkHalD.

	PkHalD, ornithine-bound
Data collection	
Space group	P1
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	72.050, 73.226, 73.176
α , β , γ (°)	66.2385, 75.7995, 85.279
Resolution (Å)	69.84–1.95 (2.02–1.95)*
<i>R</i> _{sym} or <i>R</i> _{merge}	0.1414 (2.381)
<i>I</i> / σI	8.27 (0.91)
Completeness (%)	97.31 (95.61)
Redundancy	7.0 (6.1)
CC _{1/2}	0.998 (0.441)
Refinement	
Resolution (Å)	69.84–1.95
No. reflections	661848 (56863)
<i>R</i> _{work} / <i>R</i> _{free}	0.18/0.21
No. atoms	9221
Protein	8442
Ligand/ion	159
Water	620
<i>B</i> -factors	
Protein	43.32
Ligand/ion	62.40
Metal	37
Cl	32
Ornithine	28
α KG	35
Water	48.45
R.m.s. deviations	
Bond lengths (Å)	0.007
Bond angles (°)	1.22

*Values in parentheses are for highest-resolution shell.

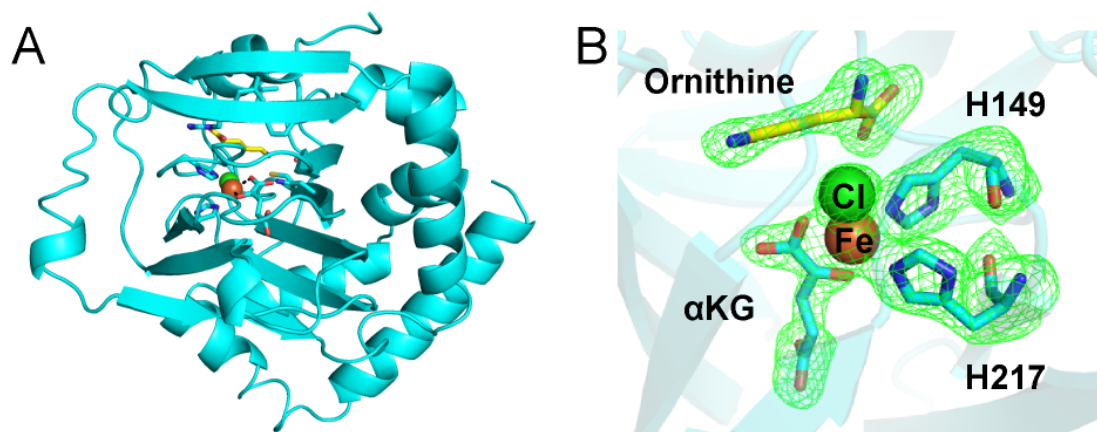


Figure 5.13 (A) Overall structure of the ornithine halogenase PkHalD. (B) $F_o - F_c$ omit map (green mesh, contoured at 2σ) for ornithine (yellow), a bound metal (modeled as Fe, brown), and chloride (green).

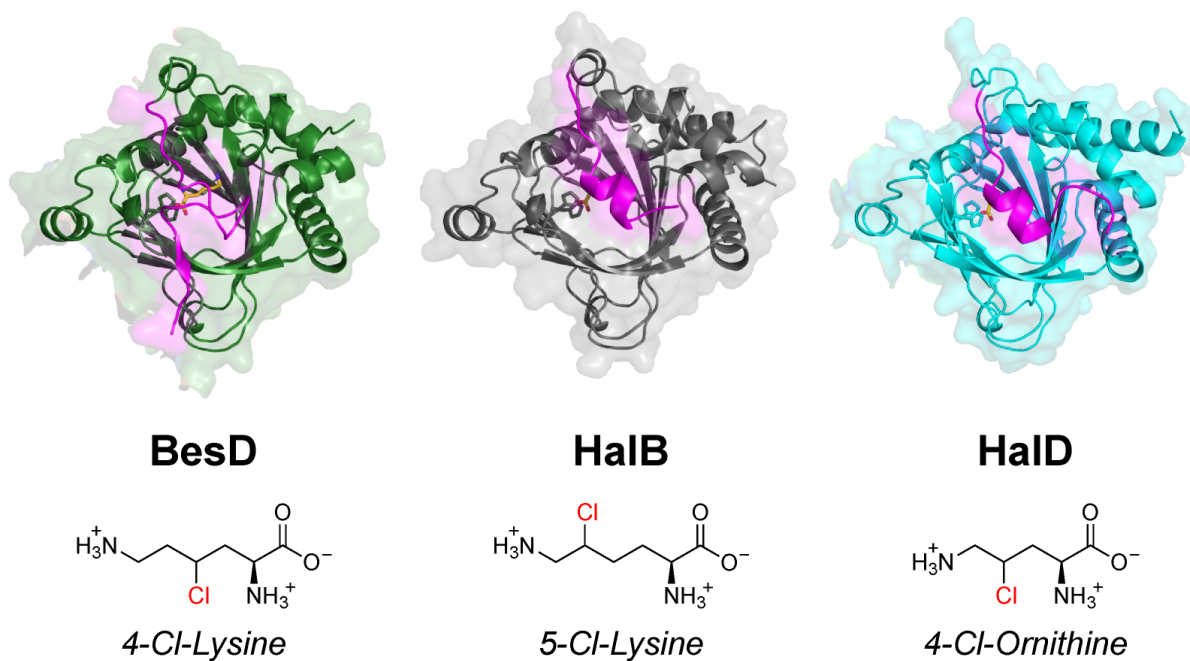


Figure 5.14. Overall fold comparison of BesD (green), HalB (grey), and HalD (cyan) with substrate covering lids shown in magenta. The substrate covering lid of PkHalD is more structurally similar to the lid in SwHalB than to the lid in BesD.

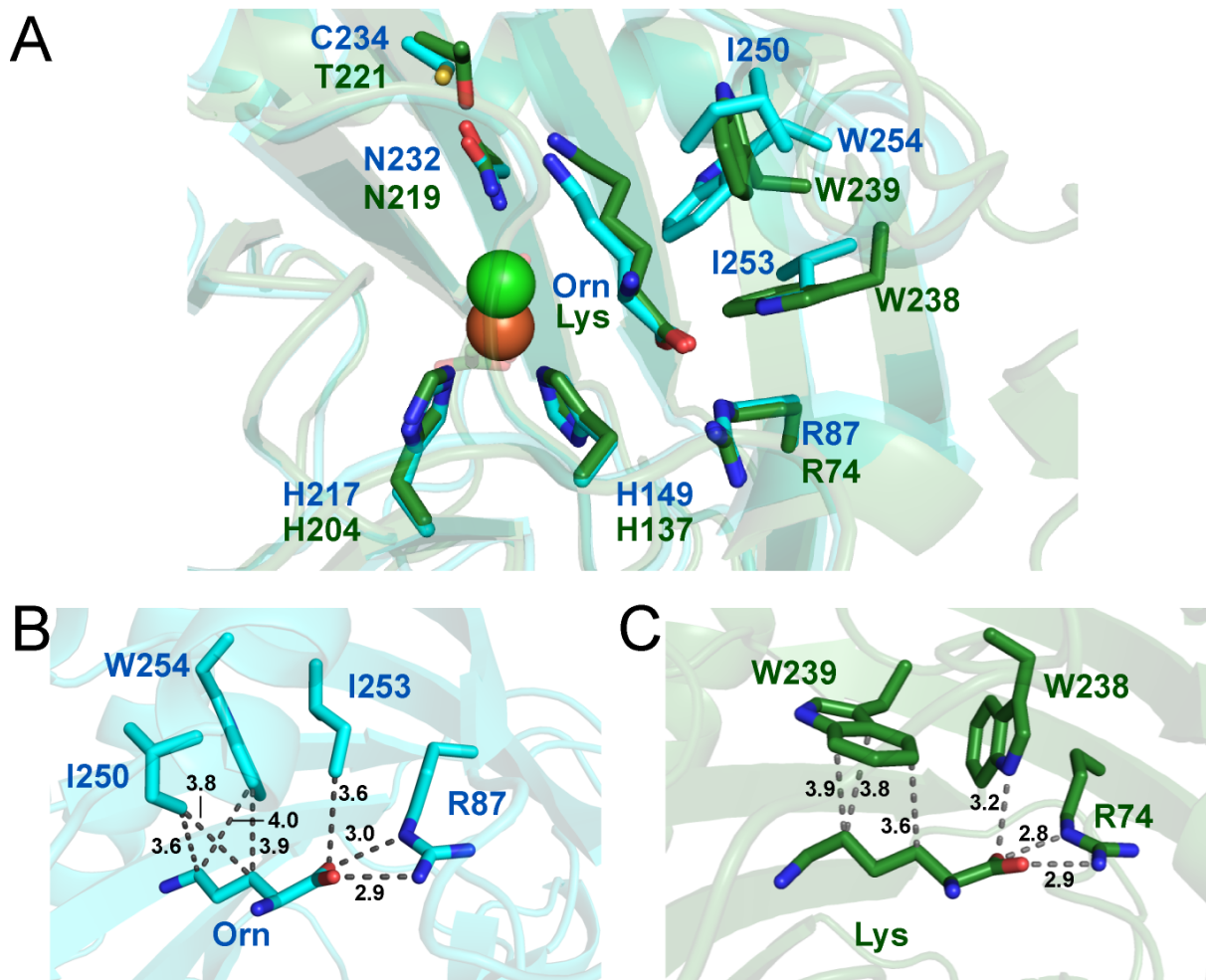


Figure 5.15. Active site comparison of HalD (cyan) and BesD (green). (A) An overlay of the active sites of PkHalD and BesD reveals high similarity in the carboxylate-binding and ϵ -amine-binding residues. In both cases, the substrate is modified at C₄. However, the structure of the substrate binding pocket that lines the aliphatic sidechain differs between the two enzymes. (B) In the ornithine halogenase, PkHalD, the substrate binding pocket is composed of Ile250, Ile253, and Trp254. (C) In the lysine halogenase, BesD, the substrate binding pocket is composed of Trp238 and Trp239.

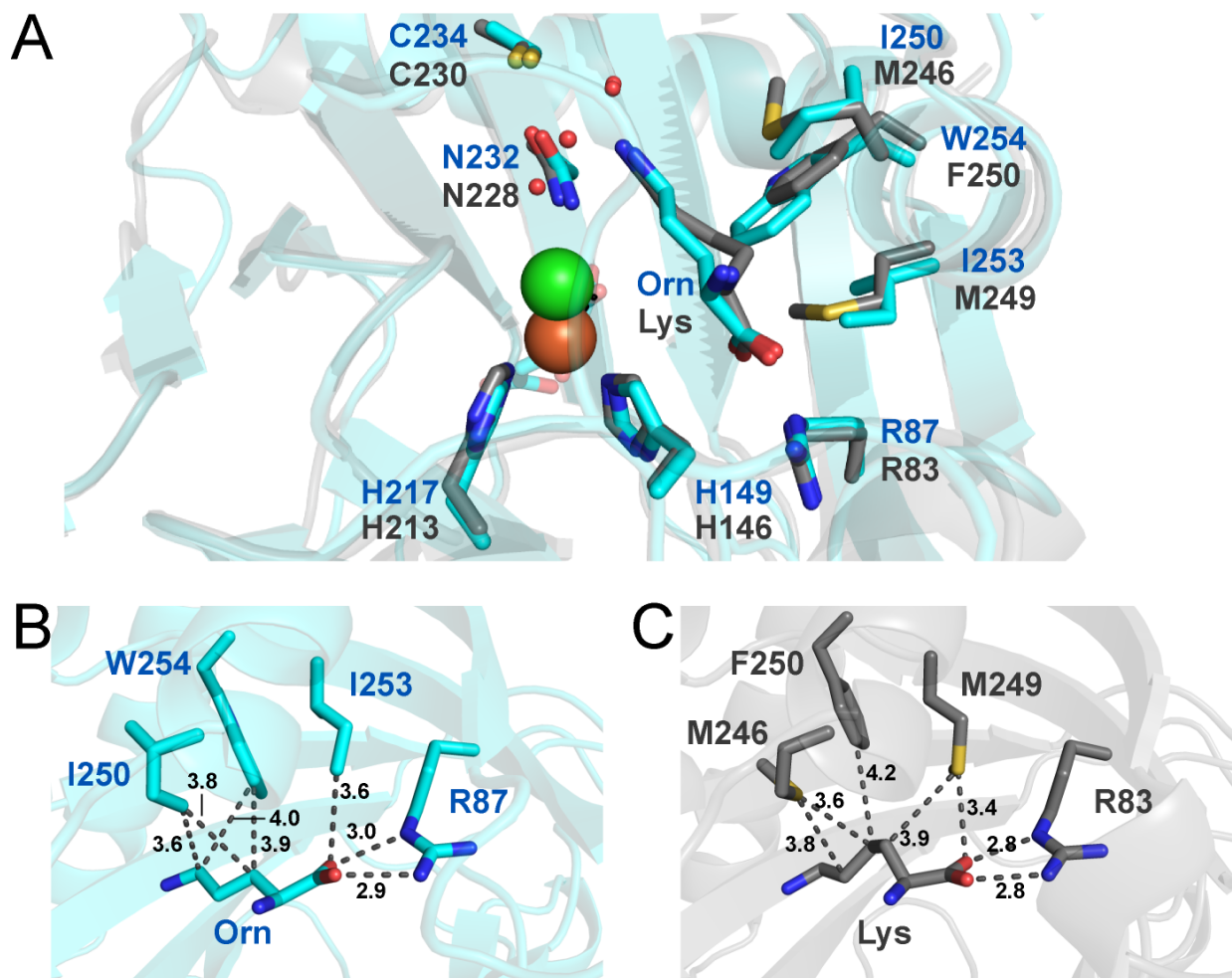


Figure 5.16. Active site comparison of HalD (cyan) and HalB (grey). (A) An overlay of the active sites of PkHalD and SwHalB reveals high similarity in the carboxylate-binding and ϵ -amine-binding residues. However, while ornithine is in an extended conformation, lysine is puckered into a substrate binding pocket formed by three residues on the C-terminal substrate binding lid. (B) In the ornithine halogenase, PkHalD, the substrate binding pocket is composed of Ile250, Ile253, and Trp254. (C) In the lysine halogenase, SwHalB, the substrate binding pocket is composed of Met246, Met249, and Phe250.

Through the structural investigation of this enzyme family, we solved a total of three different enzyme crystal structures that support the hypothesis of a conformational change upon substrate binding. In addition to SwHalB, which produces 5-Cl-lysine, we solved an apo crystal structure of BesD from *S. lavanduligriseus* (SlBesD) that lacks density for the C-terminus and also demonstrates the structural differences at the key antiparallel beta strands when substrate is not bound (*Table 5.4, Figure 5.18*). Since SlBesD is less than 50% identical in sequence to SwHalB, chlorinates lysine at C₄ and not C₅, and crystallizes under different conditions, this finding suggests that a conformational change is not specific to SwHalB. Furthermore, when we solved the crystal structure of a closely related lysine hydroxylase (Hydrox, *Chapter 4*), we found that Hydrox also displays a helix-to-sheet transition similar to SwHalB. These discoveries suggest that a conformational change upon substrate binding may be operational throughout the BesD family of non-heme Fe enzymes. Future work should investigate whether these changes are important for substrate binding in solution.

Table 5.3. Data collection and refinement parameters for glycine-bound and apo SwHalB showing a conformational change upon substrate binding.

	SwHalB glycine-bound (ChainA) and apo (ChainB)
Data collection	
Space group	P6 ₂ 22
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	165.31, 165.31, 105.24
α , β , γ (°)	90, 90, 120
Resolution (Å)	82.65–2.03 (2.103–2.03)*
<i>R</i> _{sym} or <i>R</i> _{merge}	0.01841 (0.5973)
<i>I</i> / σ <i>I</i>	31.02 (1.29)
Completeness (%)	99.98 (99.93)
Redundancy	2.0 (2.0)
CC _{1/2}	1 (0.889)
Refinement	
Resolution (Å)	82.65–2.03
No. reflections	54981 (5407)
<i>R</i> _{work} / <i>R</i> _{free}	.189/0.209
No. atoms	4474
Protein	3880
Ligand/ion	8
Water	586
<i>B</i> -factors	43.22
Protein	42.54
Ligand/ion	38.31
Glycine	41
succinate	38
Water	47.80
R.m.s. deviations	
Bond lengths (Å)	0.015
Bond angles (°)	1.31

*Values in parentheses are for highest-resolution shell.

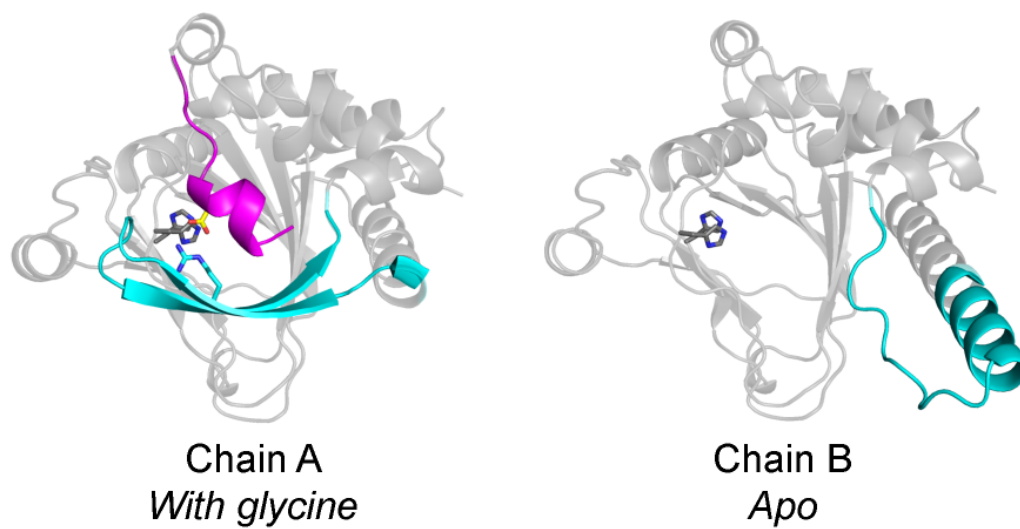
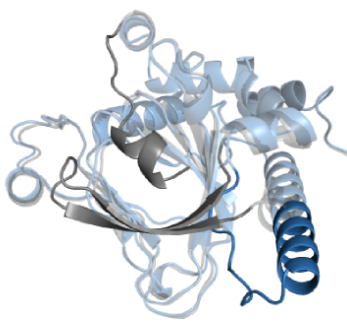


Figure 5.17. A crystal structure of SwHalB with two molecules per asymmetric unit reveals a potential conformational change upon substrate binding. Chain A has density for the amino acid glycine (yellow) in the enzyme active site. A pair of antiparallel beta strands (cyan) containing Arg83, which binds the carboxylate of glycine, closes over the active site of SwHalB in chain A. In addition, chain A has well-defined density for the substrate covering lid (magenta). In contrast, chain B lacks density for the substrate and for the substrate covering lid. In addition, the region of the protein containing Arg83 (cyan) adopts an alpha helical formation rather than antiparallel beta strands. Note that glycine is bound in chain A, rather than the native substrate, lysine. This is likely due to the high concentration of glycine (21.8 mM) present in the crystallization buffer.

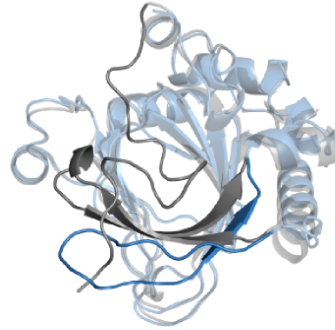
Table 5.4. Data collection and refinement parameters for apo SlBesD.

Apo SlBesD	
Data collection	
Space group	P3 ₁ 21
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	88.688, 88.688, 148.129
α , β , γ (°)	90, 90, 120
Resolution (Å)	53.31–2.601 (2.694–2.601)*
<i>R</i> _{sym} or <i>R</i> _{merge}	0.2501 (6.736)
<i>I</i> / σI	9.04 (0.42)
Completeness (%)	98.86 (89.85)
Redundancy	18.9 (13.1)
CC _{1/2}	0.999 (0.228)
Refinement	
Resolution (Å)	53.31–2.601
No. reflections	21081 (1895)
<i>R</i> _{work} / <i>R</i> _{free}	0.2919/0.326
No. atoms	3415
Protein	3415
<i>B</i> -factors	81.31
Protein	81.31
R.m.s. deviations	
Bond lengths (Å)	0.004
Bond angles (°)	1.01

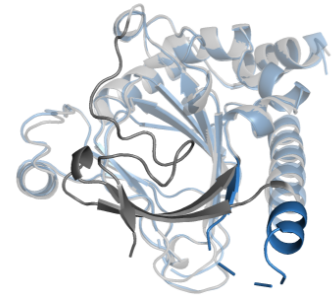
*Values in parentheses are for highest-resolution shell.



HalB



BesD



Hydrox

Figure 5.18. A comparison of substrate bound (grey) and apo (blue) structures of SwHalB, SlBesD, and Hydrox reveal structural differences upon substrate binding. In all three cases, the substrate-bound structures possess defined density for a substrate covering lid that is absent in the apo structures. All three structures also demonstrate structural differences in the region of the protein that forms antiparallel beta strands upon substrate binding.

5.4. Conclusion

The newly discovered halogenases (HalA-E) directly chlorinate amino acids, greatly increasing the number of compounds accessible through enzymatic halogenation. Furthermore, the simple substrates of these enzymes make them promising candidates for engineering novel halogenases with an expanded substrate scope. Toward this goal, we have investigated the mechanisms by which enzymes in the BesD family evolved diverse regioselectivity and substrate selectivity. In particular, we performed structural and biochemical comparisons of BesD (4-Cl-lysine product), HalB (5-Cl-lysine product) and HalD (4-Cl-ornithine product). These three enzymes are able to achieve subtly different reactions on very similar substrates.

BesD and HalB achieve divergent regioselective outcomes by controlling the conformation of the lysine substrate within the active site. A comparison of the lysine-bound structures of HalB and BesD reveals differences in the substrate covering lids that serve to orient the substrate to achieve the desired regioselectivity. In particular, Met246 shapes the substrate pocket of HalB such that C₅ (lysine) is most proximal to the catalytic site. Mutagenesis of Met246 to the smaller nonpolar residues Leu, Ile, and Val relieves the steric constraints and disrupts the regioselectivity, leading to some halogenation at C₄ as well.

The substrate covering lid of HalD enables selectivity of the smaller, 5-carbon substrate, ornithine, over the larger 6-carbon substrate, lysine. The covering lid of HalD is structurally similar to that of HalB but varies in the identity of residues that form the substrate binding pocket. The pocket of HalD, which contains Trp254 rather than the Phe250 found in HalB, is smaller, and therefore fails to accommodate lysine as effectively as it does ornithine. These structural differences likely contribute to the 40-fold difference in K_M when comparing the activity of HalD on ornithine and lysine.

Finally, we have solved crystal structures that reveal conformational changes in the enzymes upon substrate binding. In structures of BesD, HalB, and Hydrox, the substrate covering lid becomes ordered and closes over the active site upon substrate binding. Taken together, this work suggests that the covering lid is a promising target for future engineering efforts within the BesD halogenase family.

5.5. References

1. Kan, S. B. J., Lewis, R. D., Chen, K. & Arnold, F. H. Directed evolution of cytochrome c for carbon–silicon bond formation: Bringing silicon to life. *Science* **354**, 1048–1051 (2016) doi:10.1126/science.aah6219.
2. Kan, S. B. J., Huang, X., Gumulya, Y., Chen, K. & Arnold, F. H. Genetically programmed chiral organoborane synthesis. *Nature* **552**, 132 (2017) doi:10.1038/nature24996.
3. Zhang, J., Huang, X., Zhang, R. K. & Arnold, F. H. Enantiodivergent α -Amino C–H Fluoroalkylation Catalyzed by Engineered Cytochrome P450s. *J. Am. Chem. Soc.* **141**, 9798–9802 (2019) doi:10.1021/jacs.9b04344.
4. Prier, C. K., Zhang, R. K., Buller, A. R., Brinkmann-Chen, S. & Arnold, F. H. Enantioselective, intermolecular benzylic C–H amination catalysed by an engineered iron-haem enzyme. *Nat. Chem.* **9**, 629–634 (2017) doi:10.1038/nchem.2783.
5. Vaillancourt, F. H., Yin, J. & Walsh, C. T. SyrB2 in syringomycin E biosynthesis is a nonheme Fe(II) α -ketoglutarate- and O₂-dependent halogenase. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 10111–10116 (2005) doi:10.1073/pnas.0504412102.
6. Hillwig, M. L. & Liu, X. A new family of iron-dependent halogenases acts on freestanding substrates. *Nat. Chem. Biol.* **10**, 6–10 (2014) doi:10.1038/nchembio.1625.
7. Pratter, S. M., Light, K. M., Solomon, E. I. & Straganz, G. D. The Role of Chloride in the Mechanism of O₂ Activation at the Mononuclear Nonheme Fe(II) Center of the Halogenase HctB. *J. Am. Chem. Soc.* **136**, 9385–9395 (2014) doi:10.1021/ja503179m.
8. Zhang, Y., Zhao, C., Yan, S., Li, Q., Zhu, H., Zhong, Z., Ye, Y. & Deng, Z. An iron(II)- and α -ketoglutarate-dependent halogenase acts on nucleotide substrates. *Angew. Chem. Int. Ed. Engl.* anie.201914994 (2020) doi:10.1002/anie.201914994.
9. Gkotsi, D. S., Dhaliwal, J., McLachlan, M. M., Mulholland, K. R. & Goss, R. J. Halogenases: powerful tools for biocatalysis (mechanisms, applications, and scope). *Curr. Opin. Chem. Biol.* **43**, 119–126 (2018) doi:10.1016/j.cbpa.2018.01.002.
10. Marchand, J. A., Neugebauer, M. E., Ing, M. C., Lin, C.-I., Pelton, J. G. & Chang, M. C. Y. Discovery of a pathway for terminal-alkyne amino acid biosynthesis. *Nature* **567**, 420–424 (2019) doi:10.1038/s41586-019-1020-y.
11. Neugebauer, M. E., Sumida, K. H., Pelton, J. G., McMurry, J. L., Marchand, J. A. & Chang, M. C. Y. A family of radical halogenases for the engineering of amino-acid-based products. *Nat. Chem. Biol.* **15**, 1009–1016 (2019) doi:10.1038/s41589-019-0355-x.
12. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004) doi:10.1186/1471-2105-5-113.
13. Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A. & Smith, H. O. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–5 (2009) doi:10.1038/nmeth.1318.
14. Bryksin, A. & Matsumura, I. Overlap extension PCR cloning: a simple and reliable way to create recombinant plasmids. *Biotechniques* **48**, 463–465 (2010) doi:10.2144/000113418.
15. Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D. & Bairoch, A. ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res.* **31**, 3784–8 (2003) doi:10.1093/nar/gkg563.
16. Kabsch, W. XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 125–32 (2010) doi:10.1107/S0907444909047337.
17. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D. Biol. Crystallogr.* **69**, 1204–14 (2013) doi:10.1107/S0907444913000061.

18. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G. W., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A. & Wilson, K. S. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D. Biol. Crystallogr.* **67**, 235–42 (2011) doi:10.1107/S0907444910045749.
19. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of *Coot*. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 486–501 (2010) doi:10.1107/S0907444910007493.
20. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D. Biol. Crystallogr.* **66**, 213–21 (2010) doi:10.1107/S0907444909052925.
21. Agarwal, V., Miles, Z. D., Winter, J. M., Eustáquio, A. S., El Gamal, A. A. & Moore, B. S. Enzymatic halogenation and dehalogenation reactions: pervasive and mechanistically diverse. *Chem. Rev.* **117**, 5619–5674 (2017) doi:10.1021/acs.chemrev.6b00571.
22. Wong, S. D., Srnec, M., Matthews, M. L., Liu, L. V., Kwak, Y., Park, K., Bell, C. B., Alp, E. E., Zhao, J., Yoda, Y., Kitao, S., Seto, M., Krebs, C., Bollinger, J. M. & Solomon, E. I. Elucidation of the Fe(IV)=O intermediate in the catalytic cycle of the halogenase SyrB2. *Nature* **499**, 320–323 (2013) doi:10.1038/nature12304.
23. Blasiak, L. C., Vaillancourt, F. H., Walsh, C. T. & Drennan, C. L. Crystal structure of the non-haem iron halogenase SyrB2 in syringomycin biosynthesis. *Nature* **440**, 368–371 (2006) doi:10.1038/nature04544.
24. Matthews, M. L., Neumann, C. S., Miles, L. a, Grove, T. L., Booker, S. J., Krebs, C., Walsh, C. T. & Bollinger, J. M. Substrate positioning controls the partition between halogenation and hydroxylation in the aliphatic halogenase, SyrB2. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 17723–17728 (2009) doi:10.1073/pnas.0909649106.
25. Martinie, R. J., Livada, J., Chang, W., Green, M. T., Krebs, C., Bollinger, J. M. & Silakov, A. Experimental correlation of substrate position with reaction outcome in the aliphatic halogenase, SyrB2. *J. Am. Chem. Soc.* **137**, 6912–6919 (2015) doi:10.1021/jacs.5b03370.
26. Srnec, M. & Solomon, E. I. Frontier molecular orbital contributions to chlorination versus hydroxylation selectivity in the non-heme iron halogenase SyrB2. *J. Am. Chem. Soc.* **139**, 2396–2407 (2017) doi:10.1021/jacs.6b11995.
27. Klinman, J. P. & Offenbacher, A. R. Understanding Biological Hydrogen Transfer Through the Lens of Temperature Dependent Kinetic Isotope Effects. *Acc. Chem. Res.* **51**, 1966–1974 (2018) doi:10.1021/acs.accounts.8b00226.
28. Luo, L., Pappalardi, M. B., Tummino, P. J., Copeland, R. A., Fraser, M. E., Grzyska, P. K. & Hausinger, R. P. An assay for Fe(II)/2-oxoglutarate-dependent dioxygenases by enzyme-coupled detection of succinate formation. *Anal. Biochem.* **353**, 69–74 (2006) doi:10.1016/j.ab.2006.03.033.
29. Guo, A. C., Jewison, T., Wilson, M., Liu, Y., Knox, C., Djoumbou, Y., Lo, P., Mandal, R., Krishnamurthy, R. & Wishart, D. S. ECMD: the E. coli Metabolome Database. *Nucleic Acids Res.* **41**, D625–30 (2013) doi:10.1093/nar/gks992.
30. Hewitson, K. S., Holmes, S. L., Ehrismann, D., Hardy, A. P., Chowdhury, R., Schofield, C. J. & McDonough, M. A. Evidence That Two Enzyme-derived Histidine Ligands Are Sufficient for Iron Binding and Catalysis by Factor Inhibiting HIF (FIH). *J. Biol. Chem.* **283**, 25971 (2008) doi:10.1074/JBC.M804999200.

31. Kawamura, A., Münzel, M., Kojima, T., Yapp, C., Bhushan, B., Goto, Y., Tumber, A., Katoh, T., King, O. N. F., Passioura, T., Walport, L. J., Hatch, S. B., Madden, S., Müller, S., Brennan, P. E., Chowdhury, R., Hopkinson, R. J., Suga, H. & Schofield, C. J. Highly selective inhibition of histone demethylases by de novo macrocyclic peptides. *Nat. Commun.* **8**, 14773 (2017) doi:10.1038/ncomms14773.
32. Rydzik, A. M., Chowdhury, R., Kochan, G. T., Williams, S. T., McDonough, M. A., Kawamura, A. & Schofield, C. J. Modulating carnitine levels by targeting its biosynthesis – selective inhibition of γ -butyrobetaine hydroxylase. *Chem. Sci.* **5**, 1765–1771 (2014) doi:10.1039/C4SC00020J.

Appendix 1: *Supplementary Tables and Sequences*

Table A1.1. Strains, plasmids, oligonucleotides, and sequences used in Chapter 2. (A) Strains and plasmids used in this study. (B) Oligonucleotides used for plasmid construction and sequencing.

A1.1A. Strains and plasmids

Strain	Description	Source
BL21 Star (DE3)	<i>F-ompT hsdSB (rB⁻, mB⁻) gal dcm me131</i> (DE3)	Thermo-Fisher
GM272	<i>F- dam-3 dcm-6 hsdS2I metb1 lacYI or Z4 galK2 ga/T22 mtl-2 tonA2or A31 tsx-1 or -78 supE44 (thi-1)</i>	ATCC
B834(DE3)	<i>F⁻ ompT hsdSB(rB⁻ mB⁻) gal dcm met</i> (DE3)	Novagen
<i>S. cattleya</i>	NRRL 8057 (ATCC 35852)	ATCC
<i>S. cattleya</i> Δ besA	Δ besA::Am ^R -OriT	This study
<i>S. cattleya</i> Δ besB	Δ besB::Am ^R , -OriT	This study
<i>S. cattleya</i> Δ besC	Δ besC::Am ^R -OriT	This study
<i>S. cattleya</i> Δ besD	Δ besD::Am ^R -OriT	This study
<i>S. cattleya</i> Δ besE	Δ besE::Am ^R -OriT	This study
<i>S. cattleya</i> Δ besB+besB	Δ besB::Am ^R attB ^{ΦC31} ::ermEp*-besB-Hyg ^R	This study
<i>S. cattleya</i> Δ besC+besC	Δ besC::Am ^R attB ^{ΦC31} ::ermEp*-besC-Hyg ^R	This study
<i>S. cattleya</i> Δ besD+besD	Δ besD::Am ^R attB ^{ΦC31} ::ermEp*-besD-Hyg ^R	This study

Plasmid	Description	Source
pUZ8002	<i>tra, neo, Km^R, RP4</i>	Ref. [1]
pIJ773	<i>lacI, Cb^r, Am^R, OriT, ColE1</i>	JIC
pIJ773-BesA	<i>besA</i> flanking regions (2 kB US/DS)-Am ^R -OriT, <i>lacI</i> , Cb ^R , ColE1	This study
pIJ773-BesB	<i>besB</i> flanking regions (2 kB US/DS)-Am ^R -OriT, <i>lacI</i> , Cb ^R , ColE1	This study
pIJ773-BesC	<i>besC</i> flanking regions (2 kB US/DS)-Am ^R -OriT, <i>lacI</i> , Cb ^R , ColE1	This study
pIJ773-BesD	<i>besD</i> flanking regions (2 kB US/DS)-Am ^R -OriT, <i>lacI</i> , Cb ^R , ColE1	This study
pIJ773-BesE	<i>besE</i> flanking regions (2 kB US/DS)-Am ^R -OriT, <i>lacI</i> , Cb ^R , ColE1	This study
pSET152	Hyg ^R , <i>int(φ31)</i> , OriT, pUC18	JIC
pSET152-BesB	BesB (ErmEp*), Hyg ^R , <i>int(φ31)</i> , OriT, pUC18	This study
pSET152-BesC	BesC (ErmEp*), Hyg ^R , <i>int(φ31)</i> , OriT, pUC18	This study
pSET152-BesD	BesD (ErmEp*), Hyg ^R , <i>int(φ31)</i> , OriT, pUC18	This study
pET16b-His ₁₀ -BesA	His ₁₀ -BesA (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -BesD	His ₁₀ -BesD (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -BesE	His ₁₀ -BesE (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-Strep-BesD	Strep-BesD (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-Strep-BesD G139D	Strep-BesD G139D (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pSV272.1-His ₆ MBP-BesB	His ₆ MBP-BesB (T7), <i>lacI</i> , Km ^R , f1, ColE1	This study
pSV272.1-His ₆ MBP-BesB-opt	His ₆ MBP-BesB (T7) codon optimized, <i>lacI</i> , Km ^R , f1, ColE1	This study
pSV272.1-His ₆ MBP-BesC	His ₆ MBP-BesC (T7), <i>lacI</i> , Km ^R , f1, ColE1	This study
pETDUET-1	<i>lacI</i> , f1, Cb ^R , ColE1	Novagen
pETDUET-BesC.BesD	BesD (T7) and BesC (T7) from <i>P. fluorescens</i> , <i>lacI</i> , Cb ^R , ColE1	This study
pPra	BesD (T7) and BesC (T7) from <i>P. fluorescens</i> , MBP-sBesB (T7) from <i>S. sp.</i> NRRL S-1448, <i>lacI</i> , Cb ^R , ColE1	This study

A1.1B. Oligonucleotide sequences

Name	Sequence
LR-SCAT_0184-F	acacccccgcctcgttacccggcggtgacaacatcgatgatgtattccggggatccgtcgacctgcagttcg
LR-SCAT_0184-R	cgcggtgtcttctcccgctggcagggcgggacgcggttgatcatgtaggctggagctgcttcaagttcctata
LR-SCAT_2136-F	cgtttccgtaggttacggttccgtagccgtagccgttacgtactccgagggagctctatgtattccggggatccgtcgacctgc
LR-SCAT_2136-R	agttcg acggtcctcgtcgacccgcgccatcagcactgccgtcgttccccacgcccctcatcatgtaggctggagctgcttcaagttc ctata
LR-SCAT_4823-F	gccgcggcgggcccctgcgacatcgaacggtaggagcaccattccggggatccgtcgacc
LR-SCAT_4823-R	gcctgacctcgggtcgtgagccagggcggttgtaacaatctgtaggctggagctgcttc
LR-SCAT_p1525-F	aggctgtgtcgccgtgccggggccgggagactcgacggcaltccggggatccgtcgacc
LR-SCAT_p1525-R	taccggcgccggttcggaccggggacggtgtgacgcccgggtgtaggctggagctgcttc
K773-BesA-UP-F	gggatccactagatggaagcgctcgtcctcgttcatcc
K773-BesA-UP-R	cgcggtggcgccgctcgtggcgatgctgttggtccgc
K773-BesA-DOWN-F	acaaaagctggtaccggtacggcatcgacaaggaacacctga
K773-BesA-DOWN-R	cggaatatcaagcttgcccgcgcagacgttcaggc
K773-BesB-UP-F	cgggggatccactagtgcgggttgagcttgccgaggggtct
K773-BesB-UP-R	caccgcggtggcgccgctgacaacggtgaggaccagcgc
K773-BesB-DOWN-F	gtatcgataagcttacctgaagtcgaagctcagcaagagggaac
K773-BesB-DOWN-R	ggaatatcaagcttcacttgccgatgccacgagttcc
K773-BesC-UP-F	agcccggggatccactagtgcgcaaccaggggtacgacgac
K773-BesC-UP-R	tccaccgcggtggcgccgccctggcgagcgtacggcgg
K773-BesC-DOWN-F	agggaacaaaagctggtaccggtgcacaccgttccgccacc
K773-BesC-DOWN-R	aagcttgatattccggggatccgtctcgtgttccctcacctgtcgcg
K773-BesD-UP-F	tctagaactagtggaaacggcgccaattcacccgcaca
K773-BesD-UP-R	ggatccactagtggactgccgtccaggaagccg
K773-BesD-DOWN-F	atatcaagcttgatgaatctccgatgctcgctatgcg
K773-BesD-DOWN-R	gaattgggtaccggcaaggccgccttcgtggg
K773-BesE-UP-F	ttatccactagtctgcctccgatacctgaagtcgaagtctca
K773-BesE-UP-R	ggatccactagtctgaccagtcgagggaattcgaagacgc
K773-BesE-DOWN-F	atatcaagcttacgaagaacgtctgcccggtggc
K773-BesE-DOWN-R	gaattgggtaccggtggcgagttcgggcgcgctc
dcK-X-US-R	atgccgctcgccagtcgattggc
dcK-X-DS-F	cgtagccgggcaggataggtgaagta
dcK-BesA-US-F	ggcgactaccacaacgccgtcaac
dcK-BesA-DS-R	tcccttctacatcacggggatggcc
dcK-BesB-US-F	gtggtgtagcgggctgccc
dcK-BesB-DS-R	aagaacccggcaccggacg
dcK-BesC-US-F	gccagtactggggctgcccgtt
dcK-BesC-DS-R	cggaccctcaggaccagacacca
dcK-BesD-US-F	gcgttccacctgcggtagttggcc
dcK-BesD-DS-R	cgaccgtgacgcatagtggcga
dcK-BesE-US-F	ctcacttgccgatgccacgagtt
dcK-BesE-DS-R	aacatggcctgctcgtacgcg
dcK-0184-IN-F	tcatcgacccggtcgacgtactcggg
dcK-0184-IN-R	gtgacggcggcggagtcggg
dcK-2136-IN-F	atgaccacacagccggacgtgatcacc
dcK-2136-IN-R	tcatgccccttggcgacatcttcttacg
dcK-4823-IN-F	ttgaccgcatcgacccaccgc
dcK-4823-IN-R	acaatctcaggcggcgacgagctcctg
dcK-p1525-IN-F	atgacgattccaccgaggttcccctcctg
dcK-p1525-IN-R	gtcaccgcccgtgcggacgccttg
dcK-0184-OUT-F	cggcagggcgggacgcgtttg
dcK-0184-OUT-R	ctcgttcaccggcggtgacaacatcg
dcK-2136-OUT-F	ccgtagccgttacgtactccgagggacgtct
dcK-2136-OUT-R	catcagcactgccgtcgttccccacg
dcK-4823-OUT-F	ccctgcgacatcgaacggtaggagcacc
dcK-4823-OUT-R	cgggttcgtgagccaggcggttgtaacaatc
dcK-p1525-OUT-F	cccgtgactggcaggctgtgtcg
dcK-p1525-OUT-R	ccggttcggaccggggacggtgtg

dcK-BesA-IN-F	tcttcggttcgacgagccggc
dcK-BesA-IN-R	cgagggtctccggttcacgagc
dcK-BesB-IN-F	aagacctcgcgatcagctccgg
dcK-BesB-IN-R	ttcgaatccggcgacccggcg
dcK-BesC-IN-F	atgaccgacctgaacaccccgaa
dcK-BesC-IN-R	cttgccgatgccacgagttccc
dcK-BesD-IN-F	gtcatgttcaggatgatgcgggtggtg
dcK-BesD-IN-R	ccttcaccggttcggcatcgtcac
dcK-BesD-IN-F	acgacacaccaccacgcgacgt
dcK-BesD-IN-R	aggaccagacgaccagttcgccg
His ₁₀ -BesA F	catcacagcagcggccatatcgaaggctgcatatgactgatcccgggttcgacgaagatcctgatctacgc
His ₁₀ -BesA R	ttcctttcgggctttagtagcagccgatcctcgagtcaggcgctggggtacggcccg
MBP-BesB-SCAT-F	acctcgggatcgaggaaaacctgtatttcagggcagtcgctcgtatccccgacgtgg
MBP-BesB-SCAT-R	agtgcggccgcaagcttgcgacggagctcgaatttcagtcggggcgagggcc
MBP-BesB-S31-F	acctcgggatcgaggaaaacctgtatttcagggcagtcctatccccgacgtggcctc
MBP-BesB-S31-R	agtgcggccgcaagcttgcgacggagctcgaatttcacccggggcgagggc
MBP-BesB-LAVE-F	acctcgggatcgaggaaaacctgtatttcagggcagtcgctcgtatccccgacgtgg
MBP-BesB-LAVE-R	agtgcggccgcaagcttgcgacggagctcgaatttcagtcggggcgagggcc
MBP-BesB-S1448-F	acctcgggatcgaggaaaacctgtatttcagggcagtcacaccgagggatccggggag
MBP-BesB-S1448-R	agtgcggccgcaagcttgcgacggagctcgaatttcagagggggcgagggcc
MBP-BesB-CATE-F	acctcgggatcgaggaaaacctgtatttcagggcagtcgctcgtgccccgagg
MBP-BesB-CATE-R	agtgcggccgcaagcttgcgacggagctcgaatttcagtcgggacgcagggccc
MBP-BesC-F	caacctcgggatcgaggaaaacctgtatttcagggcagtcacgacctgaacaccccggaatccacc
MBP-BesC-R	gtgcggccgcaagcttgcgacggagctcgaatttcacttgccgatgccacgagttccgcac
His ₁₀ -BesD-F	cagcggccatgaaaacctgtatttcagggcagtcgggcagtaaccggcaggaactcaaagacgtgtg
His ₁₀ -BesD-F2	catcatcatcatcatcatcatcagcagcggccatgaaaacctgtatttcagggc
His ₁₀ -BesD-R	ttcctttcgggctttagtagcagccgatcctcgagtcagtccttgatggccgtgcccgcg
His ₁₀ -BesE-F	catcacagcagcggccatatcgaaggctgcatatgtccggcacgacacaccaccgc
His ₁₀ -BesE-R	ttcctttcgggctttagtagcagccgatcctcgagtcaggaccagacgaccagttcgccgcg
His ₆ -BesF-F	acagcagcggcctgtgtgcgcgcgccagccatgatgaagctctccaaagcgtctggg
His ₆ -BesF-R	ccgcaagcttgcgacggagctcgaatttcggatccctaccgacgtcccggagttgacgcct
His ₆ -γ-GT(e.c.)-F	acagcagcggcctgtgtgcgcgcgccagccatgatgataaaaccgacgttttac
His ₆ -γ-GT(e.c.)-R	ccgcaagcttgcgacggagctcgaatttcggatccttagtaccgcccggttaaatcatc
Strep-BesD-F	agctggagccaccgcagttcgaaaaaggtgcaggtatgggcagtaaccggcaggaactcaaagacgtgtg
Strep-BesD-F2	ataattttgttaactttaagaaggagatataccatggcaagctggagccaccgcagttcgaaaag
Strep-BesD-R	gtgcggccgcaagcttgcgacggagctcgaatttcagtccttgatggccgtgcccgcg
Strep-BesD G139D-F	ataattttgttaactttaagaaggagatataccatggcaagctggagccaccgcagttcgaaaag
Strep-BesD G139D-R	cgcgaagctgaagtcacccagtgccagccgtgggtgtcc
Strep-BesD G139D-F2	cggctggcactgggatgacttcagcttcgcgctgatctgggtgtcc
Strep-BesD G139D-R2	gtgcggccgcaagcttgcgacggagctcgaatttcagtccttgatggccgtgcccgcg
MBP-BesB-Opt-F	acctcgggatcgaggaaaacctgtatttcagggcagtcacactgaggccagtggggaa
MBP-BesB-Opt-R	tttgatctcgagtgccgcaagcttgcgacggagctcgaatttcacaaggagcgcagagcttgctc
petDUET-A-F	aataattttgttaactttaagaaggagatataccatgactgatcccgggttcgacgaa
petDUET-A-R	cgacctgcaggcgccgagctcgaatttcggatccgcgctggggtacggcccg
petDUET-E-F	tagtatattagttaagtataagaaggagatatacatgtccggcacgacacaccacca
petDUET-E-R	tttctttaccagactcgagggtaccggaccagacgaccagttcgccg
pPra-C-F	aataattttgttaactttaagaaggagatataccatgtccatcactcaagaacgtttc
pPra-C-R	gtcgacctgcaggcgccgagctcgaatttcggatcctcattggcgctctccagcgttag
pPra-D-F	tcttagtatattagttaagtataagaaggagatatacatgaactacgtacttgataggct
pPra-D-R	ttcgacagcgggttctttaccagactcgagggtaccttaggtcaagctctggctgcctc
pPra-B-F	cagcacgtcgaggcagccaagcttgacctaaaggtaccaagcttgccgcccataatgcttaa
pPra-B-R	caaatttcgcagcagcgggttctttaccagactcgagtcacaagggacgcagagcttgctca

Table A1.2. Strains, plasmids, oligonucleotides, and sequences used in Chapter 3. (A) Strains and plasmids used in this study. (B) Oligonucleotides used for plasmid construction and sequencing. (C) Sequences of synthetic genes.

A1.2A. Strains and plasmids

Strain	Description	Source
BL21 Star (DE3)	<i>F-ompT hsdSB (rB-, mB-) gal dcm rne131</i> (DE3)	Thermo-Fisher

Plasmid	Description	Source
pET16b-His ₁₀ -ScattleyaBesD	His ₁₀ -ScattleyaBesD (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -SlavenduligriseusBesD	His ₁₀ -SlavenduligriseusBesD (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -PfluorescensHalA	His ₁₀ -PfluorescensHalA (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -PorientalisHalA	His ₁₀ -PorientalisHalA (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -SwuyuanensisHalB	His ₁₀ -SwuyuanensisHalB (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -StoyocaensisHalB	His ₁₀ -StoyocaensisHalB (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -SviridosporusHalB	His ₁₀ -SviridosporusHalB (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -AawajinensisHalB	His ₁₀ -AawajinensisHalB (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -SgriseusHalB	His ₁₀ -SgriseusHalB (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -SafghaniensisHalB	His ₁₀ -SafghaniensisHalB (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -SiranensisHalB	His ₁₀ -SiranensisHalB (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -SprunicolorHalB	His ₁₀ -SprunicolorHalB (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -LanisaHalC	His ₁₀ -LanisaHalC(T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -PkilonensisHalD	His ₁₀ -PkilonensisHalD (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -PSHC52HalD	His ₁₀ -PSHC52HalD (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -PtrivialisHalD	His ₁₀ -PtrivialisHalD (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -Proot562HalE	His ₁₀ -Proot562HalE (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -PfulvaHalE	His ₁₀ -PfulvaHalE (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -SpristinaespiralisHalF	His ₁₀ -SpristinaespiralisHalF (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -MpelagiusHalG	His ₁₀ -MpelagiusHalG (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -PcorrugataHalH	His ₁₀ -PcorrugataHalH (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -PmediterraneaHalH	His ₁₀ -PmediterraneaHalH (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pSV272.1-His ₆ MBP-Proot562HalE	His ₆ MBP-Proot562HalE (T7), <i>lacI</i> , Km ^R , f1, ColE1	This study
pET16b-His ₁₀ -SIBesD R74A	His ₁₀ -SIBesD R74A (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -SIBesD W238A	His ₁₀ -SIBesD W238A (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -SIBesD W239A	His ₁₀ -SIBesD W239A(T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -SIBesD H134A	His ₁₀ -SIBesD H134A (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -SIBesD D140A	His ₁₀ -SIBesD D140A (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -SIBesD N219A	His ₁₀ -SIBesD N219A (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -SIBesD E120A	His ₁₀ -SIBesD E120A (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -SIBesD T221A	His ₁₀ -SIBesD T221A (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -RapL	His ₁₀ -RapL (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -OCD	His ₁₀ -OCD (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -LDC	His ₁₀ -LDC (T7), <i>lacI</i> , Cb ^R , ColE1	This study

pET16b-His ₁₀ -IlvE	His ₁₀ -IlvE (T7), <i>lacI</i> , Cb ^R , ColE1	This study
IVTT-PepC	PepC (T7), Cb ^R	This study
IVTT-PepA	PepA (T7), Cb ^R	This study
pET16b- His ₁₀ -PrescissionCutSite-IMDPH	His ₁₀ -PrescissionCutSite-IMDPH (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pRARE	tRNA ^{Ile} (AUA), tRNA ^{Ser} (AGG), tRNA ^{Ser} (AGA), tRNA ^{Asp} (CUA), tRNA ^{Gly} (CCC), tRNA ^{Gly} (GGA), Cm ^R	Novagen

A1.2B. Oligonucleotide sequences

Name	Sequence
His ₁₀ -ScattleyaBesD-F	agcagcgccatctagaagtgccttttcaggggcccgcatatgggcagtaaccggcaggaactcaaagacgtgtgc
His ₁₀ -ScattleyaBesD-R	ttccttcgggcttggtagcagccggatcctcagtccttgatggccgtgcccgcgacac
His ₁₀ -PfluorescensHalA-F	agcagcgccatctagaagtgccttttcaggggcccgcatatgaactacgtacttgatgaggctcgc
His ₁₀ -PfluorescensHalA-R	ttccttcgggcttggtagcagccggatcctcagtgtaggtcaagctcttggtgcct
His ₁₀ -PorientalisHalA-F	agcagcgccatctagaagtgccttttcaggggcccgcatatgaattacgcacttaacaatactattctgcagaaaca
His ₁₀ -PorientalisHalA-R	ttccttcgggcttggtagcagccggatcctcagtgcaagtcaggcttttagcagcctct
His ₁₀ -SiranensisHalB-F	agcagcgccatctagaagtgccttttcaggggcccgcatatgaccacgaatacgctacaaccg
His ₁₀ -SiranensisHalB-R	ttccttcgggcttggtagcagccggatcctcagtcaggaggagaacatcgcgctc
His ₁₀ -PkilonensisHalD-F	agcagcgccatctagaagtgccttttcaggggcccgcatatgagcgaacaaacgtccagcttg
His ₁₀ -PkilonensisHalD-R	ttccttcgggcttggtagcagccggatcctcagtcaggcttctcggtggtccg
His ₁₀ -PtrivialsHalD-F	agcagcgccatctagaagtgccttttcaggggcccgcatatgaaaactcagcaatccggcttg
His ₁₀ -PtrivialsHalD-R	ttccttcgggcttggtagcagccggatcctcagtcaggcttctgaagcttcggc
His ₁₀ -Proot562HalE-F	agcagcgccatctagaagtgccttttcaggggcccgcatatgggtgtgtaagtgcgatca
His ₁₀ -Proot562HalE-R	ttccttcgggcttggtagcagccggatcctcagtcagtcacggaagacaaaagtctcatgttc
His ₁₀ -PfulvaHalE-F	agcagcgccatctagaagtgccttttcaggggcccgcatatgtcagagctcgtctgttaaaca
His ₁₀ -PfulvaHalE-R	ttccttcgggcttggtagcagccggatcctcagtcagtcacggaagacaaaagtctcatgttc
His ₁₀ -PmediterraneaHalH-F	agcagcgccatctagaagtgccttttcaggggcccgcatatgagtaatcgaaaagcagatccgctctt
His ₁₀ -PmediterraneaHalH-R	ttccttcgggcttggtagcagccggatcctcagtgatccacagggcatcgcttcg
His ₆ MBP-Proot562HalE-F	caacctcgggatcgaggaaaacctgtatttcagggcacgtgggtgtgtaagtgcgatcag
His ₆ MBP-Proot562HalE-R	gcggccgcaagctgtgcacggagctcgaattctcatcagtcacggaagacaaaagtctcatgttc
His ₁₀ -SIBesD R74A-F	cctccagaccaccggatacaccgcgcgtcgtatccgtggtgcagagcgagacgatcg
His ₁₀ -SIBesD R74A-R	tctcgtctgcaccacggacatcgacgcgcgggtgtatccggtggtctggaggcgg
His ₁₀ -SIBesD W238A-F	gggacctgcggcgcgagctcaagggcgacgaccgcgcgtgggaggacggcgcgctg
His ₁₀ -SIBesD W238A-R	agcggctgcggcgcagctcggcgtcctccacgcgcggtcgtcgcccttgagc
His ₁₀ -SIBesD W239A-F	cggcgcgagctcaagggcgacgaccgctgggcggaggacggcgcgctgccc
His ₁₀ -SIBesD W239A-R	gtgcagcggctcggcgcagctcggcgtcctccgcccagcggctcgtcgccc
His ₁₀ -SIBesD H134A-F	caccggcaggaacggagcggggacacggcgggctggcactggggcgatttcagcttcgc
His ₁₀ -SIBesD H134A-R	ggcgaagctgaaatcgccccagtgccagcccgctgtccccgctcgttctctcgccg
His ₁₀ -SIBesD D140A-F	agcggggacacgcacggctggcactggggcgcgttcagcttcgcccctcatctggg
His ₁₀ -SIBesD D140A-R	ggcctggagcaccagatgagggcgaagctgaacgcgccccagtgccagccg
His ₁₀ -SIBesD N219A-F	tcccgtcgcgcgaggacaccaccgcatcatcctggcgatgacctgggcgggcga
His ₁₀ -SIBesD N219A-R	cggcgcaggtcccgtcgcggcccgccaggtcatcgccaggatgatgcgggtggtgtcc
His ₁₀ -SIBesD E120A-F	cggccttggggcaggggtgcagttgcggaattctcatcaccggcaggaacggagcggg
His ₁₀ -SIBesD E120A-R	ccgttctgcgggtgatgaggaattccgcgtcggccttggggcaggggtgcagtt

His ₁₀ -SiBesD T221A-F	gcgaggacaccacccgcacatcctgaacatggcgtgggcgggcgagcgg
His ₁₀ -SiBesD T221A-R	gagctcgcgccgaggtcccgcctcgcccgccacgccatgttcaggatgatgcgggtggt
His ₁₀ -SiBesD-F	gcgccatctagaagtgtcttttcaggggcccgcatatgagcagcaatgggcaggaaagcaccgtcgtca
His ₁₀ -SiBesD-R	ccttcgggctttgttagcagccgatcctcgagtcagtcacatccagtcagcggctgccggc
pET16-anneal-R	cgtatcacgaggccctttcgtcttc
His ₁₀ -LDC-F	ggccatctagaagtgccttttcaggggcccgcatatgaacatcattgccattatgggaccg
His ₁₀ -LDC-R	cttccttcgggctttgttagcagccgatccttatcccgccatttttaggactcgtacg
His ₁₀ -IlvE-F	cggccatctagaagtgtcttttcaggggcccgcatatgaccacgaagaagctgattacatttgg
His ₁₀ -IlvE-R	cttccttcgggctttgttagcagccgatccttattgattaactgatctaaccagccccattatcttc
His ₁₀ -OCD-F	cggccatctagaagtgtcttttcaggggcccgcatatgacgtatttcattgatgtccaacctatgcgg
His ₁₀ -OCD-R	cttccttcgggctttgttagcagccgatcctcatcaggcaaccgctcggatacgc
PepC-F	atggaaaccggtctaaaaacatgctgtaatgaggatcccgggaattctcgagta
PepC-R	cagcatgttttagaacgggtttccatagtatatctccttctaaagttaacaaaatttttagaccctatagtg
PepA-F	atgtctgaacgttctccaacatgaataatgaggatcccgggaattctcgagta
PepA-R	tttcatgttgaagaacgttcagacatatgtatatctccttctaaagttaacaaaatttttagaccctatagtg
IMDPH-F	tcatcatcatcatcatcacagcagcgccatctagaagtgtcttttcaggggcccgcatatgtactgatcgctaaa
IMDPH-R	gaagctctgac
	cagcttccttcgggctttgttagcagccgatcctcaggagcccagacggtagttcggg

A1.2C. Synthetic gene sequences

WP_030791981 (*Streptomyces lavenduligriseus* BesD)

AGCAGCGGCCATCTAGAAGTGCTTTTTTCAGGGCCCGCATATGAGCAGCAATGGGCGAGGAAAGCACCGTCGTCAATCC
TCTGGAGCAGGGCGCACTGCGCCGTATGGCGCATACTACCACCGGTACGGCATCGCCACCGTCACCGATCTGATTC
GGGAAGACGTCCGCAAAAACGTGCGGGCGGAGGCGGACCGGCTGCTGGAGAAATACGCCGAGCGGCGTGATCTCCGC
CTCCAGACCACCGGATACACCCGCGCTCGATGTCCGTGGTGACAGCGAGACGATCGCGGGCAACAGCGAGCTGGT
CACCTCGATCTACGCGAATCCGGAAGTCTCGGCGCGCTGGAGCGCATAGCCGGCGAGAACTGCACCCCTGCCCA
AGGCCGACGAGGAATTCCTCATCACCCGGCAGGAACGGAGCGGGACACGCACGGCTGGCACTGGGGCGATTTTCAGC
TTCGCCCTCATCTGGGTGCTCCAGGCCCGCCCATCGACATCGGCGGCATGCTCCAGTGCGTACCGCACACGGAGTG
GGACAAGTCCGATCCGCGGATCCACAGTATCTCGTCGACAATCCCATCCACACGTACCACTTCCAGTCGGGCGACG
TGTATTTTCTGCGCACCGACACACGCTCCACCGCACGGTCCCGCTGCGCGAGGACACACCCGCATCATCTGAAC
ATGACCTGGGCGGGCGAGCGGGACCTGCGGCGCGAGCTCAAGGGCGACGACCGCTGGTGGGAGGACGCCGACGTGCC
GGCAGCCGCTGCACTGGATGACTGACTCGAGGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGA

WP_016975823 (*Pseudomonas fluorescens* HalA)

ATGAACTACGTACTTGATGAGGCTCGCCTGCAAGAGCATCACGCTACTAAGTTTCTGAATCAAGCGTATTTGCACT
GCGTCATGAGTTTGCCCGCAACGGGTTTATAAAAGTGCGCAATATTGTGACGACGACCTGCGCGAAAAGATAACGC
GTGAGGTCAACAGCCTGATCGACCGCCAGCTTGAACGTGCGGACTTGACCTGGCAACCACCGACAATACGCCACGC
TACATGAGTGTAGTGCGCAGCGAGTTTATTGCTGAAAACAGCACGCTGATCAACACGCTGTCCAAGTCCAAAGGACT
GCTGGAAACCTTGTCCCAGATCGCGGGAAGTCAACTTATCGCGTCTGTTTCCAAGGATGAGGAATATTTGATCACCA
AGCAGGAGCGCAAAGGTGATACCCACGGTTGGCACTGGGGGGATTACAGCTTCGCACTGATCTGGATTATCGAAACA
CCGTCGATAGCCAAGGGCGGGATGTTGCAATGCGTACCTCATACGTATGGGATAAATCCAACCCGAGAATTCATGA
GCTGCTGTGCAGTAACCCGATTGCGACATACGGCTTCGTGACGGGGGATATTTACTTCCTGCGCACCGACACCC
TGCACCGCACTATCCCGCTTAACGAAGATGCGACGCGGATCATCCTCAACATGACGTGGGCTGCTGAAAAAGACCTC
TCTCGCAACCTGCACGGCAATGATCGCTGGTGGGAGGACCAGCACGTGAGGCAGCCAAGAGCTTGACCTAA

SDN46247 (*Streptomyces wuyuanensis* HalB)

AGCAGCGGCCATCTAGAAGTGCTTTTTTCAGGGCCCGCATATGACCGAGAACAGCATCACCGCCGCGAATGTGGAAGA
GCTGATCGCGAAGAACATCGCGGAGCGGTTGCGCGACGACCACGAGGTGCTCGGCCTCTCCAGCACTTCCGCCGCG
AGGGCTACGTCAAGCTCCCCGGCCTGGTCTCCCCGGAGGTCTTCGACGCGGTGCGCGGGAGACCCACCAGCTGATC

GACACCCACCAGAAGCGCATCGACATCCGTCTCAAGGAGACGGGGGACTCCCCGCGCTACATGTCCACCGTCGGTCA
GAAGGCGATCGCCACCGACGGCTCGCTGATCCCTGCCGTCTACGAGTCCACCGCGCTCAAGGGCTTCCTTTCCCGCC
TCGCCAAGGAGGAGGTTCATGGGCTGCCCCTGGGACGAGGAGAAGTACATCATCACCCGCCAGCACCAGAAGGGCGAC
ACCCACGGCTGGCACTGGGGCGACTTCAGCTTCACCGTCATCTGGCTGATCGAGGCCCCGTGCTGGAGTACGGCGG
CATGCTGCAGTGCATCCCGCACACCGACTGGAACAAGGACGACCCGCGCGTTCGAGGACTACCTGCAGAAGCACCCGA
TCCGCAGCTACGGCCACGCCAAGGGCGACCTCTACCTGCTGCGTTTCGGACACCACCTGCACCGCACGGTCCCGCTG
AACGCCGACAGGACCCGCATCATCCTCAACACCTGCTGGGCCAGCCGCGCGGACCAGCAGAAGGCGACCACCCACGA
GACCATGAACGCGATGTTTCGACTGATGAGGATCCGGCTGCTAACAAAGCCCCGAAAGGAAGCTG

WP_037940874 (*Streptomyces toyocaensis* HalB)

AGCAGCGGCCATCTAGAAGTGCTTTTTTCAGGGCCCGCATATGACGCAGAACGCGATCACCGCAGAGAACGTCTGAAGA
GCTCATCGCCAAGAACATCACGGAGCGGTTTCGACAACGACCACGAGGTCCTGGGGCTCTCTCAGCACTTCGCGCCGCG
AGGGCTACGTCAAGCTGCCCCGACTGGTCTCCGAGGAGGTCTTCGAGGCGGTTCGCGCCGAGACCCACCAGCTGATC
GACACGCACCAGAAGCGGATCGACATCCGTCTCAAGGAGACCGGGCGACTCGCCGCGCTTCATGTCCACCGTCGGCCA
GCGGGCGATCGCCACCGACGGGTCGCTGATCCCGGCCGTCTACGAGTCGCGGGCGCTCAAGGGCTTCCTGTCCCGCC
TCGCCAAGGAGGAGGTTCATGGACTGCCCCTGGGACGAGGAGAAGTACATCATCACCCGCCAGCACCAGAAGGGCGAC
ACCCACGGCTGGCACTGGGGCGACTTCAGCTTCACGGTCATCTGGCTGATCGAGGCGCCCTCCCTGGAGTACGGCGG
CATGCTCCAGTGCCTCCCGCACACCGACTGGAACAAGGACGACCCGCGCGTTCGAGGAGTACCTGCAGCAGCACCCGA
TCCGCTCGTACGGCCACGCGCGGGGCGATCTGTACTTCTGCGCTCGGACACCACGCTGCACCGCACGGTTCGCTG
AACGCCGACAAGACCCGCATCATCCTCAACACCTGCTGGGCGAGCCGCGACGGACAGCCAGAAGGCGACCACCCACGA
GACGATGAACGCGATGTTTCGACTGATGAGGATCCGGCTGCTAACAAAGCCCCGAAAGGAAGCTGAG

WP_004985557 (*Streptomyces viridosporus* HalB)

AGCAGCGGCCATCTAGAAGTGCTTTTTTCAGGGCCCGCATATGACGCAGAACGCGATCACCGCAGAGAACGTCTGAAGA
GCTCATCGCCAAGAACCTCACGGAGCGGTTTCGACGACGACCACGAGGTCCTCAAGCTCTCCAGCACTTCGCGCCGCG
AGGGGTACGTCAAGCTGCCCGGCTGGTCTCCGCGAGGTCTTCGAGGCGGTTCGCGCCGAGACCCACCAGCTGCTC
GACACGCACCAGAAGCGCATCGACATCCGTCTGAAGGAGACCGGGCGACTCGCCGCGCTTCATGTCCACCGTCGGGCA
GAAGGCGATCGCCGCCGACGGCTCGCTGATCCCGGCCGTCTACGAGTCGCGGGCGCTCAAGGGCTTCCTCTCCCGCC
TCGCGAAGGAGGAGGTTCGTTGACTGCCCCTGGGACGAGGAGAAGTACATCATCACCCGCCAGCACCAGAAGGGCGAC
ACCCACGGCTGGCACTGGGGAGACTTCAGCTTCACGGTCATCTGGCTGATCGAGGCGCCCGCCCTGGAGTACGGCGG
CATGCTCCAGTGCCTTCGCGCACACCGACTGGAACAAGGACGACCCGCGCGTTCGAGGAGTACCTGCAGCAGCACCCGA
TCCGCTCCTACGGCCACGCCCGGGGCGACCTGTACTTCTGCGCTCGGACACCACGCTGCACCGCACGGTCCCGCTG
AACGCCGACAAGACCCGCATCATCCTCAACACCTGCTGGGCGAGCCGCGCGGACAGCCAGAAGGCGACCACCCACGA
GACGATGAACGCGATGTTTCGACTGATGAGGATCCGGCTGCTAACAAAGCCCCGAAAGGAAGC

WP_067690966 (*Actinoplanes awajinensis* HalB)

AGCAGCGGCCATCTAGAAGTGCTTTTTTCAGGGCCCGCATATGACTGTGCGAGATTACGTCGTGAATGTCTGAAGAGCT
TATCGCGAAGCACCTCACCGAGGAGTTCGCCGACGACAACAAGGTTTCGCGCGCTGTCCAGCACTTTTCGGCGTGAGG
GCTACGTAAAGCTGTCCGGCCTGGTCTCGCCGAGATCTTCGCCGCGGTACCAAGGAAACGCACGACATCCTGGAC
CTGCACGCCAAGCGGATCGACATCCACCTCAAGGAGACCGGCAACACGCCGCGCTTCATGAGCACCGTCAGCCAGGC
CGCGATCGCCGCCGACGCGAAGCTGATCCCCGCCGTGTACGAGTCGCCGCGCTGAAGGGCTTCCTGTCCCGCATCG
CCGCCGAGGAGGTTCATGGACTGCCCCTGGGAGGGCGAGAAGTACATCATCATCCGCCAGGAGCGGGTCGGTGACACC
CACGGCTGGCACTGGGGCGACTTCCCCTTCACCGTCATCTGGATCATCGAGGCGCCGGCCACCGAGTTCGGCGGCAT
GCTCCAGACCGTGCCGCACACCGACTGGAACAAGGAAGACCCGCAGGTTCAGCGAGTACCTGATCAACAACGTGATCA
AGACCTGGCCGCACTCCACCGGCGAGCTGTACTTCTGCGGTTCGGACACCACGCTGCACCGCACGATCCCGCTGAAC
GCGGACCGCACCCGGATCATCCTCAACACCTGCTGGGGCAACGCAGCGACCGCAGGCCGCCCGCAGAGCCACGAGAC
CATGGAAGCGATGTTTCGACTGATGAGGATCCGGCTGCTAACAAAGCCCCGAAAGGAAGCTGA

WP_003966730 (*Streptomyces griseus* HalB)

AGCAGCGGCCATCTAGAAGTGCTTTTTTCAGGGCCCGCATATGTCAATTCGAGGAAATCACCCCGGCCGGTATCGAGAA
GGACCTCGCGAACCCTTCGCGGAGGCTTTTCGACGACGCCACGCTGGTCGAGCTCTCGCAGAGCTTCGCGCGCAACG
GCTACGTGAAGCTGTCAACTTCGTCTCCGAAGACCTCTTCAAGACGGTCACCAAGGAGTGCCACCGCCTCCTCGAC
CTGCACCGAGCAGCGCATCGACATCCACCTCAAGGAGACCAGCAACTCCCCGCGCTTCATGAGCACGGTGAGCCAGAA
GGCCATCGAGCAGGACAGCGAGCTCGTCCGACCGTCTACAACCTCCGAGGCGATGATGGGCGCGCTGTCCCGCATCG
CCGCCGAGCCGGTCTGCGGTGCCCCTGGGACGAGGAGAAGTACGTCTCATCCGCCAGCACAAAGCAGGGCGACACG

CACGGCTGGCACTGGGGCGACTTCAGCTTCACCGTCATCTGGATCATCGAGGCGCCGGAAGCGGAGTTCGGCGGCCA
GCTCCAGGCCGTGCCGCACACGGACTGGAACAAGGAGGACCCGCGGGTCCACGAGTACCTCGCCAAGCACCAGATCA
GCACCTACCCGCACGCCACCGGCGAGATGTACTTCCTGCGGTGCGACACGACCCTGCACCGCACGATTCCGCTGAAC
GCCGACCGGACGCGCATCATCTCAACACCTGCTGGGGCAGCGAGAAGGACCAGCAGAAGACCGCGTCCCACGAGAC
CATGCAGGCGATGTTTACGCTGATGAGGATCCGGCTGCTAACAAAGCCCCGAAAGGAAGC

WP_020275004 (*Streptomyces afghaniensis* HalB)

AGCAGCGGCCATCTAGAAGTGCTTTTTTTCAGGGCCCGCATATGTCATTTCGACGAGATTGCCCCGGCGCAGATCGAGAA
GGAGCTGGCGAACCCTTCGCGGGCGCTTTTGACGACGCCGCGCTGCGTGAGCTGTTCGAGAGCTTCCGGCGCAACG
GCTACGTGAAACTGTGCAACTTCGTCTCGAAGGACCTCTTCCAGACCATCACGAAGGAGTGCCACAGGCTGCTGGAC
CTGCACCAGCAGCGCATCGACATCCATCTGAAGGAGACCAGCAACTCGCCGCGCTTCATGAGCACGGTGAGCCAGAA
GGCCATCGAGAACGACAGCGAACTCGTCCGGGGCGTCTACAACCTCCGAGGCGATGATGGGCACTCTGTCCCGCATCG
CCGGCCAGCCGGTCTGCCCCTGCCCCTGGGACGAGGAGAAGTACGTCTCATCCGCCAGCACAAAGCAGGGTGACACG
CACGGCTGGCACTGGGGCGACTTCAGCTTCACCGTCATCTGGATCATCGAGGCGCCGGGCCCCGGAGTTCGGCGGGCA
GCTCCAGGCCGTTCGCGACACGGACTGGGACAAGGACGACCCGCGGGTCCACGAGTACCTCGCCAAGCACCAGATCA
GCACGTACCCGCACGCCACCGGCGAGATGTACTTCCTGCGGTGCGACACCACCCTGCACCGCACGATTCCGCTGAAC
GCCGACAAGACGCGGATCATCTCAACACCTGCTGGGGCAGCGAGCAGGACCAGTTGAAGAAGGCGACCCACGAAAC
CATGCAGGCGATGTTCAACTGATGAGGATCCGGCTGCTAACAAAGCCCCGAAAGGAAGC

WP_044582715 (*Streptomyces iranensis* HalB)

ATGACCACGAATACGCCTACAACCGGTGTCCGTACGCTCGACCAGGACGCATTTCGACCGTCATCTCAAGGAATTCTC
CGAGAACGCCCCCATCCTCGAGTACTCGCAAGCGTTCGCGCGGGAAGGTTACGTCAAGCTCACCGGTCTGGTCTCGG
CGGATCTGTTTCGCCGAGGTGCTCAGCGAAGTGGGCGACCTGCTGGACCAGCACGCCCAGCGTATCGACATCGAACTC
AAAGAGACCGGGAATTTCGCCGCGCAGGATGCACACGGTCAGCGCGGCCGACATCGCCCAGGACTCCCGGCTCATTTT
CGCGATTTACGACTCCGCGAGCATCCGCGACACGCTGGGCCGGATCGCCCAGGCCGACGTGCGGCCCTGCCCTGGG
AGGGCGAGAAGTACGTATCATATCCGCCAGGACACGCGGGCGACACCCATGGATGGCACTGGGGGAGACTTCAGCTAC
ACACTGATCTGGATCATCTCAGGCGCCCCGGAGGTGGGCGGGATGCTGCAGTGCCTGCCCACACCACTGGGA
CAAGCAGAACCCGCGGGTGCACGAGTACCTCCAGAACCACCCATCCGCACCTACGCCAACGCCACCGGCGACCTGT
ACTTCTGCGTTTCGACACACCCTCCACCGGACGATCCCGCTCAGCAGGCCGGCCACCCGCATCATCTCAACACC
TGCTGGGCCAGCACCGCCGACGTGCAACGTCCACCACACACGAAACCATGGACGCGATGTTCTCTCTCTGA

WP_019057205 (*Streptomyces prunicolor* HalB)

AGCAGCGGCCATCTAGAAGTGCTTTTTTTCAGGGCCCGCATATGAGCTCCACGCAGGTACCCCCGAAAACATCGATT
AATTCTCGCCGAGCACATCTCTGAGGTGTTTCGACGATGAACGCTTCAGGCGAGACTCGCAGAAATTCAGGCGCGAAG
GTTACATAAAGGTCCGCGAACTGGTTCCGGACGCCGTCTTCGAAGGGCTGCAAAAAGAAGTACGCGACATCATCGCG
CAGCACGAGAAACGCATCGACGTCCATCTGCCGAGACAGCAACTCGCCCCGCTACATGAGCACCGTGAGCGCCTC
CGCCATCATCAAGGACGGCAGCCTCATCCCGGCGGTCTACAACCTCTCGGCGCTCAGGGACGCGTTGTCCAGGCTCG
CGGCCGAGGCCGTATCGACTGTTCTGGGACGGCGAGAAATACGCGATCCTGCACCAGCACAGAAGGGCGACACC
CACGGCTGGCACTGGGGCGACTTCAGCTACGCCATCGTGTGGATCATCGAGGCCCGGACCTGAAGTACGGCGGGGC
CCTGCAGTGCATCCCGCACACGGACTGGGACAAGTCCGATCCGCGGGTGATGGACTACATCTCGAAAACCCATCC
GCACCTACGGAACGTCACCAGGGACCTGTACTTCTTCGACGGACACGACCCTGCACCGGACGATTCCCTTGTCC
GAGGACAAGACACGCATCATCTCAACACCAGTTTCGCCGCTGAACGCGACCTGAAGAAGAAGCTACCCACGAGAC
GATGGGCGCCATGTTTCGACTGATGAGGATCCGGCTGCTAACAAAGCCCCGAAAGGAAGCTG

WP_019233318 (*Legionella anisa* HalC)

AGCAGCGGCCATCTAGAAGTGCTTTTTTTCAGGGCCCGCATATGAATACATTAAATGACACTATTTCTTCTTATAGTAA
TGCTGCATTTACGGCGAAAAAATTCGAAAGTTGCGTGAAGAATTTTCTCGTCATGGCATAGTCAAAATTCCTGATA
TTGTCAGTGAACAACTCAAACAAAATATTACAATGAAATTTACGGTTTACTCGAAAATATGCCGAACGTGCGGAT
CTTCTTTTAAAACTACTGACAATACTCCGCGATACTTAAGTGTGGTGCGTAGTGAATTAATCGAAGAGCAGAGCAA
GATGATTAGGGAAGTTAGAGAATCAGATGCACTATTGCAATTTTTAGAGAAAATTACCCAAGAGAAATTATTAACCC
ATGTCAAAGAAGATGAAAAATTTGTGATTACCAAGCAAGAGTTTCAGCGGTGATACTCATGGTTGGCACTGGGGAGAT
TATAGTTTTGCTTTGATCTGGATCGTCGAAATGCCACCAGTTGATTGTGGTGGAATGTTACAATGTGTCCCTCATAC
AAGTTGGGATAAAGAAGCGCCTGACTTATATCAATATTTTTGTGATAACCCCATTTTACTCATTCATTTGTTCCCG
GCGATGTGTACTTATTGAAAGCCGATACGACCCTACATCGAACAGTACCCTTAAGATATGATGCAACTCGCATTATG

CTCAATATGACCTGGGGCAGCATACGCGATCAGGATAACCCCTTTCTATCCGTTGATGATCGATGGTGGGATAATTT
AAATGCACAGGCTGCACTCCATGTTTCACTGACGAGGATCCGGCTGCTAACAAAGCCCCGAAAGGAA

WP_046063366 (*Pseudomonas kilonensis* HalD)

ATGAGCGAACAAACGTCCAGCTTGGTTATTGAAGTAATGGAGCAGCAACTGGCCAAGCATTTCAGGCTATTTTGCA
AGATGAAAACCGGATGAAGCAGATTTCGTAATGAATTTGCGCGAGACGGTTATTTCAACTTCAAGAACTTTTCGTTTC
TACCGAAAAGGATCTTGAAAAATGTCCACGCCGAAGTCCATGCGTTATTGGATGAATATTCGGTGCGTCGCGATGTT
ACTGTTCCGTCCACAGGCAATACCTACCGCAAGATGTACAACGTCAACCAACCGGAGATCGCCGAAGGCGGGACATT
CATCCCTGCGCTCTACCAATCCGAGTCGCTGCGTAAGTTTCTGGGCAACATCGCTGGCGATGATCTGGCGTCTTGCT
GGGAGCAGGAGCAATACCTGGTCACCAAGCTGAGCCACCCGGGCGATACCCACGGCTGGCATTGGGGGGATTACCCG
TACACGATGATCTGGATCATCGAGGCGCCCCGAGGACCCGGCGATTGGCGGGCTGCTCCAGTGCGTGCCACACAGCGA
ATGGGACAAGCAGAACCCGCGAGATCTGGCAGTACATCCTCAATAACCCGATCAAGTCCTACCACCATCTCAAGGGTG
ATGTGTATTTTCTCAAGTCCGACACCACGTTGCACCACGTGCGTGGCGATTTCAGCAGGAAACCACTCGGATAATTCTC
AACACATGCTGGGCCAGCGCCCATGACCGGCGAACCGATGTGCGCCACGAAAGCATCGAAGTGATCTGGGACACCAA
GGCCCGGACCACCGAAGAAGCCTGA

WP_041021480 (*Pseudomonas* sp. *SHC52* HalD)

AGCAGCGGCCATCTAGAAGTGCTTTTTTCAGGGCCCGCATATGAGCGAACAAACATCTAGCTTGGTCATTGATGTAAT
GGAGCAGCAACTCTCCAAGCACTTCCAGGCTATTTTGCAAGATGAAAACCGTATGAAGCAGATTTCGTAATGAATTTTC
GCCGCGATGGGTATTTCAACTTCAAAAACCTTTTCGTTTCTACCGAAGCGAATCCTGGAGAATGTTTCATACCGAAGTG
CATGCATTGTTGGATGAATATTCGGTACGCCGTGATGTCACTGTTCCATCGACCGGTAATACTTATCGCAAGATGTA
TAACGTTAATCAGCCGGAATCGCCGAAGGTGGGACTTTTATTCCGGCGCTTTACAAGTCCGAGTCGCTGCGCAATT
TCCTTGGCAACATCGCAGGCGACGATCTGGCGTCTGCTGGGAGCAGGAGCAATACCTGATCACAAACTGAGCCAT
CCAGGCGACACCCACGGCTGGCACTGGGGTGACTACCCCTACACCATGATCTGGATCATCGAGGCCCCGTGAGGACCC
CGCGATCGGCGGCGTACTGCAATGCGTACCGCACAGCGAGTGGGACAAGGAGAACCCACAGATCTGGCAATACATCC
TGAACAATCCGATCAAGTCTTATCACCACCTCAAGGGTGACGTGTATTTTCTGAAGTCCGATACCACCTTGACCCAC
GTGGTGCCGATCCAAGTATCGAAGTATCTGAGGACACCAAGCCAGGACCACCGAAGAGGCCTGATGAGGATCCGGCT
GCTAACAAAGCCCCGAAAGGAAGCTGAG

WP_053557074 (*Streptomyces* sp. *pristinaespiralis* HalF)

AGCAGCGGCCATCTAGAAGTGCTTTTTTCAGGGCCCGCATATGCTCATGACGATCGAAGCCGTATCCCCCGCCATGAC
CATGCGTGAGGCGCTGTCGCGCTCCTCGCCCCAGGAAACGTCCCCGGCCCGCACGTCGACGCCGAACAGGTGCCCC
GTATCGAGCCGATGATCGCACGGCACCGGCTCACCTCGACGAGAAGCTGACGGCCGCCAAGCGGCAGTACTCGGCC
GAGCACTACGCAAGAACCACGGCGCTGCTGCCGGCCGAGCTGATCTCCGAAGTGAAGCAGAGGCGCGAAGACTCGC
GGACCTCAACACCCGGCGGATCGACATCACCGTCCCCGTACCGGTGACACGCCCGCAGGCTGGGCAGCCTGCCCA
ACTCGCTGTTCCAGGAGCAGAGCAAGCTGTTACGGCGCTCTACAACCTCCGTGGAGTTCGGCTCCGTGCTCTCCACC
GTGGTCGGTACGCCGCTCCTCGACGCCTCCTTCGAGGACGAGAAGATCACCGGCACCCACCAGAGCCGCGCGGGCGA
CACGCACGGCTGGCACTGGGGCGACACGAGTTTCGCTTCATCTTCATCGCCGAGGCTCCGTGATCGAGCACGGCG
GCATGCTCCAGGCCGTGCCGCACACGGTCTGGGACAAGAAGAACCCGGCCGTGCACGAGAACCTCGCCGGCGGCACC
ATCCGCACCTACCACCACGAGGCCGGCGAGTCGTACGTCTTCAAGACGGACACCACCCTGCACCGCACCGTCCCGAT
CCAGACCGAGGGCGTGAGCGGATCATCGTGAACCTGACCCTCGCCAGCCTCAAGGACTGGCTGGAGGAGAAGTCCT
ACGAGACGACCAACGCCGTCTACACGGAATGATGAGGATCCGGCTGCTAACAAAGCCCCGAAAGGAAGCT

WP_036209184 (*Marinobacter pelagius* HalG)

AGCAGCGGCCATCTAGAAGTGCTTTTTTCAGGGCCCGCATATGGACAACCTGGAAGGTATTTTGAGCGCACACCATCG
AGATACATTCACTCAAGAGAAGCTGAAACGCGTCTCAGAGGATTTCTCACGAAATGGCTACATCAAGTTTCGACATTG
GTGAAGTGGTCCCTAAGCAAATTTTGAAGTCAATCATCGTAGAGGCGAACAACCTGCTGGAACAGCACAGTATTGCA
CGCGATGTAAAAGTCGCTTCAACAGGTTATAGCAAGCGGTATATGAGTAATGTCACTGCCAAGGCATCCGGGATCA
TGGAGAGGTATCGAATCCTTATATAAATCCCCGCTCTGCGAGACATCCTTGGGAAGCTTGCCGGCGATGAATTGG
CGAGCTGCTGGGAAGAAGAACAACCTACCTGATCAATAAGCTTGAACACGAAGGCGATACGCACGGATGGCACTGGGGC
GACTACCCATACACCGTATCTGGATTCTGGACGCGCCAGCTATTGAACATGGCGGGATTCTACAGTGATCCCGCA
CACGACTTGGGACAAGAACAACCCACAGGTTGAAAATTATATCATTGACCACCCTATTTCGGTCTTACTACCACACCA
CCGGCCAAGCCTACTTTCTGAAATCGGACACAACGCTGCACAGGGTCACCCCGATCAAAGGAGAGGCTACCCGGATC

ATCCTTAACACATGCTGGGCGAGTGCGAACGACCGGCGTGAAAATGTCGAGCATGAAACCATGGTAGCCGCTTTTCAT
TTAACGAGGATCCGGCTGCTAACAAAGCCCCGAAAGGAA

WP_024777807 (*Pseudomonas corrugata* HalH)

AGCAGCGGCCATCTAGAAGTGCTTTTTTCAGGGCCCCGCATATGAGTAATCGAATAGCAGGTCTGGACTTGGCGGATGT
GGAACAGCAACTTTTCCAAGCACTTCCGGGATTTTCTGGACAATGAAATCCGGATGAAACAGGTCCGTAACGCGTTTC
ATCAAGATGGCTATTTTCACTTCAAGGACTTTTCATTTCTACCTGGCGAGATGTGGGAAGAGCTTCATACCGAAGTG
CATGGGTTGCTCGATGAGTATTCGGCAGGACACGAGGCGACGGTCCTGTTTATCCATGATCCCCACCGCAAGGTTCA
TATGATCCCTCAAGCGCAGATCGCCGAGGGCGCAGCCGTGGTTCCGGCGCTCTATCACTCAGAGTCGCTACGCCACT
TCCTCAGCTGCATCGCAGACGAAGAACTGGCGTCTTGCCCGGACAGCGGACAGTACCTGATCAGCAGGCTGAATCAT
CCAGATGATACCCATGATTGGCATTGGGGTGATTACCCTTACACCCTGGTCTGGATCATTGAGGCTCCTGAAGGCCC
GGAGGGCGGCGATGTATTGCAATGCGTGCCCTACAGTAATGGGAACAAGGATGACTTGACAGTCTGGGAATACAGCC
TGGGCAACCCCGGCAAGGCCTGTCACCACCTCAGGGGTGATGTGTATTTCTCAAGTCCGGTACCACGCTGCATCAT
CTGGTGCCAATTACAGCGAAACGCCACGCGCATCATTCTCAGTACCAGTTGGCTGGTGGCTGCGGCGGGCGAAGCGAT
GTTCTGTGGATAATGAGGATCCGGCTGCTAACAAAGCCCCGAAAGGAAGCTGA

WP_047703910 (*Pseudomonas mediterranea* HalH)

ATGAGTAATCGAAAAGCAGATCCGCTCTTGGTAAATGTGGCGCAGCAACTTTCCAGGCACTTCCAGGATTTTCTGAA
CAATGAAATCCGGATGAAGCAGGTCCGTAACGCGTTTCATCGGGATGGCTATTTCCACTTCAAGGATTTTTTCATTTT
TACCTGGCGAGATGTGGGAAGAGCTTCATACCGAAGTGATGGGCTGCTCAATGAATATTCAGTGGGTACGAAGCG
ACCGTCCTGTTTATCCATGATCCCCACCGCAAAATCCATGTGATCCCGCAAGCGCAGATCGCCGAAGGCGGAGCCGT
GGTTCCGGCGCTTTATCATTCCGAGTCGCTACGCCGCTTCCTCGGCAGTATCGCAGGGGAAGAGCTGGTGTCTGCTGCC
AGGATCACGGTCAGTACCTGATCAGCAGGCTGAATCATCCCGACGATACCCATGGCTGGCATTGGGGCGATTATCCC
TACACCCTGGTCTGGATCGTCGAGGCGCCTGAAGACCCGGAGGGCGGCGACGTCCTGCAATGCGTGCCCTATAGCAA
CGGCAACAAGGATGACTTGACGTTTGGGAATACAGCCTGGACGATCCTGGCAAGGCCTGTCCCCACCTCAAGGGTG
ATGTGTATTTCTCAAGTCCGGTACCACGCTGCACCATCTGGTGCCGATTCAACGAAACGCTACGCGCATTATCCTC
AGTACCAGTTGGCAAGTAGCTCCGGCTGGCGAAGCGATGCCCTGTGGATAA

Lysine cyclodeaminase (RapL) [2]

CGGCCATCTAGAAGTGCTTTTTTCAGGGCCCCGCATATGCAGACTAAAGTGTTGTGCCAGCGCGATATCAAGCGTATTT
TATCTGTAGTGGGGCGCGATGTCATGATGGACCGCCTTATTAGCGAAGTTCATGCAGGATTTGCCCGCTTGGGCCGT
GGTGAGACAGACGAACCACCGCCTCGCACAGGATTTGCACGCGGGGGGGATGTACCCGGAGTAATTGAATTCATGCC
TCATCGCGCATCCGGTATTGGGGTTACAATGAAGACAGTCAGCTATTTCGCCACAGAATTTTCGAGCGCTTCAATCTTC
CTACTATCGTCGGGACTGTAAGCCGTTTAGATGACGATAGTGGATCAATGGTCGCTCTTGAGATGCTGCGACGATC
ACGGCGATGCGTACAGGAGCGGTAGCGGCTGTTGCAACACGTCTGCTTGCTCGTCCAGGATCAACGACCTTGGCATT
GATCGGAGCCGGCGCTCAGGCGGTGACGCAAGCGCACGCCCTGAGTCGCGTTCTGCCGTTAGAGCGCATCTTGATTA
GCGACATCAAAGCTGAACATGCGGAATCTTTTGCAGGCCGCGTAGCATTCTGGAACCTTCCCGTTGAGGTAACAGAC
GCGGCCACCGCGATGGCTACAGCTGACGTATTATGTACGGTCACTTCCGTGCCTGTGGGGGGTGGACCTGTAGTGCC
GGCCGAGCCTCGTCAGGCTCATCTGCATGTGAATGGCATTGGAGCTGATGAGCAGGGCAAGACAGAGTTACCGAAAG
CGTTGCTGGATGATGCTTTTTATTTGTGTTGATCACCTGGGCAGGCCCGCGCAGAAGGGGAGTTTCAACAGCTGCCG
GATCGTGAATTAGGACCTAGCCTGGCTGATTTATGTGCAGCCCCAGAAATTGCCGCACCACATCCTGAGCGCTTATC
GGTGTGTTGACAGCACCGGCTCGGCCTTCGCAGACCATATCGCTTTAGATGTACTGCTGGGATTTCGCGGATGAGTTGG
GTTTAGGACACAAAATGTCTATCGAATCGACCCCAGAGGATGTTTTAGATCCATACTCGCTGTGAGGATCCGGCTGC
TAACAAAGCCCCGAAAGGAAG

Table A1.3. Strains, plasmids, oligonucleotides, and sequences used in Chapter 4. (A) Strains and plasmids used in this study. (B) Oligonucleotides used for plasmid construction and sequencing. (C) Sequences of synthetic genes. (D) Untested plasmids related to Chapter 4. (E) Cloning table for untested constructs. (F) Oligonucleotides use to clone untested constructs.

A1.3A. Strains and plasmids

Strain	Description	Source
BL21 Star (DE3)	<i>F-ompT hsdSB (rB-, mB-) gal dcm rne131</i> (DE3)	Thermo-Fisher

Plasmid	Description	Source
pET16b-His ₁₀ -Hydrox	His ₁₀ -Hydrox (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -Hal	His ₁₀ -Hal (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -Hal_opt	His ₁₀ -Hal_opt (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -Hydrox D144G	His ₁₀ -Hydrox D144G (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -Hydrox D144G V225N	His ₁₀ -Hydrox D144G V225N (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -Hydrox D144G N151I	His ₁₀ -Hydrox D144G N151I(T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -Hydrox D144G N151I V225N	His ₁₀ -Hydrox D144G N151I V225N (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -Hydrox D144G N151I V225N F150L	His ₁₀ -Hydrox D144G N151I V225N F150L (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -Hydrox D144G N151I V225N GAV218-220 to ST218-219	His ₁₀ -Hydrox D144G N151I V225N GAV218-220 to ST218-219 (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -Hydrox D144G N151I V225N M223I	His ₁₀ -Hydrox D144G N151I V225N M223I(T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -Hydrox D144G N151I V225N V226M S227T	His ₁₀ -Hydrox D144G N151I V225N V226M S227T (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -Hydrox D144G N151I V225N S228A	His ₁₀ -Hydrox D144G N151I V225N S228A (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -Hydrox D144G N151I V225N E231K G235E V237D	His ₁₀ -Hydrox D144G N151I V225N E231K G235E V237D (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -Chimera	His ₁₀ -Chimera (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -Chimera L150F	His ₁₀ -Chimera L150F (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -Chimera ST218-219GAV	His ₁₀ -Chimera ST218-219GAV (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -Chimera I222M	His ₁₀ -Chimera I222M (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -Chimera M225V T226S	His ₁₀ -Chimera M225V T226S (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -Chimera A228S	His ₁₀ -Chimera A228S (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -Chimera K230E E234G D236V	His ₁₀ -Chimera K230E E234G D236V (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b- His ₁₀ -PrescissionCutSite-IMDPH	His ₁₀ -PrescissionCutSite-IMDPH (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pACYC-DUET-BesC.MBP-BesB	BesC (T7) and MBP-BesB (T7) from <i>P. fluorescens</i> , <i>lacI</i> , Cm ^R , ColE1	This study
pPra	BesD (T7) and BesC (T7) from <i>P. fluorescens</i> , MBP-sBesB (T7) from <i>S. sp.</i> NRRL S-1448, <i>lacI</i> , Cb ^R , ColE1	Ref. [3]
pACYC-DUET	<i>lacI</i> , f1, Cm ^R , ColE1	Novagen
pRARE	tRNA ^{Ile} (AUA), tRNA ^{Ser} (AGG), tRNA ^{Ser} (AGA), tRNA ^{Asp} (CUA), tRNA ^{Gly} (CCC), tRNA ^{Gly} (GGA), Cm ^R	Novagen

A1.3B. Oligonucleotide sequences

Name	Sequence
HalHydrox-F	atcatcatcacagcagcgccatctagaagtgttttcagggcccgcat
HalHydrox-R	cagcttcctttcgggctttgttagcagccggatc
Hydrox D144G-F	ctggcactgggatgactactcgtttgctttgatttgggcct
Hydrox D144G-R	aaagcaaacgagtagtcatcccagtgccagccgtgagtgtc
Hydrox V225N-F	gtgcgtacaatgctgaatgtcagttggcggcagaacgcgact
Hydrox V225N-R	tgccgaccaactgacattcagcattgtacgcacagctccatc
Hydrox D144G N151I-F	ggtgattatagtttcgcctttatttgggctcttcgatgccg
Hydrox D144G N151I-R	aagagcccaaataaaggcgaaactataatcaccccagtgcc
ChP1-R	gagcccaaatcaaggcgaaactataatcaccccagtgccacccatgtgtgc
ChP2-F	ggtgattatagtttcgccttgatttgggctcttcgatgccgccgattgcctc
ChP2-R	cgcccagggtcatgttcagtatagtcctagttgaatcttcagttaacgggaccgtgcg
ChP3-F	taggactatactgaacatgacctgggcgcaaaacgcgacttggaaaaggatcttgttg
HydroxTM F150L-F	ggcactgggggtgattatagtttcgccttgatttgggctcttcgatgccg
HydroxTM F150L-R	aagagcccaaataaaggcgaaactataatcaccccagtgcc
HydroxTM GAVtoST-F	cgcacgggtcccgttaactgaagattcaactcgtacaatgctgaatgtcagttggtcg
HydroxTM GAVtoST-R	ttgtacgagttgaatcttcagttaacgggaccgtgcg
HydroxTM M223I-F	actgaagatggagctgtgcgtacaatactgaatgtcagttggcggcagaac
HydroxTM M223I-R	ctgacattcagattgtacgcacagctccatcttcagt
HydroxTM NVS to NMT-F	tggagctgtgcgtacaatgctgaatgacctggcggcagaacgcgact
HydroxTM NVS to NMT-R	ccgaccagggtcatattcagcattgtacgcacagctcca
HydroxTM S228A-F	tgcgtaaatgctgaatgtcagttggcgccagaacgcgacttgggaaagg
HydroxTM S228A-R	ttctgccgcccaactgacattcagcattgtacgca
HydroxTM E230K G234E V236D	tgctgaatgtcagttggcggcaaaacgcgacttgaaaaggatcttacccg
HydroxTM E230K G234E V236D	ttccaagtcgcgttttcgccaccaactgacattcagca
Ch L150F-F	gcactgggggtgattatagtttcgccttcatttgggctcttcgatgccg
Ch L150F-R	gagcccaaatgaaggcgaaactataatcaccccagtgcc
Ch STtoGAV-F	cgcacgggtcccgttaactgaagatggagctgtgaggactatactgaacatgacctgggcg
Ch STtoGAV-R	cctcacagctccatcttcagttaacgggaccgtgcg
Ch I222M-F	ggtcccgttaactgaagattcaactaggactatgctgaacatgacctgggcggc
Ch I222M-R	agtcctagttgaatcttcagttaacgggacc
Ch NMT to NVS-F	ccgtaactgaagattcaactaggactatactgaacgtcagttggcgccaaacgcgac
Ch NMT to NVS-R	gttcagtatagtcctagttgaatcttcagttaacgg
Ch A228S-F	tcaactaggactatactgaacatgacctggcggcaaaacgcgacttggaaaaggatct
Ch A228S-R	gttttgcggaccagggtcatgttcagtatagtcctagttgaatct
Ch K230E E234G D236V-F	ctaggactatactgaacatgacctggcgccagaacgcgacttgggaaaggttctac
Ch K230E E234G D236V-R	tgccgccagggtcatgttca
pET16-anneal-R	cgtatcacgaggcccttctgccttc

pBesCB-F	taattttgtttaactttaataaggagatataaccccatgtccatcactcaagaaacgtttca
pBesCB-R	gcagcagcggtttcttaccagactcgagggtacctcacaagggacgcagagcttgct
IMDPH-F	tcatcatcatcatcatcacagcagcgccatctagaagtgttttcagggcccgcatatgtactatcg ctaaagaagctctgac
IMDPH-R	cagcttcctttcggtctttagcagccgatcctcaggagcccagacggtagtctggg
Preclission-seq-F	ctagaagtgttttcagggccg

A1.3C. Synthetic gene sequences

WP_107105619 (*Streptomyces* sp. MBT76 Hydrox, *E. coli* codon optimized)

AGCAGCGGCCATCTAGAAGTGCTTTTTTCAGGGCCCGCATATGGACGTGCATGAAATTGACGAAACCTTAGAAAAATT
CTTAGCGGAGAACTATACTCCGGAACGCGTTTCAGCAACTGGCAGATCGTTTCCAGCGTACCGGATTTCGTGAAGTTTCG
ACTCTCATATGCGCATCGTGCCTGAAGAACTGATCACCGCTGTTTCGCGCGGAGGCCGATCGCCTGGTTTCGCGAGCAT
AAAGAGCGTCGTGACCTTGTGTTAGGTACAACCTGGTGGGACGCCCCGTAATTTAAGCGTAGTCAAATCACAGGACGT
GGAGCAATCAGACTTAATTCGCGCCGTCCTCGCTCAGAGGTGTTGCTTACGTTTTTGGCTGGGATCACCCGTGAAC
GCATTATCCCTGAGGTGTCTGATGATGAGCGCTACTTGATCACTACCAAGAGTTTGCAGGTGACACACATGGGTGG
CACTGGGACGATTATAGTTTTCGCCTTTAATTGGGCTCTTCGCATGCCGCCGATTGCCTCTGGAGGGATGGTGCAGGC
TGTCCACACACCCATTGGGATAAGAACGCGCCTCGTATCAATGAAACGCTGTGCGAGCGTCAGATCGACACGTACG
GGCTTGTGTCGGGCGACTTGTATCTGCTTCGTTTCAGATACCACGATGCACCGCACGGTCCCGTTAACTGAAGATGGA
GCTGTGCGTACAATGCTGGTTGTGAGTTGGTGGCAGAACGCGACTTGGGAAAGGTTCTTACCGGGAATGATCGCTG
GTGGGAGAATCCAGAAGCAGGCGCCGCTCAACCAGTGCATCGTGCGGGATGAGGATCCGGCTGCTAACAAAGCCCGA
AAGGAA

WP_122981682 (*Actinoplanes teichomyceticus* Hal, *E. coli* codon optimized)

AGCAGCGGCCATCTAGAAGTGCTTTTTTCAGGGCCCGCATATGAACGTAGAGCAGATCGACGAAAATTTAGCCAAATT
TCTTGCGGAACGCTATAACCCCTGAGTCCGTGGCACAGTTGGCCGACCGTTTCCATCGCTTCGGTTTTGTAAAGTTTG
ACGCCGCTAATCGCCTGGTTCCAGACGAATTGCAAACCTGCTGTGCGCGAAGAATGCGATTTACTTATCGAACAGCAT
AAAGAACGCCGTAACCTTATTACTTTCTACAACCTGGCAATACCCCGCGTCGTATGTCCGTTGTAAAAAGCGAGGAAAT
TGAGAAGTCCGAGCTTATCTCTACATTGTGCGCTCCGAGGTTTTACTTGGCTTTTTGGCGGGTATCACCCGTGAGG
AAATTATCCCTGAAGTGTATCGGATGAGCGTTATCTTATTACACATCAAGAATTCAAATCCGACACTCACGGCTGG
CACTGGGGTGACTACTCGTTTGCTTTGATTTGGGCTTACGCATGCCACCAATCGAGCACGGTGGTATGCTTCAAGC
AGTCCCACATACACACTGGGATAAATCAAACCCGCGCATCAATCAAACATTGTGTGAACGCGAGATTAACACCCATG
GGTTAGAAAGTGGTGATTTATATTTACTTTCGTACAGACACCACACTGCATCGTACAGTCCCGCTTTCCGAGGATTCT
ACCCGTACCATCCTGAACATGACTTGGGCGCGAAGCGCGACTTGGAAAAGGACTTAGTGGGGAATGATCGCTGGTG
GGAAAATCCCGAGGCCGAAGCGGCACGCGCTGTAGACAATGCGTGAGGATCCGGCTGCTAACAAAGCCCGAAAGGAA

WP_122981682_optimized (*Actinoplanes teichomyceticus* Hal, optimized to match Hydrox)

AGCAGCGGCCATCTAGAAGTGCTTTTTTCAGGGCCCGCATATGAACGTGGAGCAAATTGACGAAAACCTTAGCAAAATT
CTTAGCGGAGCGCTATACTCCGGAATCCGTTGCGCAACTGGCAGATCGTTTCCATCGTTTCGGATTTCGTGAAGTTTCG
ACGCTGCTAATCGCCTCGTGCCTGACGAACCTGCAAACCGCTGTTTCGCGAGGAGTGCGATCTCCTGATTGAGCAGCAT
AAAGAGCGTCGTAACTTCTGTAAAGTACAACCTGGTAATACGCCCCGTCGTATGAGCGTAGTCAAATCAGAGGAAAT
TGAGAAATCAGAGTTAATTAGCACCCCTCTCTCGCTCAGAGGTGTTGCTTGGGTTTTTGGCTGGGATCACCCGTGAAG
AAATTATCCCTGAGGTGTCTTCTGATGAGCGCTACTTGATCACTACCAAGAGTTTAAGAGTGACACACATGGGTGG
CACTGGGGTGATTATAGTTTTCGCTTGGCTTTCGCTATGCCGCCGATTGAGCATGGAGGGATGCTGCAGGC
TGTCCACACACCCATTGGGATAAGTCCAACCTCGTATCAATCAAACGCTGTGCGAGCGTGAGATCAACACGCACG
GGCTTGAGTCGGGCGACTTGTATCTGCTTCGTACAGATACCAGCTGCACCGCACGGTCCCGTTATCTGAAGATTCA
ACTAGGACTATACTGAACATGACCTGGGCGGCAAAACGCGACTTGGAAAAGGATCTTGTGGGAATGATCGCTGGTG
GGAGAATCCAGAAGCAGAAGCCGCTCGAGCAGTGGATAATGCGTGAGGATCCGGCTGCTAACAAAGCCCGAAAGGAA

A1.3D. Untested plasmids associated with Chapter 4

Number	Plasmid	Description
1	pET16b-His ₁₀ -Hal I151N	His ₁₀ -Hal I151N (T7), <i>lacI</i> , Cb ^R , ColE1
2	pET16b-His ₁₀ -Hal G144D	His ₁₀ -Hal G144D (T7), <i>lacI</i> , Cb ^R , ColE1
3	pET16b-His ₁₀ -Hal N225V	His ₁₀ -Hal N225V (T7), <i>lacI</i> , Cb ^R , ColE1
4	pET16b-His ₁₀ -Hal G144D I151N	His ₁₀ -Hal G144D I151N (T7), <i>lacI</i> , Cb ^R , ColE1
5	pET16b-His ₁₀ -Hal G144D L150F I151N	His ₁₀ -Hal G144D L150F I151N (T7), <i>lacI</i> , Cb ^R , ColE1
6	pET16b-His ₁₀ -Hal I151NNK	His ₁₀ -Hal I151NNK (T7), <i>lacI</i> , Cb ^R , ColE1
7	pET16b-His ₁₀ -Hydrox N151I	His ₁₀ -Hydrox N151I (T7), <i>lacI</i> , Cb ^R , ColE1
8	pET16b-His ₁₀ -Hydrox V225N N151I	His ₁₀ -Hydrox V225N N151I (T7), <i>lacI</i> , Cb ^R , ColE1
9	pET16b-His ₁₀ -SiBesD N219V	His ₁₀ -SiBesD N219V (T7), <i>lacI</i> , Cb ^R , ColE1
10	pET16b-His ₁₀ -SiBesD I151N	His ₁₀ -SiBesD I151N (T7), <i>lacI</i> , Cb ^R , ColE1
11	pET16b-His ₁₀ -Hal G144D I151N N225V	His ₁₀ -Hal G144D I151N N225V (T7), <i>lacI</i> , Cb ^R , ColE1
12	pET16b-His ₁₀ -Hal I151N N225V	His ₁₀ -Hal I151N N225V (T7), <i>lacI</i> , Cb ^R , ColE1
13	pET16b-His ₁₀ -Hal G144D N225V	His ₁₀ -Hal G144D N225V (T7), <i>lacI</i> , Cb ^R , ColE1
14	pET16b-His ₁₀ -Hal I151A	His ₁₀ -Hal I151A (T7), <i>lacI</i> , Cb ^R , ColE1
15	pET16b-His ₁₀ -Hal I151V	His ₁₀ -Hal I151V (T7), <i>lacI</i> , Cb ^R , ColE1
16	pET16b-His ₁₀ -Hal G144D L150F I151N N225V	His ₁₀ -Hal G144D L150F I151N N225V (T7), <i>lacI</i> , Cb ^R , ColE1
17	pET16b-His ₁₀ -Hydrox V225N	His ₁₀ -Hydrox V225N (T7), <i>lacI</i> , Cb ^R , ColE1

A1.3E. Cloning table for untested plasmids. All constructs were made from a 3-piece Gibson reaction with piece A, piece B, and NdeI/HindIII-cut pET16b-His₁₀-Prescission-IMDPH vector. Numbers correspond to the plasmids in A1.3D.

	PCR template	Forward primer	Reverse primer
1a	pET16b-His ₁₀ -Hal	HalHydrox-F	Hal I151N-R
1b	pET16b-His ₁₀ -Hal	Hal I151N-F	pET16-anneal-R
2a	pET16b-His ₁₀ -Hal	HalHydrox-F	Hal G144D-R
2b	pET16b-His ₁₀ -Hal	Hal G144D-F	pET16-anneal-R
3a	pET16b-His ₁₀ -Hal	HalHydrox-F	Hal N225V-R
3b	pET16b-His ₁₀ -Hal	Hal N225V-F	pET16-anneal-R
4a	pET16b-His ₁₀ -Hal G144D	HalHydrox-F	Hal G144D I151N-R
4b	pET16b-His ₁₀ -Hal G144D	Hal G144D I151N-F	pET16-anneal-R
5a	pET16b-His ₁₀ -Hal G144D	HalHydrox-F	Hal G144D L150F I151N-R
5b	pET16b-His ₁₀ -Hal G144D	Hal G144D L150F I151N-F	pET16-anneal-R
6a	pET16b-His ₁₀ -Hal	HalHydrox-F	Hal I151NNK-F
6b	pET16b-His ₁₀ -Hal	Hal I151NNK-R	pET16-anneal-R
7a	pET16b-His ₁₀ -Hydrox	HalHydrox-F	Hydrox N151I-R
7b	pET16b-His ₁₀ -Hydrox	Hydrox N151I-F	pET16-anneal-R
8a	pET16b-His ₁₀ -Hydrox V225N	HalHydrox-F	Hydrox N151I-R
8b	pET16b-His ₁₀ -Hydrox V225N	Hydrox N151I-F	pET16-anneal-R
9a	pET16b-His ₁₀ -SiBesD	HalHydrox-F	SiBesD N219V-R
9b	pET16b-His ₁₀ -SiBesD	SiBesD N219V-F	pET16-anneal-R
10a	pET16b-His ₁₀ -SiBesD	HalHydrox-F	SiBesD I151N-R
10b	pET16b-His ₁₀ -SiBesD	SiBesD I51N-F	pET16-anneal-R
11a	pET16b-His ₁₀ -Hal G144D N225V	HalHydrox-F	Hal G144D I151N-R
11b	pET16b-His ₁₀ -Hal G144D N225V	Hal G144D I151N-F	pET16-anneal-R
12a	pET16b-His ₁₀ -Hal N225V	HalHydrox-F	Hal I151N-R
12b	pET16b-His ₁₀ -Hal N225V	Hal I151N-F	pET16-anneal-R

13a	pET16b-His ₁₀ -Hal G144D	HalHydrox-F	Hal N225V-R
13b	pET16b-His ₁₀ -Hal G144D	Hal N225V-F	pET16-anneal-R
16a	pET16b-His ₁₀ -Hal G144D N225V	HalHydrox-F	Hal G144D L150F I151N-R
16b	pET16b-His ₁₀ -Hal G144D N225V	Hal G144D L150F I151N-F	pET16-anneal-R
17a	pET16b-His ₁₀ -Hal G144D N225V	HalHydrox-F	Hydrox V225N-R
17b	pET16b-His ₁₀ -Hal G144D N225V	Hydrox V225N -F	pET16-anneal-R

A1.3F. Oligonucleotide sequences for untested plasmids. Numbers correspond to the plasmids in *A1.3D*.

Associated construct	Primer	Sequence
1, 12	Hal I151N-F	gcactggggtgactactcggttgcttgaattgggccttacgcatgccacc
1, 12	Hal I151N-R	caaagcaaacgagtagtcacccagtg
2	Hal G144D-F	ctggcactgggatgactactcggttgcttgaattgggcct
2	Hal G144D-R	aaagcaaacgagtagtcacccagtgccagcggtagtg
3, 13	Hal N225V-F	cgtaccatcctggtatgacttggccgcgaagcgcgac
3, 13	Hal N225V-R	gcggcccaagtcataaccaggatggtacgggtagaatcctcg
4, 11	Hal G144D I151N-F	gcactgggatgactactcggttgcttgaattgggccttacgcatgccacc
4, 11	Hal G144D I151N-R	caaagcaaacgagtagtcacccagtg
5, 16	Hal G144D L150F I151N-F	gcactgggatgactactcggttgcttgaattgggccttacgcatgccacc
5, 16	Hal G144D L150F I151N-R	aaaagcaaacgagtagtcacccagtg
6, 14, 15	Hal I151NNK-F	actggggtgactactcggttgcttgnnktgggccttacgcatgccacc
6, 14, 15	Hal I151NNK-R	caaagcaaacgagtagtcacccagtg
7	Hydrox N151I-F	ttatagtttcgccttatttgggctcttcgcatgccg
7	Hydrox N151I-R	cgaagagcccaaataaaggcgaaactataatcgccagtg
8	Hydrox N151I-F	ttatagtttcgccttatttgggctcttcgcatgccg
8	Hydrox N151I-R	cgaagagcccaaataaaggcgaaactataatcgccagtg
9	SIBesD N219V-F	ggacaccaccgcatcatcctggtcatgacctggcgggcgag
9	SIBesD N219V-R	ctgcccggccaggatcatgaccaggatgatgcgggtggtgtcc
10	SIBesD I51N-F	ggcgatttcagcttcgccctcaactgggtgtccaggcccc
10	SIBesD I151N-R	ggggcctggagcaccagttgaggcggaagctgaaatcgcc
17	Hydrox V225N-F	gtgcgtacaatgctgaatgtcagttggtcggcagaacgcgact
17	Hydrox V225N-R	tgccgaccaactgacattcagcattgtacgcacagctccatc

Table A1.4. Strains, plasmids, oligonucleotides, and sequences used in Chapter 5. (A) Strains and plasmids used in this study. (B) Oligonucleotides used for plasmid construction and sequencing. (C) Sequences of synthetic genes.

A1.4A. Strains and plasmids

Strain	Description	Source
BL21 Star (DE3)	<i>F-ompT hsdSB (rB-, mB-) gal dcm rne131</i> (DE3)	Thermo-Fisher

Plasmid	Description	Source
pET16b-His ₁₀ -SlavenduligriseusBesD	His ₁₀ -SlavenduligriseusBesD (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -SwuyuanensisHalB	His ₁₀ -SwuyuanensisHalB (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -SwHalB M246NNK	His ₁₀ -SIBesD R74A (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -SwHalB F250NNK	His ₁₀ -SIBesD W238A (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -SwHalB M246V F250NNK	His ₁₀ -SIBesD W239A(T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -SwHalB M246L F250NNK	His ₁₀ -SIBesD H134A (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -SwHalB M246I F250NNK	His ₁₀ -SIBesD D140A (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -PrescissionCutSite-IMDPH	His ₁₀ -PrescissionCutSite-IMDPH (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pET16b-His ₁₀ -PkilonensisHalD	His ₁₀ -PkilonensisHalD (T7), <i>lacI</i> , Cb ^R , ColE1	This study
pACYC-DUET-BesC.MBP-BesB	BesC (T7) and MBP-BesB (T7) from <i>P. fluorescens</i> , <i>lacI</i> , Cm ^R , ColE1	This study
pPra	BesD (T7) and BesC (T7) from <i>P. fluorescens</i> , MBP-sBesB (T7) from <i>S. sp.</i> NRRL S-1448, <i>lacI</i> , Cb ^R , ColE1	Ref. [3]
pACYC-DUET	<i>lacI</i> , f1, Cm ^R , ColE1	Novagen
pRARE	tRNA ^{Ile} (AUA), tRNA ^{Ser} (AGG), tRNA ^{Ser} (AGA), tRNA ^{Asp} (CUA), tRNA ^{Gly} (CCC), tRNA ^{Gly} (GGA), Cm ^R	Novagen

A1.4B. Oligonucleotide sequences

Name	Sequence
His ₁₀ -SIBesD-F	gcggccatctagaagtgccttttcaggggcccgcatatgagcagcaatgggcaggaaagcaccgctgtca
His ₁₀ -SIBesD-R	cctttcgggctttagcagccggatcctcgagtcagtcacagtcagcggtgcccggc
His ₁₀ -SwHalB-F	agcagcggccatctagaagtgccttttcaggggcccgcatatgaccgagaacagcatcacccgc
His ₁₀ -SwHalB M246NNK-R	cagcttccttcgggctttagtagcagccggatcctcatcagtcgaacatcgcggtmnnnggtctggtgggtggtcgcc
His ₁₀ -SwHalB F250NNK-R	cagcttccttcgggctttagtagcagccggatcctcatcagtcmmnncatcggttcatggtctcggtgggtggtcgcc
His ₁₀ -SwHalB M246V F250NNK-R	cagcttccttcgggctttagtagcagccggatcctcatcagtcmmnncatcgcggttaacgggtctcggtgggtggtcgcc
His ₁₀ -SwHalB M246L F250NNK-R	cagcttccttcgggctttagtagcagccggatcctcatcagtcmmnncatcgcggttaacgggtctcggtgggtggtcgcc
His ₁₀ -SwHalB M246I F250NNK-R	cagcttccttcgggctttagtagcagccggatcctcatcagtcmmnncatcgcggttaacgggtctcggtgggtggtcgcc
pET16-anneal-R	cgtatcacgaggcccttctgtcttc
His ₁₀ -PkilonensisHalD-F	agcagcggccatctagaagtgccttttcaggggcccgcatatgagcgaacaaacgtccagcttg
His ₁₀ -PkilonensisHalD-R	ttccttcgggctttagtagcagccggatcctcgagtcaggcttctcggtggtccg
ChDB1-F	gcggccatctagaagtgccttttcaggggcccgcatatgagcagcaatgggcaggaaagcaccgctgtca
ChDB1-R	ctggaggcggagatcaatccgctcgccgtatttctccagcagccggtc
ChDB2-F	aatacgcgagcggattgatctccgctccagaccaccggatacaccg
ChDB2-R	gataatgaattcctcgctggccaggggcagggtgcagtttctcgccgg

ChDB3-F	ccccctgggcccgcagaggaattcattatcacccggcaggaacggagcgggg
ChDB3-R	ggctcgtgggtggtcgcttctgctggtcccgctcgccgcccaggtc
ChDB4-F	gaccagcagaaggcgaccacccacgagacc
ChDB4-R	cgtatcacgaggcccttctgtcttc
pBesCB-F	taattttgtttaactttaataaggagatataccccatgtccatcactcaagaaacgtttca
pBesCB-R	gcagcagcggtttcttaccagactcgagggtagctcacaagggacgcagagcttgct
IMDPH-F	tcatcatcatcatcatcacagcagcgccatctagaagtgttttcagggcccgcatatgtacgtatcgctaaaga
IMDPH-R	agctctgac
	cagcttccttgcggcctttagtagcagccggatcctcaggagcccagacggtagttcggg

A1.4C. Synthetic gene sequences

WP_030791981 (*Streptomyces lavenduligriseus* BesD)

AGCAGCGGCCATCTAGAAGTGCTTTTTTCAGGGCCCCGCATATGAGCAGCAATGGGCAGGAAAGCACCGTCGTCAATCC
TCTGGAGCAGGGCGCACTGCGCCGTATGGCGCATCACTACCACCGGTACGGCATCGCCACCGTCACCGATCTGATTC
GGGAAGACGTCCGCAAAAACGTGCGGGCGGAGGCGGACCGGCTGCTGGAGAAATACGCCGAGCGGCGTGATCTCCGC
CTCCAGACCACCGGATACACCCGCCGCTCGATGTCCGTGGTGAGAGCGAGACGATCGCGGGCAACAGCGAGCTGGT
CACCTCGATCTACGCGAATCCGGAAGTCTCGGCGCGCTGGAGCGCATAGCCGGCGAGAACTGCACCCCTGCCCA
AGGCCGACGAGGAATTCCTCATCACCCGGCAGGAACGGAGCGGGACACGCACGGCTGGCACTGGGGCGATTTTCAGC
TTCGCCCTCATCTGGGTGCTCCAGGCCCGCCCATCGACATCGGCGGCATGCTCCAGTGCGTACCGCACACGGAGTG
GGACAAGTCCGATCCGCGGATCCACAGTATCTCGTCGACAATCCCATCCACACGTACCACTTCCAGTCGGGCGACG
TGTATTTTCTGCGCACCGACACCACGCTCCACCGCACGGTCCCGCTGCGCGAGGACACCACCCGCATCATCTGAAC
ATGACCTGGGCGGGCGAGCGGGACCTGCGGCGCGAGCTCAAGGGCGACGACCGCTGGTGGGAGGACGCCGACGTGCC
GGCAGCCGCTGCACTGGATGACTGACTCGAGGATCCGGCTGCTAACAAAGCCCCGAAAGGAAGCTGA

SDN46247 (*Streptomyces wuyuanensis* HalB)

AGCAGCGGCCATCTAGAAGTGCTTTTTTCAGGGCCCCGCATATGACCGAGAACAGCATCACCGCCGCGAATGTGAAGA
GCTGATCGCGAAGAATCGCGGAGCGGTTCCGCCGACGACCACGAGGTGCTCGGCCTCTCCCAGCACTTCCGCCGCG
AGGGCTACGTCAAGCTCCCCGGCCTGGTCTCCCCGGAGGTCTTCGACGCGGTGCGCGGGAGACCCACCAGCTGATC
GACACCCACCAGAAGCGCATCGACATCCGTCTCAAGGAGACGGGGGACTCCCCGCGCTACATGTCCACCGTCGGTCA
GAAGGCGATCGCCACCGACGGCTCGCTGATCCCTGCCGTCTACGAGTCCACCGCGCTCAAGGGCTTCTTTTCCGCC
TCGCCAAGGAGGAGGTATGGGCTGCCCCTGGGACGAGGAGAAGTACATCATCACCCGCCAGCACCAGAAGGGCGAC
ACCCACGGCTGGCACTGGGGCGACTTCAGCTTCACCGTCATCTGGCTGATCGAGGCCCCGTCGCTGGAGTACGGCGG
CATGCTGCAGTGCATCCCGCACACCGACTGGAACAAGGACGACCCGCGCGTTCGAGGACTACCTGCAGAAGCACCCGA
TCCGCAGCTACGGCCACGCCAAGGGCGACCTCTACCTGCTGCGTTTCGGACACCACCCTGCACCGCACGGTCCCGCTG
AACGCCGACAGGACCCGCATCATCTCAACACCTGCTGGGCCAGCCGCGCGGACCAGCAGAAGGCGACCCACGA
GACCATGAACGCGATGTTTCGACTGATGAGGATCCGGCTGCTAACAAAGCCCCGAAAGGAAGCTG

WP_046063366 (*Pseudomonas kilonensis* HalD)

ATGAGCGAACAACAGTCCAGCTTGGTTATTGAAGTAATGGAGCAGCAACTGGCCAAGCATTTCAGGCTATTTTGCA
AGATGAAAACCGGATGAAGCAGATTCTGAATGAATTTCCGCCGAGACGGTTATTTCAACTTCAAGAACTTTTCGTTTC
TACCGAAAAGGATCTTGGAAAATGTCCACGCCGAAGTCCATGCGTTATTGGATGAATATTTCGGTGCCTCGCGATGTT
ACTGTTCCGTCCACAGGCAATACCTACCGCAAGATGTACAACGTCACCAACCGGAGATCGCCGAAGGCGGGACATT
CATCCCTCGCTCTACCAATCCGAGTCGCTGCGTAAGTTCTTGGGCAACATCGCTGGCGATGATCTGGCGTCTTGCT
GGGAGCAGGAGCAATACCTGGTACCAAGCTGAGCCACCCGGGCGATACCCACGGCTGGCATTGGGGGGATTACCCG
TACACGATGATCTGGATCATCGAGGCGCCCGAGGACCCGGCGATTGGCGGCGTGCTCCAGTGCCTGCCACACAGCGA
ATGGGACAAGCAGAACCCGCAGATCTGGCAGTACATCTCAATAACCCGATCAAGTCTTACCACCATCTCAAGGGTG
ATGTGTATTTCTCAAGTCCGACACCACGTTGCACCACGTCGTGCCGATTTCAGCAGGAAACCACTCGGATAATTCTC
AACACATGCTGGGCCAGCGCCCATGACCGGCGAACCGATGTGCCCCACGAAAGCATCGAAGTGATCTGGGACACCAA
GGCCCGGACCCGGAAGAAGCCTGA

ChimeraDB coding sequence

ATGAGCAGCAATGGGCAGGAAAGCACCGTCGTCAATCCTCTGGAGCAGGGCGCACTGCGCCGTATGGCGCATCACTA
CCACCGGTACGGCATCGCCACCGTCACCGATCTGATTTCGGGAAGACGTCCGCAAAAACGTGCGGGCGGAGGCGGACC
GGCTGCTGGAGAAATACGCCGAGCGGATTGATCTCCGCCTCCAGACCACCGGATACACCCGCCGCTCGATGTCCGTG
GTGCAGAGCGAGACGATCGCGGGCAACAGCGAGCTGGTCACCTCGATCTACGCGAATCCGGAAGTCTCGGCGCGCT
GGAGCGCATAGCCGGCGAGAAACTGCACCCCTGCCCCTGGGCCGACGAGGAATTCATTATCACCCGGCAGGAACGGA
GCGGGGACACGCACGGCTGGCACTGGGGCGATTTTCAGCTTCGCCCTCATCTGGGTGCTCCAGGCCCCGCCCATCGAC
ATCGGCGGCATGCTCCAGTGCGTACCGCACACGGAGTGGGACAAGTCCGATCCGCGGATCCACCAGTATCTCGTCGA
CAATCCCATCCACACGTACCACTTCCAGTCGGGCGACGTGTATTTCTGCGCACCGACACCACGCTCCACCGCACGG
TCCCGCTGCGCGAGGACACCACCGCATCATCCTGAACATGACCTGGGCGGGCGAGCGGGACCAGCAGAAGGCGACC
ACCCACGAGACCATGAACGCGATGTTCTGACTGA

Appendix 2: *Supplementary Figures*

Figure A2.1. Phylogenetic tree of BesD homologs. The tree was generated using MEGA [4] from an alignment made in MUSCLE [5]. Along with BesD, Hal and Hydrox from Chapter 4 are annotated.

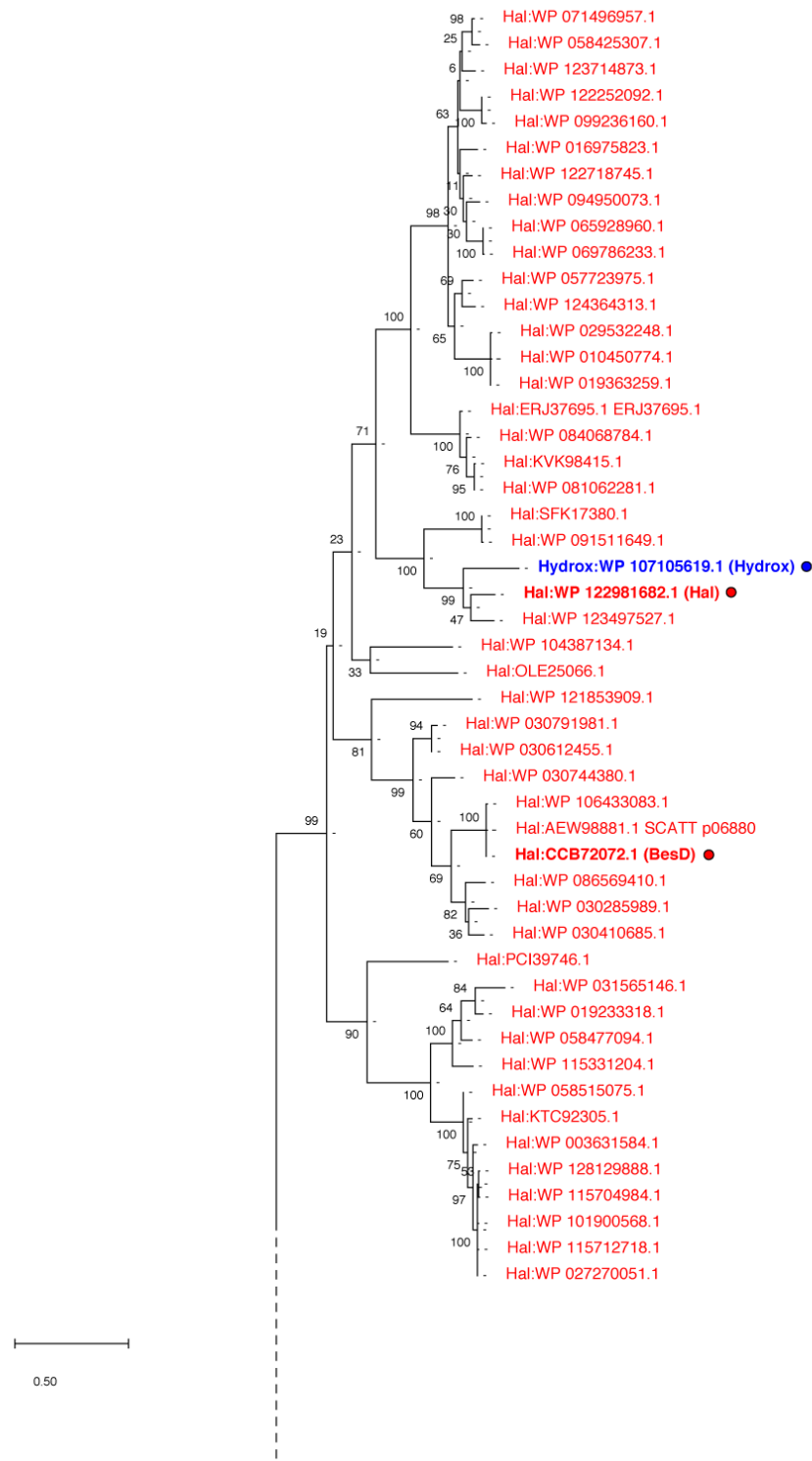


Figure A2.1 (continued). Phylogenetic tree of BesD homologs. The tree was generated using MEGA [4] from an alignment made in MUSCLE [5].

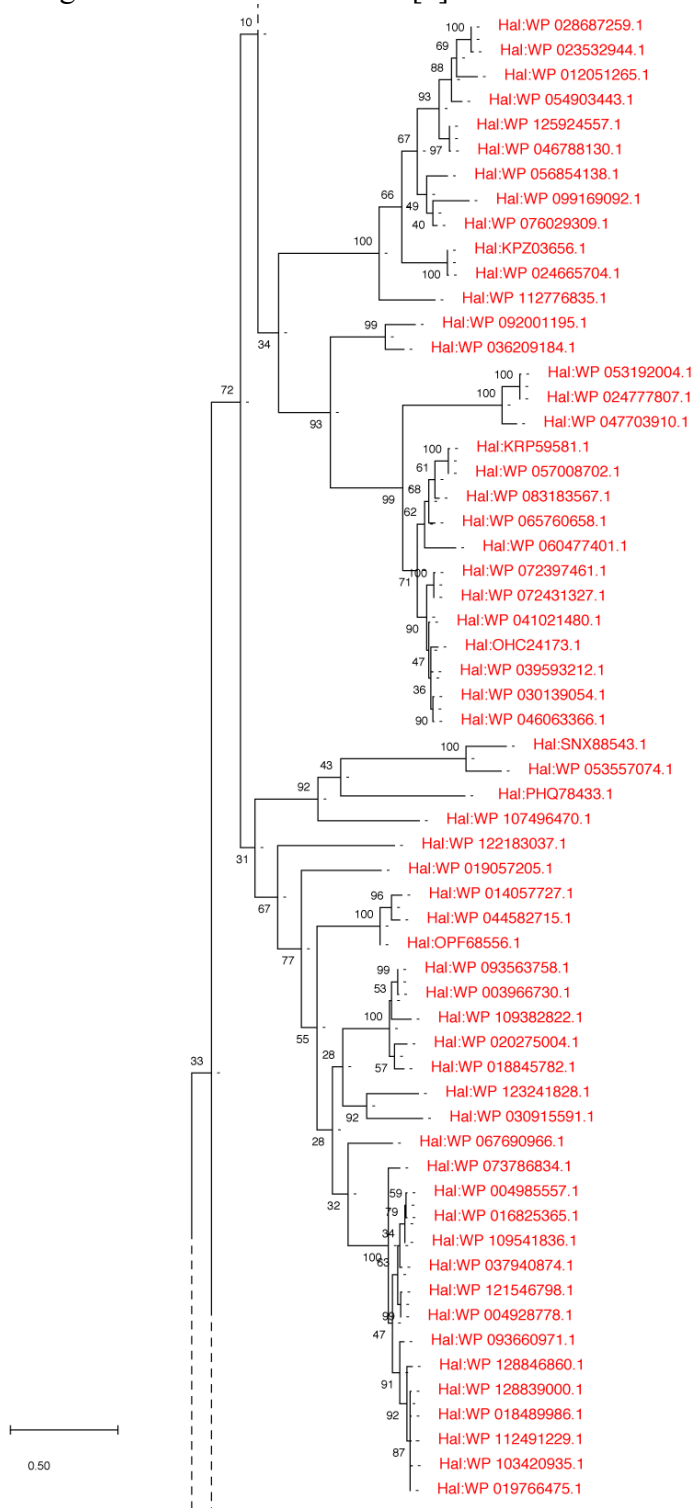


Figure A2.1 (continued). Phylogenetic tree of BesD homologs. The tree was generated using MEGA [4] from an alignment made in MUSCLE [5].

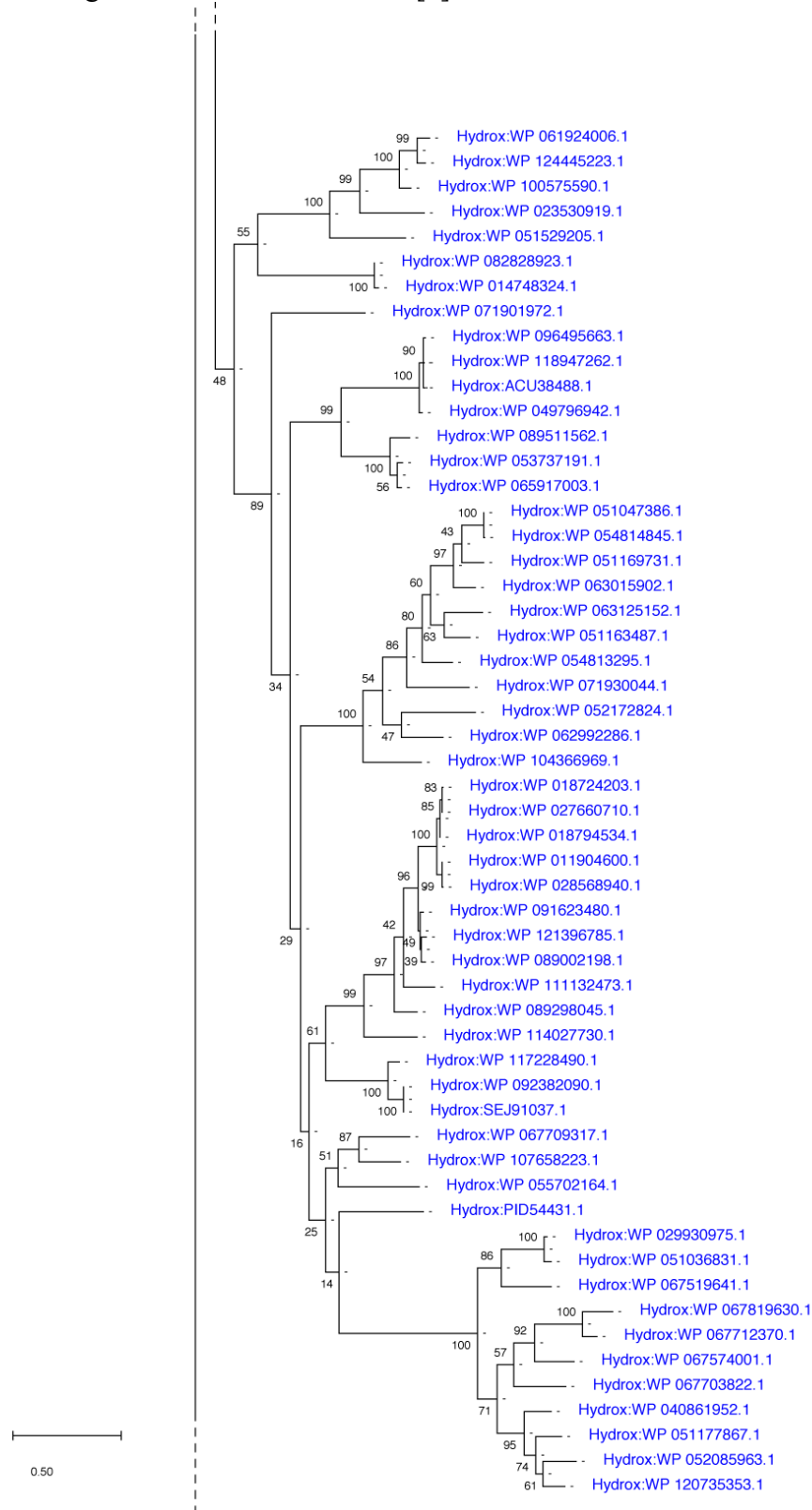


Figure A2.1 (continued). Phylogenetic tree of BesD homologs. The tree was generated using MEGA [4] from an alignment made in MUSCLE [5].

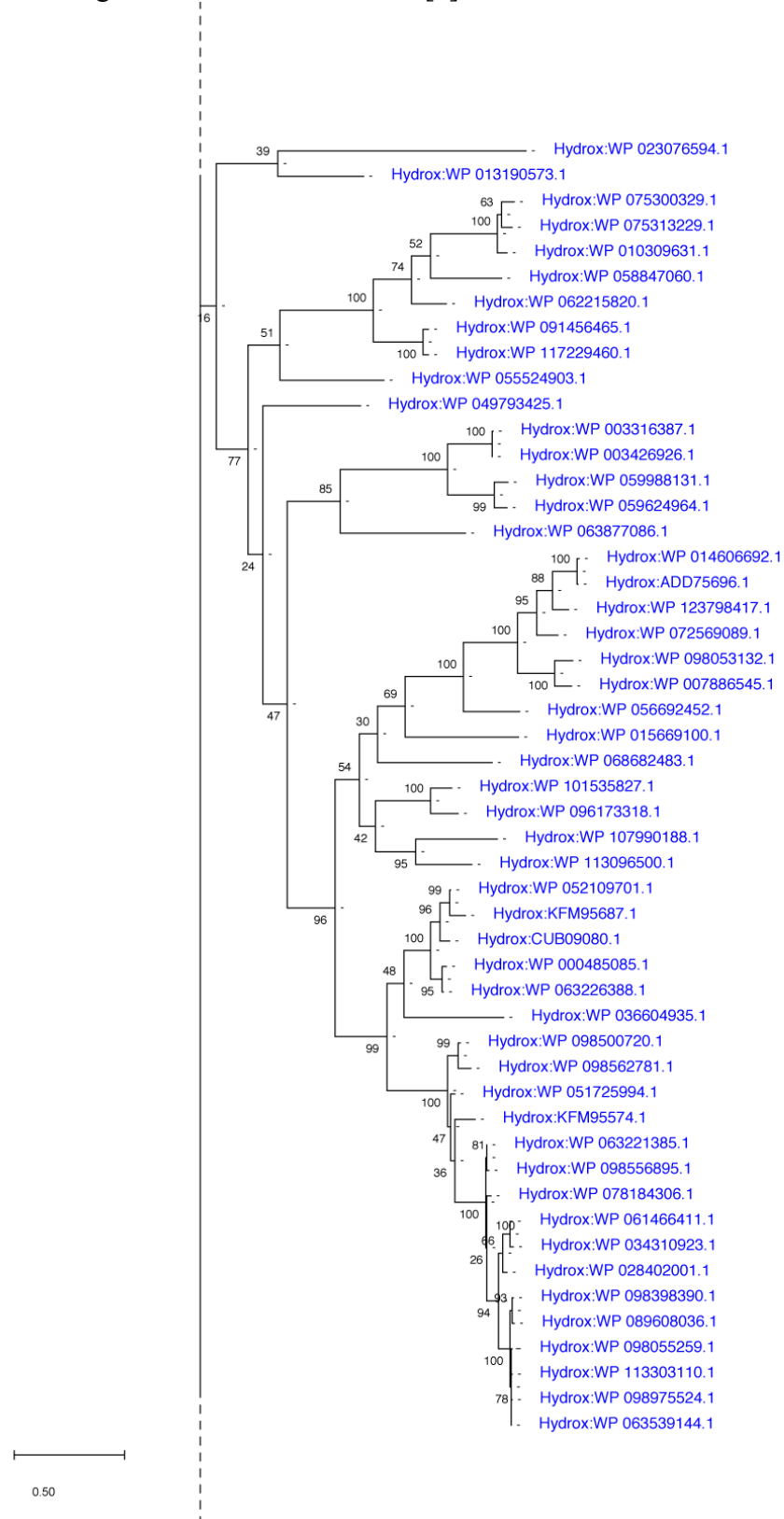
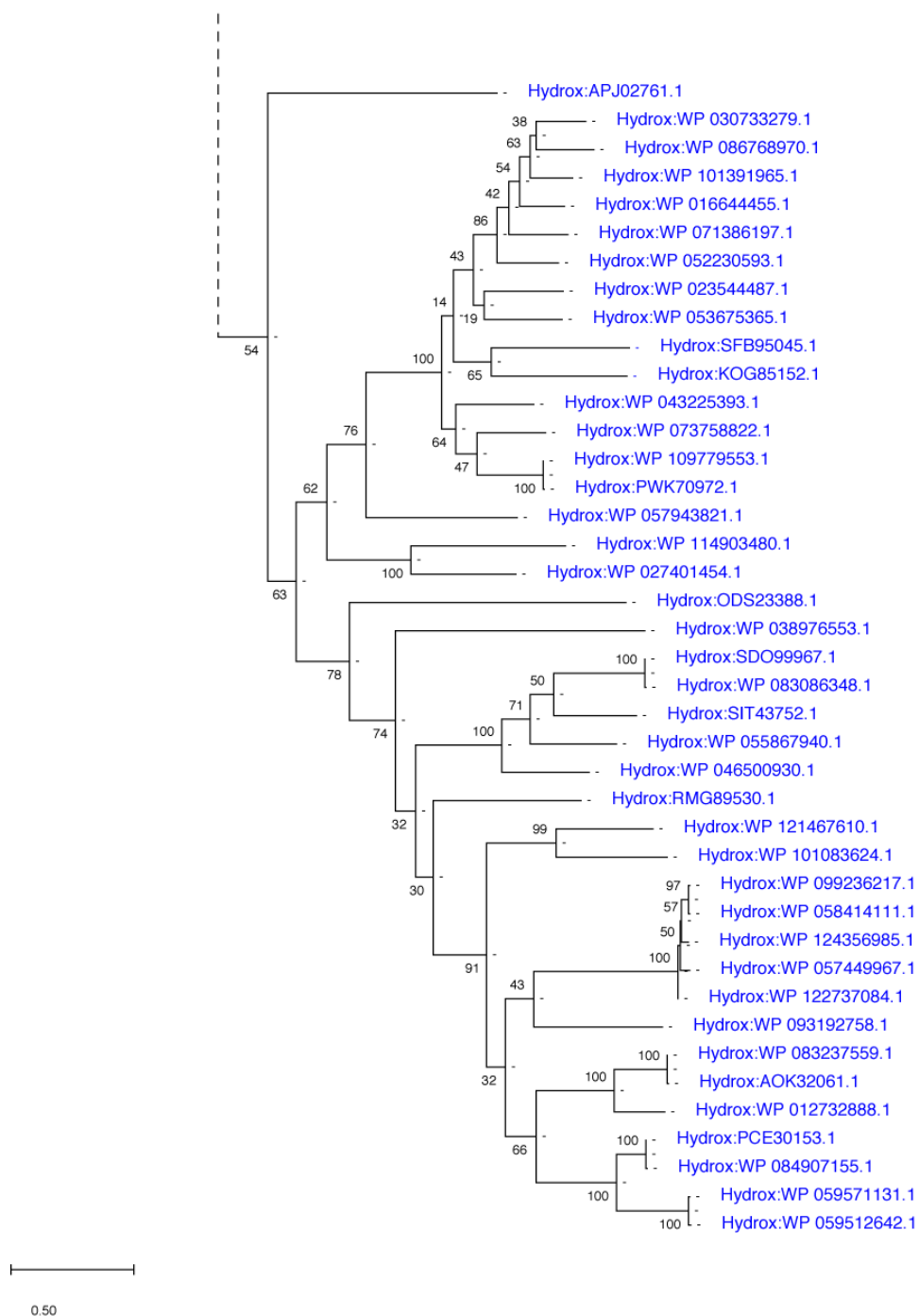


Figure A2.1 (continued). Phylogenetic tree of BesD homologs. The tree was generated using MEGA [4] from an alignment made in MUSCLE [5].



Hydrox D144G

```
MDVHEIDETLEKFLAENYTTPESVQQLADRFRQTGFVKFDSHMRIVPELITAVRAEADRLVREHKERRDLVLTGTTGTPRNLSVVKSQDV
Hal
MNVEQIDENLAKFLAERYTTPESVQQLADRFHRFGVKFDAANRLVPDELQTAVREECDLLIEQHKERNLLSTTGNTPRRMSVVKSEBI
C1_1
-----XXLAENYTTPESVRVQQLADRFQRTGFVKFDSHMRIVPELITAVRAEADRLVREHKERRDLVLTGTTGTPRNLSVVKSQDV
C1_2
---XXXXXDLXSKFLAERYTTPESVQQLADRFRQTGFVKFDSHMRIVPELITAVRAEADRLVREHKERRDLVLTGTTGTPRNLSVVKSQDV
C1_3
---XXXXXDLXSKFLAENYTTPESVRVQQLADRFQRTGFVKFDSHMRIVPELITAVRAEADRLVREHKERRDLVLTGTTGTPRNLSVVKSQDV
C1_4
---XXXXXXLRKFLAENYTTPESVRVQQLADRFQRTGFVKFDSHMRIVPELITAVRAEADRLVREHKERRDLVLTGTTGTPRNLSVVKSQDV
C1_5
---XXXXXTLXKFLAENYTTPESVRVQQLADRFQRTGFVKFDSHMRIVPELITAVRAEADRLVREHKERRDLVLTGTTGTPRNLSVVKSQDV
C1_6
---XXXXXTNLRXFLAENYTTPESVRVQQLADRFQRTGFVKFDSHMRIVPELITAVRAEADRLVREHKERRDLVLTGTTGTPRNLSVVKSQDV
C1_7
---XXXXXTNLRKFLAENYTTPESVRVQQLADRFQRTGFVKFDSHMRIVPELITAVRAEADRLVREHKERRDLVLTGTTGTPRNLSVVKSQDV
C1_8
---XXXXXTLXKFLAENYTTPESVRVQQLADRFQRTGFVKFDSHMRIVPELITAVRAEADRLVREHKERRDLVLTGTTGTPRNLSVVKSQDV
C1_9
---XXXXXDLXSKFLAERYTTPESVQQLADRFHRFGVKFDAANRLVPDELQTAVREECDLLIEQHKERNLLSTTGNTPRRMSVVKSEBI
C1_10
---XXXXXTNLXFLAENYTTPESVRVQQLADRFQRTGFVKFDAANRLVPDELQTAVREECDLLIEQHKERNLLSTTGNTPRRMSVVKSEBI
C1_11
---XXXXXTLXSKFLAERYTTPESVQQLADRFQRTGFVKFDSHMRIVPELITAVREECDLLIEQHKERNLLSTTGNTPRRMSVVKSEBI
C1_12
---XXXXXTLXSKFLAERYTTPESVQQLADRFQRTGFVKFDSHMRIVPELITAVREECDLLIEQHKERNLLSTTGNTPRRMSVVKSEBI
C1_13
---XXXXXTLXSKFLAERYTTPESVQQLADRFHRFGVKFDAANRLVPDELQTAVREECDLLIEQHKERRDLVLTSTTGNTPRRMSVVKSEBI
C1_14
---XXXXXDLXSKFLAERYTTPESVQQLADRFHRFGVKFDAANRLVPDELQTAVREECDLLIEQHKERNLLSTTGNTPRRMSVVKSEBI
C1_15
---XXXXXXLXSKFLAERYTTPESVQQLADRFHRFGVKFDAANRLVPDELQTAVREECDLLIEQHKERNLLSTTGNTPRRMSVVKSEBI
C1_16
---XXXXXTNLXFLAENYTTPESVRVQQLADRFHRFGVKFDAANRLVPDELQTAVREECDLLIEQHKERNLLSTTGNTPRRMSVVKSEBI
C1_17
---XXXXXTNLXFLAERYTTPESVQQLADRFHRFGVKFDAANRLVPDELQTAVREECDLLIEQHKERNLLSTTGNTPRRMSVVKSEBI
C1_18
---XXXXIDLXSKFLAERYTTPESVQQLADRFHRFGVKFDAANRLVPDELQTAVREECDLLIEQHKERNLLSTTGNTPRRMSVVKSEBI
C1_19
---XXXXXTLXSKFLAERYTTPESVQQLADRFHRFGVKFDAANRLVPDELQTAVREECDLLIEQHKERNLLSTTGNTPRRMSVVKSEBI
C1_20
---XXXXXTLXSKFLAERYTTPESVQQLADRFHRFGVKFDAANRLVPDELQTAVREECDLLIEQHKERNLLSTTGNTPRRMSVVKSEBI
C1_21
---XXXXXTLXSKFLAERYTTPESVQQLADRFHRFGVKFDAANRLVPDELQTAVREECDLLIEQHKERNLLSTTGNTPRRMSVVKSEBI
C1_22
---XXXXXTLXSKFLAERYTTPESVQQLADRFHRFGVKFDAANRLVPDELQTAVREECDLLIEQHKERNLLSTTGNTPRRMSVVKSEBI
C1_23
---XXXXXXLXSKFLAERYTTPESVQQLADRFHRFGVKFDAANRLVPDELQTAVREECDLLIEQHKERNLLSTTGNTPRRMSVVKSEBI
C1_24
---XXXXXDLXKFLAENYTTPESVRVQQLADRFHRFGVKFDSHMRIVPELITAVRAEADRLVREHKERNLLSTTGNTPRRMSVVKSEBI
C1_25
---XXXXXTLXSKFLAERYTTPESVRVQQLADRFQRTGFVKFDSHMRIVPELITAVRAEADRLVREHKERNLLSTTGNTPRRMSVVKSEBI
C1_26
---XXXXDLXSKFLAENYTTPESVRVQQLADRFQRTGFVKFDAANRLVPDELQTAVREECDLLIEQHKERNLLSTTGNTPRRMSVVKSEBI
C1_27
---XXXXXTLXFLAERYTTPESVQQLADRFHRFGVKFDAANRLVPDELQTAVREECDLLIEQHKERNLLSTTGNTPRRMSVVKSEBI
C1_28
---XXXXXTLXSKFLAERYTTPESVQQLADRFHRFGVKFDAANRLVPDELQTAVREECDLLIEQHKERRDLVLTGTTGTPRNLSVVKSQDV
C1_29
---XXXXXTLXSKFLAERYTTPESVQQLADRFHRFGVKFDAANRLVPDELQTAVREECDLLIEQHKERNLLSTTGNTPRNLSVVKSQDV
C1_30
---XXXXXXLXSKFLAERYTTPESVQQLADRFHRFGVKFDAANRLVPDELQTAVREECDLLIEQHKERNLLSTTGNTPRNLSVVKSQDV
*****
```

* * * * *

Figure A2.2 (continued). Halogenating shuffle library sequencing results. Sequence alignments of Hydrox D144G, Hal, and 30 halogenating hits from the shuffle screen (Cl_1 through Cl_30). Residues corresponding to the wild type Hal are shown in red. Residues corresponding to the wild type Hydrox are shown in blue. Residues that do not match either parent and have arisen through error prone assembly are shown in magenta.

[illegible]

[illegible]

Figure A2.3 (continued). Non-halogenating shuffle library sequencing results. Sequence alignment of Hydrox D144G, Hal, and 33 library of the members that did *not* yield fluorescence upon CuAAC of the supernatant (OH_1 through OH_33). Alignments were performed in MUSCLE [5]. 46 constructs were sequenced in total, but 13 were partial sequences, empty vector, or poor reads. Residues corresponding to the wild type Hal are shown in red. Residues corresponding to the wild type Hydrox are shown in blue. Residues that do not match either parent and have arisen through error prone assembly are shown in magenta.

Hydrox D144G

	181	191	201	211	221	231	241	251																												
Hal	NETL	CERQ	IDTY	GLV	SGD	LYLL	SDTT	MHRT	VPL	TED	GA	VR	TML	VVS	SA	ER	DL	GK	VL	TG	ND	R	W	W	N	E	A	A	Q	P	V	H	R	A	G	
OH_1																																				
OH_2																																				
OH_3																																				
OH_4																																				
OH_5																																				
OH_6																																				
OH_7																																				
OH_8																																				
OH_9																																				
OH_10																																				
OH_11																																				
OH_12																																				
OH_13																																				
OH_14																																				
OH_15																																				
OH_16																																				
OH_17																																				
OH_18																																				
OH_19																																				
OH_20																																				
OH_21																																				
OH_22																																				
OH_23																																				
OH_24																																				
OH_25																																				

Appendix 3: *Python scripts used for data analysis*

Script A3.1. Python script for annotation of BesD homologs as putative halogenases or hydroxylases based on the presence of the HXA/G (Halogenase) or HXD/E (Hydroxylase) sequence motifs.

```
###This file takes in an alignment (FASTA FORMAT) and the sequence position of the facial triad.
It then re- outputs the alignment labeled with 'Hal' or 'Hydrox' accordingly
import Bio
import csv
from Bio import AlignIO
from Bio.Align import MultipleSeqAlignment
from Bio.SeqRecord import SeqRecord
from Bio.Seq import Seq
from Bio.Alphabet import generic_dna
from Bio.Alphabet import generic_protein

#Specify the PFAM alignemnt file of interest and note the HXD positions.
myali = "CD-Hit-100-BlastE5_MUSCLE-manual_curation_deleting_gaps.txt" #input file
posH = 197 #position of His in HXD/E/A/E motif
posD = 199 #position of Asp/Glu/Ala/Gly in HXD/E/A/G motif
StorageFile = 'CD100-BlastE5May' #name of output file

#Read in the alignment
alignment = AlignIO.read(myali, "fasta")#reading in a single alignment with multiple sequences

#Print as a test to be sure the file is ok
print alignment

#Make a new alignment object to store putative hydroxylases
HalAlignment = MultipleSeqAlignment([])

count = 0 #Initiate the count for the number of HXG/As

HalAlign = [] #Create a list to store alignments that meet the criteria
Resi = [] #Create a list for storing the redidues that are not D/E

#Loop through the alignment and identify HXG/A residues
for record in alignment:
    #Print record[posH], record[posD]
    if record[posD] != 'D' and record[posD] != 'E' and record[posD] != '-' and record[posH] == 'H'
    and (record[posD] == 'A' or record[posD] == 'G'):
        count = count + 1 #count the hits
        #Change the record id to include label 'Hal'
        title = 'Hal:%s' %record.id
        record.id = title
        HalAlignment.append(record) #Add hits to the list
        Resi.append(record[posD])
        HalAlign.append(title) #Add the halogenase hits to a separate list
    else:
        #Change the record id to include label 'Hydrox'
        title = 'Hydrox:%s' %record.id
        record.id = title
        HalAlignment.append(record) #Add hits to the list

AlignIO.write(HalAlignment,StorageFile,"fasta") #Writes labels file into the storage file

print 'The number of putative halogenases HXG/A is:', count
print 'The halogenase alignment file has been printed to: ', StorageFile

#This code prints a list of the halogenases
thefile = open('CD100-BlastE5HalogenasesMay.txt', 'w')
for item in HalAlign:
    [accessionNo,seq,x] = item.partition('/')
    thefile.write("%s\n" %accessionNo)#Prints the label as well as the accession number
thefile.close()
```

Script A3.2. Python script for shuffle library analysis.

```
#This is step 1. Purpose: Translate each sequencing read into amino acids
#Input: save sequencing results directly as a multifasta file
import Bio
import csv
from Bio import AlignIO
from Bio import SeqIO
from Bio.Align import MultipleSeqAlignment
from Bio.SeqRecord import SeqRecord
from Bio.Seq import Seq
from Bio.Seq import MutableSeq
from Bio.Alphabet import generic_dna
from Bio.Alphabet import generic_protein

#Specify the file of interest
mySeq = '200201_Hydroxylases.txt'

#Specify the file where you will save translation output. CHANGE AS NEEDED
saveIt = '200517_Hydroxylases_TRANSLATE'

#Make an empty file for storing the sequence translations
sequences = []
seqIDs = []

#Go sequence by sequence to translate and check that it's in-frame using the HGWHWG motif
for record in SeqIO.parse(mySeq, "fasta"):
    addn = record.seq

    #Make a loop for trying all three possible frames for translation
    for x in range(0,3):
        temp = addn[x:] #Hold the translation at a given starting position
        tsn = temp.translate().split('*') #translate
        #Parse the split to find the reading frame we care about
        for i in tsn:
            if 'HGWHWG' in i: #Check if in-frame
                sequences.append(i)
                seqIDs.append(record.id)
                print i #Prints the in-frame read

#Take sequence file with translations. Export as fasta, then align in MUSCLE
thefile = open(saveIt, 'w')
count = 0
for item in sequences:
    thefile.write(">%s\n%s\n" % (seqIDs[count],item))
    count = count + 1
thefile.close()
```

```

#This is step 2 for the shuffle analysis
#Inputs are amino acid alignments from MUCSLE. Hal and Hydrox are provided for comparison
import Bio
import csv
from Bio import AlignIO
from Bio import SeqIO
from Bio.Align import MultipleSeqAlignment
from Bio.SeqRecord import SeqRecord
from Bio.Seq import Seq
from Bio.Alphabet import generic_dna
from Bio.Alphabet import generic_protein

myAlign = '200326_Hydroxylase_Translate_Ali.txt' #This is the input alignment file
HalAli = 'HalAli2.txt' #Sequence of the WT halogenase
HydroxAli = 'HydroxAli2.txt' #Sequence of the WT hydroxylase

saveIt = '200516_Hydroxylase_Results' #File name for where to save results

#Read in the fasta file of Alignment
alignment = AlignIO.read(myAlign, "fasta") #Reading in a single alignment with multiple sequences
Hal = AlignIO.read(HalAli, "fasta")[0] #Reading in the Hal sequence
Hydrox = AlignIO.read(HydroxAli, "fasta")[0] #Reading in the hydrox sequence
print alignment #To check proper import
print Hal #To check proper import
print Hydrox #To check proper import

count = 0 #Initialize the count
seqIDs = [] #For storing ID of hits
Counting = [0]*len(Hal) #Each position will count how many of each residue there is
#Now I have to check each sequence against the Hal and Hydrox. If it's the Hal residue, I will +1
#to the value stored at a given position
for record in alignment:
    current = record.seq
    seqIDs.append(record.id)
    holder = ''
    for i in range(0, len(current)):
        numHold = Counting[i]
        if current[i]==Hal.seq[i] and current[i]==Hydrox.seq[i]: #If hal and hydrox have same
                                                                    residue anyway
            Counting[i] = numHold+0 #Do not increase the count at that position
        elif current[i]==Hal.seq[i] and current[i]!=Hydrox.seq[i]: #Else if the residue
                                                                    matches Hal
            Counting[i] = numHold+1 #Add one to count the halogenase-like residues
        elif current[i]!=Hal.seq[i] and current[i]==Hydrox.seq[i]: #Else if the residue
                                                                    matches Hydrox
            Counting[i] = numHold+0 #Do not change the count
        else: #If there is an X there because of the sequencing reads gave 'N' due to low
            confidence base
            Counting[i] = numHold+0 #Do not change the count

print Counting #Print the results of the count at each position

thefile = open(saveIt, 'w') #Prepare to write in the file

for item in Counting:
    thefile.write("%s\n" %item)
thefile.close()

```


Appendix 4: *Strain tables for cloned constructs*

Table A4.1. Strain list for plasmids in Chapter 2.

Plasmid	Description	Strain
pET16b-His ₁₀ -BesA	His ₁₀ -BesA (T7), <i>lacI</i> , Cb ^R , ColE1	2680
pET16b-His ₁₀ -BesD	His ₁₀ -BesD (T7), <i>lacI</i> , Cb ^R , ColE1	3526
pET16b-His ₁₀ -BesE	His ₁₀ -BesE (T7), <i>lacI</i> , Cb ^R , ColE1	2677
pET16b-Strep-BesD	Strep-BesD (T7), <i>lacI</i> , Cb ^R , ColE1	2674
pET16b-Strep-BesD G139D	Strep-BesD G139D (T7), <i>lacI</i> , Cb ^R , ColE1	3524
pSV272.1-His ₆ MBP-BesC	His ₆ MBP-BesC (T7), <i>lacI</i> , Km ^R , f1, ColE1	3525
pET16b-His ₁₀ -BesC	His ₁₀ -BesC (T7), <i>lacI</i> , Cb ^R , ColE1	3489

Table A4.2. Strain list for plasmids in Chapter 3.

Plasmid	Description	Strain
pET16b-His ₁₀ -ScattleyaBesD	His ₁₀ -ScattleyaBesD (T7), <i>lacI</i> , Cb ^R , ColE1	3351
pET16b-His ₁₀ -SlavenduligriseusBesD	His ₁₀ -SlavenduligriseusBesD (T7), <i>lacI</i> , Cb ^R , ColE1	3352
pET16b-His ₁₀ -PfluorescensHalA	His ₁₀ -PfluorescensHalA (T7), <i>lacI</i> , Cb ^R , ColE1	2872
pET16b-His ₁₀ -PorientalisHalA	His ₁₀ -PorientalisHalA (T7), <i>lacI</i> , Cb ^R , ColE1	2910
pET16b-His ₁₀ -SwuyuanensisHalB	His ₁₀ -SwuyuanensisHalB (T7), <i>lacI</i> , Cb ^R , ColE1	2939
pET16b-His ₁₀ -StoyocaensisHalB	His ₁₀ -StoyocaensisHalB (T7), <i>lacI</i> , Cb ^R , ColE1	3527
pET16b-His ₁₀ -SviridosporusHalB	His ₁₀ -SviridosporusHalB (T7), <i>lacI</i> , Cb ^R , ColE1	3258
pET16b-His ₁₀ -AawajinensisHalB	His ₁₀ -AawajinensisHalB (T7), <i>lacI</i> , Cb ^R , ColE1	3529
pET16b-His ₁₀ -SgriseusHalB	His ₁₀ -SgriseusHalB (T7), <i>lacI</i> , Cb ^R , ColE1	3530
pET16b-His ₁₀ -SafghaniensisHalB	His ₁₀ -SafghaniensisHalB (T7), <i>lacI</i> , Cb ^R , ColE1	2913
pET16b-His ₁₀ -SiranensisHalB	His ₁₀ -SiranensisHalB (T7), <i>lacI</i> , Cb ^R , ColE1	2879
pET16b-His ₁₀ -SprunicolorHalB	His ₁₀ -SprunicolorHalB (T7), <i>lacI</i> , Cb ^R , ColE1	3531
pET16b-His ₁₀ -LanisaHalC	His ₁₀ -LanisaHalC(T7), <i>lacI</i> , Cb ^R , ColE1	2914
pET16b-His ₁₀ -PkilonensisHalD	His ₁₀ -PkilonensisHalD (T7), <i>lacI</i> , Cb ^R , ColE1	2873
pET16b-His ₁₀ -PSHC52HalD	His ₁₀ -PSHC52HalD (T7), <i>lacI</i> , Cb ^R , ColE1	2941
pET16b-His ₁₀ -PtrivialisHalD	His ₁₀ -PtrivialisHalD (T7), <i>lacI</i> , Cb ^R , ColE1	2909
pET16b-His ₁₀ -Proot562HalE	His ₁₀ -Proot562HalE (T7), <i>lacI</i> , Cb ^R , ColE1	2911
pET16b-His ₁₀ -PfulvaHalE	His ₁₀ -PfulvaHalE (T7), <i>lacI</i> , Cb ^R , ColE1	2878
pET16b-His ₁₀ -SpristinaespiralisHalF	His ₁₀ -SpristinaespiralisHalF (T7), <i>lacI</i> , Cb ^R , ColE1	2931
pET16b-His ₁₀ -MpelagiusHalG	His ₁₀ -MpelagiusHalG (T7), <i>lacI</i> , Cb ^R , ColE1	2915
pET16b-His ₁₀ -PcorrugataHalH	His ₁₀ -PcorrugataHalH (T7), <i>lacI</i> , Cb ^R , ColE1	3532
pET16b-His ₁₀ -PmediterraneaHalH	His ₁₀ -PmediterraneaHalH (T7), <i>lacI</i> , Cb ^R , ColE1	2874
pSV272.1-His ₆ MBP-Proot562HalE	His ₆ MBP-Proot562HalE (T7), <i>lacI</i> , Km ^R , f1, ColE1	3488
pET16b-His ₁₀ -SIBesD R74A	His ₁₀ -SIBesD R74A (T7), <i>lacI</i> , Cb ^R , ColE1	3534
pET16b-His ₁₀ -SIBesD W238A	His ₁₀ -SIBesD W238A (T7), <i>lacI</i> , Cb ^R , ColE1	3535
pET16b-His ₁₀ -SIBesD W239A	His ₁₀ -SIBesD W239A(T7), <i>lacI</i> , Cb ^R , ColE1	3536
pET16b-His ₁₀ -SIBesD H134A	His ₁₀ -SIBesD H134A (T7), <i>lacI</i> , Cb ^R , ColE1	3356
pET16b-His ₁₀ -SIBesD D140A	His ₁₀ -SIBesD D140A (T7), <i>lacI</i> , Cb ^R , ColE1	3355
pET16b-His ₁₀ -SIBesD N219A	His ₁₀ -SIBesD N219A (T7), <i>lacI</i> , Cb ^R , ColE1	3354
pET16b-His ₁₀ -SIBesD E120A	His ₁₀ -SIBesD E120A (T7), <i>lacI</i> , Cb ^R , ColE1	3358

pET16b-His ₁₀ -SIBesD T221A	His ₁₀ -SIBesD T221A (T7), <i>lacI</i> , Cb ^R , ColE1	3353
pET16b-His ₁₀ -RapL	His ₁₀ -RapL (T7), <i>lacI</i> , Cb ^R , ColE1	3344
pET16b-His ₁₀ -OCD	His ₁₀ -OCD (T7), <i>lacI</i> , Cb ^R , ColE1	3343
pET16b-His ₁₀ -LDC	His ₁₀ -LDC (T7), <i>lacI</i> , Cb ^R , ColE1	3332
pET16b-His ₁₀ -IlvE	His ₁₀ -IlvE (T7), <i>lacI</i> , Cb ^R , ColE1	3334
IVTT-PepC	PepC (T7), Cb ^R	3275
IVTT-PepA	PepA (T7), Cb ^R	3456
pET16b- His ₁₀ -PrescissionCutSite-IMDPH	His ₁₀ -PrescissionCutSite-IMDPH (T7), <i>lacI</i> , Cb ^R , ColE1	2378

Table A4.3. Strain list for plasmids in Chapter 4.

Plasmid	Description	Strain
pET16b-His ₁₀ -Hydrox	His ₁₀ -Hydrox (T7), <i>lacI</i> , Cb ^R , ColE1	3336
pET16b-His ₁₀ -Hal	His ₁₀ -Hal (T7), <i>lacI</i> , Cb ^R , ColE1	3335
pET16b-His ₁₀ -Hal_opt	His ₁₀ -Hal_opt (T7), <i>lacI</i> , Cb ^R , ColE1	3533
pET16b-His ₁₀ -Hydrox D144G	His ₁₀ -Hydrox D144G (T7), <i>lacI</i> , Cb ^R , ColE1	3339
pET16b-His ₁₀ -Hydrox D144G V225N	His ₁₀ -Hydrox D144G V225N (T7), <i>lacI</i> , Cb ^R , ColE1	3342
pET16b-His ₁₀ -Hydrox D144G N151I	His ₁₀ -Hydrox D144G N151I(T7), <i>lacI</i> , Cb ^R , ColE1	3449
pET16b-His ₁₀ -Hydrox D144G N151I V225N	His ₁₀ -Hydrox D144G N151I V225N (T7), <i>lacI</i> , Cb ^R , ColE1	3445
pET16b-His ₁₀ -Hydrox D144G N151I V225N F150L	His ₁₀ -Hydrox D144G N151I V225N F150L (T7), <i>lacI</i> , Cb ^R , ColE1	3480
pET16b-His ₁₀ -Hydrox D144G N151I V225N GAV218-220 to ST218-219	His ₁₀ -Hydrox D144G N151I V225N GAV218-220 to ST218-219 (T7), <i>lacI</i> , Cb ^R , ColE1	3481
pET16b-His ₁₀ -Hydrox D144G N151I V225N M223I	His ₁₀ -Hydrox D144G N151I V225N M223I(T7), <i>lacI</i> , Cb ^R , ColE1	3482
pET16b-His ₁₀ -Hydrox D144G N151I V225N V226M S227T	His ₁₀ -Hydrox D144G N151I V225N V226M S227T (T7), <i>lacI</i> , Cb ^R , ColE1	3483
pET16b-His ₁₀ -Hydrox D144G N151I V225N S228A	His ₁₀ -Hydrox D144G N151I V225N S228A (T7), <i>lacI</i> , Cb ^R , ColE1	3484
pET16b-His ₁₀ -Hydrox D144G N151I V225N E231K G235E V237D	His ₁₀ -Hydrox D144G N151I V225N E231K G235E V237D (T7), <i>lacI</i> , Cb ^R , ColE1	3486
pET16b-His ₁₀ -Chimera	His ₁₀ -Chimera (T7), <i>lacI</i> , Cb ^R , ColE1	3473
pET16b-His ₁₀ -Chimera L150F	His ₁₀ -Chimera L150F (T7), <i>lacI</i> , Cb ^R , ColE1	3474
pET16b-His ₁₀ -Chimera ST218-219GAV	His ₁₀ -Chimera ST218-219GAV (T7), <i>lacI</i> , Cb ^R , ColE1	3475
pET16b-His ₁₀ -Chimera I222M	His ₁₀ -Chimera I222M (T7), <i>lacI</i> , Cb ^R , ColE1	3476
pET16b-His ₁₀ -Chimera M225V T226S	His ₁₀ -Chimera M225V T226S (T7), <i>lacI</i> , Cb ^R , ColE1	3477
pET16b-His ₁₀ -Chimera A228S	His ₁₀ -Chimera A228S (T7), <i>lacI</i> , Cb ^R , ColE1	3478
pET16b-His ₁₀ -Chimera K230E E234G D236V	His ₁₀ -Chimera K230E E234G D236V (T7), <i>lacI</i> , Cb ^R , ColE1	3479
pET16b- His ₁₀ -PrescissionCutSite-IMDPH	His ₁₀ -PrescissionCutSite-IMDPH (T7), <i>lacI</i> , Cb ^R , ColE1	2378
pACYC-DUET-BesC.MBP-BesB	BesC (T7) and MBP-BesB (T7) from <i>P. fluorescens</i> , <i>lacI</i> , Cm ^R , ColE1	3345

Table A4.4. Strain list for untested plasmids related to Chapter 4.

Plasmid	Description	Strain
pET16b-His ₁₀ -Hal I151N	His ₁₀ -Hal I151N (T7), <i>lacI</i> , Cb ^R , ColE1	3502
pET16b-His ₁₀ -Hal G144D	His ₁₀ -Hal G144D (T7), <i>lacI</i> , Cb ^R , ColE1	3337
pET16b-His ₁₀ -Hal N225V	His ₁₀ -Hal N225V (T7), <i>lacI</i> , Cb ^R , ColE1	3338
pET16b-His ₁₀ -Hal G144D I151N	His ₁₀ -Hal G144D I151N (T7), <i>lacI</i> , Cb ^R , ColE1	3503
pET16b-His ₁₀ -Hal G144D L150F I151N	His ₁₀ -Hal G144D L150F I151N (T7), <i>lacI</i> , Cb ^R , ColE1	3504
pET16b-His ₁₀ -Hydrox N151I	His ₁₀ -Hydrox N151I (T7), <i>lacI</i> , Cb ^R , ColE1	3516
pET16b-His ₁₀ -Hydrox V225N N151I	His ₁₀ -Hydrox V225N N151I (T7), <i>lacI</i> , Cb ^R , ColE1	3517
pET16b-His ₁₀ -SiBesD N219V	His ₁₀ -SiBesD N219V (T7), <i>lacI</i> , Cb ^R , ColE1	3518
pET16b-His ₁₀ -SiBesD I151N	His ₁₀ -SiBesD I151N (T7), <i>lacI</i> , Cb ^R , ColE1	3519
pET16b-His ₁₀ -Hal G144D I151N N225V	His ₁₀ -Hal G144D I151N N225V (T7), <i>lacI</i> , Cb ^R , ColE1	3505
pET16b-His ₁₀ -Hal I151N N225V	His ₁₀ -Hal I151N N225V (T7), <i>lacI</i> , Cb ^R , ColE1	3507
pET16b-His ₁₀ -Hal G144D N225V	His ₁₀ -Hal G144D N225V (T7), <i>lacI</i> , Cb ^R , ColE1	3341
pET16b-His ₁₀ -Hal I151A	His ₁₀ -Hal I151A (T7), <i>lacI</i> , Cb ^R , ColE1	3520
pET16b-His ₁₀ -Hal I151V	His ₁₀ -Hal I151V (T7), <i>lacI</i> , Cb ^R , ColE1	3521
pET16b-His ₁₀ -Hal G144D L150F I151N N225V	His ₁₀ -Hal G144D L150F I151N N225V (T7), <i>lacI</i> , Cb ^R , ColE1	3506
pET16b-His ₁₀ -Hydrox V225N	His ₁₀ -Hydrox V225N(T7), <i>lacI</i> , Cb ^R , ColE1	3340

Table A4.5. Strain list for plasmids in Chapter 5.

Plasmid	Description	Strain
pET16b-His ₁₀ -SlavenduligriseusBesD	His ₁₀ -SlavenduligriseusBesD (T7), <i>lacI</i> , Cb ^R , ColE1	3352
pET16b-His ₁₀ -SwuyuanensisHalB	His ₁₀ -SwuyuanensisHalB (T7), <i>lacI</i> , Cb ^R , ColE1	2939
pET16b- His ₁₀ -PrescissionCutSite-IMDPH	His ₁₀ -PrescissionCutSite-IMDPH (T7), <i>lacI</i> , Cb ^R , ColE1	2378
pET16b-His ₁₀ -PkilonensisHalD	His ₁₀ -PkilonensisHalD (T7), <i>lacI</i> , Cb ^R , ColE1	2873
pACYC-DUET-BesC.MBP-BesB	BesC (T7) and MBP-BesB (T7) from <i>P. fluorescens</i> , <i>lacI</i> , Cm ^R , ColE1	3345

Appendix References

1. Gust, B., Chandra, G., Jakimowicz, D., Yuqing, T., Bruton, C. J. & Chater, K. F. λ Red-Mediated Genetic Manipulation of Antibiotic-Producing *Streptomyces*. *Advances in applied microbiology*. **54**, 107–128 (2004) doi:10.1016/S0065-2164(04)54004-2.
2. Gatto, G. J., Boyne, M. T., Kelleher, N. L. & Walsh, C. T. Biosynthesis of pipecolic acid by RapL, a lysine cyclodeaminase encoded in the rapamycin gene cluster. *J. Am. Chem. Soc.* **128**, 3838–3847 (2006) doi:10.1021/ja0587603.
3. Marchand, J. A., Neugebauer, M. E., Ing, M. C., Lin, C.-I., Pelton, J. G. & Chang, M. C. Y. Discovery of a pathway for terminal-alkyne amino acid biosynthesis. *Nature* **567**, 420–424 (2019) doi:10.1038/s41586-019-1020-y.
4. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018) doi:10.1093/molbev/msy096.
5. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **5**, 113 (2004) doi:10.1186/1471-2105-5-113.