THE UNIVERSITY OF CHICAGO

ENGINEERING Fe(II)- AND α-KETOGLUTARATE-DEPENDENT OXYGENASES FOR SELECTIVE C–H FUNCTIONALIZATION

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF CHEMISTRY

BY

CHRISTIAN GOMEZ

CHICAGO, ILLINOIS

AUGUST 2023

In memory of my “abuela”, Bernarda Perez

1925 - 2018

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

APPENDIX I

APPENDIX II

LIST OF SCHEMES

CHAPTER 1

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

LIST OF TABLES

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 5

APPENDIX III

ACKNOWLEDGEMENTS

I would like to thank those who have contributed to the scientific findings described in this dissertation. I thank my advisor, Prof. Jared Lewis, who has assisted me in navigating my time as a graduate student and whose mentorship has helped me grow as a scientist. My dissertation committee members, Prof. John Anderson and Prof. Joseph Piccirilli, have provided insightful suggestions and comments regarding my research. Members of the Lewis lab Dr. Natalie Chan, Dr. Dibyendu Mondal, Qian Du, Anusha Manganahalli, Jake Gaines, Harumi Shimano, and Ayantika Ghosh have contributed to my graduate research, and I thank them for their work. Finally, our external collaborators, Prof. Heather Kulik, Prof. Philip Romero, Dr. Sameer D'Costa, and Vyshnavi Vennelakanti, have improved my research with their contributions.

I would like to also thank all members of the Lewis lab, both former and current, since they have impacted my development as a scientist in many ways. I thank Dr. Ketaki Belsare, Dr. Mary Andorfer, Dr. Brian Fisher, Dr. Krysten Jones, and Alan Swartz for mentoring me in different areas during the early years of my graduate education. I would also like to thank the rest of the group members I met when I joined the University of Chicago, Dr. Joseph Gair, Dr. Yifan Gu, Dr. Kat Ellis-Guardiola, Dr. Natalie Chan, Dr. David Upp, Dr. Harrison Snodgrass, Atreyi Bhattacharya, and Paul Butkovich, for providing a welcoming environment. I thank members that joined the group at Indiana University, Dr. Rui Huang, Dr. Dibyendu Mondal, Dr. Xinhang Yang, Dr. Bingqing Liu, Dr. Jianbin Li, Dr. Vikas Thakur, Dr. Autumn Kim, Dr. Dipankar Sahoo, Yasmine Zubi, Yuhua Jiang, Anusha Manganahalli, Qian Du, Amardeep Kumar, Caitlin Roof, Payal, Prabir Saha, Jake Gaines, Ayantika Ghosh, and Harumi Shimano. It has been a pleasure to study and work alongside these people.

The work presented in this dissertation would not have been possible without the personal support of friends and family. I am grateful for the friends I made during my time at the University of Chicago, who provided support during challenging times, including Dr. Benjamin Slaw, Dr. Tessa Lynch-Colameta, Samantha Duarte, Dr. Charles Cole, Dr. Laura Tociu, and Dr. Julia Murphy. In particular, I want to thank Atreyi Bhattacharya and Dr. Harrison Snodgrass for their continuing friendship and support, and I am looking forward to making more memories together. I also thank Dr. Brian Fisher, Dr. David Upp, and Dr. Natalie Chan for the countless memories during our time together in the Lewis Lab, including our transition to Indiana University. Moving from Chicago was a challenging process, but I am grateful for the friends I made in Bloomington, including Dr. Edward Sheetz, Dr. Jake Huerfano, Dr. Veronica Carta, Ana Sánchez-Arroyo, and Jan Bents. I thank everyone who supported me, and I appreciate every interaction we had, including dinners, trips, movies, and D&D nights. I would also like to thank those friends from before graduate school. My childhood friends, Gabriel Bianchini, Natalia Geloz, Tamara Juarez, Natalin Laurencio, and Anabel Pangiarella have constantly brought joy to my life, and I will always be grateful for them and their families. My college friends, in particular Yamila Pérez Sirkin and Agustin Millet, have been crucial in my personal and professional development, so I thank them.

I am very grateful for my family and their constant support. I would not be where I am today without them. My parents, Flavia Lafuente and Julio Gomez, have supported me and showed me unconditional love. I thank them for all the opportunities they have provided for me. I thank my brother, Alejandro Gomez, for being a constant role model. My sisters, Maylen Avila, Sayen Avila, and Valentina Gomez are an endless source of joy, and I will be forever grateful for their presence in mi life. I would also like to thank my grandparents, Bernarda Perez, Shirley

Barrios, and Carlos Lafuente, for all their love and support. I thank Alejandro Avila and Florencia Barreto for being wonderful additions to my family. I am very thankful to my aunt, great-aunt, uncles, and cousins, for making my life better by being part of it. Being away from my family has been one of the most challenging aspects of my graduate studies, and no words can describe how much I have missed them. Finally, I would like to thank my partner, Julian Morris, who has been my biggest supporter for the past five years. His love and understanding through these challenging times have been crucial to my success, and I will be forever grateful for that.

ABSTRACT

This dissertation describes the development and characterization of engineered Fe(II)- and α-ketoglutarate-dependent oxygenases (FeDOs) as site-selective biocatalysts for native and non-native C–H functionalization reactions. These enzymes provide an alternative to conventional organic synthesis by acting on unactivated C–H bonds with high selectivities under aqueous and mild conditions. Functionalization occurs via a C–H abstraction step followed by radical rebound of the new functional group. Previous attempts at non-native catalysis using FeDOs relied on the presence of an exogenous ligand that participates in the rebound step. Controlling the selectivity of rebound between native and non-native groups has been challenging, with most cases of non-native FeDO biocatalysis resulting in large amounts of the native product. The study presented herein addresses this challenge by engineering FeDOs with novel levels of selectivity towards non-native reactions.

Chapter 1 provides a general background of site-selective enzymatic C–H functionalization, protein engineering, and directed evolution. Additionally, the family of FeDOs is introduced and described, with a focus on hydroxylating FeDOs and Fe(II)- α-ketoglutarate-dependent halogenases (FeDHs). Finally, several examples are presented in which these enzymes have been used as site-selective biocatalysts, with the goal of depicting the current state of the field and its challenges.

Chapter 2 introduces the hydroxylase SadA and its variant SadA D157G. SadA D157G had been previously reported to carry out non-native chlorination of *N*-succinyl-L-leucine, albeit with poor chemoselectivity. This variant was engineered for improved expression via MBP-fusion to give SadX. Additionally, the activity of SadX and SadX variants in the presence of

different exogenous anions is characterized. Finally, the reaction of SadX in the presence of NaOCN is shown to enable carbamate formation.

Chapter 3 discusses the directed evolution of SadX for non-native reactivities, including chlorination and azidation. Conditions for screening SadX libraries in the presence of exogenous anions are established. This effort resulted in SadX variants with an increased chemoselectivity towards azidation of *N*-succinylated compounds. The resulting evolution campaign also offered insight into some of the underlying features of the SadX scaffold that control activity, chemoselectivity, and site selectivity.

Chapter 4 focuses on exploring the native hydroxylase activity of SadA. An improved hydroxylase with higher activity on the model substrate *N*-succinyl-L-leucine is described. Additionally, the effect of mutations accumulated during the directed evolution described in Chapter 3 is evaluated when these mutations are introduced in the wild-type SadA scaffold. A series of engineered hydroxylases is obtained, with notable levels of activity and selectivity on a variety of *N*-succinylated compounds.

Chapter 5 describes a novel methodology to incorporate next-generation sequencing and machine learning into directed evolution, with a focus on cost-efficient techniques with low experimental burden. This workflow is validated by applying it to the directed evolution libraries discussed in Chapter 3. A small set of predicted sequences is experimentally evaluated, resulting in a variant with improved azidation activity.

Chapter 1: Introduction

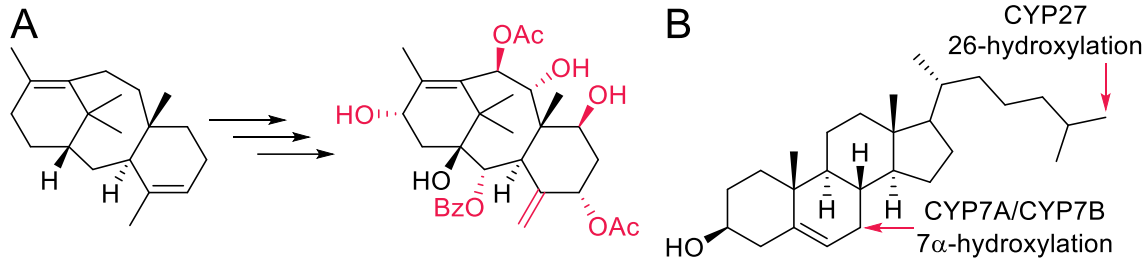
1.1 Site-selective enzymatic C–H functionalization

The direct functionalization of C–H bonds represents an attractive strategy in organic synthetic chemistry. This approach could remove the need for pre-functionalization of starting materials, reduce the number of functional group manipulation steps, improve atom economy, and provide new routes to build up structural complexity. However, given the energetic characteristics and ubiquity of C–H bonds in organic molecules, these transformations present the challenges of having to activate stable bonds and functionalize them in a selective manner. Most common approaches to overcome these challenges involve the use of organometallic catalysts that coordinate to a preexisting directing group on the starting material to guide the functionalization of a specific C–H bond.[1] Other approaches exploit differences in acidity, bond-dissociation energy, and other electronic effects.[2] Although these methods provide ways to expand the scope of C–H functionalization reactions, they rely on preexisting features of the starting material to direct selectivity.

The most effective catalysts for non-directed C–H functionalization reactions can be found in nature in the form of enzymes.[3] These biocatalysts are common in the biosyntheses of primary and secondary metabolites and can exert control over the selectivity of these transformations. One of the most cited examples that showcases the remarkable potential of site selective enzymes is the biosynthesis of the diterpenoid paclitaxel (Taxol®), in which multiple secondary, tertiary, and allylic C–H bonds in the precursor taxadiene are differentiated (Scheme

1.1A).[4,5] Highly selective hydroxylation reactions are common in steroid biosynthesis and metabolism, where enzymes are able to distinguish multiple C–H bonds with similar intrinsic reactivities (Scheme 1.1B).[6] This type of outstanding site selectivity is achieved in the absence of directing groups through complex and dynamic molecular recognition that results from multiple interactions between the respective substrates and enzyme active sites.

Scheme 1.1: Examples of site selectivity in enzyme catalysis



A) Site-selective steps in the transformation of taxadiene to baccatin III as part of the biosynthesis of paclitaxel.[4,5] B) Site- and enantioselective cytochrome P450 hydroxylases involved in distinct functionalization steps as part of primary bile acid synthesis from cholesterol.[7]

The field of biocatalysis in general not only grants access to high levels of selectivity, but also provides benefits from environmental and economical points of view when compared to conventional chemocatalysis.[8–11] Enzymes are obtained via heterologous protein expression in hosts such as *Escherichia coli*,[12] offering a lower cost of production and a lower environmental impact due to the renewable nature of the process. Additionally, it is frequent that enzyme-catalyzed reactions are performed in aqueous media instead of organic solvents, making the process more amenable to industrial scale and reducing waste disposal costs. Finally, the replacement of catalysts that rely on precious metals by enzymes represents a significant sustainability and economical benefit.

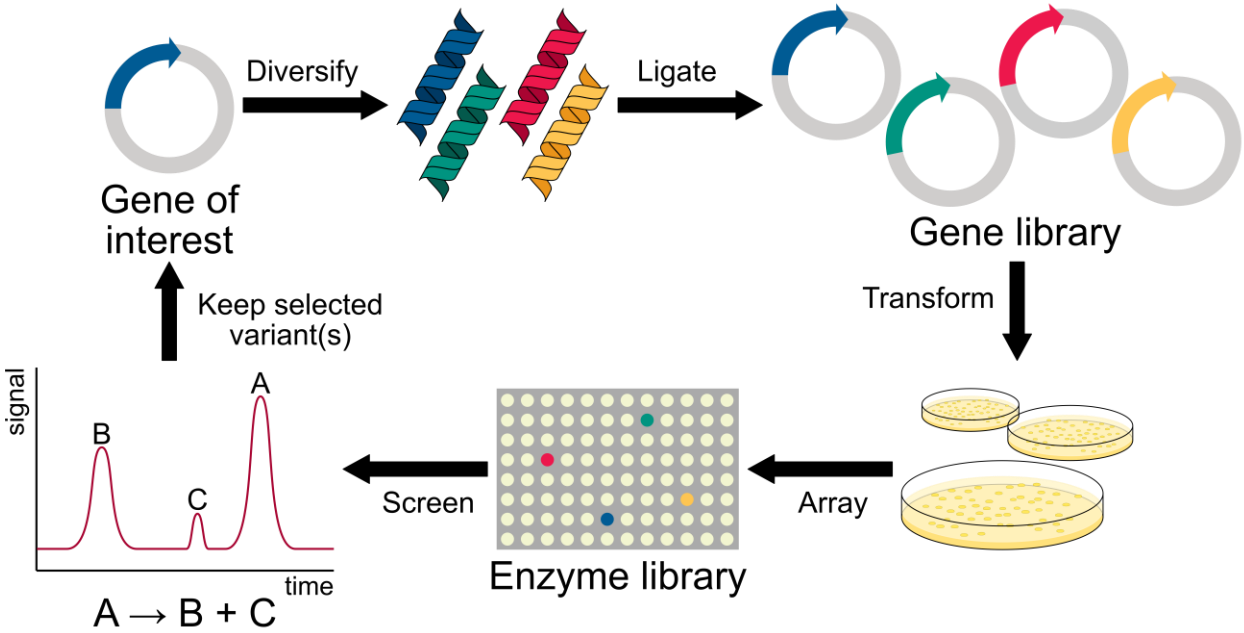
1.2 Protein engineering and directed evolution

Despite all the advantages to biocatalysis described above, the field has had to overcome several challenges.[13,14] The earlier examples of wild-type enzymes used for chemistry were

considered to have narrow substrate scope and low stability when compared with conventional catalysts. In the 1980s, advances in protein engineering based on enzyme structure information led to improvements that addressed some of the inherent challenges of enzyme catalysis. In the 1990s, new protein engineering and high-throughput screening techniques resulted in the development of directed evolution, pioneered by Frances Arnold and co-workers.[15]

Directed evolution consists of iterative rounds of gene diversification, protein expression, and screening, during which the desired feature is gradually improved (Figure 1.1). Although relevant enzymatic information such as structure and mechanism can be used to inform directed evolution campaigns, it is not strictly needed. Non-targeted approaches to the genetic diversification step can help identify beneficial mutations that would have been hard to predict and provide information about the mutational landscape of a given biocatalyst.[16,17] Early examples of successful directed evolution campaigns focused on evolving hydrolases,[18] but myriad other chemistries have been improved by this technique, through improvements in catalytic activity, thermostability, substrate scope, and selectivity.[19,20]

Figure 1.1: General scheme of directed evolution



The gene encoding for the protein of interest is subjected to a diversification step, in which a library of mutants is formed. This library is then introduced into a host organism (e.g. *Escherichia coli*), overexpressed, and screened for the desired feature. Variants that show the desired level of improvement are used as the starting point for the next iteration.

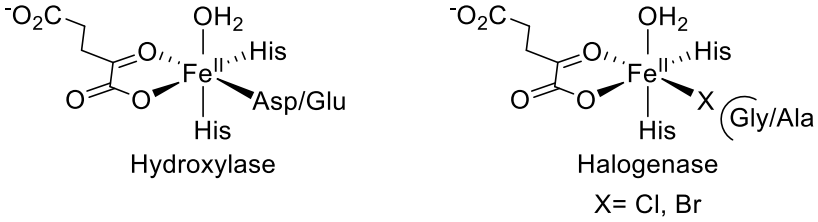
1.3 Introduction to Fe(II)- and α-ketoglutarate-dependent oxygenases

The superfamily of Fe(II)- and α-ketoglutarate-dependent oxygenases (FeDOs) consists of enzymes with a broad range of oxidative reactivities, including hydroxylation, halogenation, desaturation, epimerization, ring closure, ring expansion, and epoxidation.[21,22] The earliest example of a characterized FeDO was an animal prolyl hydroxylase involved in collagen biosynthesis[23]. Since then, FeDOs involved in DNA and RNA modification, antibiotic and natural product biosynthesis, and post-translational protein modification have been identified in bacteria, fungi, plants, and vertebrates.[21,24–27] This introduction will hereafter focus on the hydroxylase and halogenase subgroups given their ability to carry out C–H functionalization reactions and their relevance to the present work.

Structural characterization of FeDOs established the presence of a conserved double-stranded β-helix or cupin fold, inside of which a metal-binding motif weakly coordinates an

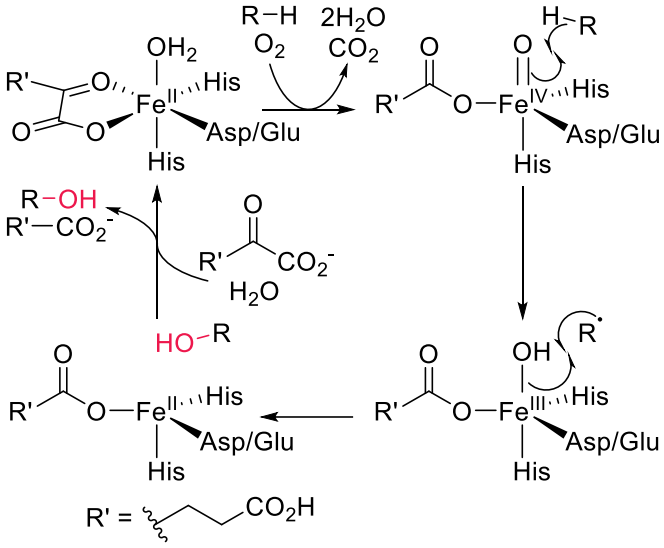
Fe(II) ion (Figure 1.2).[28] This highly-conserved motif is a facial triad consisting of two histidine residues and an aspartate or glutamate residue in a His–X–Asp/Glu–X_n–His sequence. Fe(II)- and α-ketoglutarate-dependent halogenases (FeDHs) contain a variation of this facial triad in which the aspartate/glutamate residue is instead glycine or alanine.[29] In both cases, α-ketoglutarate acts as a co-substrate, coordinating the Fe(II) with its C-1 carboxylate and C-2 ketone, and gets oxidized to succinate and CO_2 in the reaction.

Figure 1.2: Conserved FeDO and FeDH active sites



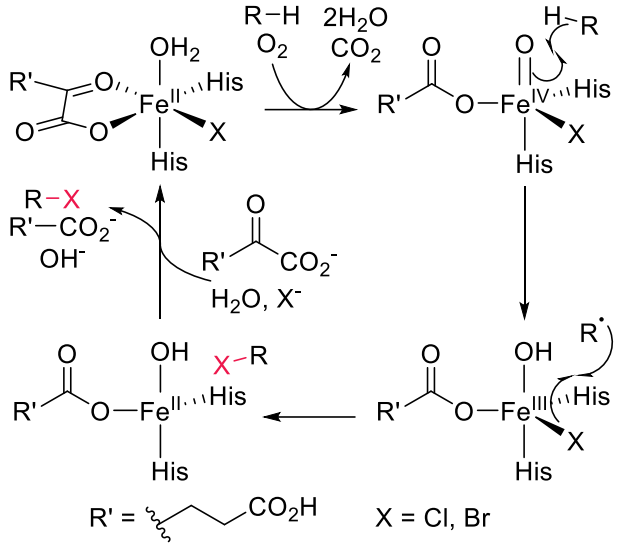
The consensus hydroxylation mechanism involves displacement of a water molecule by substrate binding in the active site followed by O_2 binding and activation to form a Fe(III)-superoxo intermediate (Figure 1.3).[30] Oxidative decarboxylation of α-ketoglutarate results in the formation of a high-valent Fe(IV)-oxo intermediate. This highly reactive intermediate is responsible for the hydrogen abstraction of a C–H bond in the substrate. The resulting radical substrate and Fe(III)-hydroxo intermediate react through a radical rebound step to form the hydroxylated product and an Fe(II) center.

Scheme 1.2: Consensus mechanism for FeDO-catalyzed hydroxylation



The presence of the smaller glycine/alanine residue in the binding motif of FeDHs results in an open coordination site that is occupied by a halide ligand.[31,32] The mechanism of FeDO-catalyzed halogenation follows the same path described for hydroxylases until the C–H abstraction step by the Fe(IV)-oxo intermediate. Hypothetically, the substrate radical could react with either hydroxyl or halide ligands via radical rebound. Native FeDHs, however, have evolved to favor a halide radical rebound step, resulting in the corresponding halogenated product. This control over chemoselectivity (i.e. halogenation against hydroxylation) has been linked to careful substrate positioning in the active site, placing the C–H bond to be functionalized in a way that disfavors rebound with the hydroxyl ligand once the radical substrate is formed.[33,34]

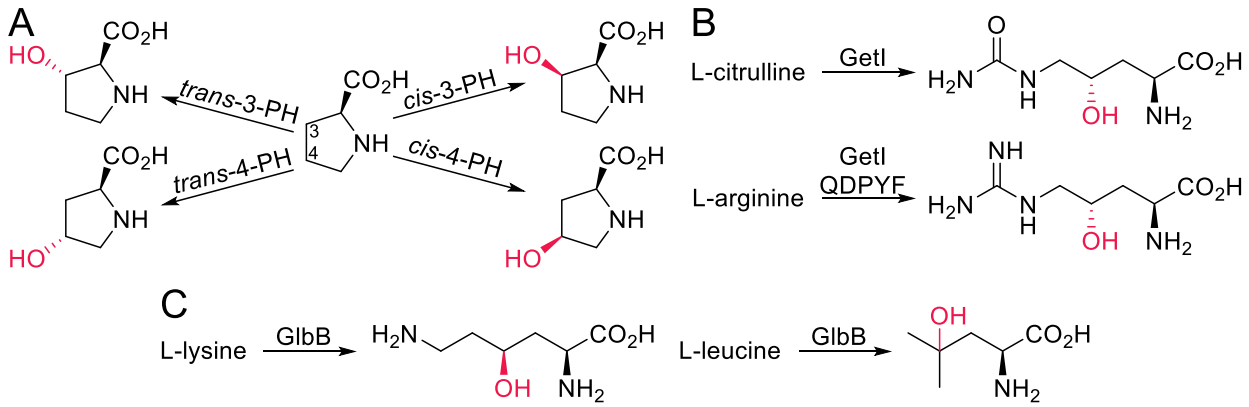
Scheme 1.3: Consensus mechanism for FeDH-catalyzed halogenation



1.4 Biocatalysis and Fe(II)- and α-ketoglutarate-dependent oxygenases

Hydroxylating FeDOs have been used most frequently as site-selective biocatalyst in the context of amino acid hydroxylation. However, when compared with another highly developed family of Fe(II)-containing hydroxylating enzymes, cytochromes P450, FeDOs tend to suffer from narrower substrate scope and often act only on compounds similar to their native substrates. For example, bacterial and fungal L-proline hydroxylases can selectively produce all four isomers of 3- and 4-hydroxy-L-proline (Scheme 1.4A) and have shown activity on related cyclic amino acids such as L-pipecolic acid.[35,36] The native L-citrulline hydroxylase GetI was engineered via the introduction of five mutations to selectively hydroxylate L-arginine and produce 4-hydroxy-L-arginine.[37] Lysine hydroxylases have also been widely characterized, with GlbB being able to hydroxylate the non-native substrate L-leucine.[38]

Scheme 1.4: Examples of site-selective FeDO-catalyzed hydroxylation reactions



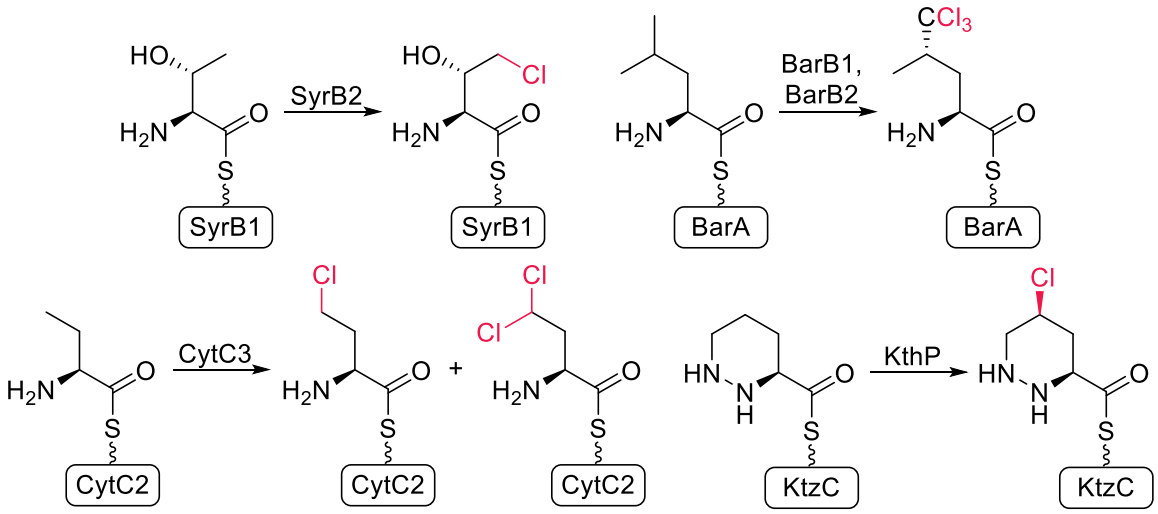
A) Proline hydroxylation by different proline hydroxylases (PH) with distinct site- and enantioselectivity.[35] B) Engineered L-citrulline hydroxylase GetI accepts non-native substrate L-arginine.[37] C) Lysine hydroxylase GlbB also shows activity on L-leucine.[38]

While enzymes such as flavin-dependent halogenases and vanadiuym-dependent haloperoxidases can catalyze halogenation of C–H bonds, their substrate scope in such reactions is limited to electron-rich aromatic compounds.[39] The halogenation of unactivated aliphatic C–H bonds is characteristic of FeDHs, making them a distinct and attractive platform to access this and other types of aliphatic C–H functionalization reactions. Since the first FeDH[40] was characterized almost four decades after the report of the first prolyl hydroxylase[23], the field of enzymatic aliphatic halogenation is in an earlier stage than other halogenases, and even other FeDOs. Despite this, significant advances have taken place in the last two decades, and relevant ones will be highlighted next.

The first characterized FeDH was SyrB2,which catalyzes chlorination of a L-threonine moiety in the biosynthesis of syringomycin E by *Pseudomonas syringae* (Scheme 1.5).[40] SyrB2 requires a carrier protein partner, SyrB1, to bring the thioester linked L-threonine to the active site, a feature shared among the early examples of characterized FeDHs. The crystal structure of SyrB2 provided meaningful insights into the mechanism of FeDHs early on, making it one of the best studied enzymes of this class.[41] The activity of other FeDHs in conjunction with carrier

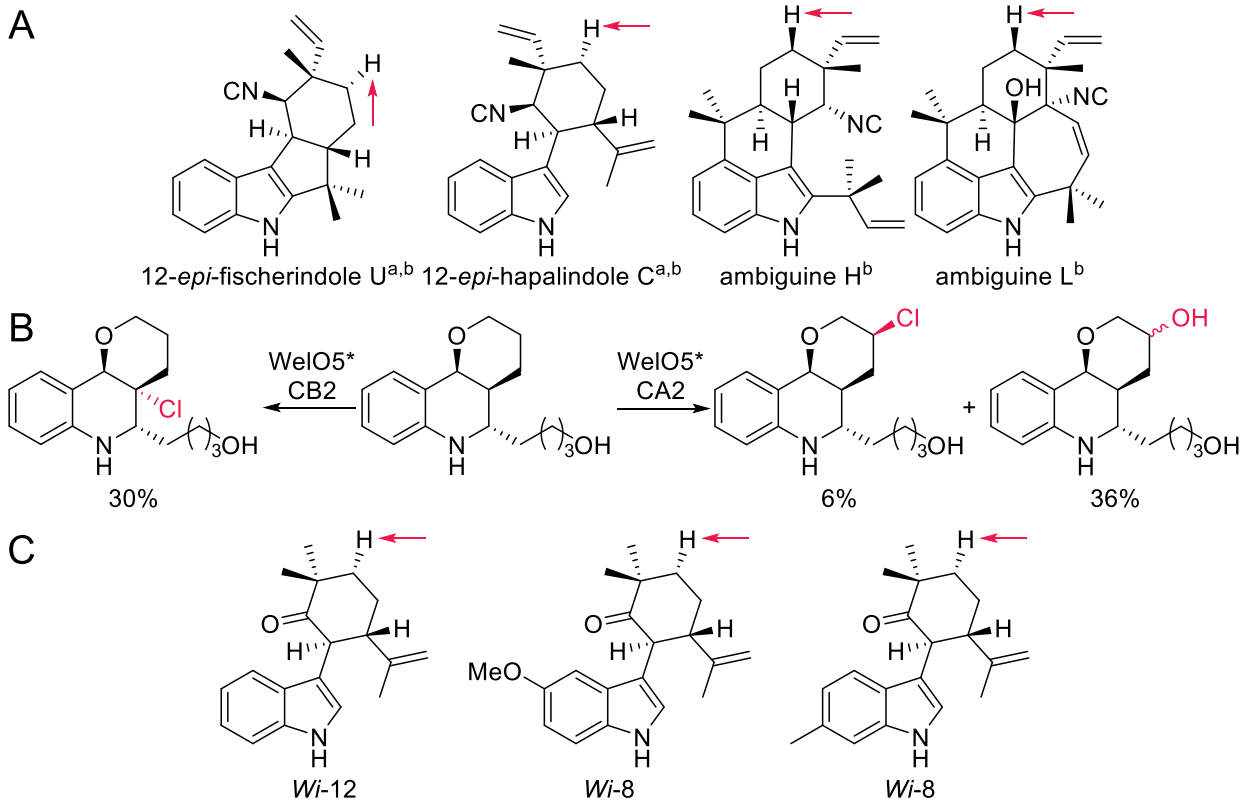
protein partners, including BarB1/BarB2 and BarA, CytC3 and CytC2, and KthP and KtzC, was characterized during this early stage, (Scheme 1.5).[42–44]

Scheme 1.5: Examples of early FeDHs acting on substrates linked to carrier proteins



WelO5 was the first characterized FeDH to act on a free-standing substrate.[45] This enzyme was found as part of the welwitindolinone biosynthesis in *Hapalosiphon welwitschii* UTEX B1830 and was shown to have activity on 12-*epi*-fischerindole U and 12-*epi*-hapalindole C (Scheme 1.6A). A more promiscuous FeDH, AmbO5, was characterized to not only have activity on the native substrates of WelO5, but also on a series of ambiguine compounds (Scheme 1.6A).[46] Efforts to engineer these FeDHs started with expanding the substrate scope of WelO5*, a WelO5 homologue, to act on a martinelline-related substrate. Variants CA2 and CB2 showed improvements in chlorination but with poor chemoselectivity and changes in site selectivity, respectively (Scheme 1.6B).[47] A more recent example is the directed evolution of the related enzyme *Wi*-WelO15 to accept non-native substrates derived from 12-*epi*-hapalindole C, where the isonitrile functionality was replaced by a ketone (Scheme 1.6C).[48] Variants with high site-, chemo-, and enantioselectivity on a subset of compounds were obtained.

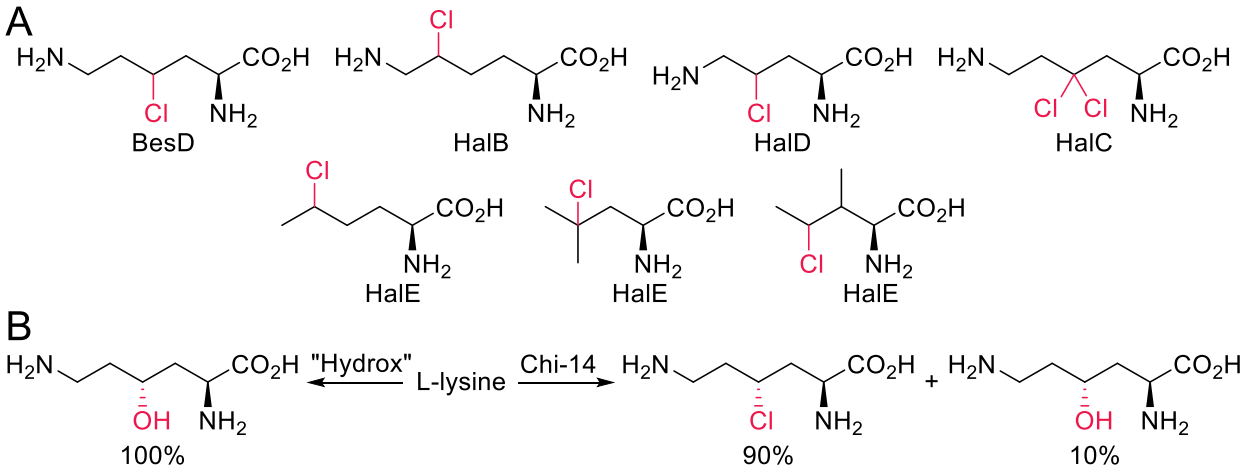
Scheme 1.6: Examples of reactions catalyzed by WelO5, AmbO5, and related FeDHs



A) Substrates accepted by WelO5 (a) and AmbO5 (b). Halogenation sites are indicated with an arrow. B) Halogenation with WelO5* engineered variants. Percentages indicate conversion values. C) Evolved variants of *Wi*-WelO15 and their non-native substrates. Halogenation sites are indicated with an arrow.

A family of FeDHs that acts on free-standing amino acids was more recently discovered. BesD, a lysine chlorinase, was the first enzyme of this family, being found as part of the biosynthesis of β-ethynylserine in *Streptomyces cattleya* (Scheme 1.7A).[49] A sequence-based homology search resulted in four other related halogenases, HalB, HalC, HalD, and HalE, with activity on a set of aliphatic amino acids (Scheme 1.7A).[50] Further exploration of related sequences resulted in a hydroxylase/halogenase pair, "Hydrox" and "Hal", respectively, that had 71% of sequence identity.[51] A DNA shuffling approach combining both sequences allowed for the engineering of a "Hydrox" mutant, Chi-14, that achieved levels of chlorination comparable to "Hal" with even higher chemoselectivity (Scheme 1.7B).

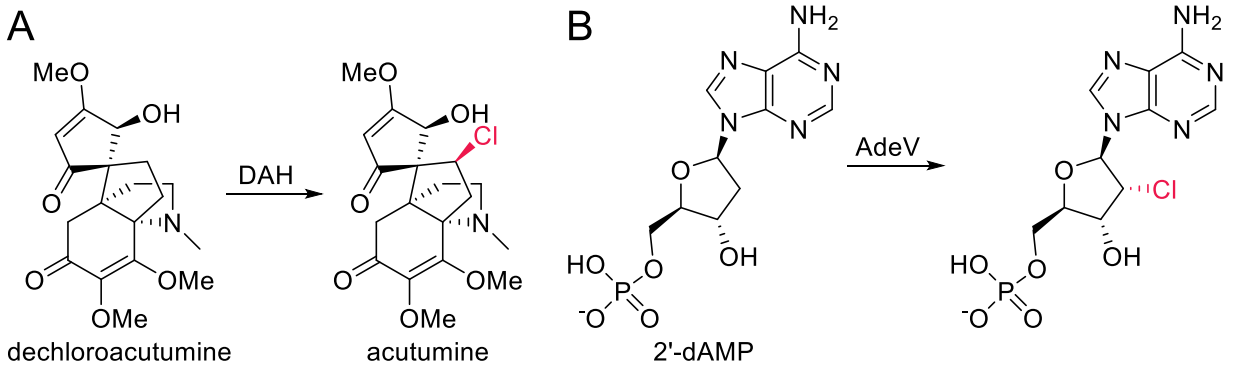
Scheme 1.7: Reactions catalyzed by FeDHs that act on free-standing amino acids



A) Representative products formed by BesD-related FeDHs.[50] B) Engineered "Hydrox" variant Chi-14 achieves non-native chlorination reaction with high levels of chemoselectivity. Percentages indicate product distribution.[51]

Recently, the first FeDH of plant origin, DAH, was discovered and shown to chlorinate dechloroacutumine, while showing no activity on a series of other alkaloids (Scheme 1.8A).[52] Finally, the FeDH, AdeV was recently reported to halogenate nucleotide substrates, such as 2'-deoxyadenosine monophosphate (2'-dAMP, Scheme 1.8B).[53] These examples, along with the other FeDHs described above, highlight the current state in the field and showcase the increasing number of enzymes being discovered and the limitations associated with them. Although this class of enzymes is capable of natively functionalizing a diverse set of substrates (e.g., amino acids, nucleotides, alkaloids), most individual enzymes possess narrow substrate scopes, complicating synthetic applications. The examples of protein engineering and directed evolution represent a promising path to developing more attractive and directly applicable biocatalyst for aliphatic C–H functionalization.

Scheme 1.8: Other FeDHs that act on free-standing non-amino acid substrates



A) Chlorination reaction by dechloroacutumine halogenase DAH.[52] B) AdeV-catalyzed chlorination of nucleotide 2'-deoxyadenosine monophosphate (2'-dAMP).[53]

1.5 Conclusions

Enzymatic C–H functionalization provides an attractive alternative to traditional synthetic routes. Developing biocatalysts with excellent selectivity and the expansion of available reactivities to those that are not native is key to fully exploiting these natural systems. FeDOs represent a platform with the potential for a myriad of non-native activities since they decouple the C–H activation and C–X bond formation involved in C–H functionalization. Better understanding of these biocatalysts, their behavior in the presence of non-native ligands in their active site, as well as the factors that determine the site and chemoselectivity of these transformations is crucial for their development as biocatalysts.

1.6 References

(1) Engle, K. M.; Mei, T. S.; Wasa, M.; Yu, J. Q. Weak Coordination as a Powerful Means for Developing Broadly Useful C-H Functionalization Reactions. *Acc. Chem. Res.* **2012**, *45* (6), 788–802. https://doi.org/10.1021/ar200185g.

(2) Newhouse, T.; Baran, P. S. If C-H Bonds Could Talk: Selective C-H Bond Oxidation. *Angewandte Chemie - International Edition*. John Wiley & Sons, Ltd April 4, 2011, pp 3362–3374. https://doi.org/10.1002/anie.201006368.

(3) Lewis, J. C.; Coelho, P. S.; Arnold, F. H. Enzymatic Functionalization of Carbon-Hydrogen Bonds. *Chem. Soc. Rev.* **2011**, *40* (4), 2003–2021. https://doi.org/10.1039/c0cs00067a.

(4) Kaspera, R.; Croteau, R. Cytochrome P450 Oxygenases of Taxol Biosynthesis. *Phytochem. Rev.* **2006**, *5* (2–3), 433–444. https://doi.org/10.1007/S11101-006-9006-4/METRICS.

(5) Ishihara, Y.; Baran, P. S. Two-Phase Terpene Total Synthesis: Historical Perspective and Application to the Taxol® Problem. *Synlett*. 2010, pp 1733–1745. https://doi.org/10.1055/s-0030-1258123.

(6) Niwa, T.; Murayama, N.; Imagawa, Y.; Yamazaki, H. Regioselective Hydroxylation of Steroid Hormones by Human Cytochromes P450. *Drug Metabolism Reviews*. Informa Healthcare May 1, 2015, pp 89–110. https://doi.org/10.3109/03602532.2015.1011658.

(7) Schwarz, M.; Lund, E. G.; Lathe, R.; Björkhem, I.; Russell, D. W. Identification and Characterization of a Mouse Oxysterol 7α-Hydroxylase CDNA. *J. Biol. Chem.* **1997**, *272* (38), 23995–24001. https://doi.org/10.1074/jbc.272.38.23995.

(8) Wohlgemuth, R. Biocatalysis — Key to Sustainable Industrial Chemistry. *Curr. Opin. Biotechnol.* **2010**, *21* (6), 713–724. https://doi.org/10.1016/J.COPBIO.2010.09.016.

(9) Burk, M. J.; Van Dien, S. Biotechnology for Chemical Production: Challenges and Opportunities. *Trends Biotechnol.* **2016**, *34* (3), 187–190. https://doi.org/10.1016/J.TIBTECH.2015.10.007.

(10) Truppo, M. D. Biocatalysis in the Pharmaceutical Industry: The Need for Speed. *ACS Med. Chem. Lett.* **2017**, *8* (5), 476–480. https://doi.org/10.1021/acsmedchemlett.7b00114.

(11) Abdelraheem, E. M. M.; Busch, H.; Hanefeld, U.; Tonin, F. Biocatalysis Explained: From Pharmaceutical to Bulk Chemical Production. *React. Chem. Eng.* **2019**, *4* (11), 1878–1894. https://doi.org/10.1039/C9RE00301K.

(12) Pines, O.; Inouye, M. Expression and Secretion of Proteins in E. Coli. *Appl. Biochem. Biotechnol. - Part B Mol. Biotechnol.* **1999**, *12* (1), 25–34. https://doi.org/10.1385/MB:12:1:25/METRICS.

(13) Khersonsky, O.; Roodveldt, C.; Tawfik, D. S. Enzyme Promiscuity: Evolutionary and Mechanistic Aspects. *Curr. Opin. Chem. Biol.* **2006**, *10* (5), 498–508. https://doi.org/10.1016/J.CBPA.2006.08.011.

(14) Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K. Engineering the Third Wave of Biocatalysis. *Nat. 2012 4857397* **2012**, *485* (7397), 185–194. https://doi.org/10.1038/nature11117.

(15) Arnold, F. H. Design by Directed Evolution. *Acc. Chem. Res.* **1998**, *31* (3), 125–131. https://doi.org/10.1021/ar960017f.

(16) Shimotohno, A.; Oue, S.; Yano, T.; Kuramitsu, S.; Kagamiyama^, H. Demonstration of the Importance and Usefulness of Manipulating Non-Active-Site Residues in Protein Design 1. *J. Biochem* **2001**, *129*, 943–948.

(17) Romero, P. A.; Arnold, F. H. Exploring Protein Fitness Landscapes by Directed Evolution. *Nat. Rev. Mol. Cell Biol. 2009 1012* **2009**, *10* (12), 866–876. https://doi.org/10.1038/nrm2805.

(18) Moore, J. C.; Arnold, F. H. Directed Evolution of a Para-Nitrobenzyl Esterase for Aqueous-Organic Solvents. *Nat. Biotechnol. 1996 144* **1996**, *14* (4), 458–467. https://doi.org/10.1038/nbt0496-458.

(19) Bell, E. L.; Finnigan, W.; France, S. P.; Green, A. P.; Hayes, M. A.; Hepworth, L. J.; Lovelock, S. L.; Niikura, H.; Osuna, S.; Romero, E.; Ryan, K. S.; Turner, N. J.; Flitsch, S. L. Biocatalysis. *Nat. Rev. Methods Prim. 2021 11* **2021**, *1* (1), 1–21. https://doi.org/10.1038/s43586-021-00044-z.

(20) Hanefeld, U.; Hollmann, F.; Paul, C. E. Biocatalysis Making Waves in Organic Chemistry. *Chem. Soc. Rev.* **2022**, *51* (2), 594–627. https://doi.org/10.1039/D1CS00100K.

(21) Wu, L. F.; Meng, S.; Tang, G. L. Ferrous Iron and α-Ketoglutarate-Dependent Dioxygenases in the Biosynthesis of Microbial Natural Products. *Biochim. Biophys. Acta - Proteins Proteomics* **2016**, *1864* (5), 453–470. https://doi.org/10.1016/J.BBAPAP.2016.01.012.

(22) Hausinger, R. P. Biochemical Diversity of 2-Oxoglutarate-Dependent Oxygenases. *RSC Met.* **2015**, *2015-January* (3), 1–58. https://doi.org/10.1039/9781782621959-00001.

(23) Hutton, J. J.; Tappel, A. L.; Udenfriend, S. Requirements for α-Ketoglutarate, Ferrous Ion and Ascorbate by Collagen Proline Hydroxylase. *Biochem. Biophys. Res. Commun.* **1966**, *24* (2), 179–184. https://doi.org/10.1016/0006-291X(66)90716-9.

(24) Fedeles, B. I.; Singh, V.; Delaney, J. C.; Li, D.; Essigmann, J. M. The AlkB Family of Fe(II)/α-Ketoglutarate-Dependent Dioxygenases: Repairing Nucleic Acid Alkylation Damage and Beyond. *J. Biol. Chem.* **2015**, *290* (34), 20734–20742. https://doi.org/10.1074/JBC.R115.656462.

(25) Kawarada, L.; Suzuki, T.; Ohira, T.; Hirata, S.; Miyauchi, K.; Suzuki, T. ALKBH1 Is an RNA Dioxygenase Responsible for Cytoplasmic and Mitochondrial TRNA Modifications. *Nucleic Acids Res.* **2017**, *45* (12), 7401–7415. https://doi.org/10.1093/NAR/GKX354.

(26) Gao, S. S.; Naowarojna, N.; Cheng, R.; Liu, X.; Liu, P. Recent Examples of α-Ketoglutarate-Dependent Mononuclear Non-Haem Iron Enzymes in Natural Product Biosyntheses. *Nat. Prod. Rep.* **2018**, *35* (8), 792–837. https://doi.org/10.1039/C7NP00067G.

(27) Mitchell, A. J.; Weng, J. K. Unleashing the Synthetic Power of Plant Oxygenases: From Mechanism to Application. *Plant Physiol.* **2019**, *179* (3), 813–829. https://doi.org/10.1104/PP.18.01223.

(28) Hausinger, R. P. Fe(II)/α-Ketoglutarate-Dependent Hydroxylases and Related Enzymes. *Crit. Rev. Biochem. Mol. Biol.* **2004**, *39* (1), 21–68. https://doi.org/10.1080/10409230490440541.

(29) Straganz, G. D.; Nidetzky, B. Variations of the 2-His-1-Carboxylate Theme in Mononuclear Non-Heme Fe II Oxygenases. *ChemBioChem*. 2006, pp 1536–1548. https://doi.org/10.1002/cbic.200600152.

(30) Martinez, S.; Hausinger, R. P. Catalytic Mechanisms of Fe(II)- and 2-Oxoglutarate-Dependent Oxygenases. *J. Biol. Chem.* **2015**, *290* (34), 20702–20711. https://doi.org/10.1074/jbc.R115.648691.

(31) Timmins, A.; De Visser, S. P. A Comparative Review on the Catalytic Mechanism of Nonheme Iron Hydroxylases and Halogenases. *Catal. 2018, Vol. 8, Page 314* **2018**, *8* (8), 314. https://doi.org/10.3390/CATAL8080314.

(32) Voss, M.; Honda Malca, S.; Buller, R. Exploring the Biocatalytic Potential of Fe/α-Ketoglutarate-Dependent Halogenases. *Chem. – A Eur. J.* **2020**, *26* (33), 7336–7345. https://doi.org/10.1002/CHEM.201905752.

(33) Mehmood, R.; Qi, H. W.; Steeves, A. H.; Kulik, H. J. The Protein's Role in Substrate Positioning and Reactivity for Biosynthetic Enzyme Complexes: The Case of SyrB2/SyrB1. *ACS Catal.* **2019**, *9* (6), 4930–4943. https://doi.org/10.1021/acscatal.9b00865.

(34) Martinie, R. J.; Livada, J.; Chang, W. C.; 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.* **2015**, *137* (21), 6912–6919. https://doi.org/10.1021/JACS.5B03370/SUPPL_FILE/JA5B03370_SI_001.PDF.

(35) Klein, C.; Hüttel, W. A Simple Procedure for Selective Hydroxylation of L-Proline and L-Pipecolic Acid with Recombinantly Expressed Proline Hydroxylases. *Adv. Synth. Catal.* **2011**, *353* (8), 1375–1383. https://doi.org/10.1002/ADSC.201000863.

(36) Chen, H.; Bong, Y. K.; Cabirol, F. L.; Prafulchandra, A. G.; Li, T.; Moore, J. C.;

Quintanar-Audelo, M.; Hong, Y.; Collier, S. J.; Smith, D. Biocatalysts and Methods for Hydroxylation of Chemical Compounds, 2015.

(37) Zwick, C. R.; Sosa, M. B.; Renata, H. Characterization of a Citrulline 4-Hydroxylase from Nonribosomal Peptide GE81112 Biosynthesis and Engineering of Its Substrate Specificity for the Chemoenzymatic Synthesis of Enduracididine. *Angew. Chemie Int. Ed.* **2019**, *58* (52), 18854–18858. https://doi.org/10.1002/ANIE.201910659.

(38) Amatuni, A.; Renata, H. Identification of a Lysine 4-Hydroxylase from the Glidobactin Biosynthesis and Evaluation of Its Biocatalytic Potential. *Org. Biomol. Chem.* **2019**, *17* (7), 1736–1739. https://doi.org/10.1039/c8ob02054j.

(39) 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. *Chemical Reviews*. American Chemical Society April 26, 2017, pp 5619–5674. https://doi.org/10.1021/acs.chemrev.6b00571.

(40) Vaillancourt, F. H.; Yin, J.; Walsh, C. T. SyrB2 in Syringomycin E Biosynthesis Is a Nonheme FeII α-Ketoglutarate- and O2-Dependent Halogenase. *Proc. Natl. Acad. Sci.* **2005**, *102* (29), 10111–10116. https://doi.org/10.1073/PNAS.0504412102.

(41) Blasiak, L. C.; Vaillancourt, F. H.; Walsh, C. T.; Drennan, C. L. Crystal Structure of the Non-Haem Iron Halogenase SyrB2 in Syringomycin Biosynthesis. *Nat. 2006 4407082* **2006**, *440* (7082), 368–371. https://doi.org/10.1038/nature04544.

(42) 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.* **2006**, *128* (12), 3900–3901. https://doi.org/10.1021/ja060151n.

(43) 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.* **2006**, *13* (11), 1183–1191. https://doi.org/10.1016/J.CHEMBIOL.2006.09.012.

(44) 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* **2011**, *50* (27), 6063–6072. https://doi.org/10.1021/bi200656k.

(45) Hillwig, M. L.; Liu, X. A New Family of Iron-Dependent Halogenases Acts on Freestanding Substrates. *Nat. Chem. Biol.* **2014**, *10* (11), 921–923. https://doi.org/10.1038/nchembio.1625.

(46) 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. **2016**, *55* (19), 5780–5784. https://doi.org/10.1002/anie.201601447.

(47) 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.* **2019**, *58* (51), 18535–18539. https://doi.org/10.1002/ANIE.201907245.

(48) Duewel, S.; Schmermund, L.; Faber, T.; Harms, K.; Srinivasan, V.; Meggers, E.; Hoebenreich, S. Directed Evolution of an FeII-Dependent Halogenase for Asymmetric C(Sp3)–H Chlorination. *ACS Catal.* **2019**, *10* (2), 1272–1277. https://doi.org/10.1021/ACSCATAL.9B04691.

(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* **2019**. https://doi.org/10.1038/s41586-019-1020-y.

(50) 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* (10), 1009–1016. https://doi.org/10.1038/s41589-019-0355-x.

(51) Neugebauer, M. E.; Kissman, E. N.; Marchand, J. A.; Pelton, J. G.; Sambold, N. A.; Millar, D. C.; Chang, M. C. Y. Reaction Pathway Engineering Converts a Radical Hydroxylase into a Halogenase. https://doi.org/10.1038/s41589-021-00944-x.

(52) Kim, C. Y.; Mitchell, A. J.; Glinkerman, C. M.; Li, F.-S.; Pluskal, T.; Weng, J.-K. The Chloroalkaloid (−)-Acutumine Is Biosynthesized via a Fe(II)-and 2-Oxoglutarate-Dependent Halogenase in Menispermaceae Plants. https://doi.org/10.1038/s41467-020-15777-w.

(53) Zhao, C.; Yan, S.; Li, Q.; Zhu, H.; Zhong, Z.; Ye, Y.; Deng, Z.; Zhang, Y. An Fe2+- and α-Ketoglutarate-Dependent Halogenase Acts on Nucleotide Substrates. *Angew. Chemie* **2020**, *132* (24), 9565–9571. https://doi.org/10.1002/ANGE.201914994.

Chapter 2: Optimization and characterization of Fe(II)- and α-ketoglutarate-dependent oxygenase catalysis with exogenous anions

2.1 Introduction

2.1.1 Non-native activity in FeDHs

Regardless of the high levels of activity and selectivity an enzyme may have, the synthetic applicability of these biocatalysts will depend on whether they are able to catalyze a given transformation of interest. Synthetic challenges may lie beyond what enzymes evolved to do, creating the need for finding or developing new biocatalysts with novel activities, selectivities, and substrate scopes. Protein engineering and directed evolution have been used to alter or develop new enzymatic features, and examples where the substrate scope and site selectivity of FeDOs were targeted were highlighted in Chapter 1.

Additionally, non-native activities have been explored in FeDHs. The halogenase SyrB2 was shown to catalyze non-native azidation and nitration reactions in the presence of azide (N_3^-) and nitrite (NO_2^-) ions, respectively, albeit at low levels and with low levels of contaminant chloride ions (Cl^-) being able to produce significant amounts of chlorinated product.[1] A single mutation improved these non-native activities and selectivity relative to chlorination, but the yields remained low. Although no additional functionalized products were detected, this study also showed evidence of exogenous bromide (Br^-), cyanide (CN^-), cyanate (OCN^-), bisulfide (HS^-), and formate (HCO_2^-) anions binding to the enzyme-Fe(II)-α-ketoglutarate complex.

More recently, FeDHs acting on free-standing substrates have been reported to catalyze non-native reactions using exogenous anions. The BesD-related halogenase SwHalB from *Streptomyces wuyuanensis* catalyzes the bromination and azidation of L-lysine in the presence of Br^- and N_3^-, respectively, although no conversion or yield values were reported.[2] Plant FeDH SaDAH is capable of azidation of its native substrate dechloroacutumine in the presence of N_3^-.[3] A small fraction of a hydroxylated side product was detected under these conditions, as well as the chlorinated acutumine product in equivalent levels to the azidated product. This chlorinated compound was the product of trace levels of Cl^- in the protein storage buffer. Finally, WelO5* variants engineered to chlorinate non-native polyketide substrate soraphen A were also capable of bromination, azidation, and nitration in the presence of Br^-, N_3^-, and NO_2^-, respectively, although mixtures of functionalization at two different sites were observed, and no data on yields or chemoselectivity was provided.[4]

2.1.2 Expanding the collection FeDH with modified hydroxylases

The use of FeDHs as biocatalysts to perform non-native reactions suffers from several limitations, such as reduced activity, poor chemoselectivity in the presence of trace amounts of Cl^- anion, emergence of hydroxylation as a side reaction, and the relatively low number of FeDHs characterized to date. For this reason, expanding the scope of enzymes to carry out these transformations is of great importance. The mechanistic similarities between hydroxylases and halogenases of the FeDO superfamily suggest that one could look at the latter group to access more biocatalysts capable of C–H functionalization with non-native anions.

The fact that the facial triads of hydroxylases and halogenases differ only in one amino acid (Figure 1.2) prompts the question of whether hydroxylases could be rationally converted to halogenases by introducing the corresponding Asp/Glu→Gly/Ala mutations. This strategy has

the potential to open the vast library of hydroxylating FeDOs to halogenation and other non-native reactions. However, applying this approach successfully has proved to be challenging. An early study attempted to explore the halogenation potential of facial triad mutants of prolyl 4-hydroxylase, but both Asp→Gly and Asp→Ala mutants showed loss of hydroxylase activity and no halogenase activity.[5] Notably, the addition of other anions such as N_3^-, HCO_2^-, and $CH_3CO_2^-$ was not sufficient to rescue the native hydroxylase activity either.

The facial triad of taurine hydroxylase TauD has been thoroughly investigated.[6] The halogenating ability of facial triad mutant D101A was tested with different concentrations of Cl^- ions in solution. Although no chlorination was detected, and hydroxylase activity was lost under these conditions, the use of HCO_2^- ions in solution resulted in partial rescue of the native hydroxylation in mutant D101A. Finally, the factor-inhibiting hypoxia inducible factor (FIH), a hydroxylating-FeDO, was showed loss of activity when introducing Asp→Gly/Ala mutations in the facial triad D201G and D201A.[7] When tested against halide ions (i.e. F^-, Cl^-, Br^-, and I^-), only the presence of Cl^- was able to partially rescue the native hydroxylase activity in FIH D201G, and no halogenated products were detected for any of these mutant/halide combinations.

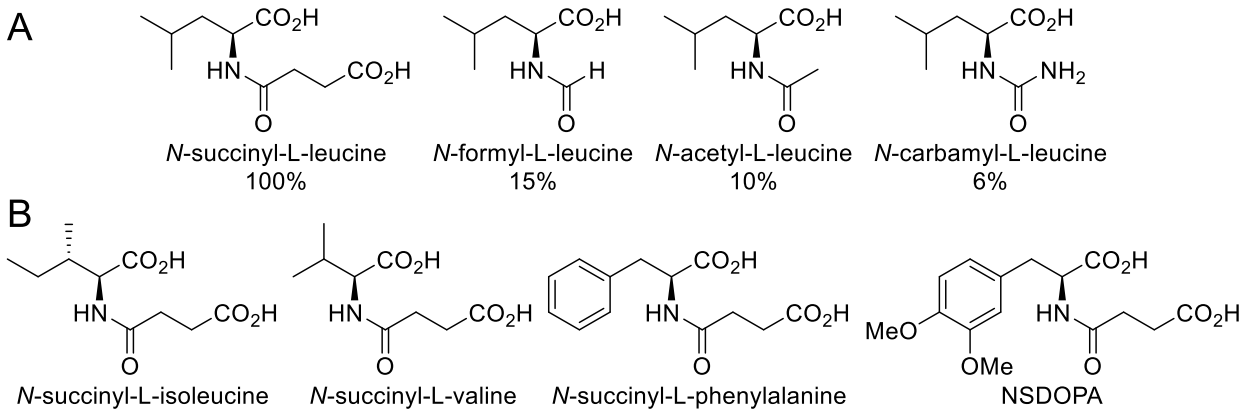
Despite the above-mentioned cases, where non-native halogenation was not enabled by facial triad Asp/Glu→Gly/Ala mutations, two enzymes were able to achieve this goal. L-proline *cis*-4-hydroxylase from *Sinorhizobium meliloti* (SmP4H) was engineered to introduce the Asp→Gly facial triad mutation D108G.[8] This new variant, SmP4H-0, was able to produce a mixture of *cis*-4-hydroxy-L-proline and *cis*-3-chloro-L-proline in the presence of L-proline and Cl^-. The chemoselectivity of this reaction, however, favored the native hydroxylation in a 24:1 hydroxylation to chlorination ratio. A directed evolution campaign produced variant SmP4H-7, with a 18.7-fold increase in chlorination compared to SmP4H-0, but with a chemoselectivity that

still favored the native hydroxylation reaction, with a 12:1 hydroxylation to chlorination ratio. The other successful example of engineering halogenase activity from a hydroxylase is the FeDO SadA, and it will be described in more detail in the following section.[9]

2.1.3 FeDO SadA and engineered FeDH SadA D157G

SadA was first characterized as part of a search for novel microbial biocatalysts able to produce L-*threo*-β-hydroxyleucine.[10] Although no enzymes in this study showed activity on L-leucine, SadA from *Burkholderia ambifaria* AMMD hydroxylated several *N*-substituted L-leucine derivatives, with *N*-succinyl-L-leucine being the substrate that showed the highest activity (Figure 2.1A). Furthermore, SadA was shown to accept other *N*-succinylated amino acids, including *N*-succinyl-L-isoleucine, *N*-succinyl-L-valine, *N*-succinyl-L-phenylalanine, and *N*-succinyl-L-3,4-dimethoxyphenylalanine (NSDOPA, Figure 2.1B).[11,12] Engineering of site G79 and F261 led to improved activity on NSDOPA.

Figure 2.1: Substrate scope of SadA

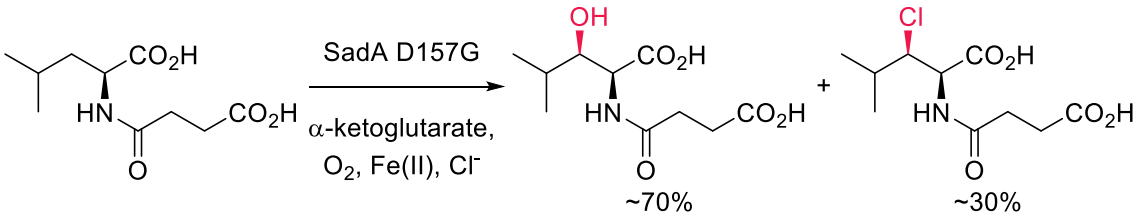


A) Relative hydroxylase activity of SadA on different *N*-substituted L-leucine substrates.[10] B) Other *N*-succinylated amino acids accepted by SadA.[11,12]

A structure similarity search based on native FeDH WelO5 conducted by the Boal and Liu groups found that SadA had the highest similarity score.[9] This led them to investigate the Asp→Gly facial triad mutant of SadA, SadA D157G, and its potential to act as a FeDH. This new mutant not only was able to chlorinate *N*-succinyl-L-leucine in the presence of Cl^-, but also

brominate it in the presence of Br^-, albeit with lower conversion (Scheme 2.1). These non-native activities had poor chemoselectivity, however, with the chlorinated product being ~30% of the total product pool and the remaining ~70% being the native hydroxylated product. Despite these drawbacks, the fact that SadA D157G was the first example of a facial triad mutant to be able to accept multiple exogenous anions (i.e., Cl^- and Br^-) and perform non-native halogenation reactions suggested distinct and flexible scope of anionic ligands. These features, combined with the scope of substrates accepted by wild-type SadA, suggested that these enzymes were a promising platform for developing biocatalysts for a range of non-native activities. Efforts to explore the potential of said platform have been summarized and published.[13]

Scheme 2.1: Initial halogenase activity of SadA D157G[a]



[a] Percentages represent fractions of the total product pool.

Authorship

Reaction conditions for the anion screen experiment described in section 2.2.2 were performed by Dr. Natalie H. Chan. Dr. Chan also performed an earlier version of the anion screen depicted in Figure 2.8, but the experiment was repeated to include an additional anion and enzyme (i.e., NCO^- and SadXL, respectively). Additionally, Dr. Chan contributed with the isolation and structural characterization of products **1c** and **1e**$_{N3}$.

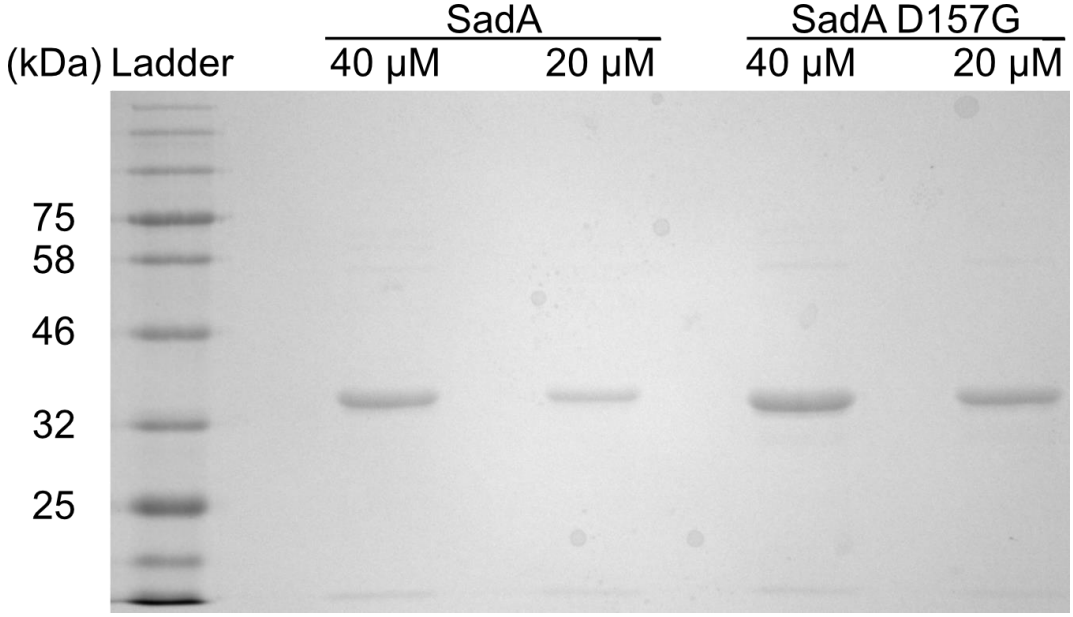
2.2 Results and discussion

2.2.1 Optimization of SadA D157G soluble expression

We started by heterologously expressing both wild-type SadA and the mutant SadA D157G in *E. coli* and purifying these enzymes using Ni-NTA affinity chromatography.

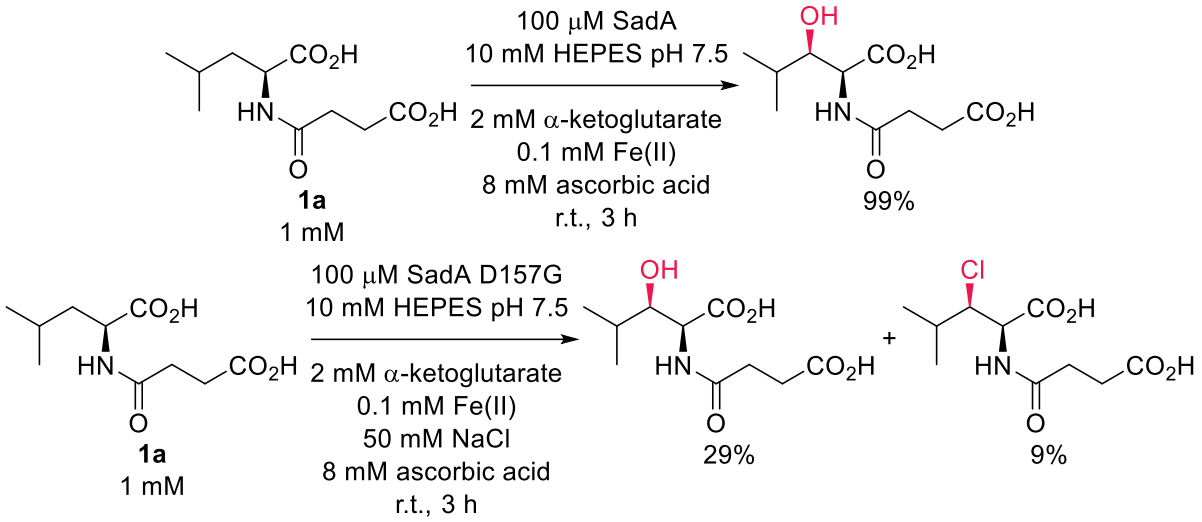
Acceptable yield levels (30 – 60 mg/L) and good purity of protein samples were obtained (Figure 2.2). The activity of these enzymes was tested using *N*-succinyl-L-leucine (**1a**) as substrate, and levels of chlorination and hydroxylation comparable to the previously reported values were obtained from *in vitro* bioconversions as determined by LC-MS (Scheme 2.2).[9] However, envisioning future directed evolution efforts in which variants would be screened in 96-well plates using cell lysate instead of purified enzymes, we attempted chlorination reactions with SadA D157G lysate in 96-well plates. Even though optimization of reaction conditions was attempted, very low levels of chlorinated product were observed under these conditions. These results also suffered from poor reproducibility due to the low signal-to-noise ratio.

Figure 2.2: SDS-PAGE of purified SadA and SadA D157G[a]



[a] Lanes correspond to SadA with a N-terminal His_6-tag ("SadA", MW = 34208.22 Da) and SadA D157G with a N-terminal His_6-tag ("SadA D157G", MW = 34150.19 Da). Two different final concentrations (40 µM and 20 µM) were used for each sample.

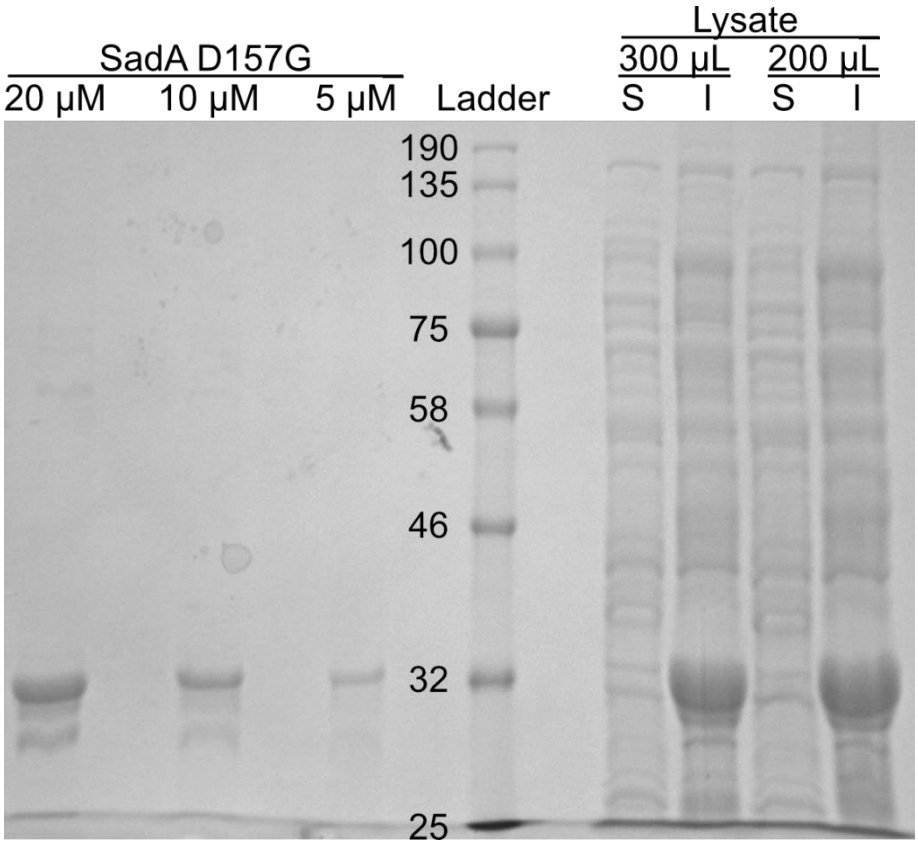
Scheme 2.2: Initial activity assays with purified SadA and SadA D157G[a]



[a] Percentages represent conversion values determined by LC-MS peak area comparison.

When the relative concentrations of SadA D157G in the soluble and insoluble fractions were analyzed in lysates of cultures grown in 24-well plates, it was clear that most of the overexpressed protein was present in the insoluble fraction (Figure 2.3). This result suggested that the soluble expression of SadA D157G would need to be improved to allow for future engineering efforts. Two well-established strategies to improve soluble expression of proteins in *E. coli* were attempted: co-expression with molecular chaperones and fusion to a solubility-enhancing tag.[14,15]

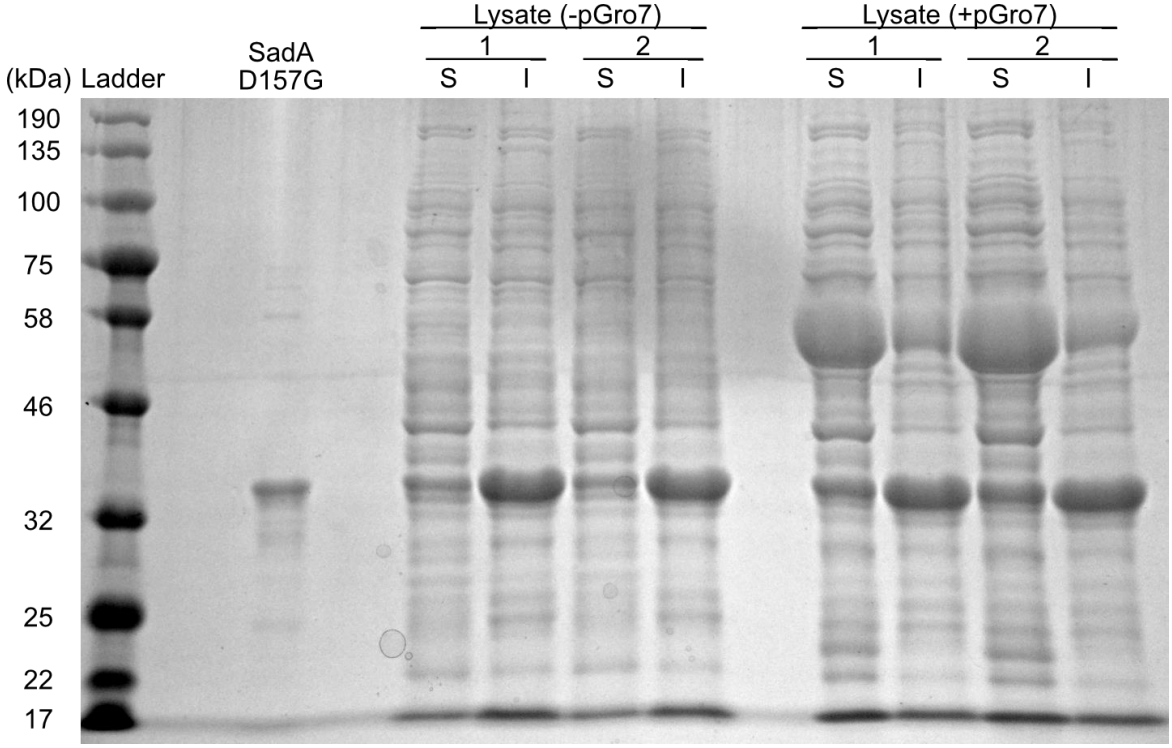
Figure 2.3: SDS-PAGE analysis of SadA D157G soluble expression[a]



[a] Lanes correspond to positive controls with purified SadA D157G ("SadA D157G") at different final concentrations (20 μM, 10 μM, and 5 μM), the soluble ("S") and insoluble ("I") fractions after lysis of a cell resuspension in different volumes (300 μL and 200 μL).

The commercially available plasmid pGro7, which encodes the chaperone protein pair GroEL-GroES, was transformed into *E. coli* cells containing the SadA D157G-encoding plasmid.[16] Co-expression of this protein system in 24-well plates resulted in a ~2.5-fold increase in the relative expression of soluble SadA D157G when analyzed by SDS-PAGE (Figure 2.4). However, the insoluble fraction still contained most of the overexpressed enzyme.

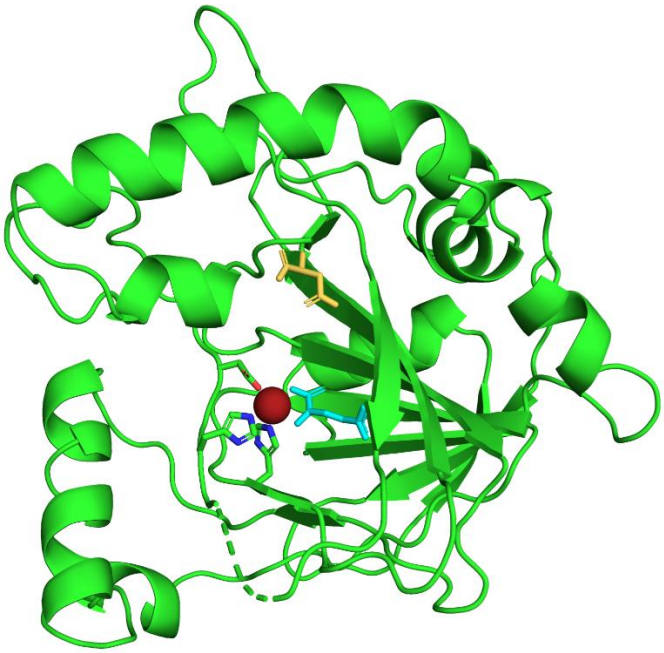
Figure 2.4: SDS-PAGE expression analysis using molecular chaperones[a]



[a] Lanes correspond to a positive control of purified SadA D157G ("SadA D157G"), the soluble ("S") and insoluble ("I") fractions after lysis of duplicate cultures expressing SadA D157G ("Lysate (-pGro7)"), and the soluble and insoluble fractions after lysis of a culture co-expressing SadA D157G, GroEL, and GroES ("Lysate (+pGro7)").

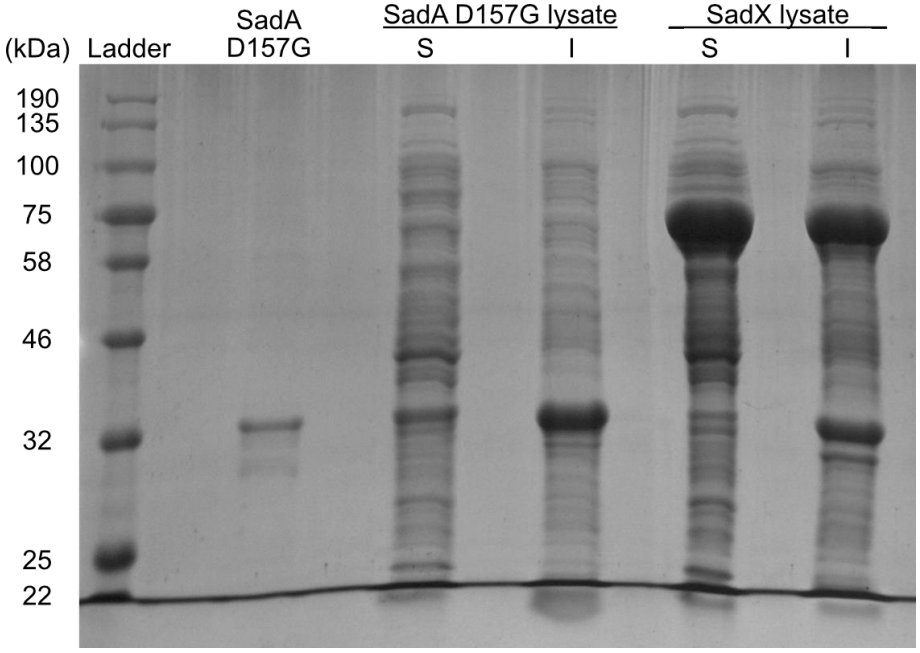
The fusion tag chosen to improve soluble expression of SadA D157G was the maltose-binding protein (MBP) from *E. coli*.[17] According to the reported crystal structure of SadA (PDB ID: 3W21),[18] the C-terminus of the protein is located proximal to the active site (Figure 2.5, shown in yellow), which suggested that a N-terminal fusion would be more likely to succeed. With this in mind, the MBP tag was added to the N-terminus of SadA D157G, making the fusion protein MBP-SadA D157G, which will be referred to as SadX from now on. Expression of this fusion in 24-well plates resulted in a significant improvement in soluble expression when analyzed by SDS-PAGE (Figure 2.6), with most of the overexpressed SadX being in the soluble fraction of the resulting lysate. When SadX was overexpressed in 1-liter cultures, yields of 200 – 300 mg/L protein were obtained.

Figure 2.5: Crystal structure of SadA (PDB ID: 3W21)[a]



[a] Facial triad residues are shown as green sticks, Zn(II) is shown as a red sphere in place of Fe(II), α-ketoglutarate co-substrate is shown as cyan sticks, and the C-terminal residue Asp_{273} is shown as yellow sticks.

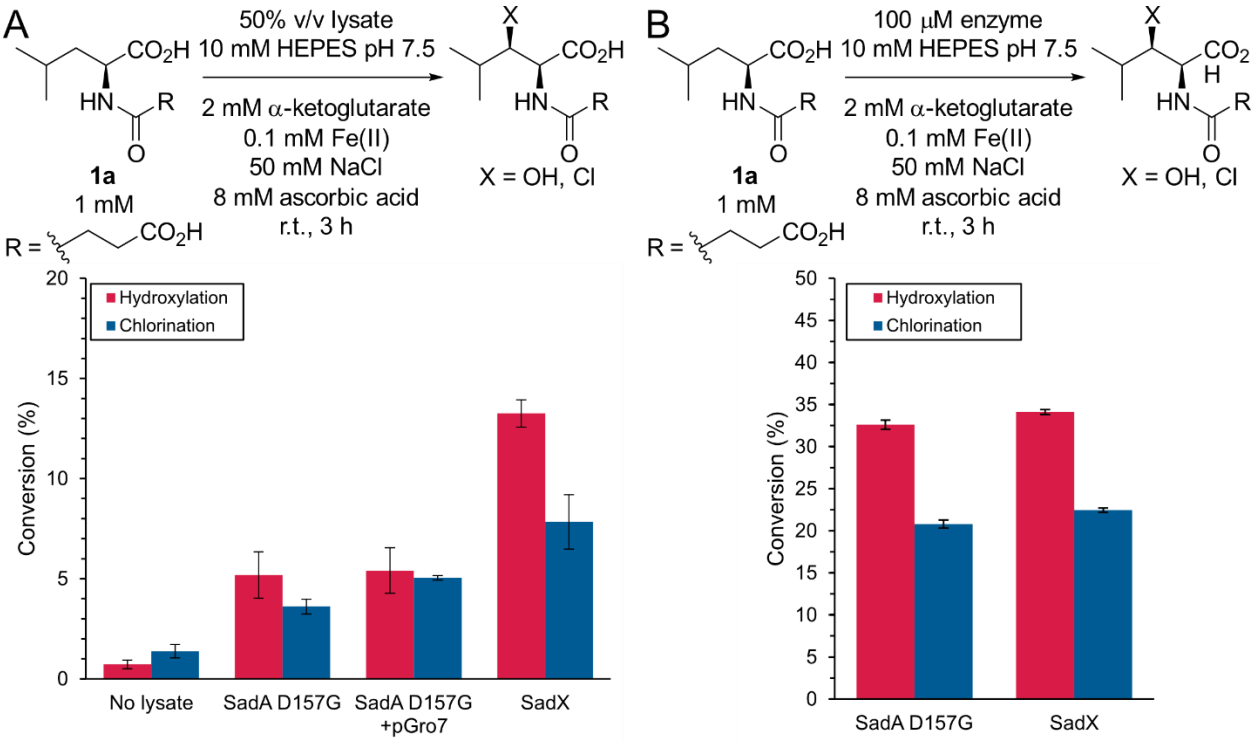
Figure 2.6: SDS-PAGE expression analysis of MBP-fused SadA D157G (SadX)[a]



[a] Lanes correspond to a positive control of purified SadA D157G ("SadA D157G"), the soluble ("S") and insoluble ("I") fractions after lysis of a culture of expressing SadA D157G ("SadA D157G lysate"), and the soluble and insoluble fractions after lysis of a culture expressing the MBP-fusion SadX ("SadX lysate", MW = 75407.40 Da).

The newly developed expression systems were compared by carrying out bioconversions using the corresponding soluble fractions of lysate (Figure 2.7A). These results showed a significant increase in the conversion to chlorinated product, to a level that was reliably reproduced through replicate experiments. Additionally, when bioconversions using purified SadA D157G and SadX were compared, no significant differences were observed, suggesting that the MBP-tag had negligible effects on the activity of these enzymes (Figure 2.7B). Given these results, SadX was selected as the starting point for exploring non-native selective C–H functionalization reactions.

Figure 2.7: Activity of SadA D157G expression systems[a]

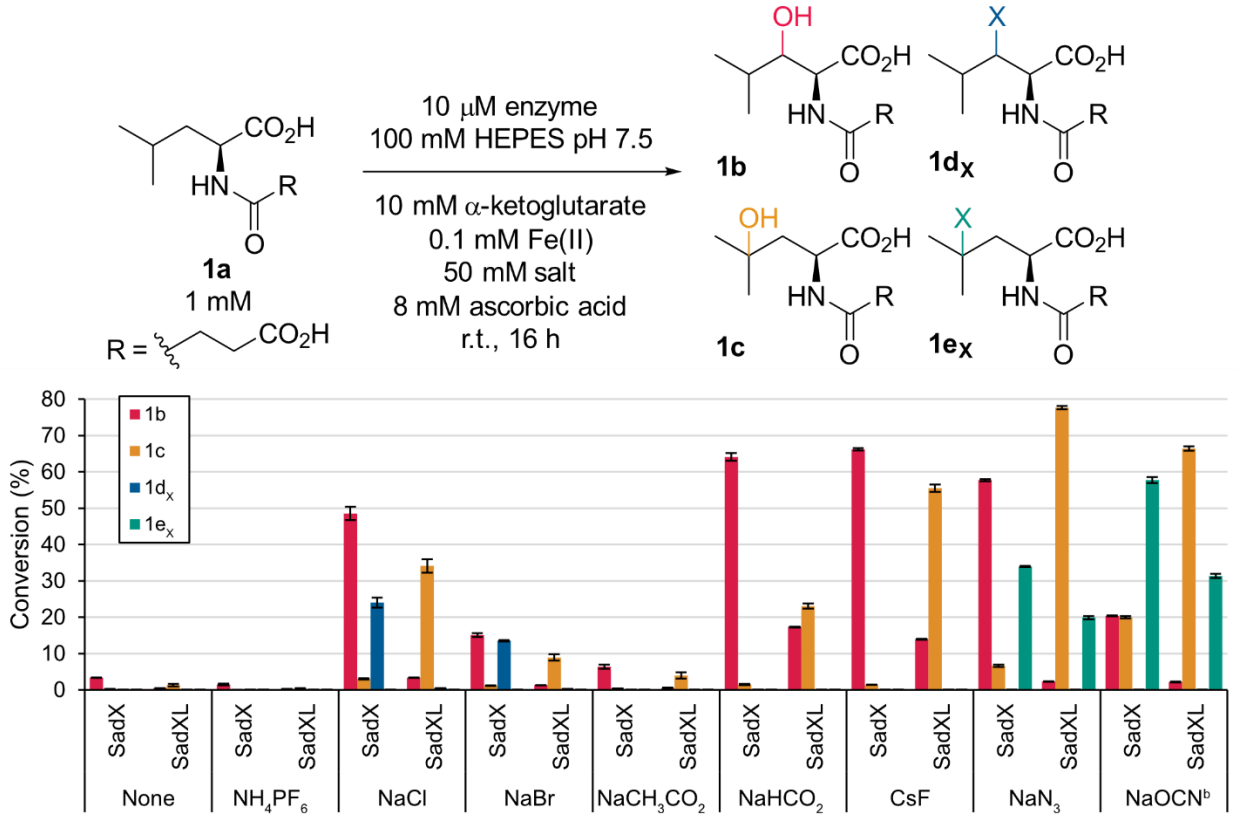


A) Conversion of **1a** using cell lysate of the SadA D157G expression systems. B) Conversion of **1a** using purified SadA D157G and SadX. [a] Conversion was calculated with LC-MS chromatogram peak areas. Each data point is the average of three replicates and error bars are standard deviations.

2.2.2 SadX activity in the presence of exogenous anions

The activity of SadX on **1a** was evaluated in the presence of different anions with the aim of studying the effect of these anionic ligands on the native hydroxylation reaction and exploring potential non-native reactivities (Figure 2.8). Based on previous screens of FeDHs with exogenous anions,[1–4] the following ionic compounds were selected: NaCl, NaBr, CsF, $NaCH_3CO_2$, $NaHCO_2$, NaN_3, NaOCN, and NH_4PF_6. Additionally, a mutant of SadX, SadXL (SadX F152L), which derived from evolution efforts detailed in Chapter 3, had been found to produce the γ-hydroxylated product **1c** instead of the native β-hydroxylated product **1b** in the presence of Cl^-. Given this interesting change in site-selectivity, this variant enzyme was also added to the anion screen experiments.

Figure 2.8: Reaction of 1a with SadX and SadXL in the presence of exogenous anions[a]



[a] Reactions in the absence ("None") and presence of anions (X). Conversion was calculated with chromatogram peak areas from LC-MS. Each data point is the average of three replicates and error bars are standard deviations. [b] Reaction with SadX results in a mixture of products.

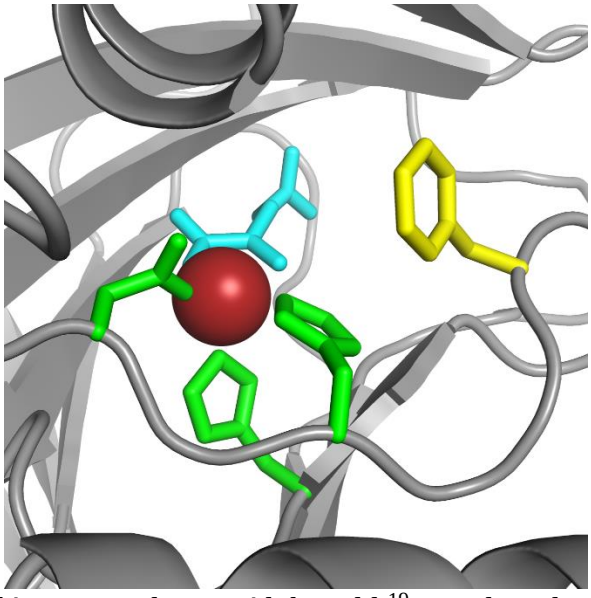
Reactions carried out in the absence of anions showed only trace conversion to products, suggesting the need for an anionic ligand to complete the facial triad of these enzymes. A similar result can be seen in the presence of the non-coordinating anion PF_6^-, implying that the ionic strength of the media had no effect on activity either. Reactions with NaCl and NaBr reproduced the original reports of SadA D157G,[9] not only showing recovery of the native hydroxylation activity, but also halogenation at the β-position of the amino acid fragment.

All other anions in this screen had not been explored with SadA D157G before this study. These anionic ligands showed different degrees of rescue of the native hydroxylase activity of SadA. Both formate and acetate, ligands that mimic the native aspartate residue, produced **1b**, although only the former did it with a high conversion value, suggesting better fit in the active site due to its smaller size. The use of CsF rescued hydroxylation with excellent conversion values, although no rebound of fluoride was observed, as opposed to the other halide ligands Cl^- and Br^-. Further characterization performed by Dr. Chan showed evidence of F^- binding to the active site both to an Fe(II) state via the analysis of the Fe(II)→α-ketoglutarate metal-to-ligand charge transfer band using UV-Vis spectroscopy, as well as to the catalytically-relevant Fe(III)-hydroxo state via studying the kinetic behavior of SadX in the presence of CsF. The lack of fluorination was studied in collaboration with Dr. Vyshnavi Vennelakanti and Prof. Heather J. Kulik, who performed DFT calculations to establish the energies associated with the rebound step with different ligands. These calculations suggest that the lack of fluorination is most likely due to the 3.5 kcal/mol $\Delta\Delta G^{\ddagger}$ between the transition states linked to rebound with hydroxyl and fluoride rebound. These results, as well as the characterization of anionic ligand binding to SadX, were summarized in a published article.[13]

Reactions with SadX in the presence of NaN_3 and NaOCN resulted in hydroxylated products **1b** and **1c**, with similar conversion values in the latter case. This change in the site selectivity of SadX is anion-dependent, with all other cases favoring hydroxylation at the β-position exclusively. Additionally, non-native functionalization of **1a** was observed; γ-azidated product $\mathbf{1e_{N3}}$ was obtained from reactions with N_3^-, and a mixture of products was obtained in the presence of NCO^-. Even though azidation had been reported on other FeDHs,[1–4] these results are the first example of cyanate incorporation by a FeDH, as well as the first FeDO facial triad mutant to carry out azidation reactions. The site selectivity of these transformations differed from the native β-hydroxylated **1b**, as well as halogenated products $\mathbf{1d_{Cl}}$ and $\mathbf{1d_{Br}}$. Further discussion of products from reactions in the presence of azide and cyanate can be found in chapter 3 and section 2.2.3, respectively.

Introducing mutation F152L in variant SadXL had a significant effect on product distribution. The absence of this residue in SadA's reported crystal structure limits the structural analysis of this mutation. However, a model of SadA obtained using AlphaFold locates residue Phe_{152} in close proximity to the active site and the substrate binding pocket (Figure 2.9).[19] This mutation showed a high selectivity for γ-hydroxylated product **1c** over the native β-hydroxylated **1b** in all cases, except in reactions with $NaHCO_2$. Additionally, this variant was incapable of halogenating to form β-functionalized products $\mathbf{1d_{Cl}}$ and $\mathbf{1d_{Br}}$. These two results suggest an enzyme that is worse at accessing the β-position in substrate **1a**. Finally, in cases where non-native functionalization was observed, the chemoselectivity of SadXL towards hydroxylation was higher than SadX's, as it can be seen by the relative production of **1c** over $\mathbf{1e_X}$ in SadXL reactions compared to **1b** over $\mathbf{1e_X}$ in SadX reactions.

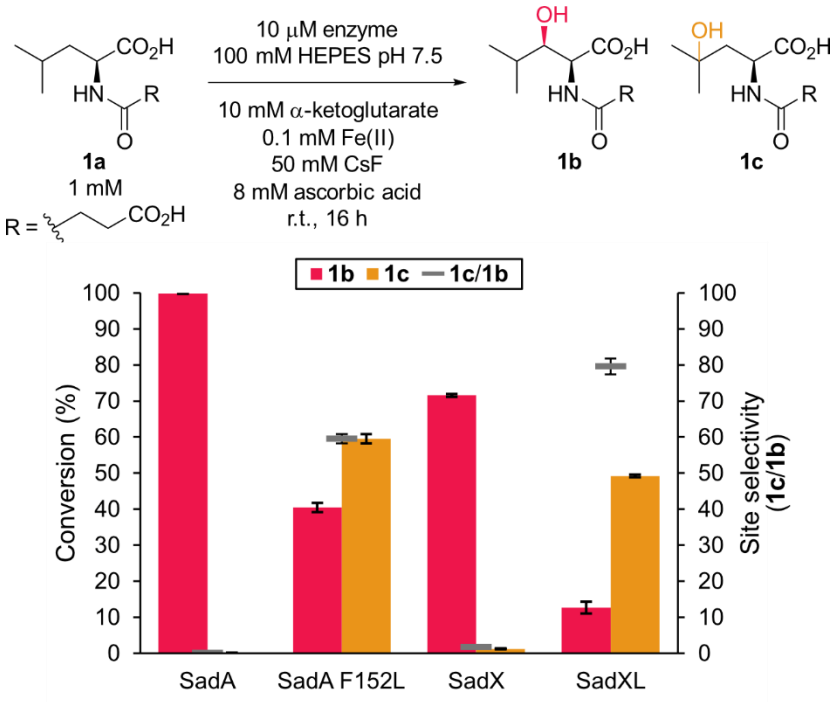
Figure 2.9: Location of Phe$_{152}$ in SadA's active site[a]



[a] Structural model of SadA generated using AlphaFold.[19] Facial triad residues are shown as green sticks, Zn(II) is shown as a red sphere in place of Fe(II), α-ketoglutarate co-substrate is shown as cyan sticks, and phenylalanine residue 152 is shown as yellow sticks.

The effects of mutation F152L on site selectivity were further investigated. A series of reactions were carried out in the presence of CsF using SadX, SadXL, SadA, and SadA F152L, with the latter being the facial triad reversion of SadXL (Figure 2.10). The ratio of conversion to products **1c**/**1b** was used to analyze the site selectivity of these bioconversions. While the **1c**/**1b** ratio was ~4 in reactions with SadXL, this value was only ~1.5 in the case of SadA F152L. These results suggest that the high selectivity of SadXL for γ-hydroxylation is not caused by the enzyme scaffold on its own, and that the presence of the exogenous F^- anion is responsible as well.

Figure 2.10: Site selectivity in SadA F152L variants[a]



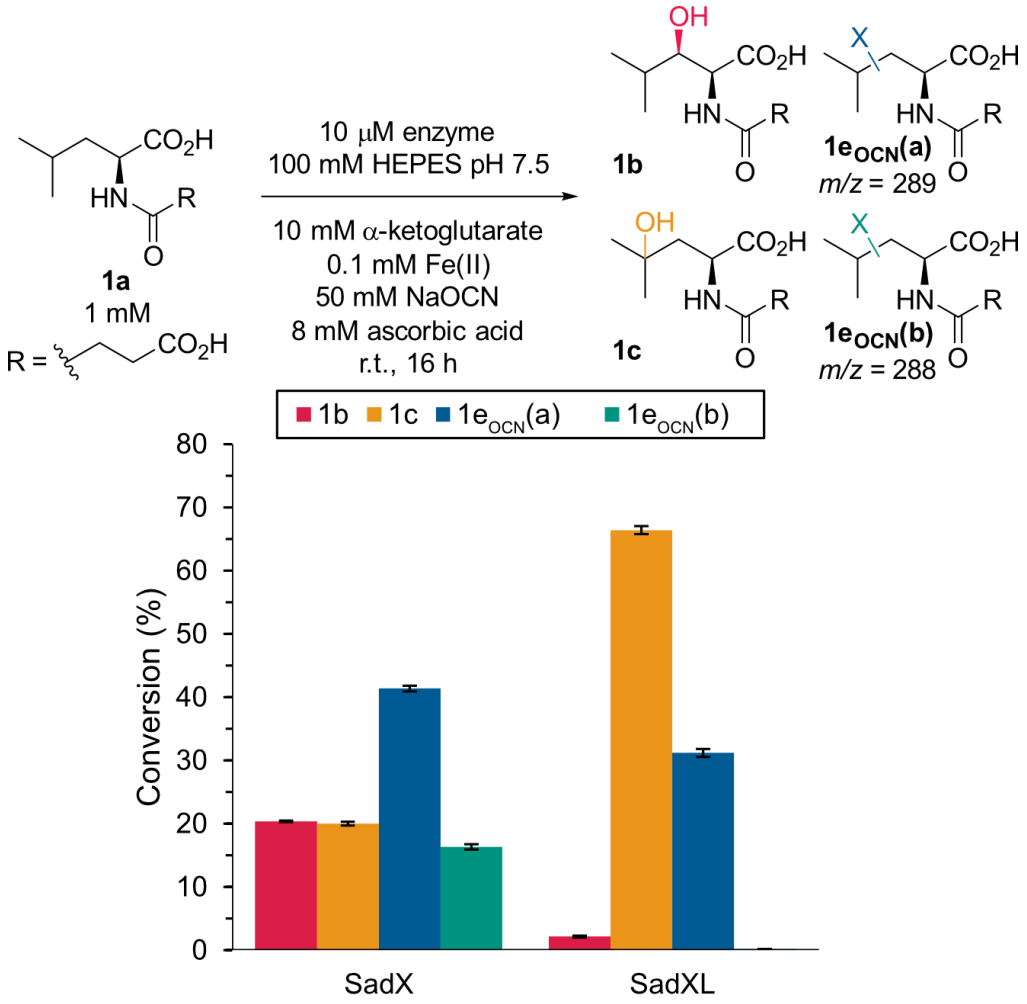
[a] Conversion of **1a** to products by SadA variants in the presence of CsF. Conversion values were calculated with extracted ion chromatogram peak areas of products and the substrate from LC-MS experiments. **1c/1b** is the ratio of conversion to product **1c** over **1b**. Each data point corresponds to the average of three replicates and error bars represent standard deviations.

2.2.3 Non-native cyanate rebound

As discussed in section 2.2.2, the incorporation of NCO^- via non-native rebound with a FeDO detected in this study has not been reported before, so further characterization of this transformation was carried out. Reactions of **1a** with SadA in the presence of NaOCN were analyzed via LC-MS (Figure 2.11). Besides hydroxylated products **1b** and **1c**, two new products, **1e$_{OCN}$(a)** and **1e$_{OCN}$(b),** were detected. The base peaks in the mass spectra of these compounds did not have *m/z* values corresponding to cyanate or isocyanate derivatives of **1a**, but instead they were ~289 and ~288 for **1e$_{OCN}$(a)** and **1e$_{OCN}$(b)**, respectively. Additionally, the product distribution favored **1e$_{OCN}$(a)** with approximately twice as much conversion to that product when compared to **1e$_{OCN}$(b)**. The product distribution of this transformation using SadXL showed no

conversion to product **1e$_{OCN}$(b)** and an overall lower chemoselectivity towards NCO^- rebound when compared to conversion to hydroxylated product **1c**. These results highlighted once again the effect of mutation F152L on site and chemoselectivity.

Figure 2.11: Bioconversions with SadX variants in the presence of NaOCN[a]

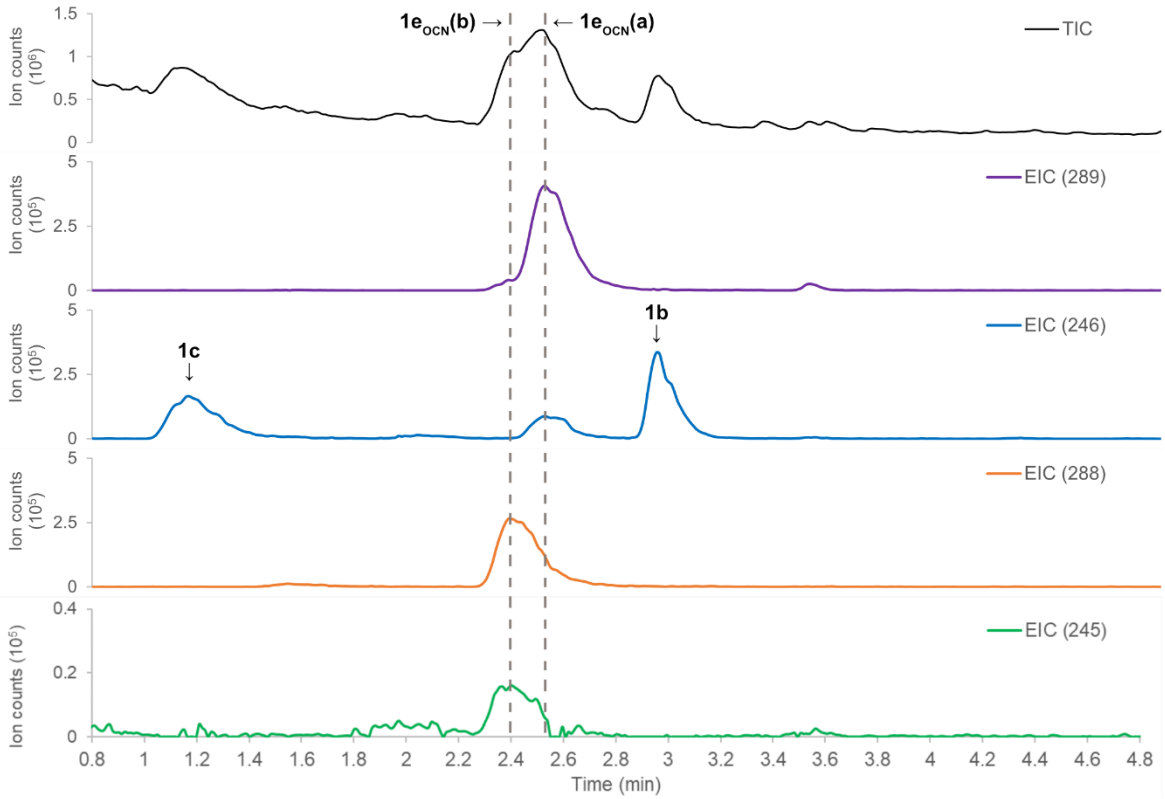


[a] Conversion of **1a** to products by SadX and SadXL in the presence of NaOCN. Conversion values were calculated with extracted ion chromatogram peak areas of products and the substrate from LC-MS experiments. The *m/z* values shown correspond to the base peaks in negative ionization mass spectra of the novel products **1e$_{OCN}$(a)** and **1e$_{NCO}$(b)**. Each data point corresponds to the average of three replicates and error bars represent standard deviations.

Analysis of LC-MS chromatograms and corresponding mass spectra was carried out to begin to characterize these two novel products. The mass spectrum of product **1e$_{OCN}$(a)** contained a fragment ion with *m/z* of 246, matching the $[M - H]^-$ ion for hydroxylated products

1b/**1c**, and this fragment co-eluted with the peak associated with **1e$_{OCN}$(a)** (Figure 2.12). These observations suggested that **1e$_{OCN}$(a)** contained a new C–O bond that persisted after fragmentation in the mass spectrometer. Analogously, the mass spectrum of **1e$_{OCN}$(b)** contained a fragment ion with *m/z* of 245, which could be associated with an ion containing a new C–N bond. When considered together, these results suggested products that correspond to both NCO$^{\bullet}$ and OCN$^{\bullet}$ rebound, as well as secondary reactions that these products undergo to result in compounds **1e$_{OCN}$(a)** and **1e$_{OCN}$(b)**.

Figure 2.12: Chromatographic analysis of SadX bioconversion in the presence of NaOCN[a]

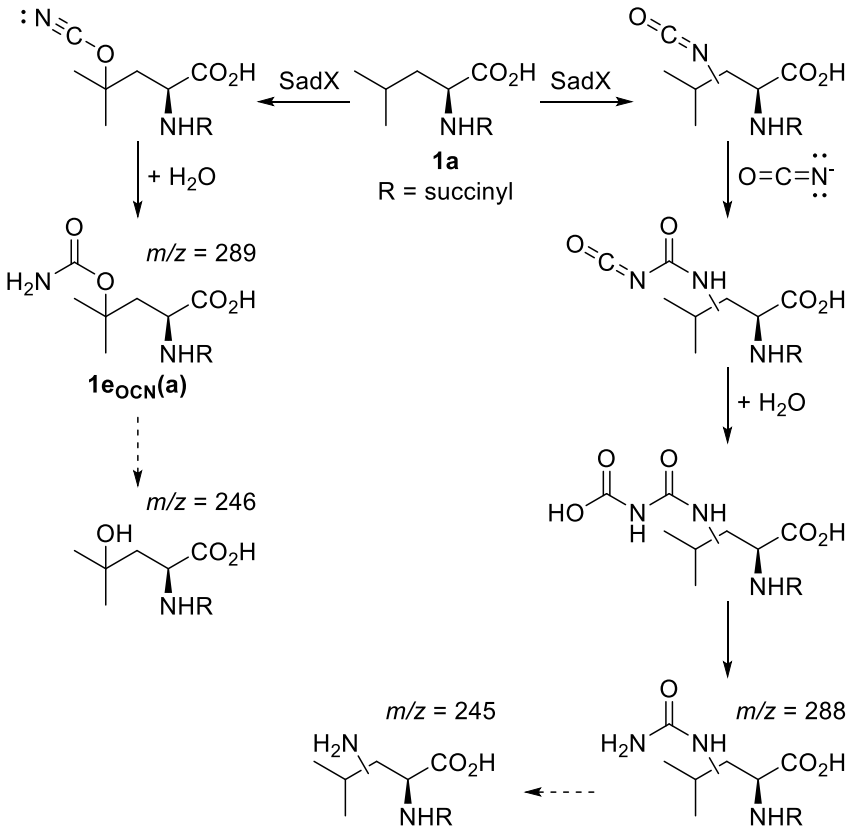


[a] Negative mode LC-MS chromatograms of a representative reaction of **1a** with SadX in the presence of NaOCN. Extracted ion chromatograms (EIC) for ions with *m/z* values of 289, 246, 288, and 245 are shown under total ion chromatogram (TIC). Vertical dashed lines highlight two peaks associated with **1e$_{OCN}$(a)** and **1e$_{OCN}$(b)**.

To further characterize **1e$_{OCN}$(a)** and **1e$_{OCN}$(b)**, the purification of SadX bioconversions was attempted using semi-preparative HPLC, but separation of these two products proved difficult. However, when bioconversions of **1a** with SadXL in the presence of NaOCN were

analyzed, only **1e$_{OCN}$(a)** was detected, alongside major product **1c**¸ simplifying the isolation of this novel product (Figure 2.11). Once characterized, product **1e$_{OCN}$(a)** was identified as the γ-functionalized primary carbamate derivative of **1a** (Scheme 2.3). Although no further characterization was performed on **1e$_{OCN}$(b)**, the rebound of a OCN• radical, followed by nucleophilic attack of an additional equivalent of OCN^- and subsequent hydration would result in a carbamide derivative of **1a** which would match the observed *m/z* value. Regardless of the true nature of product **1e$_{OCN}$(b)**, it can be concluded that SadX is able to accommodate a cyanate ligand in its active site while remaining active, as well as incorporate this non-native ligand into substrate **1a** resulting in a primary carbamate and a different novel product that suggest the formation of a C–N bond.

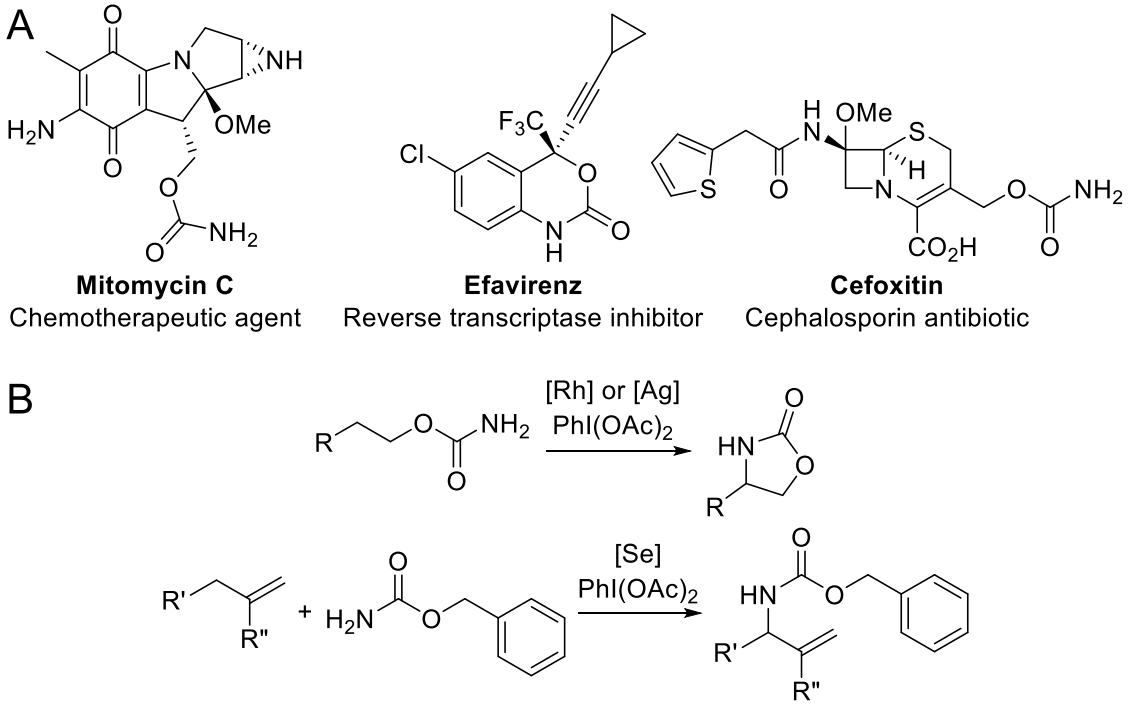
Scheme 2.3: Putative formation of products and fragments from bioconversions in the presence of NaOCN[a]



[a] *m/z* values shown correspond to the $[M - H]^-$ ions of the displayed structures. Dashed arrows represent fragmentation in the ionization chamber of the mass spectrometer.

Given the unprecedented biocatalytic C–H carbamate formation, we considered its potential synthetic utility. The relevance of carbamate synthesis can be illustrated by the prevalence of this functional group in industries such as agrochemicals, peptide and polymer synthesis, and medicinal chemistry (Figure 2.13A).[20] The relatively high chemical stability and higher membrane permeability make the carbamate bond an ideal unnatural peptide bond analogue for drug discovery. Specifically, primary carbamates have been used as starting materials or intermediates in different synthetic strategies, including intramolecular and intermolecular C–H insertion reactions, highlighting the potential of this platform to be incorporated into the chemoenzymatic synthesis of valuable target molecules (Figure 2.13B).[21–24] Currently, further exploration of the substrate scope of SadX-catalyzed carbamate synthesis is being carried out, as well as efforts to engineer chemoselective enzymes for this transformation.

Figure 2.13: Therapeutic and synthetic relevance of organic carbamates.



A) Carbamate-containing therapeutic agents.[20] B) Examples of intra- and intermolecular C–H functionalization using primary carbamates.[21–24]

2.3 Conclusions

The FeDO variant SadA D157G was improved for soluble expression by fusion with a solubility-enhancing MBP tag for the purpose of aiding future engineering and evolution efforts. The activity of this new variant, SadX, was studied in the presence of substrate *N*-succinyl-L-leucine and a panel of non-native anions. All coordinating anions were able to rescue the native hydroxylation activity, and C–H functionalization resulting from chloride, bromide, azide, and cyanate rebound was detected. Additionally, the effect of active site mutation F152L was investigated, resulting in a novel site selectivity that favored γ-hydroxylation over the native β-hydroxylation. Finally, the unprecedented incorporation of cyanate by a FeDO was further explored, identifying one of the products as a primary carbamate. These studies not only showcase the ability of SadA's active site to accommodate non-native ligands while remaining active, but also provide an understanding of non-native rebound chemistry in FeDO that will aid the engineering of these enzymes as biocatalyst for non-native reactions.

2.4 Experimental

2.4.1 Materials

Unless otherwise noted, all reagents were obtained from commercial suppliers and used without further purification. A gene encoding SadA D157G was obtained as a synthetic gene from Integrated DNA Technologies. *E. coli* BL21-Gold (DE3) cells (#230132) were purchased from Agilent Technologies. BamHI-HF (NEB #R3136) and HindIII-HF (NEB #R3104) restriction enzymes, Q5 HF DNA polymerase (NEB # M0491), dNTPs (NEB #N0447), Dpn I (NEB #R0176), and Gibson Assembly Master Mix (NEB #E2611) were purchased from New England Biolabs. Oligonucleotides were obtained from Integrated DNA Technologies. PrimeSTAR Max DNA Polymerase was purchased from Takara Bio USA. Gene confirmation by

Sanger sequencing for SadA, SadA D157G, and SadX was performed at the University of Chicago Comprehensive Cancer Center DNA Sequencing & Genotyping Facility (900 E. 57th Street, Room 1230H, Chicago, IL 60637), and all remaining genes were sequenced by Quintarabio (625 Mt Auburn St, Suite 105, Cambridge, MA 02138). Ni-nitrilotriacetic acid (Ni-NTA) resin and Pierce BCA Protein Assay Kits were purchased from Fisher Scientific International, Inc., and the manufacturer's instructions were followed when using both products. Dowex™ 50WX8 strong cation exchange resin was purchased from Sigma-Aldrich. A gene encoding desuccinylase LasA was obtained as a synthetic gene from Twist Bioscience and expressed following published protocols.[12]

2.4.2 General procedures

PCR amplification was carried out using an Applied Biosystems ProFlex PCR System thermocycler. DNA transformation by electroporation was performed using a Bio-Rad MicroPulser. Protein concentrations were measured using a Tecan Infinite 200 PRO plate reader on a Tecan NanoQuant plate or with Bradford assay. Calculated extinction coefficients (Benchling [Biology Software], https://benchling.com) were used when determining protein concentration using absorbance at 280 nm (36,440 $M^{-1}cm^{-1}$ for SadA, 104,280 $M^{-1}cm^{-1}$ for all SadX variants, and 22,920 $M^{-1}cm^{-1}$ for LasA). NMR spectra (^{1}H and ^{13}C) were obtained using a Varian 500 MHz Inova NMR spectrometer, or a Varian 400 MHz Inova NMR Spectrometer at room temperature. Chemical shifts are reported in ppm and coupling constants are reported in Hz. LC-ESI-MS analysis was performed using an Agilent 1290 system equipped with an Agilent ZORBAX Eclipse Plus C18 column (2.1 x 50 mm, 1.8 µm).

The LC-MS method applied to these studies used as solvents 0.1% formic acid in water (solvent A), and 0.1% formic acid in acetonitrile (solvent B), a flow rate of 0.4 mL/min, and ESI

negative scan mode detection. The following solvent gradient was employed: 5% B (0-1.25 min), 5 to 15% B (1.25-3.3 min), 15% B (3.3-5 min), 15 to 95% B (5-8 min), 95% B (8-10 min), 95 to 5% B (10-12 min). Table 2.1 shows the *m/z* being tracked for the relevant species and their elution times in each LC-MS method used. An *m/z* value of 184 was tracked for **1d$_{Cl}$** and **1d$_{Br}$**, which corresponds to the dominant fragment ion.[9]

Table 2.1: Dominant ions for relevant species monitored via LC-MS

Species	***m/z***
1a	230
1b/**1c**	246
1d$_{Cl}$/**1d$_{Br}$**	184
1e$_{N3}$	271
1e$_{OCN}$(a)	289
1e$_{OCN}$(b)	288

2.4.3 Molecular cloning

Standard molecular cloning procedures were used throughout this study.[25] Details for the cloning of relevant constructs are detailed below.

pET28a(SadA D157G)

The synthetic gene for SadA D157G was cloned into a pET28a plasmid vector using BamHI and HindIII restriction sites, downstream of an N-terminal His_6 tag.

pET28a(SadA)

The wild-type enzyme SadA was obtained by reverting the D157G mutation in SadA D157G via SOE PCR.[26] A pET28a(SadA D157G) plasmid was used to generate the desired insert and the linearized vector with oligonucleotides shown in Table 2.2.

Table 2.2: Oligonucleotides used to generate pET28a(SadA) via SOE PCR

#	Name	Sequence
1	SadAGenFP	5’ – AGC AAA TGG GTC GCG GAT CCA TGC AGC ATA CCT ATC CGG C – 3’
2	SadAG157DMutRP	5’ – GTA TCA CGA CCA TAG CTA ACA TCA TCA CAA TGC GGT GCA AAG TTA AAC GGT G – 3’
3	SadAG157DMutFP	5’ – GTT AGC TAT GGT CGT GAT ACC GTT AAT TGG CCT CTG AAA CGT AGC TTT CC – 3’
4	SadAGenRP	5’ – TCG AGT GCG GCC GCA AGC TTT CAA TCA AAC ATA CGC CAA CC – 3’
5	SadAVectorFP	5’ – GTT GGC GTA TGT TTG ATT GAA AGC TTG CGG CCG CAC TCG A – 3’
6	SadAVectorRP	5’ – GCC GGA TAG GTA TGC TGC ATG GAT CCG CGA CCC ATT TGC T – 3’

The SadA insert was obtained by a first step including two fragmentation reactions using the same template but different oligonucleotide pairs (i.e., fragment 1 with primers **1** and **2**, and fragment 2 with primers **3** and **4**). The vector was linearized using primers **5** and **6.** The amplification conditions were the following: 1 ng/µL template DNA, 1x Q5 reaction buffer, 0.8 mM dNTPs each, 1 µM forward primer, 1 µM reverse primer, and 0.02 U/µL Q5 HF DNA polymerase, with a final volume of 50 µL. The PCR program for the fragmentations was the following: 98 °C for 2 minutes; 24 cycles of 98 °C for 30 seconds, 56 °C for 30 seconds, and 72 °C for 45 seconds; and ending with 72 °C for 10 minutes. The PCR program for the amplification of the vector was the following: 98 °C for 2 minutes; 24 cycles of 98 °C for 30 seconds, 65 °C for 30 seconds, and 72 °C for 3.5 minutes; and ending with 72 °C for 10 minutes. All PCR products were digested with 10 units of DpnI at 37 °C for 2 hours and purified through 1% agarose gel electrophoresis.

The SadA fragments were assembled via PCR using oligonucleotides **1** and **4**. The assembly PCR conditions were as follows: 2 ng/µL of each fragment, 1x Q5 reaction buffer, 0.8 mM dNTPs each, 1 µM forward primer, 1 µM reverse primer, and 0.02 U/µL Q5 HF DNA polymerase, with the final reaction volumes being 50 µL. The amplification program consisted

of the following: 98 °C for 2 minutes; 24 cycles of 98 °C for 30 seconds, 56 °C for 30 seconds, and 72 °C for 45 seconds; and ending with 72 °C for 10 minutes. The PCR product was digested with 10 units of DpnI at 37 °C for 2 hours and purified with 1% agarose gel electrophoresis.

The resulting SadA insert and pET28a linearized vector were joined using Gibson Assembly, incubating at 50 °C for 1 hour and using a vector:insert ratio of 1:10. The resulting product was purified and concentrated using a Zymo PCR clean and concentrator kit.

pET28a(SadX)

The gene encoding the MBP tag was amplified from the pLIC(MBP-RebF) plasmid,[27] and the linearized vector was amplified from a pET28a(SadA D157G) plasmid using the oligonucleotides shown in Table 2.3. The amplification conditions were the following: 1 ng/µL template DNA, 1x Q5 reaction buffer, 0.8 mM dNTPs each, 1 µM forward primer, 1 µM reverse primer, and 0.02 U/µL Q5 HF DNA polymerase, with a final volume of 50 µL. The PCR program for the amplification of the MBP insert (using oligonucleotides **7** and **8**) was the following: 98 °C for 2 minutes; 30 cycles of 98 °C for 30 seconds and 72 °C for 90 seconds; and ending with 72 °C for 10 minutes. The PCR program for the amplification of the vector (using oligonucleotides **9** and **10**) was the following: 98 °C for 2 minutes; 30 cycles of 98 °C for 30 seconds and 72 °C for 4 minutes; and ending with 72 °C for 10 minutes. All PCR products were digested with 10 units of DpnI at 37 °C for 2 hours and purified with 1% agarose gel electrophoresis. The resulting MBP insert and pET28a(SadA D157G) linearized vector were joined using Gibson Assembly, incubating at 50 °C for 1 hour and using a vector:insert ratio of 1:10. The resulting product was purified and concentrated using a Zymo PCR clean and concentrator kit.

Table 2.3: Oligonucleotides used to generate pET28a(SadX)

#	Name	Sequence
7	NTermMBPInsertFP	5’ – GCG GCA GCC ATA TGG CTA GCA TGC ACC ATC ACC ATC ACC ATG G – 3’
8	NTermMBPInsertRP	5’ – TAG GTA TGC TGC ATG GAT CCG GCT GCT CCC TGG AAA TAC AGG – 3’
9	NTermMBPVectorFP	5’ – GTA TTT CCA GGG AGC AGC CGG ATC CAT GCA GCA TAC CTA TCC GG – 3’
10	NTermMBPVectorRP	5’ – CCA TGG TGA TGG TGA TGG TGC ATG CTA GCC ATA TGG CTG CCG C – 3’

pET28a(SadXL)

The variant SadX F152L (SadXL) was obtained by site-directed mutagenesis of SadX following a modified QuickChange™ protocol.[28] The pET28(SadX) plasmid was used as template for amplification using the oligonucleotides shown in Table 2.4. The amplification conditions were the following: 2 ng/µL template DNA, 0.3 µM forward primer, 0.3 µM reverse primer, and 1x PrimeSTAR Max DNA Polymerase, with the final volume being 50 µL. The PCR program for the amplifications was the following: 98 °C for 10 seconds; 30 cycles of 98 °C for 10 seconds, 55 °C for 5 seconds, and 72 °C for 45 seconds; and ending with 72 °C for 1 minute. The PCR products were digested with 10 units of DpnI at 37 °C for 1 hour and purified via 1% agarose gel electrophoresis.

Table 2.4: Oligonucleotides used to introduce mutation G157D in SadX and SadXL

#	Name	Sequence
11	SadXF152LFP	5’ – CGT GTT TAT GAA CCG TTT GAA GCA CCG TTT AAC TTA GCA CCG CAT TGT GG – 3’
12	SadXF152LRP	5’ – GTG CTT CAA ACG GTT CAT AAA CAC GAC GGG TTG CCA GCA GAA ATT CAA C – 3’

pET28a(SadXL G157D)

The modified QuickChange™ protocol was used to revert the active site of SadXL to the wild-type sequence by introducing the G157D mutation. The pET28a(SadXL) plasmid obtained

before was used as template for amplification using the oligonucleotides detailed in Table 2.5.

The procedure was analogous to the one used to obtain pET28a(SadXL).

Table 2.5: Oligonucleotides used to introduce G157D mutation in SadXL

#	Name	Sequence
13	SadXLG157DFP	5’ – GTT AGC TAT GGT CGT GAT ACC GTT AAT TGG CCT CTG AAA CGT AGC TTT CC – 3’
14	SadXLG157DRP	5’ – GTA TCA CGA CCA TAG CTA ACA TCA TCA CAA TGC GGT GCT AAG TTA AAC GGT G – 3’

2.4.4 Protein sequences

SadA
MQHTYPAQLMRFGTAARAEHMTIAAAIHALDADEADAIVMDIVPDGERDAWWDDEGF
SSSPFTKNAHHAGIVATSVTLGQLQREQGDKLVSKAAEYFGIACRVNDGLRTTRFVRLF
SDALDAKPLTIGHDYEVEFLLATRRVYEPFEAPFNFAPHCDDVSYGRDTVNWPLKRSFP
RQLGGFLTIQGADNDAGMVMWDNRPESRAALDEMHAEYRETGAIAALERAAKIMLKP
QPGQLTLFQSKNLHAIERCTSTRRTMGLFLIHTEDGWRMFD*

MBP-SadA
MGSSHHHHHHSSGLVPRGSHMASMHHHHHHGKIEEGKLVIWINGDKGYNGLAEVGKK
FEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQ
DKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKS
ALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKH
MNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGV
LSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAAT
MENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSNNNNTSEN
LYFQGAAGSMQHTYPAQLMRFGTAARAEHMTIAAAIHALDADEADAIVMDIVPDGERD
AWWDDEGFSSSPFTKNAHHAGIVATSVTLGQLQREQGDKLVSKAAEYFGIACRVNDGL
RTTRFVRLFSDALDAKPLTIGHDYEVEFLLATRRVYEPFEAPFNFAPHCDDVSYGRDTV
NWPLKRSFPRQLGGFLTIQGADNDAGMVMWDNRPESRAALDEMHAEYRETGAIAALE
RAAKIMLKPQPGQLTLFQSKNLHAIERCTSTRRTMGLFLIHTEDGWRMFD*

SadX
MGSSHHHHHHSSGLVPRGSHMASMHHHHHHGKIEEGKLVIWINGDKGYNGLAEVGKK
FEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQ
DKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKS
ALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKH
MNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGV
LSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAAT
MENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSNNNNTSEN
LYFQGAAGSMQHTYPAQLMRFGTAARAEHMTIAAAIHALDADEADAIVMDIVPDGERD
AWWDDEGFSSSPFTKNAHHAGIVATSVTLGQLQREQGDKLVSKAAEYFGIACRVNDGL
RTTRFVRLFSDALDAKPLTIGHDYEVEFLLATRRVYEPFEAPFNFAPHCGDVSYGRDTV

NWPLKRSFPRQLGGFLTIQGADNDAGMVMWDNRPESRAALDEMHAEYRETGAIAALE
RAAKIMLKPQPGQLTLFQSKNLHAIERCTSTRRTMGLFLIHTEDGWRMFD*

SadXL

MGSSHHHHHHSSGLVPRGSHMASMHHHHHHGKIEEGKLVIWINGDKGYNGLAEVGKK
FEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQ
DKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKS
ALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKH
MNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGV
LSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAAT
MENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSNNNNTSEN
LYFQGAAGSMQHTYPAQLMRFGTAARAEHMTIAAAIHALDADEADAIVMDIVPDGERD
AWWDDEGFSSSPFTKNAHHAGIVATSVTLGQLQREQGDKLVSKAAEYFGIACRVNDGL
RTTRFVRLFSDALDAKPLTIGHDYEVEFLLATRRVYEPFEAPFNLAPHCGDVSYGRDTV
NWPLKRSFPRQLGGFLTIQGADNDAGMVMWDNRPESRAALDEMHAEYRETGAIAALE
RAAKIMLKPQPGQLTLFQSKNLHAIERCTSTRRTMGLFLIHTEDGWRMFD*

SadXL G157D

MGSSHHHHHHSSGLVPRGSHMASMHHHHHHGKIEEGKLVIWINGDKGYNGLAEVGKK
FEKDTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQ
DKLYPFTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKS
ALMFNLQEPYFTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKH
MNADTDYSIAEAAFNKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGV
LSAGINAASPNKELAKEFLENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAAT
MENAQKGEIMPNIPQMSAFWYAVRTAVINAASGRQTVDEALKDAQTNSSSNNNNTSEN
LYFQGAAGSMQHTYPAQLMRFGTAARAEHMTIAAAIHALDADEADAIVMDIVPDGERD
AWWDDEGFSSSPFTKNAHHAGIVATSVTLGQLQREQGDKLVSKAAEYFGIACRVNDGL
RTTRFVRLFSDALDAKPLTIGHDYEVEFLLATRRVYEPFEAPFNLAPHCDDVSYGRDTV
NWPLKRSFPRQLGGFLTIQGADNDAGMVMWDNRPESRAALDEMHAEYRETGAIAALE
RAAKIMLKPQPGQLTLFQSKNLHAIERCTSTRRTMGLFLIHTEDGWRMFD*

2.4.4 Gene expression and cell culture

Plasmid or nicked plasmid DNA from the cloning procedures outlined above were used to transform electrocompetent *E.* coli BL21-Gold (DE3). Cells were recovered in 1 mL of SOC media, incubated at 37 °C for 1 hour, and 200 µL were plated onto a LB/agar plate containing 50 µg/mL kanamycin. After incubating the plate at 37 °C overnight, a single colony was picked and used to inoculate 5 mL of TB media with 50 µg/mL kanamycin. This culture was incubated at 37 °C and 250 rpm overnight. The resulting culture was partially used to generate a glycerol stock by combining 500 µL of the culture with 500 µL of 50% v/v water/glycerol and being stored

at -80 °C, while the remaining culture was used to extract the protein-encoding plasmid and confirm DNA sequence by Sanger sequencing.

An overnight culture was prepared by inoculating kanamycin-containing TB (7.5 mL) and shaking the culture at 250 rpm, 37 °C overnight. The overnight culture was used to inoculate 750 mL of fresh kanamycin-containing LB media in a 2.8 L Fernbach flask. The culture was incubated at 37 °C, 250 rpm until OD_{600} of the culture reached 0.6 – 0.8 (at ~t = 4 h), after which the culture was cooled to 18 °C without shaking. Protein expression was then induced by 1 mM IPTG. The induced culture was allowed to grow for 16 hours at 18 °C, 250 rpm. The cells were harvested afterwards by centrifugation at 4 °C, 4000 rpm for 20 minutes.

2.4.5 Evaluation of expression levels

Analysis of expression levels via SDS-PAGE

The following protocol was used to compare expression levels of SadA D157G, SadA D157G expressed in the presence of molecular chaperones GroEL/GroES, and the fusion SadX. An overnight culture was prepared by inoculating kanamycin-containing TB (5 mL) and shaking the culture at 250 rpm, 37 °C overnight. 60 µL of the resulting overnight culture were used to inoculate 6 mL of LB media containing 50 µg/mL kanamycin in a well of a 24-deep well plate. The culture was incubated at 37 °C, 215 rpm until OD_{600} of the culture reached 0.6 – 0.8, after which the culture was cooled to 18 °C and protein expression was induced by 1 mM IPTG. The induced culture was allowed to grow for 18 hours at 18 °C, 215 rpm. The cells were harvested by centrifugation at 4 °C, 3600 rpm for 15 minutes. The resulting cell pellets were resuspended with 400 µL of 1 mg/mL lysozyme in 10 mM HEPES pH 7.5 and incubated for 45 minutes at 37 °C, 250 rpm. The plate containing the cell suspension was flash frozen in liquid N_2 to promote complete cell lysis, and then thawed at room temperature. 40 µL of 1 mg/mL DNase in 10 mM

HEPES pH 7.5 were added and the cell suspension was incubated for 15 minutes at 37 °C, 250 rpm. The insoluble and soluble fractions of the generated lysate were separated by centrifugation at 4 °C, 3600 rpm for 15 minutes. The supernatant was transferred to a separate vessel ("soluble fraction"), while the pellet was resuspended in 440 µL of 10 mM HEPES pH 7.5 ("insoluble fraction").

The lysate fractions were analyzed via SDS-PAGE. 12 µL of each fraction were diluted with 12 µL of 2x Laemmli Sample Buffer (Bio-Rad) prepared according to manufacturer's instructions and incubated at 95 °C for 10 minutes. 6 µL of the resulting mixture were loaded onto a 12% acrylamide SDS-PAGE gel and run at 120 V for 95 minutes using a Bio-Rad PowerPac Basic power supply.

Activity assay in lysate

An overnight culture was prepared by inoculating kanamycin-containing TB (5 mL) and shaking the culture at 250 rpm, 37 °C overnight. 50 µL of the resulting overnight culture were used to inoculate 1 mL of LB media containing 50 µg/mL kanamycin in a well of a 96-deep well plate. The culture was incubated at 37 °C, 215 rpm until OD_{600} of the culture reached 0.6 – 0.8, after which the culture was cooled to 18 °C and protein expression was induced by 1 mM IPTG. The induced culture was incubated for 18 hours at 18 °C, 215 rpm. The cells were harvested by centrifugation at 4 °C, 3600 rpm for 15 minutes. The resulting cell pellets were resuspended with 125 µL of 1 mg/mL lysozyme in 10 mM HEPES pH 7.5 and incubated for 45 minutes at 37 °C, 250 rpm. The plate containing the cell suspension was flash frozen in liquid N_2 to promote complete cell lysis, and then thawed at room temperature. 10 µL of 1 mg/mL DNase in 10 mM HEPES pH 7.5 were added and the cell suspension was incubated for 15 minutes at 37 °C, 250

rpm. The insoluble and soluble fractions of the generated lysate were separated by centrifugation at 4 °C, 3600 rpm for 15 minutes.

Bioconversions were set up in microtiter v-bottom 96-well plates by combining 25 µL/well of solution A (50 mL containing 8 mM α-ketoglutarate, 200 mM NaCl, 32 mM ascorbic acid, and 0.4 mM $Fe(NH_4)_2(SO_4)_2$ in 10 mM HEPES pH 7.5), 25 µL/well of solution B (30 mL of 4 mM **1a** in 10 mM HEPES pH 7.5), and 50 µL/well of clarified lysate. The final reaction conditions were 100 µL of 1 mM **1a**, 2 mM α-ketoglutarate, 50 mM NaCl, 8 mM ascorbic acid, 0.1 mM $Fe(NH_4)_2(SO_4)_2$, and 50% v/v lysate in 10 mM HEPES pH 7.5. The plates were sealed with breathable film and shaken at room temperature and 750 rpm for 3 hours, after which 100 µL of methanol were added to quench the reactions and precipitated proteins were removed by centrifugation at 3600 rpm for 10 minutes. 50 µL of the resulting supernatant were diluted with 100 µL of water, filtered through 0.22 µm filter plates by centrifugation at 3600 rpm for 10 minutes, and analyzed via LC-MS.

2.4.6 Protein purification

Protein purification was carried out using Ni-NTA affinity chromatography. Cell pellets were re-suspended in Ni-NTA binding buffer (100 mM HEPES, 20 mM imidazole, pH 7.5, 30 mL) and sonicated (40 amplitude, 30 second burst, 10 minute total process). The lysed culture was then clarified at 15000 rpm, 4 °C for 60 minutes. The supernatant was tumbled with Ni-NTA resin (5 mL) pre-equilibrated with an equilibration buffer (20 mM sodium phosphate pH 7.4, containing 10 mM imidazole, and 50 mM NaCl) at 4 °C for 1 h. After the initial elution of sample, the resin was washed with 5 CV of a wash buffer (20 mM sodium phosphate pH 7.4, containing 20 mM imidazole, and 50 mM NaCl). The desired enzyme was eluted with 5 CV of elution buffer (20 mM sodium phosphate pH 7.4, containing 500 mM imidazole, and 50 mM

NaCl), with each CV of eluent collected as a separate fraction. Fraction purity was confirmed by SDS-PAGE (12% acrylamide). Pure fractions were combined, and buffer exchanged into 5 mM EDTA in 100 mM HEPES pH 7.5 to remove any residual iron ions. The purified protein was then buffer exchanged into 100 mM HEPES pH 7.5 three times, concentrated to ~ 1 mM, and stored as a frozen stock at -80 °C.

2.4.7 Activity assays

*Small-scale bioconversions of **1a** with SadA*

Initially, the reported activity of SadA was reproduced small-scale bioconversions. Reactions were set up by combining in a 96-well plate 25 µL/well of solution A (25 mL containing 8 mM α-ketoglutarate, 32 mM ascorbic acid, and 0.4 mM $Fe(NH_4)_2(SO_4)_2$ in 10 mM HEPES pH 7.5), 25 µL/well of solution B (2 mL of 4 mM **1a** in 10 mM HEPES pH 7.5), and 50 µL/well of solution C (200 µL of 0.2 mM of SadA). The final reactions (100 µL) had final concentrations of 1 mM **1a**, 2 mM α-ketoglutarate, 8 mM ascorbic acid, 0.1 mM $Fe(NH_4)_2(SO_4)_2$, and 0.1 mM SadA in 10 mM HEPES pH 7.5. The plate was sealed with breathable film and shaken at room temperature and 750 rpm for 3 hours, after which 100 µL of methanol were added to quench the reactions and precipitated proteins were removed by centrifugation at 3600 rpm for 10 minutes. Finally, 80 µL of the resulting supernatant were filtered through 0.22 µm filter plates by centrifugation at 3600 rpm for 10 minutes and analyzed via LC-MS.

*Small-scale bioconversions of **1a** with SadA D157G/SadX*

The reaction setup was analogous to hydroxylation bioconversions with SadA, with the exception of solution A containing 8 mM α-ketoglutarate, 200 mM NaCl, 32 mM ascorbic acid, and 0.4 mM $Fe(NH_4)_2(SO_4)_2$, and solution C being 200 µL of 0.2 mM SadA D157G or SadX.

This resulted in reactions (100 μL) with final concentrations of 1 mM **1a**, 2 mM α-ketoglutarate, 50 mM NaCl, 8 mM ascorbic acid, 0.1 mM $Fe(NH_4)_2(SO_4)_2$, and 0.1 mM SadA D157G or SadX in 10 mM HEPES pH 7.5.

*Reaction of **1a** with SadX/SadXL in the presence of exogenous anions*

SadX and SadXL activity was analyzed in reactions prepared from three stock solutions. Solution A (1 mL containing 3 mM **1a**, 30 μM enzyme, 0.3 mM $Fe(NH_4)_2(SO_4)_2$, and 24 mM ascorbic acid in 100 mM HEPES pH 7.5) was prepared by combining 300 μL of 10 mM **1a**, 300 μL of 80 mM ascorbic acid, and 4.5 μL of 2 mM enzyme, all in 100 mM HEPES pH 7.5, with 365.5 μL of 100 mM HEPES (pH 7.5) and 30 μL of 10 mM $Fe(NH_4)_2(SO_4)_2$ in LC-MS-grade water (and stored on ice to avoid rapid air oxidation in the absence of ascorbic acid). Solutions B and C, containing 150 mM of the desired salt and 30 mM α-ketoglutarate, respectively, were prepared in 100 mM HEPES (pH 7.5). 30 μL of solutions A-C were added to a 96-well v-bottom plate. The final reaction volume was 90 μL, and the reactions contained 1 mM substrate, 10 μM SadX (or variant), 50 mM salt (or no salt as a control), 0.1 mM $Fe(NH_4)_2(SO_4)_2$, 8 mM ascorbic acid, and 10 mM α-ketoglutarate. Plates were sealed with a breathable film and shaken at 750 rpm at room temperature overnight. The reactions were quenched the following day by the addition of methanol (90 μL). The precipitated protein was then removed by centrifugation. The supernatant (60 μL) was diluted with water (100 μL), filtered, and analyzed by LC-MS with an injection volume of 0.5 μL. Products and substrate peak areas in the resulting extract ion chromatograms were integrated on the Agilent MassHunter software. Product conversions were calculated using relative peak area ratios between product(s) and the substrate.

2.4.8 Synthesis and Biocatalysis

*Synthesis of substrate N-succinyl-L-leucine (**1a**)*

A 250 mL round bottom flask was charged with a stir bar and L-Leucine (10 g, 60.5 mmol, 1 eq). Water (50 mL) and NaOH (5 M, 12 mL) was added to leucine and the resulting suspension was stirred. Succinic anhydride (6.3 g, 63 mmol, 1.03 eq) was added as solids to the suspension. NaOH (5 M, 20 mL) was added to the suspension again. The reaction was heated to 47 °C for 2 hours, after which the reaction was cooled to room temperature. The reaction was then acidified to pH 1 with 6 M HCl, diluted with water (300 mL), and extracted with ethyl acetate (400 mL x 2). The combined organic layers were washed with brine (300 mL x 1), dried over $MgSO_4$, and filtered. The filtrate was concentrated to wet white solids under reduced pressure. Slow evaporation of residual solvent at room temperature afforded white crystalline solids over a few days. The product was washed with hexanes and dried *in vacuo* with some heating with a heat gun. The resulting sticky solids were lyophilized to afford dry white solids. Yield: 7.4 g, 53%. The ^{1}H-NMR spectrum of product matched the spectrum reported in literature.[9]

*Isolation of **1e$_{N3}$** via SadX-3 reaction with **1a***

A 1 L Erlenmeyer flask was charged with **1a** (50 mM, 30 mL, 1.5 mmol), SadX-3 (SadX V39I/F152L) (0.1 mol%, 0.75 mM, 2.0 mL, 1.5 μmol), $Fe(NH_4)_2(SO_4)_2$ (100 mM, 0.15 mL, 15 μmol), ascorbic acid (80 mM, 30 mL, 2.4 mmol), NaN_3 (50 mM, 30 mL, 1.5 mmol), and α-ketoglutarate (200 mM, 30 mL, 6 mmol) in 100 mM HEPES, pH 7.5. Buffer was added to make up to a 150 mL reaction volume. All stocks were made in 100 mM HEPES, and pH adjusted to pH 7.5 with 10 M NaOH prior to addition, with the exception of the stock of $Fe(NH_4)_2(SO_4)_2$ which was prepared in water. The flask was sealed with a breathable film and

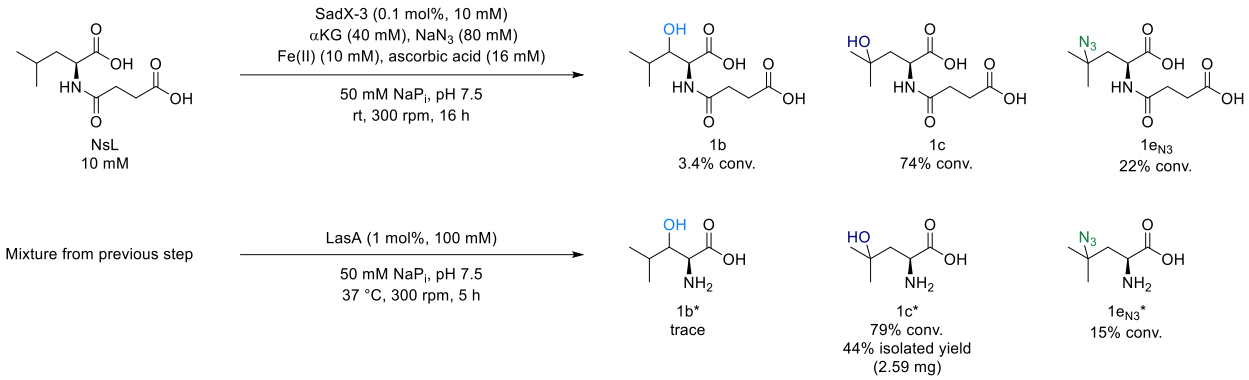
shaken at 60 rpm overnight at 30 °C. Complete consumption of **1a** on the next day was confirmed by LC-MS of the reaction mixture. The reaction was lyophilized and then resuspended in methanol (150 mL). The suspension was clarified by centrifugation at 4,000 rpm for 5 minutes. The supernatant was mixed with celite (54 g) and methanol was removed under reduced pressure. A quarter of the celite mixture (13.5 g) was loaded onto a 60 g reverse phase Biotage column (Biotage Sfar C18 D) and **1e$_{N3}$** was purified in a Biotage Isolera One with the following gradient: 4 CV of 100% A, 4 CV of 10% B, 4 CV from 10 to 60% B, 2 CV of 100% B, where solvent A is water with 0.1% formic acid and solvent B is LC-MS-grade ACN with 0.1% formic acid. Two other rounds of reverse phase purification were performed on the remaining celite mixture using the same method. Fraction purity was checked using the LC-MS. Fractions containing **1e$_{N3}$** were consolidated, dried, embedded with celite, loaded onto the Biotage column, and purified via flash chromatography using the gradient described before. Pure fractions from all rounds of purification were combined, concentrated under reduced pressure, and lyophilized. Yield: 42.6 mg, 10.4%.

^{1}H-NMR (500 MHz, CD_3OD) δ 4.56 (dt, *J* = 9.2, 3.0 Hz, 1H), 2.69 – 2.46 (m, 4H), 2.10 (dd, *J* = 14.7, 3.4 Hz, 1H), 1.86 (dd, *J* = 14.6, 9.1 Hz, 1H), 1.34 (s, 3H), 1.32 (s, 3H).

^{13}C-NMR (126 MHz, CD_3OD) δ C 176.10, C 175.30, C 174.23, C 61.57, CH 50.57, CH_2 43.37, CH_2 31.52, CH_2 30.03, CH_3 26.47, CH_3 26.24.

Peaks were assigned based on multiplicity-edited ^{1}H-^{13}C HSQC (Figure A1.1).

*Isolation of 4-azido-L-leucine (**1c***) via tandem SadX-3 and LasA reactions with **1a***



The following reaction was set up with four replicates. A 20 mL scintillation vial was charged with **1a** (50 mM, 200 µL, 10 µmol), SadX-3 (0.1 mol%, 0.59 mM, 17 µL, 10 nmol), $Fe(NH_4)_2(SO_4)_2$ (1 mM, 10 µL, 10 nmol), ascorbic acid (80 mM, 200 µL, 16 µmol), and α-ketoglutarate (200 mM, 200 µL, 40 µmol) in 50 mM sodium phosphate buffer pH 7.5. Buffer was added to make up to a 1 mL reaction volume. All stocks were made in 50 mM sodium phosphate buffer and adjusted to pH 7.5 with 10 M NaOH prior to addition, with the exception of the stock of $Fe(NH_4)_2(SO_4)_2$ prepared in water and added last to avoid precipitation with phosphate ions. The vial was sealed with a breathable film and shaken at 300 rpm overnight at 23 °C. Complete consumption of **1a** on the next day was confirmed by LC-MS of the reaction mixture. The four reactions were transferred to four 2 mL microcentrifuge tubes. LasA (1.07 mM, 93.4 µL, 0.1 µmol) in 50 mM Tris/HCl at pH 7.5 was added to each reaction. Reactions were shaken at 300 rpm and 37 °C for five hours. Complete consumption of **1e$_{N3}$** and close to complete consumption of **1c** to their respective desuccinylated forms were observed via LC-MS analysis. The four reactions were quenched with 1 mL of methanol each, and the precipitated protein was removed via centrifugation at 21,000 g for 10 minutes. The supernatants were combined, and methanol was removed with a rotary evaporator before acidifying to pH 1 with 6 M HCl. The acidified supernatant was then purified on a cation exchange resin following the

procedure described below. The lyophilized post-cation exchange product (7.6 mg) was characterized by ^{1}H-NMR and it contained mostly **1c***, **1e$_{N3}$***, and amine-based buffers (Tris and HEPES) from the original protein stocks. This mixture was purified by semi-preparative HPLC in an Agilent 1100 HPLC equipped with a Supelco Discovery C18 semi-preparative column (25 cm × 10 mm, 5 μm particle size) at 3 mL/min with the following method: 1% ACN in water from 0 to 7 minutes, 1 to 6% ACN from 7 to 20.7 minutes, 6 to 80% ACN from 20.7 to 25 minutes. UV absorbance of eluent was detected at 210 nm. Fractions were collected manually, and their purity confirmed by LC-MS. **1c*** eluted at ~4 minutes. Pure **1c*** fractions were combined, concentrated under reduced pressure, and lyophilized. Yield: 2.59 mg, 44%.

^{1}H-NMR (500 MHz, D_2O) δ 3.96 (dd, *J* = 10.1, 3.3 Hz, 1H), 2.13 (dd, *J* = 15.3, 3.3 Hz, 1H), 1.95 (dd, *J* = 15.3, 10.2 Hz, 1H), 1.36 (s, 3H), 1.35 (s, 3H).

^{13}C-NMR (126 MHz, D_2O) δ 174.95, 70.92, 52.46, 41.91, 29.74, 26.53.

Peaks were assigned based on multiplicity-edited ^{1}H-^{13}C HSQC (Figure A1.2).

Cation exchange protocol

A 100 mL chromatography column was slurry-packed with 11 g of DOWEX™ 50WX8 resin in methanol. The resin was equilibrated by washing with 200 mL of methanol, 300 mL of water, 1 M HCl until the pH of the flow through was lower than 2, and water until the pH of the flow through was ~6-7. The protein-free supernatant from a SadX-3/LasA reaction was added to the column and the resin was washed with 500 mL of water. 1 M NH_4OH was added until product no longer eluted from the column as confirmed by LC-MS analysis. The fractions that contained products were combined and dried by lyophilization.

*Isolation of **1e$_{OCN}$(a)** via SadXL reaction with **1a***

A 250 mL polypropylene beaker was charged with **1a** (5 mM, 6 mL), SadXL (1.5 mM, 1 mL), $Fe(NH_4)_2(SO_4)_2$ (20 mM, 0.3 mL), ascorbic acid (80 mM, 6 mL), NaOCN (40 mM, 6 mL), and α-ketoglutarate (20 mM, 6 mL) in 100 mM HEPES, pH 7.5. Buffer was added to make up to a 30 mL reaction volume. All stocks were made in 100 mM HEPES, and pH adjusted to pH 7.5 with 10 M NaOH prior to addition, with the exception of the stock of $Fe(NH_4)_2(SO_4)_2$ which was prepared in water. The final concentrations were 1 mM 1a, 50 μM SadXL, 200 μM $Fe(NH_4)_2(SO_4)_2$, 16 mM ascorbic acid, 8 mM NaOCN, and 4 mM α-ketoglutarate. The beaker was sealed with a breathable film and shaken at 125 rpm and 30 °C for 3 hours. Full consumption of **1a** on was confirmed by LC-MS of the reaction mixture. The reaction was quenched by addition of 30 mL of methanol and clarified by centrifugation at 15,000 rpm for 15 minutes. The supernatant was filtered, concentrated down by rotary evaporation, and purified via semipreparative HPLC. Fraction purity was checked using LC-MS, and those containing **1e$_{OCN}$(a)** were consolidated, dried, and analyzed via NMR.

^{1}H-NMR (500 MHz, CD_3OD) δ 4.55 (dd, *J* = 9.41, 3.37 Hz, 1H), 2.63 – 2.48 (m, 4H), 2.22 (dd, *J* = 14.69, 3.34 Hz, 1H), 2.14 (dd, *J* = 14.73, 9.30 Hz, 1H), 1.50 (s, 3H), 1.49 (s, 3H).

^{13}C-NMR (126 MHz, CD_3OD) δ C 177.94, C 177.00, C 173.94, C 159.22, C 81.17, CH 51.97, CH_2 44.34, CH_2 32.10, CH_2 30.87, CH_3 26.65, CH_3 26.62.

Peaks were assigned based on multiplicity-edited ^{1}H-^{13}C HSQC (Figure A1.3).

2.5 References

(1) Matthews, M. L.; Chang, W.; Layne, andrew; Miles, L.; Krebs, C.; Martin Bollinger Jr, J. Direct Nitration and Azidation of Aliphatic Carbons by an Iron-Dependent Halogenase. *Nat. Chem. Biol.* | **2014**, *10*. https://doi.org/10.1038/nCHeMBIO.1438.

(2) 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* (10), 1009–1016. https://doi.org/10.1038/s41589-019-0355-x.

(3) Kim, C. Y.; Mitchell, A. J.; Glinkerman, C. M.; Li, F.-S.; Pluskal, T.; Weng, J.-K. The Chloroalkaloid (−)-Acutumine Is Biosynthesized via a Fe(II)-and 2-Oxoglutarate-Dependent Halogenase in Menispermaceae Plants. https://doi.org/10.1038/s41467-020-15777-w.

(4) Büchler, J.; Honda Malca, S.; Patsch, D.; Voss, M.; Turner, N. J.; Bornscheuer, U. T.; Allemann, O.; Le Chapelain, C.; Lumbroso, A.; Loiseleur, O.; Buller, R. Algorithm-Aided Engineering of Aliphatic Halogenase WelO5* for the Asymmetric Late-Stage Functionalization of Soraphens. https://doi.org/10.1038/s41467-022-27999-1.

(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* **2009**, *4* (11), e7635. https://doi.org/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.* **2007**, *101* (5), 797–808. https://doi.org/10.1016/J.JINORGBIO.2007.01.011.

(7) Chaplin, V. D.; Hangasky, J. A.; Huang, H. T.; Duan, R.; Maroney, M. J.; Knapp, M. J. Chloride Supports O2 Activation in the D201G Facial Triad Variant of Factor-Inhibiting Hypoxia Inducible Factor, an α-Ketoglutarate Dependent Oxygenase. *Inorg. Chem.* **2018**, *57* (20), 12588–12595. https://doi.org/10.1021/acs.inorgchem.8b01736.

(8) Papadopoulou, A.; Meierhofer, J.; Meyer, F.; Hayashi, T.; Schneider, S.; Sager, E.; Buller, R. Re-Programming and Optimization of a L-Proline Cis-4-Hydroxylase for the Cis-3-Halogenation of Its Native Substrate. *ChemCatChem* **2021**, *13* (18), 3914–3919. https://doi.org/10.1002/CCTC.202100591.

(9) Mitchell, A. J.; Dunham, N. P.; Bergman, J. A.; Wang, B.; Zhu, Q.; Chang, W. C.; Liu, X.; Boal, A. K. Structure-Guided Reprogramming of a Hydroxylase to Halogenate Its Small Molecule Substrate. *Biochemistry* **2017**, *56* (3), 441–444. https://doi.org/10.1021/acs.biochem.6b01173.

(10) Hibi, M.; Kawashima, T.; Kasahara, T.; Sokolov, P. M.; Smirnov, S. V.; Kodera, T.; Sugiyama, M.; Shimizu, S.; Yokozeki, K.; Ogawa, J. A Novel Fe(II)/α-Ketoglutarate-Dependent Dioxygenase from Burkholderia Ambifaria Has β-Hydroxylating Activity of N

-Succinyl l-Leucine. *Lett. Appl. Microbiol.* **2012**, *55* (6), 414–419. https://doi.org/10.1111/j.1472-765X.2012.03308.x.

(11) Qin, H. M.; Miyakawa, T.; Nakamura, A.; Hibi, M.; Ogawa, J.; Tanokura, M. Structural Optimization of SadA, an Fe(II)- and α-Ketoglutarate- Dependent Dioxygenase Targeting Biocatalytic Synthesis of N-Succinyl-L-Threo-3, 4-Dimethoxyphenylserine. *Biochem. Biophys. Res. Commun.* **2014**, *450* (4), 1458–1461. https://doi.org/10.1016/j.bbrc.2014.07.008.

(12) Hibi, M.; Kasahara, T.; Kawashima, T.; Yajima, H.; Kozono, S.; Smirnov, S. V.; Kodera, T.; Sugiyama, M.; Shimizu, S.; Yokozeki, K.; Ogawa, J. Multi-Enzymatic Synthesis of Optically Pure β-Hydroxy α-Amino Acids. *Adv. Synth. Catal.* **2015**, *357* (4), 767–774. https://doi.org/10.1002/adsc.201400672.

(13) Chan, N. H.; Gomez, C. A.; Vennelakanti, V.; Du, Q.; Kulik, H. J.; Lewis, J. C. Non-Native Anionic Ligand Binding and Reactivity in Engineered Variants of the Fe(II)- and α-Ketoglutarate-Dependent Oxygenase, SadA. *Inorg. Chem.* **2022**, *61* (36), 14477–14485. https://doi.org/10.1021/acs.inorgchem.2c02872.

(14) Mogk, A.; Mayer, M. P.; Deuerling, E. Mechanisms of Protein Folding: Molecular Chaperones and Their Application in Biotechnology**. *ChemBioChem* **2002**, *3*, 807–814. https://doi.org/10.1002/1439-7633.

(15) Esposito, D.; Chatterjee, D. K. Enhancement of Soluble Protein Expression through the Use of Fusion Tags. *Curr. Opin. Biotechnol.* **2006**, *17* (4), 353–358. https://doi.org/10.1016/J.COPBIO.2006.06.003.

(16) Nishihara, K.; Kanemori, M.; Kitagawa, M.; Yanagi, H.; Yura, T. Chaperone Coexpression Plasmids: Differential and Synergistic Roles of DnaK-DnaJ-GrpE and GroEL-GroES in Assisting Folding of an Allergen of Japanese Cedar Pollen, Cryj2, in Escherichia Coli. *Appl. Environ. Microbiol.* **1998**, *64* (5), 1694–1699. https://doi.org/10.1128/aem.64.5.1694-1699.1998.

(17) Kapust, R. B.; Waugh, D. S. Escherichia Coli Maltose-Binding Protein Is Uncommonly Effective at Promoting the Solubility of Polypeptides to Which It Is Fused . *Protein Sci.* **1999**, *8* (8), 1668–1674. https://doi.org/10.1110/PS.8.8.1668.

(18) Miyakawa, H.-M.; Jia, T.; Nakamura, M. Z.; Ohtsuka, A. Crystal Structure of a Novel N-Substituted L-Amino Acid Dioxygenase from Burkholderia Ambifaria AMMD. *PLoS One* **2013**, *8* (5), 63996. https://doi.org/10.1371/journal.pone.0063996.

(19) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nat. 2021 5967873* **2021**, *596* (7873), 583–589.

https://doi.org/10.1038/s41586-021-03819-2.

(20) Ghosh, A. K.; Brindisi, M. Organic Carbamates in Drug Design and Medicinal Chemistry. *Journal of Medicinal Chemistry*. American Chemical Society April 9, 2015, pp 2895–2940. https://doi.org/10.1021/jm501371s.

(21) Espino, C. G.; Du Bois, J. A Rh-Catalyzed C-H Insertion Reaction for the Oxidative Conversion of Carbamates to Oxazolidinones. *Angew. Chemie - Int. Ed.* **2001**, *40* (3), 598–600. https://doi.org/10.1002/1521-3773(20010202)40:3<598::AID-ANIE598>3.0.CO;2-9.

(22) Cui, Y.; He, C. A Silver-Catalyzed Intramolecular Amidation of Saturated C-H Bonds. *Angew. Chemie - Int. Ed.* **2004**, *43* (32), 4210–4212. https://doi.org/10.1002/anie.200454243.

(23) Rigoli, J. W.; Weatherly, C. D.; Alderson, J. M.; Vo, B. T.; Schomaker, J. M. Tunable, Chemoselective Amination via Silver Catalysis. *J. Am. Chem. Soc.* **2013**, *135* (46), 17238–17241. https://doi.org/10.1021/ja406654y.

(24) Obenschain, D. C.; Tabor, J. R.; Michael, F. E. Metal-Free Intermolecular Allylic C–H Amination of Alkenes Using Primary Carbamates. *ACS Catal.* **2023**, *13*, 4369–4375. https://doi.org/10.1021/acscatal.3c00807.

(25) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning: A Laboratory Manual, 2nd Edn.*; Cold Spring Laboratory Press, 1989.

(26) Ho, S. N.; Hunt, H. D.; Horton, R. M.; Pullen, J. K.; Pease, L. R. Site-Directed Mutagenesis by Overlap Extension Using the Polymerase Chain Reaction. *Gene* **1989**, *77* (1), 51–59. https://doi.org/10.1016/0378-1119(89)90358-2.

(27) Payne, J. T.; Poor, C. B.; Lewis, J. C. Directed Evolution of RebH for Site-Selective Halogenation of Large Biologically Active Molecules. *Angew. Chemie Int. Ed.* **2015**, *54* (14), 4226–4230. https://doi.org/10.1002/ANIE.201411901.

(28) Liu, H.; Naismith, J. H. An Efficient One-Step Site-Directed Deletion, Insertion, Single and Multiple-Site Plasmid Mutagenesis Protocol. *BMC Biotechnol.* **2008**, *8* (1), 91. https://doi.org/10.1186/1472-6750-8-91.

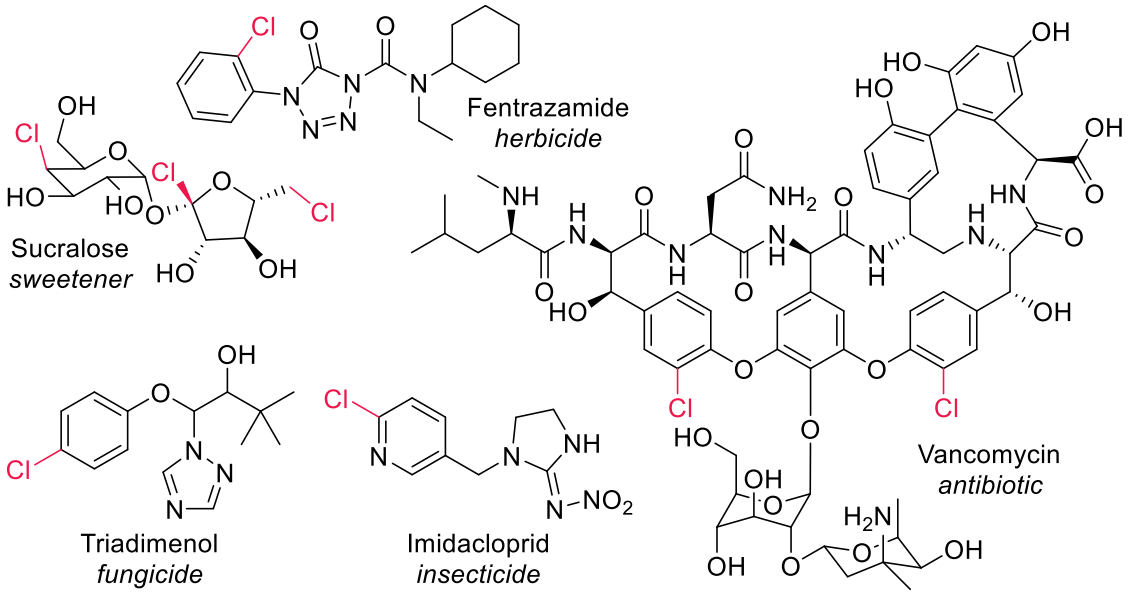
Chapter 3: Directed evolution of SadX for non-native azidation

3.1 Introduction

3.1.1 The relevance of C–H halogenation

Halogenated compounds are highly prevalent as pharmaceuticals, agrochemicals, synthetic building blocks, and biologically active molecules (Figure 3.1).[1–3] Replacing C–H with C-X bonds, where X = F, Cl, Br, or I, has been shown to increase the efficacy and biological activity of antibiotics, fungicides, herbicides, and insecticides.[2,4] Additionally, naturally occurring halogenated products have been shown to possess interesting biological activities, including antimicrobial and antitumor activities.[5,6] The discovery of new natural products and the development of efficient synthetic routes to access halogenated compounds are therefore of importance to a wide range of industries and fields.

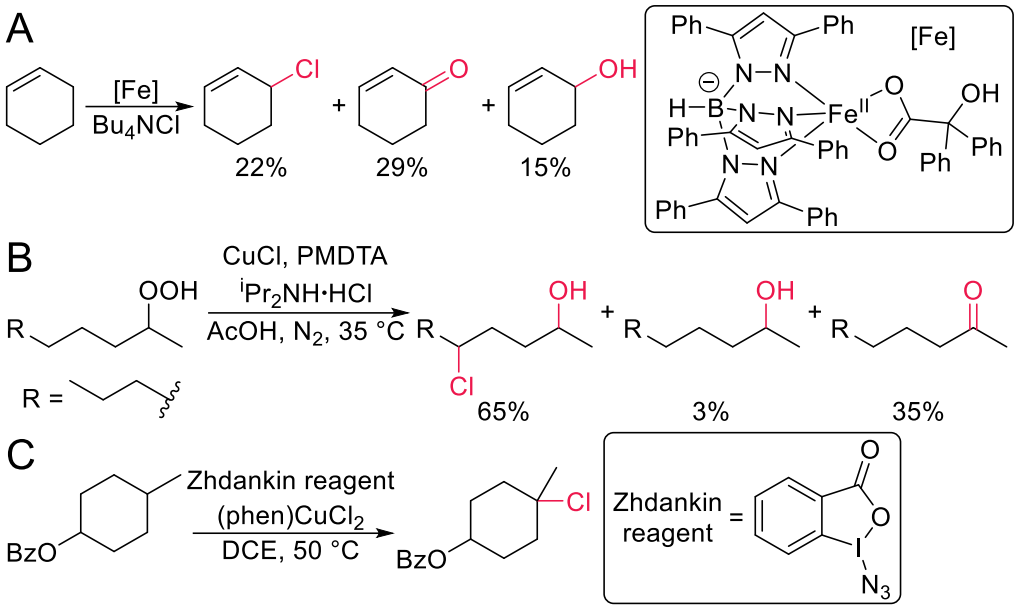
Figure 3.1: Representative halogenated compounds and their activities



Direct C–H halogenation is an intuitive and efficient strategy to access halogenated compounds. These reactions can involve the use of metal catalysts, oxidative reagents, and photocatalysts, resulting in the need for harsh conditions and hazardous chemicals.[7–11] Additionally, the selectivity of these reactions is typically substrate-controlled, resulting in low selectivities in the absence of the appropriate directing groups.[12] This effect is of particular importance in the case of aliphatic C–H halogenation given the ubiquity of this functionality. Consequently, there is a need for establishing synthetic strategies that can produce halogenated compounds in an efficient and selective manner.

Recent developments in aliphatic C–H halogenation strategies include methods inspired by halogenating enzymatic systems. For example, the halogenation of several small molecules, such as adamantane, toluene, and cyclohexene, was achieved with the use of a biomimetic Fe(II) complex in the presence of oxygen and under acidic conditions (Scheme 3.1A).[13] The proposed mechanism for this reaction involves the formation of an Fe(IV)-oxo-chloro complex followed by radical processes analogous to those of native FeDHs (Scheme 1.3). Although site-selective halogenation occurs, the formation of hydroxylated and oxidized products is also observed, attributed to be most likely due to the strong iron-chloro bond.[13]

Scheme 3.1: Representative examples of aliphatic C–H halogenation reactions



A) Biomimetic Fe(II) catalyst produces a mixture of halogenated and oxidized products.[13] B) Copper-catalyzed directed C–H halogenation selective for chlorination of alkyl substrates. PMDTA = *N,N,N',N'',N''*-pentamethyldiethylenetriamine.[14] C) Non-directed C–H chlorination of alkyl substrates based on the use of Zhdankin's azidoiodinane reagent.[15]

As mentioned above, the use of directing groups is a common strategy to overcome the site selectivity issue of these reactions. One example of such an approach is the copper-catalyzed halogenation of hydroperoxides (Scheme 3.1B).[14] The mechanism of this reaction is proposed to involve the formation of an alkoxy radical, a 1,5-hydrogen atom transfer, and a chlorine atom transfer. Several primary and secondary alkyl hydroperoxides were successfully halogenated in this study to produce the corresponding chlorinated alcohols, but oxidized byproducts were also detected.

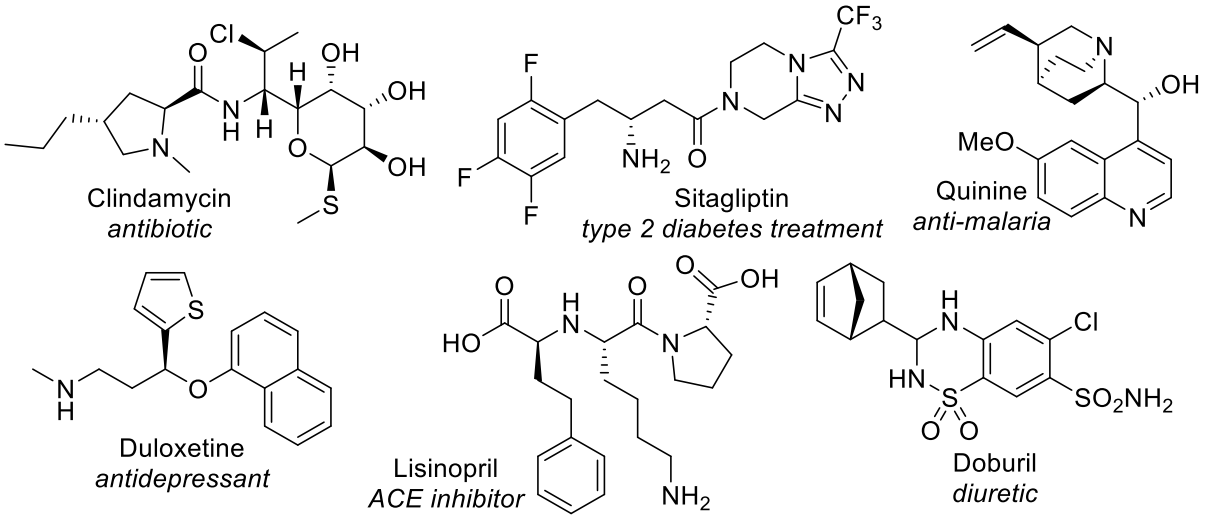
A recent study reported the non-directed and site-selective chlorination of mostly tertiary and benzylic C–H bonds in the presence of Zhdankin's azidoiodinane reagent and dichloro(1,10-phenanthroline)copper(II), an inexpensive copper(II) chloride complex (Scheme 3.1C).[15] This reaction was successfully demonstrated in more than 30 diverse compounds, including complex small molecules like botulin and gibberellic acid. The proposed mechanism starts with the

formation of an *o*-iodobenzoyloxy radical from Zhdankin's reagent, followed by hydrogen atom abstraction, and chlorine atom transfer from the copper chloride complex to the alkyl radical. Although no byproducts were reported, this strategy did not work on compounds with functionalities such as secondary alcohols, and double or triple C–C bonds.

3.1.2 The importance of C–N bond forming reactions

Aliphatic amines are one of the most prevalent functional groups in biologically active compounds and pharmaceutical candidates (Figure 3.2).[16,17] The presence of an amino group usually increases the solubility of a compound, while lipophilicity can me modulated by the *N*-substituents. This feature makes them ideal for compounds meant to cross cell membranes.[18] The ubiquity of aliphatic amines in pharmaceutical agents and preclinical candidates can be exemplified by a 2011 study that analyzed the most common reactions and functional groups used by medicinal chemists.[19] This study revealed that 42.9% of the compounds in their search contained at least one aliphatic amine.

Figure 3.2: Representative compounds containing aliphatic amines and their therapeutic use

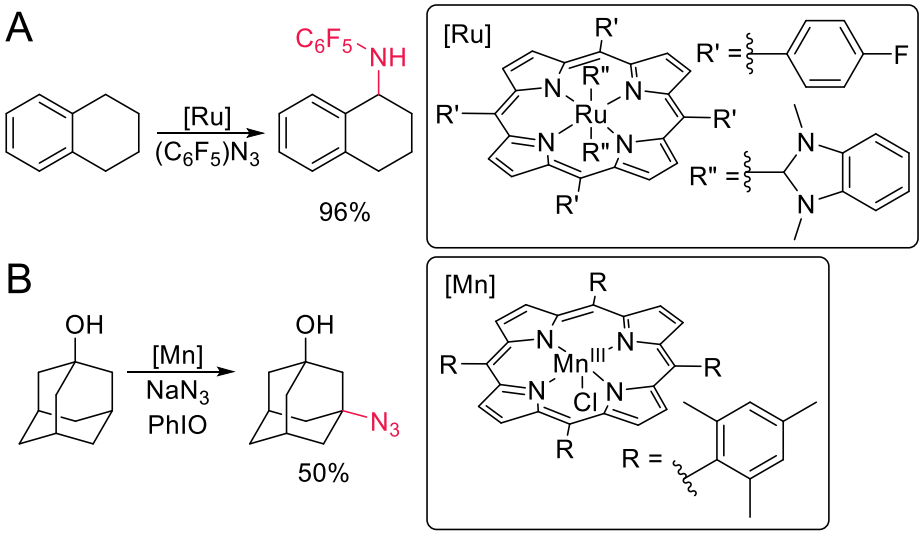


The most traditional synthetic strategies to obtain aliphatic amines are direct *N*-alkylation and reductive amination.[19] *N*-alkylation is the reaction of an amine with an alkyl halide to

produce an amine with higher substitution. Even though this strategy is widely used, it suffers from poor selectivity, with overalkylation being common. The resulting mixture of secondary and tertiary amines and ammonium salts creates downstream issues associated with product purification. Reductive amination consists of the condensation between an aldehyde or ketone with a primary or secondary amine, resulting in an imine or an iminium ion, respectively. The corresponding intermediate is then reduced (e.g., with $NaBH_3CN$) to form a new aliphatic amine. This approach, however, can suffer from reactivity issues, with sterically hindered amines and ketones resulting in low reactivity, and the opposite effect being observed with small amines and aldehydes.[20,21]

As with other types of direct C–H functionalization, direct C–N bond formation reactions can provide a more efficient way to obtain these highly valued compounds.[22,23] Several approaches using transition metal catalysts have been developed, including the use of metalloporphyrin complexes.[24] One example of these reactions include the use of a ruthenium-porphyrin complex in nitrene insertion reactions in secondary, benzylic, and allylic C–H bonds (Scheme 3.2A).[25] Another representative study in this field demonstrated the use of a manganese-porphyrin catalyst to azidate secondary, tertiary, and benzylic C–H bonds through a proposed radical mechanism (Scheme 3.2B).[26] Although several functional groups were tolerated, secondary alcohols were oxidized to ketones under these conditions.

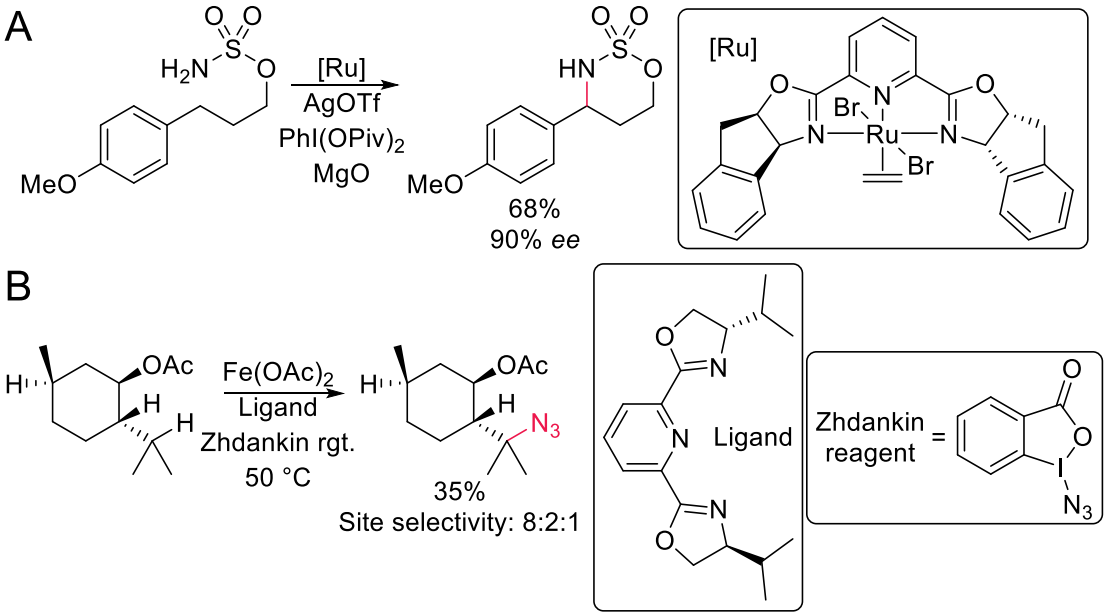
Scheme 3.2: Representative C–N bond forming reactions catalyzed by metalloporphyrin complexes



A) Amination via nitrene insertion catalyzed by a ruthenium-porphyrin complex.[25] B) Aliphatic C–H azidation catalyzed by a manganese-porphyrin complex.[26]

Non-porphyrin transition metal catalysts have also been developed for direct C–N bond formation. One example of such a catalyst is a ruthenium(II)-pybox complex capable of catalyze intramolecular C–H amination via nitrene insertion (Scheme 3.3A).[27] This study showed the successful functionalization of benzylic and allylic C–H bonds in a series of sulfamate esters, while functionalization of an aliphatic secondary C–H bond resulted in poor yields. Another approach relied on the use of Zhdankin's azidoiodinane reagent, iron(II) acetate, and a tridentate pybox-type ligand to achieve azidation of tertiary and benzylic C–H bonds (Scheme 3.3B).[28] When challenged with multiple sites for functionalization, this system yielded site selectivities as low as 5:1 and as high as 14:1. It is worth mentioning that the azidation of complex molecules, such as the steroid oestrone and a gibberellic acid derivative, was successfully achieved, highlighting the utility of this platform for late-stage functionalization.

Scheme 3.3: Representative C–N bond forming reaction catalyzed by transition metals

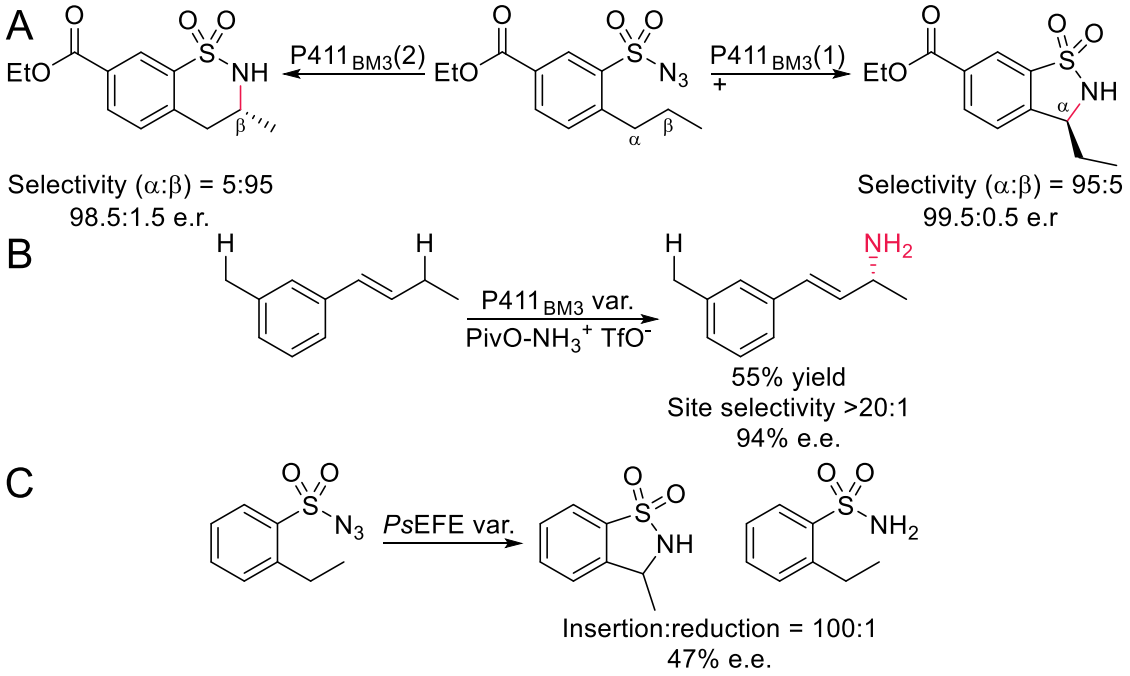


A) Intramolecular C–H amination catalyzed by a ruthenium(II)-pybox complex.[27] B) Site-selective azidation of tertiary C–H bonds with the use of Zhdankin's azidoiodinane reagent and an iron(II)-pybox complex.[28]

Enzymes have been engineered to catalyze C–N bond forming reactions as well. Heme-dependent enzymes have been developed as biocatalysts for intramolecular amination.[29–32] One of these studies engineered enzyme $P411_{BM3}$ into variants with divergent site selectivity for the formation of a six- or five-membered ring (Scheme 3.4A). Additionally, this family of enzymes is capable of intermolecular amination of benzylic, allylic, and propargylic C–H bonds using sulfonyl azides or *N*-hydroxylamine esters.[33–36] The use of engineered P411 variants with a hydroxylamine ester precursor resulted in the synthesis of primary amines at benzylic and allylic positions, with site selectivities ranging from 1.6:1 to >20:1, depending on the substrate (Scheme 3.4B). Non-heme Fe(II)-dependent enzymes have also been used to carry out intramolecular C–H amination in sulfonyl azides via nitrene insertion.[37,38] A recent example used engineered variants of the FeDO *Pseudomonas savastanoi* ethylene-forming enzyme (*Ps*EFE) to catalyze

the intramolecular cyclization of 2-ethylbenzenesulfonyl azide to the corresponding sultam, with high selectivity over the reduction side reaction (Scheme 3.4C).

Scheme 3.4: Enzymatic C–H amination reactions via non-native nitrene insertion

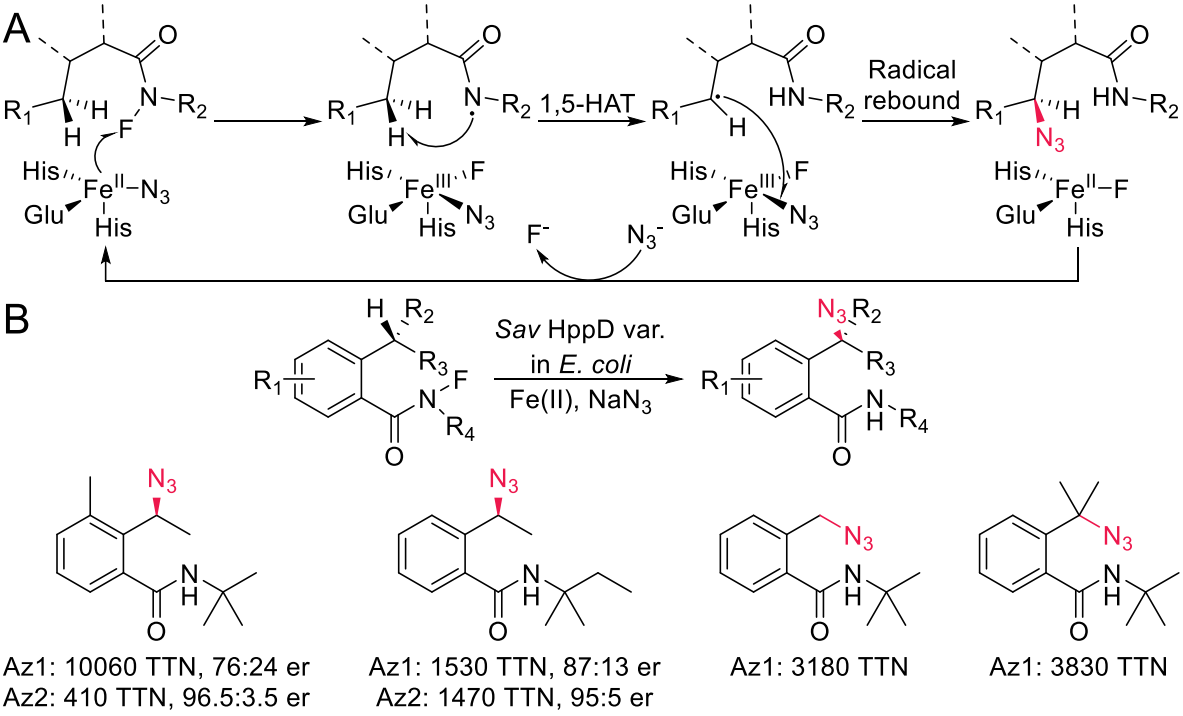


A) Heme-dependent $P411_{BM3}$ variants with complementary site selectivity for intramolecular amination. $P411_{BM3}(1)$ = $P411_{BM3}$-T268A-F87A; $P411_{BM3}(2)$ = $P411_{BM3}$-CIS-T438S-I263F.[32] B) Primary amine synthesis by a $P411_{BM3}$ engineered variant with high site selectivity.[34] C) Engineered non-heme Fe(II)-dependent *Ps*EFE catalyzes intramolecular cyclization of sulfonyl azide.[37]

Finally, a recent study engineered the non-heme Fe(II)-dependent pyruvate dioxygenase from *Streptomyces avermitilis*, *Sav* HppD, to carry out non-native azidation of *N*-fluoroamides at benzylic C–H bonds via a radical relay mechanism (Scheme 3.5A).[39] This process is proposed to occur via the formation of an amidyl radical and a Fe(III)(F)(N_3) intermediate. A 1,5-hydrogen atom transfer step follows, resulting in a new carbon centered radical and a subsequent radical rebound of azide to yield the final azidated product. Notably, the C–H abstraction step in this mechanism is proposed to be carried out by the substrate itself, and not the Fe center. The proposed intermediates are analogous to previously reported fluorination reactions with Fe(II) salts.[40] This study yielded two engineered variants, Az1 and Az2, that were able to successfully

azidate primary, secondary, and tertiary benzylic positions with TTN of up to 10060 and e.r. as high as 96.5:3.5 when using *E. coli* cells harboring these biocatalysts (Scheme 3.5B).

Scheme 3.5: Benzylic azidation via non-native radical relay by non-heme Fe(II)-dependent pyruvate dioxygenase[39]



A) Proposed radical relay mechanism for non-native azidation. B) Representative reactions of two *Sav* HppD evolved variants.

The methods described above represent a wide range of strategies currently used to approach the highly valued but challenging C–N bond forming reactions. Most of them, however, require the use of pre-functionalized substrates or pre-activated nitrogen sources, such as hypervalent iodine reagents and nitrene precursors. There is a high need for techniques that allow for the direct introduction of ammonia equivalents from simple and readily available sources. The examples of non-native azidation carried out by FeDHs highlighted in chapter 2 represent a promising alternative, requiring only the use of sodium azide as the nitrogen source and oxygen as stoichiometric oxidant.[41–44] Furthermore, the product of C–H azidation can be used to access primary amines, amides, lactams, and *N*-heterocycles, as well as grant access to

azide-alkyne cycloaddition chemistry (i.e., "click chemistry").[45–47] Efforts to investigate the use of engineered FeDO SadX for non-native azidation were summarized and published.[48]

Authorship

The synthesis of compounds **2a**, **3a**, **4a**, **5a**, **6a**, and **7a** was carried out by Dr. Dibyendu Mondal. Dr. Natalie H. Chan established the conditions under which the kinetic characterization of evolved enzymes in section 3.2.5 was performed. Dr. Chan also contributed to this study with the synthesis of compound **1a**. The expression and purification of enzymes needed for the isolation of products shown in Scheme 3.6 was carried out with assistance from Qian Du.

3.2 Results and discussion

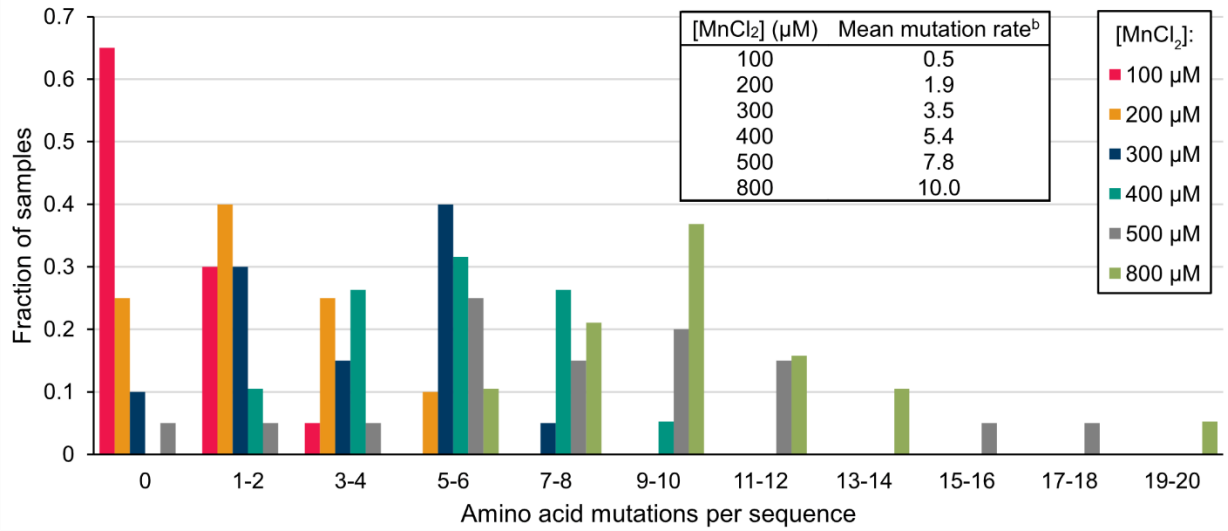
3.2.1 Conditions for directed evolution of SadX

The activity of the MBP-fusion SadX for the chlorination of *N*-succinyl-L-leucine (**1a**) in the presence of NaCl was previously established using both clarified cell lysate and purified enzyme (Figure 2.7). The product of native hydroxyl rebound was still the main product of these reactions. The first goal of this study was to explore the extent of improvement in chemoselectivity towards non-native chlorination that could be achieved via directed evolution. The mutagenesis required for the diversification step needed to be established, and the screening conditions needed to be validated.

Given the fact that the reported crystal structure for SadA is missing a region that is believed to be in contact with the active site and substrate binding pocket, a non-targeted mutagenesis approach was chosen.[49] To investigate the effect of different mutagenic conditions, several error-prone DNA polymerase chain reactions (epPCR) were set up using SadA as template and different concentrations of $MnCl_2$.[50] Twenty samples from each epPCR condition were sequenced and the number of non-silent mutation (i.e., amino acid mutations) introduced in

each sequence was quantified (Figure 3.3). The average mutation rate for reactions with 300 µM $MnCl_2$ was 3.5 amino acids per sequence, which was selected as a figure that would allow for enough diversity to be generated, while maintaining the overall mutagenic burden on the resulting enzyme under levels that would be deleterious.

Figure 3.3: Mutation rate under different epPCR conditions[a]

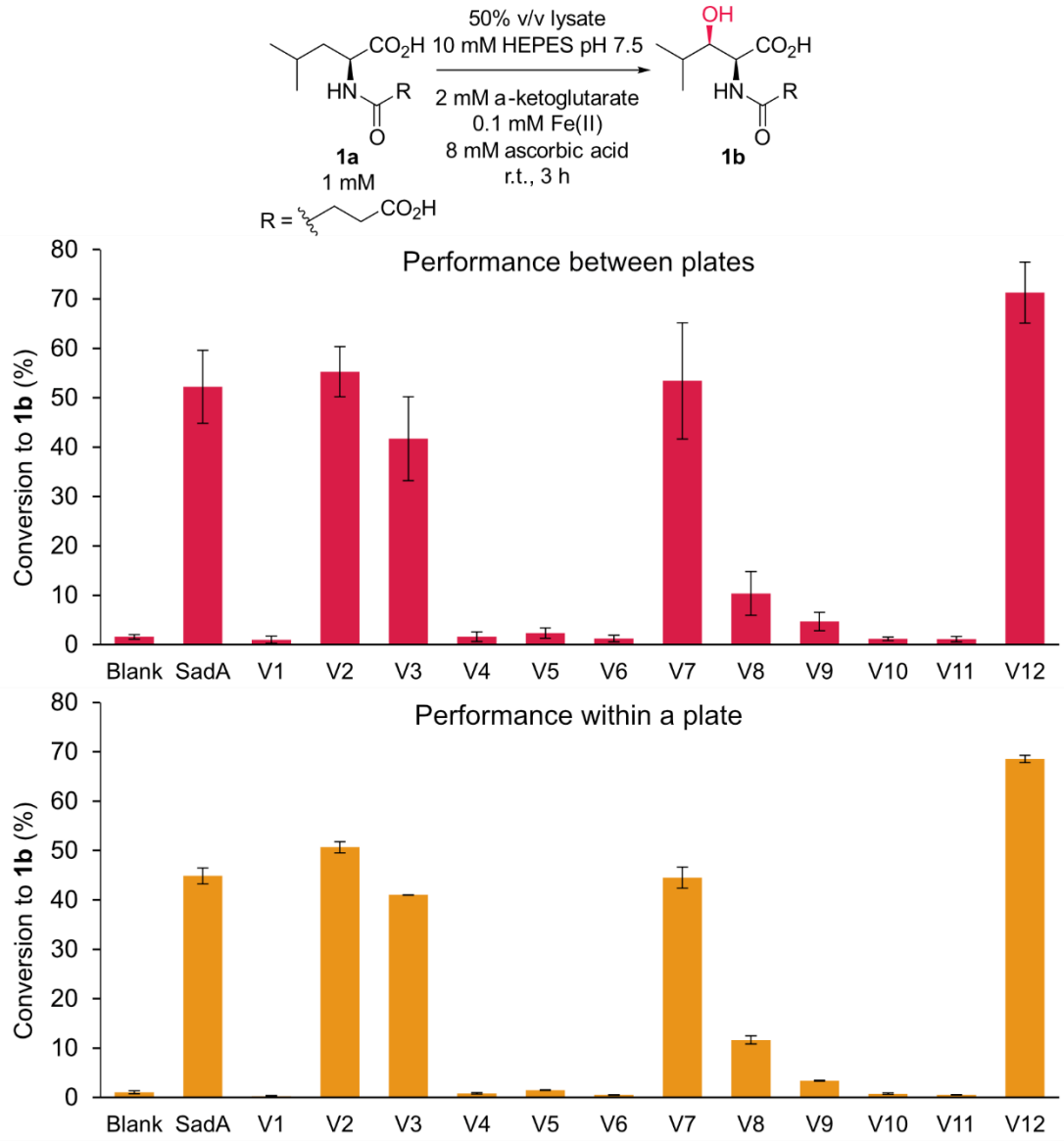


[a] Distribution of amino acid mutations per sequence in samples obtained via epPCR in the presence of different concentrations of $MnCl_2$. Twenty samples were submitted for each epPCR condition. [b] Mean mutation rate expressed as the average amino acid mutation per sequence.

Additionally, the reproducibility of the screening assay was evaluated. Twelve random variants from the SadA epPCR library using 300 µM $MnCl_2$ were used to inoculate cultures in 96-deep well plates. These twelve variants, along with SadA, were expressed in triplicate in three separate plates, with the goal to evaluate plate-to-plate variability, and the resulting clarified *E. coli* lysates were used to carry out hydroxylation bioconversions (Figure 3.4, top). This random set of variants resulted in a diverse range of conversion values, but most importantly, the variability observed for each sample was considered within acceptable levels for the screening of libraries under these conditions. Additionally, one of the analyzed plates contained triplicate cultures within itself, and the analysis of these samples suggested a high reproducibility of the

analytical method when cultures are grown in the same plate. These results not only validate the screening method, but also highlight the importance of having proper control samples in each plate, to minimize the effect of variability between different plates.

Figure 3.4: Evaluation of variability in lysate-based screening assay[a]

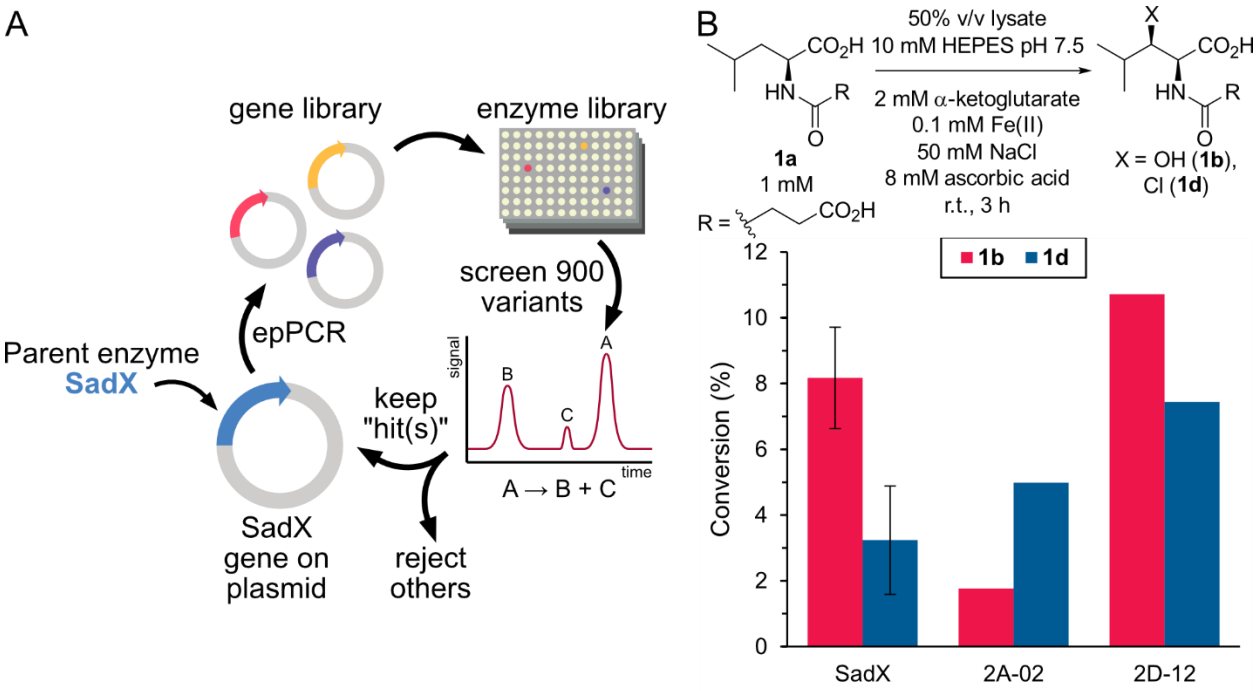


[a] Conversion of **1a** to **1b** using clarified *E. coli* lysate from twelve randomly selected SadA variants obtained via epPCR (V1-V12). Negative control "Blank" contained lysate from a culture of *E. coli* harboring an empty pET28a plasmid. Conversion to products calculated with LC-MS extracted ion chromatogram peak areas of product and substrate. Each data point corresponds to the area of three replicate lysates obtained from cultures grown in independent 96-well plates (top) or three replicate lysates from cultures grown in the same 96-well plate (bottom). Error bars represent the corresponding standard deviations.

3.2.2 Engineering SadX for improved chlorination

With mutagenesis and screening conditions defined, one round of directed evolution on SadX was carried out using epPCR with an average of 3-4 mutations per sequence (Figure 3.5A). A total of 900 variants were screened under chlorination conditions using substrate **1a**, and two potential candidates were identified: variant 2A-02, which showed overall lower conversion under screening conditions but higher chemoselectivity towards the chlorinated product, and variant 2D-12, which showed higher conversion values but no improvement in chemoselectivity (Figure 3.5B). These two variants were selected for further validation and characterization.

Figure 3.5: Overview of a single round of evolving SadX for non-native chlorination

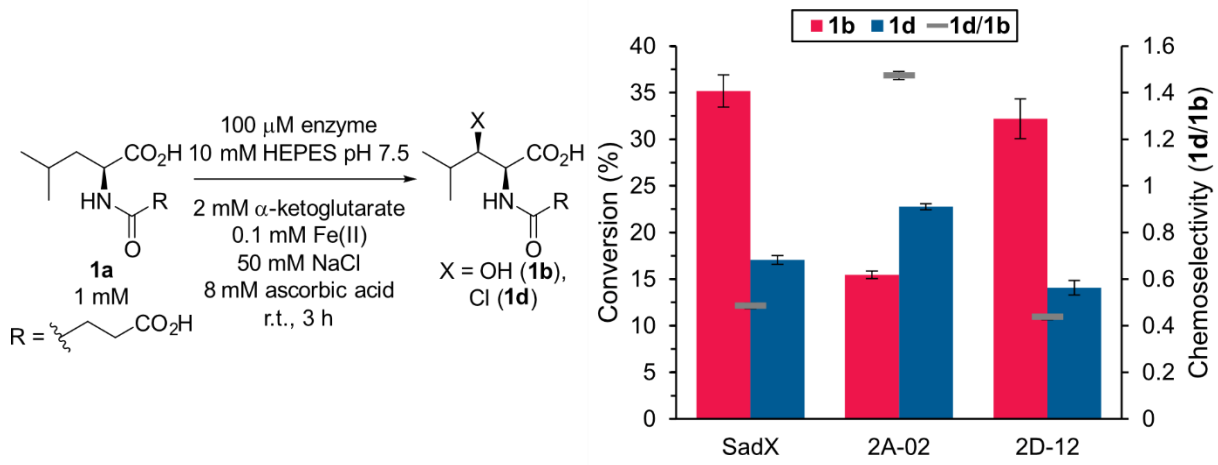


A) General directed evolution strategy for improving non-native chlorinase activity in SadX. A single round of evolution was carried out. B) Screening results for the conversion of **1a** to **1b** and **1d** for two potential "hits". Conversion to products calculated with LC-MS extracted ion chromatogram peak areas of products and substrate. Data for parent SadX corresponds to the area of four replicate lysates obtained from the same plate as 2A-02 and 2D-12, other data points are a single measurements. Error bars represent the corresponding standard deviations.

Sanger sequencing determined the introduction of two mutations in 2A-02 (I71V/R172H) and three mutations in 2D-12 (R48C/I99V/E208G). The following step was the validation of the

potential “hits” identified during library screening. This is done by expressing and purifying any selected variants and running bioconversions under fully defined conditions (Figure 3.6). This step allows comparison of the conversion and selectivity of enzymes without interfering factors, such as different expression levels or undefined reaction media present in lysate reactions. Analyzing both the conversion to the halogenated product **1d** and the chemoselectivity towards chlorination (expressed as the ratio of **1d** over **1b**) revealed that variant 2D-12 did not improve either of these features with respect to parent SadX. The performance of 2A-02, however, was substantially better. Even though the conversion to **1d** improved only by ~7%, the chemoselectivity of the reaction improved by 3-fold, inverting the native selectivity of SadX, and making **1d** the major product.

Figure 3.6: Validation of candidates from SadX evolution for chlorinase activity[a]



[a] Conversion of **1a** to products **1b** and **1d** by SadX variants identified during screening of an epPCR library. Conversion values were calculated with extracted ion chromatogram peak areas of products and the substrate from LC-MS experiments. The chemoselectivity of these variants is expressed as the ratio of conversion to product **1d** over **1b**. Each data point corresponds to the average of three replicates and error bars represent standard deviations.

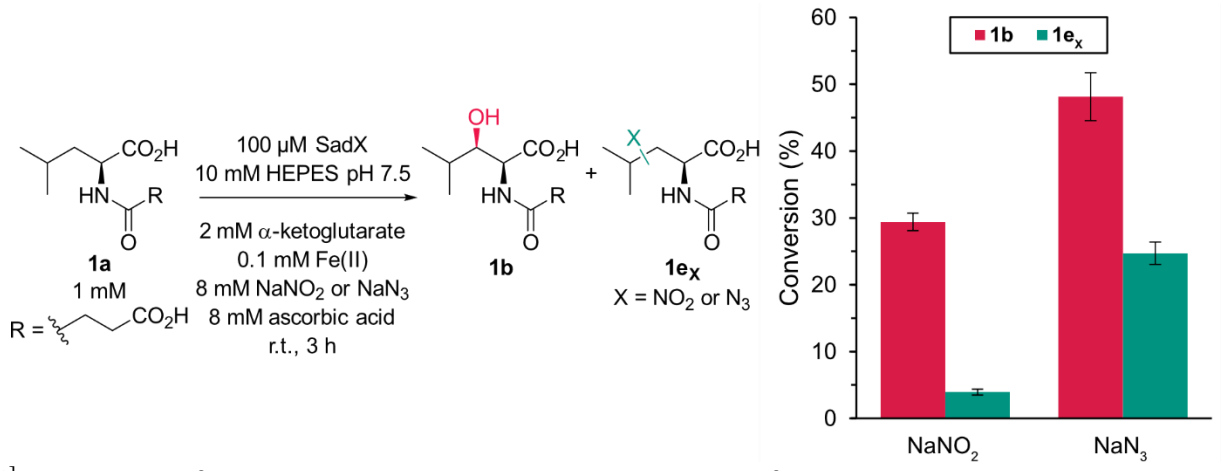
Rationalizing the activity changes imparted by 2A-02 by analysis of the crystal structure of SadA proved difficult. Residue Arg_{172} is located on the outer surface of the protein, distal to the active site and binding pocket. Residue Ile_{71} is located in a region that is missing from the

reported structure, near the substrate binding pocket.[49] The change in size from mutation I71V could result in a change in substrate positioning with respect to the Fe(III)(OH)(Cl) intermediate, favoring chloride rebound over hydroxyl rebound. Beyond this speculation, the results shown above suggested that engineering of SadX towards a non-native activity could be achieved using directed evolution.

3.2.3 Site selectivity in the directed evolution of SadX

Simultaneously with the engineering efforts described in section 3.2.2, other non-native functionalization activities were explored using SadX. Inspired by direct nitration and azidation reports using SyrB2, we investigated the activity of SadX in the presence of $NaNO_2$ and NaN_3 (Figure 3.7).[41] Azidation of substrate **1a** using SadX was observed, reproducing the results presented in chapter 2, with the major product being the native hydroxylated **1b**. A new product was detected in the presence of $NaNO_2$, with a mass corresponding to the product of the nitration of **1a**. This novel product was obtained with a conversion of ~4%, and the chemoselectivity of this reaction was notably lower towards the non-native rebound when compared to the azidation case.

Figure 3.7: Non-native nitration and azidation activities of SadX[a]

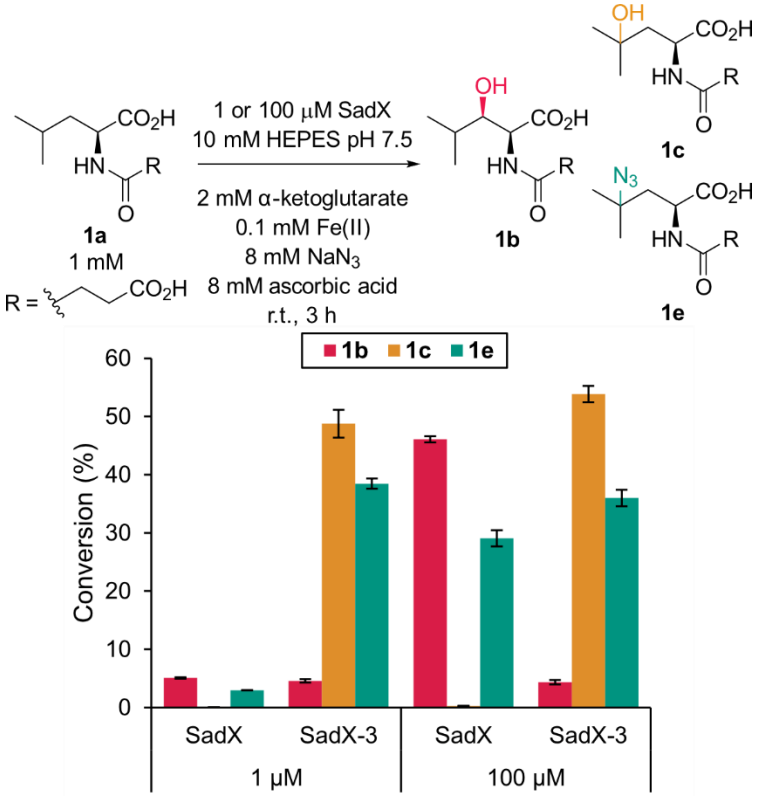


[a] Conversion of **1a** to products **1b** and **1ex** in the presence of $NaNO_2$ and NaN_3. Conversion was calculated with extracted ion chromatogram peak areas of from LC-MS experiments. Each data point corresponds to the average of three replicates and error bars represent standard deviations.

The next step was to evolve SadX for non-native azidation of **1a**. A first attempt was carried out by making targeted libraries at positions 79 and 261. Mutations at these positions were reported to improve the activity of wild-type SadA towards the hydroxylation of *N*-succinyl-L-3,4-dimethoxyphenylalanine by ~6-fold.[51] Two libraries that individually targeted these positions in SadX with a degenerate NNK codon were prepared and screened for improved azidation of **1a**. This approach did not yield any improved variants, suggesting that these mutations had no beneficial effects towards the azidation of **1a** when introduced into SadX.

SadX was subjected to epPCR with 3-4 mutations per sequence and the resulting library was screened focusing on improved conversion of **1a** to the azidated product **1e**. The result of this process was the double mutant termed SadX-3, which contained mutations V39I and F152L and provided improved conversion of **1a** to **1e** under screening conditions. When purified and investigated under standard conditions (100 μM enzyme), SadX-3 exhibited only a modest improvement in the conversion to **1e** (Figure 3.8). The use of a 100-fold lower enzyme loading revealed a similar conversion to **1e** for SadX-3, while SadX had significantly lower levels of azidation. Under these conditions, SadX-3 exhibited a ~11-fold improvement in total turnovers when compared to SadX. Product **1e** would be later characterized as γ-azidated *N*-succinyl-L-leucine, suggesting that the hydrogen atom abstraction step was happening at the γ-position of the substrate instead of the native β-position.

Figure 3.8: Azidation with evolved variant SadX-3 with different enzyme concentrations[a]



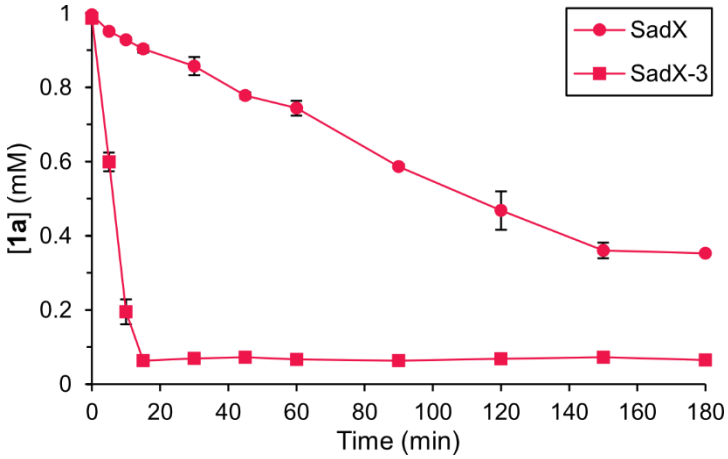
[a] Conversion of **1a** to hydroxylated products **1b** and **1c**, and azidated product **1e** by SadX and SadX-3 at two different concentrations (1 µM and 100 µM). Conversion values were calculated with extracted ion chromatogram peak areas of products and the substrate from LC-MS experiments. Each data point corresponds to the average of three replicates and error bars represent standard deviations.

The chemoselectivity of SadX-3 was very similar to SadX, with the native hydroxyl rebound being favored. The main product, however, was the result functionalization of the γ-position of the amino acid fragment of substrate **1a**, generating this novel hydroxylated product **1c**. Around 95% of the product pool was found to be γ-functionalized compounds, suggesting that productive C–H abstraction was happening almost exclusively at the γ-position. Residue Val_{39} is located distal to the active site in the crystal structure of SadA, so the effect of mutation V39I is difficult to rationalize from this structural information. Residue Phe_{152} is part of a missing fragment in the reported crystal structure of SadA, similar to the residue Ile_{71} mentioned in section 3.2.2. This missing fragment is proximal to the substrate binding pocket and active

site, which could suggest a direct re-shaping of this space and a change in the orientation in which the substrate is positioned, leading to an almost exclusive productive C–H abstraction step at the γ-position.

To further characterize this new variant that showed a notable improvement in total turnovers, reactions with SadX-3 were analyzed in function of time. The consumption of substrate **1a** was monitored via LC-MS in reactions using 1 mM substrate and 10 µM SadX or SadX-3 (Figure 3.9). These results showed a significant difference in the behavior of these two enzymes, with SadX-3 having a higher rate of substrate consumption. Reactions with SadX stopped after ~2.5 hours of reaction time with a final total conversion of ~65%. In the case of SadX-3, reactions progressed to >90% total conversion in the first 15 minutes. This remarkable improvement allowed SadX-3 to be employed in larger scale bioconversions to isolate and characterize product **1e**, as reported in chapter 2.

Figure 3.9: Substrate consumption in bioconversions using SadX and SadX-3[a]



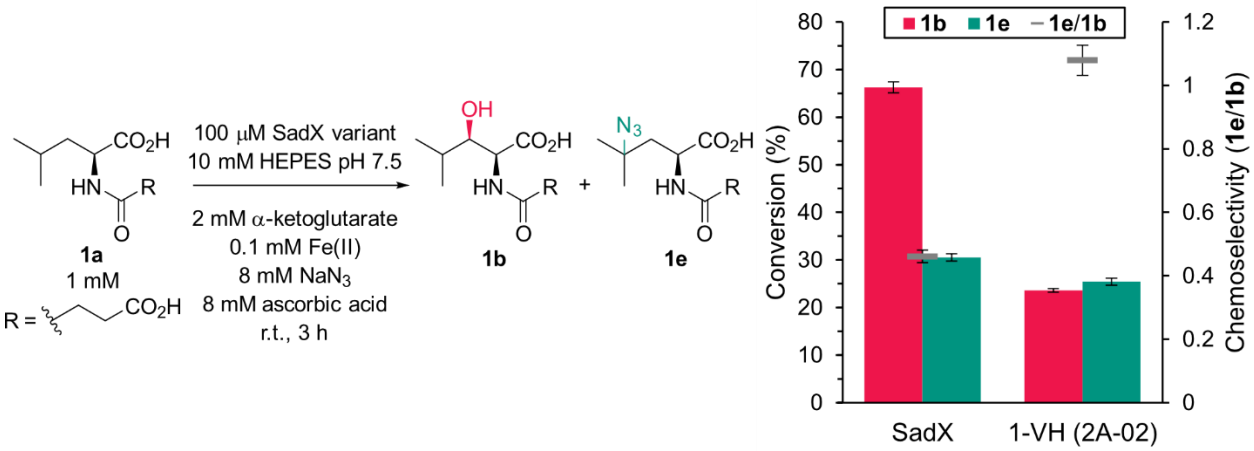
[a] Remaining concentration of substrate **1a** in bioconversions with SadX (circles) and SadX-3 (squares). Different small-scale bioconversions were stopped at different times to analyze reaction progress. Reaction conditions were 10 µM enzyme, 1 mM **1a**, 2 mM α-ketoglutarate, 0.1 mM $Fe(NH_4)_2(SO_4)_2$, 8 mM NaN_3, and 8 mM ascorbic acid in 10 mM HEPES pH 7.5. Each data point corresponds to the average of three replicates and error bars represent standard deviations. Several error bars are not visible due to sufficiently small deviations.

Even though the improvement in azidation obtained with SadX-3 was modest, further engineering could have taken place to improve the chemoselectivity of the enzyme for the azidation of **1a**. However, a different evolution effort that was being carried out simultaneously was continued instead, since it showed more promising results. This directed evolution lineage will be discussed in the following sections.

3.2.4 Directed evolution of SadX for site-selective azidation

Variant 2A-02 was the result of a round of directed evolution aimed at improving chlorination activity of SadX. A significant improvement in chemoselectivity was obtained with this new variant, with the non-native chlorinated compound **1d** being the major product (Figure 3.6). This variant was tested under azidation conditions to evaluate its potential to be a starting point for directed evolution (Figure 3.10). Only β-hydroxylated **1b** and γ-azidated **1e** were detected as products. Conversion to **1e** showed no improvement with 2A-02 when compared to SadX, but chemoselectivity was greatly increased, with a >2-fold rise in the **1e** to **1b** ratio when compared to the parent enzyme.

Figure 3.10: Azidase activity of evolved chlorinase 1-VH (2A-02)[a]

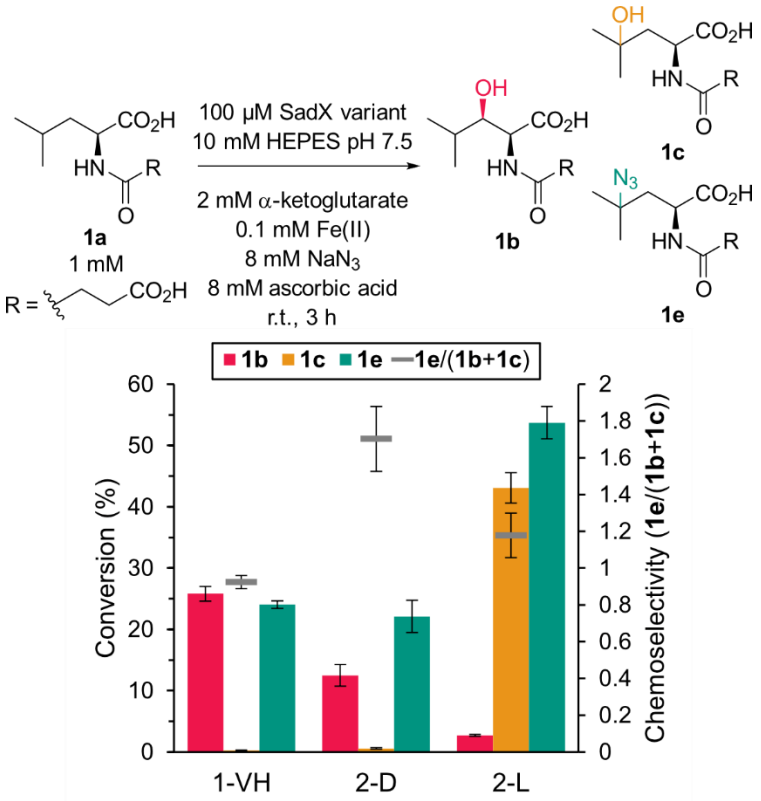


[a] Conversion of **1a** to products **1b** and **1e** using SadX and 2A-02 (renamed 1-VH). Conversion values were calculated with extracted ion chromatogram peak areas of products and the substrate from LC-MS experiments. The chemoselectivity of these variants is expressed as the ratio of conversion to product **1e** over **1b**. Each data point corresponds to the average of three replicates and error bars represent standard deviations.

These levels of improvement in chemoselectivity were not observed previously for azidation reactions using engineered SadX variants. For that reason, 2A-02 was selected as the first variant in a new directed evolution lineage and it was used as a parent for a subsequent run. This variant was renamed "1-VH" to follow the naming convention that will be applied to all future variants from this lineage, using the number of the generation in which the variant was found followed by the single letter code for the amino acids that were introduced in that generation (e.g., Val_{79} and His_{172} in the case of 1-VH).

The next round of directed evolution consisted of another epPCR library in which 900 variants were screened. Two enzymes were identified with attractive improvements with respect to the parent 1-VH. Variant 2-D (1-VH N65D) showed a ~1.8-fold improvement in chemoselectivity with respect to its parent enzyme, continuing the trend of decreasing the amount of β-hydroxylated product **1b** while maintaining constant levels of azidation (Figure 3.11). Variant 2-L (1-VH F152L) resulted in a lower increase in chemoselectivity, but a >2-fold improvement in conversion to product **1e**. This effect was accompanied by production of γ-hydroxylated product **1c**, resembling the behavior of SadX-3 (Figure 3.8). Unsurprisingly, the mutation introduced in 2-L was also present in SadX-3, strengthening the hypothesis that position Phe_{152} is important in defining the site selectivity of SadX. The combination of mutations F152L and N65D was unsuccessful in generating a variant with increased chemoselectivity and conversion to **1e**.

Figure 3.11: Azidation reactions by second generation SadX variants[a]

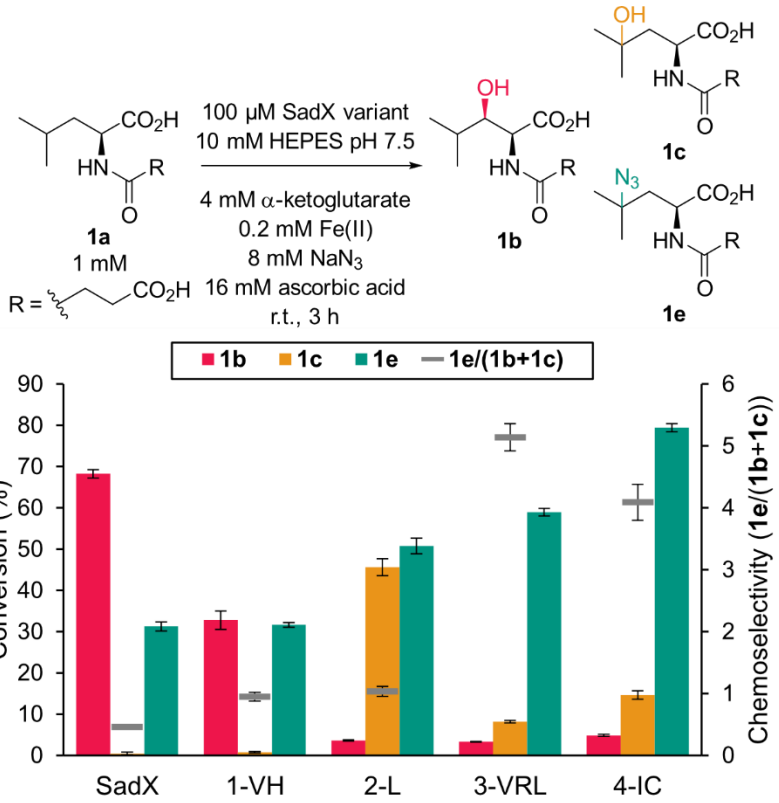


[a] Conversion of **1a** to products **1b**, **1c**, and **1e** by SadX variants 1-VH, 2-D, and 2-L. Conversion values were calculated with extracted ion chromatogram peak areas of products and the substrate from LC-MS experiments. The chemoselectivity of these variants is expressed as the ratio of conversion to product **1e** over the sum of conversions to products **1b** and **1c**. Each data point corresponds to the average of three replicates and error bars represent standard deviations.

The two new variants 2-D and 2-L represented two potential pathways to engineer a highly active and selective azidase. While 2-D provided a significant improvement in chemoselectivity, further engineering to increase conversion was needed. On the other hand, 2-L showed improved conversion to **1e**, so engineering its chemoselectivity was needed. It was unclear which pathway would successfully arrive at the desired enzyme faster, so both of these variants were used to make independent epPCR libraries, each containing 900 variants. While variants with improved conversion to **1e** were found in the library that used 2-D as parent, none of them performed better than 2-L. Additionally, no improvements in chemoselectivity were found in the 2-D library.

The library based on 2-L, however, produced a new triple mutant with significantly increased performance. This new variant, 3-VRL (2-L I38V/Q233R/F261L), not only resulted in higher conversion to **1e**, but also showed a ~5-fold improvement in chemoselectivity when compared to 2-L, and a ~11-fold improvement when compared to the starting point SadX (Figure 3.12). This leap in selectivity was produced by a decrease in production of **1c**, suggesting that selectivity towards non-native azide rebound was being favored. While residues Ile_{38} and Gln_{233} are located distal to the active site, residue Phe_{261} is located in the active site, facing the Fe(II) center. This fact suggests a direct impact on substrate positioning and rebound selectivity.

Figure 3.12: Conversion and product distribution of evolved variant 4-IC and its lineage[a]

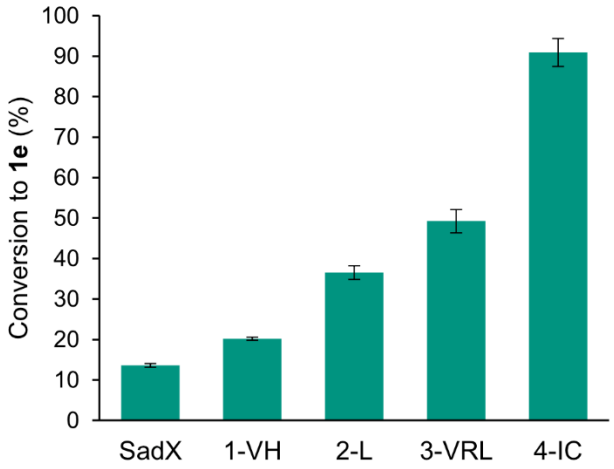


[a] Conversion of **1a** to products **1b**, **1c**, and **1e** by variants in the lineage of 4-IC. Conversion values were calculated with extracted ion chromatogram peak areas of products and the substrate from LC-MS experiments. The chemoselectivity of these variants is expressed as the ratio of conversion to product **1e** over the sum of conversions to products **1b** and **1c**. Each data point corresponds to the average of three replicates and error bars represent standard deviations.

While the selectivity of 3-VRL towards non-native azidation was the highest in this series of enzymes, the overall conversion of **1a** decreased. Optimization of the reaction conditions led to higher activity when increasing the concentrations of α-ketoglutarate, Fe(II), and ascorbic acid. Under these conditions 2-L was able to convert >99% of substrate to products, while 3-VRL had an overall conversion of ~70% (Figure 3.12). In order to further improve the conversion of **1a**, an additional round of directed evolution was carried out with an epPCR library of 900 variants and conditions that led to an average of 1 to 2 mutations per sequence. Screening of this library resulted in several potential improved enzymes being identified. Recombination of new mutations with reversion of 3-VRL mutations resulted in variant 4-IC (3-VRL V38I/R48C). This new variant showed ~99% conversion of substrate **1a** and a chemoselectivity towards azidation that was lower than that of 3-VRL, but still ~9-fold higher with respect to SadX.

The analysis of reactions catalyzed by SadX variants to obtain conversion values for all product relied on the comparison of LC-MS chromatogram areas of substrate **1a** and all products. This analysis assumes homogenous ionization efficiency among all analytes, which may not necessarily be the case. Azidated product **1e** was isolated from scaled-up bioconversions using SadX-3 and used to construct a standard calibration curve to accurately quantify the conversion to **1e** (Figure 3.13). Under this assay, variant 4-IC achieved 91% conversion of **1a** to **1e**. The discrepancies between conversion values from these two analyses highlight differences in ionization efficiencies between compounds **1a**-**1e**. Isolation of product **1c** was not possible, making it impossible to have standard curves for all compounds. Without a standard curve, comparison of absolute conversions is not valid, but this approach still allows for comparison of relative conversions across the different variants.

Figure 3.13: Conversion to azidated product by SadX variants quantified with a standard curve[a]



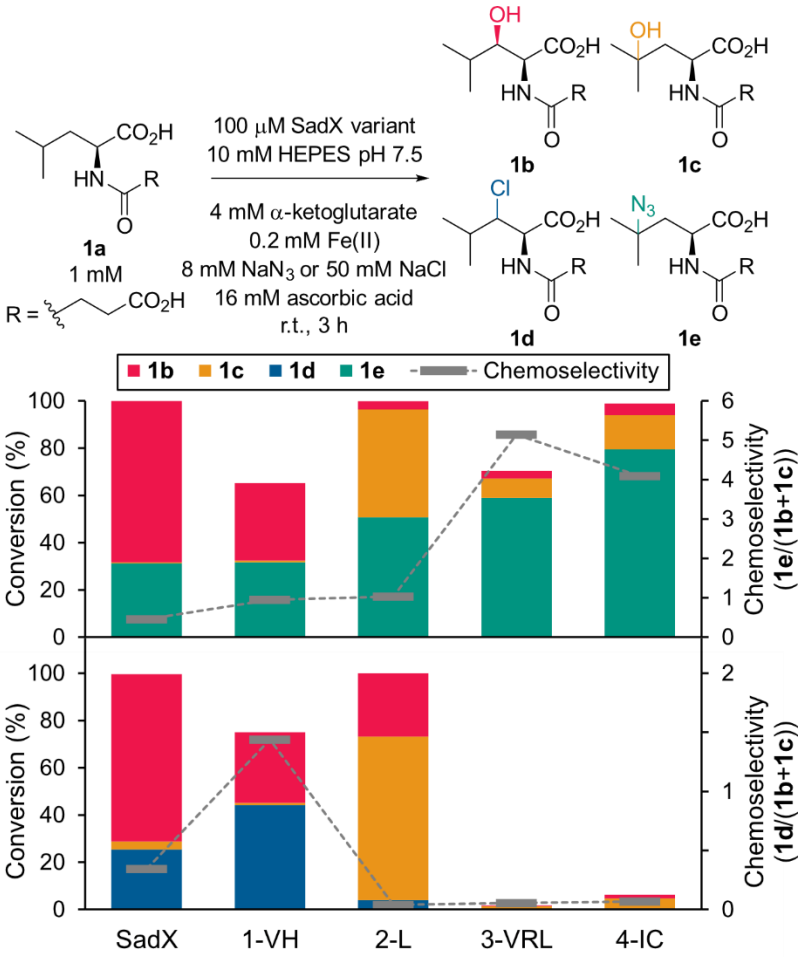
[a] Conversion values were calculated with the extracted ion chromatogram peak area of product **1e** relative to internal standard *N*-acetyl-L-valine from LC-MS experiments and a standard curve constructed with isolated **1e**. Reaction conditions were 100 µM enzyme, 1 mM **1a**, 4 mM α-ketoglutarate, 0.2 mM $Fe(NH_4)_2(SO_4)_2$, 8 mM NaN_3, and 16 mM ascorbic acid in 10 mM HEPES pH 7.5. Each data point corresponds to the average of three replicates and error bars represent standard deviations.

3.2.5 Analysis and characterization of evolved SadX variants

The product distributions for evolved SadX variants under azidation and chlorination conditions reveal a subtle relationship between overall conversion, chemoselectivity, and site selectivity (Figure 3.14). The fact that SadX is able to form β-chlorinated product **1d** and γ-azidated product **1e** suggests that the Fe(IV)-oxo intermediate is able to abstract both β and γ C–H bonds in substrate **1a**. The fact that no β-azidated product is detected implies that rebound of the azide ligand may be controlled by the relative orientation of substrate radical, as previously proposed for the chemoselectivity of hydroxyl and chloride rebound in FeDH SyrB2.[52] Variant 1-VH resulted in a decrease of β-hydroxylated product **1b**, with higher chemoselectivities for both chlorination and azidation. The next variant 2-L showed a significant increase in azidation and γ-hydroxylation, as well as an almost complete absence of chlorination and β-hydroxylation. This effect seems to align with an active site that has been evolved to favor

productive C–H abstraction at the γ-position, since only a small percentage of β-functionalized products were detected. Consequently, the following two variants in the lineage, which further improved γ-azidation, showed no improved activity under chlorination conditions. The lack of product **1c** under those conditions suggests that variants 3-VRL and 4-IC must have been evolved to position the substrate in a way that hydroxyl rebound to the γ-position is disfavored.

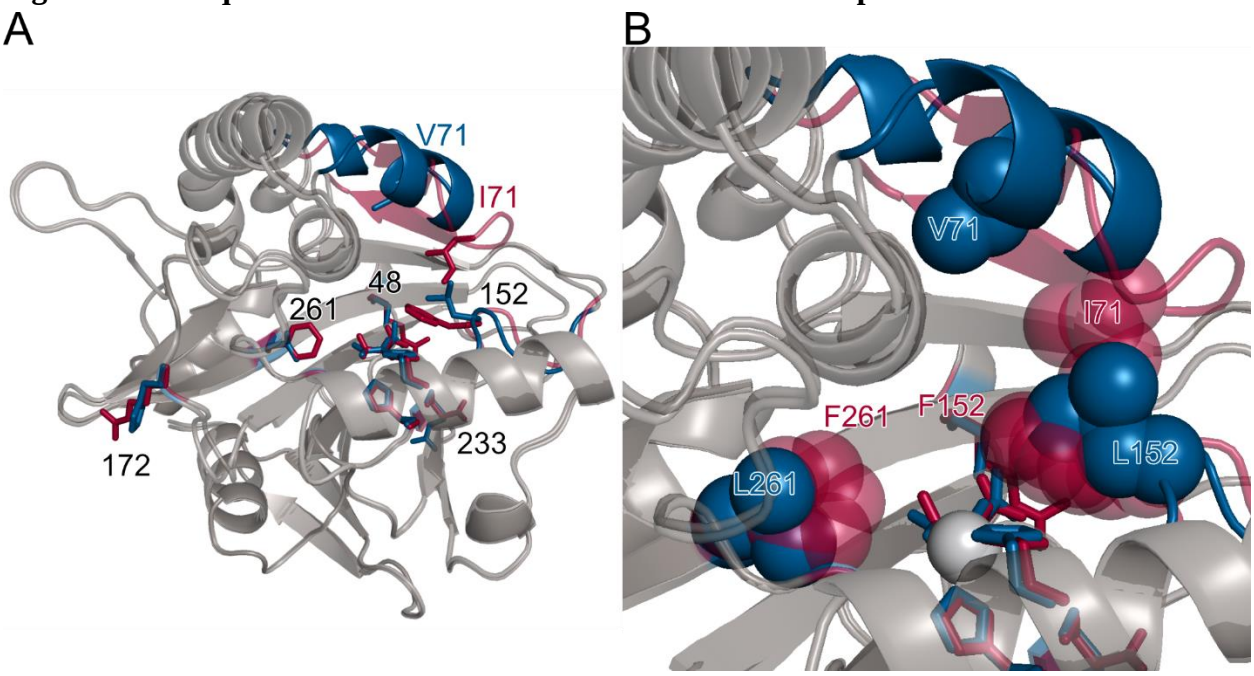
Figure 3.14: Conversion and product distribution of evolved SadX variants under azidation and chlorination conditions[a]



[a] Conversion of **1a** to products **1b**-**1e** by variants in the lineage of 4-IC. Reactions were set up under azidation (top) and chlorination (bottom) conditions. Conversion values were calculated with extracted ion chromatogram peak areas of products and the substrate from LC-MS experiments. The chemoselectivity of these variants is expressed as the ratio of conversion to product **1d** or **1e** over the sum of conversions to products **1b** and **1c**. Each data point corresponds to the average of three replicates.

It is worth noting that the two instances in which chemoselectivity towards azidated product **1e** increased significantly, with variants 1-VH and 3-VRL, a decrease in the total conversion of substrate was observed. This apparent trade-off between chemoselectivity and overall activity of the SadX variants once again seemed to support the proposed mechanisms that control chemoselectivity in native FeDHs, but further analysis of these evolved SadX variants was needed. Modeled structures of SadX and 4-IC were obtained using AlphaFold (Figure 3.15).[53] Other models were constructed using SWISS-MODEL[54] and Rosetta[55], but both had poorly packed structures for the Ser_{59}-Ser_{75} and Glu_{147}-Ala_{153} missing segments. The AlphaFold SadX model predicted a well-packed structure for these segments, with the missing Ser_{59}-Ser_{75} loop being modeled as a β-hairpin-like structure above the active site, and residue Phe_{152} facing the active site (Figure 3.15A). The mutations that had the most impact on the active site, according to these models, were I71V, F152L, and F261L (Figure 3.15B). Mutation I71V is predicted to change the structure of the Ser_{59}-Ser_{75} loop, and both F152L and F261L are predicted to increase the volume of the active site. Noteworthy, the introduction of F152L in 2-L led to a substantial change in site-selectivity, with all major products being γ-functionalized, and introduction of F261L in 3-VRL led to an improvement in the specificity towards azidation. These predictions seem to support the idea of a change in substrate positioning in the active site leading to a change in site and chemoselectivity. It is worth mentioning that, as described in section 3.2.2, Phe_{261} in SadX was subjected to site saturation mutagenesis but no improved variants were successfully identified. This suggests that the effects of mutation F261L are specific to improving azidation over γ-hydroxylation, while hydroxylation in SadX is exclusively at the β-position.

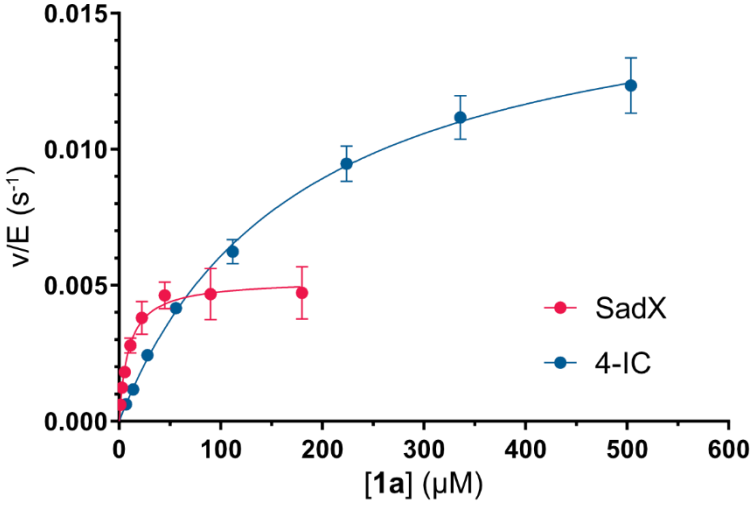
Figure 3.15: Acquired mutations overlaid in SadX and 4-IC AlphaFold models



A) Location of accumulated mutations in the respective structures of SadX (red) and 4-IC (blue). Only segments where both models showed notable differences when aligned, both histidine residues from the facial triad, and α-ketoglutarate are colored. B) Accumulated active site mutations and their effect on the active site structure.

Analysis of steady state kinetics for azidation of substrate **1a by** SadX and 4-IC was carried out (Figure 3.15). 4-IC showed a ~3-fold and ~20-fold increase in k_{cat} and K_M, respectively, as well as a ~6-fold decrease in k_{cat}/K_M. This decrease in catalytic efficiency seems to be caused by the higher K_M value, suggesting that the binding of **1a** needed to achieve the improvements in chemoselectivity is less favorable than in the parent enzyme. This would be another observation that supports previously reported insights on the mechanism of halogenase SyrB2, where selective chlorination of the native substrate is achieved by reduction of the overall activity of the enzyme.[52]

Figure 3.16: Steady state kinetic characterization of azidation by SadX and 4-IC[a]



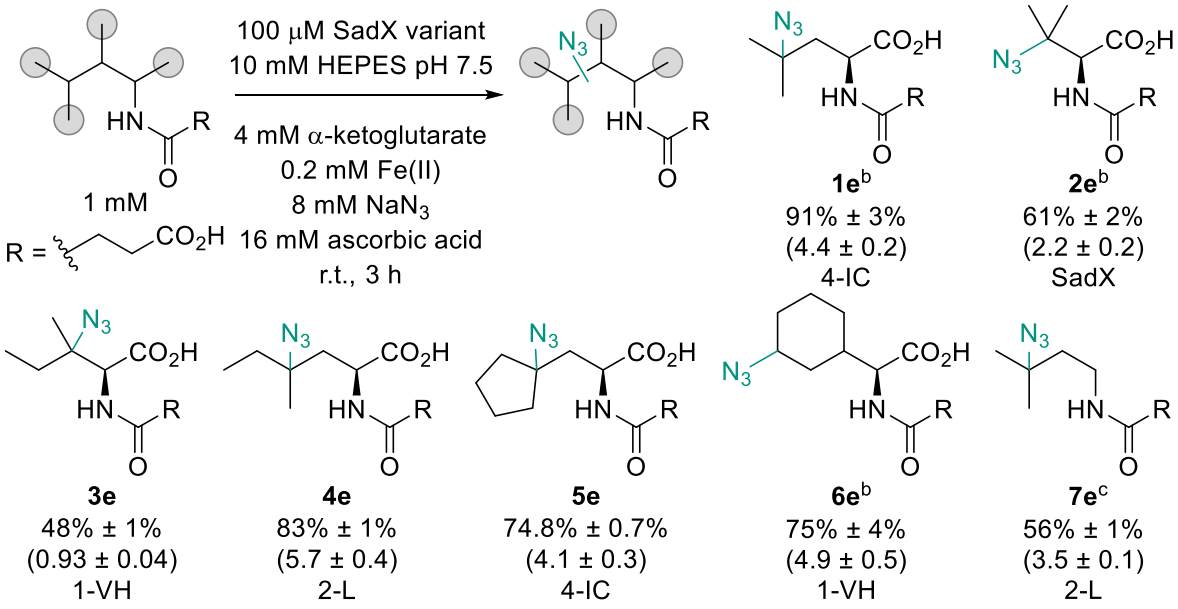
Enzyme	[**1a**] (μM)	k_{cat} (10^{-3} s^{-1})	K_M (mM)	k_{cat}/K_M (M^{-1} s^{-1})
SadX	1.4 – 180	5.2	0.009	577.8
4-IC	7 – 504	16.8	0.177	94.9

[a] Each data point corresponds to the average of triplicate samples and error bars represent the corresponding standard deviations.

Exploration of the substrate scope of the evolved SadX variants revealed good-to-high conversion of *N*-succinylated β- and γ-branched amino acids, as well as *N*-succinylated amines, to the corresponding azidated products (Scheme 3.6). Conversion values with a subset of representative substrates ranged between 48% to 91%, and chemoselectivity ratios of azidation over hydroxylation were between 0.93 and 5.7. Most of the characterized products were functionalized at a tertiary C–H bond, but secondary C–H azidation was observed in product **6e**, indicating that this preference can be overridden. Substrates that lacked the *N*-succinyl moiety were not functionalized, suggesting the importance of interactions between this functional group and the binding pocket of the evolved azidases. Production of **7e**, however, shows the potential to functionalize non-amino acid-derived substrates. No single enzyme in the 4-IC lineage showed high levels of conversion and selectivity for all substrates examined, but instead different enzymes were able to better functionalize specific substrates. This fact highlights the importance

of precise substrate positioning to enable selective functionalization, as previously proposed for native FeDHs.[52,56]

Scheme 3.6: Representative substrate scope of evolved azidases[a]

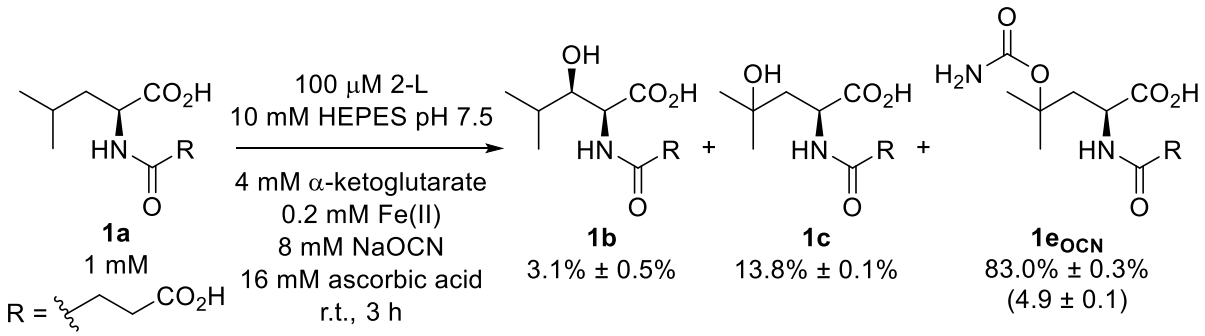


[a] Percent conversion azidated products **1e** – **7e** by evolved SadX variants. Unless otherwise specified, conversion values were calculated with extracted ion chromatogram peak areas of products and substrates from LC-MS experiments. Chemoselectivity towards azidation (in parenthesis) expressed as the ratio of azidated products over hydroxylated products. Each value is the average and standard deviation of three replicate experiments. [b] Conversion values were quantitated by LC-MS relative to *N*-acetyl-L-valine internal standard and a standard curve of the corresponding isolated product. [c] Standard reaction conditions with 8 mM α-ketoglutarate.

Additionally, the activity of these evolved SadX variants was evaluated in the presence of NCO^-. As previously discussed in chapter 2, variant SadXL was able to functionalize substrate **1a** and produce primary carbamate **1e**$_{OCN}$ with a 31% conversion and 0.45 chemoselectivity ratio over hydroxylated products (Figure 2.11). The best enzyme in the 4-IC lineage to catalyze this functionalization reaction selectively was 2-L (Scheme 3.7). Under these conditions, product **1e**$_{OCN}$ was the major product, with a chemoselectivity ratio of almost 5:1 over hydroxylated products and a total conversion of 83% towards the carbamate product. Both SadXL and 2-L share mutation F152L, but the additional mutations that the latter enzyme accumulated (i.e., I71V and R172H) further improve the chemoselectivity of this transformation. Thus, evolving

SadX for selective azidation allowed to establish 2-L as an unprecedented selective enzyme for primary carbamates synthesis from direct C–H functionalization.

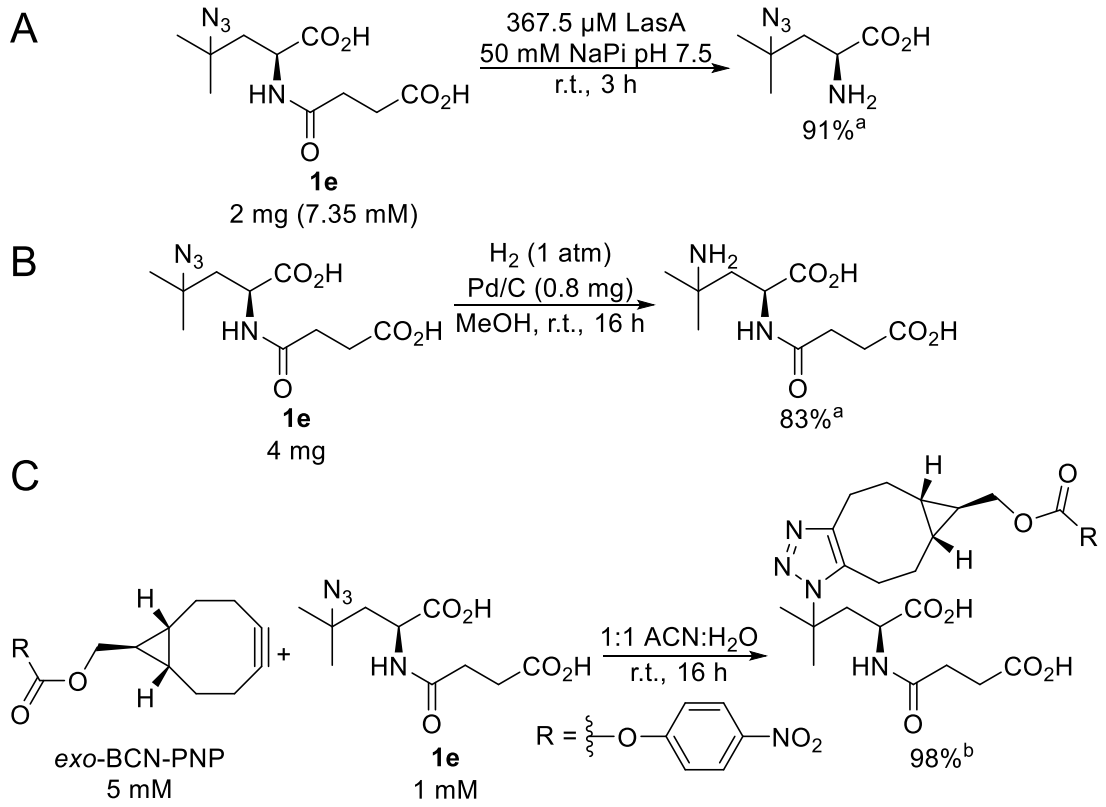
Scheme 3.7: Bioconversion of 1a by variant 2-L in the presence of NaOCN[a]



[a] Percent conversion values were calculated with extracted ion chromatogram peak areas of products and substrate from LC-MS experiments. Chemoselectivity towards product **1e$_{OCN}$** (in parenthesis) expressed as the ratio of **1e$_{OCN}$** over hydroxylated products **1b** and **1c**. Each value is the average and standard deviation of three replicate experiments.

Finally, to demonstrate the synthetic potential of these evolved azidases, azidated product **1e** was isolated and used as substrate for reactions. The succinyl functional group was removed from **1e** to produce 4-azido-L-leucine (Scheme 3.8A). This was achieved using *N*-succinyl-L-amino acid desuccinylase LasA, an enzyme from the same gene cluster as SadA. LasA was previously used in conjunction with SadA to successfully produce β-hydroxyleucine from *N*-succinyl-L-leucine.[57] LasA was able to accept **1e** and the free azidated amino acid was isolated with 91% yield. Additionally, **1e** was reduced in the presence of H_2 and Pd/C, producing the corresponding primary amine with an isolated yield of 83% (Scheme 3.8B). Lastly, a strain-promoted azide-alkyne cycloaddition (SPAAC) was carried out using azide **1e** and *exo*-bicyclo[6.1.0]nonyne *p*-nitrophenyl carbonate (*exo*-BCN-PNP) to produce the corresponding triazole linkage with a 98% conversion (Scheme 3.8C). These reactions aim to illustrate the diversity of applications of the azidated products generated by SadX variants evolved for non-native azidation.

Scheme 3.8: Synthetic applications of evolved azidases



A) Biocatalytic desuccinylation of azidated compound **1e**. B) Reduction of azide in **1e** to primary amine. C) Strain-promoted azide-alkyne cycloaddition between azide **1e** and alkyne *exo*-BCN-PNP. [a] Isolated yield. [b] Conversion based on LC-MS extracted ion chromatograms.

3.3 Conclusions

The study presented in this chapter aimed to address the challenging task of achieving highly selective C–H functionalization via rebound of non-native anionic ligands using engineered FeDOs. Azidation was the focus of this study, but chlorination and carbamate insertion were also examined. The conditions for the directed evolution of engineered FeDH SadX were established and an improved chlorinase, 1-VH, was obtained via random mutagenesis. Further evolution of 1-VH towards azidation led to a series of enzymes capable of selectively functionalizing *N*-succinylated-L-amino acids and *N*–succinylated amines. Analysis of the acquired mutations suggests a change in substrate binding that allows for the observed change in the chemoselectivity of the azidation of substrate **1a**. Kinetic studies revealed that the

increase in chemoselectivity was accompanied by a decrease in the catalytic efficiency of the most evolved variant 4-IC with respect to SadX, mainly caused by a less favorable substrate binding. Finally, one of the obtained azidated products was used in an enzymatic desuccinylation reaction to obtain the free azidated amino acid, in a reduction to produce the corresponding primary amine, and in a strain-promoted azide-alkyne cycloaddition. These results represent a successful use of directed evolution to achieve novel levels of chemoselectivity towards non-native rebound in FeDOs. Additionally, this study establishes a platform for further engineering of SadX variants in order to expand the substrate scope and the range of non-native reactivities catalyzed by these enzymes.

3.4 Experimental

3.4.1 Materials

SadX (MBP-SadA D157G fusion) was expressed from a previously constructed pET28(SadX) plasmid, as described in chapter 2.[58] LasA was obtained as a synthetic gene from Twist Bioscience and expressed according to published literature.[57] BL21-Gold (DE3) *E. coli* cells (# 230132) were purchased from Agilent Technologies. Lysozyme from chicken egg white (#L6876) and deoxyribonuclease I from bovine pancreas (DNase, #DN25) were purchased from MilliporeSigma. Restriction enzymes XmaI (#R0180) and DpnI (#R0176), Taq DNA polymerase (#M0273), T4 DNA ligase (#M0202), and dNTPs (#N0447) were purchased from New England Biolabs. PrimeSTAR Max DNA polymerase mix (#R045A) was purchased from Takara Bio USA. Oligonucleotides were purchased from MilliporeSigma. Miniprep kits (QIAprep Spin Miniprep Kit, #27104) and agarose gel extraction kits (QIAquick Gel Extraction Kit, #28704) were purchased from QIAGEN and used following the manufacturer's protocols. DNA clean and concentrator kits (#D4003) were purchased from Zymo Research and used following the

manufacturer's instructions. Terrific broth (TB, #T15000) and Luria broth (LB, #L24040) were purchased from Research Products International and prepared following the manufacturer's protocols.

Kanamycin monosulfate (#J61272) was purchased from Alfa Aesar and prepared into a 50 mg/mL aqueous x1000 stock solution. BCN-PNP (exo) (#CP-6048) was purchased from Conju-Probe. Isopropyl-β-D-thiogalactopyranoside (IPTG, #00194) was purchased from Chem-Impex International. Ni-NTA resin (HisPur™ Ni-NTA Resin, #88223) was purchased from Thermo Scientific. Dowex 50WX8 cation exchange resin (#335331000) was purchased from Acros Organics.

3.4.2 General Procedures

Colony picking, library expression, and screening reactions were set up using an automation system consisting of a Thermo Scientific Spinnaker robotic arm and a Multidrop Combi liquid dispenser, a NorgrenSystems CP-7200 colony picker, and a Hamilton Nimbus liquid handler. The automation system was controlled by Thermo Scientific Momentum software. DNA and protein concentrations were measured with a Tecan Infinite 200 PRO plate reader. DNA amplification by PCR was performed on a Applied Biosystems ProFlex PCR System thermocycler. DNA transformation was carried out by electroporation with a Bio-Rad MicroPulser. Cell lysis by sonication was performed using a QSonica S-4000 sonicator with a 0.5" horn.

^{1}H NMR, ^{13}C NMR, and ^{1}H-^{13}C HSQC spectra were obtained using a Bruker 500 MHz Avance Neo NMR spectrometer.

LC-MS analysis was carried out using an Agilent 1290 Infinity II system connected to a ZORBAX Eclipse Plus C18 column (2.1 mm x 50 mm, 1.8 µm particle size), an Agilent Jet

Stream (AJS-ES) ion source, and a 6135 single quadrupole mass selective detector. Purification of bioconversions for product characterization was performed using an Agilent 1100 HPLC with a Supelco Discovery C18 semi-preparative column (10 mm x 25 cm, 5 µm particle size).

Data fitting for steady-state kinetic studies was performed using the software GraphPad Prism 7 (Dotmatics). HPLC and LC-MS data analysis was carried out with MassHunter (Agilent) and OpenLab Chemstation (Agilent) softwares.

3.4.3 Chromatographic methods

All chromatographic methods described below used 0.1 % formic acid in water (A) and 0.1 % formic acid in acetonitrile (B) as mobile phase components. Table 3.1 shows the *m/z* values used to obtain EICs from each experiment.

LC-MS 1

Solvent gradient: 10 %B (0 – 0.4 min), 20 to 26.25 %B (0.4 – 1.75 min), 90 %B (1.75 – 2.5 min), 10 %B (2.5 – 3.25 min). Flow rate: 0.4 mL/min. ESI negative mode.

LC-MS 2

Solvent gradient: 5 %B (0 – 1.25 min), 5 to 14 %B (1.25 – 2 min), 14 %B (2 – 4.5 min), 14 to 90 %B (4.5 – 6.8 min), 90 %B (6.8 – 7 min), 5 %B (7 – 7.25 min). Flow rate: 0.4 mL/min. ESI negative mode.

LC-MS 3

Solvent gradient: 2 %B (0 – 1.25 min), 2 to 10 %B (1.25 – 2 min), 10 %B (2 – 4.5 min), 10 to 90 %B (4.5 – 6.8 min), 90 %B (6.8 – 7 min), 2 %B (7 – 7.25 min). Flow rate: 0.4 mL/min. ESI positive mode.

LC-MS 4

Solvent gradient: 5 %B (0 – 1.25 min), 5 to 15 %B (1.25 – 4 min), 15 to 25 %B (4 – 6 min), 25 to 90 %B (6 – 8 min), 90 %B (8 – 10 min), 10 %B (10 – 12 min). Flow rate: 0.4 mL/min. ESI negative mode. SIM for *m/z* = 158 and 271.

LC-MS 5

Solvent gradient: 10 %B (0 – 0.5 min), 10 to 14.3 %B (0.5 – 2 min), 14.3 to 30.4 %B (2 – 3.5 min), 45 %B (3.5 – 5.7 min), 95 %B (5.7 – 7.5 min), 10 %B (7.5 – 8.0 min). ESI positive & negative mode. UV-Vis monitoring at 270 nm.

HPLC 1

Solvent gradient: 2 %B (0 – 3 min), 25 %B (3 – 16 min), 98 %B (16 – 21 min), 2 %B (21 – 25.5 min). Flow rate: 3 mL/min. UV-Vis monitoring at 230 nm.

HPLC 2

Solvent gradient: 2 %B (0 – 3 min), 2 to 65 %B (3 – 16 min), 98 %B (16 – 21 min), 2 %B (21 – 25.5 min). Flow rate: 3 mL/min. UV-Vis monitoring at 230 nm.

HPLC 3

Solvent gradient: 2 %B (0 – 3 min), 2 to 35 %B (3 – 7 min), 35 %B (7 – 15 min), 35 to 98 %B (15 – 16 min), 98 %B (16 – 21 min), 2 %B (21 – 25.5 min). Flow rate: 3 mL/min. UV-Vis monitoring at 230 nm.

HPLC 4

Solvent gradient: 2 %B (0 – 3 min), 2 to 30 %B (3 – 7 min), 30 %B (7 – 15 min), 35 to 98 %B (15 – 16 min), 98 %B (16 – 21 min), 2 %B (21 – 25.5 min). Flow rate: 3 mL/min. UV-Vis monitoring at 230 nm.

HPLC 5

Solvent gradient: 2 %B (0 – 3 min), 2 to 40 %B (3 – 7 min), 40 %B (7 – 15 min), 40 to 95 %B (15 – 16 min), 95 %B (16 – 22 min), 2 %B (22 – 26.5 min). Flow rate: 3 mL/min. UV-Vis monitoring at 230 nm.

HPLC 6

Solvent gradient: 1 %B (0 – 7 min), 1 to 6 %B (7 – 20.7 min), 6 to 80 %B (20.7 – 25 min), 80 %B (25 – 28 min), 80 to 1 %B (28 – 30 min), 1 %B (30 – 33 min). Flow rate: 3 mL/min. UV-Vis monitoring at 230 nm.

Table 3.1: Dominant ions for relevant species monitored via LC-MS

Species	***m/z***	**Species**	***m/z***
1a, 3a	230	**4$_{OH}$**	260
1b, 1c, 3$_{OH}$	246	**4e**	285
1d	184	**5a, 6a**	256
1e, 3e	271	**5$_{OH}$, 6$_{OH}$**	272
1e$_{OCN}$	289	**5e, 6e**	297
2a	216	**7a**	186
2$_{OH}$	230	**7$_{OH}$**	202
2e	257	**7e**	227
4a	244		

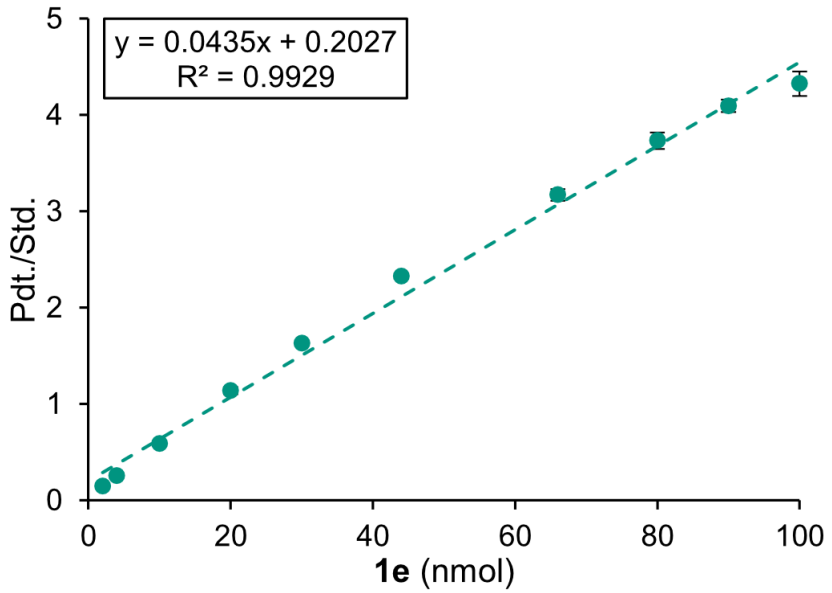
Extracted ion chromatograms (EICs) were extracted for **1a**, **1b**, **1c**, and **1d**/**1e** for each sample according to Table 3.1. The area of the corresponding peaks was integrated and used to represent the amount of each analyte. Percent conversion to product **1i** (**i** = **b**, **c**, **d**/**e**) was calculated as the peak area of product **1i** divided by the sum of the peak area of **1a** and the peak areas of all products (1). Chemoselectivity was defined as the ratio of halogenated/azidated product to hydroxylated products (2). An analogous analysis was done for bioconversions with substrates **2a**–**7a**.

$$\%Conversion\ to\ \mathbf{1}\boldsymbol{i} = 100\% \times \frac{Area_{\mathbf{1}\boldsymbol{i}}}{\left(Area_{\mathbf{1}\boldsymbol{a}} + Area_{\mathbf{1}\boldsymbol{b}} + Area_{\mathbf{1}\boldsymbol{c}} + Area_{\mathbf{1}\boldsymbol{d}/\mathbf{1}\boldsymbol{e}}\right)} \quad (1)$$

$$Chemoselectivity = \frac{Area_{\mathbf{1}\boldsymbol{d}/\mathbf{1}\boldsymbol{e}}}{(Area_{\mathbf{1}\boldsymbol{b}} + Area_{\mathbf{1}\boldsymbol{c}})} \quad (2)$$

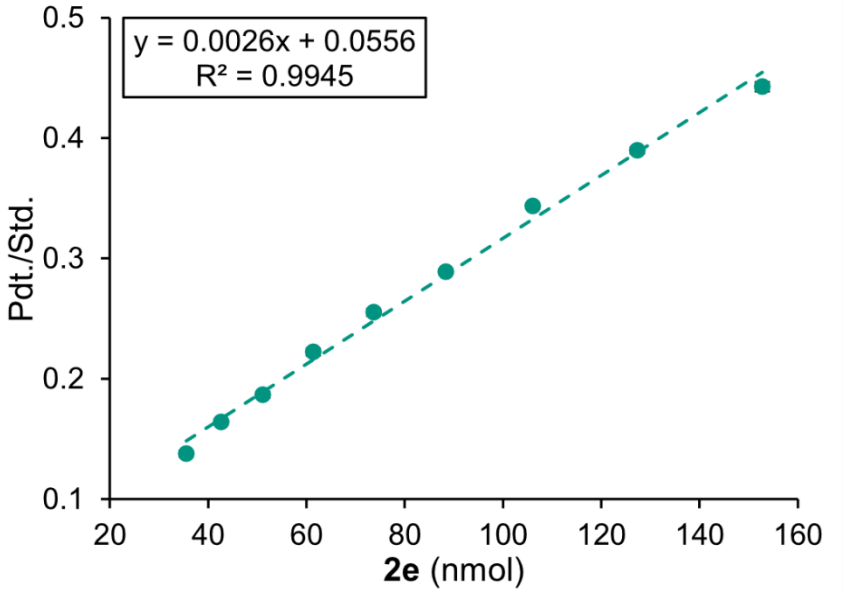
Conversion values to azidated product **1e** shown in Figure 3.13 were determined by comparison of the ratio of areas of **1e** and internal standard (*N*-acetyl-L-valine) with a calibration curve using purified **1e** (Figure 3.17). An analogous process was carried out for products **2e** and **6e** shown in Scheme 3.3 (Figures 3.18 and 3.19). Due to variations in ionization efficiencies between different LC-MS experiments, calibration curves were prepared and analyzed immediately before samples were quantified.

Figure 3.17: Representative calibration curve for LC-MS quantitation of product 1e[a]



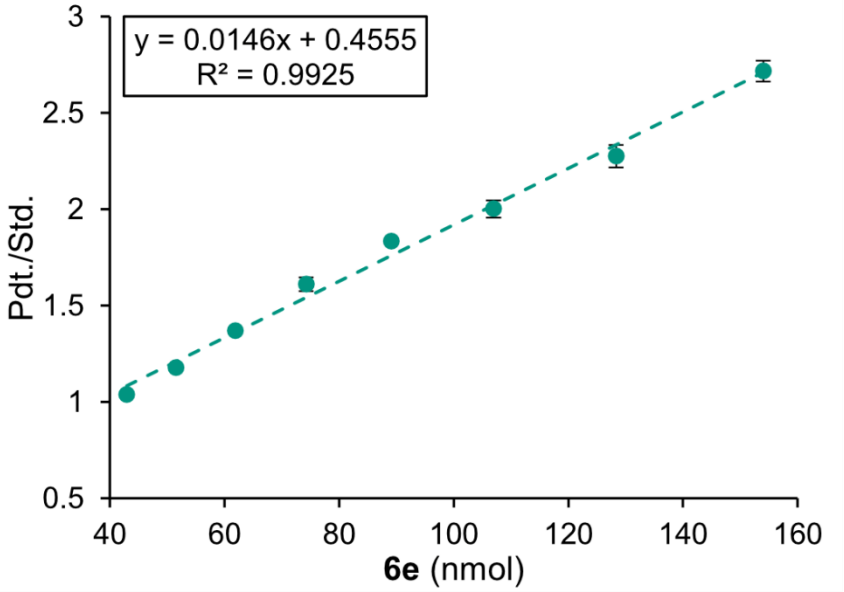
[a] Linear regression of the relationship between the ratio of LC-MS EIC areas of product **1e** over *N*-acetyl-L-valine internal standard ("Pdt./Std.") and the amount of product **1e**. Each data point corresponds to the average of three replicates and the error bars represent standard deviations. Several error bars are not visible due to sufficiently small deviations.

Figure 3.18: Representative calibration curve for LC-MS quantitation of product 2e[a]



[a] Linear regression of the relationship between the ratio of LC-MS EIC areas of product **2e** over *N*-acetyl-L-valine internal standard ("Pdt./Std.") and the amount of product **2e**. Each data point corresponds to the average of three replicates and the error bars represent standard deviations. Several error bars are not visible due to sufficiently small deviations.

Figure 3.19: Representative calibration curve for LC-MS quantitation of product 6e[a]



[a] Linear regression of the relationship between the ratio of LC-MS EIC areas of product **6e** over *N*-acetyl-L-valine internal standard ("Pdt./Std.") and the amount of product **6e**. Each data point corresponds to the average of three replicates and the error bars represent standard deviations. Several error bars are not visible due to sufficiently small deviations.

3.4.4 Molecular cloning

Standard molecular cloning procedures were used throughout this study.[59] Details for the cloning of the relevant libraries are detailed below.

Error-prone PCR libraries

Libraries of SadX variants for directed evolution were made using error-prone PCR (epPCR) and prolonged overlap extension PCR (POE-PCR).[60] A pET28a plasmid containing the corresponding parent variant was used as template for the cloning involved. The oligonucleotide primers used are shown in Table 3.2.

Table 3.2: Oligonucleotides used to generate epPCR libraries of SadX and SadA variants

#	Name	Sequence
1	SadXInsertFP	5' – GTA TTT CCA GGG AGC AGC CGG ATC CAT GCA GCA TAC CTA TCC GG – 3'
2	SadXInsertRP	5' – TCG AGT GCG GCC GCA AGC TTT CAA TCA AAC ATA CGC CAA CC – 3'
3	SadXVectorFP	5' – GTT GGC GTA TGT TTG ATT GAA AGC TTG CGG CCG CAC TCG A – 3'
4	SadXVectorRP	5' – TAG GTA TGC TGC ATG GAT CCG GCT GCT CCC TGG AAA TAC AGG – 3'
5	SadAInsertFP	5' – AGC AAA TGG GTC GCG GAT CCA TGC AGC ATA CCT ATC CGG C– 3'
6	SadAInsertRP	5' – TCG AGT GCG GCC GCA AGC TTT CAA TCA AAC ATA CGC CAA CC – 3'
7	SadAVectorFP	5' – GTT GGC GTA TGT TTG ATT GAA AGC TTG CGG CCG CAC TCG A – 3'
8	SadAVectorRP	5' – GCC GGA TAG GTA TGC TGC ATG GAT CCG CGA CCC ATT TGC T– 3'

Generation of the mutated SadX variant insert gene was carried out in a 50 µL reactions using the following conditions: 1 ng/µL template DNA, 0.2 mM dNTPs, 0.2 µM oligonucleotides **1** and **2**, 0.025 U/µL Taq DNA Polymerase, 1x Taq Standard Buffer, and 300 µM $MnCl_2$. The PCR program for this amplification was 95 °C for 30 seconds, 30 cycles of 95 °C for 30 seconds and 68 °C for 90 seconds, and 68 °C for 10 minutes. The product was digested with 10 units of DpnI at 37 °C for 1 hour and purified via 1% agarose gel electrophoresis.

Linearized pET28a(MBP) vector was amplified in 50 µL reactions under the following conditions: 1 ng/µL template DNA, 0.3 µM oligonucleotides **3** and **4**, and 1x PrimeSTAR Max DNA polymerase mix. DNA amplification was carried out with the following program: 98 °C for 10 seconds; 30 cycles of 98 °C for 10 seconds, 55 °C for 5 seconds, and 72 °C for 45 seconds; and 72 °C for 1 minute. The PCR product was digested with 10 units of DpnI at 37 °C for 1 hour and purified via 1% agarose gel electrophoresis.

Inserts and linearized vector were assembled into multimers via POE-PCR in 50 µL reactions using 250 ng of vector DNA, 33.3 ng of insert DNA (1:1 insert:vector molar ratio), and 1x PrimeSTAR Max DNA polymerase mix. Reaction conditions were the following: 98 °C for 2 minutes; 20 cycles of 98 °C for 10 seconds, 59 °C for 15 seconds, and 72 °C for 4 minutes; 15 cycles of 98 °C for 10 minutes, 59 °C for 15 seconds, and 72 °C for 8 minutes; and 72 °C for 10 minutes.

The resulting 50 µL multimer solution was digested with XmaI to obtain insert-vector monomers by setting up a 300 µL reaction with 0.17 U/µL XmaI (5 µL of 10 U/µL) in 1x CutSmart buffer. The digestion was incubated overnight at 37 °C and purified using a DNA clean and concentrator kit, eluting with 50 µL of molecular biology grade water. The digested monomers were ligated in a 100 µL reaction using 12 U/µL T4 DNA ligase (3 µL of 400 U/µL) in 1x T4 DNA ligase buffer. The ligation reaction was incubated at 22 °C for 3 hours, heat inactivated at 65 °C for 10 minutes, and purified into 10 µL of molecular biology grade water using a DNA clean and concentrator kit.

Libraries of SadA were constructed analogously except for the generation of the mutagenized SadA gene, which was obtained in 50 µL reactions with 1 ng/µL template DNA, 0.2 mM dNTPs, 0.2 µM oligonucleotides **5** and **6**, 0.025 U/µL Taq DNA Polymerase, 1x Taq

Standard Buffer, and 100 - 800 µM $MnCl_2$, as shown in Figure 3.3. Linearized pET28a vector was obtained in in 50 µL reactions with 1 ng/µL template DNA, 0.3 µM oligonucleotides **7** and **8**, and 1x PrimeSTAR Max DNA polymerase mix.

Targeted site saturation mutagenesis libraries

Targeted site saturation mutagenesis (SSM) libraries at positions Gly_{79} and Phe_{261} were obtained via site-directed mutagenesis of SadX following a modified QuickChange™ protocol.[61] The pET28(SadX) plasmid was used as template for amplification using the oligonucleotides shown in Table 3.3, with oligos **9** and **10** being used for Gly_{79}, and **11** and **12** for Phe_{261}. The amplification conditions were the following: 2 ng/µL template DNA, 0.3 µM forward primer, 0.3 µM reverse primer, and 1x PrimeSTAR Max DNA Polymerase, with the final volume being 50 µL. The PCR program for the amplifications was the following: 98 °C for 10 seconds; 30 cycles of 98 °C for 10 seconds, 55 °C for 5 seconds, and 72 °C for 45 seconds; and ending with 72 °C for 1 minute. The PCR products were digested with 10 units of DpnI at 37 °C for 1 hour and purified via 1% agarose gel electrophoresis.

Table 3.3: Oligonucleotides used for construction of SSM libraries of SadX

#	Name	Sequence
9	SadXG79NNKFP	5’ – GTA TTG TTG CAA CCA GCG TTA CCC TGN NKC AGC TGC AGC GTG AAC AGG GTG – 3’
10	SadXG79NNKRP	5’ – GTA ACG CTG GTT GCA ACA ATA CCT GCA TGA TGT GCA TTT TTG GTA AAC GGG CTG – 3’
11	SadXF261NNKFP	5’ – CAG CAC CCG TCG TAC CAT GGG TCT GNN KCT GAT TCA TAC CGA AGA TGG TTG – 3’
12	SadXF261NNKRP	5’ – CCA TGG TAC GAC GGG TGC TGG TAC AAC GTT CAA TGG CAT GCA GAT TTT TG – 3’

3.4.5 Sequences

SadA (nucleotide)

This sequence includes the SadA gene and the 5’ and 3’ flanking region of the vector were primers **5**, **6**, **7**, and **8** anneal (<u>underlined</u>).

AGCAAATGGGTCGCGGATCCATGCAGCATACCTATCCGGCACAGCTGATGCGTTTTG
GCACCGCAGCACGTGCAGAACATATGACCATTGCAGCAGCAATTCATGCACTGGAT
GCAGATGAAGCAGATGCAATTGTTATGGATATTGTTCCGGATGGTGAACGTGATGCA
TGGTGGGATGATGAAGGTTTTAGCAGCAGCCCGTTTACCAAAAATGCACATCATGCA
GGTATTGTTGCAACCAGCGTTACCCTGGGTCAGCTGCAGCGTGAACAGGGTGATAA
ACTGGTTAGCAAAGCAGCAGAATATTTTGGTATTGCCTGCCGTGTTAATGATGGTCT
GCGTACCACCCGTTTTGTTCGTCTGTTTAGTGATGCCCTGGATGCCAAACCGCTGACC
ATTGGTCATGATTATGAAGTTGAATTTCTGCTGGCAACCCGTCGTGTTTATGAACCGT
TTGAAGCACCGTTTAACTTTGCACCGCATTGTGATGATGTTAGCTATGGTCGTGATA
CCGTTAATTGGCCTCTGAAACGTAGCTTTCCGCGTCAGCTGGGTGGTTTTCTGACCAT
TCAGGGTGCAGATAATGATGCCGGTATGGTTATGTGGGATAATCGTCCGGAAAGCC
GTGCAGCGCTGGATGAAATGCATGCAGAATATCGTGAAACCGGTGCAATTGCCGCA
CTGGAACGTGCAGCCAAAATCATGCTGAAACCGCAGCCTGGCCAGCTGACACTGTT
TCAGAGCAAAAATCTGCATGCCATTGAACGTTGTACCAGCACCCGTCGTACCATGGG
TCTGTTTCTGATTCATACCGAAGATGGTTGGCGTATGTTTGATTGAGTTGGCGTATGT
TTGATTGAAAGCTTGCGGCCGCACTCGA

SadA (amino acid)

The sequence of the His_6 tag is *italicized*.

*MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGS*MQHTYPAQLMRFGTAARAEHMTIA
AAIHALDADEADAIVMDIVPDGERDAWWDDEGFSSSPFTKNAHHAGIVATSVTLGQLQ
REQGDKLVSKAAEYFGIACRVNDGLRTTRFVRLFSDALDAKPLTIGHDYEVEFLLATRR
VYEPFEAPFNFAPHCDDVSYGRDTVNWPLKRSFPRQLGGFLTIQGADNDAGMVMWDN
RPESRAALDEMHAEYRETGAIAALERAAKIMLKPQPGQLTLFQSKNLHAIERCTSTRRT
MGLFLIHTEDGWRMFD*

SadX (nucleotide)

This sequence includes the MBP tag (*italicized*), the SadA D157G gene, and the 3' flanking region of the vector were primers **2** and **3** anneal (underlined).

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCA
TATGGCTAGCATGCACCATCACCATCACCATGGAAAAATCGAAGAAGGTAAACTGGTAATC
TGGATTAACGGCGATAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAA
GATACCGGAATTAAAGTCACCGTTGAGCATCCGGATAAACTGGAAGAGAAATTCCCACAG
GTTGCGGCAACTGGCGATGGCCCTGACATTATCTTCTGGGCACACGACCGCTTTGGTGGC
TACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTTCCAGGACAAGCTG
TATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCGCTG
TTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGA

AGAGATCCCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAA
CCTGCAAGAACCGTACTTCACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAA
GTATGAAAACGGCAAGTACGACATTAAAGACGTGGGCGTGGATAACGCTGGCGCGAAAG
CGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACACATGAATGCAGACACCGATTA
CTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTG
GGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTT
CAAGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAG
TCCGAACAAAGAGCTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTG
GAAGCGGTTAATAAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAG
TTGGCGAAAGATCCACGTATTGCCGCCACTATGGAAAACGCCCAGAAAGGTGAAATCATG
CCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCC
GCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTAATTCGAGCTC
*GAACAACAACAACACTAGTGAAAACCTGTATTTCCAGGGAGCAGCCGGATCC*ATGCAGCA
TACCTATCCGGCACAGCTGATGCGTTTTGGCACCGCAGCACGTGCAGAACATATGAC
CATTGCAGCAGCAATTCATGCACTGGATGCAGATGAAGCAGATGCAATTGTTATGG
ATATTGTTCCGGATGGTGAACGTGATGCATGGTGGGATGATGAAGGTTTTAGCAGCA
GCCCGTTTACCAAAAATGCACATCATGCAGGTATTGTTGCAACCAGCGTTACCCTGG
GTCAGCTGCAGCGTGAACAGGGTGATAAACTGGTTAGCAAAGCAGCAGAATATTTT
GGTATTGCCTGCCGTGTTAATGATGGTCTGCGTACCACCCGTTTTGTTCGTCTGTTTA
GTGATGCCCTGGATGCCAAACCGCTGACCATTGGTCATGATTATGAAGTTGAATTTC
TGCTGGCAACCCGTCGTGTTTATGAACCGTTTGAAGCACCGTTTAACTTTGCACCGC
ATTGTGGCGATGTTAGCTATGGTCGTGATACCGTTAATTGGCCTCTGAAACGTAGCT
TTCCGCGTCAGCTGGGTGGTTTTCTGACCATTCAGGGTGCAGATAATGATGCCGGTA
TGGTTATGTGGGATAATCGTCCGGAAAGCCGTGCAGCGCTGGATGAAATGCATGCA
GAATATCGTGAAACCGGTGCAATTGCCGCACTGGAACGTGCAGCCAAAATCATGCT
GAAACCGCAGCCTGGCCAGCTGACACTGTTTCAGAGCAAAAATCTGCATGCCATTG
AACGTTGTACCAGCACCCGTCGTACCATGGGTCTGTTTCTGATTCATACCGAAGATG
GTTGGCGTATGTTTGATTGA<u>AAGCTTGCGGCCGCACTCGA</u>

SadX (amino acid)

The sequence of MBP is *italicized*. All further amino acid indices refer to the original enzyme SadA numbering.

MGSSHHHHHHSSGLVPRGSHMASMHHHHHHGKIEEGKLVIWINGDKGYNGLAEVGKKFEK
DTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYP
FTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPY
FTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAF
NKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFL
ENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYA
*VRTAVINAASGRQTVDEALKDAQTNSSSNNNNTSENLYFQGAAGS*MQHTYPAQLMRFGTAA
RAEHMTIAAAIHALDADEADAIVMDIVPDGERDAWWDDEGFSSSPFTKNAHHAGIVAT
SVTLGQLQREQGDKLVSKAAEYFGIACRVNDGLRTTRFVRLFSDALDAKPLTIGHDYEV
EFLLATRRVYEPFEAPFNFAPHCGDVSYGRDTVNWPLKRSFPRQLGGFLTIQGADNDAG

MVMWDNRPESRAALDEMHAEYRETGAIAALERAAKIMLKPQPGQLTLFQSKNLHAIER
CTSTRRTMGLFLIHTEDGWRMFD*

SadX-3
SadX V39I (I = ATT) F152L (L = TTA)

1-VH
SadX I71V (V = GTT) R172H (H = CAT)

2-L
1-VH F152L (L = CTT)

2-D
1-VH N65D (D = GAT)

3-VRL
2-L I38V (V = GTT) Q233R (R = CGC) F261L (L = CTT)

4-IC
3-VRL V38I (I = ATT) R48C (C = TGT)

3.4.6 Gene expression and protein purification

Libraries in 96-deep well plates

A 5 µL aliquot of the resulting plasmid or nicked plasmid DNA was used to transform 50 µL of BL21-Gold (DE3) *E. coli* via electroporation. Cells were resuspended with 945 µL of SOC media and shaken at 37 °C and 250 rpm for 45 minutes. Up to 400 µL of the culture were transferred to 86 mm x 128 mm rectangular LB/agar plates with 50 µg/mL kanamycin. The plates were incubated overnight at 37 °C. Single colonies were picked and placed in 1 mL 96-deep well plates containing 300 µL/well of TB media with 50 µg/mL kanamycin to generate primary cultures. Each plate contained 90 library variants, 4 parent samples (picked and inoculated using the same method as the library variants), and 2 blank wells that were not inoculated and served as controls of cross contamination. The resulting plates were incubated at

37 °C and 215 rpm overnight. 50 µL of each culture were used to make glycerol stocks of the library by combining them with 50 µL of 50% glycerol in microtiter 96-well plates.

Expression cultures were made using 50 µL of primary culture to inoculate 2 mL 96-deep well plates containing 1 mL/well of LB media with 50 µg/mL kanamycin. Expression plates were incubated at 37 °C and 215 rpm until OD_{600} reached 0.6 – 0.8, at which point the cultures were brought to a temperature of 18 °C, induced with a final concentration of 1 mM IPTG, and incubated at 18 °C and 215 rpm for 18 hours. Cells were harvested by centrifugation at 3600 rpm for 10 minutes and stored at -80 °C until use.

Large scale protein expression

A starter culture was prepared by inoculating 5 mL of TB media with 50 µg/mL kanamycin from glycerol stocks made during library expression and incubating at 37 °C and 250 rpm overnight. The overnight culture was used to generate an expression culture by inoculating 750 mL of LB media with 50 µg/mL kanamycin. The expression culture was incubated at 37 °C and 250 rpm until OD_{600} reached 0.6 – 0.8, after which the culture was brought to 18 °C, induced with 1 mM IPTG, and incubated at 18 °C and 250 rpm for 18 hours. Cells were harvested by centrifugation at 4700 rpm for 10 minutes and stored at -80 °C until use.

The cell pellet was resuspended in a 50 mL conical tube with 30 mL of 20 mM imidazole in 10 mM HEPES pH 7.5 and sonicated at 40 W with 0.5 min ON/0.5 min OFF cycles for 5 min total cycle time. The resulting cell lysate was clarified by centrifugation at 4 °C and 15000 rpm for 45 min using a high-speed fixed-angle rotor. The supernatant was transferred to a 10 mL polypropylene frit-bottomed gravity flow column containing 5 mL of Ni-NTA resin pre-equilibrated with equilibration buffer (20 mM phosphate, 300 mM NaCl, 10 mM imidazole, pH 7.4). After the lysate was allowed to drain, 50 mL of wash buffer (20 mM phosphate, 300 mM

NaCl, 25 mM imidazole, pH 7.4) were added to the resin. Finally, 15 mL of elution buffer (20 mM phosphate, 300 mM NaCl, 250 mM imidazole, pH 7.4) were used to elute purified protein into a 50 mL conical tube. The protein solution was transferred to a 15 mL Amicon spin filter 30K MWCO and concentrated down to 0.5 – 1.0 mL at 4700 rpm. In order to remove endogenous Fe(II) 15 mL of 50 mM EDTA in 10 mM HEPES pH 7.5 were used to dilute the protein sample and concentrate to 0.5 – 1.0 mL at 4700 rpm. Three additional buffer exchanges with 15 mL of 10 mM HEPES in pH 7.5 were performed. Protein concentration was measured using absorbance at 280 nm using a calculated extinction coefficient of 104,280 $M^{-1}cm^{-1}$ for all SadX variants (Benchling [Biology Software], https://benchling.com). The protein sample was flash frozen in liquid nitrogen and stored at -80 °C until used.

3.4.7 Activity assays

Screening azidation assay in lysate

Cell pellets in 2 mL 96-deep well plates were suspended with 125 µL of 1 mg/mL lysozyme in 10 mM HEPES pH 7.5 and incubated at 37 °C and 250 rpm for 45 minutes to achieve lysis, after which the lysates were flash frozen in liquid nitrogen and thawed in a water bath at 37 °C. 10 µL of 1 mg/mL DNase in 10 mM HEPES pH 7.5 were added and the lysates were incubated at 37 °C and 250 rpm for 15 minutes, followed by centrifugation at 3600 rpm for 15 minutes.

Screening reactions were set up in microtiter v-bottom 96-well plates by combining 25 µL/well of solution A (50 mL containing 8 mM α-ketoglutarate, 32 mM NaN_3, 32 mM ascorbic acid, and 0.4 mM $Fe(NH_4)_2(SO_4)_2$ in 10 mM HEPES pH 7.5), 25 µL/well of solution B (30 mL of 4 mM **1a** in 10 mM HEPES pH 7.5), and 50 µL/well of clarified lysate. The final reaction conditions were 100 µL of 1 mM **1a**, 2 mM α-ketoglutarate, 8 mM NaN_3, 8 mM ascorbic acid,

0.1 mM $Fe(NH_4)_2(SO_4)_2$, and 50% v/v lysate in 10 mM HEPES pH 7.5. The plates were sealed with breathable film and shaken at room temperature and 750 rpm for 3 hours, after which 100 µL of methanol were added to quench the reactions and precipitated proteins were removed by centrifugation at 3600 rpm for 10 minutes. 50 µL of the resulting supernatant were diluted with 100 µL of water, filtered through 0.22 µm filter plates by centrifugation at 3600 rpm for 10 minutes, and analyzed via LC-MS using LC-MS method 1.

The conversion values to **1e** for the four parent samples in each plate were averaged and used as comparison for every other variant in the same plate. Library variants were first evaluated so that only variants with activity higher than the sum of average parent activity and 2.5 standard deviations of parent activity were considered as potential hits. Chemoselectivity was then used to further evaluate the reduced list of potential hits and to select up to 10 variants to be sequenced and validated with bioconversions using purified protein.

Screening chlorination assay in lysate

Screening chlorination reactions were carried out analogously to screening azidation reactions except for solution A containing NaCl instead of NaN_3 (8 mM α-ketoglutarate, 200 mM NaCl, 32 mM ascorbic acid, and 0.4 mM $Fe(NH_4)_2(SO_4)_2$ in 10 mM HEPES pH 7.5). The final reaction conditions were 100 µL of 1 mM **1a**, 2 mM α-ketoglutarate, 50 mM NaCl, 8 mM ascorbic acid, 0.1 mM $Fe(NH_4)_2(SO_4)_2$, and 50% v/v lysate in 10 mM HEPES pH 7.5.

Hydroxylation reactions to evaluate screening method

Reactions shown in Figure 3.4 were set up analogously to screening azidation reactions except for solution A not containing NaN_3 (8 mM α-ketoglutarate, 32 mM ascorbic acid, and 0.4 mM $Fe(NH_4)_2(SO_4)_2$ in 10 mM HEPES pH 7.5). The final reaction conditions were 100 µL of 1

mM **1a**, 2 mM α-ketoglutarate, 8 mM ascorbic acid, 0.1 mM $Fe(NH_4)_2(SO_4)_2$, and 50% v/v lysate in 10 mM HEPES pH 7.5.

*Small-scale azidation of **1a***

Enzyme activity was determined by small-scale bioconversions using purified protein. Reactions were set up by combining in a 96-well plate 25 µL/well of solution A (25 mL containing 8 mM α-ketoglutarate, 32 mM NaN_3, 32 mM ascorbic acid, and 0.4 mM $Fe(NH_4)_2(SO_4)_2$ in 10 mM HEPES pH 7.5), 25 µL/well of solution B (2 mL of 4 mM **1a** in 10 mM HEPES pH 7.5), and 50 µL/well of solution C (200 µL of 0.2 mM of SadX variant). The final reactions (100 µL) had final concentrations of 1 mM **1a**, 2 mM α-ketoglutarate, 8 mM NaN_3, 8 mM ascorbic acid, 0.1 mM $Fe(NH_4)_2(SO_4)_2$, and 0.1 mM SadX variant in 10 mM HEPES pH 7.5. The plate was sealed with breathable film and shaken at room temperature and 750 rpm for 3 hours, after which 100 µL of methanol were added to quench the reactions, 50 µL of a 2 mM *N*-acetyl-L-valine solution in water (100 nmol) were added as internal standard, and precipitated proteins were removed by centrifugation at 3600 rpm for 10 minutes. Finally, 80 µL of the resulting supernatant were filtered through 0.22 µm filter plates by centrifugation at 3600 rpm for 10 minutes and analyzed via LC-MS using LC-MS method 2.

Further reaction optimization led to conditions shown in Figures 3.12, 3.13, and 3.14. The corresponding reaction setup was analogous to the one described above except for solution A containing 16 mM α-ketoglutarate, 32 mM NaN_3, 64 mM ascorbic acid, and 0.8 mM $Fe(NH_4)_2(SO_4)_2$ in 10 mM HEPES pH 7.5. The final reaction conditions were 1 mM **1a**, 4 mM α-ketoglutarate, 8 mM NaN_3, 16 mM ascorbic acid, 0.2 mM $Fe(NH_4)_2(SO_4)_2$, and 0.1 mM SadX variant in 10 mM HEPES pH 7.5.

*Small-scale chlorination of **1a***

Reaction setup was analogous to azidation reactions, with the exception of solution A containing 16 mM α-ketoglutarate, 200 mM NaCl, 64 mM ascorbic acid, and 0.8 mM $Fe(NH_4)_2(SO_4)_2$ resulting in reactions (100 µL) with final concentrations of 1 mM **1a**, 4 mM α-ketoglutarate, 50 mM NaCl, 16 mM ascorbic acid, 0.2 mM $Fe(NH_4)_2(SO_4)_2$, and 0.1 mM SadX variant in 10 mM HEPES pH 7.5.

*Small-scale azidation of **2a-7a***

Conversion values to **2e–7e** were obtained by carrying out small-scale bioconversions. Reactions with substrates **2a**–**6a** were set up identically as reactions with **1a** previously described resulting in 100 µL of 1 mM substrate, 4 mM α-ketoglutarate, 8 mM NaN_3, 16 mM ascorbic acid, 0.2 mM $Fe(NH_4)_2(SO_4)_2$, and 0.1 mM SadX variant in 10 mM HEPES pH 7.5. Reactions with substrate **7a** were prepared by combining 25 µL/well of solution D (50 mL containing 32 mM α-ketoglutarate, 32 mM NaN_3, 64 mM ascorbic acid, and 0.8 mM $Fe(NH_4)_2(SO_4)_2$ in 10 mM HEPES pH 7.5), 25 µL/well of solution B (15 mL of 4 mM **7a** in 10 mM HEPES pH 7.5), and 50 µL/well of solution C (15 mL of 0.4 mM 2-L in 10 mM HEPES pH 7.5). The final reaction conditions were 100 µL of 1 mM **7a**, 8 mM α-ketoglutarate, 8 mM NaN_3, 16 mM ascorbic acid, 0.2 mM $Fe(NH_4)_2(SO_4)_2$, and 0.1 mM 2-L in 10 mM HEPES pH 7.5. All bioconversions with **2a**–**7a** were treated as described before and analyzed using LC-MS method 2.

Small-scale carbamate formation reactions

The activity of SadX variants towards **1a** in the presence of NaOCN was evaluated using small-scale bioconversions. These reactions were set up analogously to small-scale azidations of **1a** except for solution A containing NaOCN instead of NaN_3 (16 mM α-ketoglutarate, 32 mM NaOCN, 64 mM ascorbic acid, and 0.8 mM $Fe(NH_4)_2(SO_4)_2$ in 10 mM HEPES pH 7.5). The

final reaction conditions were 1 mM **1a**, 4 mM α-ketoglutarate, 8 mM NaOCN, 16 mM ascorbic acid, 0.2 mM $Fe(NH_4)_2(SO_4)_2$, and 0.1 mM SadX variant in 10 mM HEPES pH 7.5.

*Small-scale nitration of **1a***

The activity of SadX towards **1a** in the presence of $NaNO_2$ was evaluated using small-scale bioconversions. These reactions were set up analogously to small-scale azidations of **1a** except for solution A containing $NaNO_2$ instead of NaN_3 (8 mM α-ketoglutarate, 32 mM $NaNO_2$, 32 mM ascorbic acid, and 0.4 mM $Fe(NH_4)_2(SO_4)_2$ in 10 mM HEPES pH 7.5). The final reaction conditions were 1 mM **1a**, 2 mM α-ketoglutarate, 8 mM $NaNO_2$, 8 mM ascorbic acid, 0.1 mM $Fe(NH_4)_2(SO_4)_2$, and 0.1 mM SadX in 10 mM HEPES pH 7.5.

*Scaled-up azidation of **1a***

In order to obtain several milligrams of azide **1e** to carry out further reactions, a modified protocol from the one reported in chapter 2 was used.[58] A reaction was set up in a 250-mL beaker by adding **1a** (6 mL of 5 mM stock, 0.03 mmol), α-ketoglutarate (6 mL of 20 mM stock, 0.12 mmol), NaN_3 (6 mL of 40 mM stock, 0.24 mmol), ascorbic acid (6 mL of 80 mM, 0.48 mmol), $Fe(NH_4)_2(SO_4)_2$ (300 µL of 20 mM), and 4-IC (750 µL of 2 mM stock, 1.5 nmol, 5 mol%). All stock solutions were prepared in 10 mM HEPES, and the pH was adjusted to 7.5 using NaOH except for the solution of $Fe(NH_4)_2(SO_4)_2$ which was prepared in MilliQ water. Buffer 10 mM HEPES pH 7.5 was added to the beaker so that the final volume was 30 mL. The final concentrations were 1 mM **1a**, 4 mM α-ketoglutarate, 8 mM NaN_3, 16 mM ascorbic acid, 0.2 mM $Fe(NH_4)_2(SO_4)_2$, and 50 µM 4-IC in 10 mM HEPES pH 7.5. The beaker was covered with breathable film and shaken in a horizontal incubator at 125 rpm and 30 °C for 3 hours, and the resulting reaction mixture was quenched by addition of 30 mL of methanol.

The sample was divided into two 50-mL conical tubes and the precipitated protein was removed by centrifugation at 15000 rpm for 15 minutes. The supernatant was concentrated down in a rotary evaporator and purified using semi-preparative HPLC using HPLC method 1. The fractions containing product **1e** were confirmed by LC-MS, combined, and dried by rotary evaporation. Spectral ^{1}H- and ^{13}C-NMR data matched the one reported in chapter 2 (Figure A1.1).[58]

*Scaled-up carbamate formation with **1a***

A larger scale bioconversion with **1a** and 2-L in the presence of NaOCN was carried out to isolate product **1e$_{OCN}$** was conducted using a modified protocol from the one reported in chapter 2.[58] A 250 mL polypropylene beaker was charged with **1a** (5 mM, 6 mL), 2-L (1.5 mM, 1 mL), $Fe(NH_4)_2(SO_4)_2$ (20 mM, 0.3 mL), ascorbic acid (80 mM, 6 mL), NaOCN (40 mM, 6 mL), and α-ketoglutarate (20 mM, 6 mL) in 10 mM HEPES, pH 7.5. Buffer was added to make up to a 30 mL reaction volume. All stocks were made in 10 mM HEPES and pH-adjusted to pH 7.5 with 10 M NaOH prior to addition, with the exception of the stock of $Fe(NH_4)_2(SO_4)_2$ which was prepared in water. The final concentrations were 1 mM **1a**, 50 μM 2-L, 200 μM $Fe(NH_4)_2(SO_4)_2$, 16 mM ascorbic acid, 8 mM NaOCN, and 4 mM α-ketoglutarate. The beaker was sealed with a breathable film and shaken at 125 rpm and 30 °C for 3 hours. Full consumption of **1a** on was confirmed by LC-MS of the reaction mixture. The reaction was quenched by addition of 30 mL of methanol and clarified by centrifugation at 15,000 rpm for 15 minutes. The supernatant was filtered, concentrated down by rotary evaporation, and purified via semipreparative HPLC. Fraction purity was checked using LC-MS, and those containing **1e$_{OCN}$** were consolidated, dried, and analyzed via NMR. Spectral ^{1}H NMR, ^{13}C NMR, and ^{1}H-^{13}C HSQC spectral data matched the one reported in chapter 2 (Figure A1.3).[58]

3.4.8 Steady-state kinetic characterization

The kinetic parameters for azidation of **1a** with SadX and 4-IC were determined to further evaluate the effects of the directed evolution performed. Bioconversions were set up at different concentrations of **1a** (Figure 3.16) and analyzed in triplicate via LC-MS (LC-MS method 4). Stock solutions of reagents were made in 10 mM HEPES pH7.5 buffer, except for $Fe(NH_4)_2(SO_4)_2$ stock solutions that were prepared in MilliQ water. To minimize errors in response factors due to instrumental fluctuations, samples were analyzed alongside freshly made calibration curves.

For SadX kinetics, a master stock solution of 504 µM **1a** was diluted in 10 mM HEPES pH 7.5 with a 2x dilution factor to generate eight stock solutions (540 µM, 270 µM, 135 µM, 67.5 µM, 33.75 µM, 16.88 µM, 8.44 µM, and 4.22 µM **1a**). 30 µL of each stock solution were added to five wells of a 96-well plate, and to each of these wells were added 30 µL of a solution containing 1mM NaN_3, 0.3 mM $Fe(NH_4)_2(SO_4)_2$, 24 mM ascorbic acid, and 0.3 µM SadX. Two additional 96-well plates were set up in an identical manner to have triplicate samples. Samples for t=0 were prepared by adding additional 30 µL of 10 mM HEPES pH 7.5 buffer and 90 µL of methanol. Bioconversions were started in the rest of the samples by addition of 30 µL of a 30 mM α-ketoglutarate solution while being shaken at 750 rpm. The final reactions consisted of 90 µL of 1.41 – 180 µM **1a**, 10 mM α-ketoglutarate, 0.5 mM NaN_3, 0.1 µM SadX, 0.1 mM Fe(II), and 8 mM ascorbic acid. The reactions were quenched at 2, 4, 6, and 8 minutes by the addition of 90 µL of methanol, followed by addition of 20 µL of 10 µM *N*-acetyl-L-valine (200 pmol) as internal standard and centrifugation at 3600 rpm for 5 minutes to remove any precipitated protein. The final reactions were filtered through a 0.22 µm filter plate and analyzed via LC-MS.

Kinetics for 4-IC were performed identically, with the exceptions the enzyme being used (4-IC instead of SadX) and the substrate concentration range. The master stock solution of **1a** was a 1,512 µM solution, and the serial dilutions resulted in stock solutions of 1,512 µM, 1,008 µM, 672 µM, 336 µM, 168 µM, 84 µM, 42 µM, and 21 µM. The final reactions contained 7 – 540 µM of **1a**.

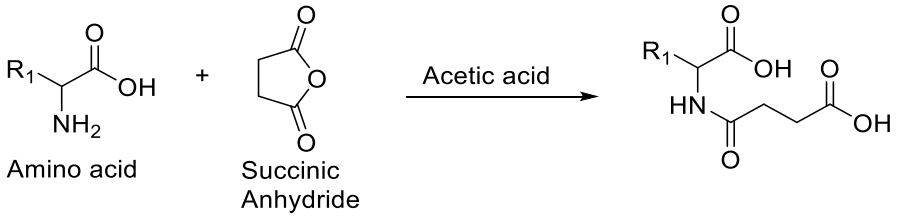
Calibration curves were constructed by preparing samples with varying known pmol of isolated **1e** and 200 pmol of *N*-acetyl-L-valine. These samples were analyzed using the same LC-MS method.

Conversion values were obtained by comparing the ratio of areas of **1e** over internal standard with the calibration curve, and these values were used to obtained initial rates of formation of **1e** per unit of enzyme (v/E) via a linear fit of [**1e**]/[Enzyme] as a function of time. K_M and k_{cat} values were obtained by fitting v/E as a function of [**1a**] following Michaelis-Menten kinetics.

3.4.9 Synthesis of starting materials

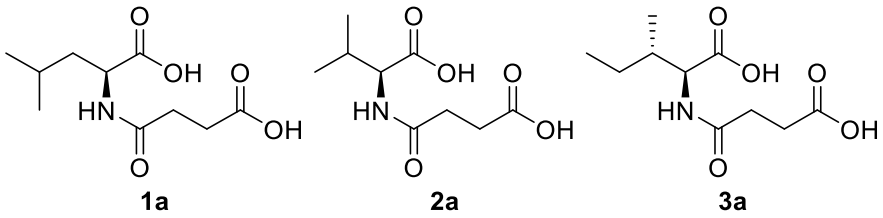
*General synthesis of N-succinylated amino acids **1a** – **6a***

Scheme 3.9: Synthesis of *N*-succinylated amino acids



To a 100 mL flask was added amino acid (1 eq), succinic anhydride (1.05 eq), and acetic acid. The mixture was heated to 70 °C for 8 hr. After complete removal of the solvent, the solid residue was recrystallized from ethyl acetate and ethanol to obtain semi purified product. The solid residue was purified in a CombiFlash system using water and acetonitrile.

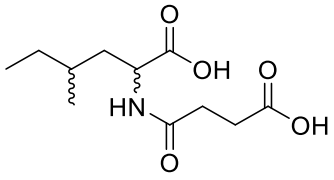
*Substrates **1a**, **2a**, and **3a***



1a **2a** **3a**

^{1}H- and ^{13}C-NMR spectral data matched literature spectra.[62]

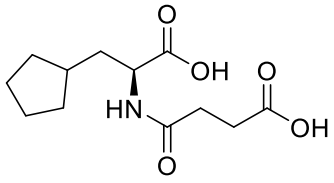
*Substrate **4a***



4a
Mixture of diastereomers

^{1}H NMR (500 MHz, CD_3OD) δ 4.45 – 4.35 (m, 1H), 2.63 – 2.45 (m, 4H), 1.82 – 1.11 (m, 5H), 0.96 – 0.84 (m, 6H). ^{13}C NMR (126 MHz, CD_3OD) δ C 176.19, C 176.05, C 174.48, CH 53.32, CH 38.94, CH_2 38.00, CH_2 33.77, CH_2 33.10, CH_2 31.35, CH_2 30.24, CH_2 26.08, CH_2 25.88. NMR spectra are shown in Figures AII.1, AII.2, and AII.3

*Substrate **5a***

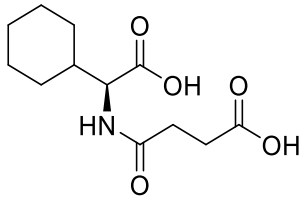


5a

^{1}H NMR (400 MHz, CD_3OD) δ 4.38 (dd, *J* = 9.45, 5.08 Hz, 1H), 2.66 – 2.47 (m, 4H), 1.93 (h, *J* = 7.85 Hz, 1H), 1.86 – 1.71 (m, 4H), 1.68 – 1.61 (m, 2H), 1.59 – 1.51 (m, 2H), 1.19 – 1.10 (m, 2H). ^{13}C NMR (126 MHz, CD_3OD) δ C 176.19, C 176.05, C 174.48, CH 53.32, CH

38.94, CH_2 38.00, CH_2 33.77, CH_2 33.10, CH_2 31.35, CH_2 30.24, CH_2 26.08, CH_2 25.88. NMR spectra are shown in Figures AII.4 and AII.5.

*Substrate **6a***



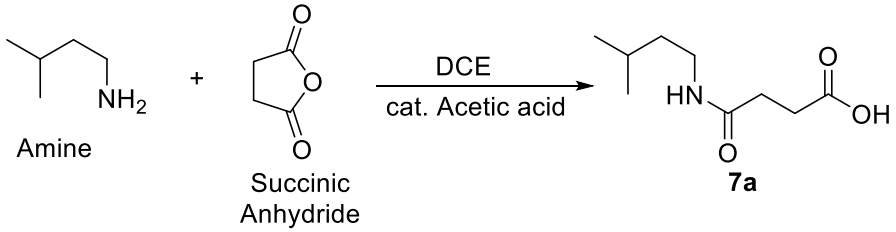
6a

^{1}H NMR (500 MHz, CD_3OD) δ 4.31 (d, *J* = 5.99 Hz, 1H), 2.65 – 2.47 (m, 4H), 1.85 – 1.62 (m, 6H), 1.36 – 1.07 (m, 5H). ^{13}C NMR (126 MHz, CD_3OD) δ C 176.24, C 174.94, C 174.60, CH 58.70, CH 41.47, CH_2 31.34, CH_2 30.77, CH_2 30.31, CH_2 29.47, CH_2 27.21, CH_2 27.20, CH_2 27.17. NMR spectra are shown in Figures AII.6 and AII.7.

*Substrate **7a***

To a 100 mL flask attached to a reflux condenser was added amine (5 g, 1 eq), succinic anhydride (6.1 g, 1.05 eq), a catalytic amount of acetic acid, and 50 mL of dichloroethane. The mixture was heated to 40 °C for 12 hr. After removal of the solvent, the solid residue was recrystallized from toluene to obtain a white solid in a yield of 92%.

Scheme 3.10: Synthesis of *N*-succinylated amine 7a



7a

^{1}H NMR (500 MHz, CD_3OD) δ 3.19 (t, *J* = 7.4 Hz, 2H), 2.58 (t, *J* = 6.98 Hz, 2H), 2.45 (t, *J* = 7.00 Hz, 2H), 1.62 (nonet, *J* = 6.66 Hz, 1H), 1.38 (q, *J* = 7.05 Hz, 2H). 0.92 (d, *J* = 6.68 Hz, 6H). ^{13}C NMR (126 MHz, CD_3OD) δ C 176.20, C 174.31, CH 58.70, CH_2 39.28, CH_2 38.70,

CH_2 31.57, CH_2 30.36, CH 26.84, CH_3 22.81, CH_3 22.81. NMR spectra are shown in Figures AII.8 and AII.9.

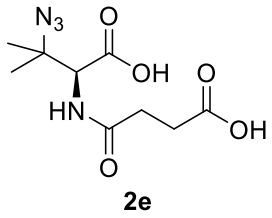
3.4.10 Isolation of azidated products via bioconversions

Small-scale reactions with **2a**–**7a** were carried out according to the protocols described before in parallel with a total amount of starting material of 10 – 15 mg. The plates were sealed with breathable film and shaken at room temperature and 750 rpm for 3 hours, after which the resulting reaction mixtures were combined, quenched with 1 equivalent volume of methanol, and the precipitated protein was removed by centrifugation at 4700 rpm for 15 minutes. The supernatant was then concentrated down in a rotary evaporator and purified using semi-preparative HPLC according to Table 3.4. The collected fractions were combined and dried in a rotary evaporator.

Table 3.4: HPLC methods used to purify compounds 2e – 7e

Compound	HPLC method
2e	1
3e	2
4e	3
5e	4
6e	4
7e	5

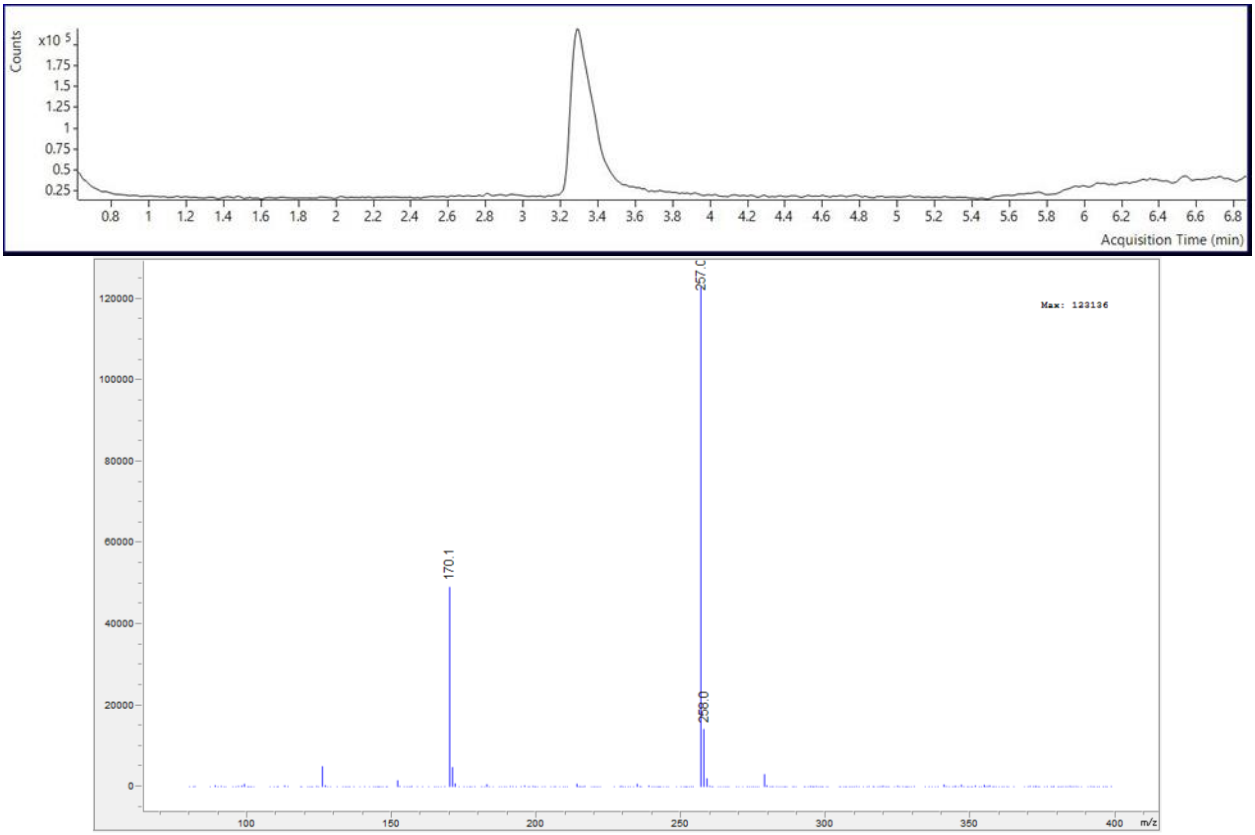
*Product **2e** via reaction of **2a** with SadX*



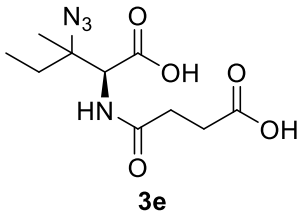
2e

^{1}H NMR (500 MHz, CD_3OD) δ 4.46 (s, 1H), 2.62 – 2.54 (m, 4H), 1.41 (s, 3H), 1.35 (s, 3H). ^{13}C NMR (126 MHz, CD_3OD) δ C 176.50, C 174.30, C 173.78, CH 63.60, C 61.62, CH_2 31.76, CH_2 30.53, CH_3 24.65, CH_3 23.74. NMR spectra shown in Figures AII.10 and AII.11.

Figure 3.20: LC-MS ESI- of product 2e



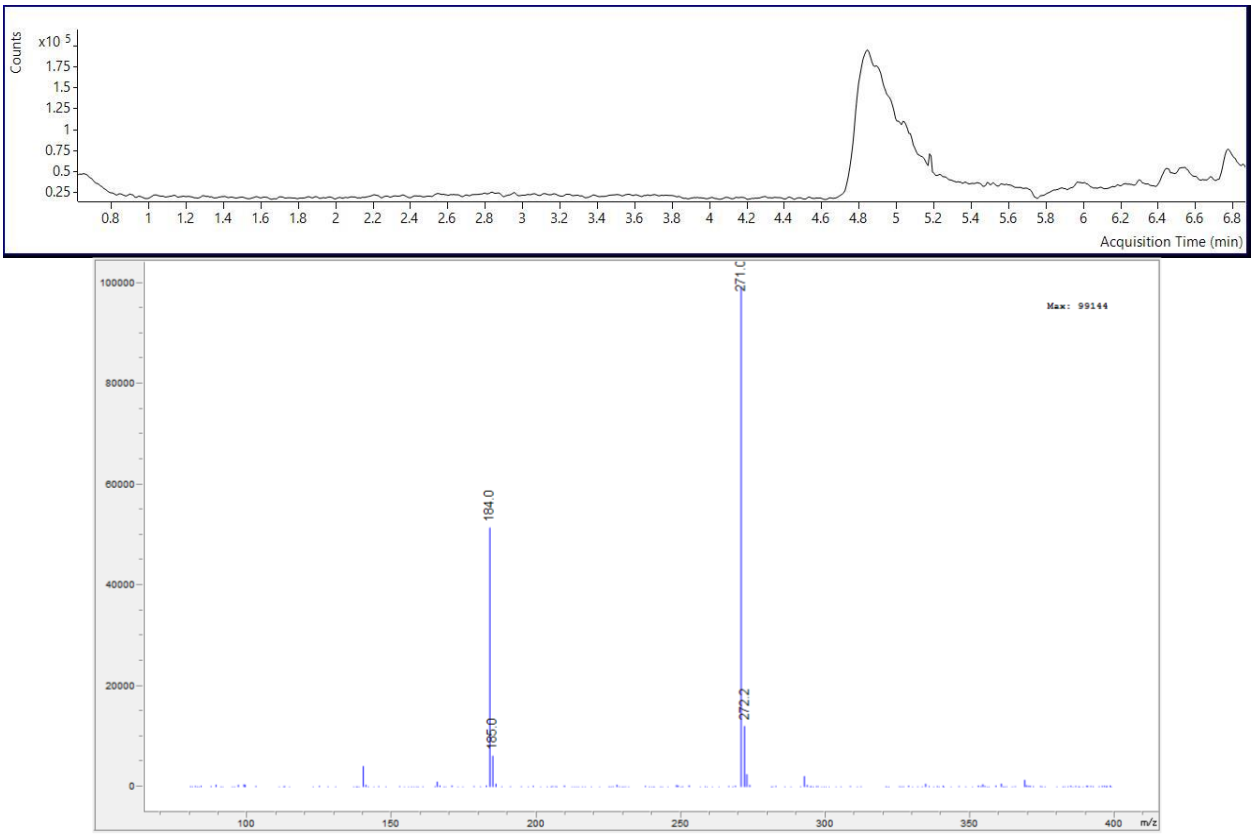
*Product **3e** via reaction of **3a** with 1-VH*



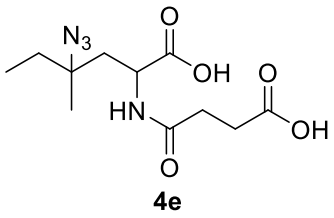
3e

^{1}H NMR (500 MHz, CD_3OD) δ 4.49 (s, 1H), 2.63 – 2.51 (m, 4H), 1.72 – 1.65 (qd, *J* = 7.15, 3.28 Hz, 2H), 1.38 (s, 3H), 0.97 (t, *J* = 7.39 Hz, 3H). ^{13}C NMR (126 MHz, CD_3OD) δ C 66.69, CH 61.51, CH_2 32.39, CH_2 32.38, CH_2 31.57, CH_3 19.60, CH_3 8.60. NMR spectra shown in Figures AII.12, AII.13, and AII.14.

Figure 3.21: LC-MS ESI- of product 3e



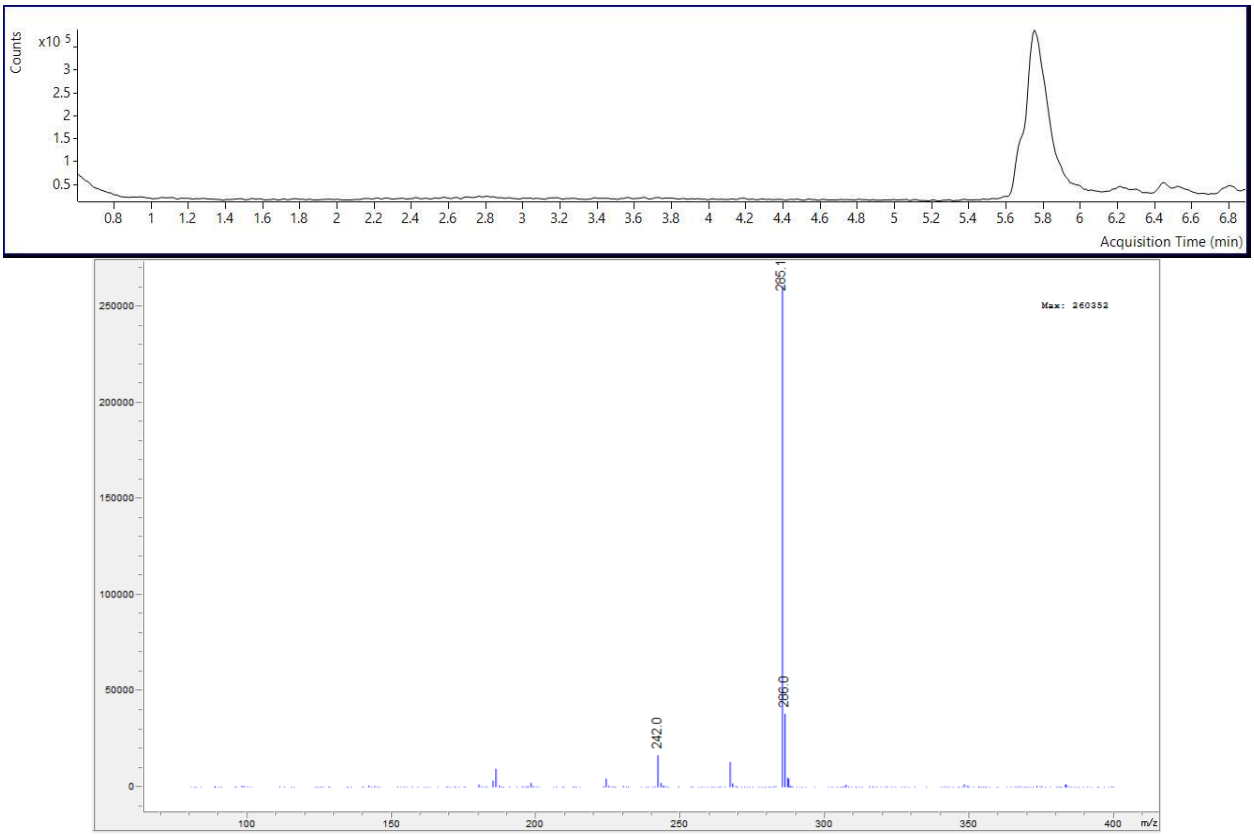
*Product **4e** via reaction of **4a** with 2-L*



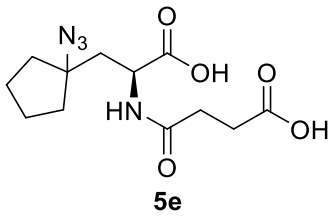
4e

^{1}H NMR (500 MHz, CD_3OD) δ 4.48 – 4.45 (m, 1H), 2.61 – 2.45 (m, 4H), 2.14 – 2.10 (m, 1H), 1.85 – 1.77 (m, 1H), 1.63 (q, J = 7.32 Hz, 2H), 1.31 – 1.29 (m, 6H). ^{13}C NMR (126 MHz, CD_3OD) δ C 65.01, CH 52.68, CH_2 42.53, CH_2 33.21, CH_2 33.02, CH_2 32.51, CH_3 23.11, CH_3 22.57. NMR spectra shown in Figures AII.15, AII.16, and AII.17.

Figure 3.22: LC-MS ESI- of compound 4e



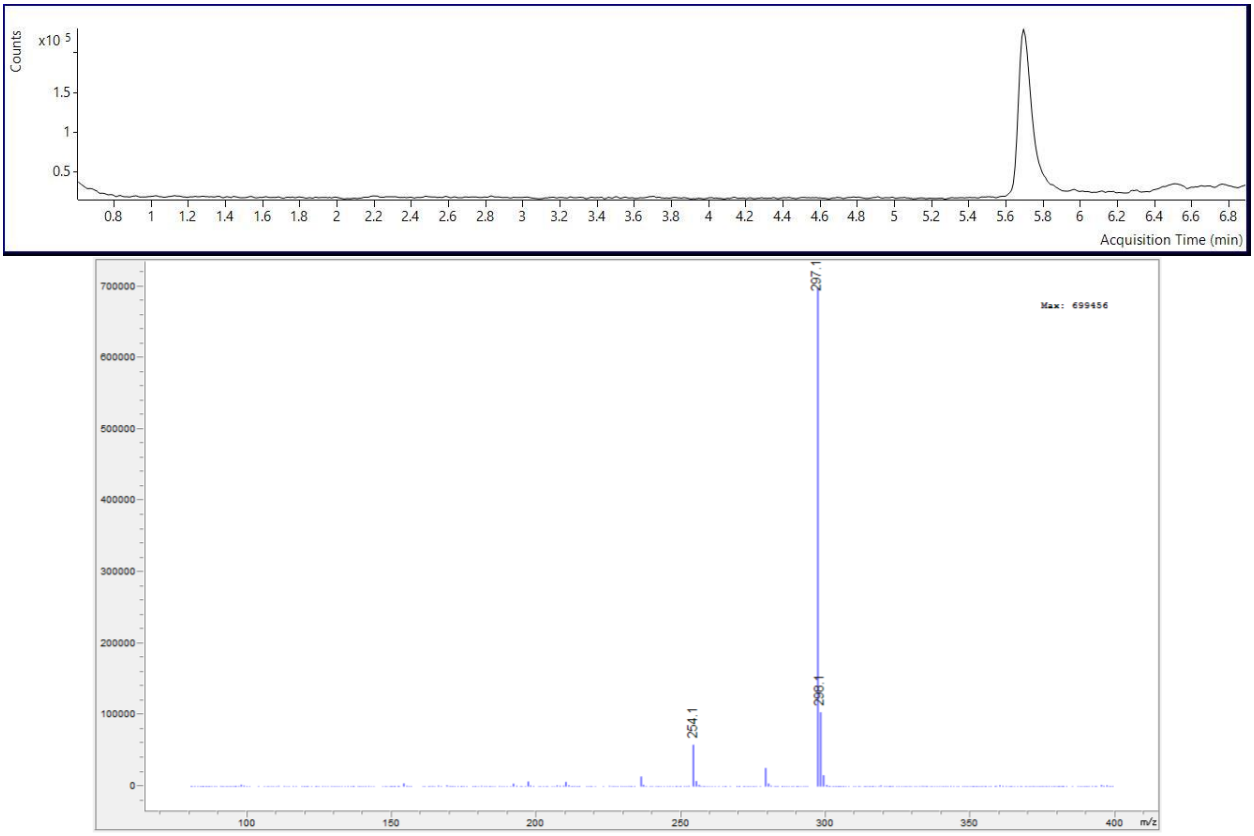
*Product **5e** via reaction of **5a** with 4-IC*



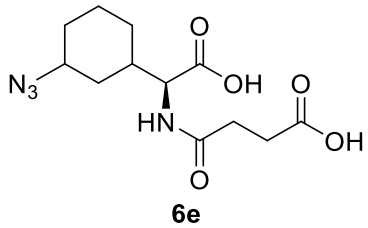
5e

^{1}H NMR (500 MHz, CD_3OD) δ 4.45 (dd, *J* = 9.08, 3.53 Hz, 1H), 2.61 – 2.52 (m, 4H), 2.29 (dd, *J* = 14.69, 3.52 Hz, 1H), 1.98 (dd, *J* = 14.68, 9.12 Hz, 1H), 1.91 – 1.81 (m, 2H), 1.77 – 1.71 (m, 4H), 1.70 – 1.64 (m, 2H). ^{13}C NMR (126 MHz, CD_3OD) δ C 176.98, C 174.08, C 174.09, C 73.61, CH 52.46, CH_2 41.49, CH_2 37.91, CH_2 37.88, CH_2 32.06, CH_2 30.73, CH_2 24.60, CH_2 24.08. NMR spectra shown in Figures AII.18, AII.19, and AII.20.

Figure 3.23: LC-MS ESI- of compound 5e



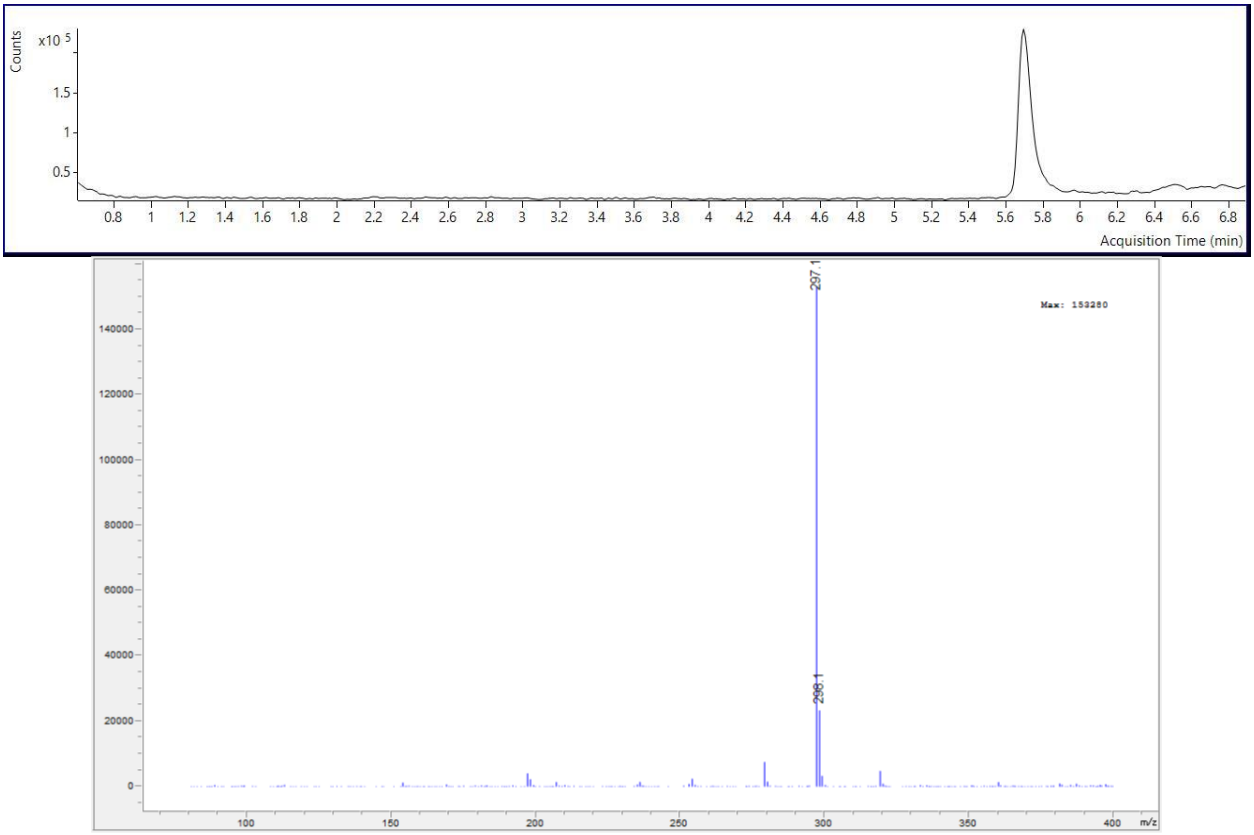
*Product **6e** via reaction of **6a** with 1-VH*



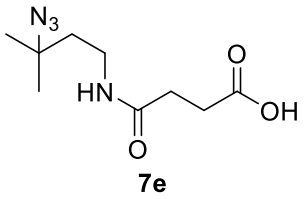
6e

^{1}H NMR (500 MHz, CD_3OD) δ 4.31 (d, *J* = 5.55 Hz, 1H), 3.97 (p, *J* = 3.30 Hz, 1H), 2.61 – 2.51 (m, 4H), 2.21 – 2.16 (m, 1H), 1.82 – 1.75 (m, 2H), 1.67 – 1.64 (m, 1H), 1.60 – 1.55 (m, 2H), 1.55 – 1.51 (m, 1H), 1.48 – 1.42 (m, 1H), 1.23 – 1.18 (m, 1H). ^{13}C NMR (126 MHz, CD_3OD) δ C 133.80, C 133.13, C 129.94, CH 59.34, CH 58.82, CH 36.42, CH_2 34.26, CH_2 31.93, CH_2 30.88, CH_2 30.36, CH_2 28.19, CH_2 21.49. NMR spectra shown in Figures AII.21, AII.22, and AII.23.

Figure 3.24: LC-MS ESI- of compound 6e



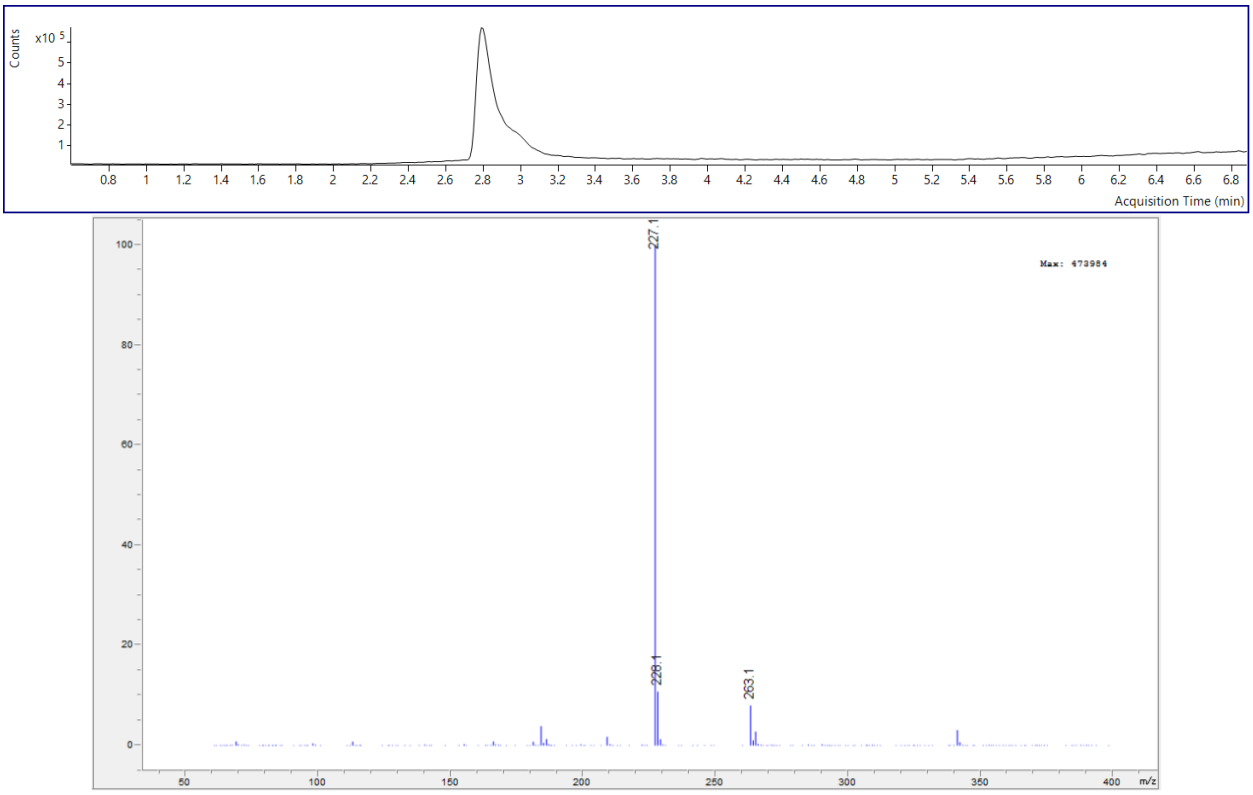
*Product **7e** via reaction of **7a** with 2-L*



7e

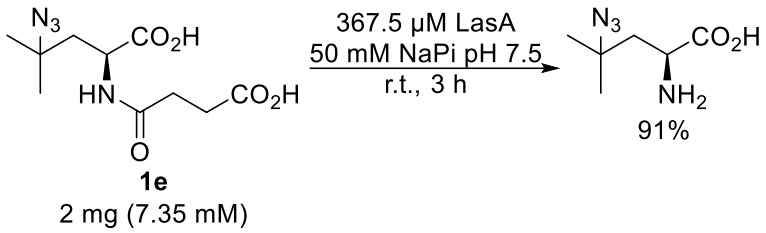
^{1}H NMR (500 MHz, CD_3OD) δ 3.27 (dd, *J* = 15.48, 7.88 Hz, 2H), 2.56 (t, *J* = 7.06, 2H), 2.44 (t, *J* = 7.09 Hz, 2H), 1.70 (dd, *J* = 15.54, 7.79 Hz, 2H), 1.31 (s, 6H). ^{13}C NMR (126 MHz, CD_3OD) δ C 177.26, C 174.74, C 61.49, CH_2 41.31, CH_2 36.33, CH_2 32.18, CH_2 31.29, CH_3 26.14. NMR spectra shown in Figures AII.24, AII.25, and AII.26.

Figure 3.25: LC-MS ESI- of compound 7e



3.4.11 Transformations with product 1e

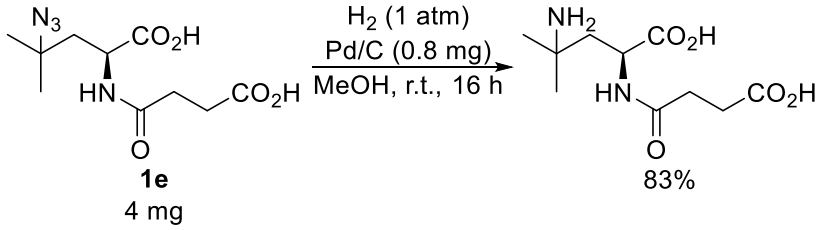
*Desuccinylation of **1e***



LasA bioconversions were set up in duplicate by combining in a 1.5-mL microcentrifuge tube **1e** (2 mg, 7.35 µmol) and LasA (112 µL of a 3.28 mM stock in 50 mM Tris-HCl buffer pH 7.5, 367.5 µmol) in 50 mM sodium phosphate buffer pH 7.5 (888 µL). LasA stock concentration was determined by absorbance at 280 nm using a calculated extinction coefficient of 22,920 $M^{-1}cm^{-1}$ (Benchling [Biology Software], https://benchling.com). The final reactions were 1 mL in volume with 7.35 mM **1e** and 367.5 µM LasA in 50 mM sodium phosphate buffer pH 7.5. The tubes were shaken at 750 rpm at room temperature for 3 hours. Complete conversion of starting material was confirmed by LC-MS (LC-MS method 3). Both 1-mL reactions were combined and quenched by addition of 2 mL of methanol. Protein was removed by centrifugation at 4700 rpm for 15 minutes and methanol was removed by rotary evaporation. The sample was acidified with 6M HCl to pH ~1 and loaded onto a 100 mL chromatography column containing 11 g of DOWEX™ 50WX8 resin previously equilibrated by washing with 200 mL of methanol, 300 mL of water, 1 M HCl until the pH of the flowthrough was ~1, and water until the pH of the flowthrough was 6–7. Once the sample was loaded, the column was washed with 500 mL of water and the cationic species were eluted with 1 M NH_4OH until no more product was detected by LC-MS (200 mL). The fractions containing product were combined and concentrated down by rotary evaporation. The resulting sample was then purified via semi-preparative HPLC using HPLC method 6 and dried by rotary evaporation. Yield: 91%

^{1}H NMR (500 MHz, CD_3OD) δ 3.63 (dd, *J* = 10.7, 2.45 Hz, 1H), 2.19 (dd, *J* = 15.22, 2.44 Hz, 1H), 1.84 (dd, *J* = 15.26, 10.07 Hz, 1H), 1.45 (s, 3H), 1.44 (s, 3H). ^{13}C NMR (126 MHz, CD_3OD) δ C 174.15, C 61.77, CH 53.32, CH_2 43.42, CH_3 27.39, CH_3 24.47. NMR spectra shown in Figures AII.27, AII.28, and AII.29.

*Reduction of **1e***

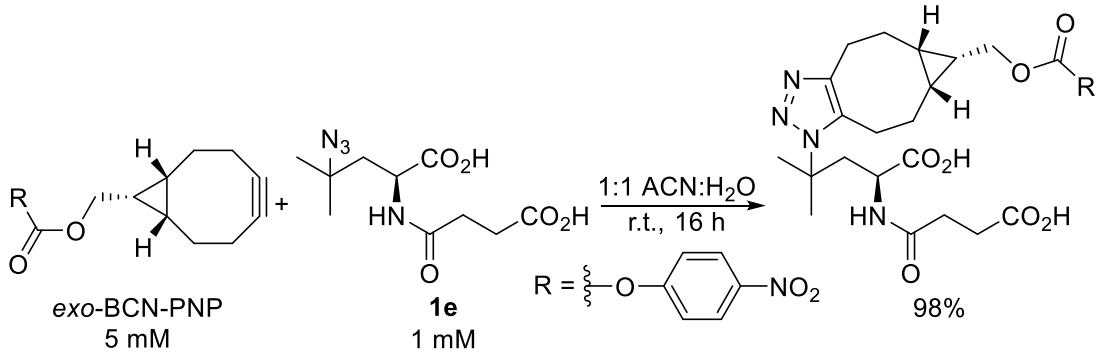


A 20-mL scintillation vial was charged with Pd/C (0.8 mg, 7.5 µmol), sealed, and cycled through vacuum and a nitrogen atmosphere three times to remove air. A solution of **1e** (4 mg, 14.7 µmol) in 2 mL of methanol was added to the vial using a syringe and hydrogen was bubbled through the for 15 minutes. The reaction mixture was stirred at room temperature under one atmosphere of hydrogen for 16 hours. The Pd/C was removed through filtration through a celite plug, and the filtrate was concentrated down by rotary evaporation. Purification was carried out by semi-preparative HPLC (HPLC method 6), and fractions containing the amine product were identified via LC-MS (LC-MS method 3). Yield: 83%

^{1}H NMR (500 MHz, CD_3OD) δ 4.39 (dd, *J* = 7.10, 5.99 Hz, 1H), 2.64 – 2.46 (m, 4H), 2.09 (dd, *J* = 15.02, 7.31 Hz, 1H), 1.88 (dd, *J* = 15.04, 5.89 Hz, 1H), 1.46 (s, 3H), 1.38 (s, 3H). ^{13}C NMR (126 MHz, CD_3OD) δ C 174.57, C 169.70, C 169.57, C 53.96, CH 52.08, CH_2 44.38,

CH_2 32.48, CH_2 31.38, CH_3 27.81, CH_3 25.60. NMR spectra shown in Figures AII.30, AII.31, and AII.32.

*Strain-promoted azide-alkyne cycloaddition of **1e** and exo-BCN-PNP*



To 1.5-mL microcentrifuge tubes were added 100 µL of a 2 mM **1e** aqueous solution and 100 µL of a solution of 10 mM *exo*-BCN-PNP in acetonitrile. The final concentrations in the 200 µL reactions were 5 mM *exo*-BCN-PNP and 1 mM **1e** in 1:1 water:acetonitrile. These reactions were shaken at 750 rpm at room temperature for 24 hours, after which they were filtered through 0.22 µM filter plates and analyzed via LC-MS (LC-MS method 5). Percent conversion was determined by comparison of areas in the EICs of substrates and products.

3.5 References

(1) Naumann, K. Influence of Chlorine Substituents on Biological Activity of Chemicals. *Adv. Synth. Catal.* **1999**, *341* (5), 417–435.

(2) Jeschke, P. The Unique Role of Halogen Substituents in the Design of Modern Agrochemicals. *Pest Manag. Sci.* **2010**, *66* (1), 10–27. https://doi.org/10.1002/ps.1829.

(3) Müller, K.; Faeh, C.; Diederich, F. Fluorine in Pharmaceuticals: Looking beyond Intuition. *Science*. American Association for the Advancement of Science September 28, 2007, pp 1881–1886. https://doi.org/10.1126/science.1131943.

(4) Harris, C. M.; Kannan, R.; Kopecka, H.; Harris, T. M. The Role of the Chlorine Substituents in the Antibiotic Vancomycin: Preparation and Characterization of Mono- and Didechlorovancomycin. *J. Am. Chem. Soc.* **1985**, *107* (23), 6652–6658.

(5) Gribble, G. W.; Kauffman, G. B. Natural Organohalogens: A New Frontier for Medicinal Agents? Products of Chemistry Edited By. *Classr. www.JCE.DivCHED.org* • **2004**, *81* (10), 1441.

(6) Gribble, G. W. *Naturally Ocurring Organohalogen Compounds - a Comprehensive Update*; Springer Science & Business Media, 2009.

(7) Vigalok, A.; Kaspi, A. W. Late Transition Metal-Mediated Formation of Carbon-Halogen Bonds. *Top. Organomet. Chem.* **2010**, *31*, 19–38. https://doi.org/10.1007/978-3-642-12073-2_2/SCHEMES/32.

(8) Liu, W.; Groves, J. T. Manganese Catalyzed C−H Halogenation. **2015**. https://doi.org/10.1021/acs.accounts.5b00062.

(9) Petrone, D. A.; Ye, J.; Lautens, M. Modern Transition-Metal-Catalyzed Carbon-Halogen Bond Formation. *Chemical Reviews*. American Chemical Society July 27, 2016, pp 8003–8104. https://doi.org/10.1021/acs.chemrev.6b00089.

(10) Podgoršek, A.; Zupan, M.; Iskra, J. Oxidative Halogenation with "Green" Oxidants: Oxygen and Hydrogen Peroxide. *Angew. Chemie Int. Ed.* **2009**, *48* (45), 8424–8450. https://doi.org/10.1002/ANIE.200901223.

(11) Tu, H.; Zhu, S.; Qing, F. L.; Chu, L. Visible-Light-Induced Halogenation of Aliphatic CH Bonds. *Tetrahedron Lett.* **2018**, *59* (3), 173–179. https://doi.org/10.1016/J.TETLET.2017.12.023.

(12) Smith, K.; El-Hiti, G. A. Regioselective Control of Electrophilic Aromatic Substitution Reactions. *Curr. Org. Synth.* **2005**, *1* (3), 253–274. https://doi.org/10.2174/1570179043366747.

(13) Chatterjee, S.; Paine, T. K. Hydroxylation versus Halogenation of Aliphatic C−H Bonds by a Dioxygen-Derived Iron–Oxygen Oxidant: Functional Mimicking of Iron

Halogenases. *Angew. Chemie* **2016**, *128* (27), 7848–7853. https://doi.org/10.1002/ANGE.201509914.

(14) Kundu, R.; Ball, Z. T. Copper-Catalyzed Remote Sp3 C-H Chlorination of Alkyl Hydroperoxides. *Org. Lett.* **2010**, *12* (11), 2460–2463. https://doi.org/10.1021/ol100472t.

(15) Fawcett, A.; Keller, M. J.; Herrera, Z.; Hartwig, J. F. Site Selective Chlorination of C(Sp3)−H Bonds Suitable for Late-Stage Functionalization. *Angew. Chemie Int. Ed.* **2021**, *60* (15), 8276–8283. https://doi.org/10.1002/ANIE.202016548.

(16) Cushnie, T. P. T.; Cushnie, B.; Lamb, A. J. Alkaloids: An Overview of Their Antibacterial, Antibiotic-Enhancing and Antivirulence Activities. *International Journal of Antimicrobial Agents*. Elsevier November 1, 2014, pp 377–386. https://doi.org/10.1016/j.ijantimicag.2014.06.001.

(17) Kittakoop, P.; Mahidol, C.; Ruchirawat, S. Alkaloids as Important Scaffolds in Therapeutic Drugs for the Treatments of Cancer, Tuberculosis, and Smoking Cessation. *Curr. Top. Med. Chem.* **2014**, *14* (2), 239–252. https://doi.org/10.2174/1568026613666131216105049.

(18) Mayol-Llinàs, J.; Nelson, A.; Farnaby, W.; Ayscough, A. Assessing Molecular Scaffolds for CNS Drug Discovery. *Drug Discov. Today* **2017**, *22* (7), 965–969. https://doi.org/10.1016/J.DRUDIS.2017.01.008.

(19) Roughley, S. D.; Jordan, A. M. The Medicinal Chemist's Toolbox: An Analysis of Reactions Used in the Pursuit of Drug Candidates. *Journal of Medicinal Chemistry*. American Chemical Society May 26, 2011, pp 3451–3479. https://doi.org/10.1021/jm200187y.

(20) Szardenings, A. K.; Burkoth, T. S.; Look, G. C.; Campbell, D. A. A Reductive Alkylation Procedure Applicable to Both Solution- and Solid-Phase Syntheses of Secondary Amines. *J. Org. Chem.* **1996**, *61* (19), 6720–6722. https://doi.org/10.1021/jo960124n.

(21) Yagafarov, N. Z.; Kolesnikov, P. N.; Usanov, D. L.; Novikov, V. V.; Nelyubina, Y. V.; Chusov, D. The Synthesis of Sterically Hindered Amines by a Direct Reductive Amination of Ketones. *Chem. Commun.* **2016**, *52* (7), 1397–1400. https://doi.org/10.1039/C5CC08577B.

(22) Park, Y.; Kim, Y.; Chang, S. Transition Metal-Catalyzed C-H Amination: Scope, Mechanism, and Applications. *Chemical Reviews*. American Chemical Society July 12, 2017, pp 9247–9301. https://doi.org/10.1021/acs.chemrev.6b00644.

(23) Trowbridge, A.; Walton, S. M.; Gaunt, M. J. New Strategies for the Transition-Metal Catalyzed Synthesis of Aliphatic Amines. *Chemical Reviews*. American Chemical Society March 11, 2020, pp 2613–2692. https://doi.org/10.1021/acs.chemrev.9b00462.

(24) Singh, R.; Mukherjee, A. Metalloporphyrin Catalyzed C-H Amination. *ACS Catal.* **2019**, *9* (4), 3604–3617. https://doi.org/10.1021/acscatal.9b00009.

(25) Chan, K.-H.; Guan, X.; Lo, V. K.-Y.; Che, C.-M. Elevated Catalytic Activity of Ruthenium(II)-Porphyrin-Catalyzed Carbene/Nitrene Transfer and Insertion Reactions with N-Heterocyclic Carbene Ligands. *Angew. Chemie* **2014**, *126* (11), 3026–3031. https://doi.org/10.1002/ange.201309888.

(26) Huang, X.; Bergsten, T. M.; Groves, J. T. Manganese-Catalyzed Late-Stage Aliphatic C-H Azidation. *J. Am. Chem. Soc.* **2015**, *137* (16), 5300–5303. https://doi.org/10.1021/jacs.5b01983.

(27) Milczek, E.; Boudet, N.; Blakey, S. Enantioselective C-H Amination Using Cationic Ruthenium(II)-Pybox Catalysts. *Angew. Chemie - Int. Ed.* **2008**, *47* (36), 6825–6828. https://doi.org/10.1002/anie.200801445.

(28) Sharma, A.; Hartwig, J. F. Metal-Catalysed Azidation of Tertiary C–H Bonds Suitable for Late-Stage Functionalization. *Nat. 2015 5177536* **2015**, *517* (7536), 600–604. https://doi.org/10.1038/nature14127.

(29) McIntosh, J. A.; Coelho, P. S.; Farwell, C. C.; Wang, Z. J.; Lewis, J. C.; Brown, T. R.; Arnold, F. H. Enantioselective Intramolecular C-H Amination Catalyzed by Engineered Cytochrome P450 Enzymes in Vitro and in Vivo. *Angew. Chemie - Int. Ed.* **2013**, *52* (35), 9309–9312. https://doi.org/10.1002/anie.201304401.

(30) Singh, R.; Bordeaux, M.; Fasan, R. P450-Catalyzed Intramolecular Sp3 C-H Amination with Arylsulfonyl Azide Substrates. *ACS Catal.* **2014**, *4* (2), 546–552. https://doi.org/10.1021/cs400893n.

(31) Bordeaux, M.; Singh, R.; Fasan, R. Intramolecular C(Sp3)H Amination of Arylsulfonyl Azides with Engineered and Artificial Myoglobin-Based Catalysts. *Bioorganic Med. Chem.* **2014**, *22* (20), 5697–5704. https://doi.org/10.1016/j.bmc.2014.05.015.

(32) Hyster, T. K.; Farwell, C. C.; Buller, A. R.; McIntosh, J. A.; Arnold, F. H. Enzyme-Controlled Nitrogen-Atom Transfer Enables Regiodivergent C-H Amination. *J. Am. Chem. Soc.* **2014**, *136* (44), 15505–15508. https://doi.org/10.1021/ja509308v.

(33) 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. 2017 97* **2017**, *9* (7), 629–634. https://doi.org/10.1038/nchem.2783.

(34) Jia, Z. J.; Gao, S.; Arnold, F. H. Enzymatic Primary Amination of Benzylic and Allylic C(Sp3)-H Bonds. *J. Am. Chem. Soc.* **2020**, *142* (23), 10279–10283. https://doi.org/10.1021/jacs.0c03428.

(35) Athavale, S. V.; Gao, S.; Liu, Z.; Mallojjala, S. C.; Hirschi, J. S.; Arnold, F. H. Biocatalytic, Intermolecular C−H Bond Functionalization for the Synthesis of Enantioenriched Amides. *Angew. Chemie* **2021**, *133* (47), 25068–25073. https://doi.org/10.1002/ANGE.202110873.

(36) Liu, Z.; Qin, Z. Y.; Zhu, L.; Athavale, S. V.; Sengupta, A.; Jia, Z. J.; Garcia-Borràs, M.; Houk, K. N.; Arnold, F. H. An Enzymatic Platform for Primary Amination of 1-Aryl-2-Alkyl Alkynes. *J. Am. Chem. Soc.* **2022**, *144* (1), 80–85. https://doi.org/10.1021/jacs.1c11340.

(37) Goldberg, N. W.; Knight, A. M.; Zhang, R. K.; Arnold, F. H. Nitrene Transfer Catalyzed by a Non-Heme Iron Enzyme and Enhanced by Non-Native Small-Molecule Ligands. *J. Am. Chem. Soc.* **2019**, *141* (50), 19585–19588. https://doi.org/10.1021/jacs.9b11608.

(38) Vila, M. A.; Steck, V.; Rodriguez Giordano, S.; Carrera, I.; Fasan, R. C−H Amination via Nitrene Transfer Catalyzed by Mononuclear Non-Heme Iron-Dependent Enzymes. *ChemBioChem* **2020**, *21* (14), 1981–1987. https://doi.org/10.1002/CBIC.201900783.

(39) Rui, J.; Zhao, Q.; Huls, A. J.; Soler, J.; Paris, J. C.; Chen, Z.; Reshetnikov, V.; Yang, Y.; Guo, Y.; Garcia-Borràs, M.; Huang, X. Directed Evolution of Nonheme Iron Enzymes to Access Abiological Radical-Relay C(Sp3)−H Azidation. *Science (80-.).* **2022**, *376* (6595), 869–874. https://doi.org/10.1126/science.abj2830.

(40) Groendyke, B. J.; Abusalim, D. I.; Cook, S. P. Iron-Catalyzed, Fluoroamide-Directed C-H Fluorination. *J. Am. Chem. Soc.* **2016**, *138* (39), 12771–12774. https://doi.org/10.1021/jacs.6b08171.

(41) Matthews, M. L.; Chang, W.; Layne, andrew; Miles, L.; Krebs, C.; Martin Bollinger Jr, J. Direct Nitration and Azidation of Aliphatic Carbons by an Iron-Dependent Halogenase. *Nat. Chem. Biol.* **2014**, *10*. https://doi.org/10.1038/nCHeMBIO.1438.

(42) 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* (10), 1009–1016. https://doi.org/10.1038/s41589-019-0355-x.

(43) Kim, C. Y.; Mitchell, A. J.; Glinkerman, C. M.; Li, F.-S.; Pluskal, T.; Weng, J.-K. The Chloroalkaloid (−)-Acutumine Is Biosynthesized via a Fe(II)-and 2-Oxoglutarate-Dependent Halogenase in Menispermaceae Plants. https://doi.org/10.1038/s41467-020-15777-w.

(44) Büchler, J.; Honda Malca, S.; Patsch, D.; Voss, M.; Turner, N. J.; Bornscheuer, U. T.; Allemann, O.; Le Chapelain, C.; Lumbroso, A.; Loiseleur, O.; Buller, R. Algorithm-Aided Engineering of Aliphatic Halogenase WelO5* for the Asymmetric Late-Stage Functionalization of Soraphens. https://doi.org/10.1038/s41467-022-27999-1.

(45) Bräse, S.; Gil, C.; Knepper, K.; Zimmermann, V. Organic Azides: An Exploding Diversity of a Unique Class of Compounds. *Angew. Chemie Int. Ed.* **2005**, *44* (33), 5188–5240. https://doi.org/10.1002/ANIE.200400657.

(46) Hennessy, E. T.; Betley, T. A. Complex N-Heterocycle Synthesis via Iron-Catalyzed, Direct C-H Bond Amination. *Science (80-.).* **2013**, *340* (6132), 591–595. https://doi.org/10.1126/science.1233701.

(47) Lallana, E.; Riguera, R.; Fernandez-Megia, E. Reliable and Efficient Procedures for the Conjugation of Biomolecules through Huisgen Azide–Alkyne Cycloadditions. *Angew. Chemie Int. Ed.* **2011**, *50* (38), 8794–8804. https://doi.org/10.1002/ANIE.201101019.

(48) Gomez, C. A.; Mondal, D.; Du, Q.; Chan, N.; Lewis, J. C. Directed Evolution of an Iron(II)- and α-Ketoglutarate-Dependent Dioxygenase for Site-Selective Azidation of Unactivated Aliphatic C−H Bonds**. *Angew. Chemie Int. Ed.* **2023**, *62* (15), e202301370. https://doi.org/10.1002/ANIE.202301370.

(49) Qin, H. M.; Miyakawa, T.; Jia, M. Z.; Nakamura, A.; Ohtsuka, J.; Xue, Y. L.; Kawashima, T.; Kasahara, T.; Hibi, M.; Ogawa, J.; Tanokura, M. Crystal Structure of a Novel N-Substituted L-Amino Acid Dioxygenase from Burkholderia Ambifaria AMMD. *PLoS One* **2013**, *8* (5), e63996. https://doi.org/10.1371/JOURNAL.PONE.0063996.

(50) You, C.; Percival Zhang, Y. H. Easy Preparation of a Large-Size Random Gene Mutagenesis Library in Escherichia Coli. *Anal. Biochem.* **2012**, *428* (1), 7–12. https://doi.org/10.1016/j.ab.2012.05.022.

(51) Qin, H. M.; Miyakawa, T.; Nakamura, A.; Hibi, M.; Ogawa, J.; Tanokura, M. Structural Optimization of SadA, an Fe(II)- and α-Ketoglutarate- Dependent Dioxygenase Targeting Biocatalytic Synthesis of N-Succinyl-L-Threo-3, 4-Dimethoxyphenylserine. *Biochem. Biophys. Res. Commun.* **2014**, *450* (4), 1458–1461. https://doi.org/10.1016/j.bbrc.2014.07.008.

(52) 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.* **2009**, *106* (42), 17723–17728. https://doi.org/10.1073/pnas.0909649106.

(53) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nat. 2021 5967873* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.

(54) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; De Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: Homology Modelling of Protein Structures and Complexes. *Nucleic Acids Res.* **2018**, *46*. https://doi.org/10.1093/nar/gky427.

(55) Dimaio, F.; Leaver-Fay, A.; Bradley, P.; Baker, D.; André, I. Modeling Symmetric Macromolecular Structures in Rosetta3. *PLoS One* **2011**, *6* (6), 20450. https://doi.org/10.1371/journal.pone.0020450.

(56) Mehmood, R.; Qi, H. W.; Steeves, A. H.; Kulik, H. J. The Protein's Role in Substrate

Positioning and Reactivity for Biosynthetic Enzyme Complexes: The Case of SyrB2/SyrB1. *ACS Catal.* **2019**, *9* (6), 4930–4943. https://doi.org/10.1021/acscatal.9b00865.

(57) Hibi, M.; Kasahara, T.; Kawashima, T.; Yajima, H.; Kozono, S.; Smirnov, S. V.; Kodera, T.; Sugiyama, M.; Shimizu, S.; Yokozeki, K.; Ogawa, J. Multi-Enzymatic Synthesis of Optically Pure β-Hydroxy α-Amino Acids. *Adv. Synth. Catal.* **2015**, *357* (4), 767–774. https://doi.org/10.1002/adsc.201400672.

(58) Chan, N. H.; Gomez, C. A.; Vennelakanti, V.; Du, Q.; Kulik, H. J.; Lewis, J. C. Non-Native Anionic Ligand Binding and Reactivity in Engineered Variants of the Fe(II)- and α-Ketoglutarate-Dependent Oxygenase, SadA. *Inorg. Chem.* **2022**, *61* (36), 14477–14485. https://doi.org/10.1021/acs.inorgchem.2c02872.

(59) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning: A Laboratory Manual, 2nd Edn.*; Cold Spring Laboratory Press, 1989.

(60) You, C.; Percival Zhang, Y. H. Easy Preparation of a Large-Size Random Gene Mutagenesis Library in Escherichia Coli. *Anal. Biochem.* **2012**, *428* (1), 7–12. https://doi.org/10.1016/J.AB.2012.05.022.

(61) Liu, H.; Naismith, J. H. An Efficient One-Step Site-Directed Deletion, Insertion, Single and Multiple-Site Plasmid Mutagenesis Protocol. *BMC Biotechnol.* **2008**, *8* (1), 91. https://doi.org/10.1186/1472-6750-8-91.

(62) Christa, P.; Dunkel, A.; Krauss, A.; Stark, T. D.; Dawid, C.; Hofmann, T. Discovery and Identification of Tastants and Taste-Modulating N-Acyl Amino Acid Derivatives in Traditional Korean Fermented Dish Kimchi Using a Sensomics Approach. *Cite This J. Agric. Food Chem* **2022**, *2022*, 7500–7514. https://doi.org/10.1021/acs.jafc.2c02623.

Chapter 4: Hydroxylase activity of evolved SadA and SadX variants

4.1 Introduction

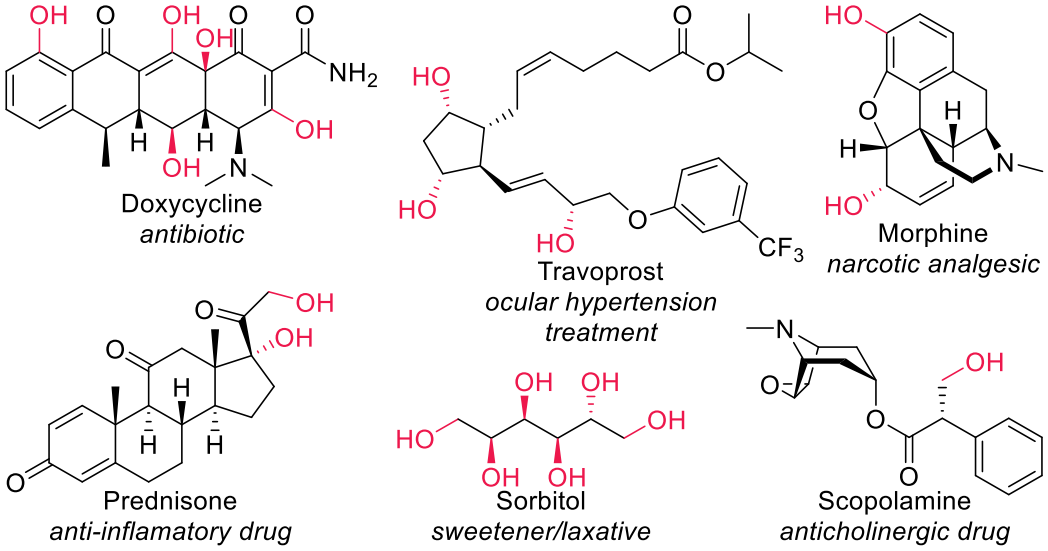
4.1.1 The relevance of C–H hydroxylation

Hydroxyl substitution of aliphatic moieties in synthetic drug candidate small molecules is relatively rare. A 2011 study that surveyed the most common drug candidate functional groups revealed that around 10% of the analyzed compounds contain an aliphatic alcohol.[1] This frequency is distinct from hydroxyl-containing natural products. A more recent study based on two different databases highlighted the difference in the frequency of occurrence of hydroxyl groups between marketed drugs that were synthetic versus those derived from natural products.[2] According to the ChEMBL database, 69% of all marketed natural-product-derived drugs contained at least one hydroxyl functional group, while only 23% of synthetic marketed drugs fulfill this criterion.[3] Likewise, according to a review evaluating drugs approved between 1981 and 2014, 54% of natural-product-derived compounds contained hydroxyl groups, whereas this was the case for only 16% of synthetic drugs.[4] These differences highlight the challenges synthetic chemists face when attempting to insert hydroxyl moieties in organic scaffolds without poor pharmacodynamic and pharmacokinetic properties.

Despite their relatively infrequent appearance in synthetic drugs, hydroxyl functional groups can still have a large impact on the behavior of bioactive molecules (Figure 4.1). Their polarized O–H bond allows them to act as hydrogen bond donors and acceptors. Drug-receptor binding can be highly dependent on such interactions and the specific orientation of the hydroxyl

groups involved.[5,6] Hydroxyl groups can also increase the aqueous solubility and polarity of molecules without significantly disrupting membrane permeability since they do not change the charge of the compound at physiological pH. Finally, alcohol moieties have been used as water molecule replacements to inhibit enzyme activity, as shown in particular for aspartic proteases. These enzymes use a catalytically active water molecule that can be replaced by a hydroxyl group in a transition-state analogue, functionally inhibiting proteolytic activity.

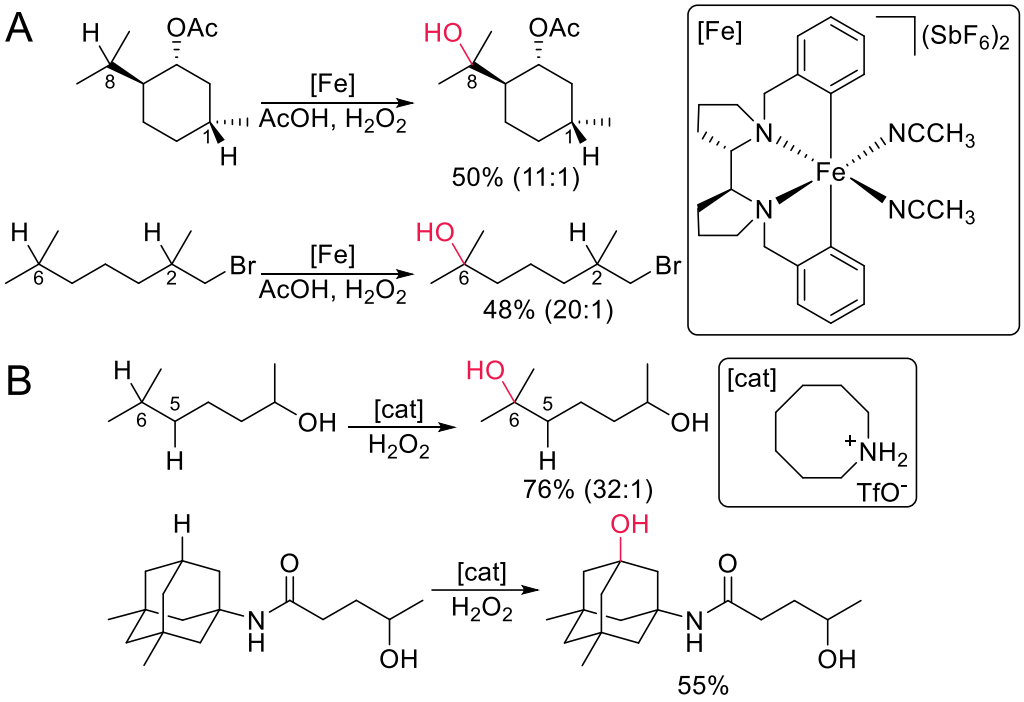
Figure 4.1: Representative hydroxyl-containing compounds and their activities



Earlier C–H functionalization strategies to introduce hydroxyl groups suffered from significant drawbacks such as the requirement of large excess of substrates, low yields, inconsistent site selectivity, and the need for stoichiometric amounts of oxidants.[7,8] Later approaches involving non-heme iron catalysts and other metalloenzyme-inspired complexes were able to achieve more predictable selectivities based on the electronic and steric properties of the C–H bonds to be functionalized (Scheme 4.1A).[9,10] Additionally, organocatalysts based on heterocycles, iminium salts, amines, and other groups, have successfully been developed to exert high levels of predictable selectivity on hydroxylation reactions (Scheme 4.1B).[11–13] Although

the ability to provide consistent patterns of selectivity greatly improves the applicability of these catalysts, this selectivity is still largely defined by the characteristics of the substrate.

Scheme 4.1: Representative small molecule catalyzed C–H hydroxylation reactions



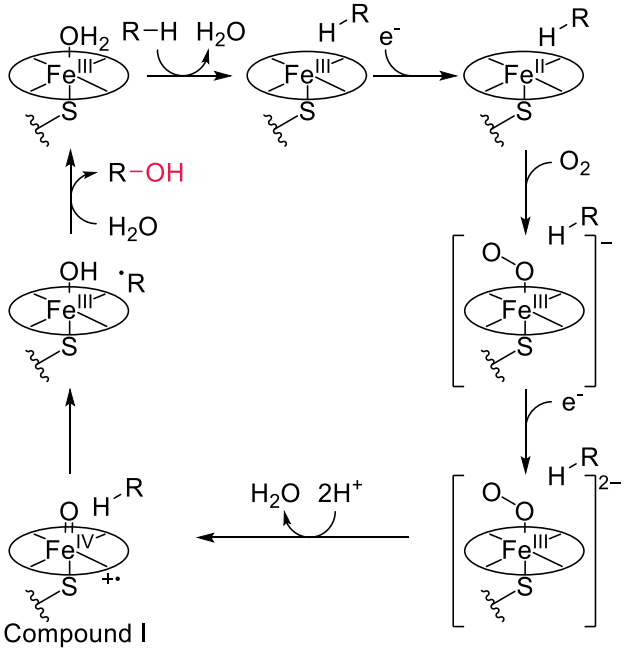
[a] Functionalized positions are numbered; percentages correspond to reaction yields; ratios in parenthesis correspond to site selectivity, with the major product being the one shown. A) Site selective hydroxylation of tertiary C–H bonds catalyzed by a non-heme iron catalyst.[9] Selectivity is determined by the steric (top) and electronic (bottom) properties of the C–H bonds. B) C–H hydroxylation reactions with high site and chemoselectivities catalyzed by an amine organocatalyst.[13]

4.1.2 Enzymatic C–H hydroxylation

A variety of enzyme classes have evolved in nature to selectively hydroxylate C–H bonds as part of natural metabolism.[14] This brief introduction intends to provide a general overview of the most common hydroxylating enzymes and is adapted from a submitted review article focused on non-native site-selective enzyme catalysis.[15] The most developed class of hydroxylating enzymes in biocatalysis is the family of cytochrome P450 monooxygenases (P450s).[16–19] These enzymes are characterized by a heme cofactor with a cysteine thiolate ligand to the iron center on the proximal face, leaving the distal face free for O_2 activation. P450-catalyzed hydroxylation

requires reduction of the heme cofactor by electron transfer from a NAD(P)H cofactor through a partner P450 reductase or ferredoxin (Scheme 4.2).[20] Molecular oxygen then binds to the heme cofactor and, through a series of intermediates, results in a highly reactive Fe(IV)-oxo intermediate named "compound I". Similar to the previously described hydroxylation mechanism of FeDOs (Scheme 1.2), this Fe(IV)-oxo intermediate abstracts a hydrogen atom from the substrate resulting in a Fe(III)-hydroxo intermediate and a radical substrate. The final step of the mechanism is a radical rebound step to generate the hydroxylated product.

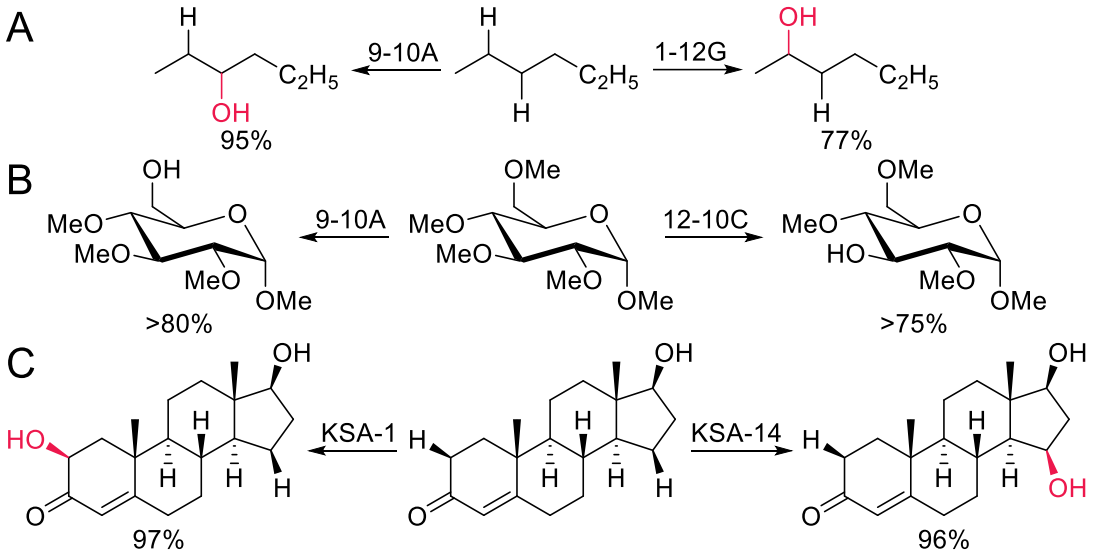
Scheme 4.2: Mechanism of P450-catalyzed C–H hydroxylation



Although several P450s have been used for site-selective hydroxylation, the long-chain fatty acid hydroxylase from *Bacillus megaterium*, $P450_{BM3}$, has been extensively studied since its fused reductase domain removes the need to provide a reductase partner.[21] This fact not only simplifies the biocatalytic system, but also translates into a more straightforward protein engineering procedure. $P450_{BM3}$ has been engineered to show selective hydroxylation of a wide variety of compounds significantly different from its native substrates. Early on, the activity of

P450$_{BM3}$ on *n*-alkanes was established and variants of this enzyme were engineered via directed evolution to achieve high selectivity (Scheme 4.3A).[22] This series of evolved biocatalysts showed a high versatility in terms of substrate scope. One example was the demethylation of permethylated monosaccharides which proceeded via hydroxylation of the methyl functional groups (Scheme 4.3B).[23] This reaction was demonstrated on a series of monosaccharides and enabled further functionalization of the single deprotected hydroxyl groups. Finally, steroids have also been used as substrates for P450$_{BM3}$-catalyzed hydroxylation, with one example engineering variants with high site selectivity for different C–H bonds in the steroid testosterone (Scheme 4.3C).[24]

Scheme 4.3: Representative site selective P450$_{BM3}$-catalyzed hydroxylation reactions[a]

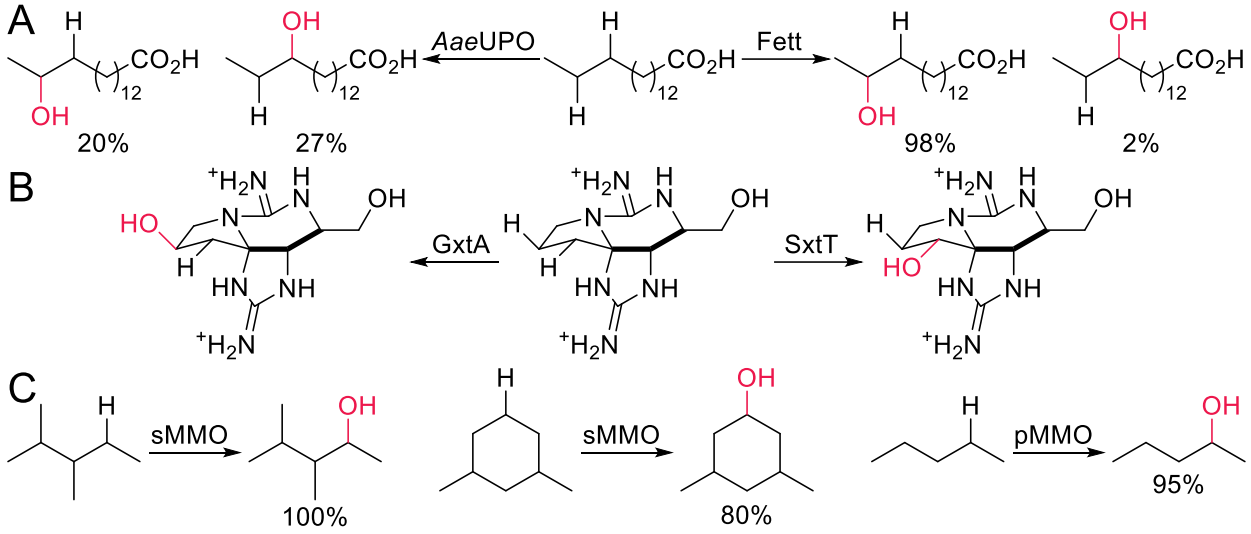


[a] Only major products are shown; site selectivity is expressed as the percentage of the major product in the product pool. A) Hydroxylation of *n*-hexane by P450$_{BM3}$ variants.[22] B) Demethylation via hydroxylation of protected monosaccharides.[23] C) Site selective hydroxylation of testosterone by P450$_{BM3}$ variants.[24]

Other types of site-selective hydroxylating metalloenzymes, besides P450s and FeDOs, include heme-dependent peroxidases, Rieske non-heme iron-dependent oxygenases, and bacterial multicomponent monooxygenases. Heme peroxidases are able to oxidize their substrates through the reduction of H_2O_2 by the heme prosthetic group.[25,26] The fungal unspecific

peroxidase (UPO) from *Agrocybe aegerita* (*Aae*UPO) has been engineered to enable high levels of selectivity in the hydroxylation of aromatic substrates, such as naphthol and a series of flavonoids.[27,28] Additionally, *Aae*UPO has been engineered to act on aliphatic C–H bonds, with a recent example of an evolved variant selectively hydroxylating subterminal positions of fatty acid substrates (Scheme 4.4A).[29]

Scheme 4.4: Representative site-selective hydroxylations catalyzed by non-P450 and non-FeDO metalloenzymes



[a] Site selectivity is expressed as the percentage of the major product in the product pool. A) Hydroxylation of palmitic acid by *Aae*UPO variants; remaining products are ω-hydroxylated palmitic acid and overoxidation products.[29] B) Site-selective monooxygenation of a saxitoxin-derived compound by Rieske monooxygenases SxtT and GxtA.[30] C) Hydroxylation of alkanes by methane monooxygenases sMMO and pMMO from *Methylococcus capsulatus* (Bath).[31–33]

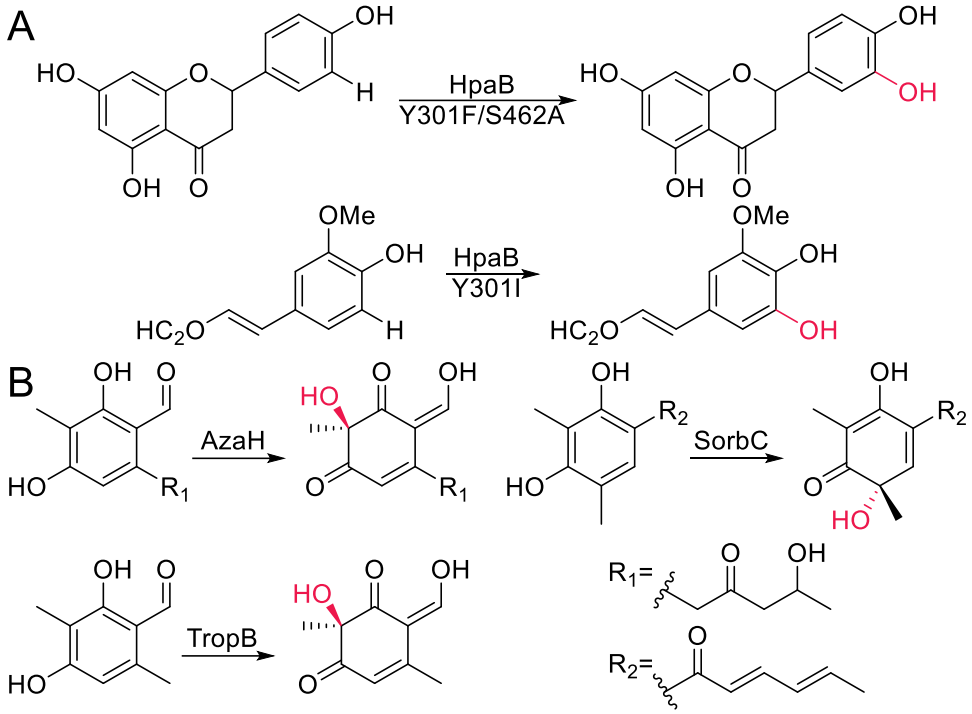
Rieske non-heme iron-dependent oxygenases are bacterial enzymes containing an iron-sulfur cluster and a catalytic non-heme iron center.[34] A reductase provides electrons through the iron-sulfur cluster to the iron center, which then forms a high-valent iron intermediate by reduction of O_2, allowing for substrate oxidation via radical intermediates. These enzymes are natively involved in the degradation of aromatic compounds via *cis*-dihydroxylation reactions. Not only have Rieske oxygenases been used for oxidation of alkenes and aromatic compounds,[35]

but they have also been used to hydroxylate non-native saxitoxin-derived tricyclic natural products (Scheme 4.4B).[30]

Bacterial multicomponent monooxygenases (BMMs) consist of three or four components: a carboxylate-bridged diiron(III)-containing hydroxylase, a NADH reductase, an effector protein to couple electron consumption and substrate oxidation, and a Rieske-type iron-sulfur ferredoxin that is not present in all BMMs.[36] As with Rieske oxygenases, BMMs have been characterized and engineered to selectively hydroxylate aromatic and aliphatic substrates.[37] Methane monooxygenases (MMOs) from methanotrophic bacteria belong to his family of enzymes. Both soluble (sMMOs) and particulate MMOs (pMMOs) have been characterized to site-selectively hydroxylate aliphatic C–H bonds (Scheme 4.4C).[31–33]

Finally, flavin-dependent monooxygenases (FMOs) are a family of enzymes that catalyze hydroxylation of aromatic C–H bonds, among other reactions.[38] Single-component FMOs contain a tightly bound FAD prosthetic group, while two-component FMOs require a reductase partner to provide the reduced flavin as a co-substrate. As opposed to most of the enzymes mentioned above, FMOs carry out oxyfunctionalization reactions via electrophilic aromatic substitution between a highly reactive hydroperoxyflavin intermediate and electron rich aromatic substrates. The 4-hydroxyphenylacetate 3-hydroxylases are two-component FMOs that natively oxidize 4-hydroxyphenylacetate to 3,4-dihydroxyphenylacetate.[39] HpaB from *E. coli* belongs to this subgroup, and it has been engineered to accept larger non-native substrates (Scheme 4.5A).[40] Other FMOs natively catalyze the oxidative dearomatization of resorcinol derivatives via C–H hydroxylation.[41–43] A series of these enzymes, AzaH, SorbC, and TropB, were characterized to have complementary substrate scope and orthogonal site selectivity towards a panel of resorcinol derivatives (Scheme 4.5B).[44]

Scheme 4.5: Representative site-selective FMO-catalyzed hydroxylation reactions[a]

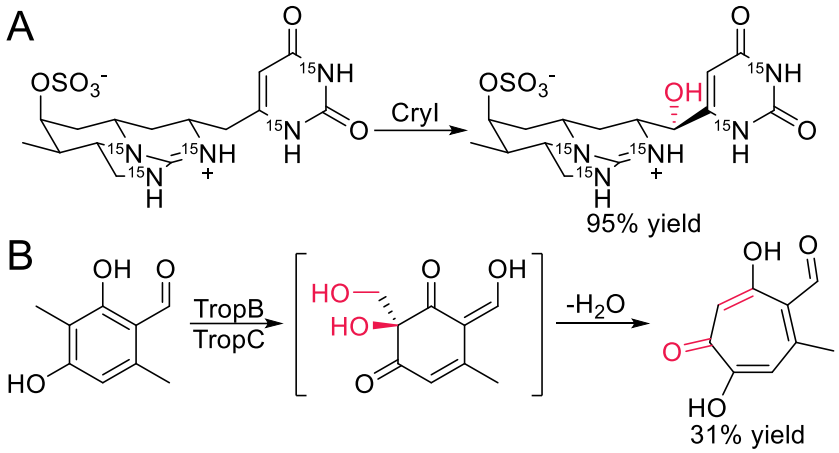


A) Site-selective hydroxylation of non-native substrates by engineered HpaB variants.[40] B) Dearomatization reactions with distinct site-selectivities.[44]

4.1.3 Recent applications of hydroxylating FeDOs

Hydroxylating FeDOs have successfully been used in drug discovery and the synthesis of commercial noncanonical amino acids.[45–47] The FeDO CryI from *Cylindrospermopsis raciborskii* is responsible for the final step in the biosynthesis of the biotoxin cylindrospermopsin. In order to improve cylindrospermopsin detection in environmental samples via LC-MS via the use of a isotopically labeled standard, the total synthesis of ^{15}N-labeled cylindrospermopsin was carried out chemoenzymatically.[48] CyrI was used to achieve late-stage functionalization of the labeled-precursor to obtain the final product (Scheme 4.6A). Additionally, the previously mentioned FMO TropB was used in conjunction with the FeDO TropC in one-pot two-enzyme cascade to synthesize stipitatic aldehyde (Scheme 4.6B).[44]

Scheme 4.6: Hydroxylating FeDOs used in the synthesis of natural products

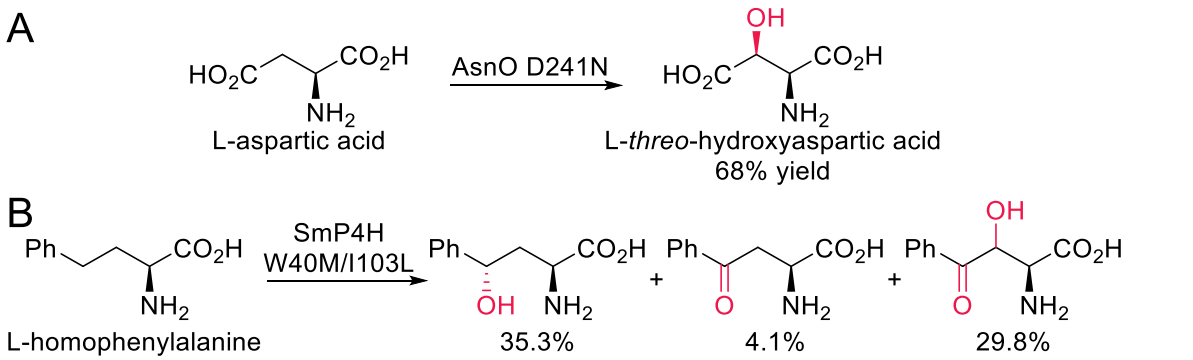


A) Late-stage hydroxylation by CryI in the chemoenzymatic total synthesis of ^{15}N-labeled cylindrospermopsin.[48] B) Synthesis of stipitatic aldehyde via a two-enzyme cascade using FMO TropB and FeDO TropC.[44]

Noncanonical amino acids, such as hydroxylated amino acids, are important synthetic building blocks, granting a functional handle for further modifications or acting as key precursors of biologically relevant molecules.[49] Besides the examples mentioned in chapter 1, hydroxylating FeDOs that act on amino acids have been engineered to expand their substrate scope. In most instances, however, these improvements are limited to close structural congeners. Asparagine oxygenase (AsnO) from *Streptomyces coelicolor*, which natively transforms L-asparagine to L-*threo*-hydroxyasparagine, was rationally engineered to accept aspartic acid as a substrate.[50] Structural analysis of AsnO revealed an interaction between the δ-amide and residue Asp_{241}. Variant AsnO D241N was able to accept L-aspartic acid and catalyze the formation of L-*threo*-hydroxyaspartic acid, although no other substrates were successfully hydroxylated (Scheme 4.7A). More recently, a proline hydroxylase from *Sinorhizobium meliloti* (SmP4H) was engineered for the hydroxylation of non-native substrate L-homophenylalanine.[51] SmP4H was identified from a panel of 36 wild-type FeDOs to have activity on L-homophenylalanine. Structural analysis using a homology model yielded several residues with potential interactions with the substrate. Site-saturation mutagenesis resulted in variant SmP4H W40M/I103L, with a

35.3% conversion to γ-hydroxylated L-homophenylalanine and a ~300-fold improvement in catalytic efficiency for this transformation when compared to wild-type SmP4H (Scheme 4.7B). It is worth noting that other oxidation products, including a γ-ketone and a β-hydroxy-γ-ketone, were observed in high amounts. Regardless, this example showed the potential of expanding substrate scope of FeDOs to compounds outside of structural congeners of the native substrate.

Scheme 4.7: Representative engineered FeDO for noncanonical amino acid synthesis



A) Engineered asparagine oxygenase AsnO catalyzes aspartic acid hydroxylation.[50] B) Proline hydroxylase SmP4H was engineered to oxidize non-native substrate homophenylalanine; percentages are LC-MS conversion values.[51]

Expanding the substrate scope of hydroxylating biocatalysts and identifying enzymes with novel site selectivities are crucial to further increasing the toolbox of hydroxylation reactions available to synthetic chemists. SadX and its evolved variants showed not only varied substrate scope of amino acid- and amine-derivatives, but also altered site selectivity for both native hydroxylation and non-native functionalizations. Efforts on characterizing this family of biocatalysts and developing improved variants for hydroxylation reactions were carried out, and the results up until this point will be discussed in this chapter.

Authorship

The synthesis of substrate **1a** was carried out by Dr. Natalie H. Chan. Substrates **2a**, **6a**, **7a**, **8a**, **9a**, **10a**, **11a**, **12a**, **13a**, **14a**, **15a**, and **16a** was performed by Dr. Dibyendu Mondal. The introduction of the G157D mutation into variants 1-VH, 2-L, 2-D, 3-VRL, and 4-IC discussed in

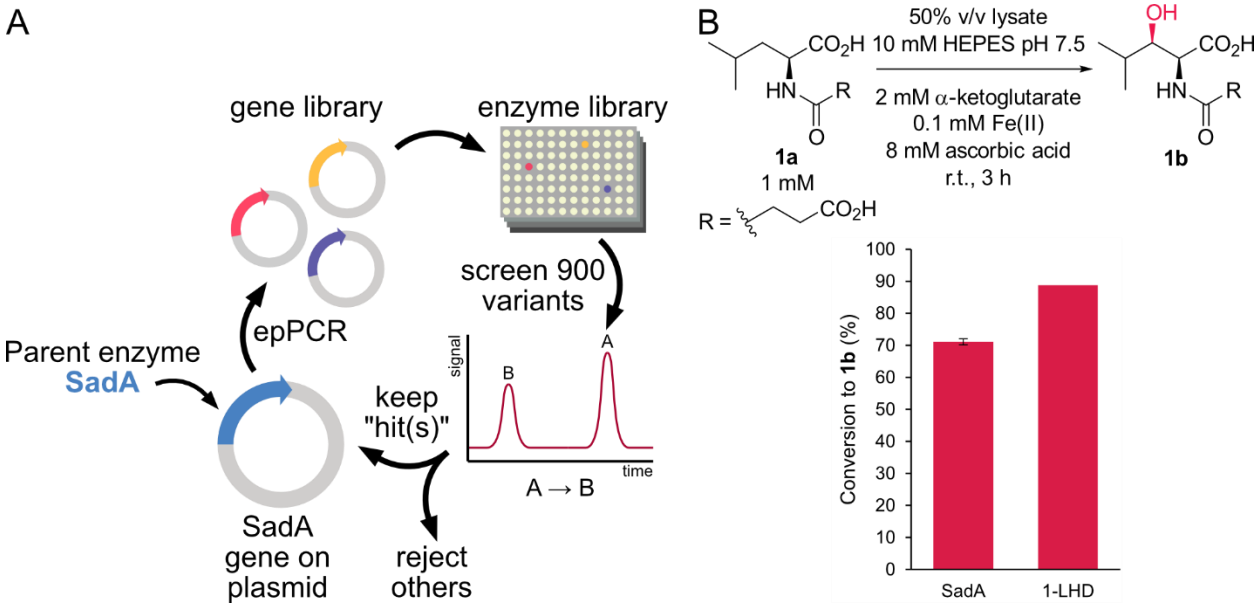
section 4.2.2, as well as the experiments illustrated in Figure 4.5, were conducted with the aid of undergraduate student Harumi Shimano.

4.2 Results and discussion

4.2.1 Directed evolution of SadA for improved hydroxylase activity

SadA variants with improved activity toward non-native transformations such as chlorination and azidation was achieved via directed evolution as discussed in chapter 3.[52] Directed evolution of wild-type SadA for improvement of its native hydroxylase activity had not been attempted before. Beneficial mutations towards hydroxylation of the model substrate *N*-succinyl-L-leucine (**1a**) could translate to other substrate and improve the non-native activities explored so far. With this goal in mind, a round of directed evolution was carried out using wild-type SadA as the starting point, and epPCR as the mutagenic step. Conditions were previously determined for an epPCR experiment with a mutation rate of 3-4 non-silent mutations per gene (Figure 3.3). This library of 900 variants was screened in cell lysate and using LC-MS to analyze the resulting bioconversions (Figure 4.2A).

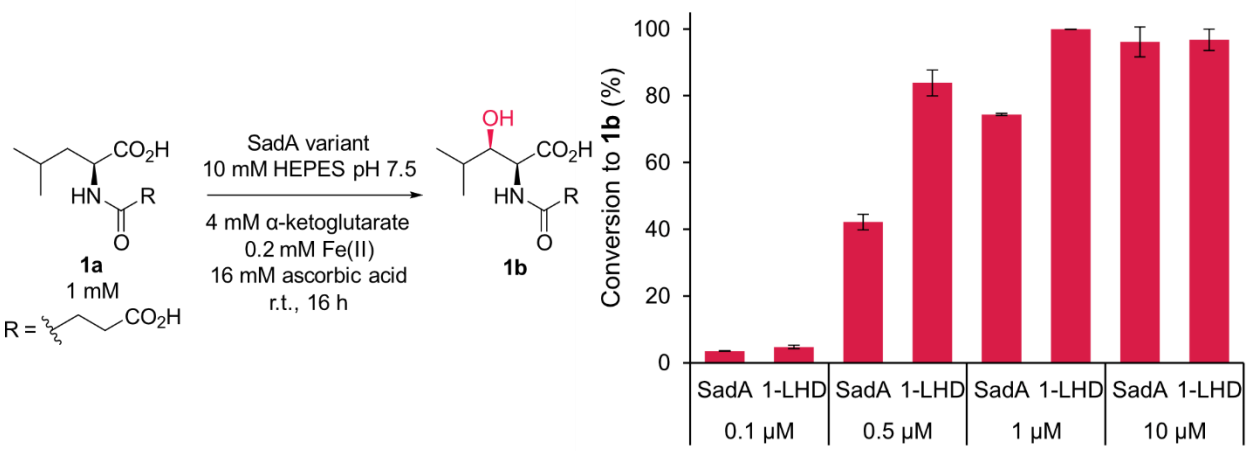
Figure 4.2: Overview of a single round of evolving SadA for native hydroxylation



A) General directed evolution strategy for improving hydroxylase activity in SadA. A single round of evolution was carried out. B) Screening results for the conversion of **1a** to **1b** for a potential "hits". Conversion to products calculated with LC-MS extracted ion chromatogram peak areas of products and substrate. Data for parent SadA corresponds to our replicate lysates obtained from the same plate as 1-LHD data for 1-LHD is a single measurement. Error bars represent the corresponding standard deviation.

One potential hit was identified from this round of evolution, showing higher conversion to β-hydroxylated product **1b** than wild-type SadA under screening conditions (Figure 4.2B). Once identified, this variant was isolated and characterized as the triple mutant SadA P61L/Q85H/E147D (1-LHD hereafter). Whereas Glu_{85} is a solvent-facing residue on the surface of the protein according to the reported crystal structure of SadA, residues Pro_{61} and Glu_{147} are in loops comprising Ser_{59}-Ser_{75} and Glu_{147}-Ala_{153}, respectively, both of which are missing in the crystal structure of SadA.[53] 1-LHD was purified, and its activity was evaluated in bioconversions with different enzyme loadings (Figure 4.3). A significant improvement was observed when using 0.5 µM and 1 µM of enzyme, with a ~2-fold improvement in total turnovers under the 0.5 µM condition, with ~840 TTN for SadA and ~1680 TTN for 1-LHD.

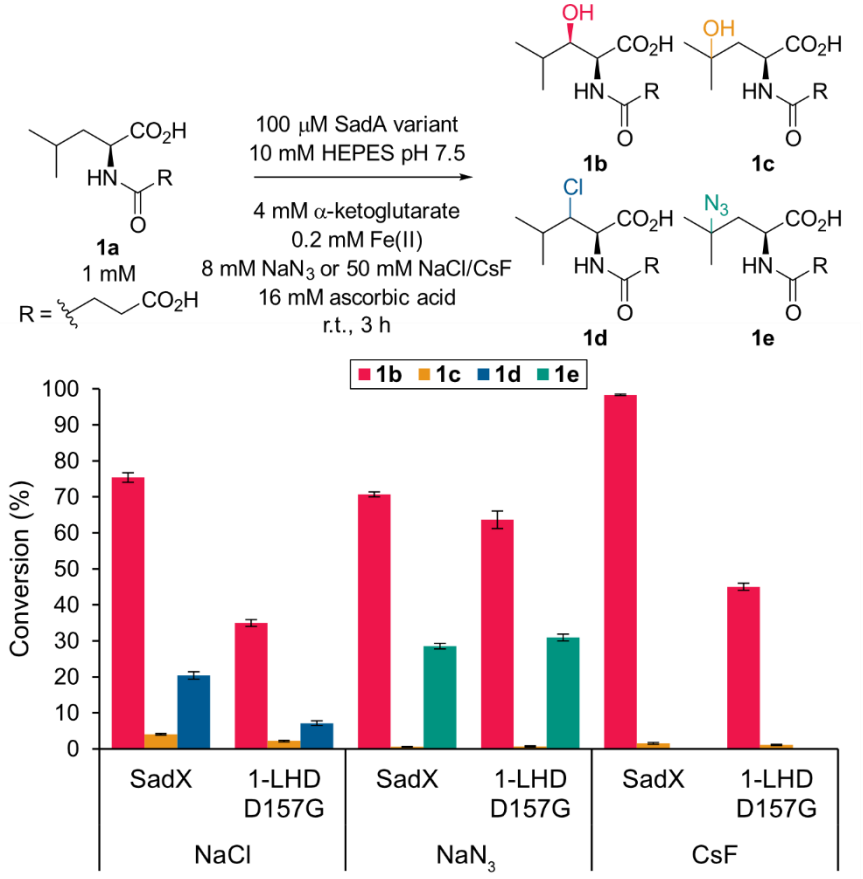
Figure 4.3: Hydroxylation with evolved variant 1-LHD at different enzyme concentrations[a]



[a] Conversion of **1a** to hydroxylated product **1b** by SadA and 1-LHD at different concentrations of enzyme. Conversion values were calculated with extracted ion chromatogram peak areas of products and the substrate from LC-MS experiments. Each data point corresponds to the average of three replicates and error bars represent standard deviations.

To evaluate if the improvement in hydroxylation observed in 1-LHD translated to non-native chlorination and azidation reactions, the facial triad mutation D157G was introduced into 1-LHD. Compared to SadX under standard chlorination and azidation conditions, these new variants showed no improvements (Figure 4.4). In the presence of NaCl, the overall conversion of **1a** decreased to ~45% with 1-LHD. In the presence of NaN_3 all conversion values were approximately equal between SadX and 1-LHD, with a slight decrease in hydroxylation. These variants were also studied in the presence of CsF, previously demonstrated to rescue native hydroxylase activity in SadA facial triad mutants (Figure 2.8). In this case, the conversion to **1b** with 1-LHD decreased to ~45% as was observed in the presence of NaCl (Figure 4.4). These results highlight once more that the identity of the anionic ligand present in the active site of SadX variants affects reactivity and selectivity. Regardless, the mutations accumulated in one round of directed evolution of SadA for its native activity did not present a beneficial effect on non-native activities in the presence of the facial triad mutation G157D.

Figure 4.4: Activity of 1-LHD G157D variant in the presence of exogenous anions[a]



[a] Conversion of **1a** to products **1b**-**1e** by SadX and 1-LHD D157G. Reactions were set up under in the presence of NaCl, NaN_3, or CsF. Conversion values were calculated with extracted ion chromatogram peak areas of products and the substrate from LC-MS experiments. Each data point corresponds to the average of three replicates and error bars represent standard deviations.

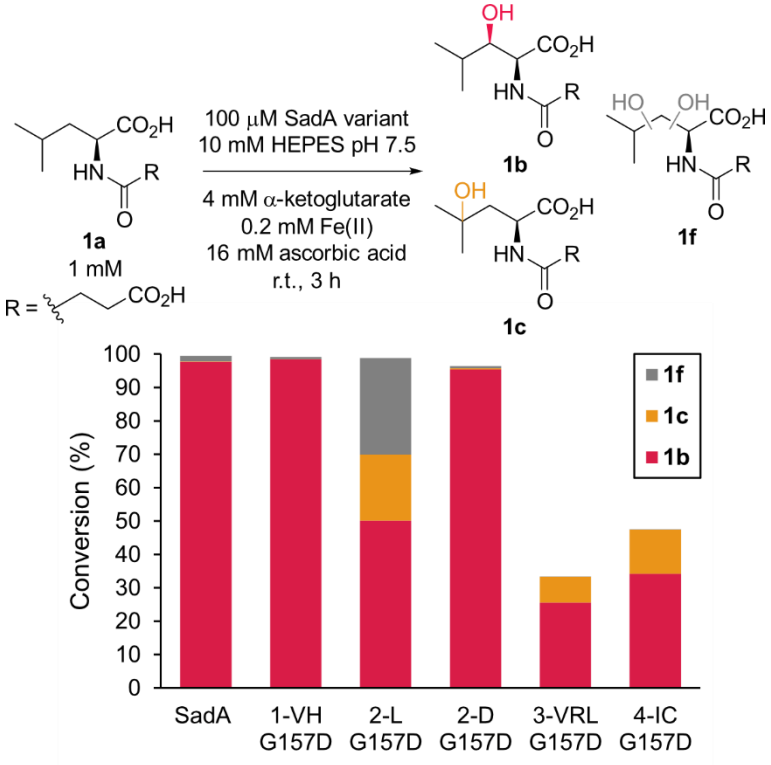
4.2.2 Hydroxylase activity of engineered SadX variants

When SadX was evolved for non-native chlorination and azidation reactions, a series of variants were generated that not only showed changes in chemoselectivity, but also distinct site selectivities for the native hydroxylation reaction (Figure 3.14). Additionally, it was shown that mutations that affected site selectivity, such as F152L, have different effects when introduced in wild-type SadA or the facial triad mutant SadX (Figure 2.10). The site selectivity of SadX variants was explored after introducing G157D mutations to revert the active site of these

enzymes to the wild-type hydroxylase sequence. Variants 1-VH, 2-L, 2-D, 3-VRL, and 4-IC were used to construct these new variants.

The new G157D variants were used to set up bioconversions using substrate **1a** (Figure 4.5). Three hydroxylated products were detected: the monohydroxylated products **1b** and **1c**, as well as a product with a *m/z* value corresponding to a dehydroxylated product **1f**. Although **1f** could be assumed to be the β,γ-dihydroxylated product, this compound was not further characterized. Production of **1f** was observed mostly with variant 2-L G157D, which also showed the highest production of **1c**, albeit with a lower site selectivity than the one observed for 2-L under azidation conditions. Both 1-VH and 2-D produced only **1b** in the presence of NaN_3, so it is not surprising that the corresponding variants 1-VH G157D and 2-D G157D also follow this trend. Finally, both 3-VRL and 4-IC showed very low conversions to hydroxylated products under azidation conditions, with **1c** being the major hydroxylated product. The G157D variants, however, showed the opposite site selectivity, favoring β-hydroxylation to produce **1b** over **1c**. The low conversion levels of 3-VRL G157D and 4-IC G157D can be rationalized since the corresponding azidases had high chemoselectivity towards rebound of the non-native ligand. This suggests that the active sites of 3-VRL G157D and 4-IC G157D are not ideal for the native functionalization of **1a**.

Figure 4.5: Hydroxylation with G157D reversions of azidase SadX variants[a]

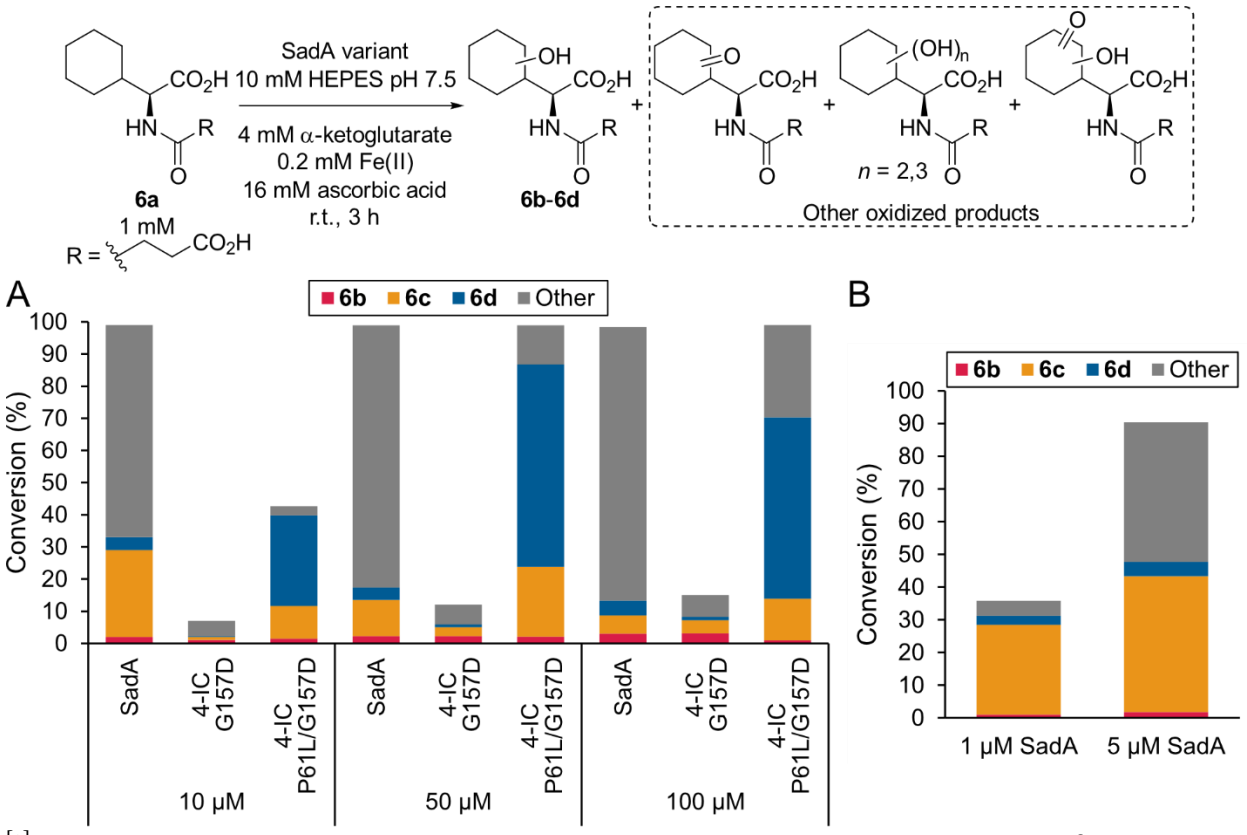


[a] Bioconversions of **1a** to hydroxylated products **1b**, **1c**, and **1f** by SadA and SadX variants with facial triad reversion G157D. Conversion values were calculated with extracted ion chromatogram peak areas of products and the substrate from LC-MS experiments. Each data point corresponds to a single measurement.

The substrate scope of this new panel of enzymes was also evaluated. First, the G157D variants were screened against substrate *N*-succinyl-L-cyclohexylglycine (**6a**) since preliminary data had shown a mixture of hydroxylated products when it was used as substrate for azidation reactions using SadX. Variants 2-L G157D and 4-IC G157D were selected for these preliminary studies, since they had shown distinct behavior when tested against substrate **1a** (Figure 4.5). Additionally, point mutations derived from improved hydroxylase 1-LHD (i.e., P61L, Q85H, and E147D) were individually introduced into 2-L G157D and 4-IC G157D, to evaluate any impact on conversion to hydroxylated products and site selectivity. Bioconversions of **6a** with wild-type SadA resulted in a mixture of oxidized products, including three monohydroxylated products (**6b**, **6c**, and **6d**), ketone-containing products, di- and tri-hydroxylated products, and products that

contained both a hydroxyl and a ketone functional groups (Figure 4.6). FeDOs have been reported to catalyze sequential oxidation of C–H bonds to produce aldehydes/ketones and even carboxylic acid, which would explain the presence of ketone-containing products.[54] SadA is able to access multiple C–H bonds in substrate **6a** increasing the mount of potential products with multiple functionalities.

Figure 4.6: Evaluating the selectivity of engineered SadA variants on *N*-succinyl-L-cyclohexylglycine[a]



[a] Conversion values were calculated with extracted ion chromatogram peak areas of products and the substrate from LC-MS experiments. Each data point corresponds to the average of three replicates. A) Conversion of **6a** to monohydroxylated products **6b**-**6d** and other oxidized products by wild-type SadA, 4-IC G157D, and 4-IC P61L/G157D under different enzyme concentration conditions. B) Bioconversions of **6a** with lower concentrations of wild-type SadA.

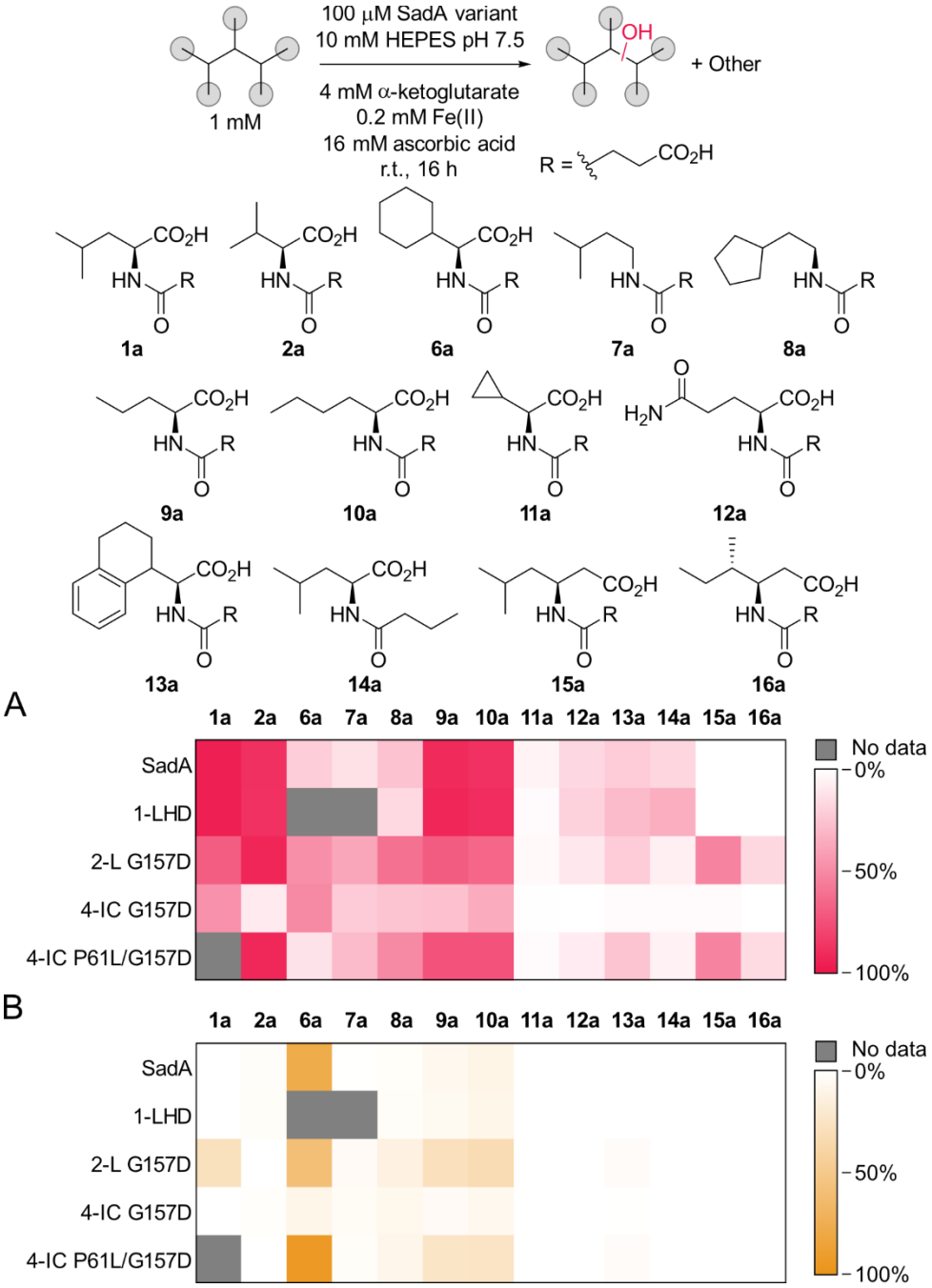
While 4-IC G157D showed more than a 6.5-fold decrease in overall conversion of **6a** under all enzyme concentration conditions explored, introduction of P61L was able to maintain conversion levels when using 50 μM of enzyme or more (Figure 4.6A). Remarkably, the main

product resulting from bioconversions with 4-IC P61L/G157D was a monohydroxylated product (**6d**) that was obtained with conversion values lower than 5% when using SadA. This could be attributed to a change in the site selectivity of the hydroxylation reaction. However, since ~65%-85% of **6a** was converted to multihydroxylated products and overoxidized compounds in the presence of SadA, it was difficult to determine if **6d** was also being the major product of SadA bioconversions but was then rapidly transformed into other products. Reactions with lower SadA concentrations were set up to minimize overoxidation and explore the product distribution of monohydroxylated products (Figure 4.6B). Using 1 μM SadA, only ~13% of the product pool corresponded to overoxidized products. Under these conditions ~76% of the product pool corresponded to monohydroxylated **6c**, while **6d** still remained a minor product. This experiment suggests that 4-IC P61L/G157D is able to hydroxylate **6a** with a novel site selectivity.

A set of *N*-succinylated amino acids and amines was used as substrates for these engineered SadA variants (Figure 4.7). The screening was focused on the conversion of substrates to monohydroxylated products, with the goal of identifying substrate-enzyme pairs that potentially resulted in novel site selectivities (Figure 4.7A). However, since monohydroxylation could be masked by further oxidation of the hydroxylated products, the conversion to these other products was also analyzed (Figure 4.7B). Most reactions showed low conversion to overoxidized products, with conversions of ~30% or lower except for substrate **6a**. Substrates **11a**, **12a**, and **13a** showed relatively low conversion values, suggesting that binding of these molecules was not conducive to C–H abstraction and hydroxyl rebound. The amide group in *N*-succinyl-L-glutamine (**12a**) could be disrupting interactions in the substrate pocket, which so far has been shown to accept aliphatic alkyl moieties. This could also been the case for **13a**, although SadA has been reported to have activity on the aromatic substrates.[55] Substrate **14** was

poorly accepted by SadA variants, highlighting the importance of the carboxylic acid of the succinyl moiety for binding.

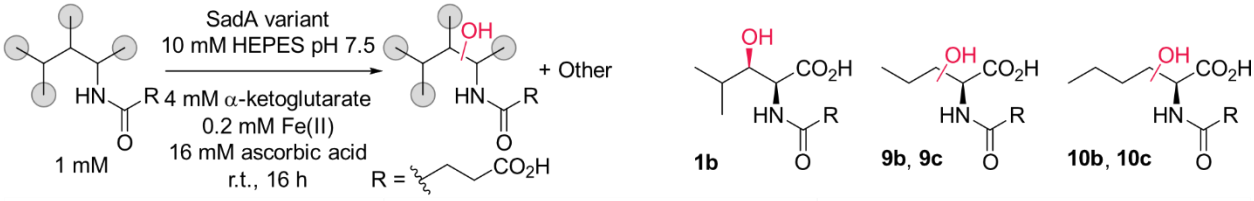
Figure 4.7: Hydroxylation substrate screen using engineered SadA variants[a]

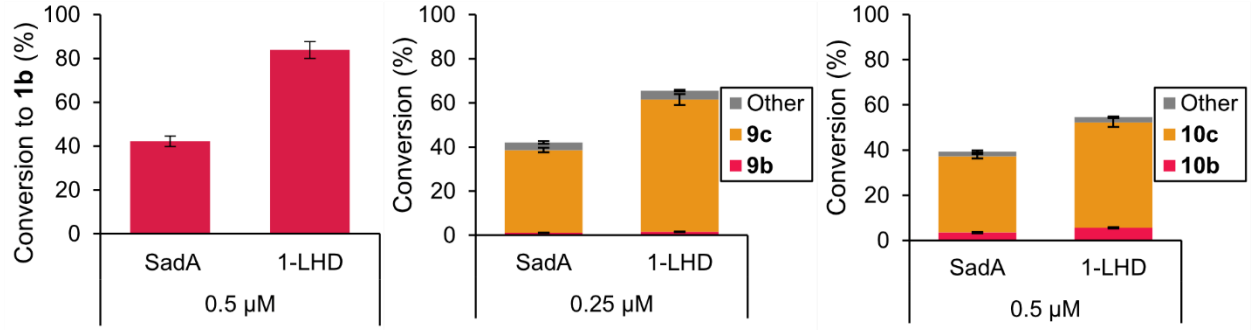


[a] Reaction products included mono-, di-, and trihydroxylated products, as well as overoxidized products with multiple hydroxyl and ketone functional groups. Conversion was calculated with extracted ion chromatogram peak areas from LC-MS experiments. Each data point corresponds to a single measurement. A) Percent conversion to monohydroxylated products by engineered SadA variants. B) Percent conversion to oxidized products other than monohydroxylated compounds.

Several substrate/enzyme pairs showed interesting results that granted more detailed observation and exploration of reaction conditions. Evolved variant 1-LHD was characterized as a better hydroxylase of substrate **1a**, and this effect was also observed with *N*-succinyl-L-norvaline (**9a**) and *N*-succinyl-L-norleucine (**10a**, Figure 4.8). While not as high as the ~2-fold improvement observed for substrate **1a**, hydroxylation of substrates **9a** and **10a** still occurred with higher TTNs with variant 1-LHD, with ~1.6- and ~1.4-fold improvements, respectively. In terms of selectivity, reactions with both **9a** and **10a** are functionalized almost exclusively at a single site. A second monohydroxylated product is observed at very low conversion values for both substrates, and overoxidation does not surpass 5% conversion in either case. Further product characterization is required to conclude if the improvements in hydroxylation observed with 1-LHD are specific to functionalization of the β-position of these substrates.

Figure 4.8: Improved hydroxylation in bioconversions with engineered 1-LHD[a]





Variant	TTN
SadA	844
1-LHD	1677

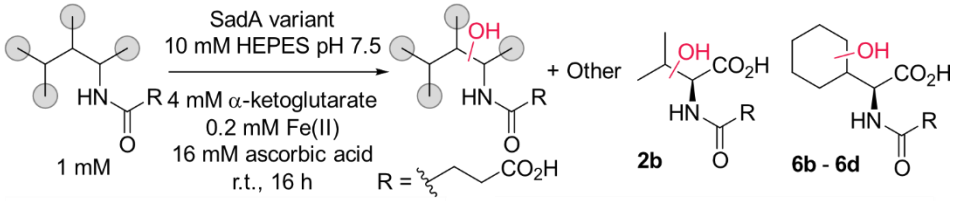
Variant	TTN (9c)
SadA	1497
1-LHD	2396

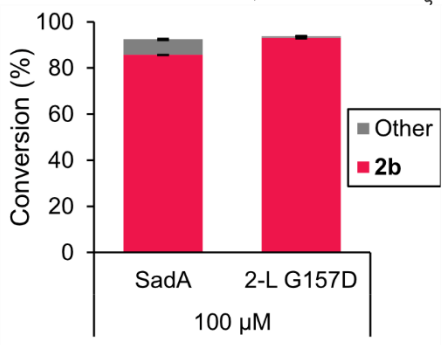
Variant	TTN (10c)
SadA	675
1-LHD	933

[a] Conversion to monohydroxylated products by SadA and 1-LHD using either 0.25 µM or 0.5 µM of biocatalyst. Other products include ketones, dihydroxylated products, and compounds containing both hydroxyl and ketone moieties. TTN are shown for the major product of each reaction. Conversion values were calculated with extracted ion chromatogram peak areas from LC-MS experiments. Each data point corresponds to the average of three replicates and error bars represent the corresponding standard deviations.

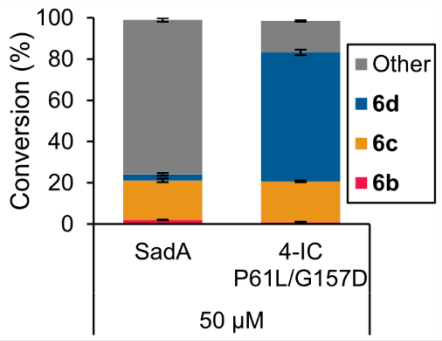
Reactions with substrates **2a** and **6a** did not show improved overall conversion values with the evaluated enzymes. However, engineered variants were found that exhibited product distributions different from wild-type SadA (Figure 4.9). In the case of **2a**, the major product in the presence of SadA was the monohydroxylated compound **2b** with a ~6.8% conversion to other oxidized products. Variant 2-L G157D was able to almost increase the selectivity towards **2b**, with more than 99% of the product pool corresponding to this monohydroxylated product. As previously discussed before, 4-IC P61L/G157D is able to produce compound **6d** as the major product of **6a** hydroxylation, an isomer which wild-type SadA is unable to produce in substantial amounts. The selectivity towards **6d** improved from 3% with SadA to 64% with 4-IC P61L/G157D. These two examples show engineered SadA variants that alter the product distribution of hydroxylation reactions.

Figure 4.9: Altered product distribution with SadA variants[a]





Variant	Selectivity (%) 2b
SadA	93
2-L G157D	>99

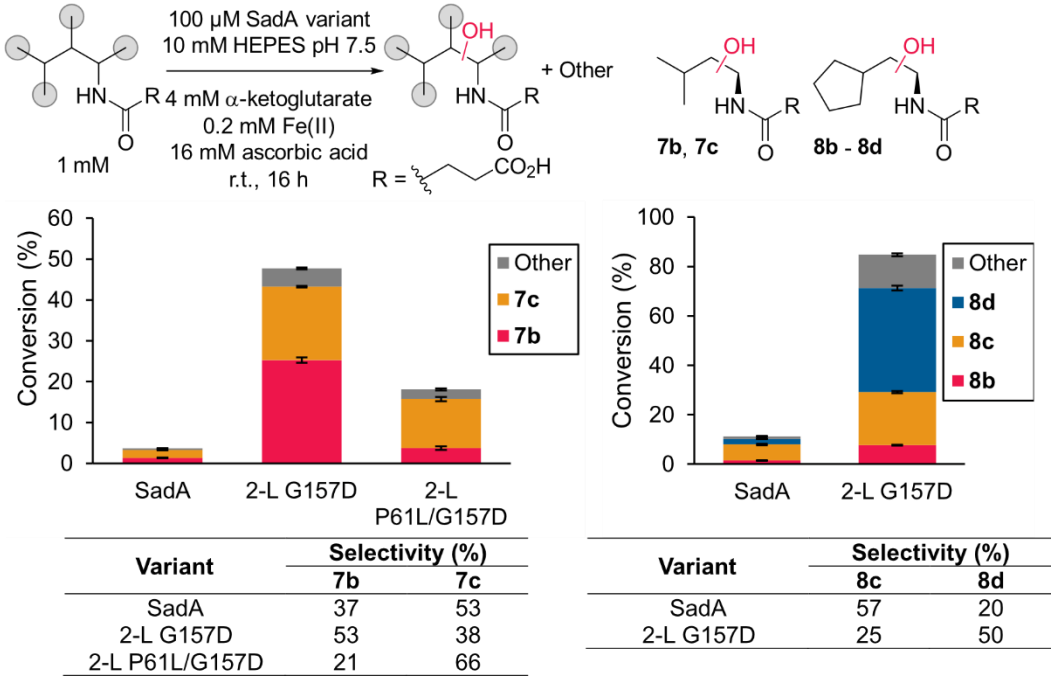


Variant	Selectivity (%) 6c	6d
SadA	19	3
4-IC P61L/G157D	20	64

[a] Conversion to monohydroxylated products by SadA and SadA engineered variants using either 50 µM or 100 µM of biocatalyst. Other products include ketones, dihydroxylated products, and compounds containing both hydroxyl and ketone moieties. Selectivity values are expressed as the fraction of a specific product in the total product pool. Conversion values were calculated with extracted ion chromatogram peak areas from LC-MS experiments. Each data point corresponds to the average of three replicates and error bars represent the corresponding standard deviations.

N-succinylated amines **7a** and **8a** showed increased conversion to hydroxylated products using engineered variants relative to SadA, and the product distribution of the corresponding reactions was altered with respect to the wild-type enzyme (Figure 4.10). The overall conversion of substrate **7a** by wild-type SadA was 3.7%, whereas that value increased to 47.7% and 18.1% with 2-L G157D and 2-L P61L/G157D, respectively. Additionally, 2-L G157D showed a distinct site selectivity, with the major product being **7b** with 53% of the product pool. 2-L P61L/G157D, on the other hand, showed a site selectivity that favored the other isomer, **7c**, with a selectivity of 66%. In the case of reactions with substrate **8a**, a similar effect was observed, with SadA only converting 11.3% of substrate to products, but 2-L G157D showing an overall conversion of 84.8%. This improvement was accompanied by a substantial increase in the selectivity towards monohydroxylated compound **8d** with it being 50% of the product pool. It is worth noting that this compound was not the major product when SadA was used. The original azidase 2-L was the variant that showed the highest conversion of substrate **7a** (Scheme 3.6), which helps rationalize that the derived 2-L G157D and 2-L P61L/G157D variants are the best suited to transform these amine derivatives.

Figure 4.10: Improved hydroxylation and altered site selectivity by engineered SadA variants[a]



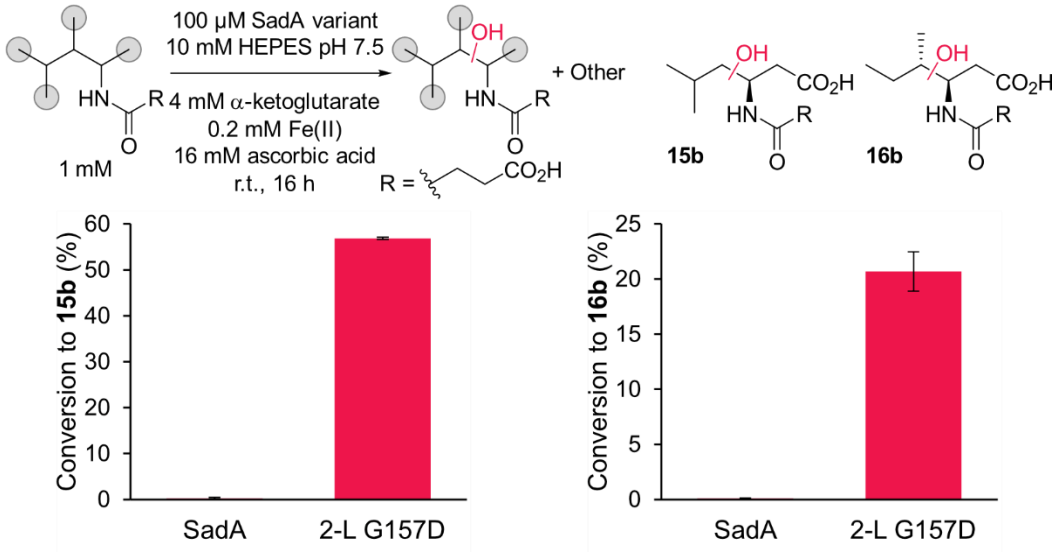
Variant	Selectivity (%)	
	7b	7c
SadA	37	53
2-L G157D	53	38
2-L P61L/G157D	21	66

Variant	Selectivity (%)	
	8c	8d
SadA	57	20
2-L G157D	25	50

[a] Conversion to monohydroxylated products by SadA and SadA engineered variants. Other products include ketones, dehydroxylated products, and compounds containing both hydroxyl and ketone moieties. Selectivity values are expressed as the fraction of a specific product in the total product pool. Conversion values were calculated with extracted ion chromatogram peak areas from LC-MS experiments. Each data point corresponds to the average of three replicates and error bars represent the corresponding standard deviations.

Finally, variant 2-L G157D enabled the functionalization of *N*-succinylated β-amino acids **15a** and **16a** which were not accepted by SadA (Figure 4.11). These compounds were hydroxylated to single monohydroxylated products **15b** and **16b** with 56.8% and 20.7% conversion values, respectively. Estimated TTNs for these reactions were low (<6) suggesting that further engineering is needed to further improve these transformations. However, the fact that 2-L G157D is able to accommodate and successfully functionalize these β-amino acid derivatives is a promising fact. These observations also highlight the promiscuity of the 2-L scaffold, which so far has been able to catalyze reactions with substrates that differ the most from the original *N*-succinylated α-amino acids among the screened panel (i.e., *N*-succinylated amines and *N*-succinylated β-amino acids).

Figure 4.11: Hydroxylation of *N*-succinylated β-amino acids by engineered 2-L G157D[a]



[a] Conversion to monohydroxylated products **15b** and **16b** by SadA and 2-L G157D. Conversion values were calculated with extracted ion chromatogram peak areas of products and the substrate from LC-MS experiments. Each data point corresponds to the average of three replicates and error bars represent the corresponding standard deviations.

Whereas further characterization of products and engineering efforts would be needed to better understand the nature of the changes in reactivity and selectivity observed in these transformations, this study aims to showcase the potential of engineered FeDOs to achieve selective C–H hydroxylation of a diverse set of compounds. The diverse set of engineered enzymes and substrates analyzed here provides enough information to fuel future efforts of SadA-derived biocatalyst development. These results complement recent efforts in engineering FeDOs for selective hydroxylation by providing a panel of biocatalysts with a broad range of selectivities and substrate scope.[51]

4.3 Conclusions

The native activity of SadA was explored in this chapter. Directed evolution was successfully applied to develop the improved *N*-succinyl-L-leucine β-hydroxylase 1-LHD. The activity of this enzyme scaffold in the presence of exogenous anionic ligands was characterized by introducing the facial triad mutation D157G. Additionally, SadX variants that were evolved

for non-native chlorination and azidation were engineered to the hydroxylase-equivalent of these variants by introducing the reversion G157D. These new engineered hydroxylases were screened against a panel of *N*-protected amino acids and amines. Substrate-enzyme pairs were identified that showed improved hydroxylase activity, altered selectivities, or a combination of both, when compared to wild-type SadA. The results presented in this study illustrate the power of SadA variants to catalyze a range of hydroxylation reactions with varying selectivities.

4.4 Experimental

4.4.1 Materials

SadX (MBP-SadA D157G fusion) and evolved variants 1-VH, 2-L, 2-D, 3-VRL, and 4-IC were obtained and expressed from a previously generated plasmids, as described in chapters 2 and 3.[52,56] BL21-Gold (DE3) *E. coli* cells (# 230132) were purchased from Agilent Technologies. Lysozyme from chicken egg white (#L6876) and deoxyribonuclease I from bovine pancreas (DNase, #DN25) were purchased from MilliporeSigma. Restriction enzymes BamHI-HF (#R3136), HindIII-HF (#R3104), and DpnI (#R0176), as well as Taq DNA polymerase (#M0273), T4 DNA ligase (#M0202), and dNTPs (#N0447) were purchased from New England Biolabs. PrimeSTAR Max DNA polymerase mix (#R045A) was purchased from Takara Bio USA. Oligonucleotides were purchased from MilliporeSigma. Miniprep kits (QIAprep Spin Miniprep Kit, #27104) and agarose gel extraction kits (QIAquick Gel Extraction Kit, #28704) were purchased from QIAGEN and used following the manufacturer's protocols. DNA clean and concentrator kits (#D4003) were purchased from Zymo Research and used following the manufacturer's instructions. Terrific broth (TB, #T15000) and Luria broth (LB, #L24040) were purchased from Research Products International and prepared following the manufacturer's protocols.

Kanamycin monosulfate (#J61272) was purchased from Alfa Aesar and prepared into a 50 mg/mL aqueous 1000x stock solution. Isopropyl-β-D-thiogalactopyranoside (IPTG, #00194) was purchased from Chem-Impex International. Ni-NTA resin (HisPur™ Ni-NTA Resin, #88223) was purchased from Thermo Scientific.

4.4.2 General procedures

Colony picking, library expression, and screening reactions were set up using an automation system consisting of a Thermo Scientific Spinnaker robotic arm and a Multidrop Combi liquid dispenser, a NorgrenSystems CP-7200 colony picker, and a Hamilton Nimbus liquid handler. The automation system was controlled by Thermo Scientific Momentum software. DNA and protein concentrations were measured with a Tecan Infinite 200 PRO plate reader. DNA amplification by PCR was performed on an Applied Biosystems ProFlex PCR System thermocycler. DNA transformation was carried out by electroporation with a Bio-Rad MicroPulser. Cell lysis by sonication was performed using a QSonica S-4000 sonicator with a 0.5” horn.

LC-MS analysis was carried out using an Agilent 1290 Infinity II system connected to a ZORBAX Eclipse Plus C18 column (2.1 mm x 50 mm, 1.8 µm particle size), an Agilent Jet Stream (AJS-ES) ion source, and a 6135 single quadrupole mass selective detector. LC-MS data analysis was carried out with MassHunter (Agilent) software.

4.4.3 Chromatographic methods

All chromatographic methods described below used 0.1 % formic acid in water (A) and 0.1 % formic acid in acetonitrile (B) as mobile phase components. Table 4.1 shows the *m/z* values used to obtain EICs from each experiment.

LC-MS 1

Solvent gradient: 10 %B (0 – 0.4 min), 20 to 26.25 %B (0.4 – 1.75 min), 90 %B (1.75 – 2.5 min), 10 %B (2.5 – 3.25 min). Flow rate: 0.4 mL/min. ESI negative mode.

LC-MS 2

Solvent gradient: 5 %B (0 – 1.25 min), 5 to 14 %B (1.25 – 2 min), 14 %B (2 – 4.5 min), 14 to 90 %B (4.5 – 6.8 min), 90 %B (6.8 – 7 min), 5 %B (7 – 7.25 min). Flow rate: 0.4 mL/min. ESI negative mode.

Table 4.1: Dominant ions for relevant substrate and monohydroxylated species.

Species	*m/z*	Species	*m/z*
1a, 10a	230	**11a**	214
1b, 1c, 10b, 10c	246	**11$_{OH}$**	230
2a, 9a	216	**12a**	245
2b, 9b, 9c	230	**12$_{OH}$**	261
6a	256	**13a**	304
6b, 6c, 6d	272	**13$_{OH}$**	320
7a	186	**14a**	200
7b, 7c	202	**14$_{OH}$**	216
8a	212	**15a, 16a**	244
8$_{OH}$	228	**15b, 16b**	260

Extracted ion chromatograms (EICs) were extracted for all species involved in a given reaction. The area of the corresponding peaks was integrated and used to represent the amount of each analyte. Percent conversion to product was calculated as the percent ratio of the area of that given product over the sum of all involved analytes.

4.4.4 Molecular cloning

Standard molecular cloning procedures were used throughout this study.[57] Details for the cloning of the relevant libraries and variants are detailed below.

SadA error-prone PCR library

A library of SadA variants for directed evolution were made using error-prone PCR (epPCR), restriction digestion, and ligation. A pET28a plasmid containing the SadA gene was used as template for the cloning involved. The oligonucleotide primers used are shown in Table 4.2.

Table 4.2: Oligonucleotides used for the preparation of an epPCR library of SadA variants

#	Name	Sequence
1	SadAInsertFP	5' – AGC AAA TGG GTC GCG GAT CCA TGC AGC ATA CCT ATC CGG C – 3'
2	SadXInsertRP	5' –TCG AGT GCG GCC GCA AGC TTT CAA TCA AAC ATA CGC CAA CC – 3'

Generation of the mutated SadA insert gene was carried out in a 50 µL reaction using the following conditions: 1 ng/µL template DNA, 0.2 mM dNTPs, 0.2 µM oligonucleotides **1** and **2**, 0.025 U/µL Taq DNA Polymerase, 1x Taq Standard Buffer, and 300 µM $MnCl_2$. The PCR program for this amplification was 95 °C for 30 seconds, 30 cycles of 95 °C for 30 seconds and 68 °C for 90 seconds, and 68 °C for 10 minutes. The product was digested with 10 units of DpnI at 37 °C for 1 hour and purified using a DNA clean and concentrator kit. The resulting DNA was then digested with 40 units of each restriction enzyme (i.e., BamHI and HindIII) in a 60 µL reaction at 37 °C for 3 hours and purified via 1% agarose gel electrophoresis.

Linearized pET28a vector was generated by digesting an empty pET28a plasmid using BamHI and HindIII in a 30 µL reaction containing ~500 ng of plasmid DNA and 30 units of each restriction enzyme. These digestions were incubated at 37 °C for 3 hours and purified via 1% agarose gel electrophoresis.

Digested insert and vector were ligated in a 20 µL reaction with a 1:10 vector:insert molar ratio. The ligation conditions were 50 ng of vector DNA, 77.6 ng of insert DNA, and 400

units of T4 DNA ligase. The ligation reaction was incubated at 16 °C overnight and purified into 7.5 µL of molecular biology grade water with a DNA clean and concentrator kit.

pET28a(1-LHD D157G)

Mutation D157G was introduced into evolved hydroxylase 1-LHD via site-directed mutagenesis of 1-LHD following a modified QuickChange™ protocol.[58] The pET28a(1-LHD) plasmid obtained from the epPCR library was used as a template for amplification using the oligonucleotides shown in Table 4.3. The amplification conditions were the following: 2 ng/µL template DNA, 0.3 µM forward primer, 0.3 µM reverse primer, and 1x PrimeSTAR Max DNA Polymerase, with the final volume being 50 µL. The PCR program for the amplifications was the following: 98 °C for 10 seconds; 30 cycles of 98 °C for 10 seconds, 55 °C for 5 seconds, and 72 °C for 45 seconds; and ending with 72 °C for 1 minute. The PCR products were digested with 10 units of DpnI at 37 °C for 1 hour and purified via 1% agarose gel electrophoresis.

Table 4.3: Oligonucleotides used to introduce mutation D157G in 1-LHD

#	Name	Sequence
3	1-LHD_D157GFP	5’ – GTT AGC TAT GGT CGT GAT ACC GTT AAT TGG CCT CTG AAA CGT AGC TTT CC – 3’
4	1-LHD_D157GRP	5’ –GTA TCA CGA CCA TAG CTA ACA TCG CCA CAA TGC GGT GCA AAG TTA AAC GGT G – 3’

Facial triad G157D reversions

Mutation G157D was introduced in evolved SadX variants 1-VH, 2-L, 2-D, 3-VRL, and 4-IC following the same QuickChange™ protocol described above, using the corresponding parent-containing pET28a plasmids as templates. The oligonucleotides used to perform these reactions are shown in Table 4.4. The introduction of G157D in 1-VH and 2-D was performed using primers **5** and **6**, while all other reactions were carried out using primers **5** and **7**.

Table 4.4: Oligonucleotides used to introduce mutation G157D in SadX variants

#	Name	Sequence
5	G157DFP	5’ – GTT AGC TAT GGT CGT GAT ACC GTT AAT TGG CCT CTG AAA CAT AGC TTT CC– 3’
6	SadXG157DRP	5’ – GTA TCA CGA CCA TAG CTA ACA TCC TCA CAA TGC GGT GCA AAG TTA AAC GGT G – 3’
7	G157DRP	5’ – GTA TCA CGA CCA TAG CTA ACA TCA TCA CAA TGC GGT GCA AGG TTA AAC GGT G – 3’

Mutations P61L, Q85H, and E147D

Mutations derived from variant 1-LHD were introduced into 2-L G157D and 4-IC G157D via the modified QuickChange™ with the oligonucleotides shown in Table 4.5. The reaction to introduce P61L into 2-L G157D was performed using oligonucleotides **8** and **9**, while those using 4-IC G157D as template were performed with **8** and **10**. Reactions to introduce Q85H and E147D were carried out using the primer pairs **11** and **12**, and **13** and **14**, respectively.

Table 4.5: Oligonucleotides used to introduced mutations P61L, Q85H, and E147D

#	Name	Sequence
8	P61LFP	5’ – GGA TGA TGA AGG TTT TAG CAG CAG CCT GTT TAC CAA AAA TGC – 3’
9	P61LRP	5’ – GCT AAA ACC TTC ATC ATC CCA CCA TGC ATC ACG TTC ACC ATC – 3’
10	4-ICP61LRP	5’ – GCT AAA ACC TTC ATC ATC CCA CCA TGC ATC ACA TTC ACC ATC – 3’
11	Q85HFP	5’ – GAT AAA CTG GTT AGC AAA GCA GCA GAA TAT TTT GGT ATT GC – 3’
12	Q85HRP	5’ – CTT TGC TAA CCA GTT TAT CAC CAT GTT CAC GCT GCA GCT GAC C – 3’
13	E147DFP	5’ – CAA CCC GTC GTG TTT ATG AAC CGT TTG ATG CAC CGT TTA AC – 3’
14	E147DRP	5’ – CAT AAA CAC GAC GGG TTG CCA GCA GAA ATT CAA CTT CAT AAT C – 3’

4.4.5 Sequences

SadA (nucleotide)

This sequence includes the SadA gene and the 5’ and 3’ flanking region of the vector (underlined).

AGCAAATGGGTCGCGGATCCATGCAGCATACCTATCCGGCACAGCTGATGCGTTTTG
GCACCGCAGCACGTGCAGAACATATGACCATTGCAGCAGCAATTCATGCACTGGAT
GCAGATGAAGCAGATGCAATTGTTATGGATATTGTTCCGGATGGTGAACGTGATGCA
TGGTGGGATGATGAAGGTTTTAGCAGCAGCCCGTTTACCAAAAATGCACATCATGCA
GGTATTGTTGCAACCAGCGTTACCCTGGGTCAGCTGCAGCGTGAACAGGGTGATAA
ACTGGTTAGCAAAGCAGCAGAATATTTTGGTATTGCCTGCCGTGTTAATGATGGTCT
GCGTACCACCCGTTTTGTTCGTCTGTTTAGTGATGCCCTGGATGCCAAACCGCTGACC
ATTGGTCATGATTATGAAGTTGAATTTCTGCTGGCAACCCGTCGTGTTTATGAACCGT
TTGAAGCACCGTTTAACTTTGCACCGCATTGTGATGATGTTAGCTATGGTCGTGATA
CCGTTAATTGGCCTCTGAAACGTAGCTTTCCGCGTCAGCTGGGTGGTTTTCTGACCAT
TCAGGGTGCAGATAATGATGCCGGTATGGTTATGTGGGATAATCGTCCGGAAAGCC
GTGCAGCGCTGGATGAAATGCATGCAGAATATCGTGAAACCGGTGCAATTGCCGCA
CTGGAACGTGCAGCCAAAATCATGCTGAAACCGCAGCCTGGCCAGCTGACACTGTT
TCAGAGCAAAAATCTGCATGCCATTGAACGTTGTACCAGCACCCGTCGTACCATGGG
TCTGTTTCTGATTCATACCGAAGATGGTTGGCGTATGTTTGATTGAGTTGGCGTATGT
TTGATTGAAAGCTTGCGGCCGCACTCGA

SadA (amino acid)

The sequence of the His_6 tag is *italicized*.

*MGSSHHHHHHSSGLVPRGSHMASMTGGQQMGRGS*MQHTYPAQLMRFGTAARAEHMTIA
AAIHALDADEADAIVMDIVPDGERDAWWDDEGFSSSPFTKNAHHAGIVATSVTLGQLQ
REQGDKLVSKAAEYFGIACRVNDGLRTTRFVRLFSDALDAKPLTIGHDYEVEFLLATRR
VYEPFEAPFNFAPHCDDVSYGRDTVNWPLKRSFPRQLGGFLTIQGADNDAGMVMWDN
RPESRAALDEMHAEYRETGAIAALERAAKIMLKPQPGQLTLFQSKNLHAIERCTSTRRT
MGLFLIHTEDGWRMFD*

SadX (nucleotide)

This sequence includes the MBP tag (*italicized*), the SadA D157G gene, and the 3’ flanking region of the vector (underlined).

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCA
TATGGCTAGCATGCACCATCACCATCACCATGGAAAAATCGAAGAAGGTAAACTGGTAATC
TGGATTAACGGCGATAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAA
GATACCGGAATTAAAGTCACCGTTGAGCATCCGGATAAACTGGAAGAGAAATTCCCACAG
GTTGCGGCAACTGGCGATGGCCCTGACATTATCTTCTGGGCACACGACCGCTTTGGTGGC
TACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTTCCAGGACAAGCTG
TATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCGCTG
TTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGA
AGAGATCCCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAA
CCTGCAAGAACCGTACTTCACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAA
GTATGAAAACGGCAAGTACGACATTAAAGACGTGGGCGTGGATAACGCTGGCGCGAAAG
CGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACACATGAATGCAGACACCGATTA
CTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTG
GGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTT
CAAGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAG
TCCGAACAAAGAGCTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTG
GAAGCGGTTAATAAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAG
TTGGCGAAAGATCCACGTATTGCCGCCACTATGGAAAACGCCCAGAAAGGTGAAATCATG
CCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCC
GCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTAATTCGAGCTC
*GAACAACAACAACACTAGTGAAAACCTGTATTTCCAGGGAGCAGCCGGATCC*ATGCAGCA
TACCTATCCGGCACAGCTGATGCGTTTTGGCACCGCAGCACGTGCAGAACATATGAC
CATTGCAGCAGCAATTCATGCACTGGATGCAGATGAAGCAGATGCAATTGTTATGG
ATATTGTTCCGGATGGTGAACGTGATGCATGGTGGGATGATGAAGGTTTTAGCAGCA
GCCCGTTTACCAAAAATGCACATCATGCAGGTATTGTTGCAACCAGCGTTACCCTGG
GTCAGCTGCAGCGTGAACAGGGTGATAAACTGGTTAGCAAAGCAGCAGAATATTTT
GGTATTGCCTGCCGTGTTAATGATGGTCTGCGTACCACCCGTTTTGTTCGTCTGTTTA
GTGATGCCCTGGATGCCAAACCGCTGACCATTGGTCATGATTATGAAGTTGAATTTC
TGCTGGCAACCCGTCGTGTTTATGAACCGTTTGAAGCACCGTTTAACTTTGCACCGC
ATTGTGGCGATGTTAGCTATGGTCGTGATACCGTTAATTGGCCTCTGAAACGTAGCT
TTCCGCGTCAGCTGGGTGGTTTTCTGACCATTCAGGGTGCAGATAATGATGCCGGTA
TGGTTATGTGGGATAATCGTCCGGAAAGCCGTGCAGCGCTGGATGAAATGCATGCA
GAATATCGTGAAACCGGTGCAATTGCCGCACTGGAACGTGCAGCCAAAATCATGCT
GAAACCGCAGCCTGGCCAGCTGACACTGTTTCAGAGCAAAAATCTGCATGCCATTG
AACGTTGTACCAGCACCCGTCGTACCATGGGTCTGTTTCTGATTCATACCGAAGATG
GTTGGCGTATGTTTGATTGA<u>AAGCTTGCGGCCGCACTCGA</u>

SadX (amino acid)

The sequence of MBP is *italicized*. All further amino acid indices refer to the original enzyme SadA numbering.

MGSSHHHHHHSSGLVPRGSHMASMHHHHHHGKIEEGKLVIWINGDKGYNGLAEVGKKFEK
DTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYP
FTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPY

FTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAF
NKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFL
ENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYA
*VRTAVINAASGRQTVDEALKDAQTNSSSNNNNTSENLYFQGAAGS*MQHTYPAQLMRFGTAA
RAEHMTIAAAIHALDADEADAIVMDIVPDGERDAWWDDEGFSSSPFTKNAHHAGIVAT
SVTLGQLQREQGDKLVSKAAEYFGIACRVNDGLRTTRFVRLFSDALDAKPLTIGHDYEV
EFLLATRRVYEPFEAPFNFAPHCGDVSYGRDTVNWPLKRSFPRQLGGFLTIQGADNDAG
MVMWDNRPESRAALDEMHAEYRETGAIAALERAAKIMLKPQPGQLTLFQSKNLHAIER
CTSTRRTMGLFLIHTEDGWRMFD*

1-VH
SadX I71V (V = GTT) R172H (H = CAT)

2-L
1-VH F152L (L = CTT)

3-VRL
2-L I38V (V = GTT) Q233R (R = CGC) F261L (L = CTT)

4-IC
3-VRL V38I (I = ATT) R48C (C = TGT)

1-LHD
SadA P61L (L = CTG) Q85H (H = CAT) E147D (D = GAT)

1-LHD G157D
1-LHD G157D (D = GAT)

G157D reversions
Corresponding parent G157D (D = GAT)

P61L variants
Corresponding parent P61L (L = CTG)

Q85H variants
Corresponding parent Q85H (H = CAT)

E147D variants
Corresponding parent E147D (D = GAT)

4.4.6 Gene expression and protein purification

Libraries in 96-deep well plates

A 5 µL aliquot of the resulting plasmid was used to transform 50 µL of BL21-Gold (DE3) *E. coli* via electroporation. Cells were resuspended with 945 µL of SOC media and shaken at 37 °C and 250 rpm for 45 minutes. Up to 400 µL of the culture were transferred to 86 mm x 128 mm rectangular LB/agar plates with 50 µg/mL kanamycin. The plates were incubated overnight at 37 °C. Single colonies were picked and placed in 1 mL 96-deep well plates containing 300 µL/well of TB media with 50 µg/mL kanamycin to generate primary cultures. Each plate contained 90 library variants, 4 parent samples (picked and inoculated using the same method as the library variants), and 2 blank wells that were not inoculated and served as controls of cross contamination. The resulting plates were incubated at 37 °C and 215 rpm overnight. 50 µL of each culture were used to make glycerol stocks of the library by combining them with 50 µL of 50% glycerol in microtiter 96-well plates.

Expression cultures were made using 50 µL of primary culture to inoculate 2 mL 96-deep well plates containing 1 mL/well of LB media with 50 µg/mL kanamycin. Expression plates were incubated at 37 °C and 215 rpm until OD_{600} reached 0.6 – 0.8, at which point the cultures were brought to a temperature of 18 °C, induced with a final concentration of 1 mM IPTG, and incubated at 18 °C and 215 rpm for 18 hours. Cells were harvested by centrifugation at 3600 rpm for 10 minutes and stored at -80 °C until use.

Large scale protein expression

A starter culture was prepared by inoculating 5 mL of TB media with 50 µg/mL kanamycin from glycerol stocks made during library expression and incubating at 37 °C and 250 rpm overnight. The overnight culture was used to generate an expression culture by inoculating

750 mL of LB media with 50 µg/mL kanamycin. The expression culture was incubated at 37 °C and 250 rpm until OD_{600} reached 0.6 – 0.8, after which the culture was brought to 18 °C, induced with 1 mM IPTG, and incubated at 18 °C and 250 rpm for 18 hours. Cells were harvested by centrifugation at 4700 rpm for 10 minutes and stored at -80 °C until use.

The cell pellet was resuspended in a 50 mL conical tube with 30 mL of 20 mM imidazole in 10 mM HEPES pH 7.5 and sonicated at 40 W with 0.5 min ON/0.5 min OFF cycles for 5 min total cycle time. The resulting cell lysate was clarified by centrifugation at 4 °C and 15000 rpm for 45 min using a high-speed fixed-angle rotor. The supernatant was transferred to a 10 mL polypropylene frit-bottomed gravity flow column containing 5 mL of Ni-NTA resin pre-equilibrated with equilibration buffer (20 mM phosphate, 300 mM NaCl, 10 mM imidazole, pH 7.4). After the lysate was allowed to drain, 50 mL of wash buffer (20 mM phosphate, 300 mM NaCl, 25 mM imidazole, pH 7.4) were added to the resin. Finally, 15 mL of elution buffer (20 mM phosphate, 300 mM NaCl, 250 mM imidazole, pH 7.4) were used to elute purified protein into a 50 mL conical tube. The protein solution was transferred to a 15 mL Amicon spin filter 30K MWCO and concentrated down to 0.5 – 1.0 mL at 4700 rpm. In order to remove endogenous Fe(II) 15 mL of 50 mM EDTA in 10 mM HEPES pH 7.5 were used to dilute the protein sample and concentrate to 0.5 – 1.0 mL at 4700 rpm. Three additional buffer exchanges with 15 mL of 10 mM HEPES in pH 7.5 were performed. Protein concentration was measured using absorbance at 280 nm using a calculated extinction coefficient of 104,280 $M^{-1}cm^{-1}$ for all SadX variants and 36,440 $M^{-1}cm^{-1}$ for all SadA variants (Benchling [Biology Software], https://benchling.com). The protein sample was flash frozen in liquid nitrogen and stored at -80 °C until used.

4.4.7 Activity assays

Screening hydroxylation assay in lysate

Cell pellets in 2 mL 96-deep well plates were suspended with 125 µL of 1 mg/mL lysozyme in 10 mM HEPES pH 7.5 and incubated at 37 °C and 250 rpm for 45 minutes to achieve lysis, after which the lysates were flash frozen in liquid nitrogen and thawed in a water bath at 37 °C. 10 µL of 1 mg/mL DNase in 10 mM HEPES pH 7.5 were added and the lysates were incubated at 37 °C and 250 rpm for 15 minutes, followed by centrifugation at 3600 rpm for 15 minutes.

Screening reactions were set up in microtiter v-bottom 96-well plates by combining 25 µL/well of solution A (50 mL containing 4 mM α-ketoglutarate, 16 mM ascorbic acid, and 0.2 mM $Fe(NH_4)_2(SO_4)_2$ in 10 mM HEPES pH 7.5), 25 µL/well of solution B (30 mL of 4 mM **1a** in 10 mM HEPES pH 7.5), and 50 µL/well of clarified lysate. The final reaction conditions were 100 µL of 1 mM **1a**, 2 mM α-ketoglutarate, 8 mM ascorbic acid, 0.1 mM $Fe(NH_4)_2(SO_4)_2$, and 50% v/v lysate in 10 mM HEPES pH 7.5. The plates were sealed with breathable film and shaken at room temperature and 750 rpm for 3 hours, after which 100 µL of methanol were added to quench the reactions and precipitated proteins were removed by centrifugation at 3600 rpm for 10 minutes. 50 µL of the resulting supernatant were diluted with 100 µL of water, filtered through 0.22 µm filter plates by centrifugation at 3600 rpm for 10 minutes, and analyzed via LC-MS using LC-MS method 1.

The conversion values to **1b** for the four parent samples in each plate were averaged and used as comparison for every other variant in the same plate. Library variants were first evaluated so that only variants with activity higher than the sum of average parent activity and

2.5 standard deviations of parent activity were considered as potential hits. Potential improved variants were selected to be sequenced and validated with bioconversions using purified protein.

Small-scale hydroxylation reactions

Enzyme activity was determined by small-scale bioconversions using purified protein. Reactions were set up by combining in a 96-well plate 25 µL/well of solution A (25 mL containing 16 mM α-ketoglutarate, 64 mM ascorbic acid, and 0.4 mM $Fe(NH_4)_2(SO_4)_2$ in 10 mM HEPES pH 7.5), 25 µL/well of solution B (2 mL of 4 mM of substrate in 10 mM HEPES pH 7.5), and 50 µL/well of solution C (200 µL of an enzyme solution with two-fold the final concentration). The final reactions (100 µL) had final concentrations of 1 mM of substrate, 4 mM α-ketoglutarate, 16 mM ascorbic acid, 0.2 mM $Fe(NH_4)_2(SO_4)_2$, and varying amounts of SadA variants in 10 mM HEPES pH 7.5. The plate was sealed with breathable film and shaken at room temperature and 750 rpm for 3 hours, after which 100 µL of methanol were added to quench the reactions, 50 µL of a 2 mM *N*-acetyl-L-valine solution in water (100 nmol) were added as internal standard, and precipitated proteins were removed by centrifugation at 3600 rpm for 10 minutes. Finally, 80 µL of the resulting supernatant were filtered through 0.22 µm filter plates by centrifugation at 3600 rpm for 10 minutes and analyzed via LC-MS using LC-MS method 2.

*Small-scale reactions of **1a** in the presence of exogenous anions*

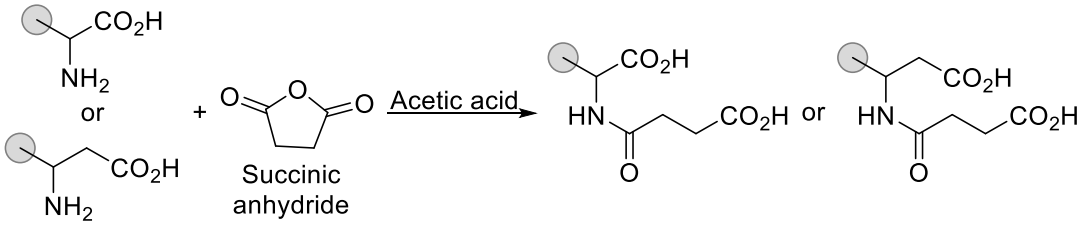
Reaction setup was analogous to hydroxylation reactions, with the exception of solution A containing 16 mM α-ketoglutarate, 64 mM ascorbic acid, 0.8 mM $Fe(NH_4)_2(SO_4)_2$, and either 200 mM NaCl, 200 mM CsF, or 32 mM NaN_3. Solution C contained 0.2 mM enzyme. The final reactions (100 µL) contained final concentrations of 1 mM **1a**, 4 mM α-ketoglutarate, 16 mM

ascorbic acid, 0.2 mM $Fe(NH_4)_2(SO_4)_2$, 0.1 mM of enzyme, and either 50 mM NaCl, 50 mM CsF, or 8 mM NaN_3 in 10 mM HEPES pH 7.5.

4.4.8 Synthesis of starting materials

General synthesis of N-succinylated amino acids

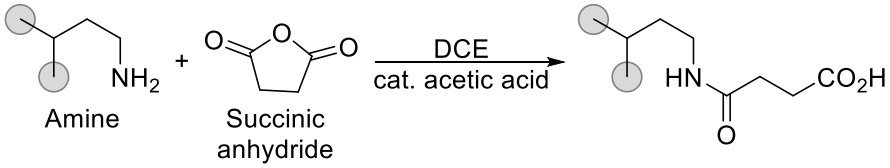
Scheme 4.8: Synthesis of *N*-succinylated amino acids



To a 100 mL flask was added amino acid (1 eq), succinic anhydride (1.05 eq), and acetic acid. The mixture was heated to 70 °C for 8 hr. After complete removal of the solvent, the solid residue was recrystallized from ethyl acetate and ethanol to obtain semi purified product. The solid residue was purified in a CombiFlash system using water and acetonitrile.

General synthesis of N-succinylated amines

Scheme 4.9: Synthesis of *N*-succinylated amines



To a 100 mL flask attached to a reflux condenser was added amine (5 g, 1 eq), succinic anhydride (6.1 g, 1.05 eq), a catalytic amount of acetic acid, and 50 mL of dichloroethane. The mixture was heated to 40 °C for 12 hr. After removal of the solvent, the solid residue was recrystallized from toluene to obtain a white solid.

4.5 References

(1) Roughley, S. D.; Jordan, A. M. The Medicinal Chemist's Toolbox: An Analysis of Reactions Used in the Pursuit of Drug Candidates. *Journal of Medicinal Chemistry*. American Chemical Society May 26, 2011, pp 3451–3479. https://doi.org/10.1021/jm200187y.

(2) Cramer, J.; Sager, C. P.; Ernst, B. Hydroxyl Groups in Synthetic and Natural-Product-Derived Therapeutics: A Perspective on a Common Functional Group. *Journal of Medicinal Chemistry*. American Chemical Society October 24, 2019, pp 8915–8930. https://doi.org/10.1021/acs.jmedchem.9b00179.

(3) Bento, A. P.; Gaulton, A.; Hersey, A.; Bellis, L. J.; Chambers, J.; Davies, M.; Krüger, F. A.; Light, Y.; Mak, L.; McGlinchey, S.; Nowotka, M.; Papadatos, G.; Santos, R.; Overington, J. P. The ChEMBL Bioactivity Database: An Update. *Nucleic Acids Res.* **2014**, *42* (D1), D1083–D1090. https://doi.org/10.1093/NAR/GKT1031.

(4) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs from 1981 to 2014. *Journal of Natural Products*. American Chemical Society March 25, 2016, pp 629–661. https://doi.org/10.1021/acs.jnatprod.5b01055.

(5) Barratt, E.; Bronowska, A.; Vondrášek, J.; Černý, J.; Bingham, R.; Phillips, S.; Homans, S. W. Thermodynamic Penalty Arising from Burial of a Ligand Polar Group Within a Hydrophobic Pocket of a Protein Receptor. **2006**. https://doi.org/10.1016/j.jmb.2006.07.067.

(6) Babine, R. E.; Bender, S. L. Molecular Recognition of Protein-Ligand Complexes: Applications to Drug Design. *Chem. Rev.* **1997**, *97* (5), 1359–1472. https://doi.org/10.1021/cr960370z.

(7) Groves, J. T.; Nemo, T. E.; Myers, R. S. Hydroxylation and Epoxidation Catalyzed by Iron-Porphine Complexes. Oxygen Transfer from Iodosylbenzene. *J. Am. Chem. Soc.* **1979**, *101* (4), 1032–1033. https://doi.org/10.1021/ja00498a040.

(8) Bovicelli, P.; Lupattelli, P.; Mincione, E.; Sapienza, L.; Moro, ple A.; Teresa Prencipe, I.; Curci, R. Oxidation of Natural Targets by Dioxiranes. 2.1 Direct Hydroxylation at the Side-Chain C-25 of Cholestane Derivatives and of Vitamin D3 Windaus-Grundmann Ketone. *J. Org. Chem* **1992**, *57*, 5052–5054.

(9) Chen, M. S.; White, M. C. A Predictably Selective Aliphatic C-H Oxidation Reaction for Complex Molecule Synthesis. *Science (80-.).* **2007**, *318* (5851), 783–787. https://doi.org/10.1126/science.1148597.

(10) Burg, F.; Gicquel, M.; Breitenlechner, S.; Pöthig, A.; Bach, T. Site- and Enantioselective C−H Oxygenation Catalyzed by a Chiral Manganese Porphyrin Complex with a Remote Binding Site. *Angew. Chemie Int. Ed.* **2018**, *57* (11), 2953–2957. https://doi.org/10.1002/ANIE.201712340.

(11) Litvinas, N. D.; Brodsky, B. H.; Du Bois, J. C-H Hydroxylation Using a Heterocyclic Catalyst and Aqueous H 2O2. *Angew. Chemie - Int. Ed.* **2009**, *48* (25), 4513–4516. https://doi.org/10.1002/anie.200901353.

(12) Johnson, S. L.; Combee, L. A.; Hilinski, M. K. Organocatalytic Atom-Transfer C(Sp3)-H Oxidation. *Synlett* **2018**, *29* (18), 2331–2336. https://doi.org/10.1055/s-0037-1610432.

(13) Hahn, P. L.; Lowe, J. M.; Xu, Y.; Burns, K. L.; Hilinski, M. K. Amine Organocatalysis of Remote, Chemoselective C(Sp3)-H Hydroxylation. *ACS Catal.* **2022**, *12* (8), 4302–4309. https://doi.org/10.1021/acscatal.2c00392.

(14) Nozaki, M. Oxygenases and Dioxygenases. *Top. Curr. Chem.* **1979**, *78*. https://doi.org/10.1007/BFb0048193.

(15) Mondal, D.; Snodgrass, H. M.; Gomez, C. A.; Lewis, J. C. Non-Native Site-Selective Enzyme Catalysis. **2023**. https://doi.org/10.26434/CHEMRXIV-2023-SQ85V.

(16) Denisov, I. G.; Makris, T. M.; Sligar, S. G.; Schlichting, I. Structure and Chemistry of Cytochrome P450. *Chemical Reviews*. American Chemical Society June 2005, pp 2253–2277. https://doi.org/10.1021/cr0307143.

(17) McIntosh, J. A.; Farwell, C. C.; Arnold, F. H. Expanding P450 Catalytic Reaction Space through Evolution and Engineering. *Current Opinion in Chemical Biology*. Elsevier Current Trends April 1, 2014, pp 126–134. https://doi.org/10.1016/j.cbpa.2014.02.001.

(18) Wei, Y.; Ang, E. L.; Zhao, H. Recent Developments in the Application of P450 Based Biocatalysts. *Current Opinion in Chemical Biology*. Elsevier Current Trends April 1, 2018, pp 1–7. https://doi.org/10.1016/j.cbpa.2017.08.006.

(19) Urlacher, V. B.; Girhard, M. Cytochrome P450 Monooxygenases in Biotechnology and Synthetic Biology. *Trends in Biotechnology*. Elsevier Current Trends August 1, 2019, pp 882–897. https://doi.org/10.1016/j.tibtech.2019.01.001.

(20) Groves, J. T. Models and Mechanisms of Cytochrome P450 Action. In *Cytochrome P450: Structure, Mechanism, and Biochemistry: Third edition*; Springer US, 2005; pp 1–43. https://doi.org/10.1007/0-387-27447-2_1.

(21) Warman, A. J.; Roitel, O.; Neeli, R.; Girvan, H. M.; Seward, H. E.; Murray, S. A.; McLean, K. J.; Joyce, M. G.; Toogood, H.; Holt, R. A.; Leys, D.; Scrutton, N. S.; Munro, A. W. Flavocytochrome P450 BM3: An Update on Structure and Mechanism of a Biotechnologically Important Enzyme. In *Biochemical Society Transactions*; Portland Press, 2005; Vol. 33, pp 747–753. https://doi.org/10.1042/BST0330747.

(22) Peters, M. W.; Meinhold, P.; Glieder, A.; Arnold, F. H. Regio- and Enantioselective Alkane Hydroxylation with Engineered Cytochromes P450 BM-3. *J. Am. Chem. Soc.* **2003**, *125* (44), 13442–13450. https://doi.org/10.1021/ja0303790.

(23) Lewis, J. C.; Bastian, S.; Bennett, C. S.; Fu, Y.; Mitsuda, Y.; Chen, M. M.; Greenberg, W.

A.; Wong, C. H.; Arnold, F. H. Chemoenzymatic Elaboration of Monosaccharides Using Engineered Cytochrome P450BM3 Demethylases. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (39), 16550–16555. https://doi.org/10.1073/PNAS.0908954106.

(24) Kille, S.; Zilly, F. E.; Acevedo, J. P.; Reetz, M. T. Regio- and Stereoselectivity of P450-Catalysed Hydroxylation of Steroids Controlled by Laboratory Evolution. *Nat. Chem.* **2011**, *3* (9), 738–743. https://doi.org/10.1038/nchem.1113.

(25) Beltrán-Nogal, A.; Sánchez-Moreno, I.; Méndez-Sánchez, D.; Gómez de Santos, P.; Hollmann, F.; Alcalde, M. Surfing the Wave of Oxyfunctionalization Chemistry by Engineering Fungal Unspecific Peroxygenases. *Current Opinion in Structural Biology*. Elsevier Current Trends April 1, 2022, p 102342. https://doi.org/10.1016/j.sbi.2022.102342.

(26) Hobisch, M.; Holtmann, D.; Gomez de Santos, P.; Alcalde, M.; Hollmann, F.; Kara, S. Recent Developments in the Use of Peroxygenases – Exploring Their High Potential in Selective Oxyfunctionalisations. *Biotechnology Advances*. Elsevier November 1, 2021, p 107615. https://doi.org/10.1016/j.biotechadv.2020.107615.

(27) Molina-Espeja, P.; Cañellas, M.; Plou, F. J.; Hofrichter, M.; Lucas, F.; Guallar, V.; Alcalde, M. Synthesis of 1-Naphthol by a Natural Peroxygenase Engineered by Directed Evolution. *ChemBioChem* **2016**, *17* (4), 341–349. https://doi.org/10.1002/cbic.201500493.

(28) Barková, K.; Kinne, M.; Ullrich, R.; Hennig, L.; Fuchs, A.; Hofrichter, M. Regioselective Hydroxylation of Diverse Flavonoids by an Aromatic Peroxygenase. *Tetrahedron* **2011**, *67* (26), 4874–4878. https://doi.org/10.1016/j.tet.2011.05.008.

(29) Gomez de Santos, P.; González-Benjumea, A.; Fernandez-Garcia, A.; Aranda, C.; Wu, Y.; But, A.; Molina-Espeja, P.; Maté, D. M.; Gonzalez-Perez, D.; Zhang, W.; Kiebist, J.; Scheibner, K.; Hofrichter, M.; Świderek, K.; Moliner, V.; Sanz-Aparicio, J.; Hollmann, F.; Gutiérrez, A.; Alcalde, M. Engineering a Highly Regioselective Fungal Peroxygenase for the Synthesis of Hydroxy Fatty Acids. *Angew. Chemie - Int. Ed.* **2023**, *62* (9), e202217372. https://doi.org/10.1002/anie.202217372.

(30) Lukowski, A. L.; Ellinwood, D. C.; Hinze, M. E.; Deluca, R. J.; Du Bois, J.; Hall, S.; Narayan, A. R. H. C-H Hydroxylation in Paralytic Shellfish Toxin Biosynthesis. *J. Am. Chem. Soc.* **2018**, *140* (37), 11863–11869. https://doi.org/10.1021/jacs.8b08901.

(31) Colby, J.; Stirling, D. I.; Dalton, H. The Soluble Methane Mono Oxygenase of Methylococcus Capsulatus (Bath). Its Ability to Oxygenate n Alkanes, n Alkenes, Ethers, and Alicyclic, Aromatic and Heterocyclic Compounds. *Biochem. J.* **1977**, *165* (2), 395–402. https://doi.org/10.1042/bj1650395.

(32) Green, J.; Dalton, H. Substrate Specificity of Soluble Methane Monooxygenase. *J. Biol. Chem.* **1989**, *264* (30), 17698–17703. https://doi.org/10.1016/s0021-9258(19)84627-6.

(33) Elliott, S. J.; Zhu, M.; Tso, L.; Nguyen, H. H. T.; Yip, J. H. K.; Chan, S. I. Regio- and Stereoselectivity of Particulate Methane Monooxygenase from Methylococcus Capsulatus

(Bath). *J. Am. Chem. Soc.* **1997**, *119* (42), 9949–9955. https://doi.org/10.1021/ja971049g.

(34) Zharikova, N. V.; Iasakov, T. R.; Zhurenko, E. I.; Korobov, V. V.; Markusheva, T. V. Bacterial Genes of Non-Heme Iron Oxygenases, Which Have a Rieske-Type Cluster, Catalyzing Initial Stages of Degradation of Chlorophenoxyacetic Acids. *Russian Journal of Genetics*. Pleiades Publishing March 1, 2018, pp 284–295. https://doi.org/10.1134/S1022795418030171.

(35) Gally, C.; Nestl, B. M.; Hauer, B. Engineering Rieske Non-Heme Iron Oxygenases for the Asymmetric Dihydroxylation of Alkenes. *Angew. Chemie - Int. Ed.* **2015**, *54* (44), 12952–12956. https://doi.org/10.1002/anie.201506527.

(36) Lippard, S. J. Hydroxylation of C-H Bonds at Carboxylate-Bridged Diiron Centres. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2005**, *363* (1829), 861–877. https://doi.org/10.1098/RSTA.2004.1532.

(37) Mitchell, K. H.; Studts, J. M.; Fox, B. G.; Ave, U. Combined Participation of Hydroxylase Active Site Residues and Effector Protein Binding in a Para to Ortho Modulation of Toluene 4-Monooxygenase Regiospecificity †. **2022**, *17*, 54. https://doi.org/10.1021/bi012036p.

(38) Chenprakhon, P.; Wongnate, T.; Chaiyen, P. Monooxygenation of Aromatic Compounds by Flavin-Dependent Monooxygenases. *Protein Sci.* **2019**, *28* (1), 8–29. https://doi.org/10.1002/pro.3525.

(39) Chaiyen, P.; Suadee, C.; Wilairat, P. A Novel Two-Protein Component Flavoprotein Hydroxylase p-Hydroxyphenylacetate Hydroxylase from Acinetobacter Baumannii. *Eur. J. Biochem.* **2001**, *268* (21), 5550–5561. https://doi.org/10.1046/j.1432-1033.2001.02490.x.

(40) Herrmann, S.; Dippe, M.; Pecher, P.; Funke, E.; Pietzsch, M.; Wessjohann, L. A. Engineered Bacterial Flavin-Dependent Monooxygenases for the Regiospecific Hydroxylation of Polycyclic Phenols. *ChemBioChem* **2022**, *23* (6). https://doi.org/10.1002/CBIC.202100480.

(41) Zabala, A. O.; Xu, W.; Chooi, Y. H.; Tang, Y. Characterization of a Silent Azaphilone Gene Cluster from Aspergillus Niger ATCC 1015 Reveals a Hydroxylation-Mediated Pyran-Ring Formation. *Chem. Biol.* **2012**, *19* (8), 1049–1059. https://doi.org/10.1016/j.chembiol.2012.07.004.

(42) Al Fahad, A.; Abood, A.; Fisch, K. M.; Osipow, A.; Davison, J.; Avramović, M.; Butts, C. P.; Piel, J.; Simpson, T. J.; Cox, R. J. Oxidative Dearomatisation: The Key Step of Sorbicillinoid Biosynthesis. *Chem. Sci.* **2014**, *5* (2), 523–527. https://doi.org/10.1039/c3sc52911h.

(43) Abood, A.; Al-Fahad, A.; Scott, A.; Hosny, A. E. D. M. S.; Hashem, A. M.; Fattah, A. M. A.; Race, P. R.; Simpson, T. J.; Cox, R. J. Kinetic Characterisation of the FAD Dependent Monooxygenase TropB and Investigation of Its Biotransformation Potential. *RSC Adv.*

2015, *5* (62), 49987–49995. https://doi.org/10.1039/c5ra06693j.

(44) Baker Dockrey, S. A.; Lukowski, A. L.; Becker, M. R.; Narayan, A. R. H. Biocatalytic Site- and Enantioselective Oxidative Dearomatization of Phenols. *Nat. Chem.* **2018**, *10* (2), 119–125. https://doi.org/10.1038/nchem.2879.

(45) Lazzarotto, M.; Hammerer, L.; Hetmann, M.; Borg, A.; Schmermund, L.; Steiner, L.; Hartmann, P.; Belaj, F.; Kroutil, W.; Gruber, K.; Fuchs, M. Chemoenzymatic Total Synthesis of Deoxy-, Epi-, and Podophyllotoxin and a Biocatalytic Kinetic Resolution of Dibenzylbutyrolactones. *Angew. Chemie - Int. Ed.* **2019**, *58* (24), 8226–8230. https://doi.org/10.1002/anie.201900926.

(46) Li, J.; Zhang, X.; Renata, H. Asymmetric Chemoenzymatic Synthesis of (−)-Podophyllotoxin and Related Aryltetralin Lignans. *Angew. Chemie - Int. Ed.* **2019**, *58* (34), 11657–11660. https://doi.org/10.1002/anie.201904102.

(47) Chen, H.; Bong, Y. K.; Cabirol, F. L.; Prafulchandra, A. G.; Li, T.; Moore, J. C.; Quintanar-Audelo, M.; Hong, Y.; Collier, S. J.; Smith, D. Biocatalysts and Methods for Hydroxylation of Chemical Compounds, 2015.

(48) Mailyan, A. K.; Chen, J. L.; Li, W.; Keller, A. A.; Sternisha, S. M.; Miller, B. G.; Zakarian, A. Short Total Synthesis of [15N5]-Cylindrospermopsins from 15NH4Cl Enables Precise Quantification of Freshwater Cyanobacterial Contamination. *J. Am. Chem. Soc.* **2018**, *140* (18), 6027–6032. https://doi.org/10.1021/jacs.8b03071.

(49) Zwick, C. R.; Renata, H. Overview of Amino Acid Modifications by Iron- and α-Ketoglutarate-Dependent Enzymes. *ACS Catal.* **2023**, *13*, 4853–4865. https://doi.org/10.1021/acscatal.3c00424.

(50) Strieker, M.; Essen, L.-O.; Walsh, C. T.; Marahiel, M. A.; Strieker,] M; Essen, L.-O.; Marahiel, M. A.; Walsh, C. T. Non-Heme Hydroxylase Engineering For Simple Enzymatic Synthesis of L-Threo-Hydroxyaspartic Acid. *ChemBioChem* **2008**, *9* (3), 374–376. https://doi.org/10.1002/CBIC.200700557.

(51) Meyer, F.; Frey, R.; Ligibel, M.; Sager, E.; Schroer, K.; Snajdrova, R.; Buller, R. Modulating Chemoselectivity in a Fe(II)/α-Ketoglutarate-Dependent Dioxygenase for the Oxidative Modification of a Nonproteinogenic Amino Acid. *ACS Catal.* **2021**, *11* (10), 6261–6269. https://doi.org/10.1021/acscatal.1c00678.

(52) Gomez, C. A.; Mondal, D.; Du, Q.; Chan, N.; Lewis, J. C. Directed Evolution of an Iron(II)- and α-Ketoglutarate-Dependent Dioxygenase for Site-Selective Azidation of Unactivated Aliphatic C−H Bonds**. *Angew. Chemie Int. Ed.* **2023**, *62* (15), e202301370. https://doi.org/10.1002/ANIE.202301370.

(53) Qin, H. M.; Miyakawa, T.; Jia, M. Z.; Nakamura, A.; Ohtsuka, J.; Xue, Y. L.; Kawashima, T.; Kasahara, T.; Hibi, M.; Ogawa, J.; Tanokura, M. Crystal Structure of a Novel N-Substituted L-Amino Acid Dioxygenase from Burkholderia Ambifaria AMMD. *PLoS One* **2013**, *8* (5), e63996. https://doi.org/10.1371/JOURNAL.PONE.0063996.

(54) Islam, M. S.; Leissing, T. M.; Chowdhury, R.; Hopkinson, R. J.; Schofield, C. J. Annual Review of Biochemistry 2-Oxoglutarate-Dependent Oxygenases. *Annu. Rev. Biochem.* **2018**, 585–620. https://doi.org/10.1146/annurev-biochem.

(55) Qin, H. M.; Miyakawa, T.; Nakamura, A.; Hibi, M.; Ogawa, J.; Tanokura, M. Structural Optimization of SadA, an Fe(II)- and α-Ketoglutarate- Dependent Dioxygenase Targeting Biocatalytic Synthesis of N-Succinyl-L-Threo-3, 4-Dimethoxyphenylserine. *Biochem. Biophys. Res. Commun.* **2014**, *450* (4), 1458–1461. https://doi.org/10.1016/j.bbrc.2014.07.008.

(56) Chan, N. H.; Gomez, C. A.; Vennelakanti, V.; Du, Q.; Kulik, H. J.; Lewis, J. C. Non-Native Anionic Ligand Binding and Reactivity in Engineered Variants of the Fe(II)- and α-Ketoglutarate-Dependent Oxygenase, SadA. *Inorg. Chem.* **2022**, *61* (36), 14477–14485. https://doi.org/10.1021/acs.inorgchem.2c02872.

(57) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning: A Laboratory Manual, 2nd Edn.*; Cold Spring Laboratory Press, 1989.

(58) Liu, H.; Naismith, J. H. An Efficient One-Step Site-Directed Deletion, Insertion, Single and Multiple-Site Plasmid Mutagenesis Protocol. *BMC Biotechnol.* **2008**, *8* (1), 91. https://doi.org/10.1186/1472-6750-8-91.

Chapter 5: Protein engineering aided by next-generation sequencing and machine learning

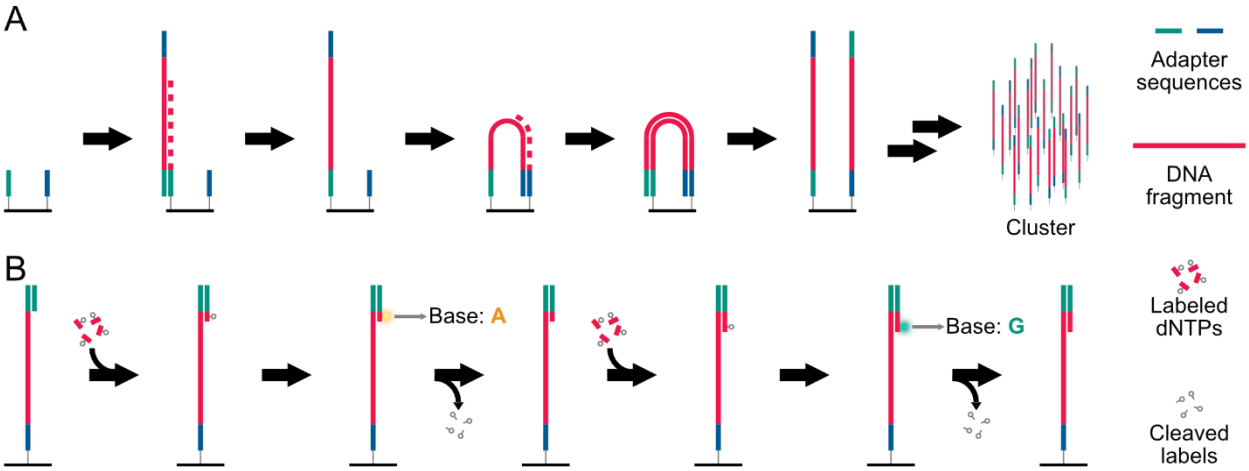
5.1 Introduction

5.1.1 Introduction to next-generation techniques

The earliest DNA sequencing method, published in the late 1970s by Sanger and coworkers, remains the standard technique for low throughput DNA sequencing.[1] This method is based on the amplification of template DNA to be sequenced by a DNA polymerase. Radioactively or fluorescently labeled chain-terminating dideoxy nucleotides (ddNTPs) are included in the amplification reactions and cause the growing chain of DNA to stop at the position where the ddNTP was incorporated. The result is a mixture of DNA fragments of varying lengths, all capped by a labeled ddNTP. The electrophoretic separation of this mixture allows for identification of the corresponding labeled ddNTPs and the reconstruction of the original sequence. Even though Sanger sequencing has been improved by technological advances, including the use of multichannel capillary electrophoresis, the throughput of this technique remains relatively low. The need for higher throughput sequencing of large genomes at relatively lower costs encouraged the development of techniques termed next-generation sequencing (NGS). These techniques encompass an array of methods based on different technologies.[2,3] A brief introduction to the most widely developed techniques will be provided in this section.

Illumina sequencing is the most widely used NGS technique from what is referred to as second generation sequencing. This technique relies on two steps: the generation of clonal clusters of DNA fragments attached to a solid support from one end (Figure 5.1A), and the use of 3'-modified reversible terminator nucleotides for sequencing (Figure 5.1B).[4] The DNA to be sequenced is first fragmented to segments of ~500 np, ligated to specific adapter sequences, and amplified using "bridged amplification" from oligonucleotides bound to the solid support that complementary to the adapter sequences. Repeated cycles of amplification result in groups of clonal DNA clustered throughout the support. Sequencing occurs via single-base amplification of the DNA clusters by the use fluorescently labeled nucleotides (dNTPs) that interrupt chain elongation. Once the modified dNTPs are incorporated, fluorescence is triggered and read, and the fluorescent label is cleaved, freeing the 3' end of the dNTP and allowing for the next step of sequencing. These sequencing steps are repeated for ~300 rounds, generating sequences that are ~300 bp long that need to be processed to reconstruct the original sequence. The main advantages of Illumina sequencing include its high base-by-base accuracy, with an error rate of 0.1%, and the ability to perform millions of parallel sequencing reactions in a single slide.

Figure 5.1: Summary of Illumina sequencing

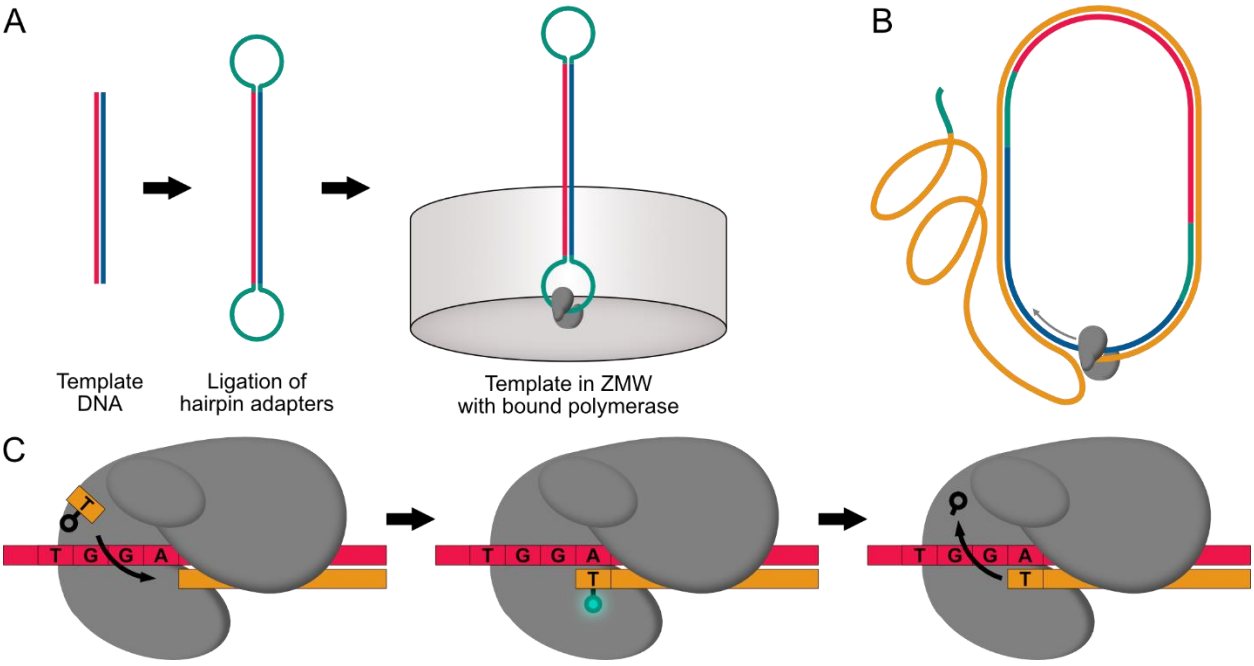


A) Schematic of "bridged amplification". Template DNA (red) anneals via an adapter sequence (green) to oligonucleotides bound to a solid support and is amplified. The newly synthesized DNA then anneals with a second oligonucleotide complementary to a second adapter sequence (blue) and is amplified once more. Several cycles result in highly dense clusters of clonal DNA. B) Sequencing primers anneal to the clonal fragments and amplification begins. Introduction of a single labeled dNTP prevents further extension. Imaging takes place and the fluorescent label is cleaved, allowing for the next step of amplification.

The third generation of NGS techniques focuses on long-read sequencing, as opposed to the short-read sequencing method employed by Illumina. One of the leaders in this field is Pacific Biosciences (PacBio) sequencing, also known as Single Molecule Real Time (SMRT) sequencing. This technique, which can sequence DNA fragments as long as 50 kb, relies on single-molecule sequencing of circularized template DNA with a DNA polymerase immobilized to the bottom of a well with characteristic dimensions (Figure 5.2).[5,6] This well is a nanoscopic structure called a zero-mode waveguide (ZMW) that causes imaging to occur only at the bottom of the well. The template preparation consists of ligating two adapter sequences, called "hairpin adapters", to the double-stranded DNA (dsDNA) template resulting in a circular single-stranded DNA (ssDNA) molecule. This ssDNA template, named SMRTbell, is then loaded onto a flow cell with many ZMWs with single DNA polymerases bound (Figure 5.2A). Sequencing is triggered by adding an oligonucleotide complementary to either adapter sequence. The DNA polymerase starts amplifying the ssDNA template and displaces the 5'-end of the primer once it

completes a full cycle. At this point, DNA polymerization continues, allowing for the synthesis of a long DNA chain that contains multiple instances of complementary fragments to both sense- and anti-sense strands of the original dsDNA (Figure 5.2B). Imaging occurs when a fluorescently labeled dNTP is incorporated into the growing DNA strand, allowing for excitation and fluorescent emission while the nucleotide is bound to the DNA polymerase (Figure 5.2C). After incorporation, the fluorescent label is cleaved and diffuses out of the ZMW detection zone, while the polymerase translocates to the next position.

Figure 5.2: Summary of SMRT sequencing (PacBio)



A) Generation of SMRTbell template. The dsDNA to be sequenced (red and blue strands, representing sense- and anti-sense strands) is ligated to hairpin adapters (green) that circularize the DNA. This ssDNA is loaded onto a SMRT cell, where single molecules enter the zero-mode waveguide (ZMW) where sequencing takes place. A single DNA polymerase immobilized to the bottom of the ZMW binds to the adapter sequence. B) Extension begins from a primer that anneals to the adapter sequence and continues producing a new strand of DNA (yellow) complementary to both sense and antisense strands. Once the polymerase reaches the 5'-end of the primer, it displaces the primer and continues with DNA extension. C) Incorporation of fluorescently labeled dNTPs generates a readout that only happens at the bottom of the ZMW, while the dNTP is bound to the DNA polymerase. The label is cleaved and diffuses out of the ZMW.

Some of the advantages of SMRT sequencing include the ability to produce long reads required for some sequencing applications. Additionally, no amplification of the template is needed during sample preparation, removing certain negative effects associated with PCR, such as the difficulty to amplify GC-rich regions or the introduction of unwanted mutations. Although the sequencing itself suffers from a relatively high error-rate, the fact that a single polymerase can generate multiple subreads of a single SMRTbell template allows the generation of highly accurate consensus sequences. This process results in highly accurate long reads, with up to 99.8% accuracy. Additionally, due to the imaging being real-time, differences in the rate of incorporation of dNTPs can be observed and assigned to specific DNA modifications, such as DNA methylation. These features, coupled to the high-throughput associated of this technology, makes SMRT sequencing a highly valuable technique in fields such as genome and epigenetics research.[7]

5.1.2 Introduction to machine learning

The field of machine learning (ML) is vast, with a broad range of applications, and ML algorithms are widely used both in science and in daily life.[8–14] Largely, ML can be described as an application of artificial intelligence focused on algorithms that allow computational systems to learn on their own.[15] This is done by either using data to develop predictive models or by identifying informative features in data. This section does not intend to exhaustively cover this subject, but instead aims to introduce key terms and relevant concepts to the work presented in this chapter.

ML can be broadly divided into supervised and unsupervised learning. Supervised learning consists of predictive models that fit data that has been labeled with an experimentally determined or arbitrarily assigned property. One example is the prediction of protein secondary

structure, where the labeled data originates from the analysis of protein crystal structure data in the Protein Data Bank (PDB).[16] One important limitation to supervised methods is the fact that often large training data sets are required to produce an accurate and efficient model. Unsupervised learning, on the other hand, does not use data with predefined labels. Instead, these methods aim to identify common patterns among the data. This process can be described as both defining labels and associating objects in the data with them. An example of such an ML approach predicted mutation effects from gene sequence co-variation.[17]

A common ML algorithm is a linear regression. Regression algorithms explore the relationship between an observed dependent variable and one or more predicted independent variables. Linear regressions assume that no relationship exists between the different independent variables. When applying regression algorithms to systems that deviate from this assumption, different methods should be considered.[18,19] While the variables being predicted in linear regression are continuous, regression algorithms can also utilize non-continuous variables. One such case is an ordinal regression model, in which the variable being predicted is discrete and has an inherent order.[20] These models are can address problems in which, for instance, different levels of performance are being evaluated.

More complex ML methods involve the use of artificial neural networks (ANNs).[21] These models are based on the connectivity and behavior of biological neural networks, with interconnected nodes arranged in a layered configuration. Each node is a mathematical function that transforms input values into output values. Thus, the output values of one layer of nodes becomes the input for the next layer. These nodes and layers can then be arranged in different configurations, or architectures. One of the most basic architectures is a fully connected layered ANN. In this configuration every node in any given layer is connected to all the nodes in the next

layer. Training of these ANNs consists of optimizing the weights associated with the connection between node pairs. A more sophisticated ANN architecture is a convolutional neural network (CNN). CNNs are applied to data with local structure information, such as 2D images or 3D structures, and focus on identifying that local structure. The input to a convolutional layer is locally subjected to a single layer fully connected ANN, or "filter". This filter is used throughout the entire input data, resulting in an output that also contains local structure information. CNNs have been applied to a diverse set of data, with one example being the AlphaFold system for the prediction of 3D protein structures by modeling the distance between pairs of residues.[22]

5.1.3 Machine learning-aided protein engineering

The goal of protein engineering is to find protein sequences with a desired function or improvement. One of the major challenges to overcome is the vast size of the protein sequence space and how to efficiently explore it. A common example to help visualize this challenge compares the 20^{100} ($\sim 10^{130}$) possible protein sequences for a 100-residue protein and the 10^{80} estimated number of atoms in the universe.[23] Experimentally exploring these sequence spaces exhaustively is impossible. Additionally, functional proteins are not frequent in a given sequence space, with most of the sequences being poorly functional or completely nonfunctional.[24] Directed evolution addresses this challenge by locally exploring the sequence space via iterative rounds of genetic diversification and selection (Figure 1.1).

However, directed evolution is not the perfect solution to the challenge of sampling large sequence spaces. For starters, when considering the fitness landscape in the sequence space, directed evolution can result in arrival to local maxima of fitness that make it very difficult to move away from.[25,26] This is a result of the fact that most evolution campaigns will make discrete movements in the sequence space around the parent sequence towards beneficial

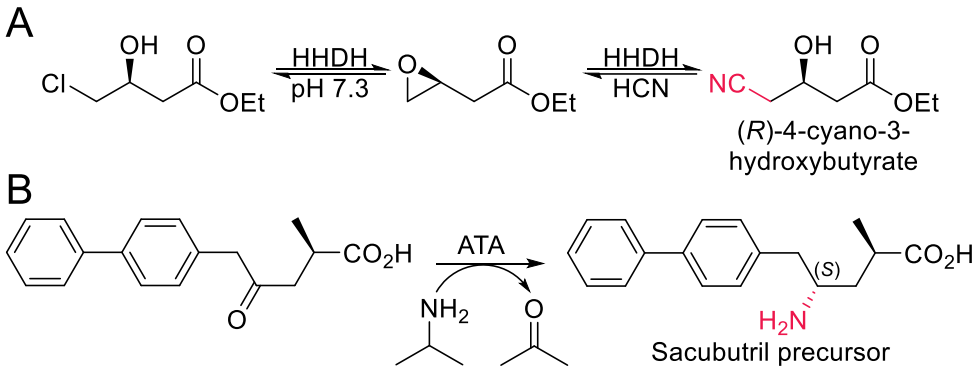
mutations. Additionally, directed evolution requires a parent enzyme with a minimum level of functionality, as well as a relatively smooth fitness landscape, since a "rugged" landscape with densely packed local maxima and minima would be challenging to map experimentally.[25] Furthermore, even if large libraries were made to exhaustively screen the sequence space, experimental screening remains the main bottleneck that limits the applicability of such approaches.

Several computational approaches to aid directed evolution have emerged, with ML being one of the most recent ones.[26–28] Directed evolution usually discards protein sequences that result in no fitness improvement and focuses only on those mutations that are beneficial. ML can instead take advantage of all the data collected during directed evolution and use it to access higher levels of fitness more rapidly. This is generally done by modeling the fitness landscape through learning from sequence-fitness data pairs. New sequences can be predicted by this ML model, allowing to further explore the sequence space, without needing to exhaustively screen said space. One of the most attractive advantages of applying ML to directed evolution is that the only associated costs correspond to DNA sequencing and computational processing, with no additional screening efforts. This is of special relevance when the screening methods are expensive or slow, or when the system being evolved has a high complexity.

An early example of ML-aided protein engineering was the evolution of a halohydrin dehalogenase for the industrial synthesis of ethyl (*R*)-4-cyano-3-hydroxybutyrate (Scheme 5.1A).[29] This study was based on protein sequence activity relationships (ProSAR) and combined the screening of combinatorial libraries with a partial least-squares (PLS) linear regression model.[30] This algorithm results in a regression coefficient for each mutation that represents its contributions to fitness independent of the effects of other mutations. During each round of

optimization, 10 to 50 mutations of interest obtained by various diversification strategies were used to generate random combinatorial libraries. After evaluating the activity of these libraries, a partial population was sequenced, with the resulting data serving as the training set for the model. The modeled coefficients were used to determine if a mutation was beneficial, potentially beneficial, neutral, or deleterious. Beneficial mutations were directly added to the next parent, potentially beneficial mutations were re-evaluated, and neutral or deleterious mutations were discarded. This process was conducted for 18 rounds, resulting in 35 mutations being introduced, and a ~4,000-fold improvement in activity.

Scheme 5.1: Reactions catalyzed by enzymes improved with the aid of ProSAR



A) Biocatalytic synthesis of (*R*)-4-cyano-3-hydroxybutyrate with an engineered halohydrin dehalogenase (HHDH).[29] B) Evolved amine transaminase (ATA) catalyzes the stereoselective synthesis of a sacubutril precursor from a non-native substrate.[31]

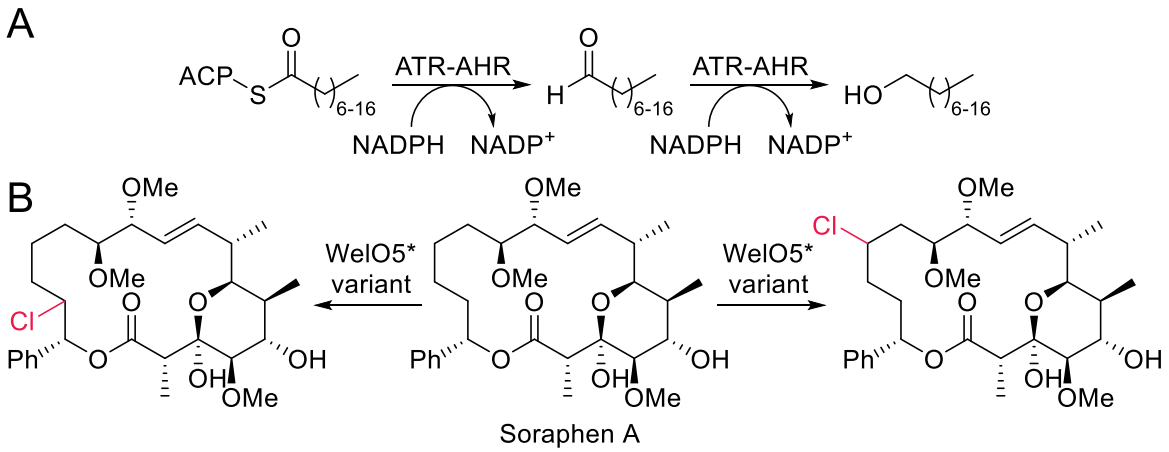
ProSAR was also used to improve the stability of a carbonic anhydrase from *Desulfovibrio vulgaris* to tolerate temperatures of up to 107 °C and a pH higher than 10.[32] This enzyme was used for large scale carbon capture from flue gas, a process that requires the enzyme to be exposed to basic conditions and a temperature of 87 – 100 °C. When this evolved variant was used in pilot scale, a 25-fold improvement in CO_2 absorption rate versus the noncatalyzed reaction was observed. The overall improvement in stability, when combining both thermostability and alkali tolerance, was ~4,000,000-fold with respect to wild-type carbonic anhydrase. Additionally, ProSAR aided the evolution of an amine transaminase (ATA) for the

synthesis of a precursor of the API sacubutril (Scheme 5.1B).[31] This reaction required activity on a non-native substrate, so an initial screen of commercial ATAs was used to identify a parent enzyme with activity on the substrate. This variant only produced trace amounts of the desired product, and its stereoselectivity produced exclusively the undesired (*R*)-product. After eleven rounds of directed evolution, conversion to product improved from <0.1% to 90%, and the stereoselectivity was enhanced to >99.9:0.1 d.r. towards the desired (*S*)-product.

Most of the reported protein engineering efforts that employ ProSAR involved screening assays with high throughput at a large scale, focusing on parallel and rapid iteration of protein optimization. There are cases in which this is not possible due to the experimental constraints associated with the screening method. A more recent study used a different ML algorithm to aided in the evolution of an acyl-ACP reductase for fatty alcohol production (Scheme 5.2A).[33] The gas chromatography assay that was employed to detect fatty alcohols was of a low enough throughput that screening a large amount of sequences was not viable. To address this challenge, this study relied on a Gaussian process (GP) that can predict the fitness landscape and estimate the model's uncertainty. An initial input of 20 experimentally determined sequence-function pairs was used to train the GP model and visualize the confidence intervals of the predictions. The next sequences to be tested were designed based on an upper-confidence bound criterion that selected both sequences predicted to be beneficial and regions of high uncertainty in the fitness landscape simultaneously.[34] This criterion was used to design 10 to 12 new sequences which would be tested experimentally. The new data was used to update the GP model, repeating the train-learn-design cycle for each round of evolution. A variant was obtained after 10 rounds of optimization that showed an increased *in vivo* fatty alcohol titer by 2-fold when compared to the

best of the wild-type parent sequences. A similar ML algorithm had been previously applied to enhancing the thermostability of cytochromes P450.[35]

Scheme 5.2: Reactions catalyzed by enzymes improved by Gaussian process models



A) An evolved chimeric acyl-thioester reductase-aldehyde reductase (ATR-AHR) acyl-ACP reductase catalyzes fatty alcohol production *in vivo* from acyl-ACP substrates.[33] B) FeDH WelO5* is evolved to chlorinate non-native substrate soraphen A with divergent site selectivities.[36]

Site selectivity has also been evolved with assistance from ML models. A recent study engineered variants of FeDH WelO5* to site selectively chlorinate the non-native substrate soraphen A (Scheme 5.2B).[36] Three residues were selected as critical to improve activity on soraphen A and complete randomization of those position was carried out. To reduce the experimental burden of screening a library with a size of 20^3 variants, a GP model was applied. Only 504 sequence-function pairs were determined experimentally and used to train this model. Predictions were made for both variants with higher activity and higher site selectivities. This process resulted in WelO5* variants with good activity on this non-native substrate and complementary site selectivity, with the most active variant showing native-like activity levels.

Finally, more complex ML-algorithms have been used to aid directed evolution. Improvement of the thermostability of hydrolases for PET depolymerization was carried out with the use of MutCompute, which is a 3D self-supervised CCN.[37,38] This algorithm does not rely on experimental sequence-fitness data sampled from a library of protein variants. Instead, it captures

biophysical and biochemical information from experimentally determined protein structures in the PDB. MutCompute is trained on more than 19,000 protein structures to learn the chemical microenvironments of amino acids and predict positions in a protein in which amino acids are not optimized. This algorithm also provides a probability distribution for the fitness of all canonical amino acids at every position. This process was applied to several PET-hydrolyzing enzymes (PETases), resulting in 29 different sequences being identified as potentially being improved. Experimental characterization of these variants yielded enzymes with an increase in T_m of up to 10 °C, increased protein yields of up to 3.8-fold, and improved activity levels of up to 29-fold.

The examples presented in this section aim to illustrate the current state of ML-guided directed evolution. A wide range of ML algorithms have been employed for different evolution campaigns, highlighting the importance of considering different approaches to ML depending on the nature of the evolution project. Factors such as the experimental burden of screening and sequence synthesis need to be evaluated when selecting the right ML technique. Additionally, novel approaches to reduce the challenges of screening and sequencing will be key to developing strategies that are affordable and efficient. Efforts to develop such a workflow, that integrates directed evolution, NGS, and ML, are summarized in the following sections.

Authorship

This work is part of a collaboration with Prof. Philip Romero. Dr. Sameer D'Costa processed the NGS data. Additionally, Dr. D'Costa constructed the ML models applied to said data and provided the corresponding predictions to be experimentally tested. Synthesis of substrate **1a** was carried out by Dr. Natalie H. Chan.

5.2 Results and discussion

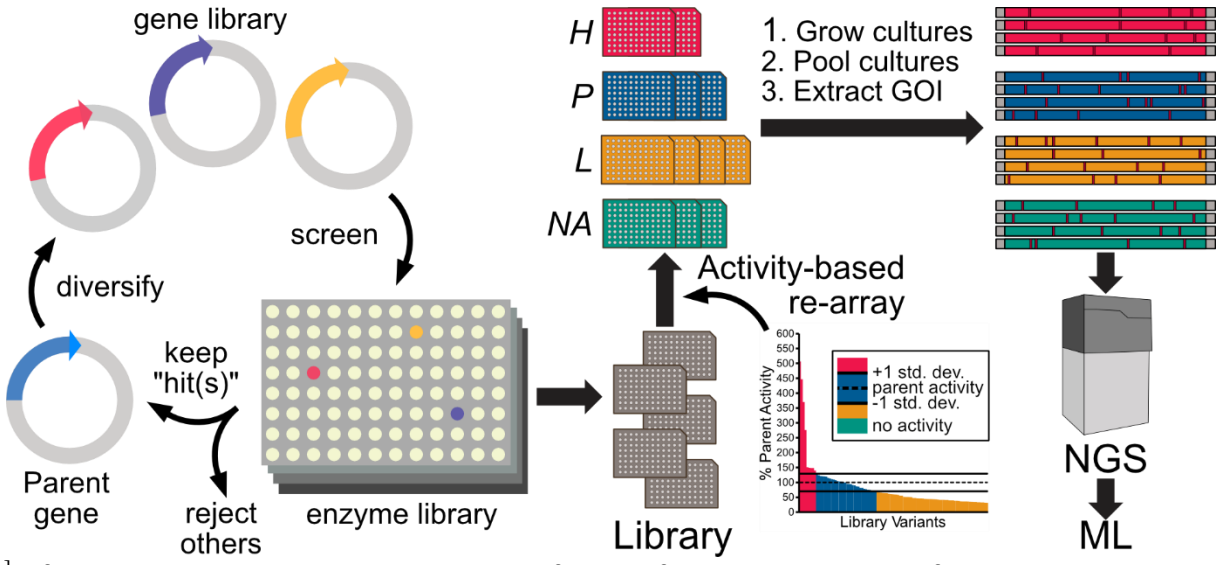
5.2.1 Workflow to incorporate NGS to directed evolution

One of the challenges of applying ML to directed evolution is the sequencing burden. Even though NGS allows for high throughput DNA sequencing, depending on library sizes, sequencing every library variant can become costly. A workflow that reduced the total number of samples to be sequenced was designed, with the main goals being to be able to reduce experimental work associated with sample preparation prior to NGS and maintain a cost-effective approach. This strategy is based on establishing an ordinal problem. Instead of having individual sequences with specific continuous labels, function data can be used to determine ordinal categories to which multiple variants belong.

In the current study, this methodology was applied to the analysis of libraries resulting from the directed evolution of engineered FeDH SadX described in chapter 3 (Figure 5.3). During this evolution campaign, variants were screened for the non-native azidation of substrate *N*-succinyl-L-leucine (**1a**) to form the corresponding γ-azidated product (**1e**). The screening data was used to categorize every sample according to the level of conversion to **1e** relative to the corresponding parent enzyme. Four different categories, or "bins", were defined using the average conversion to **1e** from parent enzyme and the corresponding standard deviation as thresholds. Variants that showed conversion to **1e** higher than the average of parent conversion (μ) plus one standard deviation (σ) were categorized as having *high* activity. Variants that fell within the range between $\mu \pm \sigma$ were placed under the *parent* activity category. Variants with conversion values lower than $\mu - \sigma$ were classified as having *low* activity. Finally, those variants with conversion to **1e** lower than 0.1% were considered to have *no activity*. The definition of these bins was arbitrary, and one can imagine different thresholds being used, as well as a

different number of bins. Furthermore, a similar binning approach could be used to analyze other features besides conversion, such as site selectivity, chemoselectivity, and enantioselectivity.

Figure 5.3: General scheme of the workflow to obtain NGS data from screened libraries to aid directed evolution[a]



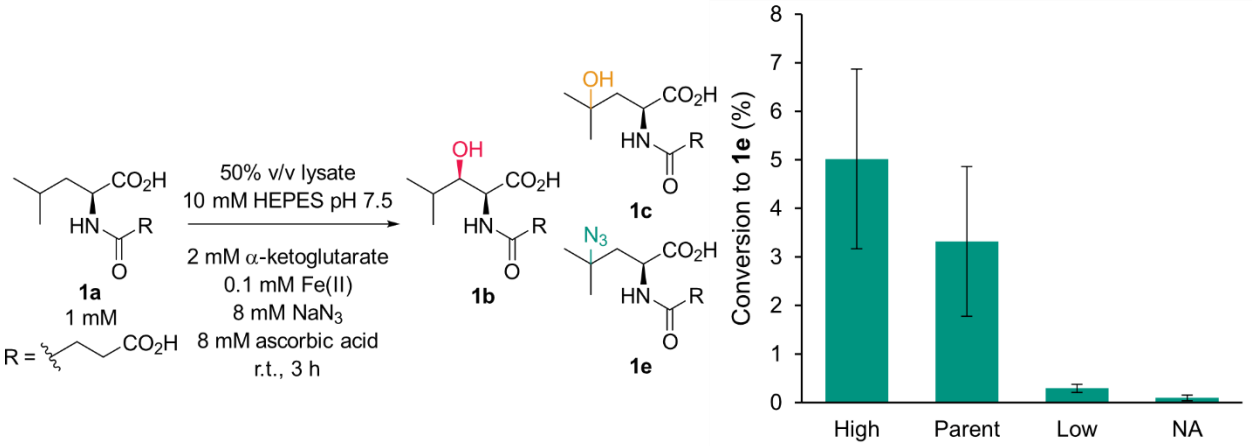
[a] After experimentally screening libraries of diversified variants as part of directed evolution, the screening data (%parent activity in this example) is used to re-array every variant into new culture plates according to defined bins (“H”, “P”, “L”, and “NA”, corresponding to high, parent, low, and no activity, respectively). The re-arrayed cultures are then incubated and pooled together, so that a single culture per bin remains. The plasmids containing the genes of interest (GOI) are extracted from these mixed samples, and the corresponding genes are isolated via restriction enzyme digestion. After purification, these samples are submitted to NGS, allowing to obtain large amounts of sequence data linked to a level of function that can be applied to train a ML model.

Once every variant was categorized, a re-array step was needed to physically sort them. This sorting was performed by using colony picker to inoculate new 96-well plates from stored cultures of the screened libraries. The instrument was programmed to only re-arrayed cultures that belonged to a given bin. Once the process was carried out with all storage plates, this was repeated for the remaining bins. The result of this step is several groups of freshly inoculated 96-well plates, one group for each defined bin, with each group containing one or more plates with the corresponding re-arrayed variants. These cultures were incubated overnight to allow for replication of the plasmids containing the genes of interest. All cultures belonging to the same

bin were pooled together into a single mixed culture after incubation, resulting in only one culture per bin. The plasmids of interest were extracted from these cultures and subjected to restriction enzyme digestion to isolate the enzyme-encoding genes. The final result is a set of four DNA samples, each containing a mixture of DNA sequences belonging to enzymes with the same level of conversion to **1e**.

The design of this workflow was concurrent with the directed evolution efforts to engineer SadX variants highly selective towards azidation. Screening an epPCR library based on variant 1-VH for improved activity on **1e** resulted in two potential improved enzymes: 2-D, with improved chemoselectivity but decreased conversion towards **1e**, and 2-L with a modest improvement in chemoselectivity but a high improvement in conversion (Figure 3.11). These two variants were used as templates for two separate epPCR libraries, and the library that originated from 2-D was used to establish the re-array workflow described above. To test the validity of this process, the entire library was re-arrayed and pooled, and the resulting pooled cultures were used to set up bioconversions with **1a** (Figure 5.4). A correlation between the conversion to **1e** and the defined level of activity was observed, with the mixed culture from the *high* bin showing the highest level of conversion, and the *low* and *no activity* bins showing the lowest levels of conversion. These results suggested that the activity levels observed during screening were reproduced to some extent when the combined activity of bins was tested.

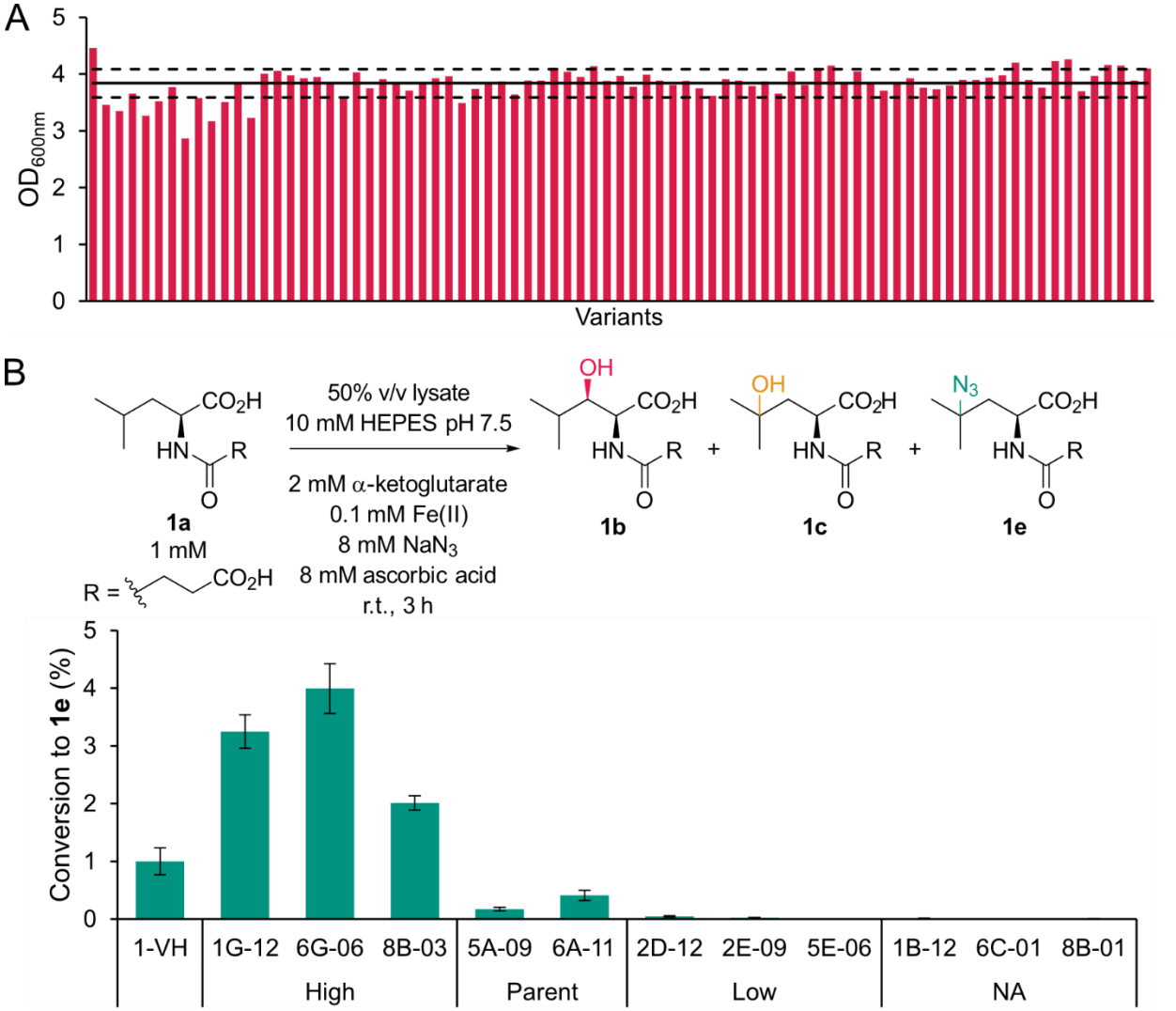
Figure 5.4: Activity of combined binned variants in lysate[a]



[a] Bioconversion of substrate **1a** to products **1b**, **1c**, and **1e** using lysates containing 2-D variants. Lysate samples contain a mixture of variants belonging to the corresponding bin. Conversion to products calculated with LC-MS extracted ion chromatogram peak areas of products and substrate. Each data point corresponds to the average of eight replicates and error bars represent the corresponding standard deviations.

Additionally, the remaining epPCR libraries that had been screened for under azidation conditions (i.e., the libraries based on variants 1-VH, 2-L, and 3-VRL) were subjected to the same workflow, and additional controls were evaluated at this time. First, differences in culture growth were investigated, with the goal to rule out a potential bias in the sequencing data due to specific cultures generating higher levels of DNA than others, misrepresenting their distribution in the pooled bin. To test this, a 96-well plate containing re-arrayed cultures from the 1-VH library was incubated overnight and allowed to grow. The optical density at 600 nm (OD_{600nm}) of the resulting cultures was obtained as a measure of bacterial growth (Figure 5.5A). These values showed a narrow distribution, with most samples resulting in OD_{600nm} values between 3.6 and 4.1. This data suggested that bacterial growth was uniform after re-array, minimizing certain sequences appearing with more frequency due to inconsistent levels of DNA replication between different samples.

Figure 5.5: Controls to validate binning of library variants



A) OD_{600nm} values for re-arrayed cultures from library 1-VH in a 96-well plate. The horizontal lines represent the average of all measurements (solid) and a standard deviation range (dashed). B) Bioconversions of **1a** to products **1b**, **1c**, and **1e** in lysate of 1-VH variants. Parent 1-VH was used as a control. Variants are grouped by bins and designated *XY-Z* according to their original location in column *Z* and row *Y* of plate *X*. *NA = no activity*. Conversion to products calculated with LC-MS extracted ion chromatogram peak areas of products and substrate. Each data point corresponds to the average of three replicates and error bars represent the corresponding standard deviations.

Additionally, random variants from all four bins from library 1-VH were selected to validate the binning process (Figure 5.5B). In contrast to the experiment shown in Figure 5.4, these bioconversions aimed to evaluate the correlation between the binning and the activity of the binned samples. This would also serve as a validation of the screening data that was used to

define the bins. Bioconversions using lysate of these randomly selected variants showed results that followed the binning definitions, with variants belonging to the *high* bin showing higher conversion to **1e** than parent 1-VH and those binned as having *parent activity* resulting in intermediate levels of conversion. While two variants from the *low* bin showed trace amounts of conversion to **1e**, the remaining one as well as all the variants from the *no activity* bin were not able to produce detectable amounts of **1e**. These controls further validated the workflow design and encouraged the sequencing of the obtained DNA samples.

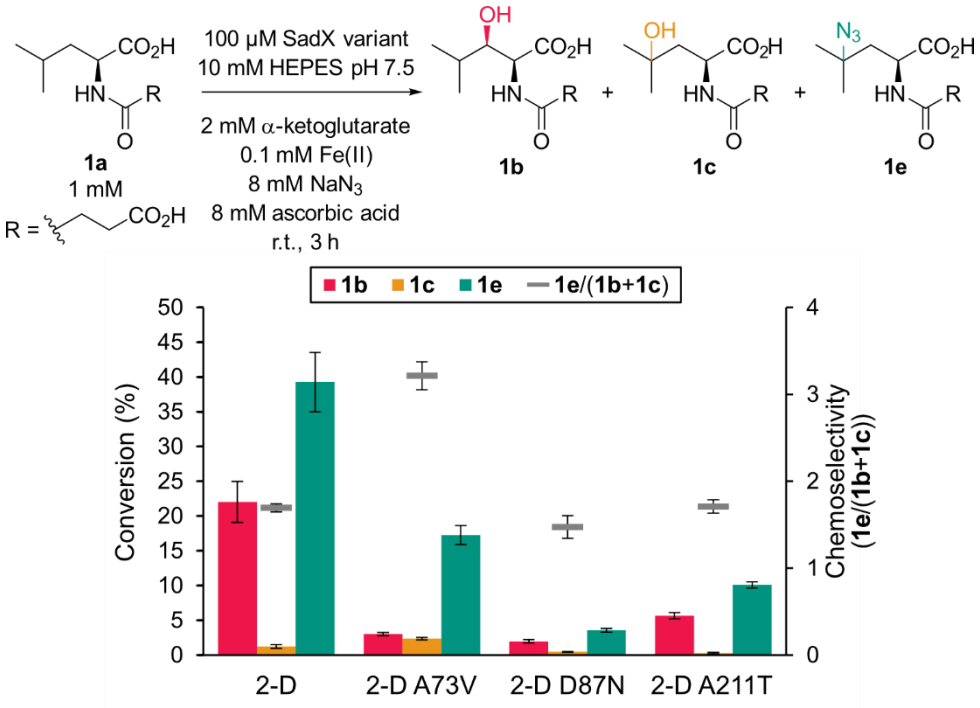
5.2.2 Validation of predicted sequences

The resulting pooled DNA samples were sequenced via SMRT (PacBio). The reason behind this decision was the ability of SMRT to generate long-reads, allowing to cover the entire sequence of SadX variants (822 bp). This results in the identification of mutations present in the different activity bins, as well as the sequence context they appear in. Additionally, the high fidelity that results from repeated subreads of the same sequence allows for accurate identification of variations as small as single mutations. Dr. Sameer D'Costa, from the Romero group, conducted the processing of sequencing data and designed an ordinal regression model with regularization trained on the variants from library 2-D. Regularization refers to techniques that are used to ensure that ML models are well behaved. The number of variables in this system is determined by the length of amino acid sequence. In the case of SadX, that is 273 positions with 20 possible amino acids, resulting in 5460 variables. The number of screened variants in each of the sequenced libraries was 900, which is substantially lower. Furthermore, regularization helps prevent the model from underfitting or overfitting the data.[39]

The model resulted in mutations that were scored against parent enzyme 2-D, with those predicted to be beneficial having a positive score, and those predicted to be detrimental having a

negative score. The expression of three different sequences containing single mutations (A73V, D87N, and A211T) was carried out to test the modeled predictions (Figure 5.6). It is worth noting that none of these mutations were previously identified during library screening since they were not present in the best performing variants that were selected for further characterization. When the activity of these variants was evaluated, no improvements in conversion to **1e** were observed, with all three variants showing lower conversion values than parent 2-D. Notably, variant 2-D A73V displayed a 1.9-fold increase in chemoselectivity towards the azidated product **1e** by decreasing the amount of β-hydroxylated product **1b** being formed. This improvement was accompanied by a decrease in conversion of ~22%.

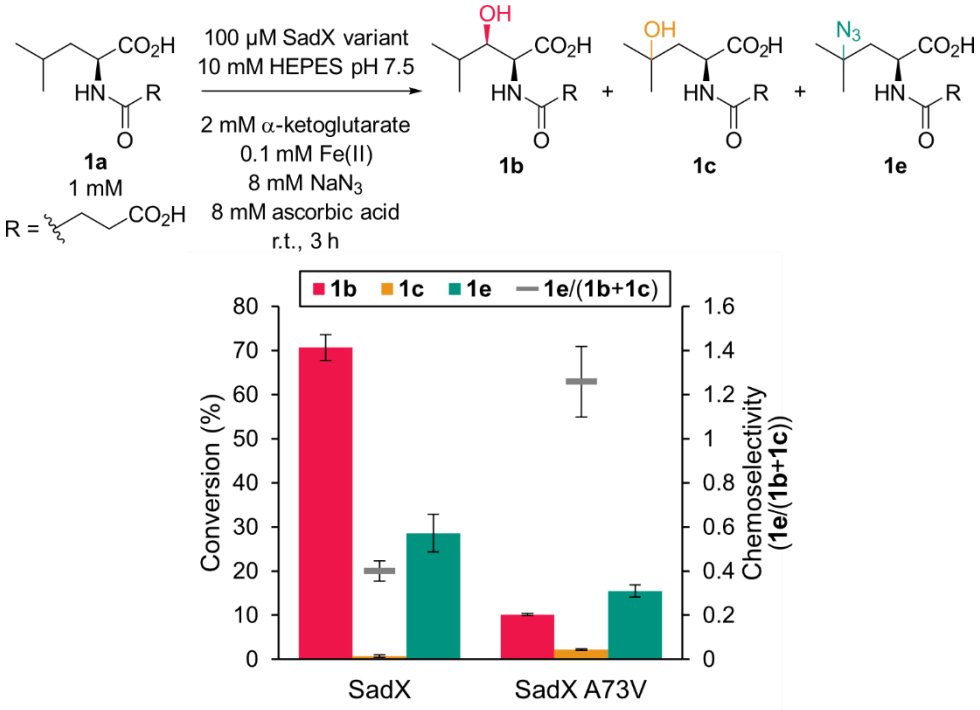
Figure 5.6: Bioconversions with predicted 2-D single mutants[a]



[a] Bioconversion of **1a** to products **1b**, **1c**, and **1e** by 2-D variants predicted to have beneficial effects. Conversion values were calculated with extracted ion chromatogram peak areas of products and the substrate from LC-MS experiments. The chemoselectivity of these variants is expressed as the ratio of conversion to product **1e** over the sum of conversions to products **1b** and **1c**. Each data point corresponds to the average of three replicates and error bars represent standard deviations.

Even though the regression model had been trained on data that was sorted according to levels of azidation, the beneficial mutation that was found had a negative impact on this feature, and instead, was beneficial towards chemoselectivity. Despite this discrepancy, the level of fold improvement in chemoselectivity was similar to that observed for the first mutant in the azidation lineage, 1-VH. Additionally, residue Ala_{73} is located near Ile_{71} which was mutated in 1-VH. To further characterize this mutation, A73V was introduced in SadX, showing an even higher improvement in chemoselectivity of 3.1-fold (Figure 5.7). This enhanced selectivity was accompanied by a ~13% decrease in conversion to **1e**, mirroring the effect that A73V displayed when introduced in 2-D. Furthermore, the identification of mutation A73C suggests that other residues in the missing loop Ser_{59}-Ser_{75} may have a strong effect on the chemoselectivity towards non-native rebound.

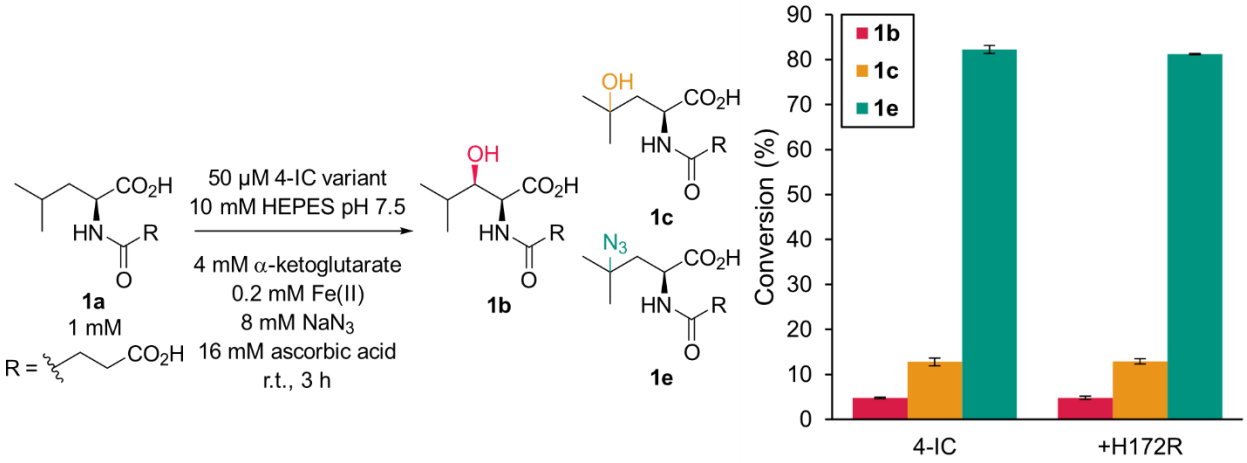
Figure 5.7: Bioconversion with new mutant SadX A73V[a]



[a] Bioconversion of **1a** by SadX and SadX A73V. Conversion values were calculated with extracted ion chromatogram peak areas from LC-MS experiments. The chemoselectivity of these variants is expressed as the conversion to product **1e** over the sum of conversions to products **1b** and **1c**. Each data point is the average of three replicates and error bars are standard deviations.

Since library 2-D resulted in a branch of directed evolution that did not provide an improved variant, the overall sequence-function data could be viewed as biased towards variants with low variability, not accurately representing the sequence space of SadX variants. In order to further this, the three remaining libraries (1-VH, 2-L, and 3-VRL) were analyzed, providing a larger number of sequences and larger changes in changes in activity. Applying the same ordinal regression to this new set of data resulted in new mutations predicted to be beneficial. Notably, mutation H172R, which was a reversion of a mutation introduced in the first round of evolution, was predicted to be beneficial. When this reversion was introduced into the most evolved azidase, 4-IC, no improvement was observed. Both 4-IC and 4-IC H172R showed high levels of conversion to **1e** with high chemoselectivity. However, this suggested that the original mutation R172H was not needed, so the reversion was kept to minimize the mutational burden on the protein scaffold.

Figure 5.8: Bioconversion with predicted reversion H172R[a]

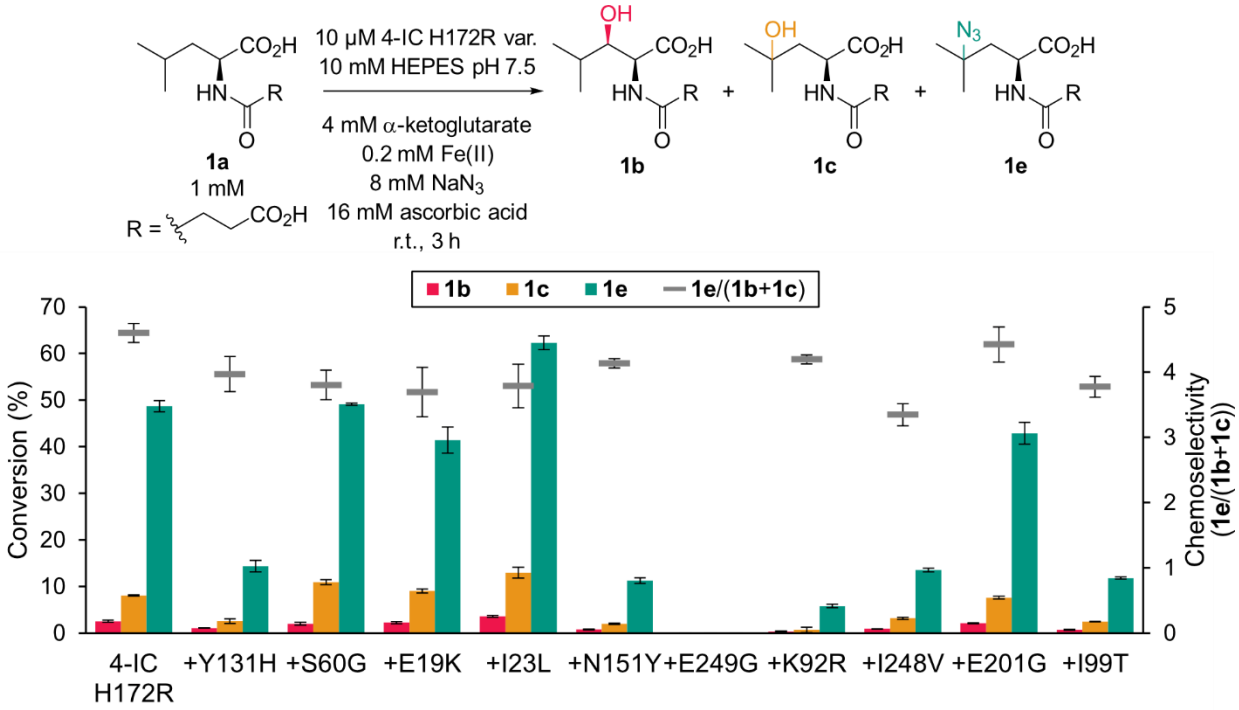


[a] Bioconversion of **1a** by 4-IC and 4-IC H172R. Conversion values were calculated with extracted ion chromatogram peak areas of products and the substrate from LC-MS experiments. Each data point corresponds to the average of three replicates and error bars represent standard deviations.

Ten additional single mutations predicted to be beneficial were introduced into the new 4-IC H172R scaffold (Figure 5.9). These sequences were modeled using individual ordinal regressions for each library, and they were selected based on to different criteria: top variants to

be predicted by any model to be beneficial (Y131H, S60G, E19K, I23L, and N151Y), and top variants predicted to be beneficial by all three models (E249G, K92R, I248V, E201G, and I99T). When bioconversions were carried out with these new variants, distinct trends could be observed for these two subgroups of predictions. In general, mutations that were predicted by all three models showed worse activity than those predicted to be beneficial by individual models. Several variants in the former group showed conversion values lower than 10%, while most in the latter showed conversions that equaled or surpassed the parent enzyme. No significant improvements in chemoselectivity were observed. Mutation I23L, however, showed a ~14% increase in conversion to **1e** with respect to the parent 4-IC H172R under the reaction conditions. This modest but significant improvement was caused by a mutation that was not identified during screening, highlighting once more the strength of ML to uncover beneficial sequences.

Figure 5.9: Bioconversions with predicted 4-IC single mutants[a]



[a] Bioconversion of **1a** by 4-IC H172R variants. Conversion values were calculated with extracted ion chromatogram peak areas from LC-MS. The chemoselectivity is expressed as the conversion to **1e** over the sum of conversions to **1b** and **1c**. Each data point is the average of three replicates and error bars are standard deviations.

Overall, however, most of the single mutations introduced into 4-IC H172R were not beneficial, with most of them resulting in an abrupt decrease in conversion to **1e**. Additionally, the one beneficial mutation gave only a low improvement in activity. Although seemingly discouraging, it is worth noting that once the ordinal regression models were constructed, the experimental screening effort to find variant 4-IC I23L/H172R was less than 2% of the effort associated with conducting another round of directed evolution via epPCR with a 900-variant library. ML-aided approaches such as the one presented in this study therefore have the potential to increase the provide additional improved sequences without requiring further experimental screening. Whereas the predicted sequences were introduced as single point mutations, these could be used combinatorially to further explore the space of predicted sequences. Furthermore, the positions of modeled mutations could be used to design site-saturation mutagenesis libraries, as an alternative to rational design. These approaches are being explored by other members of the Lewis group to further explore the application of the generated ML models. Finally, more complex ML algorithms are being applied by the Romero group, with a focus on self-supervised learning, allowing to significantly decrease the number of experimental data needed to train sufficiently accurate models.

It is worth mentioning that while this study was in progress, a report of a different methodology to incorporate NGS in directed evolution in a cost-efficient manner was published.[40] This method, named evSeq, involved the careful amplification and barcoding of every individual sequence in a library, followed by multiplexed Illumina sequencing. The strength of this technique lies in the ability to obtain sequence-function relationships for every variant, but it also presents a few challenges, mostly associated with the DNA amplification step. The introduction of a PCR step may introduce complications that are sequence dependent, such

as poor amplification of CG rich regions, as well as increase the likelihood of non-specific mutations. Additionally, careful planning of a system of barcoded primers is needed to maintain an overall low cost. The approach developed in this study and presented in this chapter has the potential to complement techniques such as evSeq, with drastically reducing the experimental requirements to prepare samples for NGS while still being able to generate data of high enough quality to construct successful ML models.

5.3 Conclusions

The development of a methodology to incorporate NGS into directed evolution was presented in this chapter. Screening data was used to define categories with different levels of activity and assign every library variant to one of those categories. Re-arraying and pooling all variants according to their corresponding category results in a drastic reduction in the number of samples while still maintaining sequence-function information. Preparation of NGS samples consisted in a simple plasmid isolation step followed by a reaction with restriction enzymes. This process was applied to four epPCR SadX libraries previously screened. The design of this workflow is part of a collaboration with Prof. Philip Romero and Dr. Sameer D'Costa, who contributed by developing an ordinal regression ML algorithm based on the NGS obtained from the processed libraries. Validation of this model was carried out, being able to identify a variant with increased chemoselectivity towards azidation of the model substrate and another variant with increased conversion to the desired azidated product. These promising results will be further developed by the Lewis and Romero groups in order to create data-efficient ML-assisted protein engineering strategy.

5.4 Experimental

5.4.1 Materials

SadX (MBP-SadA D157G fusion) and evolved variants 2-D and 4-IC were obtained and expressed from previously generated plasmids, as described in chapters 2 and 3.[41,42] BL21-Gold (DE3) *E. coli* cells (# 230132) were purchased from Agilent Technologies. Lysozyme from chicken egg white (#L6876) and deoxyribonuclease I from bovine pancreas (DNase, #DN25) were purchased from MilliporeSigma. Restriction enzymes SacI-HF (#R3156), XhoI (#R0146), and DpnI (#R0176) were purchased from New England Biolabs. PrimeSTAR Max DNA polymerase mix (#R045A) was purchased from Takara Bio USA. Oligonucleotides were purchased from MilliporeSigma. Miniprep kits (QIAprep Spin Miniprep Kit, #27104) and agarose gel extraction kits (QIAquick Gel Extraction Kit, #28704) were purchased from QIAGEN and used following the manufacturer's protocols. DNA clean and concentrator kits (#D4003) were purchased from Zymo Research and used following the manufacturer's instructions. Terrific broth (TB, #T15000) and Luria broth (LB, #L24040) were purchased from Research Products International and prepared following the manufacturer's protocols.

Kanamycin monosulfate (#J61272) was purchased from Alfa Aesar and prepared into a 50 mg/mL aqueous 1000x stock solution. Isopropyl-β-D-thiogalactopyranoside (IPTG, #00194) was purchased from Chem-Impex International. Ni-NTA resin (HisPur™ Ni-NTA Resin, #88223) was purchased from Thermo Scientific.

5.4.2 General procedures

Library re-array was set up using a Multidrop Combi liquid dispenser and a NorgrenSystems CP-7200 colony picker. DNA and protein concentrations were measured with a Tecan Infinite 200 PRO plate reader. DNA amplification by PCR was performed on an Applied

Biosystems ProFlex PCR System thermocycler. DNA transformation was carried out by electroporation with a Bio-Rad MicroPulser. Cell lysis by sonication was performed using a QSonica S-4000 sonicator with a 0.5" horn.

LC-MS analysis was carried out using an Agilent 1290 Infinity II system connected to a ZORBAX Eclipse Plus C18 column (2.1 mm x 50 mm, 1.8 µm particle size), an Agilent Jet Stream (AJS-ES) ion source, and a 6135 single quadrupole mass selective detector. LC-MS data analysis was carried out with MassHunter (Agilent) software.

5.4.3 Chromatographic method

LC-MS analysis was conducted using 0.1 % formic acid in water (A) and 0.1 % formic acid in acetonitrile (B) as mobile phase components. Table 5.1 shows the *m/z* values used to obtain EICs from each experiment.

Solvent gradient: 5 %B (0 – 1.25 min), 5 to 14 %B (1.25 – 2 min), 14 %B (2 – 4.5 min), 14 to 90 %B (4.5 – 6.8 min), 90 %B (6.8 – 7 min), 5 %B (7 – 7.25 min). Flow rate: 0.4 mL/min. ESI negative mode.

Table 5.1: Dominant ions for relevant substrate and products

Species	***m/z***
1a	230
1b, 1c	246
1e	271

Extracted ion chromatograms (EICs) were extracted for **1a**, **1b**, **1c**, and **1e** for each sample according to Table 5.1. The area of the corresponding peaks was integrated and used to represent the amount of each analyte. Percent conversion to product **1i** (**i** = **b**, **c**, **e**) was calculated as the peak area of product **1i** divided by the sum of the peak area of **1a** and the peak areas of all products (1). Chemoselectivity was defined as the ratio of azidated product to hydroxylated products (2).

$$\%Conversion\ to\ \mathbf{1i} = 100\% \times \frac{Area_{\mathbf{1i}}}{(Area_{\mathbf{1a}} + Area_{\mathbf{1b}} + Area_{\mathbf{1c}} + Area_{\mathbf{1e}})} \quad (1)$$

$$Chemoselectivity = \frac{Area_{\mathbf{1e}}}{(Area_{\mathbf{1b}} + Area_{\mathbf{1c}})} \quad (2)$$

5.4.4 Molecular cloning

Standard molecular cloning procedures were used throughout this study.[43] Point mutations were introduced via site-directed mutagenesis following a modified QuickChange™ protocol, using the corresponding parent-containing pET28a plasmids as templates.[44] The oligonucleotides used to perform these reactions are shown in Table 5.2. The amplification conditions were the following: 2 ng/µL template DNA, 0.3 µM forward primer, 0.3 µM reverse primer, and 1x PrimeSTAR Max DNA Polymerase, with the final volume being 50 µL. The PCR program for the amplifications was the following: 98 °C for 10 seconds; 30 cycles of 98 °C for 10 seconds, 55 °C for 5 seconds, and 72 °C for 45 seconds; and ending with 72 °C for 1 minute. The PCR products were digested with 10 units of DpnI at 37 °C for 1 hour and purified via 1% agarose gel electrophoresis.

Table 5.2: Oligonucleotides used to introduce predicted mutations

#	Name	Sequence
1	2-D_A73V_FP	5’ – GAT GCA CAT CAT GCA GGT GTT GTT GTG ACC AGC GTT ACC CTG GGT CAG C – 3’
2	2-D_A73V_RP	5’ – CAC CTG CAT GAT GTG CAT CTT TTG TAA ACG GGC TGC TGC TAA AAC CTT C – 3’
3	2-D_D87N_FP	5’ – CTG GGT CAG CTG CAG CGT GAA CAG GGT AAC AAA CTG GTT AGC AAA GCA GC – 3’
4	2-D_D87N_RP	5’ – CAC GCT GCA GCT GAC CCA GGG TAA CGC TGG TTG CAA CAA CAC CTG CAT G – 3’
5	2-D_A211T_FP	5’ – CGT GCA GCG CTG GAT GAA ATG CAT ACC GAA TAT CGT GAA ACC GGT GC – 3’
6	2-D_A211T_RP	5’ – CAT CCA GCG CTG CAC GGC TTT CCG GAC GAT TAT CCC ACA TAA CCA TAC C – 3’
7	SadX_A73V_FP	5’ – CCA AAA ATG CAC ATC ATG CAG GTA TTG TTG TGA CCA GCG TTA CCC TGG – 3’

Table 5.2 (*continued*)

#	Name	Sequence
8	SadX_A73V_RP	5’ – CCT GCA TGA TGT GCA TTT TTG GTA AAC GGG CTG CTG CTA AAA CCT TC – 3’
9	4-IC_H172R_FP	5’ – CTT TCC GCG TCA GCT GGG TGG TTT TCT GAC CAT TCA GGG TGC AGA TAA TG – 3’
10	4-IC_H172R_RP	5’ – CCA CCC AGC TGA CGC GGA AAG CTA CGT TTC AGA GGC CAA TTA ACG – 3’
11	Y131H_FP	5’ – GTT GAA TTT CTG CTG GCA ACC CGT CGT GTT TAT GAA CCG TTT GAA GCA C – 3’
12	Y131H_RP	5’ – GGT TGC CAG CAG AAA TTC AAC TTC ATG ATC ATG ACC AAT GGT CAGC – 3’
13	S60G_FP	5’ – GTT TAC CAA AAA TGC ACA TCA TGC AGG TGT TGT TGC AAC CAG CGT TAC C – 3’
14	S60G_RP	5’ – GCA TGA TGT GCA TTT TTG GTA AAC GGG CCG CTG CTA AAA CCT TCA TC – 3’
15	E19K_FP	5’ – GAT GCG TTT TGG CAC CGC AGC ACG TGC AAA ACA TAT GAC CAT TGC AGC – 3’
16	E19K_RP	5’ – CGT GCT GCG GTG CCA AAA CGC ATC AGC TGT GCC GGA TAG GTA TGC – 3’
17	I23L_FP	5’ – CAC CGC AGC ACG TGC AGA ACA TAT GAC CCT GGC AGC AGC AAT TCA TGC – 3’
18	I23L_RP	5’ – CAT ATG TTC TGC ACG TGC TGC GGT GCC AAA ACG CAT CAG CTG TGC CGG ATA G – 3’
19	N151Y_FP	5’ – CGT GTT TAT GAA CCG TTT GAA GCA CCG TTT TAT CTT GCA CCG CAT TG – 3’
20	N151Y_RP	5’ – GCT TCA AAC GGT TCA TAA ACA CGA CGG GTT GCC AGC AGA AAT TCA AC – 3’
21	E249G_FP	5’ – GTT GTA CCA GCA CCC GTC GTA CCA TGG GTC TGC TTC TGA TTC ATA CC– 3’
22	E249G_RP	5’ – GGT ACG ACG GGT GCT GGT ACA ACG GCC AAT GGC ATG CAG ATT TTT GC– 3’
23	K92R_FP	5’ – GCA GCA GAA TAT TTT GGT ATT GCC TGC CGT GTT AAT GAT GGT CTG CGT ACC – 3’
24	K92R_RP	5’ – GCA GGC AAT ACC AAA ATA TTC TGC TGC ACG GCT AAC CAG TTT ATC ACC – 3’
25	I248V_FP	5’ – GAG CAA AAA TCT GCA TGC CGT GGA ACG TTG TAC CAG CAC CCG TCG – 3’
26	I248V_RP	5’ – GCA TGC AGA TTT TTG CTC TGA AAC AGT GTC AGC TGG CCA GG – 3’
27	E201G_FP	5’ – GGT ATG GTT ATG TGG GAT AAT CGT CCG GGC AGC CGT GCA GCG CTG GAT G – 3’
28	E201G_RP	5’ – CGA TTA TCC CAC ATA ACC ATA CCG GCA TCA TTA TCT GCA CCC TGA ATG G – 3’

Table 5.2 (*continued*)

#	Name	Sequence
29	I99T_FP	5’ – GCA AAG CAG CAG AAT ATT TTG GTA CCG CCT GCC GTG TTA ATG ATG G – 3’
30	I99T_RP	5’ – CCA AAA TAT TCT GCT GCT TTG CTA ACC AGT TTA TCA CCC TGT TCA C – 3’

Table 5.3 shows the corresponding parent and oligonucleotide pair for each site-directed mutagenesis reaction.

Table 5.3: List of template and oligonucleotides used for variant cloning

Variant	Template	Oligonucleotides
2-D A73V	2-D	**1, 2**
2-D D87N	2-D	**3, 4**
2-D A211T	2-D	**5, 6**
SadX A73V	SadX	**7, 8**
4-IC H172R	4-IC	**9, 10**
4-IC Y131H/H172R	4-IC H172R	**11, 12**
4-IC S60G/H172R	4-IC H172R	**13, 14**
4-IC E19K/H172R	4-IC H172R	**15, 16**
4-IC I23L/H172R	4-IC H172R	**17, 18**
4-IC N151Y/H172R	4-IC H172R	**19, 20**
4-IC H172R/E249G	4-IC H172R	**21, 22**
4-IC K92R/H172R	4-IC H172R	**23, 24**
4-IC H172R/I248V	4-IC H172R	**25, 26**
4-IC H172R/E201G	4-IC H172R	**27, 28**
4-IC H172R/I99T	4-IC H172R	**29, 30**

5.4.5 Sequences

SadX (nucleotide)

This sequence includes the MBP tag (*italicized*), the SadA D157G gene, and the 3’ flanking region of the vector (<u>underlined</u>).

ATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCA
TATGGCTAGCATGCACCATCACCATCACCATGGAAAAATCGAAGAAGGTAAACTGGTAATC
TGGATTAACGGCGATAAAGGCTATAACGGTCTCGCTGAAGTCGGTAAGAAATTCGAGAAA
GATACCGGAATTAAAGTCACCGTTGAGCATCCGGATAAACTGGAAGAGAAATTCCCACAG
GTTGCGGCAACTGGCGATGGCCCTGACATTATCTTCTGGGCACACGACCGCTTTGGTGGC

TACGCTCAATCTGGCCTGTTGGCTGAAATCACCCCGGACAAAGCGTTCCAGGACAAGCTG
TATCCGTTTACCTGGGATGCCGTACGTTACAACGGCAAGCTGATTGCTTACCCGATCGCTG
TTGAAGCGTTATCGCTGATTTATAACAAAGATCTGCTGCCGAACCCGCCAAAAACCTGGGA
AGAGATCCCGGCGCTGGATAAAGAACTGAAAGCGAAAGGTAAGAGCGCGCTGATGTTCAA
CCTGCAAGAACCGTACTTCACCTGGCCGCTGATTGCTGCTGACGGGGGTTATGCGTTCAA
GTATGAAAACGGCAAGTACGACATTAAAGACGTGGGCGTGGATAACGCTGGCGCGAAAG
CGGGTCTGACCTTCCTGGTTGACCTGATTAAAAACAAACACATGAATGCAGACACCGATTA
CTCCATCGCAGAAGCTGCCTTTAATAAAGGCGAAACAGCGATGACCATCAACGGCCCGTG
GGCATGGTCCAACATCGACACCAGCAAAGTGAATTATGGTGTAACGGTACTGCCGACCTT
CAAGGGTCAACCATCCAAACCGTTCGTTGGCGTGCTGAGCGCAGGTATTAACGCCGCCAG
TCCGAACAAAGAGCTGGCAAAAGAGTTCCTCGAAAACTATCTGCTGACTGATGAAGGTCTG
GAAGCGGTTAATAAAGACAAACCGCTGGGTGCCGTAGCGCTGAAGTCTTACGAGGAAGAG
TTGGCGAAAGATCCACGTATTGCCGCCACTATGGAAAACGCCCAGAAAGGTGAAATCATG
CCGAACATCCCGCAGATGTCCGCTTTCTGGTATGCCGTGCGTACTGCGGTGATCAACGCC
GCCAGCGGTCGTCAGACTGTCGATGAAGCCCTGAAAGACGCGCAGACTAATTCGAGCTC
*GAACAACAACAACACTAGTGAAAACCTGTATTTCCAGGGAGCAGCCGGATCC*ATGCAGCA
TACCTATCCGGCACAGCTGATGCGTTTTGGCACCGCAGCACGTGCAGAACATATGAC
CATTGCAGCAGCAATTCATGCACTGGATGCAGATGAAGCAGATGCAATTGTTATGG
ATATTGTTCCGGATGGTGAACGTGATGCATGGTGGGATGATGAAGGTTTTAGCAGCA
GCCCGTTTACCAAAAATGCACATCATGCAGGTATTGTTGCAACCAGCGTTACCCTGG
GTCAGCTGCAGCGTGAACAGGGTGATAAACTGGTTAGCAAAGCAGCAGAATATTTT
GGTATTGCCTGCCGTGTTAATGATGGTCTGCGTACCACCCGTTTTGTTCGTCTGTTTA
GTGATGCCCTGGATGCCAAACCGCTGACCATTGGTCATGATTATGAAGTTGAATTTC
TGCTGGCAACCCGTCGTGTTTATGAACCGTTTGAAGCACCGTTTAACTTTGCACCGC
ATTGTGGCGATGTTAGCTATGGTCGTGATACCGTTAATTGGCCTCTGAAACGTAGCT
TTCCGCGTCAGCTGGGTGGTTTTCTGACCATTCAGGGTGCAGATAATGATGCCGGTA
TGGTTATGTGGGATAATCGTCCGGAAAGCCGTGCAGCGCTGGATGAAATGCATGCA
GAATATCGTGAAACCGGTGCAATTGCCGCACTGGAACGTGCAGCCAAAATCATGCT
GAAACCGCAGCCTGGCCAGCTGACACTGTTTCAGAGCAAAAATCTGCATGCCATTG
AACGTTGTACCAGCACCCGTCGTACCATGGGTCTGTTTCTGATTCATACCGAAGATG
GTTGGCGTATGTTTGATTGA<u>AAGCTTGCGGCCGCACTCGA</u>

SadX (amino acid)

The sequence of MBP is *italicized*. All further amino acid indices refer to the original enzyme SadA numbering.

MGSSHHHHHHSSGLVPRGSHMASMHHHHHHGKIEEGKLVIWINGDKGYNGLAEVGKKFEK
DTGIKVTVEHPDKLEEKFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYP
FTWDAVRYNGKLIAYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPY
FTWPLIAADGGYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAF
NKGETAMTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFL
ENYLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYA
*VRTAVINAASGRQTVDEALKDAQTNSSSNNNNTSENLYFQGAAGS*MQHTYPAQLMRFGTAA

RAEHMTIAAAIHALDADEADAIVMDIVPDGERDAWWDDEGFSSSPFTKNAHHAGIVAT
SVTLGQLQREQGDKLVSKAAEYFGIACRVNDGLRTTRFVRLFSDALDAKPLTIGHDYEV
EFLLATRRVYEPFEAPFNFAPHCGDVSYGRDTVNWPLKRSFPRQLGGFLTIQGADNDAG
MVMWDNRPESRAALDEMHAEYRETGAIAALERAAKIMLKPQPGQLTLFQSKNLHAIER
CTSTRRTMGLFLIHTEDGWRMFD*

1-VH
SadX I71V (V = GTT) R172H (H = CAT)

2-L
1-VH F152L (L = CTT)

2-D
1-VH N65D (D = GAT)

3-VRL
2-L I38V (V = GTT) Q233R (R = CGC) F261L (L = CTT)

4-IC
3-VRL V38I (I = ATT) R48C (C = TGT)

2-D A73V
2-D A73V (V = GTG)

2-D D87N
2-D D87N (N = AAC)

2-D A211T
2-D A211T (T = ACC)

SadX A73V
SadX A73V (V = GTG)

4-IC H172R
4-IC H172R (R = CGT)

4-IC Y131H/H172R
4-IC H172R Y131H (H = CAT)

4-IC S60G/H172R
4-IC H172R S60G (H = GGC)

4-IC E19K/H172R
4-IC H172R E19K (K = AAA)

4-IC I23L/H172R
4-IC H172R I23L (H = CTG)

4-IC N151Y/H172R
4-IC H172R N151Y (Y = TAT)

4-IC H172R/E249G
4-IC H172R E249G (G = GGC)

4-IC K92R/H172R
4-IC H172R K92R (K = CGT)

4-IC H172R/E248V
4-IC H172R E248V (V = GTG)

4-IC H172R/E201G
4-IC H172R E201G (G = GGC)

4-IC I99T/H172R
4-IC H172R I99T (T = ACC)

5.4.6 Gene expression and protein purification

Large scale protein expression

A 5 µL aliquot of the resulting plasmid was used to transform 50 µL of BL21-Gold (DE3) *E. coli* via electroporation. Cells were resuspended with 945 µL of SOC media and shaken at 37 °C and 250 rpm for 45 minutes. Up to 200 µL of the culture were transferred to circular LB/agar plates with 50 µg/mL kanamycin. The plates were incubated overnight at 37 °C. Single colonies were picked and placed in 5 mL of TB media with 50 µg/mL kanamycin. The resulting culture was incubated overnight at 37 °C and 250 rpm. Storage glycerol stocks were made by

mixing 500 µL of overnight culture with 500 µL of 50% glycerol. These stocks were stored at -80 °C until needed.

A starter culture was prepared by inoculating 5 mL of TB media with 50 µg/mL kanamycin from glycerol stocks made during library expression and incubating at 37 °C and 250 rpm overnight. The overnight culture was used to generate an expression culture by inoculating 750 mL of LB media with 50 µg/mL kanamycin. The expression culture was incubated at 37 °C and 250 rpm until OD_{600} reached 0.6 – 0.8, after which the culture was brought to 18 °C, induced with 1 mM IPTG, and incubated at 18 °C and 250 rpm for 18 hours. Cells were harvested by centrifugation at 4700 rpm for 10 minutes and stored at -80 °C until use.

The cell pellet was resuspended in a 50 mL conical tube with 30 mL of 20 mM imidazole in 10 mM HEPES pH 7.5 and sonicated at 40 W with 0.5 min ON/0.5 min OFF cycles for 5 min total cycle time. The resulting cell lysate was clarified by centrifugation at 4 °C and 15000 rpm for 45 min using a high-speed fixed-angle rotor. The supernatant was transferred to a 10 mL polypropylene frit-bottomed gravity flow column containing 5 mL of Ni-NTA resin pre-equilibrated with equilibration buffer (20 mM phosphate, 300 mM NaCl, 10 mM imidazole, pH 7.4). After the lysate was allowed to drain, 50 mL of wash buffer (20 mM phosphate, 300 mM NaCl, 25 mM imidazole, pH 7.4) were added to the resin. Finally, 15 mL of elution buffer (20 mM phosphate, 300 mM NaCl, 250 mM imidazole, pH 7.4) were used to elute purified protein into a 50 mL conical tube. The protein solution was transferred to a 15 mL Amicon spin filter 30K MWCO and concentrated down to 0.5 – 1.0 mL at 4700 rpm. In order to remove endogenous Fe(II) 15 mL of 50 mM EDTA in 10 mM HEPES pH 7.5 were used to dilute the protein sample and concentrate to 0.5 – 1.0 mL at 4700 rpm. Three additional buffer exchanges with 15 mL of 10 mM HEPES in pH 7.5 were performed. Protein concentration was measured

using absorbance at 280 nm using a calculated extinction coefficient of 104,280 $M^{-1}cm^{-1}$ for all SadX variants (Benchling [Biology Software], https://benchling.com). The protein sample was flash frozen in liquid nitrogen and stored at -80 °C until used.

Expression in 96-deep well plates

Glycerol stocks were used to inoculate 1 mL 96-deep well plates containing 300 µL/well of TB media with 50 µg/mL kanamycin to generate primary cultures. The resulting plates were incubated at 37 °C and 215 rpm overnight. Expression cultures were made using 50 µL of primary culture to inoculate 2 mL 96-deep well plates containing 1 mL/well of LB media with 50 µg/mL kanamycin. Expression plates were incubated at 37 °C and 215 rpm until OD_{600} reached 0.6 – 0.8, at which point the cultures were brought to a temperature of 18 °C, induced with a final concentration of 1 mM IPTG, and incubated at 18 °C and 215 rpm for 18 hours. Cells were harvested by centrifugation at 3600 rpm for 10 minutes and stored at -80 °C until use.

5.4.7 Activity assays

*Small-scale azidation of **1a***

Enzyme activity was determined by small-scale bioconversions using purified protein. Reactions were set up by combining in a 96-well plate 25 µL/well of solution A (25 mL containing 16 mM α-ketoglutarate, 32 mM NaN_3, 64 mM ascorbic acid, and 0.8 mM $Fe(NH_4)_2(SO_4)_2$ in 10 mM HEPES pH 7.5), 25 µL/well of solution B (2 mL of 4 mM **1a** in 10 mM HEPES pH 7.5), and 50 µL/well of solution C (200 µL of 0.2 mM of an enzyme solution with two-fold the final concentration). The final reactions (100 µL) had final concentrations of 1 mM **1a**, 4 mM α-ketoglutarate, 8 mM NaN_3, 16 mM ascorbic acid, 0.2 mM $Fe(NH_4)_2(SO_4)_2$, and varying amounts of SadX variants in 10 mM HEPES pH 7.5. The plate was sealed with breathable film and shaken at room temperature and 750 rpm for 3 hours, after which 100 µL of

methanol were added to quench the reactions, 50 µL of a 2 mM *N*-acetyl-L-valine solution in water (100 nmol) were added as internal standard, and precipitated proteins were removed by centrifugation at 3600 rpm for 10 minutes. Finally, 80 µL of the resulting supernatant were filtered through 0.22 µm filter plates by centrifugation at 3600 rpm for 10 minutes and analyzed via LC-MS.

Azidation assay in lysate

Cell pellets in 2 mL 96-deep well plates were suspended with 125 µL of 1 mg/mL lysozyme in 10 mM HEPES pH 7.5 and incubated at 37 °C and 250 rpm for 45 minutes to achieve lysis, after which the lysates were flash frozen in liquid nitrogen and thawed in a water bath at 37 °C. 10 µL of 1 mg/mL DNase in 10 mM HEPES pH 7.5 were added and the lysates were incubated at 37 °C and 250 rpm for 15 minutes, followed by centrifugation at 3600 rpm for 15 minutes.

Screening reactions were set up in microtiter v-bottom 96-well plates by combining 25 µL/well of solution A (50 mL containing 8 mM α-ketoglutarate, 32 mM NaN_3, 32 mM ascorbic acid, and 0.4 mM $Fe(NH_4)_2(SO_4)_2$ in 10 mM HEPES pH 7.5), 25 µL/well of solution B (30 mL of 4 mM **1a** in 10 mM HEPES pH 7.5), and 50 µL/well of clarified lysate. The final reaction conditions were 100 µL of 1 mM **1a**, 2 mM α-ketoglutarate, 8 mM NaN_3, 8 mM ascorbic acid, 0.1 mM $Fe(NH_4)_2(SO_4)_2$, and 50% v/v lysate in 10 mM HEPES pH 7.5. The plates were sealed with breathable film and shaken at room temperature and 750 rpm for 3 hours, after which 100 µL of methanol were added to quench the reactions and precipitated proteins were removed by centrifugation at 3600 rpm for 10 minutes. 50 µL of the resulting supernatant were diluted with 100 µL of water, filtered through 0.22 µm filter plates by centrifugation at 3600 rpm for 10 minutes, and analyzed via LC-MS.

5.4.8 Re-array and pooling workflow

Defining bins

The activity data from the high-throughput screen of directed evolution libraries was used to categorize every individual variant from the library into four distinct categories, or "bins", according to their level of azidase activity with respect to the parent enzyme. The conversion to **1e** (*%*1e) of every parent control included in the library was averaged (μ) and a standard deviation was determined (σ). The library variants with *%1e* higher than $\mu + \sigma$ were labeled as having *high* activity, those variants that followed $\mu - \sigma < \%1e < \mu + \sigma$ were labeled as having *parent* activity, variants with $0.1\% < \%1e < \mu - \sigma$ were labeled as having *low* activity, and those with *%1e* lower than 0.1% were labeled as having *no activity*. The screening data of library 2-L is shown in Table AIII.1 as an example.

Re-array of variants

Glycerol stocks of the screened library were re-arrayed according to the bins defined before into new 96-deep-well plates containing 300 µL of TB media with 50 µg/mL of kanamycin. Table AIII.2 shows a representative file used to conduct the re-array procedure. For any given library, this resulted in four sets of 96-deep-well plates, each set corresponding to an activity bin. These re-arrayed cultures were incubated overnight at 37 °C and 215 rpm. All cultures belonging to the same bin were pooled and thoroughly mixed so that all the variants would end in a single culture. This was repeated for the remaining bins.

It is worth noting that the number of variants in any given bin determined the final volume of culture after pooling, and thus, the total amount of DNA. In cases where the total number of variants in a given bin was lower than ~60, empty wells were used to inoculate

replicates of the bin members during the re-array step. This was done to prevent low DNA yields due to the low amount of culture.

DNA sample preparation

Plasmid extraction was carried out on each one of four cultures in order to obtain the purified pET-28a plasmids containing SadX variants. These plasmid samples were used to set up restriction enzyme double digests using SacI-HF and XhoI in 50 µL reactions with ~5 µg of plasmid DNA, 250 units of SacI-HF, and 250 units of XhoI. These reactions were incubated at 37 °C for 6 hours and inactivated at 65 °C for 20 minutes. The resulting mixtures were purified using 1% agarose gel electrophoresis. These purified samples were submitted to the DNA sequencing facility at the University of Wisconsin Biotechnology Center for SMRT (PacBio) sequencing as samples containing 350-400 ng of DNA in water.

5.5 References

(1) Sanger, F.; Nicklen, S.; Coulson, A. R. DNA Sequencing with Chain-Terminating Inhibitors. *Proc. Natl. Acad. Sci.* **1977**, *74* (12), 5463–5467. https://doi.org/10.1073/PNAS.74.12.5463.

(2) Slatko, B. E.; Gardner, A. F.; Ausubel, F. M. Overview of Next-Generation Sequencing Technologies. *Curr. Protoc. Mol. Biol.* **2018**, *122* (1), e59. https://doi.org/10.1002/cpmb.59.

(3) Hu, T.; Chitnis, N.; Monos, D.; Dinh, A. Next-Generation Sequencing Technologies: An Overview. *Hum. Immunol.* **2021**, *82* (11), 801–811. https://doi.org/10.1016/j.humimm.2021.02.012.

(4) Bentley, D. R.; Balasubramanian, S.; Swerdlow, H. P.; Smith, G. P.; Milton, J.; Brown, C. G.; Hall, K. P.; Evers, D. J.; Barnes, C. L.; Bignell, H. R.; Boutell, J. M.; Bryant, J.; Carter, R. J.; Keira Cheetham, R.; Cox, A. J.; Ellis, D. J.; Flatbush, M. R.; Gormley, N. A.; Humphray, S. J.; Irving, L. J.; Karbelashvili, M. S.; Kirk, S. M.; Li, H.; Liu, X.; Maisinger, K. S.; Murray, L. J.; Obradovic, B.; Ost, T.; Parkinson, M. L.; Pratt, M. R.; Rasolonjatovo, I. M. J.; Reed, M. T.; Rigatti, R.; Rodighiero, C.; Ross, M. T.; Sabot, A.; Sankar, S. V.; Scally, A.; Schroth, G. P.; Smith, M. E.; Smith, V. P.; Spiridou, A.; Torrance, P. E.; Tzonev, S. S.; Vermaas, E. H.; Walter, K.; Wu, X.; Zhang, L.; Alam, M. D.; Anastasi, C.; Aniebo, I. C.; Bailey, D. M. D.; Bancarz, I. R.; Banerjee, S.; Barbour, S. G.; Baybayan, P. A.; Benoit, V. A.; Benson, K. F.; Bevis, C.; Black, P. J.; Boodhun, A.; Brennan, J. S.; Bridgham, J. A.; Brown, R. C.; Brown, A. A.; Buermann, D. H.; Bundu, A. A.; Burrows, J. C.; Carter, N. P.; Castillo, N.; Catenazzi, M. C. E.; Chang, S.; Neil Cooley, R.; Crake, N. R.; Dada, O. O.; Diakoumakos, K. D.; Dominguez-Fernandez, B.; Earnshaw, D. J.; Egbujor, U. C.; Elmore, D. W.; Etchin, S. S.; Ewan, M. R.; Fedurco, M.; Fraser, L. J.; Fuentes Fajardo, K. V.; Scott Furey, W.; George, D.; Gietzen, K. J.; Goddard, C. P.; Golda, G. S.; Granieri, P. A.; Green, D. E.; Gustafson, D. L.; Hansen, N. F.; Harnish, K.; Haudenschild, C. D.; Heyer, N. I.; Hims, M. M.; Ho, J. T.; Horgan, A. M.; Hoschler, K.; Hurwitz, S.; Ivanov, D. V.; Johnson, M. Q.; James, T.; Huw Jones, T. A.; Kang, G. D.; Kerelska, T. H.; Kersey, A. D.; Khrebtukova, I.; Kindwall, A. P.; Kingsbury, Z.; Kokko-Gonzales, P. I.; Kumar, A.; Laurent, M. A.; Lawley, C. T.; Lee, S. E.; Lee, X.; Liao, A. K.; Loch, J. A.; Lok, M.; Luo, S.; Mammen, R. M.; Martin, J. W.; McCauley, P. G.; McNitt, P.; Mehta, P.; Moon, K. W.; Mullens, J. W.; Newington, T.; Ning, Z.; Ling Ng, B.; Novo, S. M.; O'Neill, M. J.; Osborne, M. A.; Osnowski, A.; Ostadan, O.; Paraschos, L. L.; Pickering, L.; Pike, A. C.; Pike, A. C.; Chris Pinkard, D.; Pliskin, D. P.; Podhasky, J.; Quijano, V. J.; Raczy, C.; Rae, V. H.; Rawlings, S. R.; Chiva Rodriguez, A.; Roe, P. M.; Rogers, J.; Rogert Bacigalupo, M. C.; Romanov, N.; Romieu, A.; Roth, R. K.; Rourke, N. J.; Ruediger, S. T.; Rusman, E.; Sanches-Kuiper, R. M.; Schenker, M. R.; Seoane, J. M.; Shaw, R. J.; Shiver, M. K.; Short, S. W.; Sizto, N. L.; Sluis, J. P.; Smith, M. A.; Ernest Sohna Sohna, J.; Spence, E. J.; Stevens, K.; Sutton, N.; Szajkowski, L.; Tregidgo, C. L.; Turcatti, G.; Vandevondele, S.; Verhovsky, Y.; Virk, S. M.; Wakelin, S.; Walcott, G. C.; Wang, J.; Worsley, G. J.; Yan, J.; Yau, L.; Zuerlein, M.; Rogers, J.; Mullikin, J. C.; Hurles, M. E.; McCooke, N. J.; West, J. S.; Oaks, F. L.; Lundberg, P. L.; Klenerman, D.; Durbin, R.; Smith, A. J. Accurate Whole Human Genome Sequencing

Using Reversible Terminator Chemistry. *Nature* **2008**, *456* (7218), 53–59. https://doi.org/10.1038/nature07517.

(5) Eid, J.; Fehr, A.; Gray, J.; Luong, K.; Lyle, J.; Otto, G.; Peluso, P.; Rank, D.; Baybayan, P.; Bettman, B.; Bibillo, A.; Bjornson, K.; Chaudhuri, B.; Christians, F.; Cicero, R.; Clark, S.; Dalal, R.; DeWinter, A.; Dixon, J.; Foquet, M.; Gaertner, A.; Hardenbol, P.; Heiner, C.; Hester, K.; Holden, D.; Kearns, G.; Kong, X.; Kuse, R.; Lacroix, Y.; Lin, S.; Lundquist, P.; Ma, C.; Marks, P.; Maxham, M.; Murphy, D.; Park, I.; Pham, T.; Phillips, M.; Roy, J.; Sebra, R.; Shen, G.; Sorenson, J.; Tomaney, A.; Travers, K.; Trulson, M.; Vieceli, J.; Wegener, J.; Wu, D.; Yang, A.; Zaccarin, D.; Zhao, P.; Zhong, F.; Korlach, J.; Turner, S. Real-Time DNA Sequencing from Single Polymerase Molecules. *Science (80-.).* **2009**, *323* (5910), 133–138. https://doi.org/10.1126/science.1162986.

(6) Travers, K. J.; Chin, C. S.; Rank, D. R.; Eid, J. S.; Turner, S. W. A Flexible and Efficient Template Format for Circular Consensus Sequencing and SNP Detection. *Nucleic Acids Res.* **2010**, *38* (15), e159–e159. https://doi.org/10.1093/NAR/GKQ543.

(7) Rhoads, A.; Au, K. F. PacBio Sequencing and Its Applications. *Genomics, Proteomics and Bioinformatics*. Elsevier October 1, 2015, pp 278–289. https://doi.org/10.1016/j.gpb.2015.08.002.

(8) Tarca, A. L.; Carey, V. J.; Chen, X. wen; Romero, R.; Drăghici, S. Machine Learning and Its Applications to Biology. *PLoS computational biology*. Public Library of Science 2007, p e116. https://doi.org/10.1371/journal.pcbi.0030116.

(9) Singh, S. P.; Kumar, A.; Darbari, H.; Singh, L.; Rastogi, A.; Jain, S. Machine Translation Using Deep Learning: An Overview. In *2017 International Conference on Computer, Communications and Electronics, COMPTELIX 2017*; Institute of Electrical and Electronics Engineers Inc., 2017; pp 162–167. https://doi.org/10.1109/COMPTELIX.2017.8003957.

(10) Liakos, K. G.; Busato, P.; Moshou, D.; Pearson, S.; Bochtis, D. Machine Learning in Agriculture: A Review. *Sensors (Switzerland)*. Multidisciplinary Digital Publishing Institute August 14, 2018, p 2674. https://doi.org/10.3390/s18082674.

(11) Wei, J.; Chu, X.; Sun, X. Y.; Xu, K.; Deng, H. X.; Chen, J.; Wei, Z.; Lei, M. Machine Learning in Materials Science. *InfoMat*. John Wiley & Sons, Ltd September 1, 2019, pp 338–358. https://doi.org/10.1002/inf2.12028.

(12) Zou, J.; Huss, M.; Abid, A.; Mohammadi, P.; Torkamani, A.; Telenti, A. A Primer on Deep Learning in Genomics. *Nat. Genet.* **2019**, *51* (1), 12–18. https://doi.org/10.1038/s41588-018-0295-5.

(13) Dramsch, J. S. 70 Years of Machine Learning in Geoscience in Review. In *Advances in Geophysics*; Elsevier, 2020; Vol. 61, pp 1–55. https://doi.org/10.1016/bs.agph.2020.08.002.

(14) Jadhav, O. N.; Ashwini, K. B. Movie Recommendation System Using Machine Learning

Algorithms. *J. Mach. Comput.* **2022**, *2* (2), 81–86. https://doi.org/10.53759/7669/jmc202202011.

(15) Alzubi, J.; Nayyar, A.; Kumar, A. Machine Learning from Theory to Algorithms: An Overview. In *Journal of Physics: Conference Series*; IOP Publishing, 2018; Vol. 1142, p 12012. https://doi.org/10.1088/1742-6596/1142/1/012012.

(16) Buchan, D. W. A.; Jones, D. T. The PSIPRED Protein Analysis Workbench: 20 Years On. *Nucleic Acids Res.* **2019**, *47* (W1), W402–W407. https://doi.org/10.1093/nar/gkz297.

(17) Hopf, T. A.; Ingraham, J. B.; Poelwijk, F. J.; Schärfe, C. P. I.; Springer, M.; Sander, C.; Marks, D. S. Mutation Effects Predicted from Sequence Co-Variation. *Nat. Biotechnol.* **2017**, *35* (2), 128–135. https://doi.org/10.1038/nbt.3769.

(18) Tibshirani, R. Regression Shrinkage and Selection Via the Lasso. *J. R. Stat. Soc. Ser. B* **1996**, *58* (1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

(19) Noble, W. S. What Is a Support Vector Machine? *Nature Biotechnology*. Nature Publishing Group December 2006, pp 1565–1567. https://doi.org/10.1038/nbt1206-1565.

(20) Gutiérrez, P. A.; Pérez-Ortiz, M.; Sánchez-Monedero, J.; Fernández-Navarro, F.; Hervás-Martínez, C. Ordinal Regression Methods: Survey and Experimental Study. In *IEEE Transactions on Knowledge and Data Engineering*; IEEE Computer Society, 2016; Vol. 28, pp 127–146. https://doi.org/10.1109/TKDE.2015.2457911.

(21) Lecun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature*. Nature Publishing Group May 27, 2015, pp 436–444. https://doi.org/10.1038/nature14539.

(22) Senior, A. W.; Evans, R.; Jumper, J.; Kirkpatrick, J.; Sifre, L.; Green, T.; Qin, C.; Žídek, A.; Nelson, A. W. R.; Bridgland, A.; Penedones, H.; Petersen, S.; Simonyan, K.; Crossan, S.; Kohli, P.; Jones, D. T.; Silver, D.; Kavukcuoglu, K.; Hassabis, D. Improved Protein Structure Prediction Using Potentials from Deep Learning. *Nature* **2020**, *577* (7792), 706–710. https://doi.org/10.1038/s41586-019-1923-7.

(23) Mandecki, W. The Game of Chess and Searches in Protein Sequence Space. *Trends Biotechnol.* **1998**, *16* (5), 200–202. https://doi.org/10.1016/S0167-7799(98)01188-3.

(24) Maynard Smith, J. Natural Selection and the Concept of a Protein Space. *Nature* **1970**, *225* (5232), 563–564. https://doi.org/10.1038/225563a0.

(25) Romero, P. A.; Arnold, F. H. Exploring Protein Fitness Landscapes by Directed Evolution. *Nat. Rev. Mol. Cell Biol. 2009 1012* **2009**, *10* (12), 866–876. https://doi.org/10.1038/nrm2805.

(26) Yang, K. K.; Wu, Z.; Arnold, F. H. Machine-Learning-Guided Directed Evolution for Protein Engineering. *Nat. Methods* **2019**, *16* (8), 687–694. https://doi.org/10.1038/S41592-019-0496-6.

(27) Wu, Z.; Jennifer Kan, S. B.; Lewis, R. D.; Wittmann, B. J.; Arnold, F. H. Machine Learning-Assisted Directed Protein Evolution with Combinatorial Libraries. *Proc. Natl. Acad. Sci. U. S. A.* **2019**, *116* (18), 8852–8858. https://doi.org/10.1073/PNAS.1901979116/SUPPL_FILE/PNAS.1901979116.SAPP.PDF.

(28) Patsch, D.; Buller, R. Improving Enzyme Fitness with Machine Learning. *Chimia (Aarau).* **2023**, *77* (3), 116. https://doi.org/10.2533/chimia.2023.116.

(29) Fox, R. J.; Davis, S. C.; Mundorff, E. C.; Newman, L. M.; Gavrilovic, V.; Ma, S. K.; Chung, L. M.; Ching, C.; Tam, S.; Muley, S.; Grate, J.; Gruber, J.; Whitman, J. C.; Sheldon, R. A.; Huisman, G. W. Improving Catalytic Function by ProSAR-Driven Enzyme Evolution. *Nat. Biotechnol.* **2007**, *25* (3), 338–344. https://doi.org/10.1038/nbt1286.

(30) Fox, R. Directed Molecular Evolution by Machine Learning and the Influence of Nonlinear Interactions. *J. Theor. Biol.* **2005**, *234* (2), 187–199. https://doi.org/10.1016/j.jtbi.2004.11.031.

(31) Novick, S. J.; Dellas, N.; Garcia, R.; Ching, C.; Bautista, A.; Homan, D.; Alvizo, O.; Entwistle, D.; Kleinbeck, F.; Schlama, T.; Ruch, T. Engineering an Amine Transaminase for the Efficient Production of a Chiral Sacubitril Precursor. *ACS Catal.* **2021**, *11* (6), 3762–3770. https://doi.org/10.1021/acscatal.0c05450.

(32) Alvizo, O.; Nguyen, L. J.; Savile, C. K.; Bresson, J. A.; Lakhapatri, S. L.; Solis, E. O. P.; Fox, R. J.; Broering, J. M.; Benoit, M. R.; Zimmerman, S. A.; Novick, S. J.; Liang, J.; Lalonde, J. J. Directed Evolution of an Ultrastable Carbonic Anhydrase for Highly Efficient Carbon Capture from Flue Gas. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111* (46), 16436–16441. https://doi.org/10.1073/pnas.1411461111.

(33) Greenhalgh, J. C.; Fahlberg, S. A.; Pfleger, B. F.; Romero, P. A. Machine Learning-Guided Acyl-ACP Reductase Engineering for Improved in Vivo Fatty Alcohol Production. *Nat. Commun.* **2021**, *12* (1), 1–10. https://doi.org/10.1038/s41467-021-25831-w.

(34) Srinivas, N.; Krause, A.; Kakade, S. M.; Seeger, M. W. Information-Theoretic Regret Bounds for Gaussian Process Optimization in the Bandit Setting. *IEEE Trans. Inf. Theory* **2012**, *58* (5), 3250–3265. https://doi.org/10.1109/TIT.2011.2182033.

(35) Romero, P. A.; Krause, A.; Arnold, F. H. Navigating the Protein Fitness Landscape with Gaussian Processes. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110* (3), E193–E201. https://doi.org/10.1073/pnas.1215251110.

(36) Büchler, J.; Malca, S. H.; Patsch, D.; Voss, M.; Turner, N. J.; Bornscheuer, U. T.; Allemann, O.; Le Chapelain, C.; Lumbroso, A.; Loiseleur, O.; Buller, R. Algorithm-Aided Engineering of Aliphatic Halogenase WelO5* for the Asymmetric Late-Stage Functionalization of Soraphens. *Nat. Commun. 2022 131* **2022**, *13* (1), 1–11. https://doi.org/10.1038/s41467-022-27999-1.

(37) Lu, H.; Diaz, D. J.; Czarnecki, N. J.; Zhu, C.; Kim, W.; Shroff, R.; Acosta, D. J.; Alexander, B. R.; Cole, H. O.; Zhang, Y.; Lynd, N. A.; Ellington, A. D.; Alper, H. S. Machine Learning-Aided Engineering of Hydrolases for PET Depolymerization. *Nature* **2022**, *604* (7907), 662–667. https://doi.org/10.1038/s41586-022-04599-z.

(38) Shroff, R.; Cole, A. W.; Diaz, D. J.; Morrow, B. R.; Donnell, I.; Annapareddy, A.; Gollihar, J.; Ellington, A. D.; Thyer, R. Discovery of Novel Gain-of-Function Mutations Guided by Structure-Based Deep Learning. *ACS Synth. Biol.* **2020**, *9* (11), 2927–2935. https://doi.org/10.1021/acssynbio.0c00345.

(39) Greener, J. G.; Kandathil, S. M.; Moffat, L.; Jones, D. T. A Guide to Machine Learning for Biologists. *Nature Reviews Molecular Cell Biology*. Nature Publishing Group September 13, 2022, pp 40–55. https://doi.org/10.1038/s41580-021-00407-0.

(40) Wittmann, B. J.; Johnston, K. E.; Almhjell, P. J.; Arnold, F. H. EvSeq: Cost-Effective Amplicon Sequencing of Every Variant in a Protein Library. *ACS Synth. Biol.* **2022**, *11* (3), 1313–1324. https://doi.org/10.1021/ACSSYNBIO.1C00592/ASSET/IMAGES/LARGE/SB1C00592_0005.JPEG.

(41) Gomez, C. A.; Mondal, D.; Du, Q.; Chan, N.; Lewis, J. C. Directed Evolution of an Iron(II)- and α-Ketoglutarate-Dependent Dioxygenase for Site-Selective Azidation of Unactivated Aliphatic C−H Bonds**. *Angew. Chemie Int. Ed.* **2023**, *62* (15), e202301370. https://doi.org/10.1002/ANIE.202301370.

(42) Chan, N. H.; Gomez, C. A.; Vennelakanti, V.; Du, Q.; Kulik, H. J.; Lewis, J. C. Non-Native Anionic Ligand Binding and Reactivity in Engineered Variants of the Fe(II)- and α-Ketoglutarate-Dependent Oxygenase, SadA. *Inorg. Chem.* **2022**, *61* (36), 14477–14485. https://doi.org/10.1021/acs.inorgchem.2c02872.

(43) Sambrook, J.; Fritsch, E. F.; Maniatis, T. *Molecular Cloning: A Laboratory Manual, 2nd Edn.*; Cold Spring Laboratory Press, 1989.

(44) Liu, H.; Naismith, J. H. An Efficient One-Step Site-Directed Deletion, Insertion, Single and Multiple-Site Plasmid Mutagenesis Protocol. **2008**. https://doi.org/10.1186/1472-6750-8-91.

Appendix I

NMR spectra for compounds from Chapter 2

Figure A1.1: Multiplicity-edited ^{1}H-^{13}C HSQC of product **1e$_{N3}$** in CD_3OD

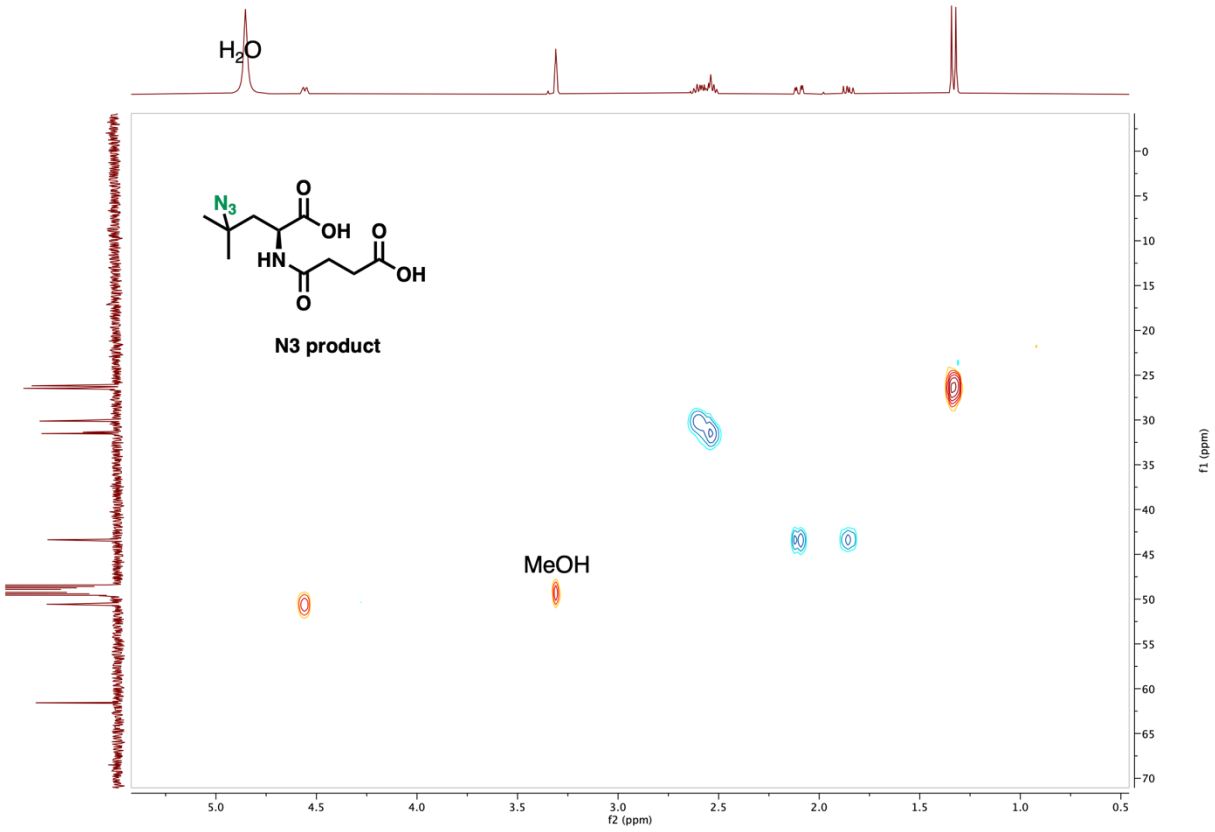


Figure A1.2: Multiplicity-edited ^{1}H-^{13}C HSQC of product **1c*** in D_2O

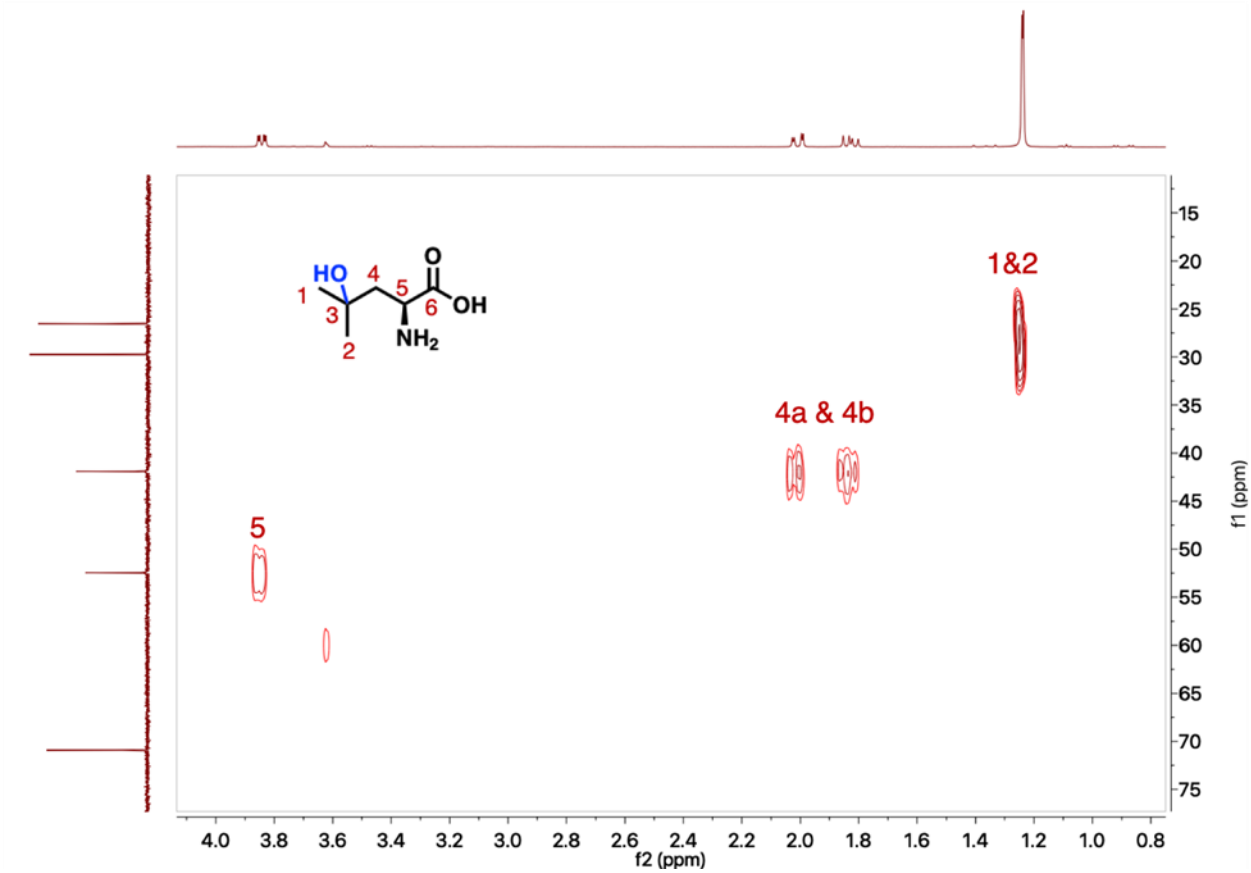
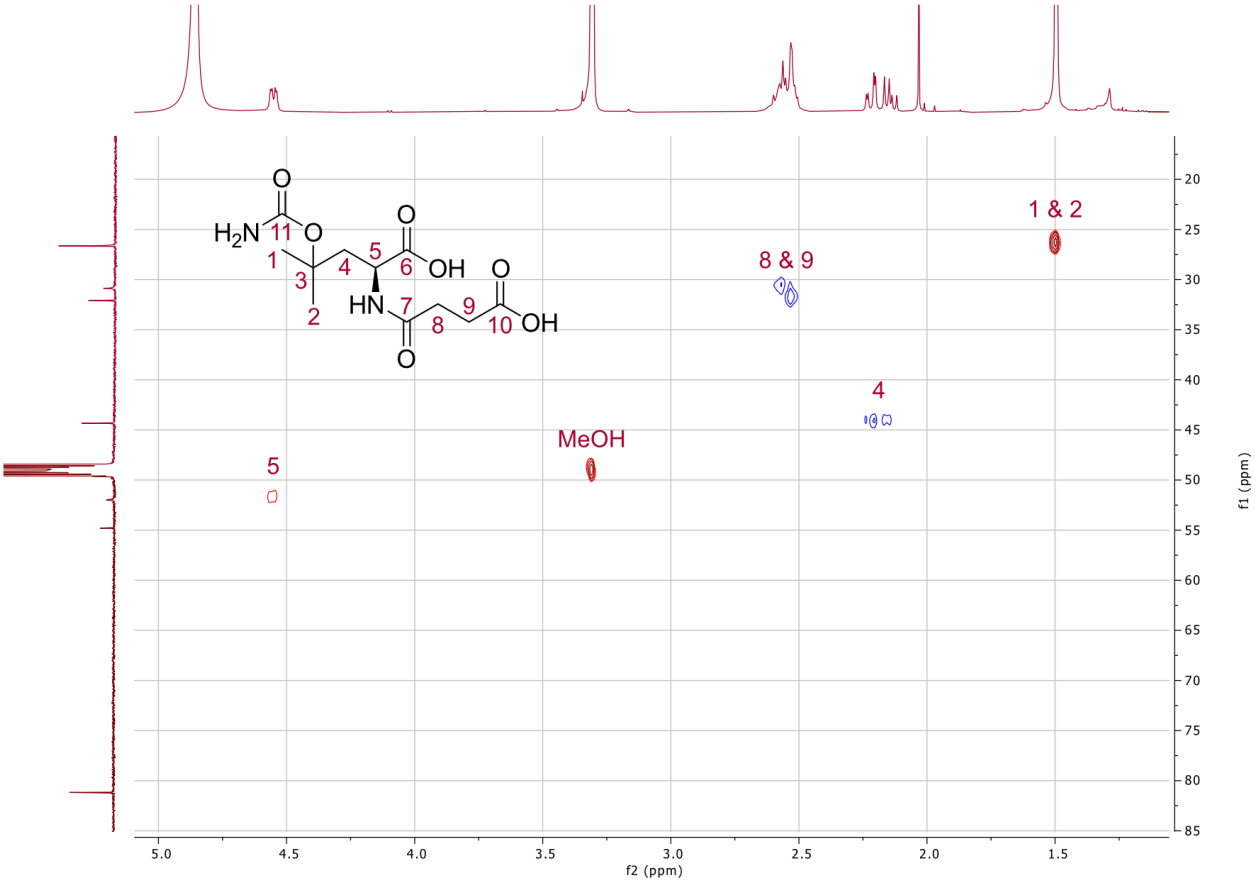


Figure A1.3: Multiplicity-edited ^{1}H-^{13}C HSQC of product **1e$_{OCN}$(a)** in D_2O



Appendix II

NMR spectra for compounds from Chapter 3

Figure AII.1: ^{1}H NMR spectrum of compound **4a** in CD_3OD

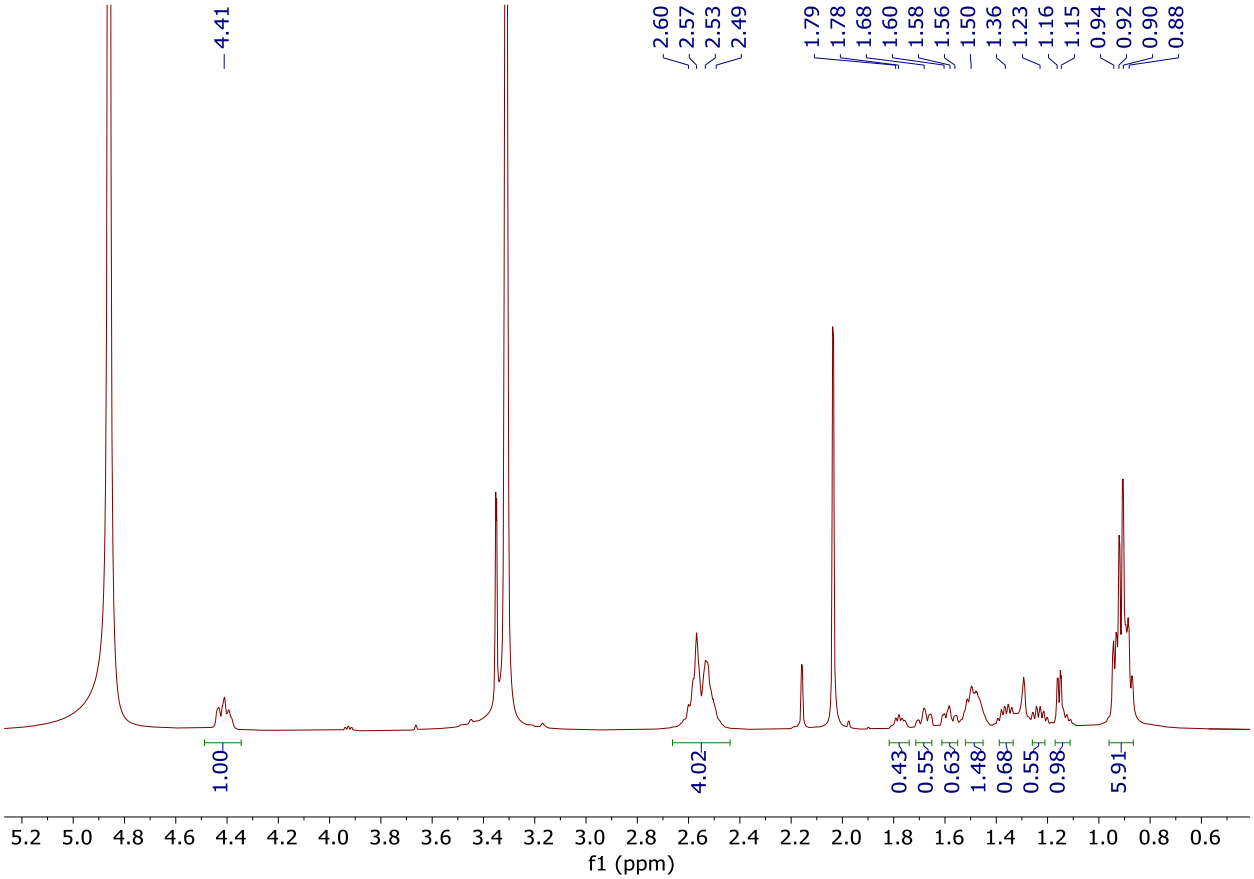


Figure AII.2: ^{13}C NMR spectrum of compound **4a** in CD_3OD

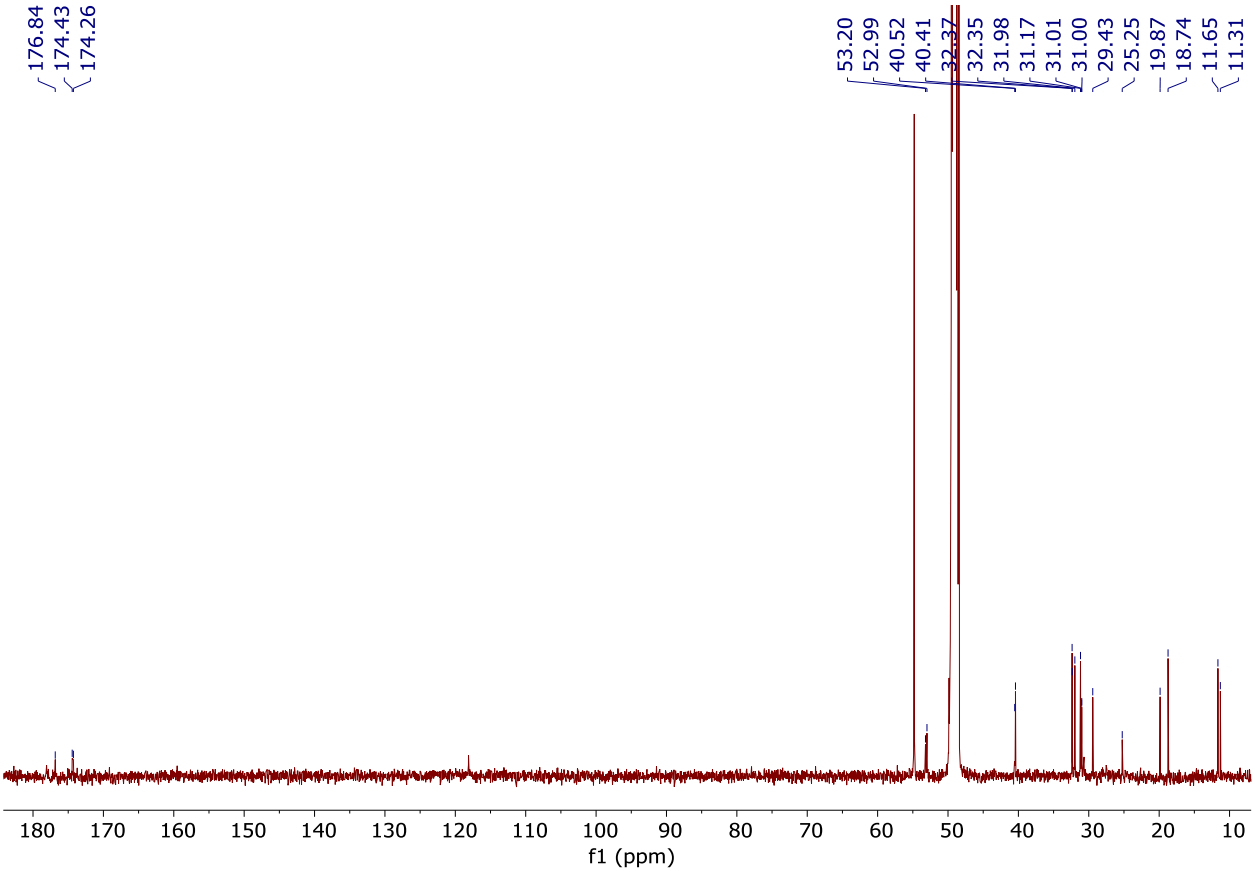


Figure AII.3: Multiplicity-edited ^{1}H-^{13}C HSQC of compound **4a** in CD_3OD

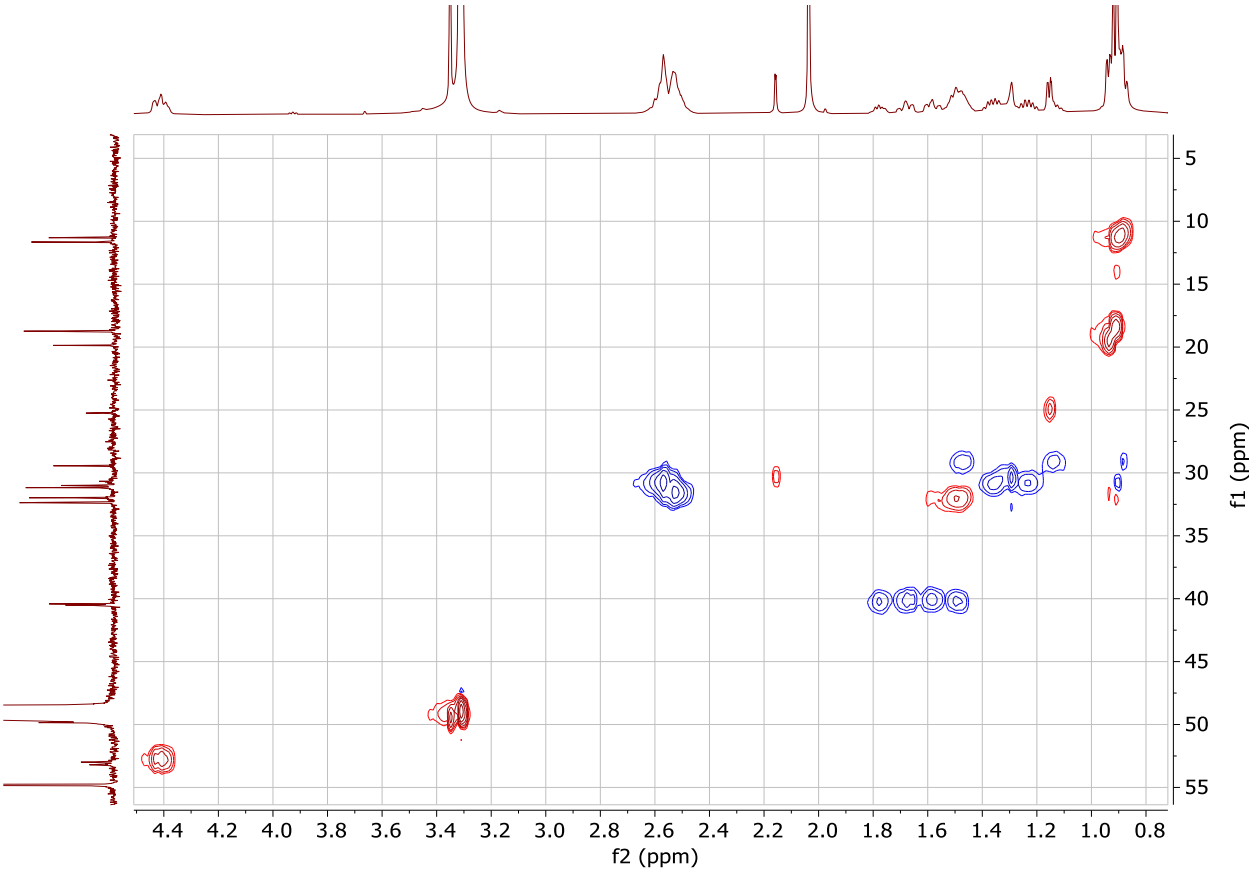


Figure AII.4: ^{1}H NMR spectrum of compound **5a** in CD_3OD

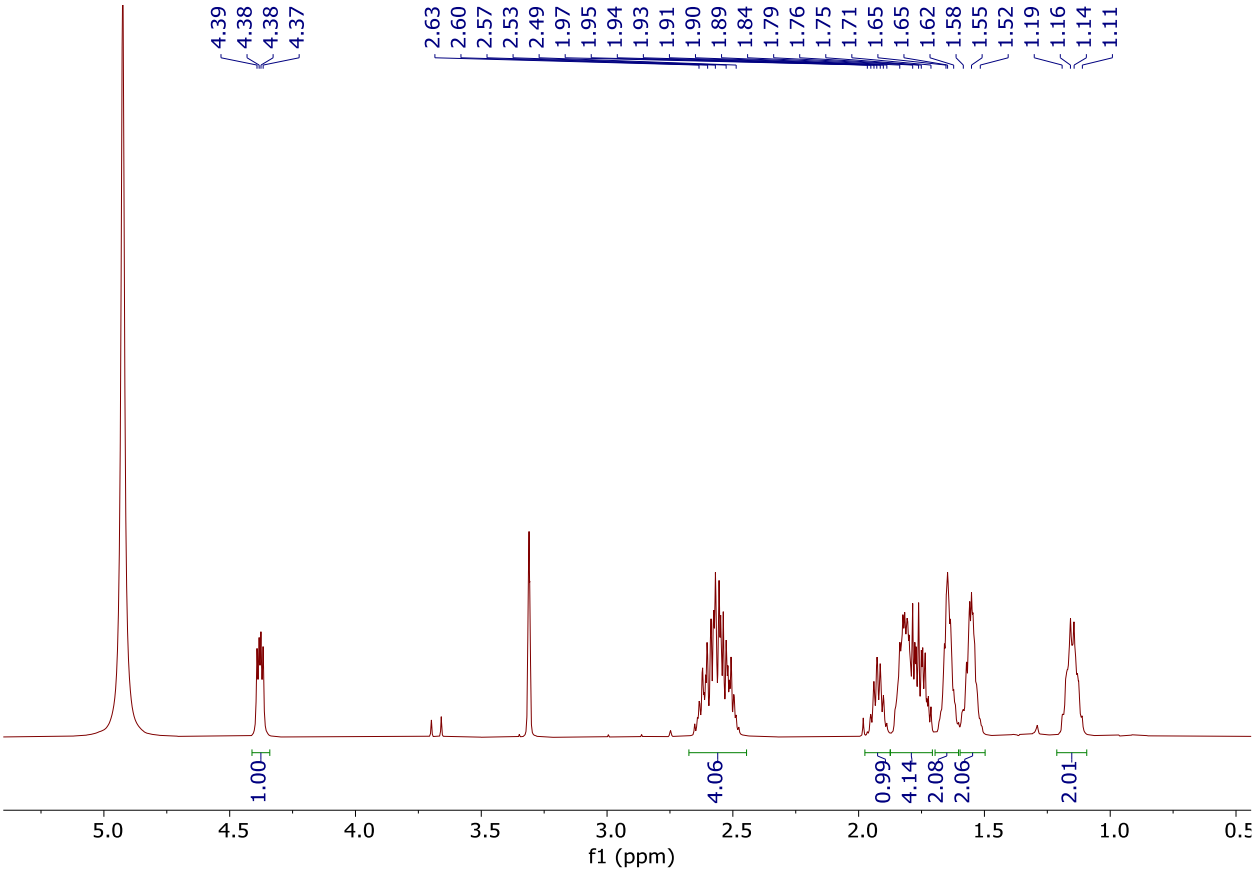


Figure AII.5: ^{13}C NMR spectrum of compound **5a** in CD_3OD

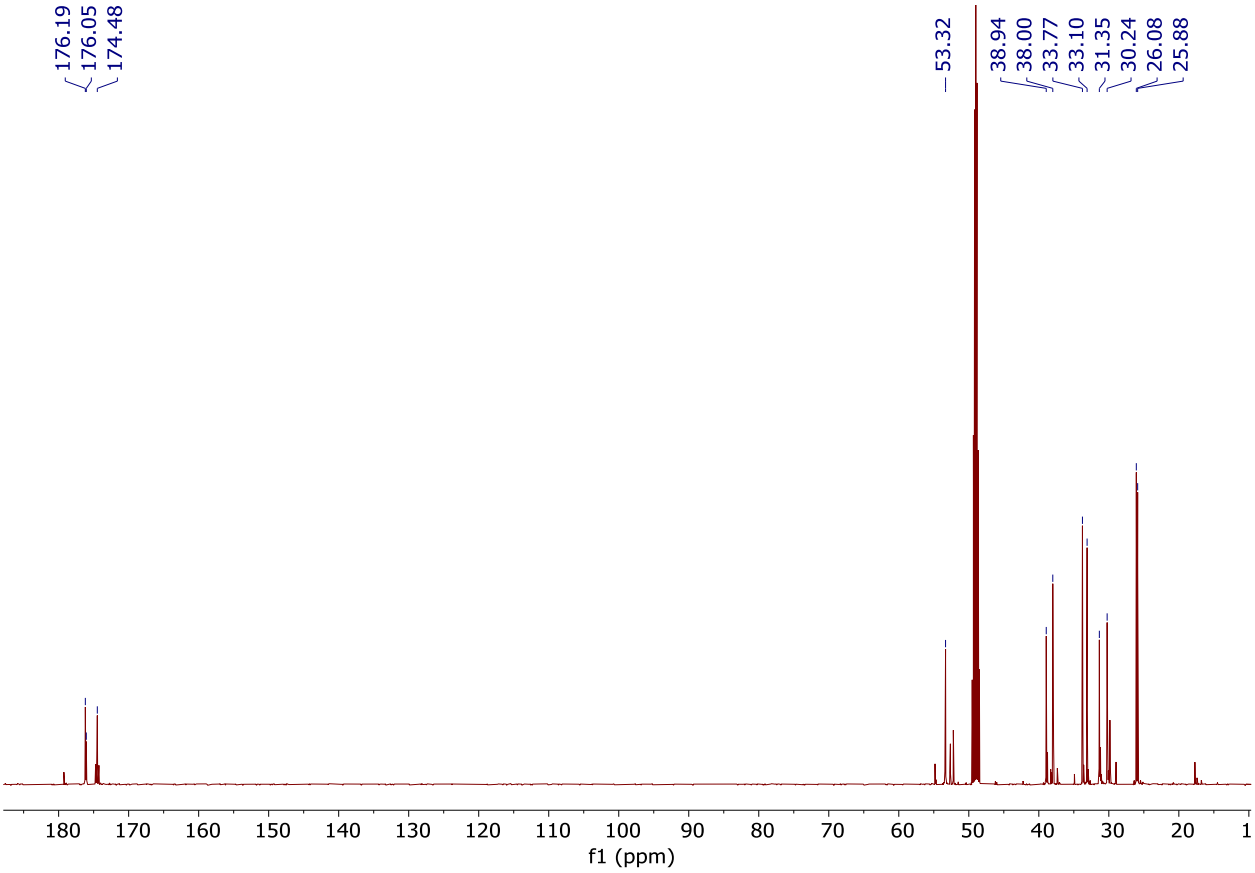


Figure AII.6: ^{1}H NMR spectrum of compound **6a** in CD_3OD

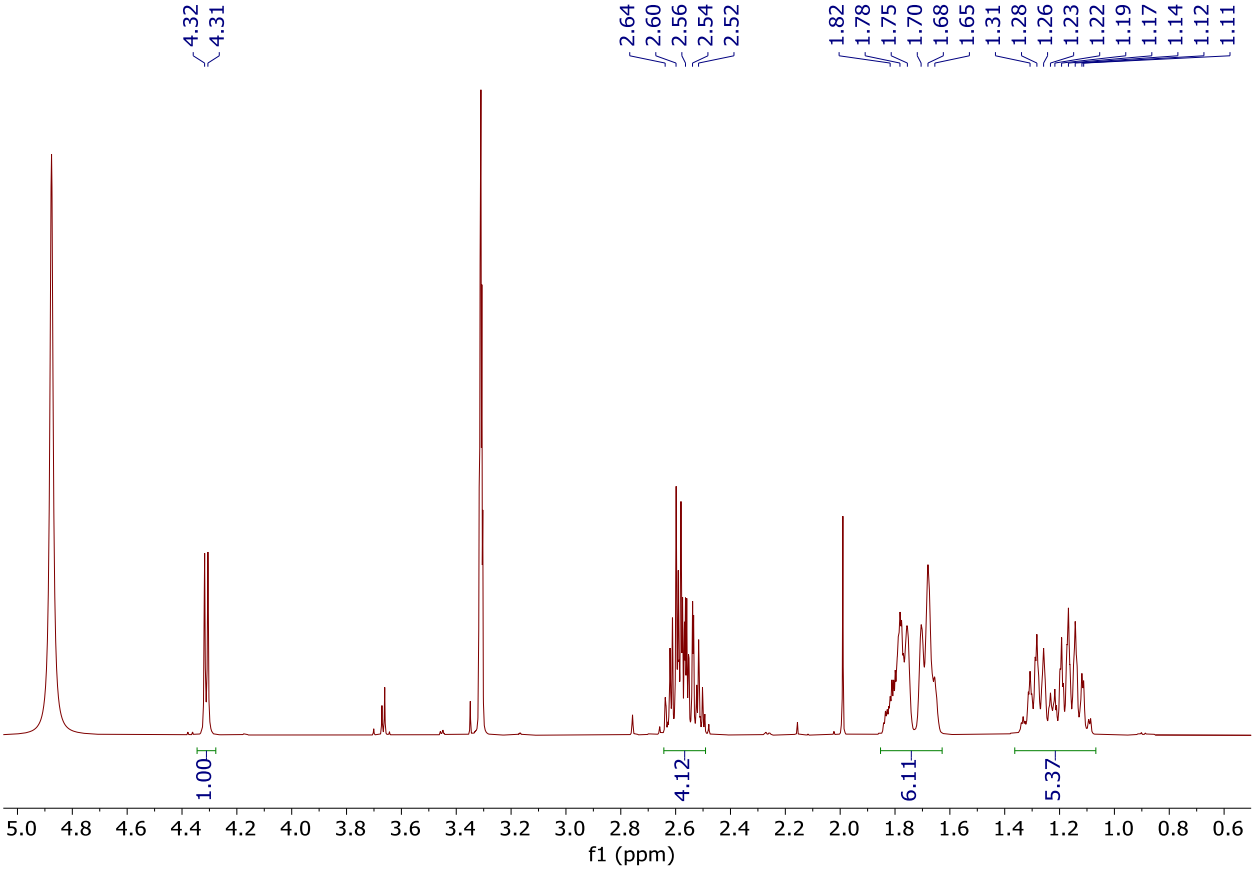


Figure AII.7: ^{13}C NMR spectrum of compound **6a** in CD_3OD

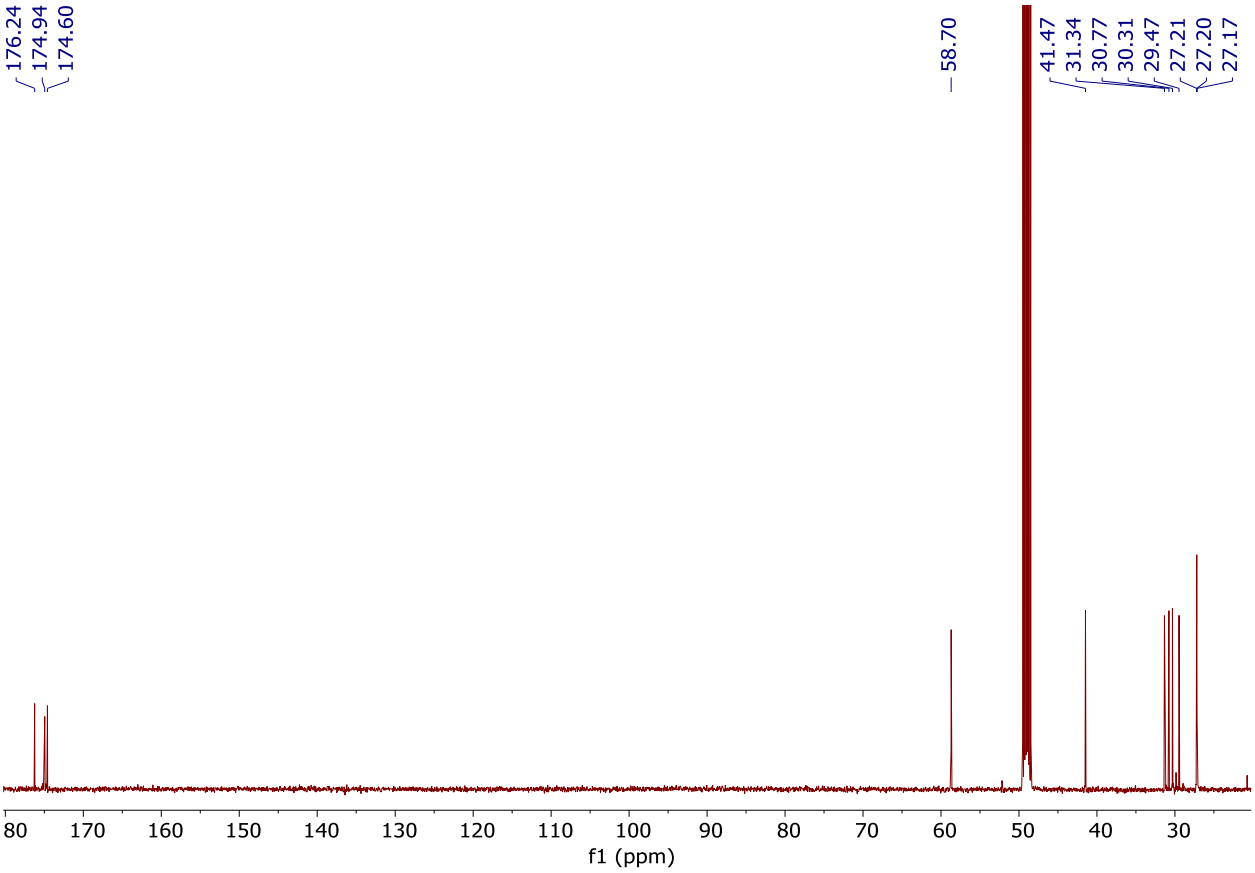


Figure AII.8: ^{1}H NMR spectrum of compound **7a** in CD_3OD

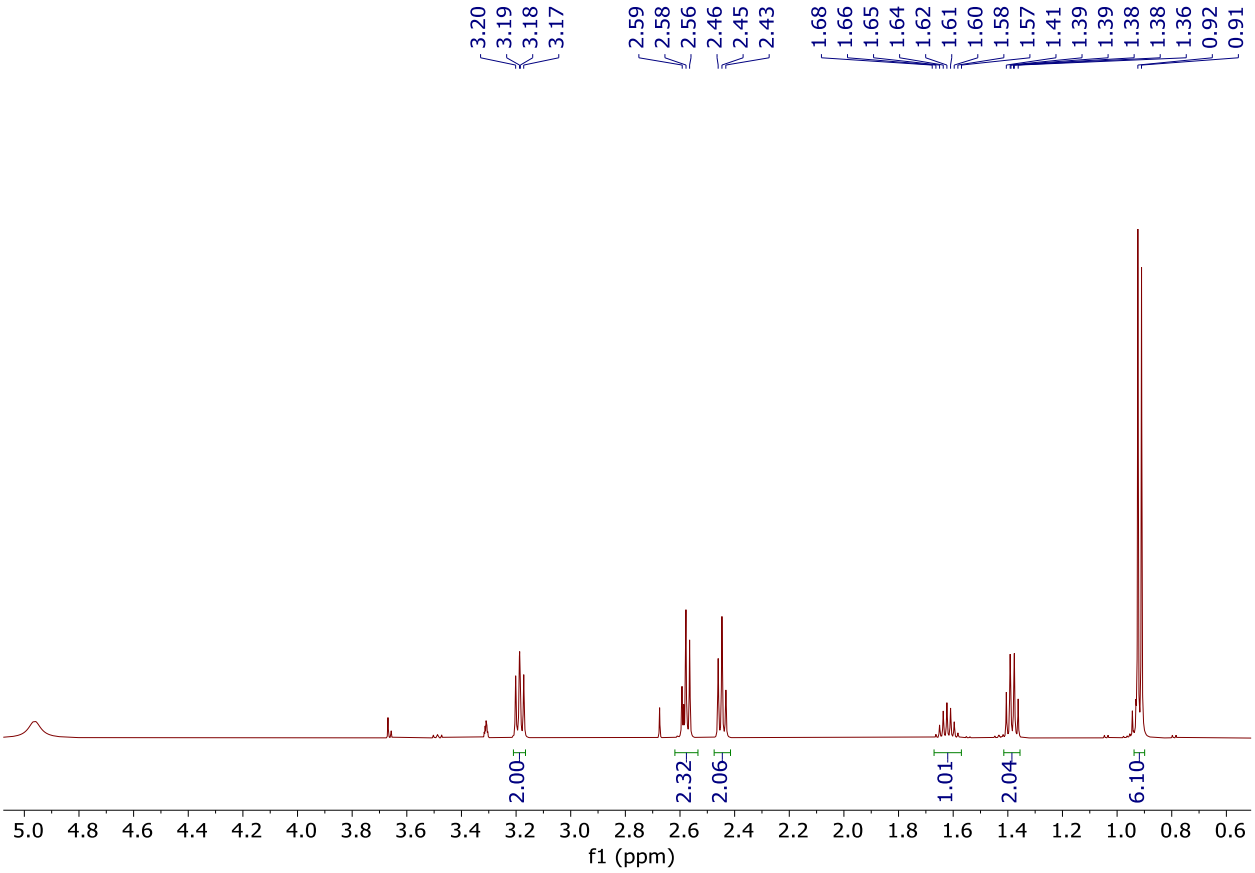


Figure AII.9: ^{13}C NMR spectrum of compound **7a** in CD_3OD

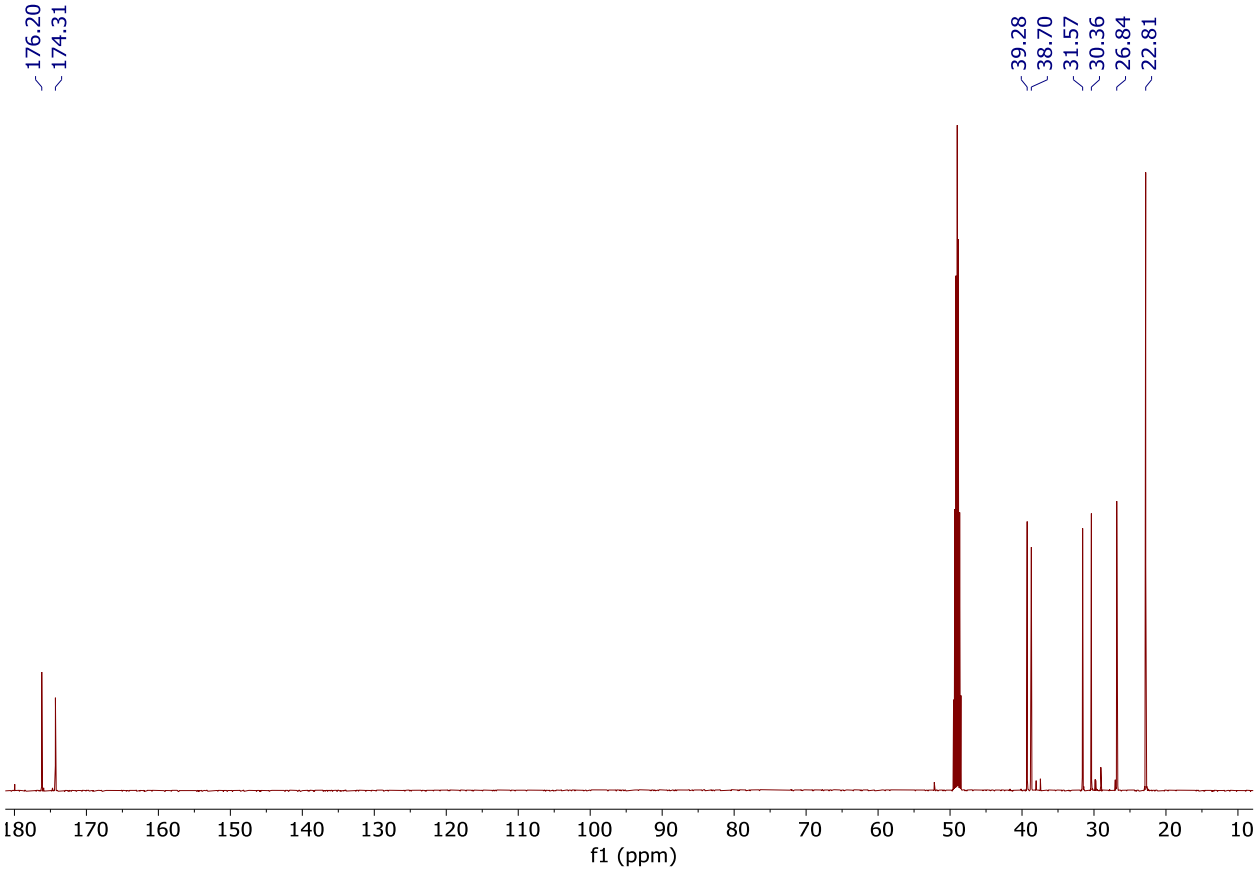


Figure AII.10: ^{1}H NMR spectrum of compound **2e** in CD_3OD

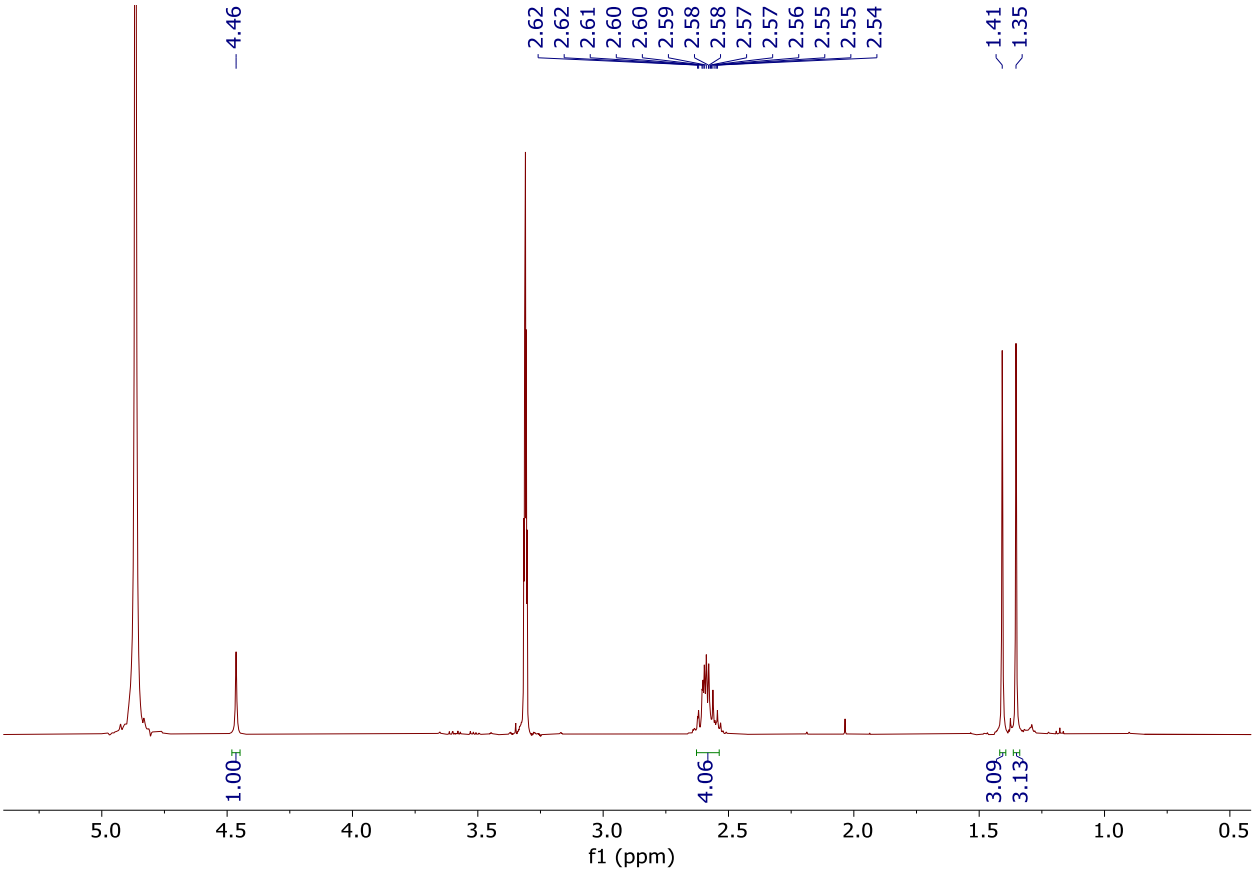


Figure AII.11: ^{13}C NMR spectrum of compound **2e** in CD_3OD

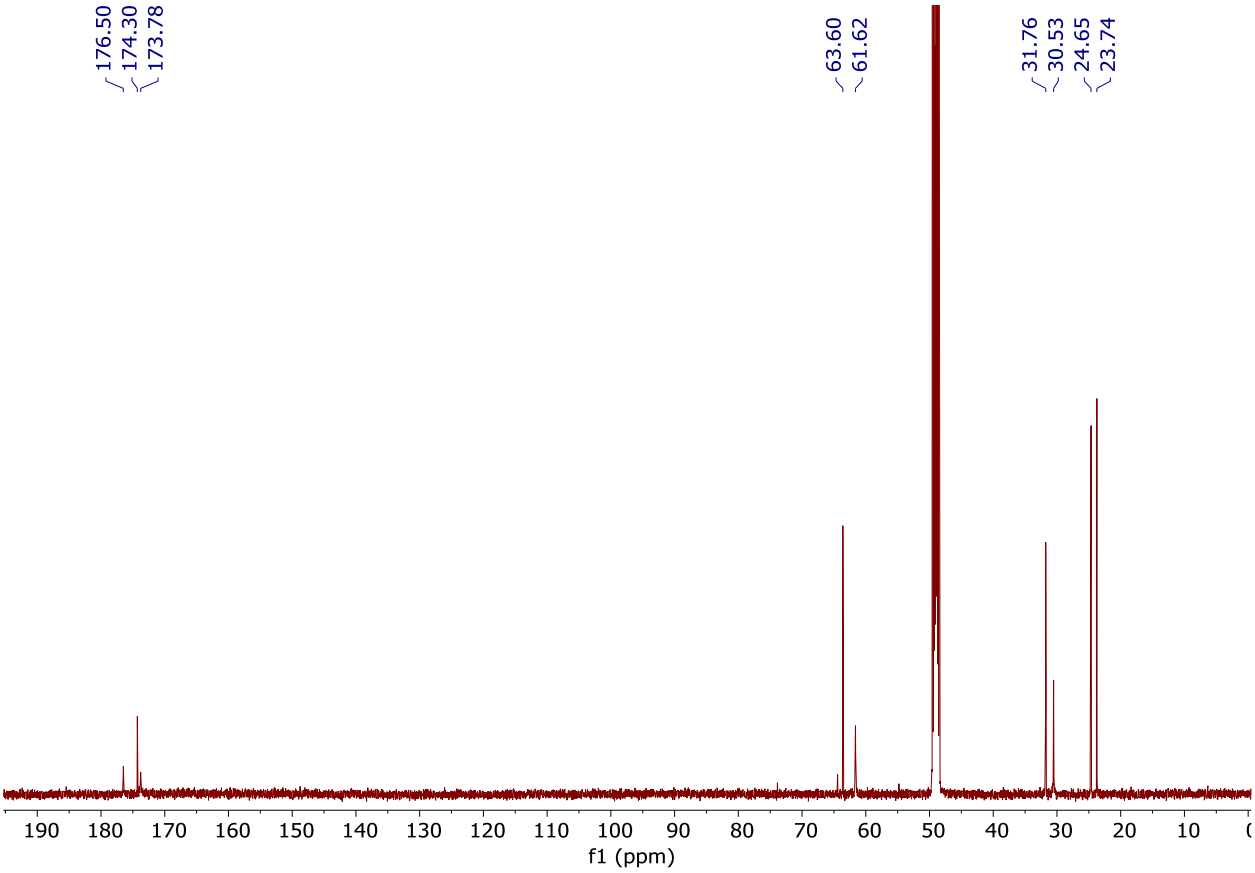


Figure AII.12: ^{1}H NMR spectrum of compound **3e** in CD_3OD

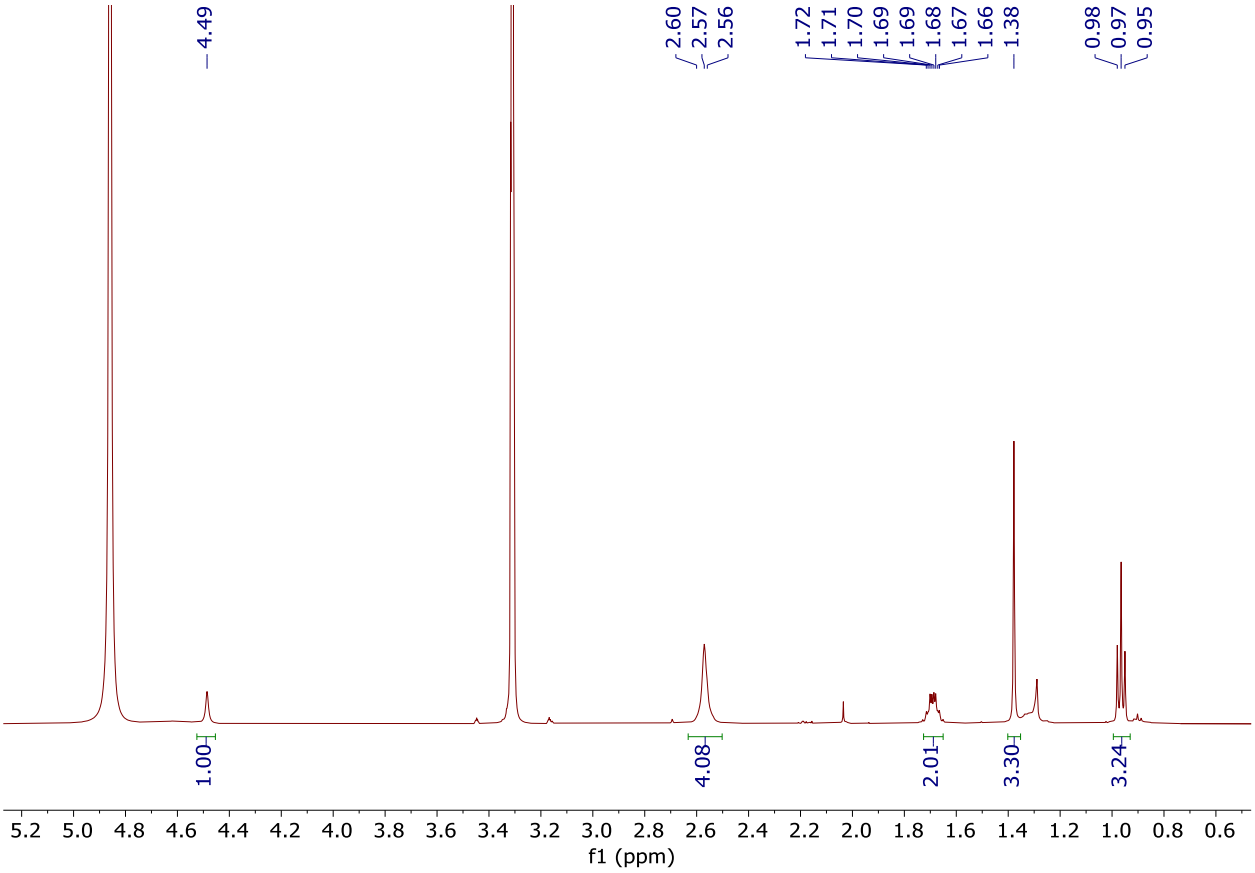


Figure AII.13: ^{13}C NMR spectrum of compound **3e** in CD_3OD

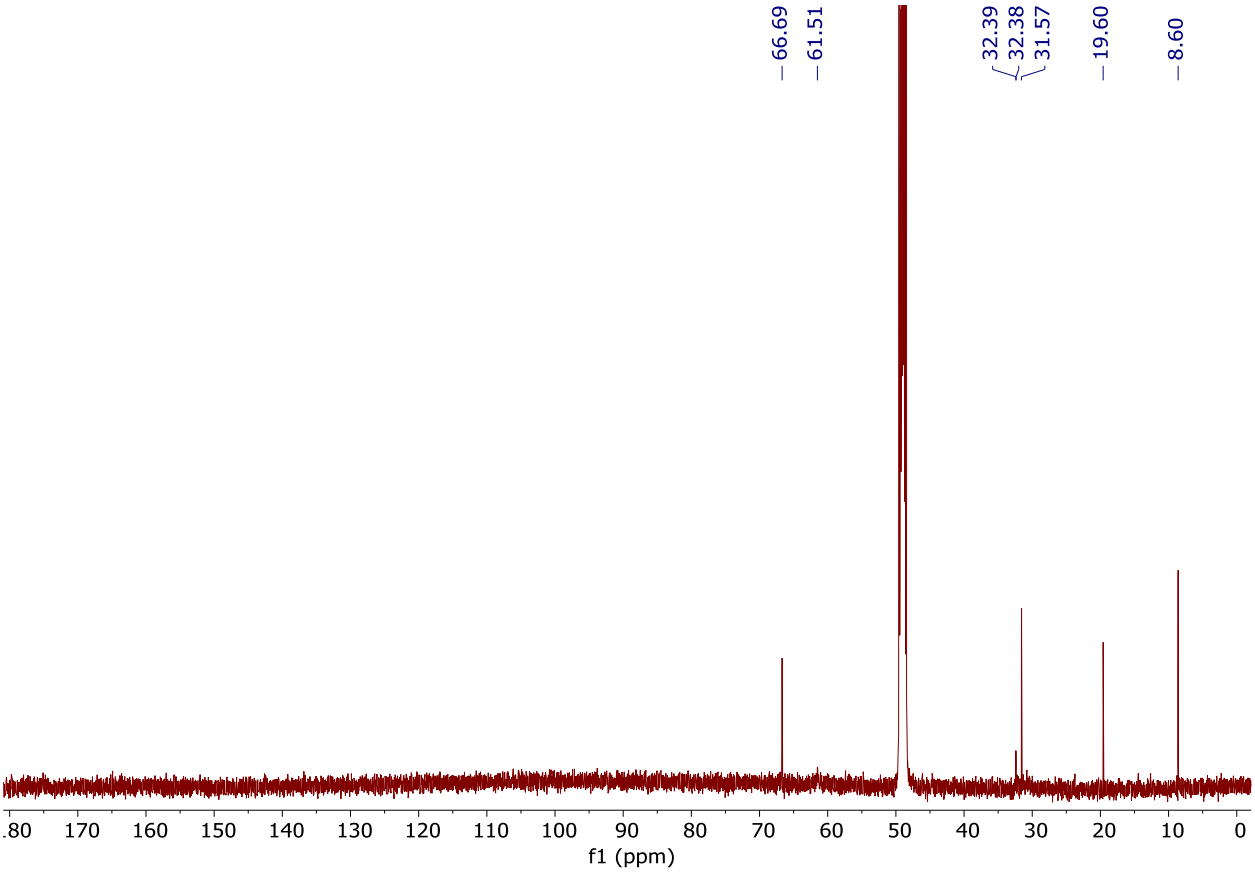


Figure AII.14: Multiplicity-edited ^{1}H-^{13}C HSQC of compound **3e** in CD_3OD

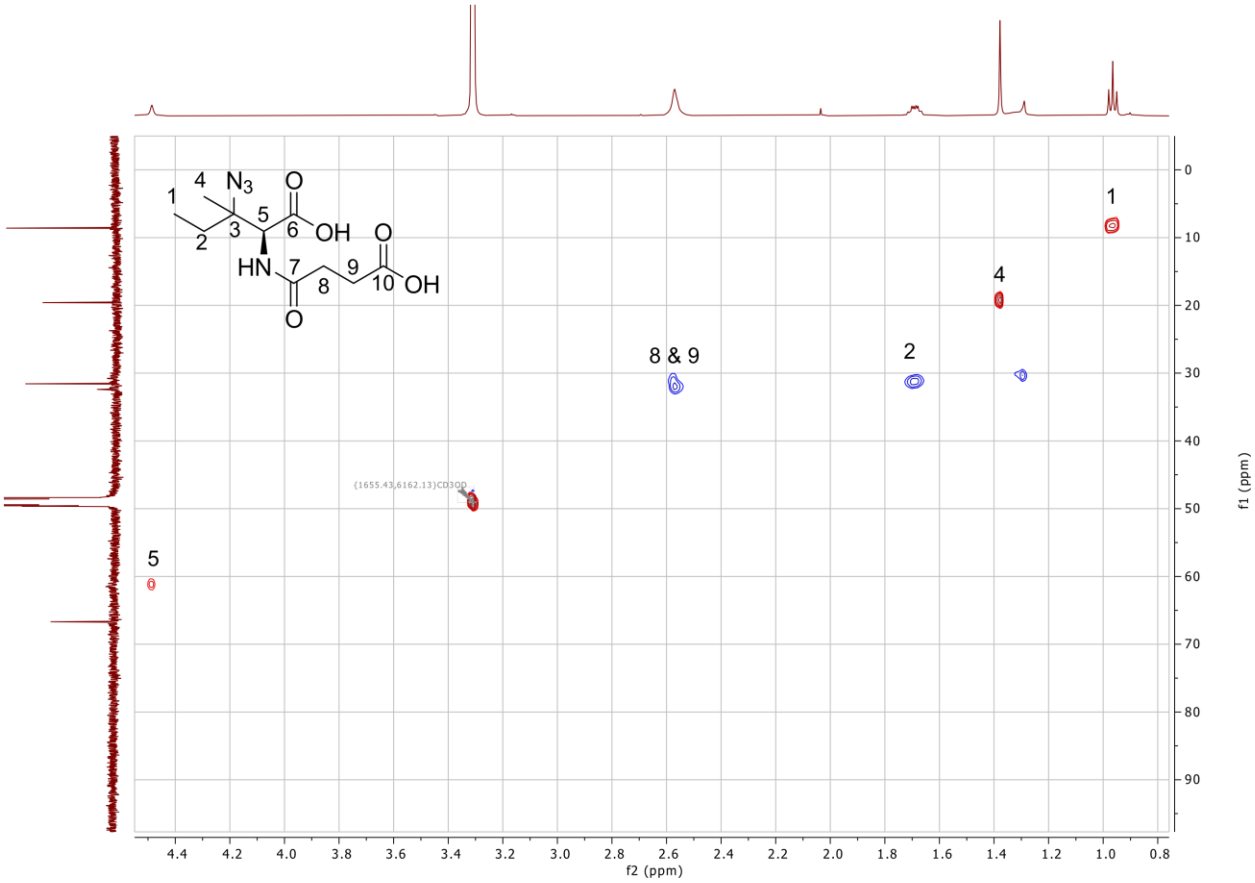


Figure AII.15: ^{1}H NMR spectrum of compound **4e** in CD_3OD

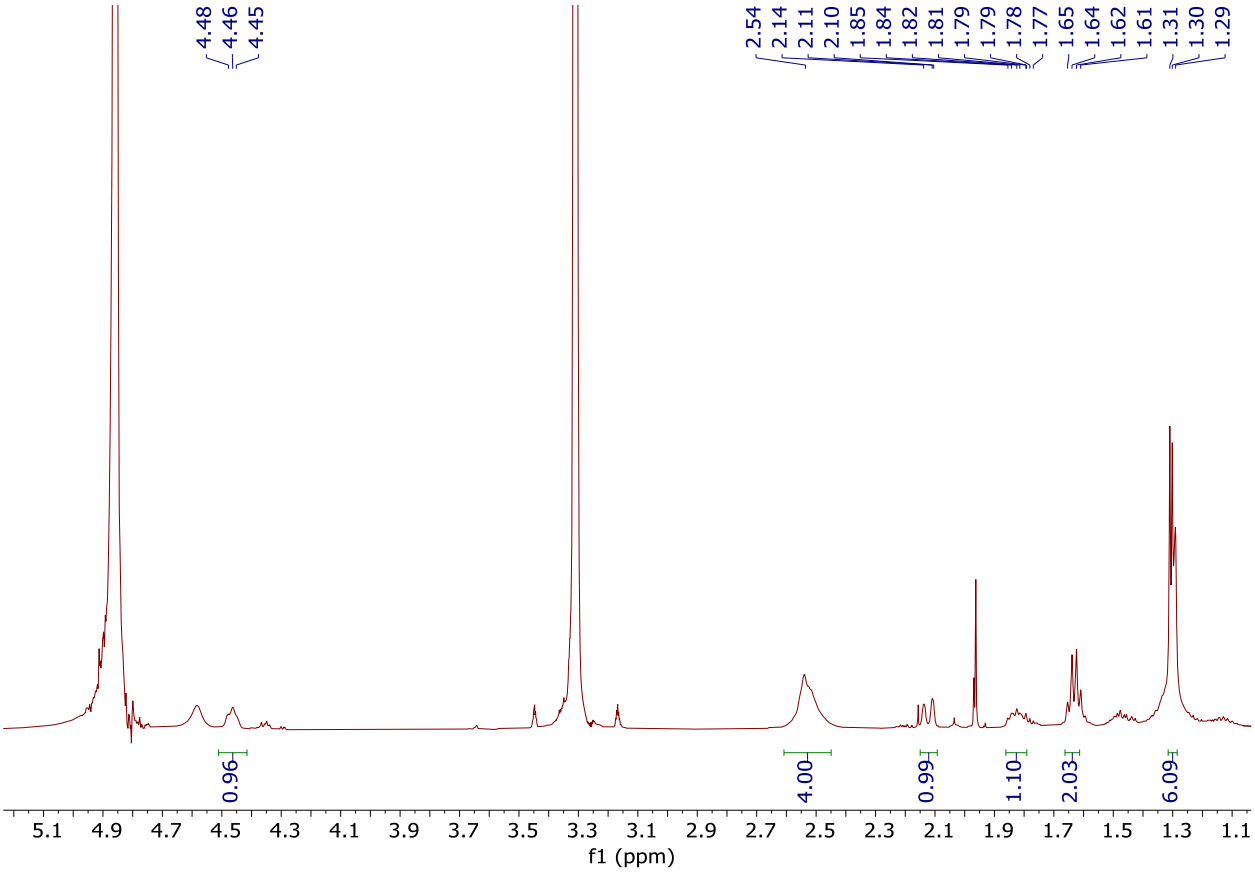


Figure AII.16: ^{13}C NMR spectrum of compound **4e** in CD_3OD

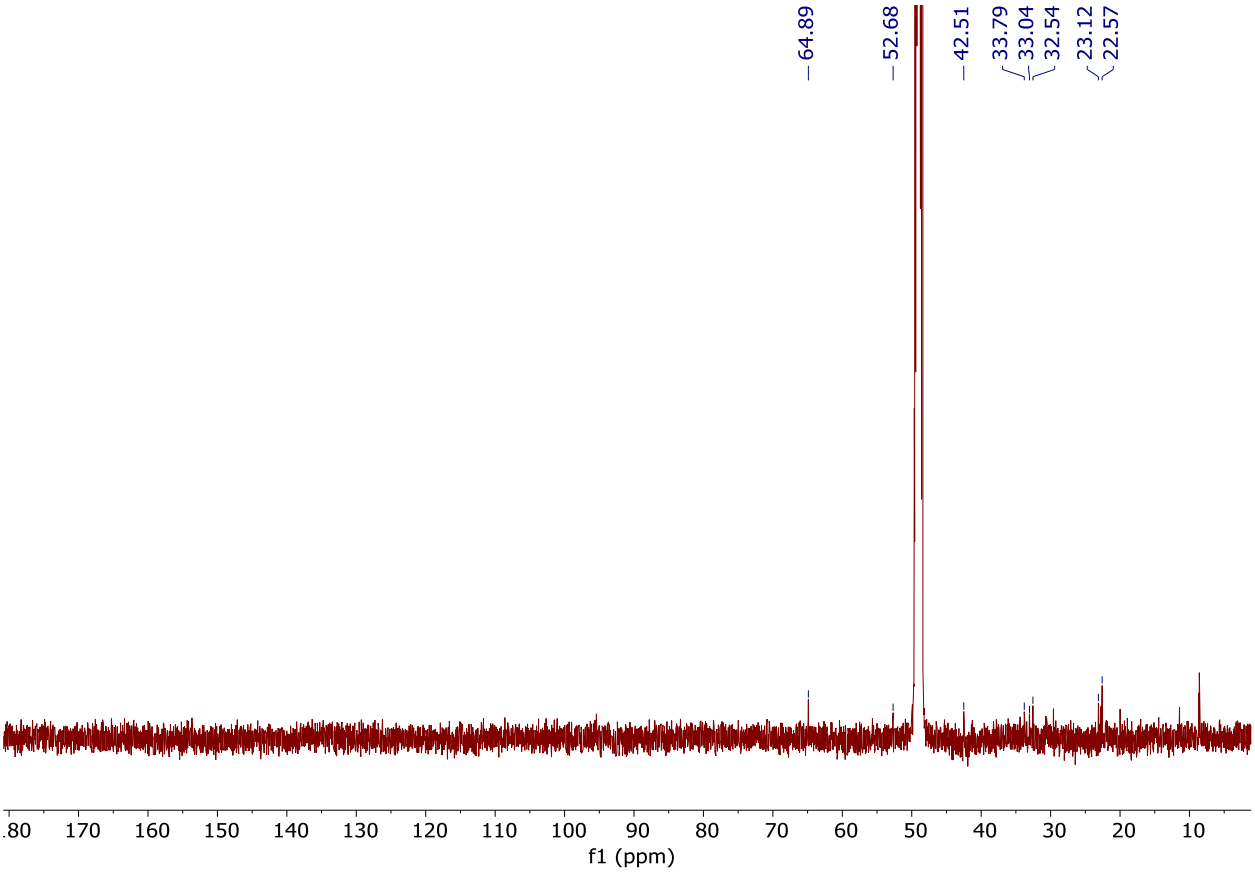


Figure AII.17: Multiplicity-edited ^{1}H-^{13}C HSQC of compound **4e** in CD_3OD

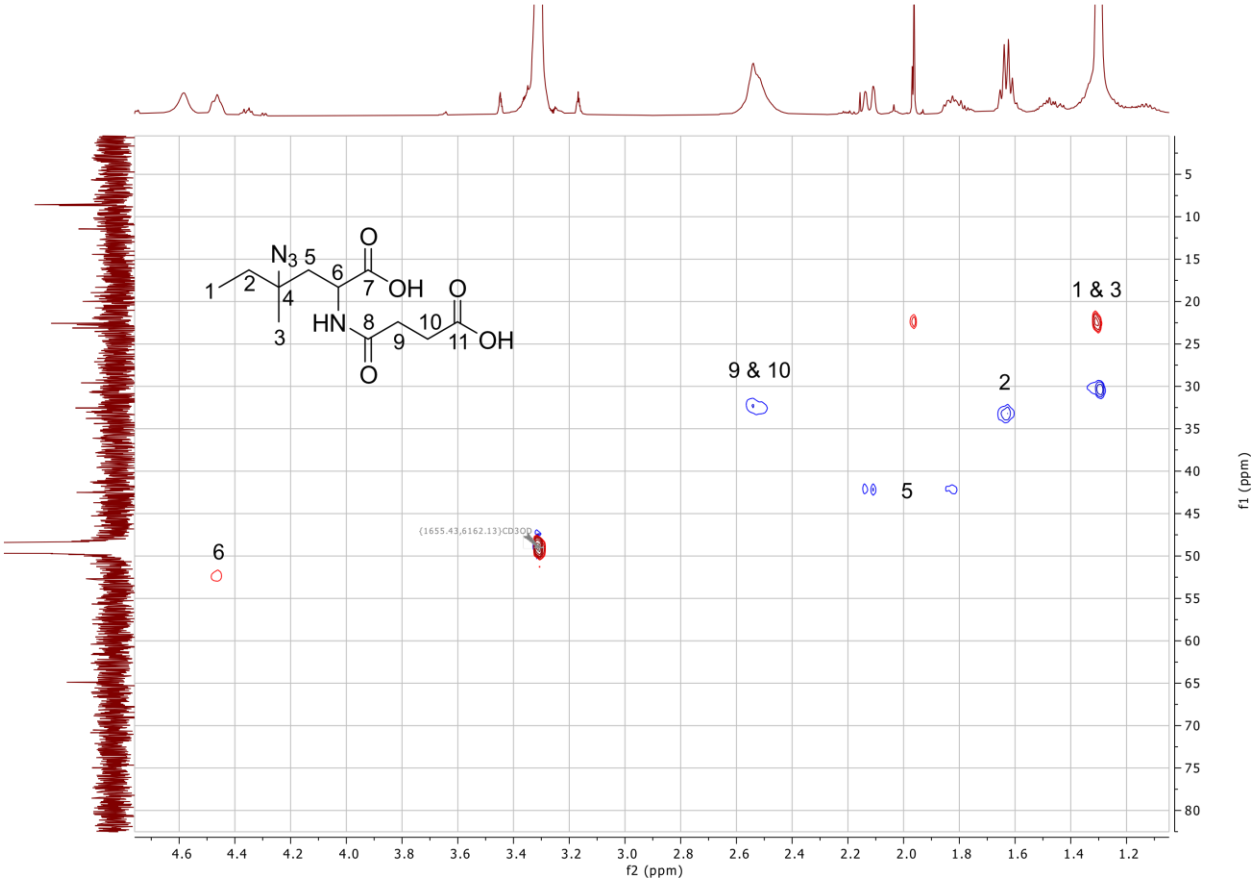


Figure AII.18: ^{1}H NMR spectrum of compound **5e** in CD_3OD

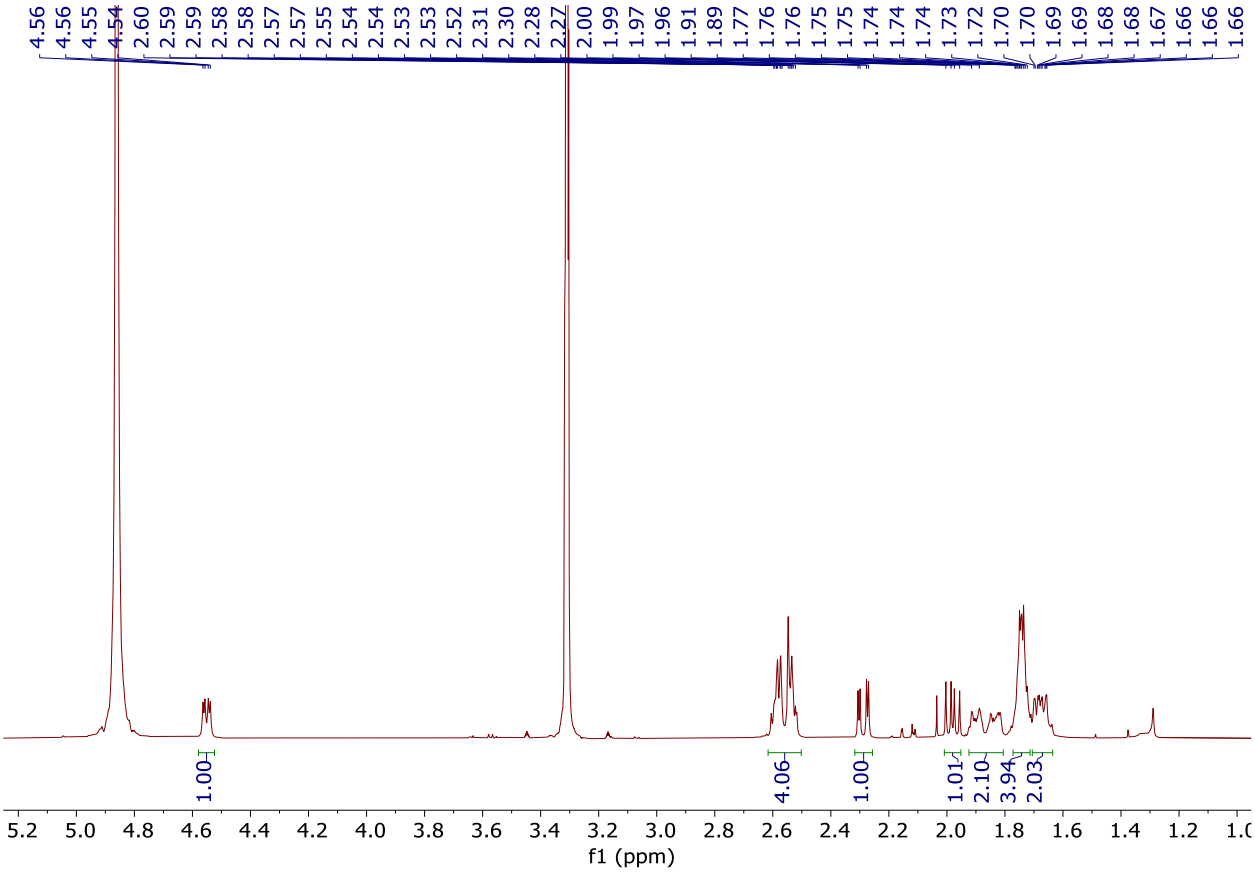


Figure AII.19: ^{13}C NMR spectrum of compound **5e** in CD_3OD

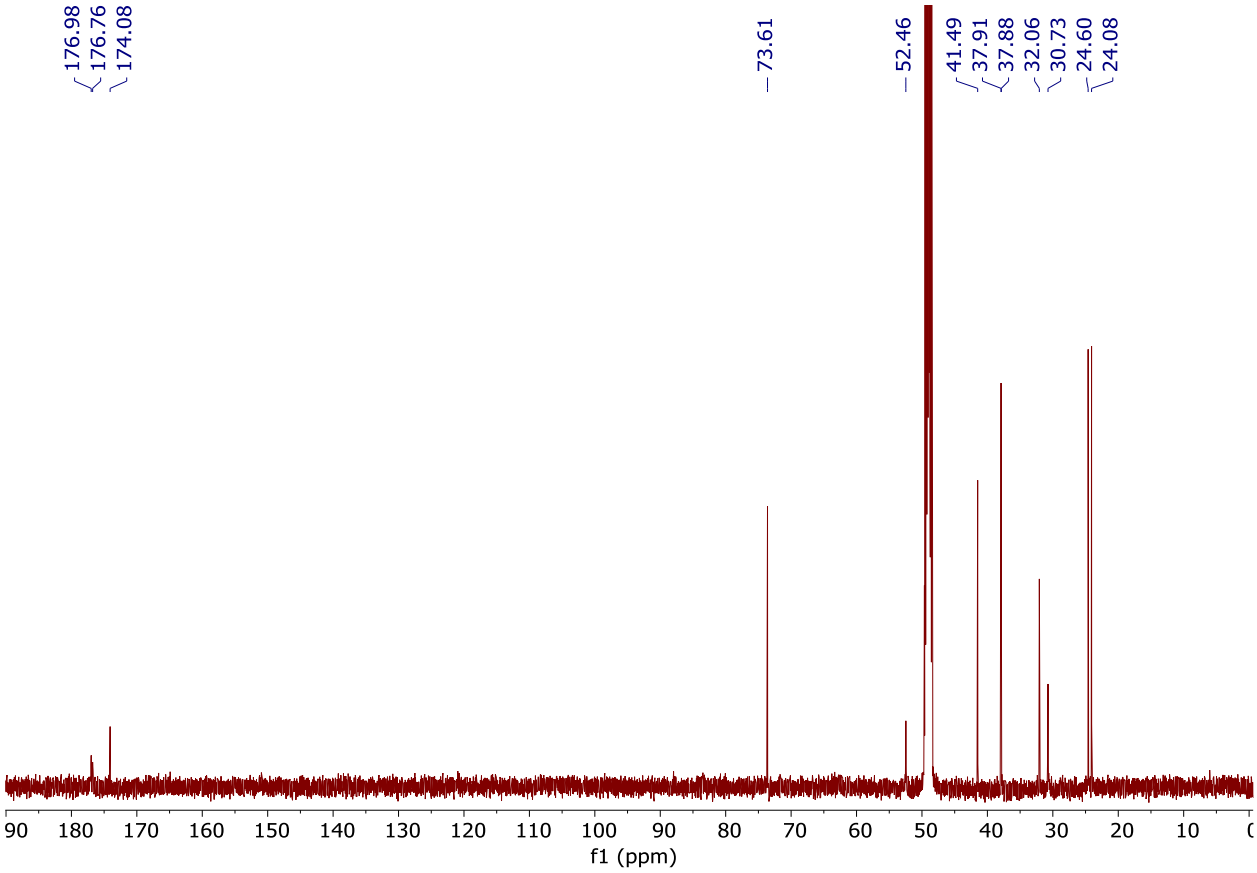


Figure AII.20: Multiplicity-edited ^{1}H-^{13}C HSQC of compound **5e** in CD_3OD

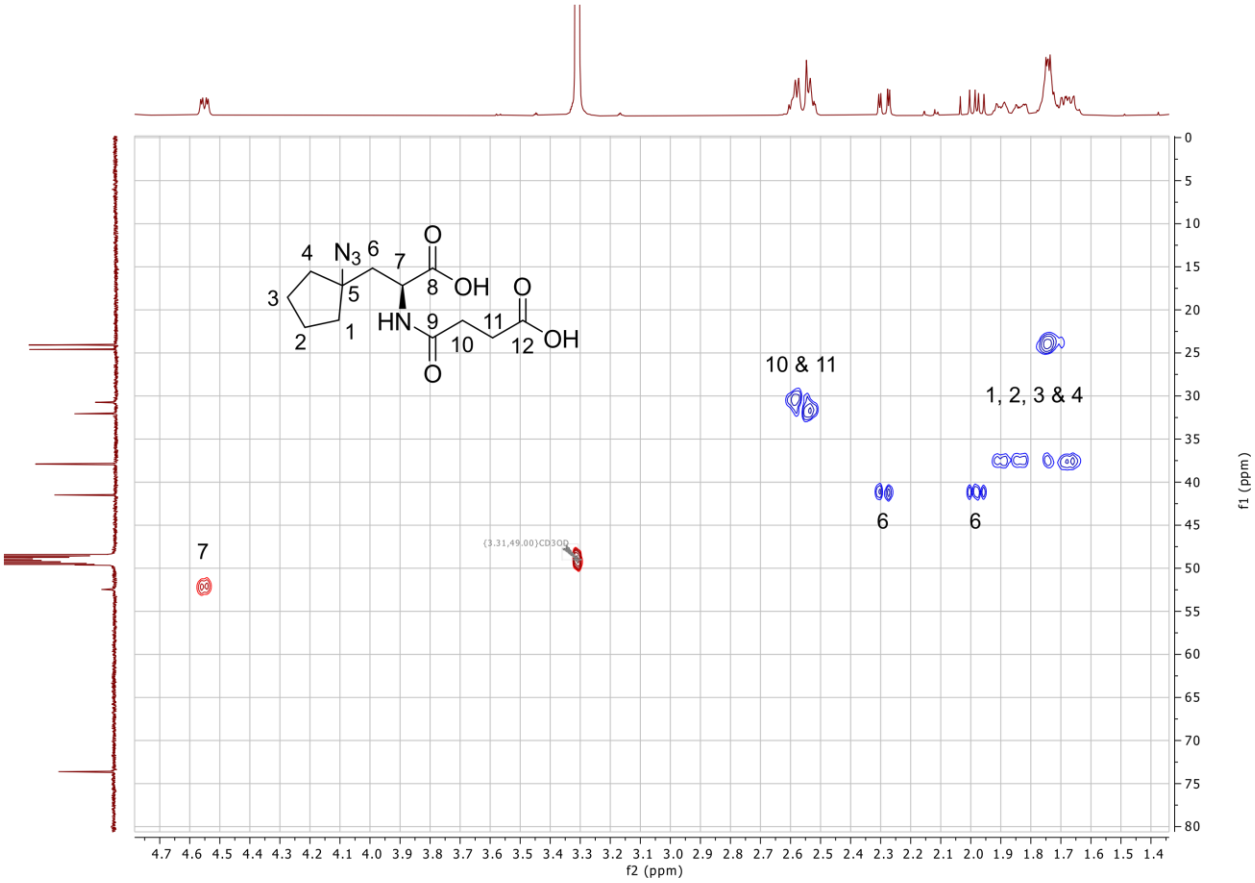


Figure AII.21: ^{1}H NMR spectrum of compound **6e** in CD_3OD

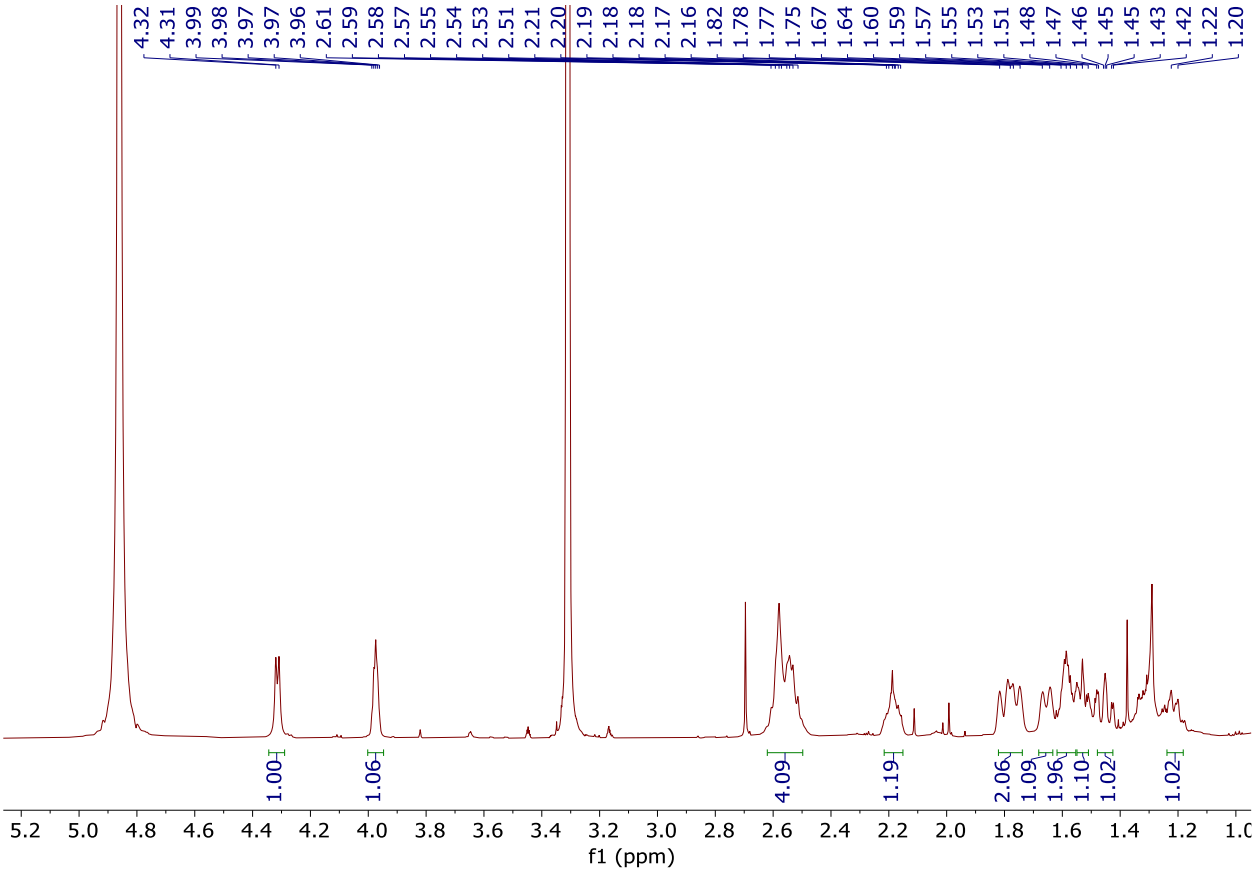


Figure AII.22: ^{13}C NMR spectrum of compound **6e** in CD_3OD

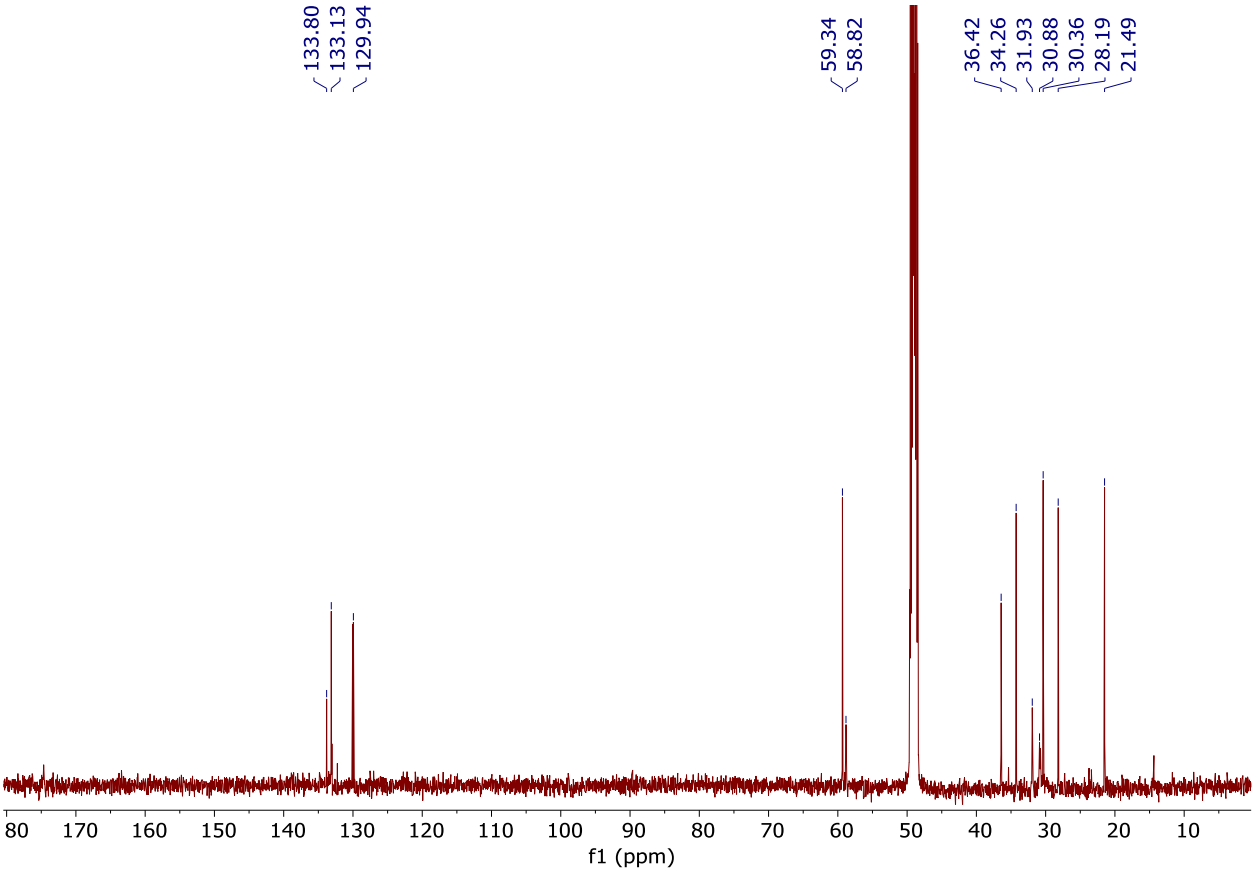


Figure AII.23: Multiplicity-edited ^{1}H-^{13}C HSQC of compound **6e** in CD_3OD

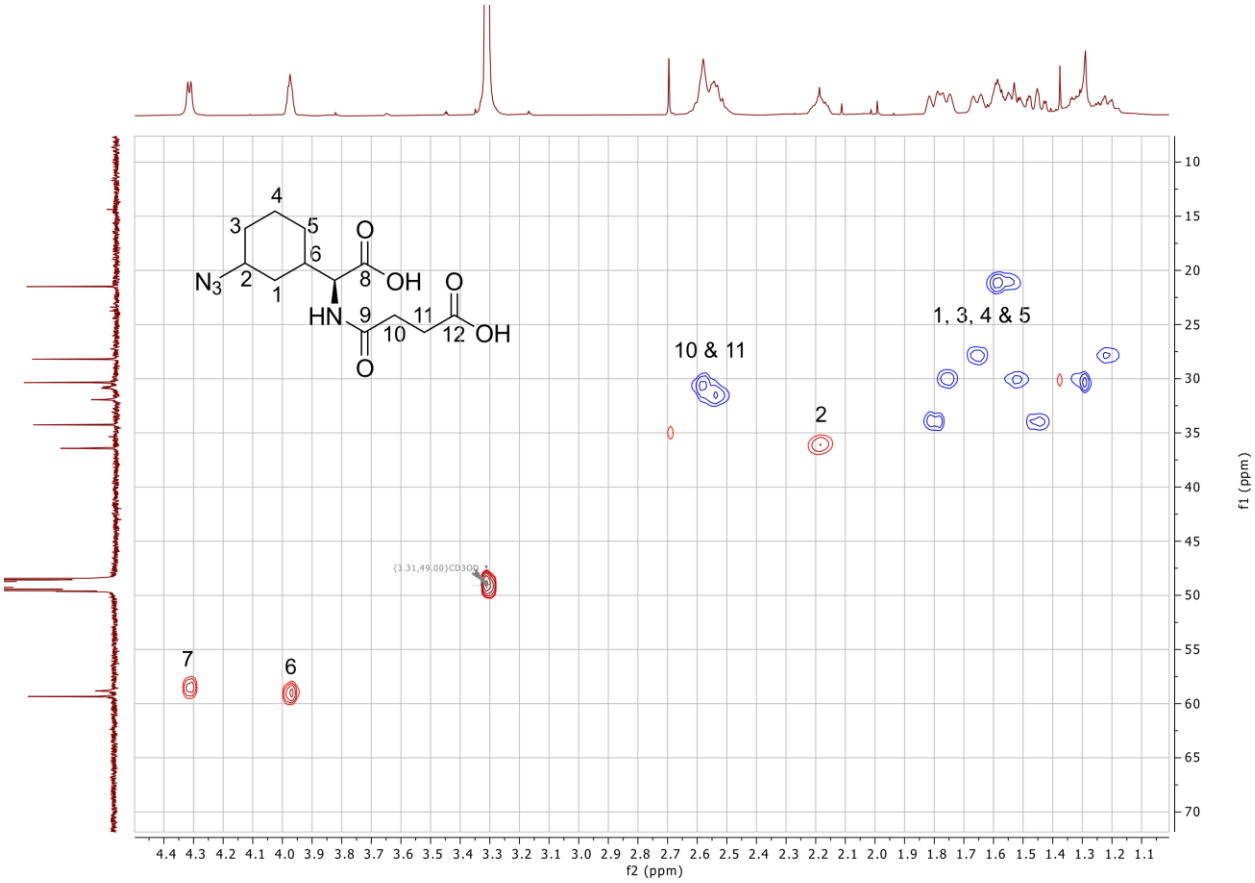


Figure AII.24: ^{1}H NMR spectra of compound **7e** in CD_3OD

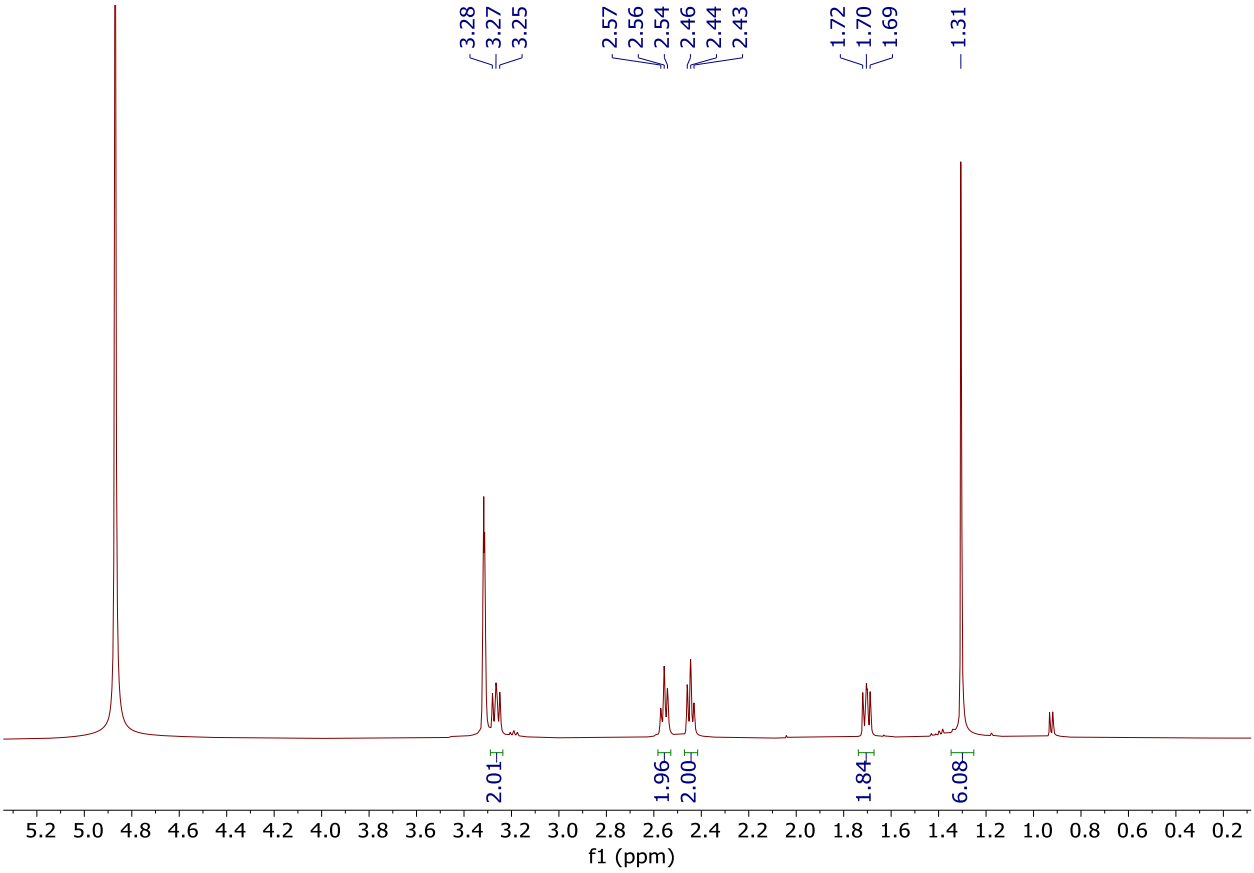


Figure AII.25: ^{13}C NMR spectra of compound **7e** in CD_3OD

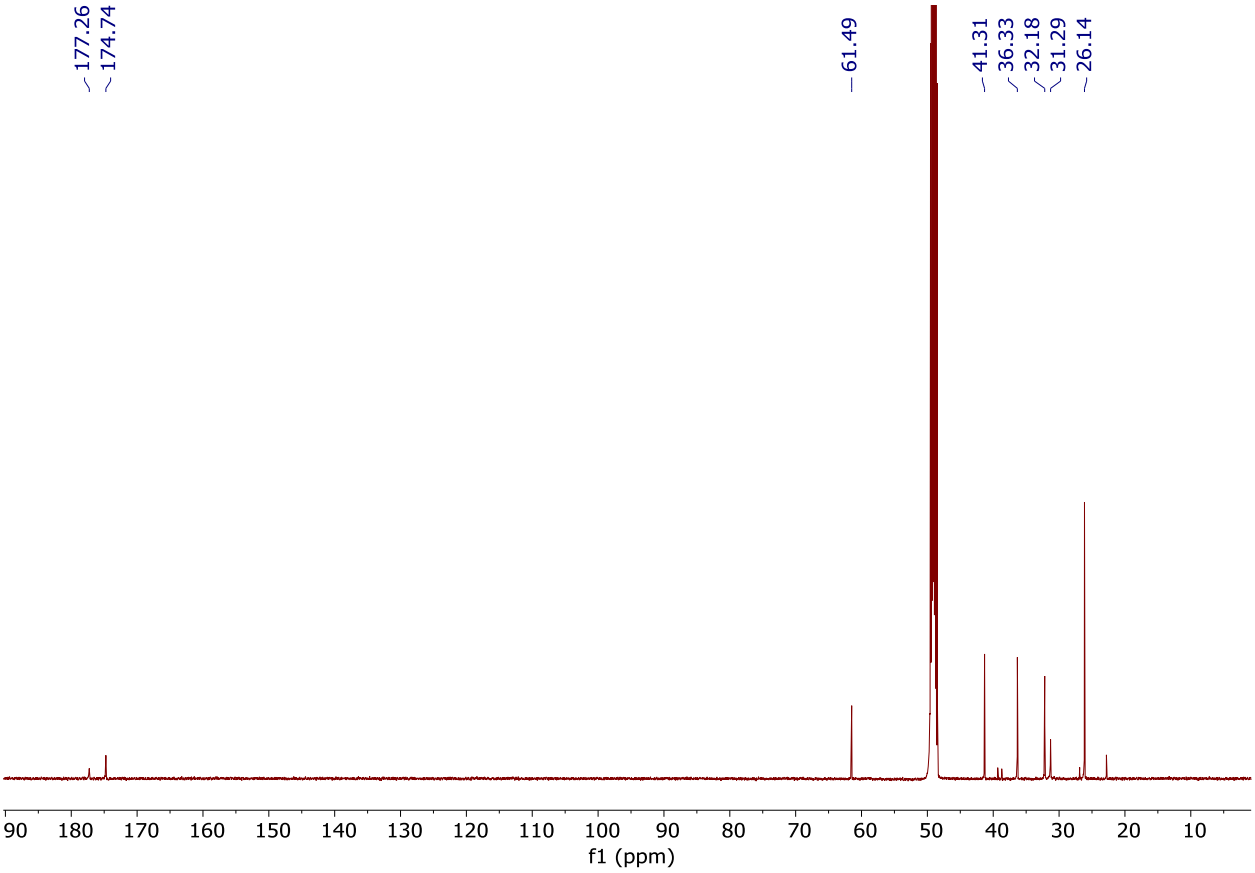


Figure AII.26: Multiplicity-edited ^{1}H-^{13}C HSQC of compound **7e** in CD_3OD

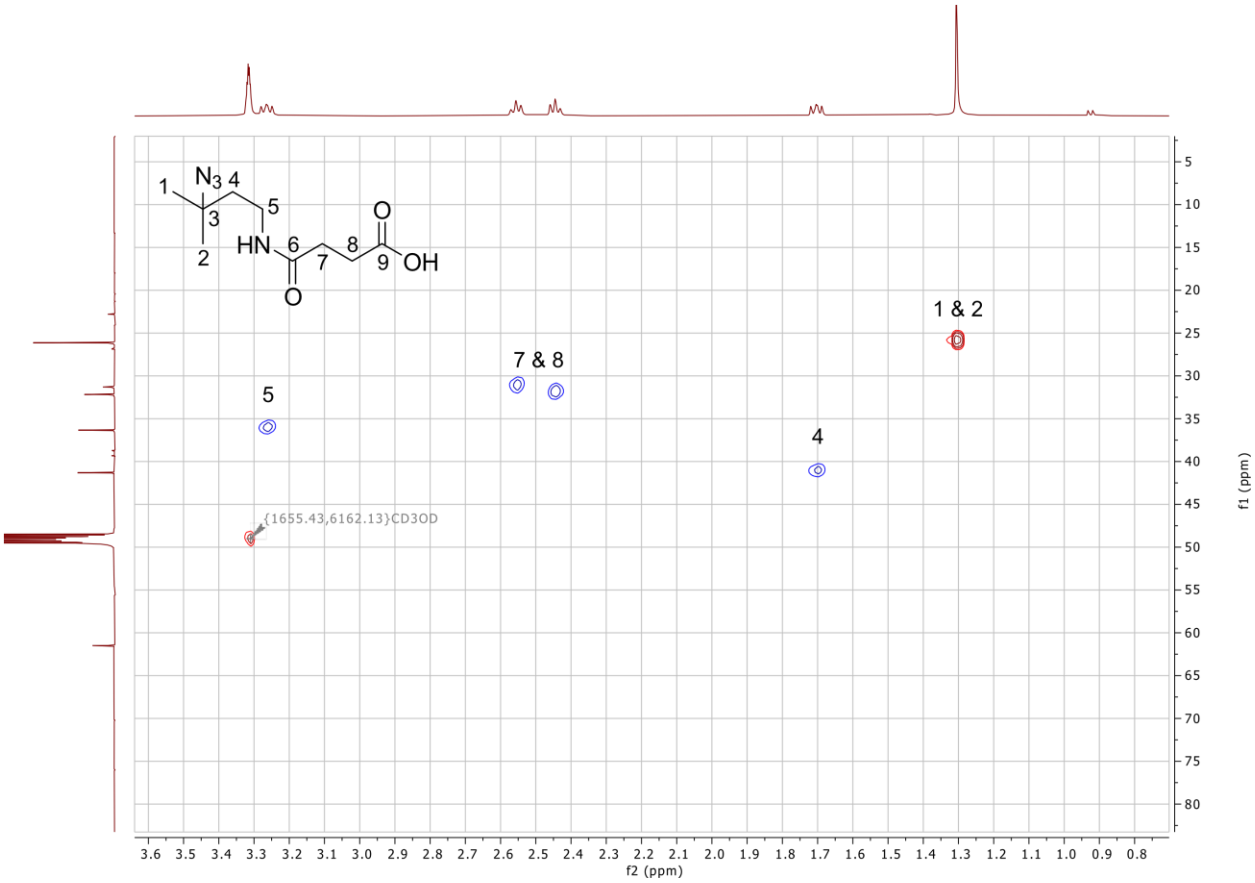


Figure AII.27: ^{1}H NMR spectrum of the product of **1e** desuccinylation in CD_3OD

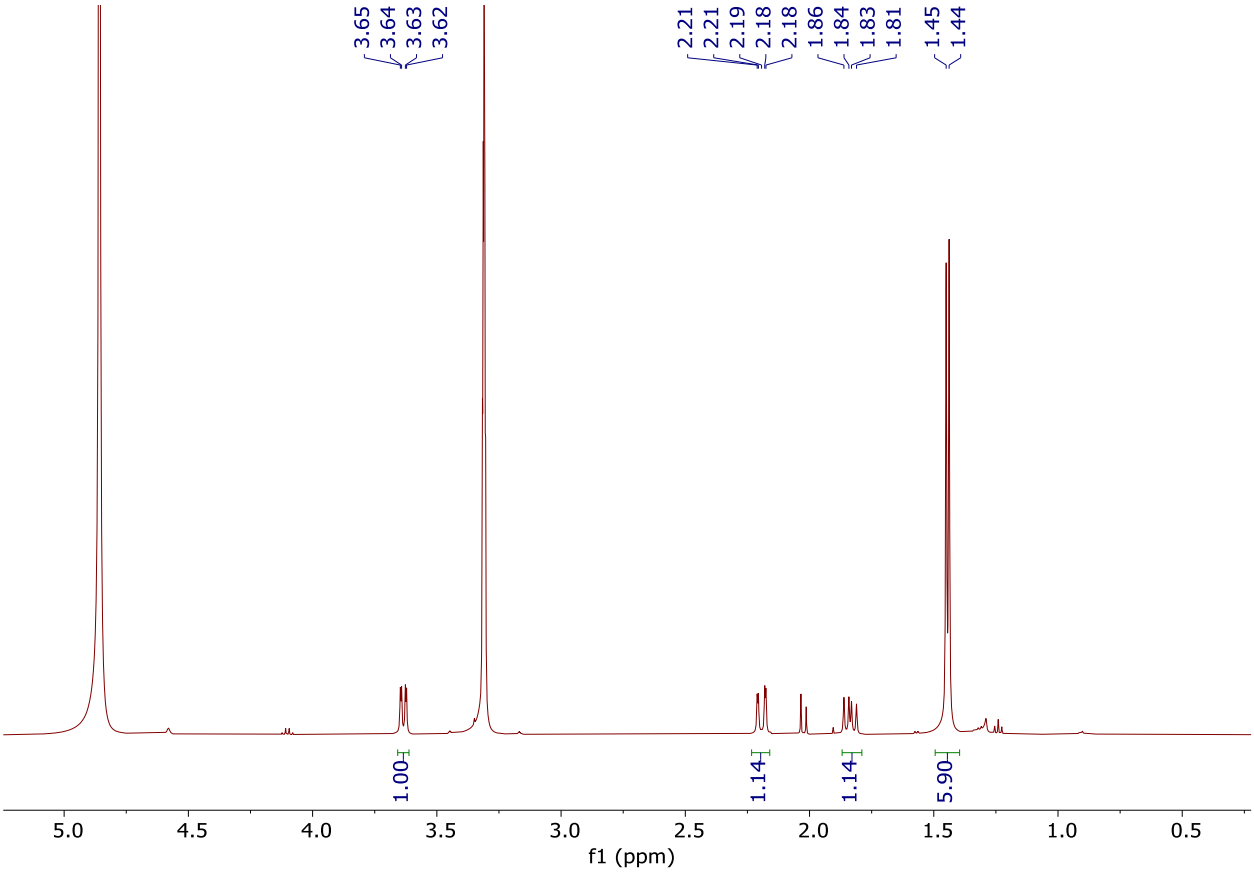


Figure AII.28: ^{13}C NMR spectrum of the product of **1e** desuccinylation in CD_3OD

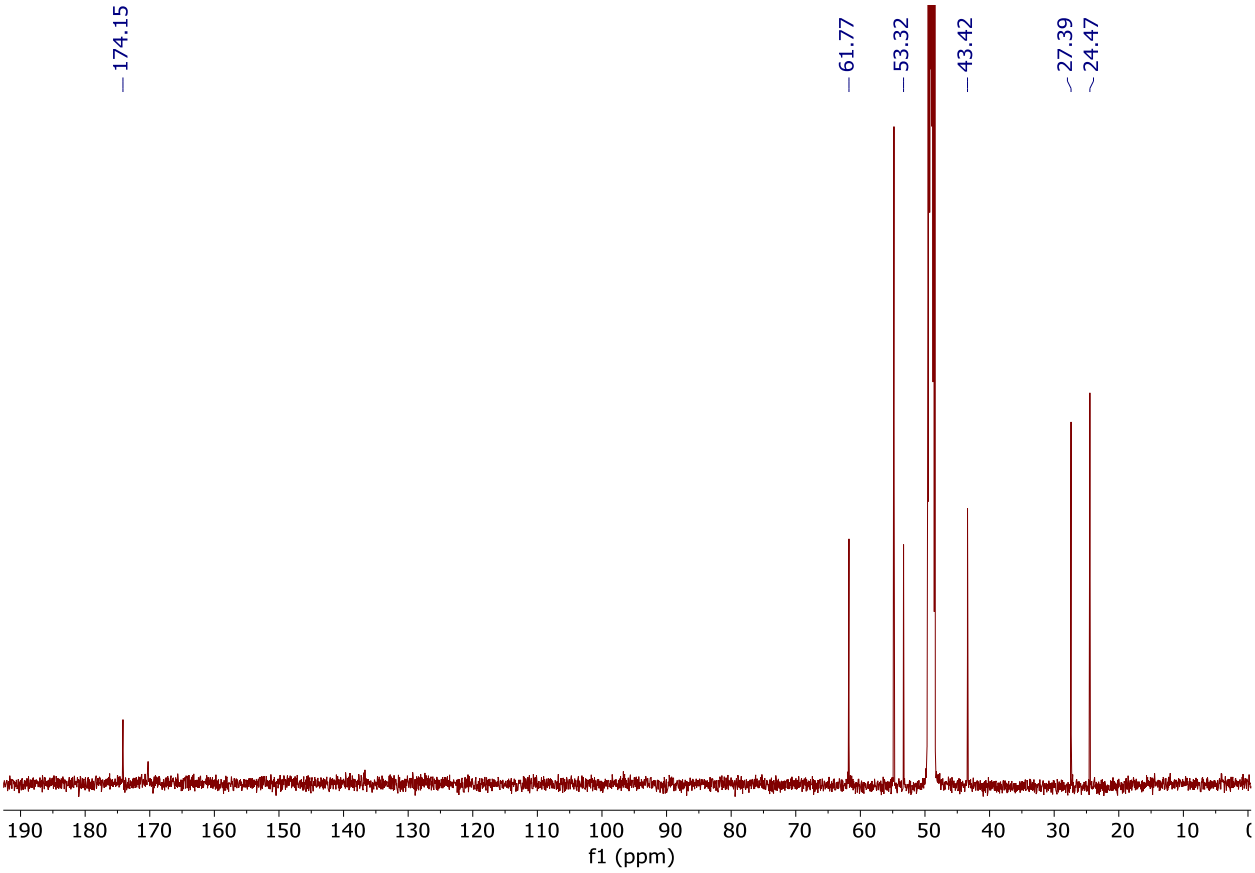


Figure AII.29: Multiplicity-edited ^{1}H-^{13}C HSQC of the product of **1e** desuccinylation in CD_3OD

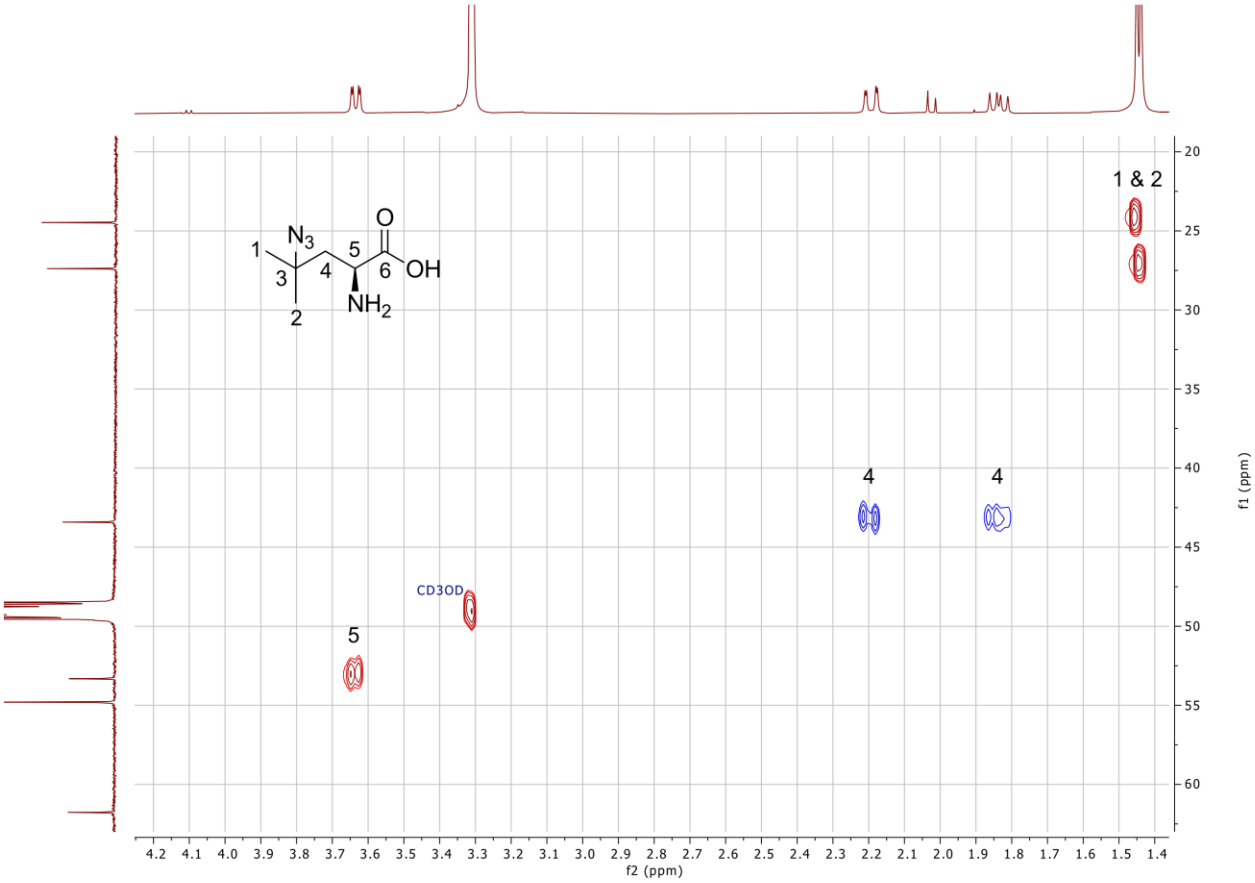


Figure AII.30: ^{1}H NMR spectrum of the product of **1e** reduction in CD_3OD

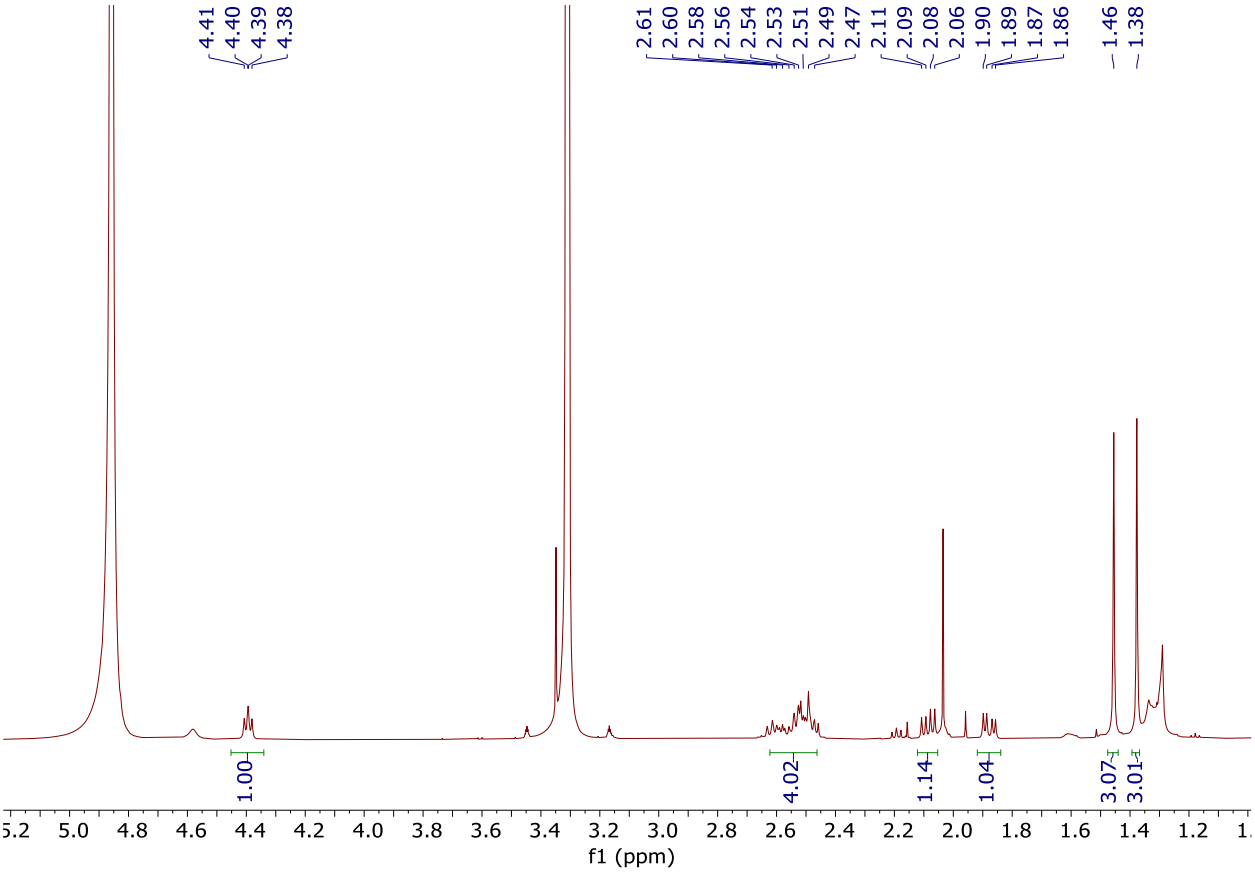


Figure AII.31: ^{13}C NMR spectrum of the product of **1e** reduction CD_3OD

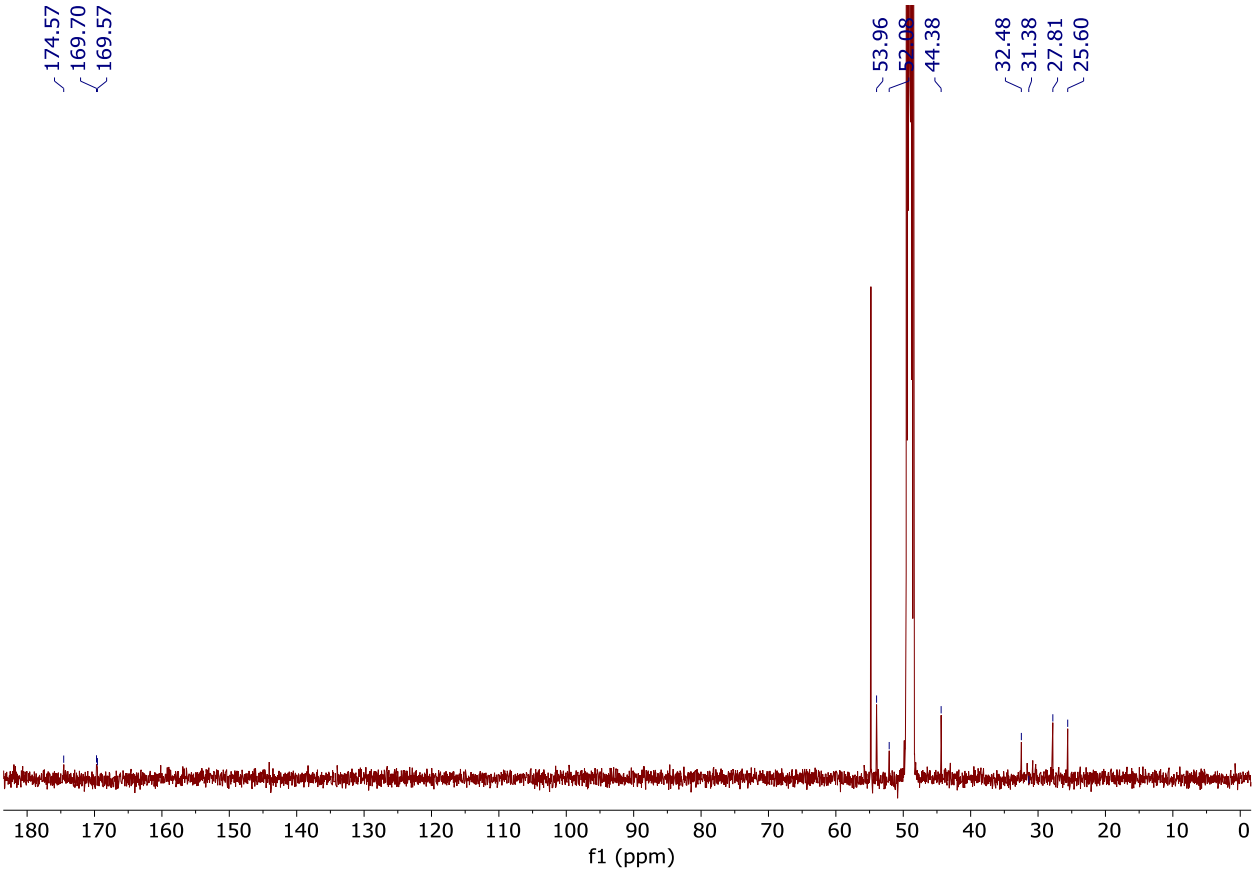
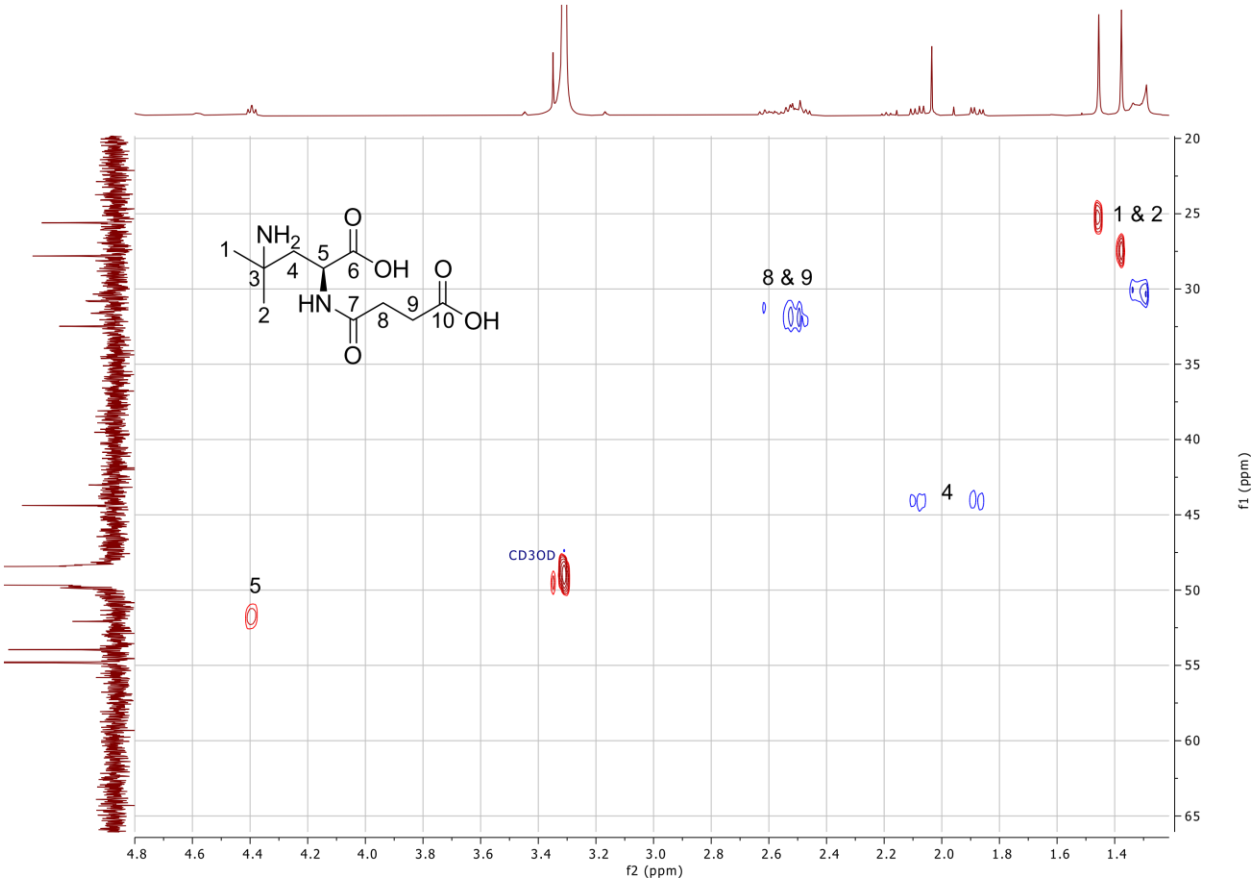


Figure AII.32: Multiplicity-edited ^{1}H-^{13}C HSQC of the product of **1e** reduction in CD_3OD



Appendix III

Representative data from Chapter 5

Table AIII.1: Processed data from screening of library 2-L

Each row corresponds to a single variant from the 2-L epPCR library. A variant's location is represented in the format *XY-Z*, with *X* = plate number, *Y* = row, and *Z* = column. Conversion to azidated product 1e is noted as *%1e*. *Parent* is the average conversion to 1e of all parent variants in a given plate, and *SD* is the corresponding standard deviation. *UB* is the upper bound defined as *Parent + SD*. *LW* is the lower bound defined as *Parent – SD*. *Bin* shows the category a variant was assigned to, between high (*H*), parent (*P*), low (*L*), and no activity (*NA*).

#	Location	%1e	Parent	SD	UB	LW	Bin	#	Location	%1e	Parent	SD	UB	LW	Bin
1	1A-01	0.593	35.507	5.512	41.019	29.995	L	451	6A-01	0.114	58.580	3.123	61.703	55.456	L
2	1A-02	0.000	35.507	5.512	41.019	29.995	NA	452	6A-02	0.089	58.580	3.123	61.703	55.456	NA
3	1A-03	34.487	35.507	5.512	41.019	29.995	P	453	6A-03	2.258	58.580	3.123	61.703	55.456	L
4	1A-04	25.131	35.507	5.512	41.019	29.995	L	454	6A-04	50.500	58.580	3.123	61.703	55.456	L
5	1A-05	0.073	35.507	5.512	41.019	29.995	NA	455	6A-05	19.470	58.580	3.123	61.703	55.456	L
6	1A-06	3.942	35.507	5.512	41.019	29.995	L	456	6A-06	1.007	58.580	3.123	61.703	55.456	L
7	1A-07	31.134	35.507	5.512	41.019	29.995	P	457	6A-07	1.591	58.580	3.123	61.703	55.456	L
8	1A-08	0.023	35.507	5.512	41.019	29.995	NA	458	6A-08	0.000	58.580	3.123	61.703	55.456	NA
9	1A-09	26.687	35.507	5.512	41.019	29.995	L	459	6A-09	38.817	58.580	3.123	61.703	55.456	L
10	1A-10	0.038	35.507	5.512	41.019	29.995	NA	460	6A-10	0.126	58.580	3.123	61.703	55.456	L
11	1A-11	1.079	35.507	5.512	41.019	29.995	L	461	6A-11	1.547	58.580	3.123	61.703	55.456	L
12	1A-12	4.480	35.507	5.512	41.019	29.995	L	462	6A-12	13.217	58.580	3.123	61.703	55.456	L
13	1B-01	0.000	35.507	5.512	41.019	29.995	NA	463	6B-01	0.040	58.580	3.123	61.703	55.456	NA
14	1B-02	7.570	35.507	5.512	41.019	29.995	L	464	6B-02	7.311	58.580	3.123	61.703	55.456	L
15	1B-03	0.276	35.507	5.512	41.019	29.995	L	465	6B-03	43.574	58.580	3.123	61.703	55.456	L
16	1B-04	2.204	35.507	5.512	41.019	29.995	L	466	6B-04	2.427	58.580	3.123	61.703	55.456	L
17	1B-05	4.884	35.507	5.512	41.019	29.995	L	467	6B-05	2.346	58.580	3.123	61.703	55.456	L
18	1B-06	0.000	35.507	5.512	41.019	29.995	NA	468	6B-06	0.561	58.580	3.123	61.703	55.456	L
19	1B-08	37.259	35.507	5.512	41.019	29.995	P	469	6B-08	0.037	58.580	3.123	61.703	55.456	NA
20	1B-09	0.053	35.507	5.512	41.019	29.995	NA	470	6B-09	0.015	58.580	3.123	61.703	55.456	NA
21	1B-10	0.037	35.507	5.512	41.019	29.995	NA	471	6B-10	0.140	58.580	3.123	61.703	55.456	L
22	1B-11	0.356	35.507	5.512	41.019	29.995	L	472	6B-11	0.453	58.580	3.123	61.703	55.456	L
23	1B-12	0.054	35.507	5.512	41.019	29.995	NA	473	6B-12	0.969	58.580	3.123	61.703	55.456	L
24	1C-01	17.329	35.507	5.512	41.019	29.995	L	474	6C-01	0.062	58.580	3.123	61.703	55.456	NA
25	1C-02	0.049	35.507	5.512	41.019	29.995	NA	475	6C-02	15.130	58.580	3.123	61.703	55.456	L
26	1C-04	0.062	35.507	5.512	41.019	29.995	NA	476	6C-04	0.147	58.580	3.123	61.703	55.456	L
27	1C-05	45.293	35.507	5.512	41.019	29.995	H	477	6C-05	0.307	58.580	3.123	61.703	55.456	L

Table AIII.1 (*continued*)

#	Location	%1e	Parent	SD	UB	LW	Bin	#	Location	%1e	Parent	SD	UB	LW	Bin
28	1C-06	6.322	35.507	5.512	41.019	29.995	L	478	6C-06	66.569	58.580	3.123	61.703	55.456	H
29	1C-07	0.224	35.507	5.512	41.019	29.995	L	479	6C-07	46.329	58.580	3.123	61.703	55.456	L
30	1C-08	41.633	35.507	5.512	41.019	29.995	H	480	6C-08	0.108	58.580	3.123	61.703	55.456	L
31	1C-09	4.015	35.507	5.512	41.019	29.995	L	481	6C-09	0.027	58.580	3.123	61.703	55.456	NA
32	1C-10	0.075	35.507	5.512	41.019	29.995	NA	482	6C-10	0.047	58.580	3.123	61.703	55.456	NA
33	1C-11	39.665	35.507	5.512	41.019	29.995	P	483	6C-11	27.154	58.580	3.123	61.703	55.456	L
34	1C-12	0.047	35.507	5.512	41.019	29.995	NA	484	6C-12	0.084	58.580	3.123	61.703	55.456	NA
35	1D-01	0.067	35.507	5.512	41.019	29.995	NA	485	6D-01	0.056	58.580	3.123	61.703	55.456	NA
36	1D-02	18.876	35.507	5.512	41.019	29.995	L	486	6D-02	0.100	58.580	3.123	61.703	55.456	NA
37	1D-03	3.398	35.507	5.512	41.019	29.995	L	487	6D-03	0.224	58.580	3.123	61.703	55.456	L
38	1D-04	8.248	35.507	5.512	41.019	29.995	L	488	6D-04	64.757	58.580	3.123	61.703	55.456	H
39	1D-05	0.225	35.507	5.512	41.019	29.995	L	489	6D-05	0.883	58.580	3.123	61.703	55.456	L
40	1D-07	0.040	35.507	5.512	41.019	29.995	NA	490	6D-07	0.124	58.580	3.123	61.703	55.456	L
41	1D-08	0.048	35.507	5.512	41.019	29.995	NA	491	6D-08	1.336	58.580	3.123	61.703	55.456	L
42	1D-09	0.178	35.507	5.512	41.019	29.995	L	492	6D-09	0.036	58.580	3.123	61.703	55.456	NA
43	1D-10	0.189	35.507	5.512	41.019	29.995	L	493	6D-10	0.004	58.580	3.123	61.703	55.456	NA
44	1D-11	0.112	35.507	5.512	41.019	29.995	L	494	6D-11	0.058	58.580	3.123	61.703	55.456	NA
45	1D-12	0.000	35.507	5.512	41.019	29.995	NA	495	6D-12	23.272	58.580	3.123	61.703	55.456	L
46	1E-01	1.698	35.507	5.512	41.019	29.995	L	496	6E-01	23.331	58.580	3.123	61.703	55.456	L
47	1E-02	0.000	35.507	5.512	41.019	29.995	NA	497	6E-02	0.021	58.580	3.123	61.703	55.456	NA
48	1E-03	5.569	35.507	5.512	41.019	29.995	L	498	6E-03	0.026	58.580	3.123	61.703	55.456	NA
49	1E-04	0.242	35.507	5.512	41.019	29.995	L	499	6E-04	0.394	58.580	3.123	61.703	55.456	L
50	1E-05	0.148	35.507	5.512	41.019	29.995	L	500	6E-05	0.882	58.580	3.123	61.703	55.456	L
51	1E-06	42.758	35.507	5.512	41.019	29.995	H	501	6E-06	0.300	58.580	3.123	61.703	55.456	L
52	1E-07	0.787	35.507	5.512	41.019	29.995	L	502	6E-07	0.071	58.580	3.123	61.703	55.456	NA
53	1E-08	0.082	35.507	5.512	41.019	29.995	NA	503	6E-08	0.147	58.580	3.123	61.703	55.456	L
54	1E-09	5.310	35.507	5.512	41.019	29.995	L	504	6E-09	0.602	58.580	3.123	61.703	55.456	L
55	1E-10	0.016	35.507	5.512	41.019	29.995	NA	505	6E-10	0.009	58.580	3.123	61.703	55.456	NA
56	1E-12	5.250	35.507	5.512	41.019	29.995	L	506	6E-12	0.063	58.580	3.123	61.703	55.456	NA
57	1F-01	0.000	35.507	5.512	41.019	29.995	NA	507	6F-01	4.177	58.580	3.123	61.703	55.456	L
58	1F-02	14.352	35.507	5.512	41.019	29.995	L	508	6F-02	42.048	58.580	3.123	61.703	55.456	L
59	1F-03	9.029	35.507	5.512	41.019	29.995	L	509	6F-03	0.198	58.580	3.123	61.703	55.456	L
60	1F-05	22.898	35.507	5.512	41.019	29.995	L	510	6F-05	0.049	58.580	3.123	61.703	55.456	NA
61	1F-06	35.398	35.507	5.512	41.019	29.995	P	511	6F-06	0.023	58.580	3.123	61.703	55.456	NA
62	1F-07	7.932	35.507	5.512	41.019	29.995	L	512	6F-07	59.425	58.580	3.123	61.703	55.456	P
63	1F-08	43.765	35.507	5.512	41.019	29.995	H	513	6F-08	0.063	58.580	3.123	61.703	55.456	NA
64	1F-09	0.000	35.507	5.512	41.019	29.995	NA	514	6F-09	0.060	58.580	3.123	61.703	55.456	NA
65	1F-10	0.050	35.507	5.512	41.019	29.995	NA	515	6F-10	0.008	58.580	3.123	61.703	55.456	NA
66	1F-11	0.112	35.507	5.512	41.019	29.995	L	516	6F-11	0.120	58.580	3.123	61.703	55.456	L
67	1F-12	0.010	35.507	5.512	41.019	29.995	NA	517	6F-12	63.433	58.580	3.123	61.703	55.456	H
68	1G-01	17.974	35.507	5.512	41.019	29.995	L	518	6G-01	53.401	58.580	3.123	61.703	55.456	L
69	1G-02	0.091	35.507	5.512	41.019	29.995	NA	519	6G-02	0.265	58.580	3.123	61.703	55.456	L
70	1G-04	30.590	35.507	5.512	41.019	29.995	P	520	6G-04	2.279	58.580	3.123	61.703	55.456	L
71	1G-05	0.189	35.507	5.512	41.019	29.995	L	521	6G-05	0.071	58.580	3.123	61.703	55.456	NA
72	1G-06	2.174	35.507	5.512	41.019	29.995	L	522	6G-06	57.254	58.580	3.123	61.703	55.456	P
73	1G-07	0.019	35.507	5.512	41.019	29.995	NA	523	6G-07	0.043	58.580	3.123	61.703	55.456	NA
74	1G-08	35.080	35.507	5.512	41.019	29.995	P	524	6G-08	0.226	58.580	3.123	61.703	55.456	L

Table AIII.1 (*continued*)

#	Location	%1e	Parent	SD	UB	LW	Bin	#	Location	%1e	Parent	SD	UB	LW	Bin
75	1G-09	0.000	35.507	5.512	41.019	29.995	NA	525	6G-09	1.193	58.580	3.123	61.703	55.456	L
76	1G-10	0.000	35.507	5.512	41.019	29.995	NA	526	6G-10	0.727	58.580	3.123	61.703	55.456	L
77	1G-11	0.136	35.507	5.512	41.019	29.995	L	527	6G-11	1.177	58.580	3.123	61.703	55.456	L
78	1G-12	0.972	35.507	5.512	41.019	29.995	L	528	6G-12	1.821	58.580	3.123	61.703	55.456	L
79	1H-01	0.116	35.507	5.512	41.019	29.995	L	529	6H-01	0.054	58.580	3.123	61.703	55.456	NA
80	1H-02	0.130	35.507	5.512	41.019	29.995	L	530	6H-02	6.832	58.580	3.123	61.703	55.456	L
81	1H-03	4.053	35.507	5.512	41.019	29.995	L	531	6H-03	5.527	58.580	3.123	61.703	55.456	L
82	1H-04	0.117	35.507	5.512	41.019	29.995	L	532	6H-04	0.743	58.580	3.123	61.703	55.456	L
83	1H-05	0.089	35.507	5.512	41.019	29.995	NA	533	6H-05	0.022	58.580	3.123	61.703	55.456	NA
84	1H-06	0.000	35.507	5.512	41.019	29.995	NA	534	6H-06	0.069	58.580	3.123	61.703	55.456	NA
85	1H-07	0.041	35.507	5.512	41.019	29.995	NA	535	6H-07	1.806	58.580	3.123	61.703	55.456	L
86	1H-08	0.046	35.507	5.512	41.019	29.995	NA	536	6H-08	1.853	58.580	3.123	61.703	55.456	L
87	1H-09	0.038	35.507	5.512	41.019	29.995	NA	537	6H-09	0.032	58.580	3.123	61.703	55.456	NA
88	1H-10	15.089	35.507	5.512	41.019	29.995	L	538	6H-10	0.048	58.580	3.123	61.703	55.456	NA
89	1H-11	0.106	35.507	5.512	41.019	29.995	L	539	6H-11	59.272	58.580	3.123	61.703	55.456	P
90	1H-12	0.166	35.507	5.512	41.019	29.995	L	540	6H-12	0.133	58.580	3.123	61.703	55.456	L
91	2A-01	0.129	22.057	3.355	25.413	18.702	L	541	7A-01	6.711	56.604	2.796	59.400	53.808	L
92	2A-02	0.151	22.057	3.355	25.413	18.702	L	542	7A-02	1.918	56.604	2.796	59.400	53.808	L
93	2A-03	2.445	22.057	3.355	25.413	18.702	L	543	7A-03	0.000	56.604	2.796	59.400	53.808	NA
94	2A-04	12.459	22.057	3.355	25.413	18.702	L	544	7A-04	0.084	56.604	2.796	59.400	53.808	NA
95	2A-05	0.037	22.057	3.355	25.413	18.702	NA	545	7A-05	0.239	56.604	2.796	59.400	53.808	L
96	2A-06	0.099	22.057	3.355	25.413	18.702	NA	546	7A-06	0.000	56.604	2.796	59.400	53.808	NA
97	2A-07	0.039	22.057	3.355	25.413	18.702	NA	547	7A-07	0.098	56.604	2.796	59.400	53.808	NA
98	2A-08	1.081	22.057	3.355	25.413	18.702	L	548	7A-08	0.472	56.604	2.796	59.400	53.808	L
99	2A-09	2.510	22.057	3.355	25.413	18.702	L	549	7A-09	24.109	56.604	2.796	59.400	53.808	L
100	2A-10	0.088	22.057	3.355	25.413	18.702	NA	550	7A-10	1.679	56.604	2.796	59.400	53.808	L
101	2A-11	0.000	22.057	3.355	25.413	18.702	NA	551	7A-11	10.560	56.604	2.796	59.400	53.808	L
102	2A-12	12.351	22.057	3.355	25.413	18.702	L	552	7A-12	10.481	56.604	2.796	59.400	53.808	L
103	2B-01	0.000	22.057	3.355	25.413	18.702	NA	553	7B-01	0.151	56.604	2.796	59.400	53.808	L
104	2B-02	22.913	22.057	3.355	25.413	18.702	P	554	7B-02	0.368	56.604	2.796	59.400	53.808	L
105	2B-03	0.119	22.057	3.355	25.413	18.702	L	555	7B-03	31.472	56.604	2.796	59.400	53.808	L
106	2B-04	13.959	22.057	3.355	25.413	18.702	L	556	7B-04	0.191	56.604	2.796	59.400	53.808	L
107	2B-05	2.065	22.057	3.355	25.413	18.702	L	557	7B-05	0.000	56.604	2.796	59.400	53.808	NA
108	2B-06	1.271	22.057	3.355	25.413	18.702	L	558	7B-06	0.058	56.604	2.796	59.400	53.808	NA
109	2B-08	0.958	22.057	3.355	25.413	18.702	L	559	7B-08	2.453	56.604	2.796	59.400	53.808	L
110	2B-09	3.474	22.057	3.355	25.413	18.702	L	560	7B-09	1.343	56.604	2.796	59.400	53.808	L
111	2B-10	1.290	22.057	3.355	25.413	18.702	L	561	7B-10	6.419	56.604	2.796	59.400	53.808	L
112	2B-11	0.815	22.057	3.355	25.413	18.702	L	562	7B-11	1.094	56.604	2.796	59.400	53.808	L
113	2B-12	1.183	22.057	3.355	25.413	18.702	L	563	7B-12	0.064	56.604	2.796	59.400	53.808	NA
114	2C-01	0.116	22.057	3.355	25.413	18.702	L	564	7C-01	1.523	56.604	2.796	59.400	53.808	L
115	2C-02	0.000	22.057	3.355	25.413	18.702	NA	565	7C-02	0.307	56.604	2.796	59.400	53.808	L
116	2C-04	0.000	22.057	3.355	25.413	18.702	NA	566	7C-04	0.145	56.604	2.796	59.400	53.808	L
117	2C-05	18.940	22.057	3.355	25.413	18.702	P	567	7C-05	56.250	56.604	2.796	59.400	53.808	P
118	2C-06	22.282	22.057	3.355	25.413	18.702	P	568	7C-06	6.248	56.604	2.796	59.400	53.808	L
119	2C-07	0.039	22.057	3.355	25.413	18.702	NA	569	7C-07	0.112	56.604	2.796	59.400	53.808	L
120	2C-08	4.044	22.057	3.355	25.413	18.702	L	570	7C-08	0.965	56.604	2.796	59.400	53.808	L
121	2C-09	0.000	22.057	3.355	25.413	18.702	NA	571	7C-09	1.315	56.604	2.796	59.400	53.808	L

Table AIII.1 (*continued*)

#	Location	%1e	Parent	SD	UB	LW	Bin	#	Location	%1e	Parent	SD	UB	LW	Bin
122	2C-10	25.202	22.057	3.355	25.413	18.702	P	572	7C-10	57.936	56.604	2.796	59.400	53.808	P
123	2C-11	0.074	22.057	3.355	25.413	18.702	NA	573	7C-11	0.038	56.604	2.796	59.400	53.808	NA
124	2C-12	0.089	22.057	3.355	25.413	18.702	NA	574	7C-12	0.081	56.604	2.796	59.400	53.808	NA
125	2D-01	0.039	22.057	3.355	25.413	18.702	NA	575	7D-01	53.356	56.604	2.796	59.400	53.808	L
126	2D-02	0.055	22.057	3.355	25.413	18.702	NA	576	7D-02	54.470	56.604	2.796	59.400	53.808	P
127	2D-03	0.082	22.057	3.355	25.413	18.702	NA	577	7D-03	19.914	56.604	2.796	59.400	53.808	L
128	2D-04	0.000	22.057	3.355	25.413	18.702	NA	578	7D-04	55.876	56.604	2.796	59.400	53.808	P
129	2D-05	0.039	22.057	3.355	25.413	18.702	NA	579	7D-05	0.068	56.604	2.796	59.400	53.808	NA
130	2D-07	0.048	22.057	3.355	25.413	18.702	NA	580	7D-07	0.326	56.604	2.796	59.400	53.808	L
131	2D-08	3.428	22.057	3.355	25.413	18.702	L	581	7D-08	0.167	56.604	2.796	59.400	53.808	L
132	2D-09	0.129	22.057	3.355	25.413	18.702	L	582	7D-09	0.024	56.604	2.796	59.400	53.808	NA
133	2D-10	15.302	22.057	3.355	25.413	18.702	L	583	7D-10	45.241	56.604	2.796	59.400	53.808	L
134	2D-11	0.014	22.057	3.355	25.413	18.702	NA	584	7D-11	0.059	56.604	2.796	59.400	53.808	NA
135	2D-12	0.067	22.057	3.355	25.413	18.702	NA	585	7D-12	10.979	56.604	2.796	59.400	53.808	L
136	2E-01	0.069	22.057	3.355	25.413	18.702	NA	586	7E-01	0.066	56.604	2.796	59.400	53.808	NA
137	2E-02	5.661	22.057	3.355	25.413	18.702	L	587	7E-02	0.070	56.604	2.796	59.400	53.808	NA
138	2E-03	0.036	22.057	3.355	25.413	18.702	NA	588	7E-03	0.232	56.604	2.796	59.400	53.808	L
139	2E-04	2.203	22.057	3.355	25.413	18.702	L	589	7E-04	1.283	56.604	2.796	59.400	53.808	L
140	2E-05	0.143	22.057	3.355	25.413	18.702	L	590	7E-05	0.148	56.604	2.796	59.400	53.808	L
141	2E-06	7.241	22.057	3.355	25.413	18.702	L	591	7E-06	55.489	56.604	2.796	59.400	53.808	P
142	2E-07	0.134	22.057	3.355	25.413	18.702	L	592	7E-07	1.035	56.604	2.796	59.400	53.808	L
143	2E-08	0.051	22.057	3.355	25.413	18.702	NA	593	7E-08	1.977	56.604	2.796	59.400	53.808	L
144	2E-09	0.049	22.057	3.355	25.413	18.702	NA	594	7E-09	2.716	56.604	2.796	59.400	53.808	L
145	2E-10	3.672	22.057	3.355	25.413	18.702	L	595	7E-10	0.173	56.604	2.796	59.400	53.808	L
146	2E-12	0.047	22.057	3.355	25.413	18.702	NA	596	7E-12	8.893	56.604	2.796	59.400	53.808	L
147	2F-01	13.190	22.057	3.355	25.413	18.702	L	597	7F-01	0.000	56.604	2.796	59.400	53.808	NA
148	2F-02	27.262	22.057	3.355	25.413	18.702	H	598	7F-02	2.576	56.604	2.796	59.400	53.808	L
149	2F-03	0.032	22.057	3.355	25.413	18.702	NA	599	7F-03	0.052	56.604	2.796	59.400	53.808	NA
150	2F-05	21.568	22.057	3.355	25.413	18.702	P	600	7F-05	0.118	56.604	2.796	59.400	53.808	L
151	2F-06	8.273	22.057	3.355	25.413	18.702	L	601	7F-06	0.000	56.604	2.796	59.400	53.808	NA
152	2F-07	0.021	22.057	3.355	25.413	18.702	NA	602	7F-07	52.445	56.604	2.796	59.400	53.808	L
153	2F-08	12.106	22.057	3.355	25.413	18.702	L	603	7F-08	0.525	56.604	2.796	59.400	53.808	L
154	2F-09	0.313	22.057	3.355	25.413	18.702	L	604	7F-09	0.556	56.604	2.796	59.400	53.808	L
155	2F-10	20.668	22.057	3.355	25.413	18.702	P	605	7F-10	0.371	56.604	2.796	59.400	53.808	L
156	2F-11	0.053	22.057	3.355	25.413	18.702	NA	606	7F-11	18.321	56.604	2.796	59.400	53.808	L
157	2F-12	14.592	22.057	3.355	25.413	18.702	L	607	7F-12	0.463	56.604	2.796	59.400	53.808	L
158	2G-01	0.059	22.057	3.355	25.413	18.702	NA	608	7G-01	0.000	56.604	2.796	59.400	53.808	NA
159	2G-02	0.038	22.057	3.355	25.413	18.702	NA	609	7G-02	0.823	56.604	2.796	59.400	53.808	L
160	2G-04	24.707	22.057	3.355	25.413	18.702	P	610	7G-04	0.213	56.604	2.796	59.400	53.808	L
161	2G-05	0.067	22.057	3.355	25.413	18.702	NA	611	7G-05	0.514	56.604	2.796	59.400	53.808	L
162	2G-06	20.971	22.057	3.355	25.413	18.702	P	612	7G-06	0.722	56.604	2.796	59.400	53.808	L
163	2G-07	0.115	22.057	3.355	25.413	18.702	L	613	7G-07	40.609	56.604	2.796	59.400	53.808	L
164	2G-08	13.713	22.057	3.355	25.413	18.702	L	614	7G-08	44.156	56.604	2.796	59.400	53.808	L
165	2G-09	6.739	22.057	3.355	25.413	18.702	L	615	7G-09	0.044	56.604	2.796	59.400	53.808	NA
166	2G-10	0.045	22.057	3.355	25.413	18.702	NA	616	7G-10	1.330	56.604	2.796	59.400	53.808	L
167	2G-11	0.091	22.057	3.355	25.413	18.702	NA	617	7G-11	1.217	56.604	2.796	59.400	53.808	L
168	2G-12	7.573	22.057	3.355	25.413	18.702	L	618	7G-12	0.122	56.604	2.796	59.400	53.808	L

Table AIII.1 (*continued*)

#	Location	%1e	Parent	SD	UB	LW	Bin	#	Location	%1e	Parent	SD	UB	LW	Bin
169	2H-01	1.757	22.057	3.355	25.413	18.702	L	619	7H-01	5.072	56.604	2.796	59.400	53.808	L
170	2H-02	0.318	22.057	3.355	25.413	18.702	L	620	7H-02	0.000	56.604	2.796	59.400	53.808	NA
171	2H-03	10.420	22.057	3.355	25.413	18.702	L	621	7H-03	0.059	56.604	2.796	59.400	53.808	NA
172	2H-04	17.093	22.057	3.355	25.413	18.702	L	622	7H-04	2.686	56.604	2.796	59.400	53.808	L
173	2H-05	0.180	22.057	3.355	25.413	18.702	L	623	7H-05	51.447	56.604	2.796	59.400	53.808	L
174	2H-06	0.086	22.057	3.355	25.413	18.702	NA	624	7H-06	1.014	56.604	2.796	59.400	53.808	L
175	2H-07	24.056	22.057	3.355	25.413	18.702	P	625	7H-07	55.650	56.604	2.796	59.400	53.808	P
176	2H-08	14.650	22.057	3.355	25.413	18.702	L	626	7H-08	0.106	56.604	2.796	59.400	53.808	L
177	2H-09	5.266	22.057	3.355	25.413	18.702	L	627	7H-09	0.057	56.604	2.796	59.400	53.808	NA
178	2H-10	0.061	22.057	3.355	25.413	18.702	NA	628	7H-10	0.070	56.604	2.796	59.400	53.808	NA
179	2H-11	0.201	22.057	3.355	25.413	18.702	L	629	7H-11	19.942	56.604	2.796	59.400	53.808	L
180	2H-12	2.538	22.057	3.355	25.413	18.702	L	630	7H-12	2.193	56.604	2.796	59.400	53.808	L
181	3A-01	0.041	35.221	12.147	47.368	23.073	NA	631	8A-01	27.075	50.305	13.349	63.654	36.956	L
182	3A-02	0.301	35.221	12.147	47.368	23.073	L	632	8A-02	0.140	50.305	13.349	63.654	36.956	L
183	3A-03	0.031	35.221	12.147	47.368	23.073	NA	633	8A-03	0.069	50.305	13.349	63.654	36.956	NA
184	3A-04	0.124	35.221	12.147	47.368	23.073	L	634	8A-04	0.956	50.305	13.349	63.654	36.956	L
185	3A-05	0.118	35.221	12.147	47.368	23.073	L	635	8A-05	56.893	50.305	13.349	63.654	36.956	P
186	3A-06	37.642	35.221	12.147	47.368	23.073	P	636	8A-06	57.983	50.305	13.349	63.654	36.956	P
187	3A-07	0.114	35.221	12.147	47.368	23.073	L	637	8A-07	0.137	50.305	13.349	63.654	36.956	L
188	3A-08	0.051	35.221	12.147	47.368	23.073	NA	638	8A-08	0.095	50.305	13.349	63.654	36.956	NA
189	3A-09	0.025	35.221	12.147	47.368	23.073	NA	639	8A-09	1.113	50.305	13.349	63.654	36.956	L
190	3A-10	0.000	35.221	12.147	47.368	23.073	NA	640	8A-10	0.000	50.305	13.349	63.654	36.956	NA
191	3A-11	20.195	35.221	12.147	47.368	23.073	L	641	8A-11	0.817	50.305	13.349	63.654	36.956	L
192	3A-12	0.261	35.221	12.147	47.368	23.073	L	642	8A-12	0.171	50.305	13.349	63.654	36.956	L
193	3B-01	0.174	35.221	12.147	47.368	23.073	L	643	8B-01	0.466	50.305	13.349	63.654	36.956	L
194	3B-02	0.269	35.221	12.147	47.368	23.073	L	644	8B-02	0.545	50.305	13.349	63.654	36.956	L
195	3B-03	0.036	35.221	12.147	47.368	23.073	NA	645	8B-03	3.564	50.305	13.349	63.654	36.956	L
196	3B-04	0.000	35.221	12.147	47.368	23.073	NA	646	8B-04	46.840	50.305	13.349	63.654	36.956	P
197	3B-05	3.171	35.221	12.147	47.368	23.073	L	647	8B-05	0.271	50.305	13.349	63.654	36.956	L
198	3B-06	6.967	35.221	12.147	47.368	23.073	L	648	8B-06	0.076	50.305	13.349	63.654	36.956	NA
199	3B-08	0.158	35.221	12.147	47.368	23.073	L	649	8B-08	0.044	50.305	13.349	63.654	36.956	NA
200	3B-09	0.352	35.221	12.147	47.368	23.073	L	650	8B-09	6.173	50.305	13.349	63.654	36.956	L
201	3B-10	9.524	35.221	12.147	47.368	23.073	L	651	8B-10	0.000	50.305	13.349	63.654	36.956	NA
202	3B-11	38.089	35.221	12.147	47.368	23.073	P	652	8B-11	0.291	50.305	13.349	63.654	36.956	L
203	3B-12	0.000	35.221	12.147	47.368	23.073	NA	653	8B-12	13.716	50.305	13.349	63.654	36.956	L
204	3C-01	5.165	35.221	12.147	47.368	23.073	L	654	8C-01	7.650	50.305	13.349	63.654	36.956	L
205	3C-02	11.467	35.221	12.147	47.368	23.073	L	655	8C-02	31.665	50.305	13.349	63.654	36.956	L
206	3C-04	6.041	35.221	12.147	47.368	23.073	L	656	8C-04	0.887	50.305	13.349	63.654	36.956	L
207	3C-05	0.943	35.221	12.147	47.368	23.073	L	657	8C-05	0.122	50.305	13.349	63.654	36.956	L
208	3C-06	59.238	35.221	12.147	47.368	23.073	H	658	8C-06	1.518	50.305	13.349	63.654	36.956	L
209	3C-07	0.000	35.221	12.147	47.368	23.073	NA	659	8C-07	6.445	50.305	13.349	63.654	36.956	L
210	3C-08	0.153	35.221	12.147	47.368	23.073	L	660	8C-08	45.223	50.305	13.349	63.654	36.956	P
211	3C-09	0.074	35.221	12.147	47.368	23.073	NA	661	8C-09	0.025	50.305	13.349	63.654	36.956	NA
212	3C-10	0.201	35.221	12.147	47.368	23.073	L	662	8C-10	0.126	50.305	13.349	63.654	36.956	L
213	3C-11	27.774	35.221	12.147	47.368	23.073	P	663	8C-11	4.331	50.305	13.349	63.654	36.956	L
214	3C-12	0.038	35.221	12.147	47.368	23.073	NA	664	8C-12	0.000	50.305	13.349	63.654	36.956	NA
215	3D-01	1.846	35.221	12.147	47.368	23.073	L	665	8D-01	0.168	50.305	13.349	63.654	36.956	L

Table AIII.1 (*continued*)

#	Location	%1e	Parent	SD	UB	LW	Bin	#	Location	%1e	Parent	SD	UB	LW	Bin
216	3D-02	0.000	35.221	12.147	47.368	23.073	NA	666	8D-02	3.496	50.305	13.349	63.654	36.956	L
217	3D-03	0.142	35.221	12.147	47.368	23.073	L	667	8D-03	0.390	50.305	13.349	63.654	36.956	L
218	3D-04	0.031	35.221	12.147	47.368	23.073	NA	668	8D-04	1.482	50.305	13.349	63.654	36.956	L
219	3D-05	0.169	35.221	12.147	47.368	23.073	L	669	8D-05	3.106	50.305	13.349	63.654	36.956	L
220	3D-07	1.715	35.221	12.147	47.368	23.073	L	670	8D-07	0.046	50.305	13.349	63.654	36.956	NA
221	3D-08	57.011	35.221	12.147	47.368	23.073	H	671	8D-08	57.658	50.305	13.349	63.654	36.956	P
222	3D-09	0.000	35.221	12.147	47.368	23.073	NA	672	8D-09	0.114	50.305	13.349	63.654	36.956	L
223	3D-10	0.437	35.221	12.147	47.368	23.073	L	673	8D-10	0.322	50.305	13.349	63.654	36.956	L
224	3D-11	2.051	35.221	12.147	47.368	23.073	L	674	8D-11	0.045	50.305	13.349	63.654	36.956	NA
225	3D-12	14.279	35.221	12.147	47.368	23.073	L	675	8D-12	0.489	50.305	13.349	63.654	36.956	L
226	3E-01	0.223	35.221	12.147	47.368	23.073	L	676	8E-01	8.854	50.305	13.349	63.654	36.956	L
227	3E-02	4.683	35.221	12.147	47.368	23.073	L	677	8E-02	20.670	50.305	13.349	63.654	36.956	L
228	3E-03	0.039	35.221	12.147	47.368	23.073	NA	678	8E-03	0.000	50.305	13.349	63.654	36.956	NA
229	3E-04	7.247	35.221	12.147	47.368	23.073	L	679	8E-04	58.164	50.305	13.349	63.654	36.956	P
230	3E-05	0.038	35.221	12.147	47.368	23.073	NA	680	8E-05	56.793	50.305	13.349	63.654	36.956	P
231	3E-06	11.018	35.221	12.147	47.368	23.073	L	681	8E-06	57.628	50.305	13.349	63.654	36.956	P
232	3E-07	1.181	35.221	12.147	47.368	23.073	L	682	8E-07	0.036	50.305	13.349	63.654	36.956	NA
233	3E-08	0.081	35.221	12.147	47.368	23.073	NA	683	8E-08	0.041	50.305	13.349	63.654	36.956	NA
234	3E-09	23.511	35.221	12.147	47.368	23.073	P	684	8E-09	4.170	50.305	13.349	63.654	36.956	L
235	3E-10	0.877	35.221	12.147	47.368	23.073	L	685	8E-10	0.713	50.305	13.349	63.654	36.956	L
236	3E-12	0.068	35.221	12.147	47.368	23.073	NA	686	8E-12	0.492	50.305	13.349	63.654	36.956	L
237	3F-01	0.127	35.221	12.147	47.368	23.073	L	687	8F-01	3.119	50.305	13.349	63.654	36.956	L
238	3F-02	0.255	35.221	12.147	47.368	23.073	L	688	8F-02	57.170	50.305	13.349	63.654	36.956	P
239	3F-03	0.070	35.221	12.147	47.368	23.073	NA	689	8F-03	0.027	50.305	13.349	63.654	36.956	NA
240	3F-05	0.262	35.221	12.147	47.368	23.073	L	690	8F-05	0.014	50.305	13.349	63.654	36.956	NA
241	3F-06	0.045	35.221	12.147	47.368	23.073	NA	691	8F-06	0.031	50.305	13.349	63.654	36.956	NA
242	3F-07	0.136	35.221	12.147	47.368	23.073	L	692	8F-07	57.789	50.305	13.349	63.654	36.956	P
243	3F-08	0.114	35.221	12.147	47.368	23.073	L	693	8F-08	0.036	50.305	13.349	63.654	36.956	NA
244	3F-09	0.070	35.221	12.147	47.368	23.073	NA	694	8F-09	55.827	50.305	13.349	63.654	36.956	P
245	3F-10	0.124	35.221	12.147	47.368	23.073	L	695	8F-10	27.777	50.305	13.349	63.654	36.956	L
246	3F-11	0.107	35.221	12.147	47.368	23.073	L	696	8F-11	0.027	50.305	13.349	63.654	36.956	NA
247	3F-12	0.151	35.221	12.147	47.368	23.073	L	697	8F-12	0.682	50.305	13.349	63.654	36.956	L
248	3G-01	54.384	35.221	12.147	47.368	23.073	H	698	8G-01	3.591	50.305	13.349	63.654	36.956	L
249	3G-02	0.023	35.221	12.147	47.368	23.073	NA	699	8G-02	0.082	50.305	13.349	63.654	36.956	NA
250	3G-04	41.938	35.221	12.147	47.368	23.073	P	700	8G-04	3.726	50.305	13.349	63.654	36.956	L
251	3G-05	0.000	35.221	12.147	47.368	23.073	NA	701	8G-05	0.046	50.305	13.349	63.654	36.956	NA
252	3G-06	2.489	35.221	12.147	47.368	23.073	L	702	8G-06	2.063	50.305	13.349	63.654	36.956	L
253	3G-07	8.777	35.221	12.147	47.368	23.073	L	703	8G-07	0.180	50.305	13.349	63.654	36.956	L
254	3G-08	0.773	35.221	12.147	47.368	23.073	L	704	8G-08	0.251	50.305	13.349	63.654	36.956	L
255	3G-09	0.934	35.221	12.147	47.368	23.073	L	705	8G-09	0.286	50.305	13.349	63.654	36.956	L
256	3G-10	1.621	35.221	12.147	47.368	23.073	L	706	8G-10	1.015	50.305	13.349	63.654	36.956	L
257	3G-11	2.581	35.221	12.147	47.368	23.073	L	707	8G-11	0.186	50.305	13.349	63.654	36.956	L
258	3G-12	0.068	35.221	12.147	47.368	23.073	NA	708	8G-12	0.348	50.305	13.349	63.654	36.956	L
259	3H-01	0.000	35.221	12.147	47.368	23.073	NA	709	8H-01	12.158	50.305	13.349	63.654	36.956	L
260	3H-02	16.112	35.221	12.147	47.368	23.073	L	710	8H-02	0.289	50.305	13.349	63.654	36.956	L
261	3H-03	0.000	35.221	12.147	47.368	23.073	NA	711	8H-03	0.793	50.305	13.349	63.654	36.956	L
262	3H-04	49.321	35.221	12.147	47.368	23.073	H	712	8H-04	58.127	50.305	13.349	63.654	36.956	P

Table AIII.1 (*continued*)

#	Location	%1e	Parent	SD	UB	LW	Bin	#	Location	%1e	Parent	SD	UB	LW	Bin
263	3H-05	8.035	35.221	12.147	47.368	23.073	L	713	8H-05	51.999	50.305	13.349	63.654	36.956	P
264	3H-06	0.000	35.221	12.147	47.368	23.073	NA	714	8H-06	0.411	50.305	13.349	63.654	36.956	L
265	3H-07	0.366	35.221	12.147	47.368	23.073	L	715	8H-07	0.870	50.305	13.349	63.654	36.956	L
266	3H-08	0.065	35.221	12.147	47.368	23.073	NA	716	8H-08	2.628	50.305	13.349	63.654	36.956	L
267	3H-09	15.326	35.221	12.147	47.368	23.073	L	717	8H-09	0.165	50.305	13.349	63.654	36.956	L
268	3H-10	0.062	35.221	12.147	47.368	23.073	NA	718	8H-10	0.146	50.305	13.349	63.654	36.956	L
269	3H-11	0.560	35.221	12.147	47.368	23.073	L	719	8H-11	2.753	50.305	13.349	63.654	36.956	L
270	3H-12	0.042	35.221	12.147	47.368	23.073	NA	720	8H-12	0.050	50.305	13.349	63.654	36.956	NA
271	4A-01	0.057	42.153	4.320	46.473	37.834	NA	721	9A-01	0.167	52.922	3.041	55.962	49.881	L
272	4A-02	0.144	42.153	4.320	46.473	37.834	L	722	9A-02	0.000	52.922	3.041	55.962	49.881	NA
273	4A-03	0.042	42.153	4.320	46.473	37.834	NA	723	9A-03	0.069	52.922	3.041	55.962	49.881	NA
274	4A-04	0.000	42.153	4.320	46.473	37.834	NA	724	9A-04	0.323	52.922	3.041	55.962	49.881	L
275	4A-05	42.009	42.153	4.320	46.473	37.834	P	725	9A-05	0.396	52.922	3.041	55.962	49.881	L
276	4A-06	0.064	42.153	4.320	46.473	37.834	NA	726	9A-06	41.557	52.922	3.041	55.962	49.881	L
277	4A-07	0.055	42.153	4.320	46.473	37.834	NA	727	9A-07	1.637	52.922	3.041	55.962	49.881	L
278	4A-08	1.780	42.153	4.320	46.473	37.834	L	728	9A-08	56.563	52.922	3.041	55.962	49.881	H
279	4A-09	19.241	42.153	4.320	46.473	37.834	L	729	9A-09	58.686	52.922	3.041	55.962	49.881	H
280	4A-10	2.326	42.153	4.320	46.473	37.834	L	730	9A-10	0.269	52.922	3.041	55.962	49.881	L
281	4A-11	34.898	42.153	4.320	46.473	37.834	L	731	9A-11	1.308	52.922	3.041	55.962	49.881	L
282	4A-12	16.816	42.153	4.320	46.473	37.834	L	732	9A-12	4.290	52.922	3.041	55.962	49.881	L
283	4B-01	0.000	42.153	4.320	46.473	37.834	NA	733	9B-01	0.079	52.922	3.041	55.962	49.881	NA
284	4B-02	0.032	42.153	4.320	46.473	37.834	NA	734	9B-02	0.580	52.922	3.041	55.962	49.881	L
285	4B-03	0.000	42.153	4.320	46.473	37.834	NA	735	9B-03	2.591	52.922	3.041	55.962	49.881	L
286	4B-04	0.334	42.153	4.320	46.473	37.834	L	736	9B-04	0.651	52.922	3.041	55.962	49.881	L
287	4B-05	0.110	42.153	4.320	46.473	37.834	L	737	9B-05	0.283	52.922	3.041	55.962	49.881	L
288	4B-06	0.193	42.153	4.320	46.473	37.834	L	738	9B-06	50.032	52.922	3.041	55.962	49.881	P
289	4B-08	2.178	42.153	4.320	46.473	37.834	L	739	9B-08	0.085	52.922	3.041	55.962	49.881	NA
290	4B-09	0.000	42.153	4.320	46.473	37.834	NA	740	9B-09	2.050	52.922	3.041	55.962	49.881	L
291	4B-10	36.749	42.153	4.320	46.473	37.834	L	741	9B-10	0.155	52.922	3.041	55.962	49.881	L
292	4B-11	0.111	42.153	4.320	46.473	37.834	L	742	9B-11	0.296	52.922	3.041	55.962	49.881	L
293	4B-12	0.576	42.153	4.320	46.473	37.834	L	743	9B-12	0.379	52.922	3.041	55.962	49.881	L
294	4C-01	0.727	42.153	4.320	46.473	37.834	L	744	9C-01	52.283	52.922	3.041	55.962	49.881	P
295	4C-02	0.188	42.153	4.320	46.473	37.834	L	745	9C-02	0.157	52.922	3.041	55.962	49.881	L
296	4C-04	14.735	42.153	4.320	46.473	37.834	L	746	9C-04	0.111	52.922	3.041	55.962	49.881	L
297	4C-05	0.042	42.153	4.320	46.473	37.834	NA	747	9C-05	0.666	52.922	3.041	55.962	49.881	L
298	4C-06	0.228	42.153	4.320	46.473	37.834	L	748	9C-06	0.845	52.922	3.041	55.962	49.881	L
299	4C-07	0.103	42.153	4.320	46.473	37.834	L	749	9C-07	0.362	52.922	3.041	55.962	49.881	L
300	4C-08	0.117	42.153	4.320	46.473	37.834	L	750	9C-08	0.355	52.922	3.041	55.962	49.881	L
301	4C-09	0.012	42.153	4.320	46.473	37.834	NA	751	9C-09	0.000	52.922	3.041	55.962	49.881	NA
302	4C-10	0.035	42.153	4.320	46.473	37.834	NA	752	9C-10	1.149	52.922	3.041	55.962	49.881	L
303	4C-11	0.101	42.153	4.320	46.473	37.834	L	753	9C-11	14.779	52.922	3.041	55.962	49.881	L
304	4C-12	0.000	42.153	4.320	46.473	37.834	NA	754	9C-12	6.936	52.922	3.041	55.962	49.881	L
305	4D-01	0.029	42.153	4.320	46.473	37.834	NA	755	9D-01	0.098	52.922	3.041	55.962	49.881	NA
306	4D-02	0.052	42.153	4.320	46.473	37.834	NA	756	9D-02	0.082	52.922	3.041	55.962	49.881	NA
307	4D-03	47.653	42.153	4.320	46.473	37.834	H	757	9D-03	0.497	52.922	3.041	55.962	49.881	L
308	4D-04	0.066	42.153	4.320	46.473	37.834	NA	758	9D-04	30.917	52.922	3.041	55.962	49.881	L
309	4D-05	1.006	42.153	4.320	46.473	37.834	L	759	9D-05	0.162	52.922	3.041	55.962	49.881	L

Table AIII.1 (*continued*)

#	Location	%1e	Parent	SD	UB	LW	Bin	#	Location	%1e	Parent	SD	UB	LW	Bin
310	4D-07	0.168	42.153	4.320	46.473	37.834	L	760	9D-07	56.188	52.922	3.041	55.962	49.881	H
311	4D-08	0.206	42.153	4.320	46.473	37.834	L	761	9D-08	0.389	52.922	3.041	55.962	49.881	L
312	4D-09	0.049	42.153	4.320	46.473	37.834	NA	762	9D-09	2.045	52.922	3.041	55.962	49.881	L
313	4D-10	0.057	42.153	4.320	46.473	37.834	NA	763	9D-10	0.536	52.922	3.041	55.962	49.881	L
314	4D-11	0.060	42.153	4.320	46.473	37.834	NA	764	9D-11	2.771	52.922	3.041	55.962	49.881	L
315	4D-12	0.064	42.153	4.320	46.473	37.834	NA	765	9D-12	23.552	52.922	3.041	55.962	49.881	L
316	4E-01	0.058	42.153	4.320	46.473	37.834	NA	766	9E-01	23.123	52.922	3.041	55.962	49.881	L
317	4E-02	0.080	42.153	4.320	46.473	37.834	NA	767	9E-02	0.434	52.922	3.041	55.962	49.881	L
318	4E-03	2.508	42.153	4.320	46.473	37.834	L	768	9E-03	0.000	52.922	3.041	55.962	49.881	NA
319	4E-04	0.000	42.153	4.320	46.473	37.834	NA	769	9E-04	24.890	52.922	3.041	55.962	49.881	L
320	4E-05	0.937	42.153	4.320	46.473	37.834	L	770	9E-05	0.374	52.922	3.041	55.962	49.881	L
321	4E-06	0.098	42.153	4.320	46.473	37.834	NA	771	9E-06	0.669	52.922	3.041	55.962	49.881	L
322	4E-07	0.112	42.153	4.320	46.473	37.834	L	772	9E-07	0.000	52.922	3.041	55.962	49.881	NA
323	4E-08	30.988	42.153	4.320	46.473	37.834	L	773	9E-08	0.034	52.922	3.041	55.962	49.881	NA
324	4E-09	0.136	42.153	4.320	46.473	37.834	L	774	9E-09	0.122	52.922	3.041	55.962	49.881	L
325	4E-10	0.135	42.153	4.320	46.473	37.834	L	775	9E-10	0.531	52.922	3.041	55.962	49.881	L
326	4E-12	0.147	42.153	4.320	46.473	37.834	L	776	9E-12	0.103	52.922	3.041	55.962	49.881	L
327	4F-01	0.677	42.153	4.320	46.473	37.834	L	777	9F-01	0.303	52.922	3.041	55.962	49.881	L
328	4F-02	3.614	42.153	4.320	46.473	37.834	L	778	9F-02	59.114	52.922	3.041	55.962	49.881	H
329	4F-03	0.130	42.153	4.320	46.473	37.834	L	779	9F-03	1.671	52.922	3.041	55.962	49.881	L
330	4F-05	0.262	42.153	4.320	46.473	37.834	L	780	9F-05	35.816	52.922	3.041	55.962	49.881	L
331	4F-06	3.736	42.153	4.320	46.473	37.834	L	781	9F-06	0.335	52.922	3.041	55.962	49.881	L
332	4F-07	0.000	42.153	4.320	46.473	37.834	NA	782	9F-07	4.312	52.922	3.041	55.962	49.881	L
333	4F-08	0.113	42.153	4.320	46.473	37.834	L	783	9F-08	2.355	52.922	3.041	55.962	49.881	L
334	4F-09	0.243	42.153	4.320	46.473	37.834	L	784	9F-09	0.281	52.922	3.041	55.962	49.881	L
335	4F-10	51.560	42.153	4.320	46.473	37.834	H	785	9F-10	0.606	52.922	3.041	55.962	49.881	L
336	4F-11	50.355	42.153	4.320	46.473	37.834	H	786	9F-11	0.258	52.922	3.041	55.962	49.881	L
337	4F-12	0.030	42.153	4.320	46.473	37.834	NA	787	9F-12	0.281	52.922	3.041	55.962	49.881	L
338	4G-01	0.050	42.153	4.320	46.473	37.834	NA	788	9G-01	0.309	52.922	3.041	55.962	49.881	L
339	4G-02	0.159	42.153	4.320	46.473	37.834	L	789	9G-02	0.087	52.922	3.041	55.962	49.881	NA
340	4G-04	0.328	42.153	4.320	46.473	37.834	L	790	9G-04	34.804	52.922	3.041	55.962	49.881	L
341	4G-05	18.669	42.153	4.320	46.473	37.834	L	791	9G-05	0.000	52.922	3.041	55.962	49.881	NA
342	4G-06	0.000	42.153	4.320	46.473	37.834	NA	792	9G-06	0.331	52.922	3.041	55.962	49.881	L
343	4G-07	0.125	42.153	4.320	46.473	37.834	L	793	9G-07	1.022	52.922	3.041	55.962	49.881	L
344	4G-08	0.070	42.153	4.320	46.473	37.834	NA	794	9G-08	0.352	52.922	3.041	55.962	49.881	L
345	4G-09	1.047	42.153	4.320	46.473	37.834	L	795	9G-09	53.661	52.922	3.041	55.962	49.881	P
346	4G-10	0.136	42.153	4.320	46.473	37.834	L	796	9G-10	3.681	52.922	3.041	55.962	49.881	L
347	4G-11	56.665	42.153	4.320	46.473	37.834	H	797	9G-11	0.077	52.922	3.041	55.962	49.881	NA
348	4G-12	0.000	42.153	4.320	46.473	37.834	NA	798	9G-12	6.015	52.922	3.041	55.962	49.881	L
349	4H-01	0.066	42.153	4.320	46.473	37.834	NA	799	9H-01	0.175	52.922	3.041	55.962	49.881	L
350	4H-02	0.009	42.153	4.320	46.473	37.834	NA	800	9H-02	0.158	52.922	3.041	55.962	49.881	L
351	4H-03	0.095	42.153	4.320	46.473	37.834	NA	801	9H-03	0.916	52.922	3.041	55.962	49.881	L
352	4H-04	0.365	42.153	4.320	46.473	37.834	L	802	9H-04	1.024	52.922	3.041	55.962	49.881	L
353	4H-05	0.064	42.153	4.320	46.473	37.834	NA	803	9H-05	0.355	52.922	3.041	55.962	49.881	L
354	4H-06	33.703	42.153	4.320	46.473	37.834	L	804	9H-06	0.072	52.922	3.041	55.962	49.881	NA
355	4H-07	9.489	42.153	4.320	46.473	37.834	L	805	9H-07	24.696	52.922	3.041	55.962	49.881	L
356	4H-08	0.119	42.153	4.320	46.473	37.834	L	806	9H-08	0.409	52.922	3.041	55.962	49.881	L

Table AIII.1 (*continued*)

#	Location	%1e	Parent	SD	UB	LW	Bin	#	Location	%1e	Parent	SD	UB	LW	Bin
357	4H-09	0.125	42.153	4.320	46.473	37.834	L	807	9H-09	0.310	52.922	3.041	55.962	49.881	L
358	4H-10	4.204	42.153	4.320	46.473	37.834	L	808	9H-10	0.045	52.922	3.041	55.962	49.881	NA
359	4H-11	0.139	42.153	4.320	46.473	37.834	L	809	9H-11	0.015	52.922	3.041	55.962	49.881	NA
360	4H-12	0.119	42.153	4.320	46.473	37.834	L	810	9H-12	4.182	52.922	3.041	55.962	49.881	L
361	5A-01	0.014	66.073	2.185	68.258	63.888	NA	811	10A-01	6.973	52.596	2.715	55.312	49.881	L
362	5A-02	0.000	66.073	2.185	68.258	63.888	NA	812	10A-02	0.073	52.596	2.715	55.312	49.881	NA
363	5A-03	0.121	66.073	2.185	68.258	63.888	L	813	10A-03	0.571	52.596	2.715	55.312	49.881	L
364	5A-04	0.036	66.073	2.185	68.258	63.888	NA	814	10A-04	0.505	52.596	2.715	55.312	49.881	L
365	5A-05	0.687	66.073	2.185	68.258	63.888	L	815	10A-05	0.126	52.596	2.715	55.312	49.881	L
366	5A-06	0.016	66.073	2.185	68.258	63.888	NA	816	10A-06	0.157	52.596	2.715	55.312	49.881	L
367	5A-07	0.050	66.073	2.185	68.258	63.888	NA	817	10A-07	18.416	52.596	2.715	55.312	49.881	L
368	5A-08	23.783	66.073	2.185	68.258	63.888	L	818	10A-08	0.092	52.596	2.715	55.312	49.881	NA
369	5A-09	0.294	66.073	2.185	68.258	63.888	L	819	10A-09	0.367	52.596	2.715	55.312	49.881	L
370	5A-10	69.280	66.073	2.185	68.258	63.888	H	820	10A-10	6.959	52.596	2.715	55.312	49.881	L
371	5A-11	0.637	66.073	2.185	68.258	63.888	L	821	10A-11	21.542	52.596	2.715	55.312	49.881	L
372	5A-12	0.379	66.073	2.185	68.258	63.888	L	822	10A-12	1.966	52.596	2.715	55.312	49.881	L
373	5B-01	2.513	66.073	2.185	68.258	63.888	L	823	10B-01	0.037	52.596	2.715	55.312	49.881	NA
374	5B-02	20.411	66.073	2.185	68.258	63.888	L	824	10B-02	29.590	52.596	2.715	55.312	49.881	L
375	5B-03	0.034	66.073	2.185	68.258	63.888	NA	825	10B-03	0.137	52.596	2.715	55.312	49.881	L
376	5B-04	3.839	66.073	2.185	68.258	63.888	L	826	10B-04	0.094	52.596	2.715	55.312	49.881	NA
377	5B-05	0.093	66.073	2.185	68.258	63.888	NA	827	10B-05	44.877	52.596	2.715	55.312	49.881	L
378	5B-06	0.034	66.073	2.185	68.258	63.888	NA	828	10B-06	56.603	52.596	2.715	55.312	49.881	H
379	5B-08	0.125	66.073	2.185	68.258	63.888	L	829	10B-08	0.978	52.596	2.715	55.312	49.881	L
380	5B-09	0.039	66.073	2.185	68.258	63.888	NA	830	10B-09	1.208	52.596	2.715	55.312	49.881	L
381	5B-10	69.602	66.073	2.185	68.258	63.888	H	831	10B-10	43.214	52.596	2.715	55.312	49.881	L
382	5B-11	68.812	66.073	2.185	68.258	63.888	H	832	10B-11	0.113	52.596	2.715	55.312	49.881	L
383	5B-12	3.046	66.073	2.185	68.258	63.888	L	833	10B-12	0.702	52.596	2.715	55.312	49.881	L
384	5C-01	0.042	66.073	2.185	68.258	63.888	NA	834	10C-01	0.248	52.596	2.715	55.312	49.881	L
385	5C-02	1.279	66.073	2.185	68.258	63.888	L	835	10C-02	0.067	52.596	2.715	55.312	49.881	NA
386	5C-04	0.337	66.073	2.185	68.258	63.888	L	836	10C-04	59.142	52.596	2.715	55.312	49.881	H
387	5C-05	0.044	66.073	2.185	68.258	63.888	NA	837	10C-05	4.070	52.596	2.715	55.312	49.881	L
388	5C-06	0.024	66.073	2.185	68.258	63.888	NA	838	10C-06	27.636	52.596	2.715	55.312	49.881	L
389	5C-07	0.012	66.073	2.185	68.258	63.888	NA	839	10C-07	3.866	52.596	2.715	55.312	49.881	L
390	5C-08	70.878	66.073	2.185	68.258	63.888	H	840	10C-08	40.535	52.596	2.715	55.312	49.881	L
391	5C-09	58.081	66.073	2.185	68.258	63.888	L	841	10C-09	0.115	52.596	2.715	55.312	49.881	L
392	5C-10	0.097	66.073	2.185	68.258	63.888	NA	842	10C-10	0.098	52.596	2.715	55.312	49.881	NA
393	5C-11	0.088	66.073	2.185	68.258	63.888	NA	843	10C-11	41.810	52.596	2.715	55.312	49.881	L
394	5C-12	0.056	66.073	2.185	68.258	63.888	NA	844	10C-12	0.246	52.596	2.715	55.312	49.881	L
395	5D-01	0.055	66.073	2.185	68.258	63.888	NA	845	10D-01	1.563	52.596	2.715	55.312	49.881	L
396	5D-02	0.041	66.073	2.185	68.258	63.888	NA	846	10D-02	43.112	52.596	2.715	55.312	49.881	L
397	5D-03	1.408	66.073	2.185	68.258	63.888	L	847	10D-03	45.119	52.596	2.715	55.312	49.881	L
398	5D-04	14.415	66.073	2.185	68.258	63.888	L	848	10D-04	0.778	52.596	2.715	55.312	49.881	L
399	5D-05	0.085	66.073	2.185	68.258	63.888	NA	849	10D-05	0.056	52.596	2.715	55.312	49.881	NA
400	5D-07	13.208	66.073	2.185	68.258	63.888	L	850	10D-07	0.086	52.596	2.715	55.312	49.881	NA
401	5D-08	0.959	66.073	2.185	68.258	63.888	L	851	10D-08	0.000	52.596	2.715	55.312	49.881	NA
402	5D-09	0.063	66.073	2.185	68.258	63.888	NA	852	10D-09	0.079	52.596	2.715	55.312	49.881	NA
403	5D-10	0.100	66.073	2.185	68.258	63.888	NA	853	10D-10	47.223	52.596	2.715	55.312	49.881	L

Table AIII.1 (*continued*)

#	Location	%1e	Parent	SD	UB	LW	Bin	#	Location	%1e	Parent	SD	UB	LW	Bin
404	5D-11	0.493	66.073	2.185	68.258	63.888	L	854	10D-11	41.194	52.596	2.715	55.312	49.881	L
405	5D-12	0.052	66.073	2.185	68.258	63.888	NA	855	10D-12	56.040	52.596	2.715	55.312	49.881	H
406	5E-01	0.070	66.073	2.185	68.258	63.888	NA	856	10E-01	0.044	52.596	2.715	55.312	49.881	NA
407	5E-02	0.020	66.073	2.185	68.258	63.888	NA	857	10E-02	0.072	52.596	2.715	55.312	49.881	NA
408	5E-03	0.260	66.073	2.185	68.258	63.888	L	858	10E-03	0.180	52.596	2.715	55.312	49.881	L
409	5E-04	67.919	66.073	2.185	68.258	63.888	P	859	10E-04	0.169	52.596	2.715	55.312	49.881	L
410	5E-05	0.026	66.073	2.185	68.258	63.888	NA	860	10E-05	0.100	52.596	2.715	55.312	49.881	NA
411	5E-06	0.156	66.073	2.185	68.258	63.888	L	861	10E-06	0.765	52.596	2.715	55.312	49.881	L
412	5E-07	61.130	66.073	2.185	68.258	63.888	L	862	10E-07	8.090	52.596	2.715	55.312	49.881	L
413	5E-08	32.792	66.073	2.185	68.258	63.888	L	863	10E-08	0.112	52.596	2.715	55.312	49.881	L
414	5E-09	0.057	66.073	2.185	68.258	63.888	NA	864	10E-09	43.473	52.596	2.715	55.312	49.881	L
415	5E-10	0.041	66.073	2.185	68.258	63.888	NA	865	10E-10	0.067	52.596	2.715	55.312	49.881	NA
416	5E-12	0.049	66.073	2.185	68.258	63.888	NA	866	10E-12	0.105	52.596	2.715	55.312	49.881	L
417	5F-01	0.024	66.073	2.185	68.258	63.888	NA	867	10F-01	0.000	52.596	2.715	55.312	49.881	NA
418	5F-02	3.265	66.073	2.185	68.258	63.888	L	868	10F-02	50.309	52.596	2.715	55.312	49.881	P
419	5F-03	2.594	66.073	2.185	68.258	63.888	L	869	10F-03	0.217	52.596	2.715	55.312	49.881	L
420	5F-05	1.907	66.073	2.185	68.258	63.888	L	870	10F-05	0.630	52.596	2.715	55.312	49.881	L
421	5F-06	0.055	66.073	2.185	68.258	63.888	NA	871	10F-06	58.463	52.596	2.715	55.312	49.881	H
422	5F-07	0.021	66.073	2.185	68.258	63.888	NA	872	10F-07	0.133	52.596	2.715	55.312	49.881	L
423	5F-08	23.071	66.073	2.185	68.258	63.888	L	873	10F-08	1.759	52.596	2.715	55.312	49.881	L
424	5F-09	4.016	66.073	2.185	68.258	63.888	L	874	10F-09	0.116	52.596	2.715	55.312	49.881	L
425	5F-10	0.643	66.073	2.185	68.258	63.888	L	875	10F-10	0.043	52.596	2.715	55.312	49.881	NA
426	5F-11	58.560	66.073	2.185	68.258	63.888	L	876	10F-11	2.212	52.596	2.715	55.312	49.881	L
427	5F-12	0.079	66.073	2.185	68.258	63.888	NA	877	10F-12	16.977	52.596	2.715	55.312	49.881	L
428	5G-01	0.068	66.073	2.185	68.258	63.888	NA	878	10G-01	0.246	52.596	2.715	55.312	49.881	L
429	5G-02	58.172	66.073	2.185	68.258	63.888	L	879	10G-02	0.075	52.596	2.715	55.312	49.881	NA
430	5G-04	0.027	66.073	2.185	68.258	63.888	NA	880	10G-04	0.421	52.596	2.715	55.312	49.881	L
431	5G-05	0.609	66.073	2.185	68.258	63.888	L	881	10G-05	31.244	52.596	2.715	55.312	49.881	L
432	5G-06	73.912	66.073	2.185	68.258	63.888	H	882	10G-06	0.067	52.596	2.715	55.312	49.881	NA
433	5G-07	0.069	66.073	2.185	68.258	63.888	NA	883	10G-07	0.138	52.596	2.715	55.312	49.881	L
434	5G-08	15.169	66.073	2.185	68.258	63.888	L	884	10G-08	0.405	52.596	2.715	55.312	49.881	L
435	5G-09	0.028	66.073	2.185	68.258	63.888	NA	885	10G-09	0.068	52.596	2.715	55.312	49.881	NA
436	5G-10	0.041	66.073	2.185	68.258	63.888	NA	886	10G-10	0.058	52.596	2.715	55.312	49.881	NA
437	5G-11	67.622	66.073	2.185	68.258	63.888	P	887	10G-11	0.122	52.596	2.715	55.312	49.881	L
438	5G-12	0.038	66.073	2.185	68.258	63.888	NA	888	10G-12	8.187	52.596	2.715	55.312	49.881	L
439	5H-01	0.044	66.073	2.185	68.258	63.888	NA	889	10H-01	0.052	52.596	2.715	55.312	49.881	NA
440	5H-02	0.034	66.073	2.185	68.258	63.888	NA	890	10H-02	5.090	52.596	2.715	55.312	49.881	L
441	5H-03	0.123	66.073	2.185	68.258	63.888	L	891	10H-03	0.108	52.596	2.715	55.312	49.881	L
442	5H-04	0.109	66.073	2.185	68.258	63.888	L	892	10H-04	0.000	52.596	2.715	55.312	49.881	NA
443	5H-05	6.465	66.073	2.185	68.258	63.888	L	893	10H-05	0.089	52.596	2.715	55.312	49.881	NA
444	5H-06	64.439	66.073	2.185	68.258	63.888	P	894	10H-06	0.416	52.596	2.715	55.312	49.881	L
445	5H-07	0.166	66.073	2.185	68.258	63.888	L	895	10H-07	47.623	52.596	2.715	55.312	49.881	L
446	5H-08	0.084	66.073	2.185	68.258	63.888	NA	896	10H-08	21.145	52.596	2.715	55.312	49.881	L
447	5H-09	2.007	66.073	2.185	68.258	63.888	L	897	10H-09	0.662	52.596	2.715	55.312	49.881	L
448	5H-10	0.046	66.073	2.185	68.258	63.888	NA	898	10H-10	0.366	52.596	2.715	55.312	49.881	L
449	5H-11	0.099	66.073	2.185	68.258	63.888	NA	899	10H-11	0.070	52.596	2.715	55.312	49.881	NA
450	5H-12	0.123	66.073	2.185	68.258	63.888	L	900	10H-12	0.000	52.596	2.715	55.312	49.881	NA

Table AIII.2: Representative re-array scheme for the *high* bin of library 2-L

Table used to program colony picker to re-array cultures between 96-well plates. The variants are re-arrayed in descending order and placed in a new 96-well plate from well A1 to A12, then from B1 to B12, etc. Note that since the total number of variants in this bin was 29, each variant was placed in triplicate to increase the volume of final culture.

#	Source	Row	Column
1	source_plate_1	C	5
2	source_plate_1	C	5
3	source_plate_1	C	5
4	source_plate_1	C	8
5	source_plate_1	C	8
6	source_plate_1	C	8
7	source_plate_1	E	6
8	source_plate_1	E	6
9	source_plate_1	E	6
10	source_plate_1	F	8
11	source_plate_1	F	8
12	source_plate_1	F	8
13	source_plate_2	F	2
14	source_plate_2	F	2
15	source_plate_2	F	2
16	source_plate_3	C	6
17	source_plate_3	C	6
18	source_plate_3	C	6
19	source_plate_3	D	8
20	source_plate_3	D	8
21	source_plate_3	D	8
22	source_plate_3	G	1
23	source_plate_3	G	1
24	source_plate_3	G	1
25	source_plate_3	H	4
26	source_plate_3	H	4
27	source_plate_3	H	4
28	source_plate_4	D	3
29	source_plate_4	D	3

#	Source	Row	Column
30	source_plate_4	D	3
31	source_plate_4	F	10
32	source_plate_4	F	10
33	source_plate_4	F	10
34	source_plate_4	F	11
35	source_plate_4	F	11
36	source_plate_4	F	11
37	source_plate_4	G	11
38	source_plate_4	G	11
39	source_plate_4	G	11
40	source_plate_5	A	10
41	source_plate_5	A	10
42	source_plate_5	A	10
43	source_plate_5	B	10
44	source_plate_5	B	10
45	source_plate_5	B	10
46	source_plate_5	B	11
47	source_plate_5	B	11
48	source_plate_5	B	11
49	source_plate_5	C	8
50	source_plate_5	C	8
51	source_plate_5	C	8
52	source_plate_5	G	6
53	source_plate_5	G	6
54	source_plate_5	G	6
55	source_plate_6	C	6
56	source_plate_6	C	6
57	source_plate_6	C	6
58	source_plate_6	D	4

#	Source	Row	Column
59	source_plate_6	D	4
60	source_plate_6	D	4
61	source_plate_6	F	12
62	source_plate_6	F	12
63	source_plate_6	F	12
64	source_plate_9	A	8
65	source_plate_9	A	8
66	source_plate_9	A	8
67	source_plate_9	A	9
68	source_plate_9	A	9
69	source_plate_9	A	9
70	source_plate_9	D	7
71	source_plate_9	D	7
72	source_plate_9	D	7
73	source_plate_9	F	2
74	source_plate_9	F	2
75	source_plate_9	F	2
76	source_plate_10	B	6
77	source_plate_10	B	6
78	source_plate_10	B	6
79	source_plate_10	C	4
80	source_plate_10	C	4
81	source_plate_10	C	4
82	source_plate_10	D	12
83	source_plate_10	D	12
84	source_plate_10	D	12
85	source_plate_10	F	6
86	source_plate_10	F	6
87	source_plate_10	F	6