Characterization and Engineering of a Promiscuous L-Threonine Transaldolase to Access Novel Amino Acid Building Blocks

By

Prasanth Kumar

A dissertation submitted in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

(Biophysics)

at the

University of Wisconsin-Madison

2022

Date of final oral examination: 9th February 2022

The dissertation is approved by the following members of the Final Oral Committee:

Andrew R. Buller, Professor, Chemistry

Amy M. Weeks, Professor, Biochemistry

Philip A. Romero, Professor, Biochemistry

Srivatsan Raman, Professor, Biochemistry

Zachary K. Wickens, Professor, Chemistry

Acknowledgments

"*It takes a village to raise a child*" – African proverb.

After almost six years into my Ph.D., I think I am a fortunate child raised by the love and care of several wonderful people. In every village, there is a wise leader who guides the young on the path to success. That figure was my Ph.D. advisor, Prof. Andrew Buller. I am forever grateful to Andrew for the kindness, patience, guidance, and support he showed me throughout my Ph.D. Andrew's relentless curiosity, thoughtful methodological approach to science, and emphasis on effective scientific communication were instrumental in my growth as a Scientist. Andrew is a leader who leads by example and has inspired me throughout my Ph.D. to learn and grow. Thanks, Andrew, for this opportunity. It's been an honor to work with you.

I thank my committee members: Prof. Amy Weeks, Prof. Dave Pagliarini, Prof. Phil Romero, Prof. Sam Gellman, Prof. Vatsan Raman, Prof. Zach Wickens, for helpful scientific discussions and feedback on my research progress.

I joined the Buller lab in the fall of 2017 with Allwin McDonald, Jon Ellis, and Lydia Perkins as the first members of the Buller lab. Alongside Andrew, they established the Buller lab as a fun place to work and an exciting place to do research, which led to more brilliant people joining the lab. Thanks, OGs, for being with me throughout my Ph.D. and inspiring me with your creativity and scientific curiosity. Working in the Buller lab was always joyful. I am thankful to all the present and past Buller Buddies for being kind, caring, and funny. I have enjoyed learning about your unique personal and scientific stories. I will cherish all the happy memories of our time together, from fun outings to talking science. I will miss you all.

I was fortunate to collaborate with several talented scientists throughout my Ph.D., and I can't imagine finishing my Ph.D. without their support. I thank my collaborators, Dr. Craig Bingman, Grace Carlson, Jon Ellis, Prof. Patrick Willoughby, Sam Bruffy, Tony Meza, and Dr. Tyler Doyon. They have collected valuable data appearing in this thesis. I also thank Aadhishree Kasat, Allwin McDonald, and Meghan Campbell, with whom I collaborated on projects not appearing in this thesis. I have enjoyed working with these wonderful folks who made me realize the power of good teamwork. Go, Team!

I thank the Morgridge Institute for Research for their generous graduate fellowship, WARF and NIH for supporting the lab's research. I also thank the Department of Chemistry, Biophysics

graduate program, and the University of Wisconsin-Madison for supporting me throughout my studies.

Ph.D. is challenging on its own. Unfortunately, I developed a few physical and mental health problems which made my life more challenging. I am thankful to my surgeon, Dr. Seth Williams, who fixed my slipped disc and gave back my pain-free life. I also thank my mental-health therapist, Dr. Dick Goldberg, Dr. Jeff Hird, and Dr. Ashley Cook, who helped me learn about myself and for teaching me valuable skills to fight the challenges that life threw at me.

During my stay at Madison, I have met numerous kind people, and a few of them turned into good friends. I thank Allwin, Alyssa, Cassie, Kaivalya, Josie, Kushal, Keeley, Kim, Munish, Mainak, Supreet, and Thejas, for being a part of my journey. I have enjoyed spending time with them and learning about their stories. They brought me joy and kept me sane through my ups and downs. I also want to thank mentor turned friends Dr. Arti Tripathi and Dr. Bob Trachman for supporting me during the critical transition periods in my Ph.D. journey.

Finally, I am grateful to my parents for their unconditional love. I would not be here if not for their hard work in ensuring that I always get the best education. Thank you, Amma and Naina.

Table of contents

Chapter 1

Introduction

Chapter 1: Introduction

1. 1. Nature creates complexity from simple building blocks

The complex biochemistry of every living cell on planet Earth arises from the interplay of simple building blocks. Deoxyribonucleic acid (DNA), which contains the essential genetic information needed to create a cell, is composed of just four nucleotide bases (Figure 1).[1] Information stored in the DNA is transferred to messenger ribonucleic acid (mRNA) which then codes for proteins, the molecular machines that control nearly every biological process. Similar to DNA, mRNA is also composed of just four nucleotide bases. Proteins perform diverse roles in biological systems, despite only being made up of 20 amino acid building blocks, many of which share similar chemical functionalities. This pattern of building complexity through simplicity is a key to evolution of new functions and a central theme in the cellular world.

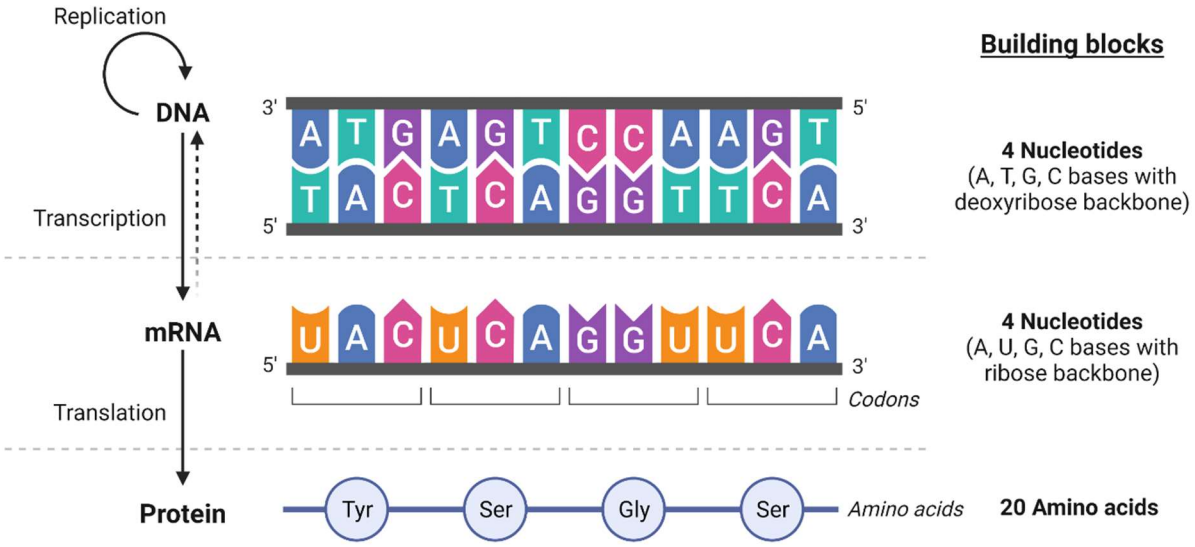


Figure 1. The central dogma of biology. DNA contains the genetic blueprint of cells and makes copies of itself through replication. Transcription transfers the information contained in the DNA to messenger ribonucleic acid (mRNA). mRNA codes for proteins through translation in which three RNA bases code for one of the 20 amino acids (or a stop codon). DNA and mRNA comprise only four nucleotide bases, and proteins contain 20 amino acids. *Figure created with BioRender.com* from a pre-existing template.

Over millions of years of evolution, certain organisms have evolved specialized cellular machinery to build complex molecules from simple cellular building blocks. These complex molecules, also known as natural products, increase the fitness of the organism in certain situations. For example, the common mold *Penicillium rubens* synthesizes penicillin, which inhibits the enzymes needed to cross-link the peptidoglycan layer, a bacterial cell wall component.[2] Without a protective peptidoglycan layer, the bacterial cell wall weakens, leading to cell lysis. Thus, penicillin acts as a potent antibacterial, giving *Penicillium rubens* a fitness advantage over other fungal species. Certain bacterial species also produce antibiotics as a means of chemical warfare against other competing bacterial species.[3]

Several of these antibiotics contain amino acids as their building blocks. Penicillin, for example, is an enzymatically modified tripeptide composed of three amino acids: L-cysteine (Cys), L-valine (Val), and L-α-aminoadipic acid (Figure 2).[4] Cys and Val are two of the 20 standard amino acids found in proteins. In addition, selenocysteine and pyrrolysine are also found in proteins. Thus, there are 22 proteogenic amino acids. L-α-aminoadipic acid is a non-proteogenic amino acid also known as non-standard amino acid (nsAA). There are hundreds of naturally occurring nsAAs formed as secondary metabolites in bacteria, fungi, plants, and marine organisms. These nsAAs possess a wide range of properties and expand the number of building blocks for natural product biosynthesis.[5]

All amino acids can also be selectively modified to bring about unique functions before being used as a building block. For example, L-tryptophan (Trp), one of the 20 proteogenic amino acids undergoes site-specific enzymatic hydroxylation to form 5'-OH-Trp. 5'-OH-Trp is then decarboxylated to form the neurotransmitter serotonin, which affects emotions and motor skills.[6] Penicillin and serotonin are two examples of how Nature creates complex biological functions from simple amino acid building blocks.

Nature's pattern of building complexity is not limited to the chemistry of amino acids. Nature uses other building blocks, such as acetyl-CoA and carbohydrates (Figure 2). Acetyl-CoA gives rise to various terpenes and polyketides natural products.[7,8] Menthol, for example, is a monoterpenoid found in the oils of corn mint and peppermint and used as a local anesthetic and for relieving minor throat irritation.[9] Lovastatin is a polyketide-derived fungal metabolite clinically used to reduce unhealthy low-density lipoprotein (LDL) cholesterol.[10] Kanamycin is a sugar-derived antibiotic used to treat severe bacterial infections and tuberculosis.[11] These natural products with diverse biological properties further demonstrates how nature creates complexity through simple building which is a recurring theme throughout the cellular world.

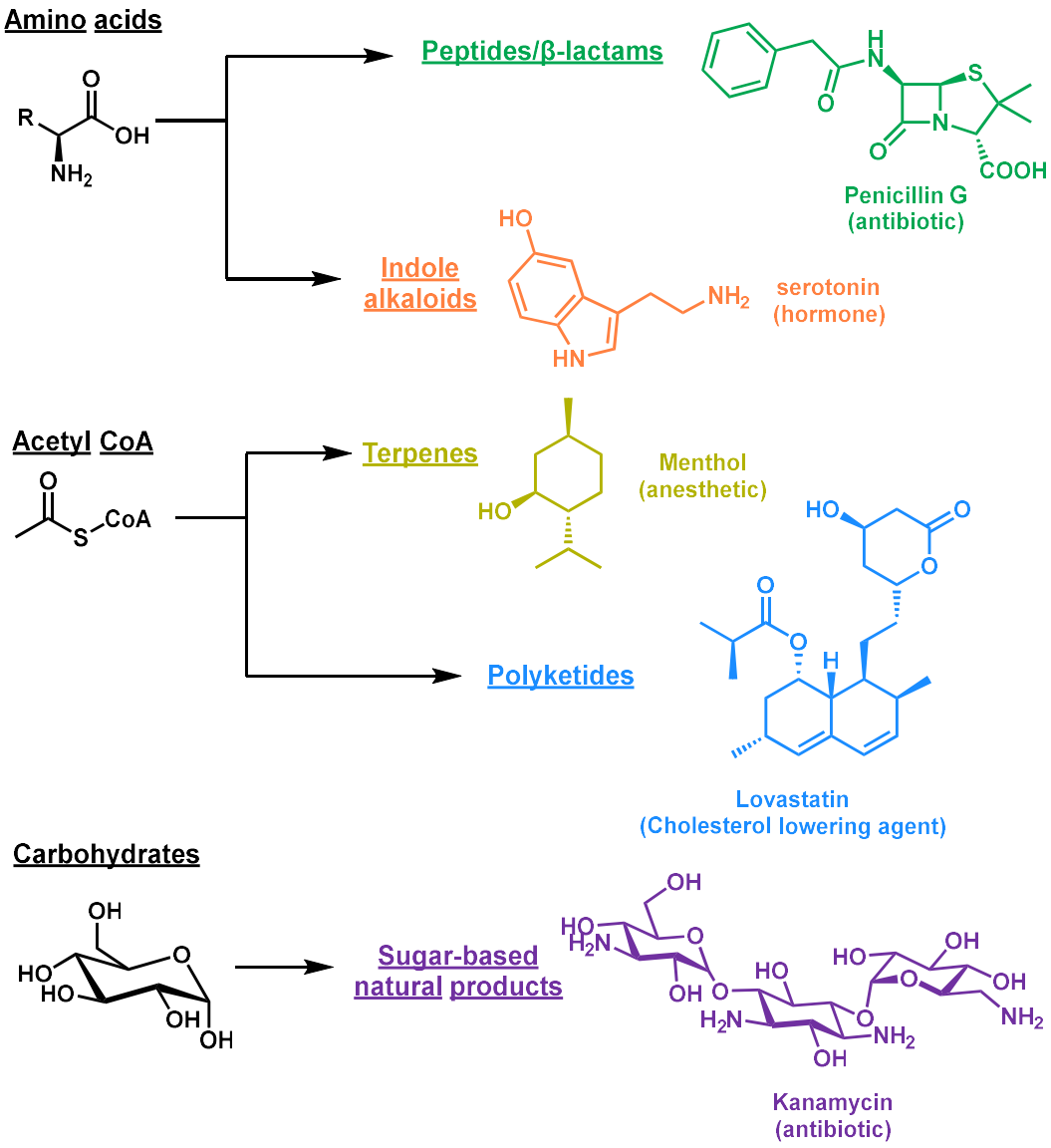


Figure 2. Common building blocks used by Nature and their corresponding natural products.

1. 2. Amino acids are part of several biologically active molecules

All amino acids contain orthogonal amino and carboxyl functional groups, which can be derivatized by established organic methodologies such as acylation, alkylation, and amidation.[12–14] The presence of diverse side chains and functional groups in a chiral configuration makes nsAAs amenable to further derivatization. The resulting three-dimensional structures contain numerous biofunctional handles, making nsAAs highly attractive as building blocks for medicinal chemistry.[5] nsAAs have applications beyond the development of peptide and small molecule therapeutics. For example, nsAAs have found several uses in genetic code expansion, in spectroscopy as probes, as post-translational modification mimics, and in site-selective labeling and alteration of proteins.[15] These wide applications have inspired biologists and organic chemists to understand how Nature makes these nsAAs and develop novel strategies to synthesize them in the lab. These efforts have led to the development of several pharmaceutical drugs that have saved millions of lives.[16]

One in three pharmaceutical drugs approved by the United States Food and Drug Administration (FDA) in 2020 contains amino acids or their derivatives.[17] Some of these drugs are the free amino acid such as Droxidopa, which is used to treat Parkinson's disease, a neuromuscular disorder. Once dosed, the amino acid drug serves as a biosynthetic precursor to the neurotransmitter norepinephrine (Figure 3).[18] Other free amino acids include the hormone thyroxin, available commercially as Synthroid, which is used to treat hypothyroidism (a condition in which the thyroid gland does not produce sufficient thyroid hormone).[19] Melphalan is a cytotoxic nsAA used clinically as a chemotherapy agent for solid tumors.[20] D-Penicillamine is used as a chelating agent to eliminate toxic metal ions and treat severe rheumatoid arthritis.[21] Radiolabeled amino acids such as 6'-18F-L-Dopa are used for PET studies to determine the spatial distribution of dopamine in the brain.[22]

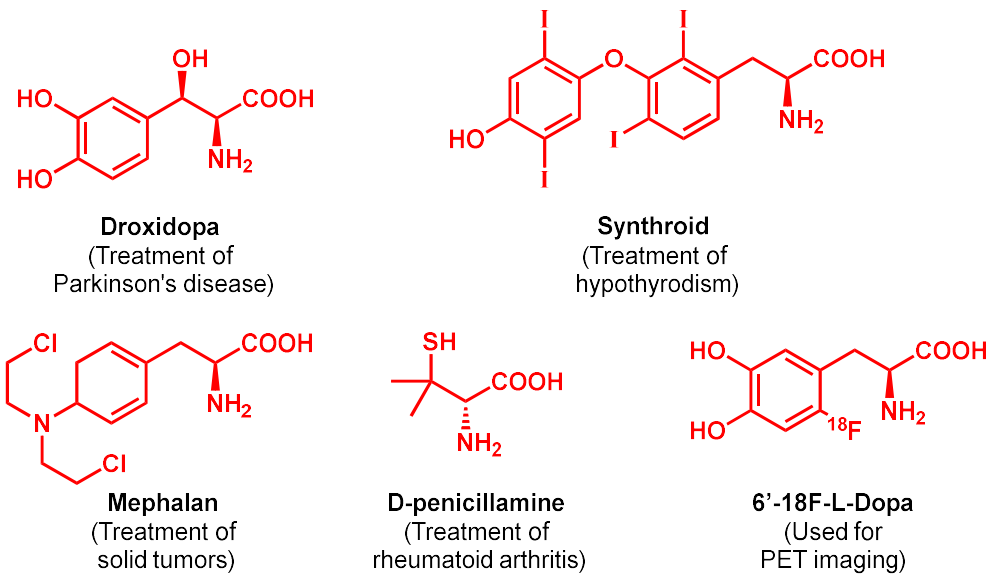


Figure 3. Non-standard amino acids (red) used as drugs and in biomedical applications.

Many drugs require modification of amino acids either at the N- or C-terminus or both to bring about therapeutic functions. For example, Lacosamide, an anticonvulsant (prevents epileptic fits or other convulsions), contains an *O*-methyl serine residue as the central scaffold and is modified at both the N- and C-terminus (Figure 4).[23] 15-Deoxyspergualin contains another analog of L-serine, an α-hydroxyl-glycine .This drug has immunosuppressant properties and is used to prevent renal graft rejection.[21] Bestatin is an investigational anticancer drug that has an alpha-hydroxylated β-amino phenylalanine core.[22] Sitagliptin, sold as Januvia, is an anti-diabetic drug developed by Merck & Co which benefits around 1.6 million patients every year.[23] Sitagliptin also has a β-amino phenylalanine core with additional elements at its C- terminus to bring about its biological activity. These drugs demonstrate the diversity of amino acids scaffolds found in live-saving therapeutics.

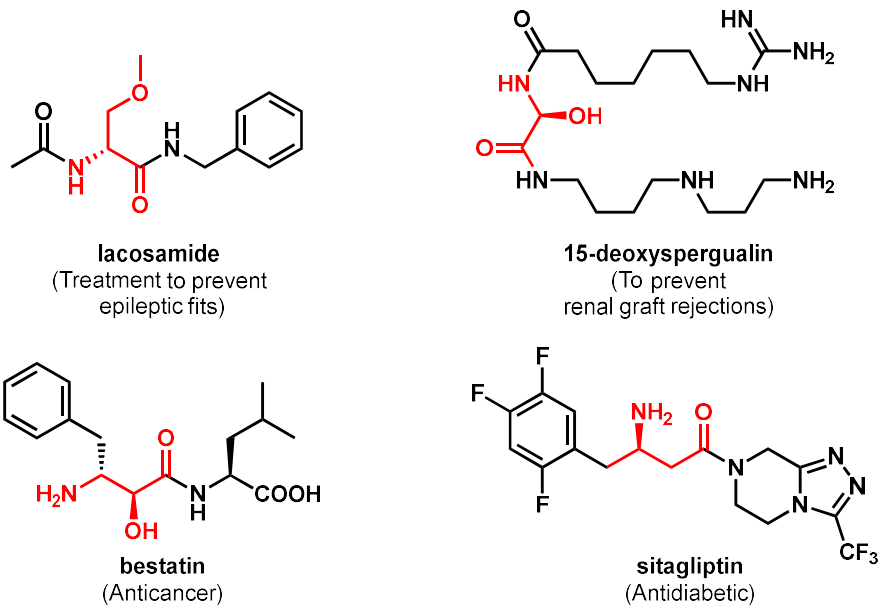


Figure 4. Non-standard amino acids (red) as components of therapeutic drugs. Amino acid backbone is colored in red.

Peptides represent another class of amino acid therapeutics. Peptides are short polymers of amino acids and are part of several hormones such as insulin, somatostatin and oxytocin. They are also part of several complex natural products, such as the peptide antibiotic actinomycin and glycopeptide antibiotic vancomycin. Cyclic peptides have gained interest due to their additional conformational constraints, potentially leading to tighter binding to their targets while simultaneously hindering proteolysis. The nonribosomal peptide synthetases (NRPS) are the biological machinery that incorporate amino acid monomers into nonribosomally synthesized peptides. The NRPS derived cyclic peptide daptomycin contains four nsAAs: L-threo-β-methyl-glutamate, kynurenine, D-alanine, and D-asparagine (Figure 5).[27] Daptomycin is used as an antibiotic against methicillin-resistant Staphylococcus aureus (MRSA) and vancomycin-resistant Enterococci (VRE).[28]

Pasireotide is a cyclic hexapeptide developed by Novartis to treat Cushing's disease (a condition in which excess cortisol is secreted, leading to weight gain and type-2 diabetes).[29]

Pasierotide is an analog of the hormone somatostatin and contains three nsAAs: *O*-benzyl Tyr, acylated hydroxyproline, and phenylglycine derivatives. Several peptide antibiotics contain β-hydroxy amino acids: amino acids with a hydroxy group at the b-carbon of the amino acid backbone. For example, the peptide antibiotic lysobactin contains three β-hydroxy amino acids: β-hydroxy-leucine, β-hydroxy-asparagine, and β-hydroxy-phenylalanine (Figure 5).[30] β-hydroxy-histidine is part of the antibiotics bleomycin and cleomcyin.[31,32] In many cases, the success of peptide drugs hinges on the nsAAs incorporated into them to improve their potency and pharmacokinetics.[5]

Amino acids have several subclasses: L-amino acids (most abundant in cells), D-amino acids (often incorporated into peptides to reduce proteolysis), deuterated amino acids (one or multiple hydrogen atoms replaced by a deuterium atom to improve the stability of drugs), α-, β-, γ-, and δ- amino acids, β- and γ- hydroxy amino acids. Irrespective of the sub-class, several life-saving drugs contain one of these amino acids. However, the emergence of new diseases and antibiotic resistance has posed a challenge, and there is an urgent need to develop novel therapeutics and improve existing drugs.[33,34] Enabling access to nsAAs is an integral part of overcoming this challenge and has been a focus of many research groups worldwide.

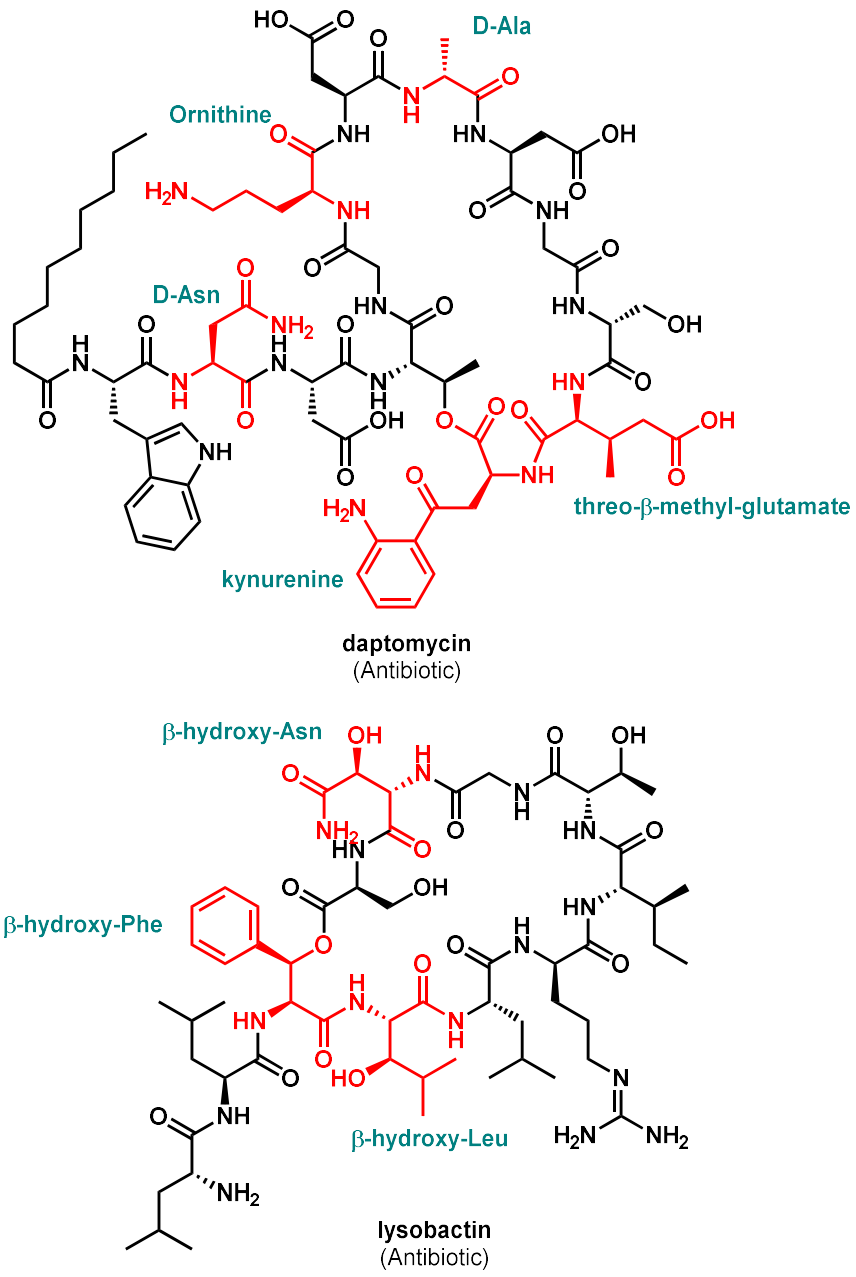


Figure 5. Non-standard amino acids are part of several peptide therapeutics. nsAAs are colored in red and labeled in green.

1. 3. β-hydroxy amino acids are valuable building blocks

Among the different classes of amino acids, the aforementioned β-hydroxy amino acids have attracted significant attention as they are valuable constituents of medicinally important compounds and complex natural products. They include naturally occurring amino acids such as serine, threonine, and 3-hydroxyproline (Figure 6). Serine and threonine are proteogenic amino acids and a key cellular metabolite. 3-hydroxyproline is one of the monomers in type III collagen biosynthesis and is essential for collagen's triple-helix formation.[35] Droxidopa which is used as a therapeutic for the treatment of Parkinon's disease is also a β-hydroxy amino acid.[18] As discussed in previous section, β-hydroxy amino acids are a central component of NRPS derived peptide antibiotic such as lysobactin and vancomycin (Figure 5, 6).[30,36] β-hydroxy amino acids are also a core constituent of sphingofungin A (antifungal),[37] salinosporamide (anticancer),[38] and cyclosporin (immunosuppressant)[39] (Figure 6).

β-hydroxy amino acids have also been used as intermediates in the biosynthesis of β-lactones such as obafluorin.[40,41] β-hydroxy-*p*-nitrophenylserine and its derivatives are the key components of amino alcohol antibiotics such as chloramphenicol and its synthetic derivatives thiamphenicol and fluorfenicol (Figure 6). β-hydroxy-tyrosine and its derivatives are important precursors for synthesizing vancomycin analogues.[42] β-hydroxy amino acids are also common synthetic precursors to β-lactams[43] and aziridine carboxylic acid derivatives[44] and as precursors to chiral bisoxazoline ligands for metal catalysis.[45] These diverse applications underlie a broad interest in developing synthetic and enzymatic methodologies to access β-hydroxy amino acids.

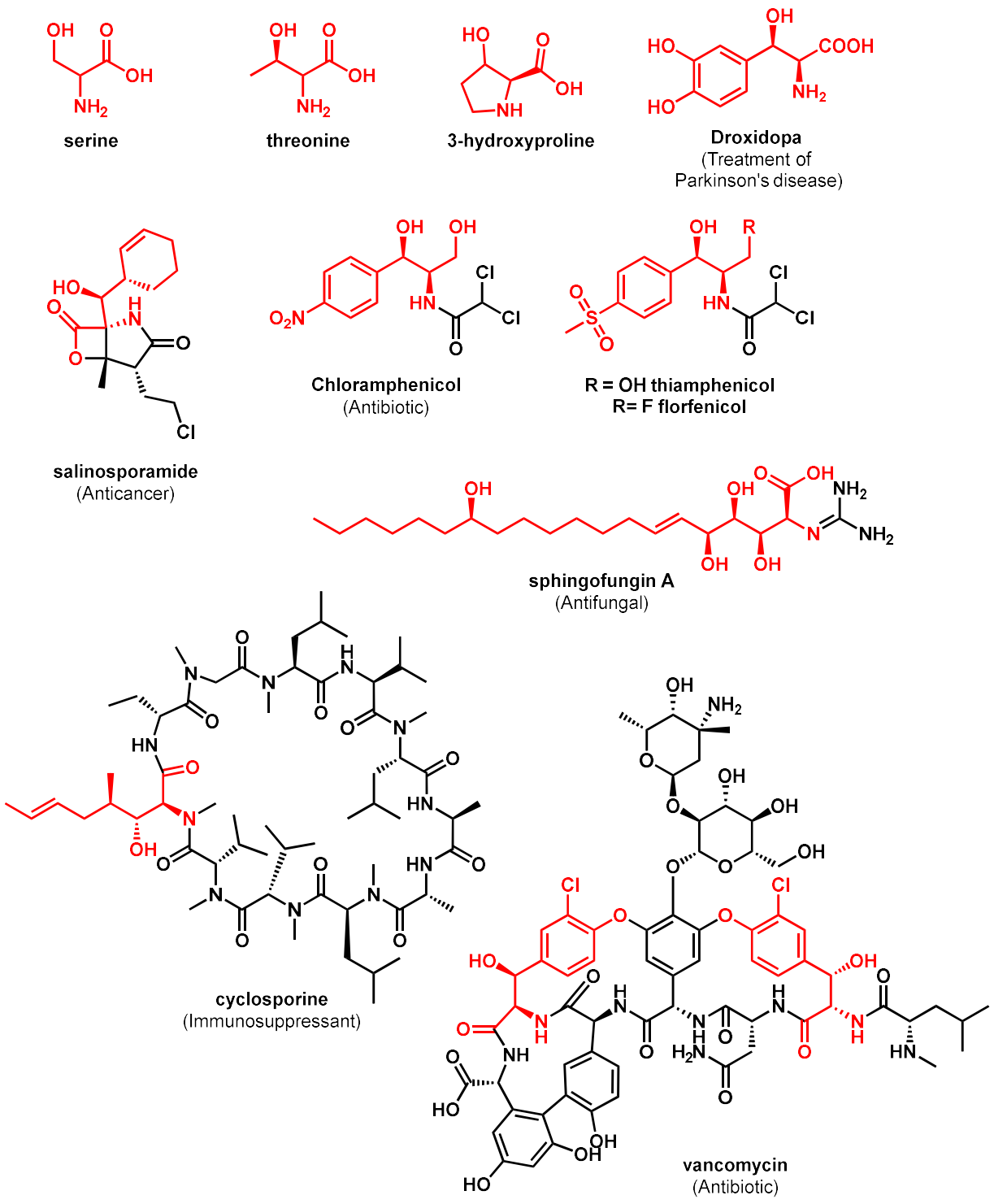


Figure 6. β-hydroxy amino acids (red) as components of natural products and in pharmaceutical drugs.

1. 4. Routes to access β-hydroxy amino acids

β-hydroxy amino acids contain two chiral centers at the Cα and Cβ positions resulting in four possible diastereomers (L-threo, D-threo, L-erythro, D-erythro), making them challenging to synthesize chemically. Several methods have been reported for asymmetric synthesis of β-hydroxy amino acids. The main challenge in each of these methods is the control of the relative and absolute stereochemistry of the asymmetric carbons. The direct asymmetric aldol reaction between glycinate Schiff bases and aldehydes is an elegant method to access β-hydroxy-α-amino esters. [46–48] Sharpless asymmetric dihydroxylation is also an effective method although it requires multi-step synthesis.[49] Recently, a stereoselective electrosynthesis using a trans-metal Schiff base complex was reported.[50] These methodologies have expanded the number of β-hydroxy amino acids that can be accessed. Nevertheless, these methods also have limitations: from the use of toxic substrates and solvents, requiring anaerobic conditions, and low carbon economy. With the current trends of improving sustainable production of small molecules, biocatalytic routes to these valuable building blocks are highly desired.

Mirroring the myriad bioactive molecules that harbor β-hydroxy amino acids, Nature has evolved multiple biosynthetic routes to produce these useful compounds. Free amino acids can undergo stereospecific Cβ hydroxylation, as illustrated by the non-heme Fe-dependent hydroxylation of homotyrosine in the biosynthesis of echinocandins (Figure 7).[51] The more common biosynthetic strategy, however, is for an amino acid to be loaded onto a non-ribosomal peptide synthetase (NRPS), whereupon a P450 enzyme catalyzes Cβ hydroxylation of an aminoacyl-*S*-enzyme intermediate en route to a peptide natural product. NRPS mediated site-specific hydroxylation occurs in the biosynthetic pathway of several peptide natural products such as vancomycin and salinosporamide.[52–55] Both of these biosynthetic strategies adds the hydroxyl group after assembly of the carbon backbone of a precursor amino acid, limiting their synthetic utility.

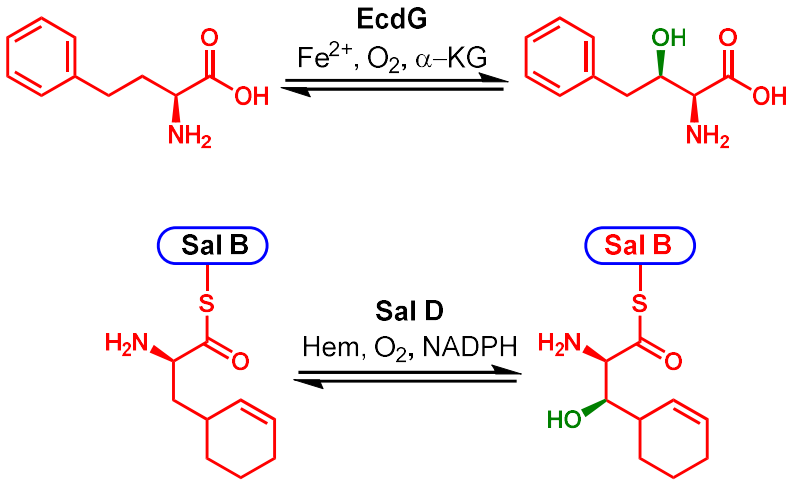


Figure 7. Site-specific hydroxylation of amino acids to access β-hydroxy amino acids in Nature. The newly formed hydroxyl group is colored in green. EcdG is a non-heme mononuclear iron oxygenase that catalyzes hydroxylation of homotyrosine in the biosynthesis of echinocandins. Sal D is the P450 monooxygenase that catalyze site-specific hydroxylation in salinosporamide biosynthesis and Sal B is the NRPS loading unit.

Researchers have therefore been drawn to the pyridoxal phosphate (PLP)-dependent threonine aldolases (TAs) and serine hydroxymethyl transferases (SHMTs) for the synthesis of β-hydroxy amino acids.[56] These enzymes natively catalyze the retro-aldol cleavage of L-threonine (Thr) and L-serine (Ser), respectively, generating glycine for use in primary metabolism (Figure 8).[57] By running these transformations in reverse with very high concentrations of glycine, TAs and SHMTs have been leveraged for the synthesis of β-hydroxy amino acids. However, most native TAs catalyze aldol reactions with low stereoselectivity at Cβ, leading to poor diastereoselectivity profiles for non-native transformations.[58–61] These challenges can be overcome in exceptional cases, either through intensive directed evolution or through the use of diastereoselective crystallization, but these routes are limited to one or a handful of substrates

before selectivity is compromised.[59,62] Consequently, a scalable and generalizable biocatalytic route for the selective synthesis of β-hydroxy amino acids has yet to be realized.

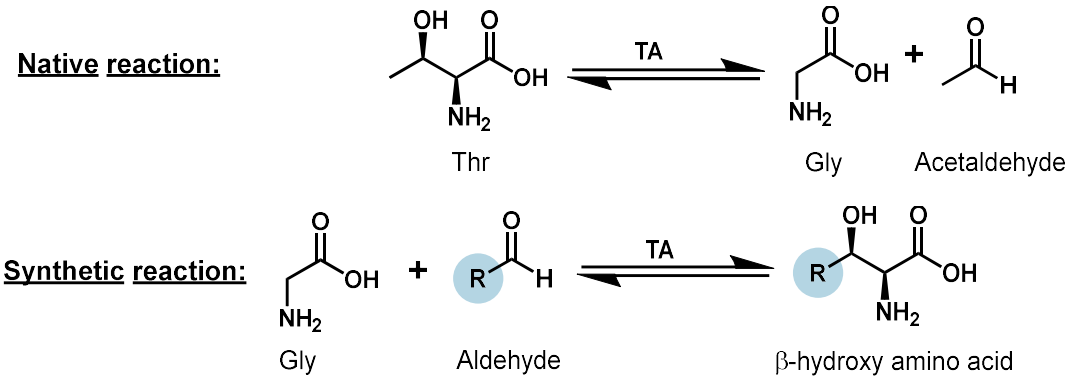


Figure 8. Native and synthetic TA reactions. TAs catalyzes the breakdown of Thr into glycine and acetaldehyde in vivo. This reaction can be reversed by addition of excess glycine and aldehyde to form diverse β-hydroxy amino acids.

There is a remarkable set of PLP dependent enzymes, L-threonine transaldolases (LTTAs), that natively synthesize β-hydroxy amino acids in a single step. LTTAs catalyze retroaldol cleavage of Thr and a subsequent aldol-like addition into an aldehyde to form a new side chain, setting the stereochemistry of the Cβ-OH group (Figure 9). Although the first LTTA, fluorothreonine transaldolase, was discovered in 2001, it took another 11 years to discover the second LTTA in 2012.[63,64] Supported by the advancement in genome sequencing, four new LTTA homologs have been discovered over the last decade during the characterization of biosynthetic pathways of bioactive natural products.[65] Fluorothreonine transaldolase is a didomain LTTA whereas the new discovered LTTAs are single-domain LTTA's, making them more amenable to heterologous expression and biochemical characterization in *E. coli*. However, none of the LTTAs have been characterized in detail. Since LTTA forms β-hydroxy amino acids stereoselectively in a single step,

the characterization of LTTAs could enable a new enzymatic route for the synthesis of β-hydroxy amino acids.

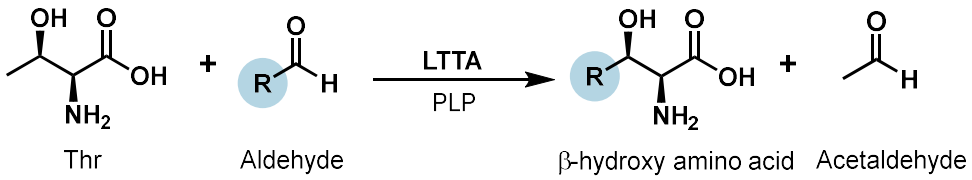


Figure 9. The general reaction catalyzed by LTTA using PLP as a cofactor.

1. 5. L-threonine transaldolases are promising biocatalysts

The first LTTA to be discovered was the fluorothreonine transaldolase (FTA) from *Streptomyces cattleya* in 2001, which forms 4-fluoro-threonine from L-threonine (Thr) and fluoroacetaldehyde (Figure 10).[63] FTA is a 634 residue PLP-dependent protein with two domains: a large N-terminal domain that shares sequence homology with SHMTs and a smaller C-domain. Another LTTA, LipK, was discovered in 2012 during the investigation of the biosynthetic pathway of A-90289, a nucleoside antibiotic. LipK transposes the side chain of Thr with the C5' aldehyde of a nucleoside using PLP as a cofactor to form 5'-C-glycyluridine, a precursor to A-90289 (Figure 10, 11).[66,67] Unlike FTA, LipK is a smaller 424 residue single domain protein. Both FTA and LipK have poor expression in non-*streptomyces* hosts such as *E. coli*.[68]

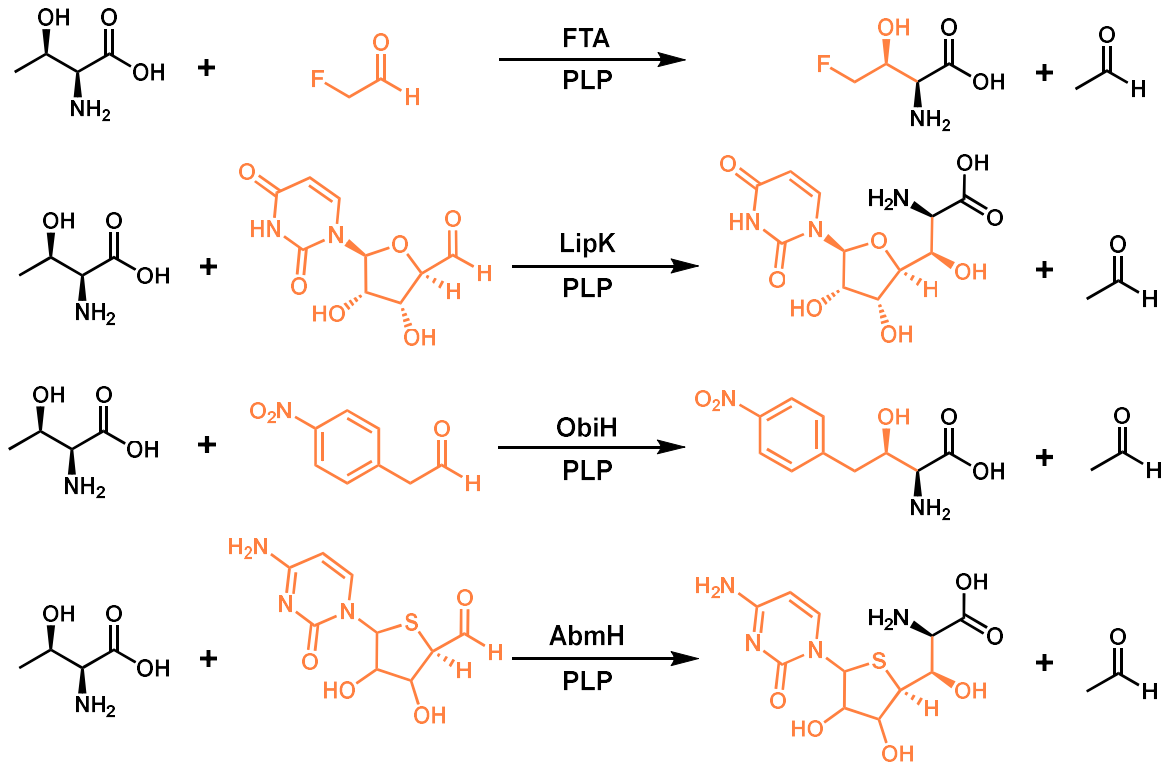


Figure 10. Biocatalytic reactions catalyzed by different LTTAs using Thr and PLP as the cofactor. LTTAs differ in their aldehyde substrates (orange) which determines the side chain of the β-hydroxy amino acid that is generated.

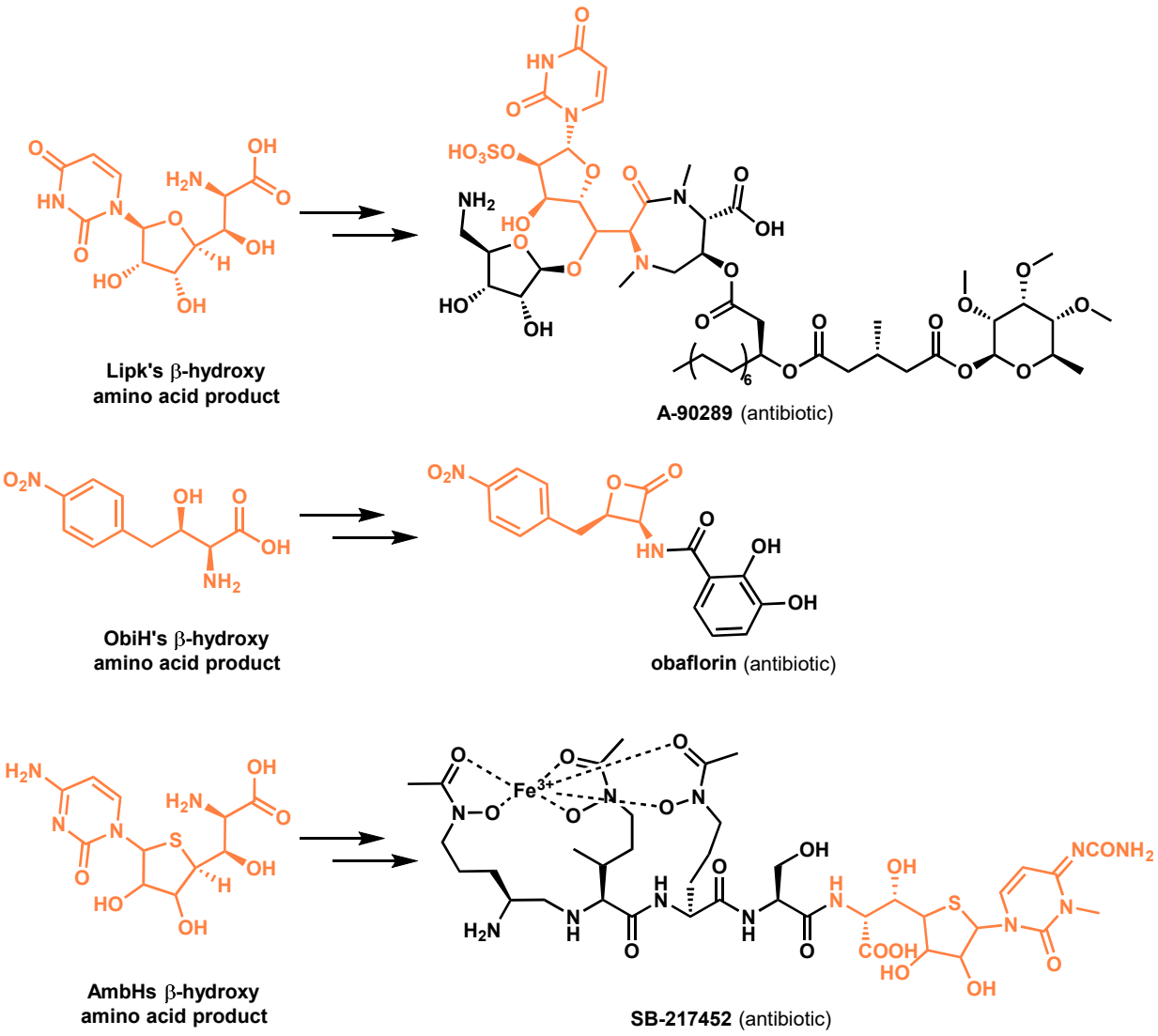


Figure 11. β-hydroxy amino acids (orange) formed by different LTTAs and their corresponding natural products.

In 2017, two groups in parallel discovered ObiH (also known as ObaG), an LTTA in the obafluorin producing microbe Pseudomonas fluorescens.[40,41] Obafluorin belongs to the chemical class of β-lactones, consisting of strained four-membered rings and are closely related to β-lactam antibiotics. Obafluorin acts as an aminoacyl tRNA synthetase inhibitor against both gram-positive

and -negative pathogens.[69] ObiH reacts with Thr and *p*-nitrophenylacetaldehyde to generate β-hydroxy-*p*-nitro-homophenylalanine en route to obafluorin (Figure 10, 11). Preliminary analytical scale reactions suggested that ObiH is also able to react with benzaldehyde, a non-native substrate to produce phenylserine .[40] However, there was no reactivity when glycine or L-serine were used as a substrate suggesting that Thr is essential for ObiH activity.[41] Since, ObiH was the first LTTA to be successfully expressed in E. coli and had apparent reactivity with non-native substrates, we were motivated to evaluate ObiH as a biocatalyst to access β-hydroxy-amino acids. Our efforts in this area are described in the upcoming chapters of this thesis.

In 2018, after we started our work with ObiH, AbmH was discovered in the biosynthetic pathway of SB-217452, an albomycin.[67] Albomycins are sulfur-containing antibiotics which consist of a 6'-amino-4'-thioheptose nucleoside and an iron chelating siderophore that are connected via amide linkage to a serine residue. Similar to LipK, AbmH catalyzes a reaction between Thr and the 5' aldehyde of a thioheptose nucleoside to yield a nucleoside amino acid which undergoes further enzymatic reactions to form SB-217452 (Figure 10, 11). Unlike LipK, AbmH was successfully expressed in *E. coli.* In 2019, another LTTA from *Pseudomonas*, PsLTTA was discovered through genome mining. However, PsLTTA has 99% sequence identity to ObiH (differs only by 2 mutations). Recent studies with PsLTTA as a biocatalyst to access β-hydroxy amino acid will be discussed in Chapter 3.

Three out of the five known LTTAs were discovered over the past five years. These homologs were originally misannotated as serine hydroxymethyl transferase (SHMT)-like enzymes by automated annotation programs. However, these newly discovered LTTAs have low homology to SHMTs (<35% sequence identity). Hence, more LTTAs homologs could be potentially discovered through genome mining. All the LTTA enzymes utilize PLP as the cofactor and Thr as the substrate to build the amino acid backbone of the β-hydroxy amino acid. LTTA enzymes differ in their aldehyde substrates which determines the side chain of the amino acids. The ability of

LTTAs to form a new C-C bond while simultaneously setting the stereochemistry at both the carbon atoms is powerful and motivates detailed biochemical characterization of these enzymes to understand their unique reactivity.

1. 6. Preface to remaining chapters

β-hydroxy amino acids are valuable building blocks for developing the next generation of small molecule and peptide therapeutics. However, existing enzymatic and synthetic routes to access β-hydroxy amino acids suffer from moderate yields, poor selectivity, and limited substrate scope. Hence, a new sustainable route to access β-hydroxy amino acids is highly desired. The remaining chapters of this thesis focus on our attempts to characterize and develop ObiH as a biocatalyst to access valuable β-hydroxy amino acids. In Chapter 2, we discuss the detailed mechanistic and structural investigation of ObiH to understand its unique reactivity. Chapter 3 discusses our efforts to synthesize structurally diverse β-hydroxy amino acids from L-threonine and commercially available aldehydes using *E. coli* overexpressing the ObiH enzyme as a whole cell biocatalyst. Finally, Chapter 4 details our ongoing efforts to engineer ObiH to improve its reactivity and expand its substrate scope. We believe these efforts will lead ObiH to be widely adopted by the synthetic community as a catalyst while simultaneously proving new building blocks for medicinal chemists.

Apart from a few secondary contributions, these were the main projects I have worked on and have resulted in the following publications:

1. **Kumar P***, Meza A*, Ellis JM, Carlson GA, Bingman CA, Buller AR. “L-Threonine Transaldolase Activity Is Enabled by a Persistent Catalytic Intermediate” *ACS Chem. Biol.* **2021**, 16, 86 – 95.
2. Doyon T*, **Kumar P***, et al. “Scalable and selective β-hydroxy-α-amino acid synthesis catalyzed by promiscuous L-threonine transaldolase ObiH”. *ChemBioChem*. **2021,** 22. (*Early access*) Designated as Very Important Paper.
3. **Kumar P***, Bruffy S*, Carlson GA, Buller AR. “Expanding the substrate scope of C-C bond forming transaldolase through directed evolution” (*Manuscript in preparation*)

1. 7. References

1. Crick, F. Central Dogma of Molecular Biology. *Nat.* **227**, 561–563 (1970).

2. Lobanovska, M. & Pilla, G. Focus: Drug Development: Penicillin's Discovery and Antibiotic Resistance: Lessons for the Future? *Yale J. Biol. Med.* **90**, 135 (2017).

3. Granato, E. T., Meiller-Legrand, T. A. & Foster, K. R. The Evolution and Ecology of Bacterial Warfare. *Curr. Biol.* **29**, R521–R537 (2019).

4. Schofield, C. J. *et al.* Proteins of the penicillin biosynthesis pathway. *Curr. Opin. Struct. Biol.* **7**, 857–864 (1997).

5. Blaskovich, M. A. T. Unusual Amino Acids in Medicinal Chemistry. *Journal of Medicinal Chemistry* **59**, 10807–10836 (2016).

6. Mohammad-Zadeh, L. F., Moses, L. & Gwaltney-Brant, S. M. Serotonin: a review. *J. Vet. Pharmacol. Ther.* **31**, 187–199 (2008).

7. Gershenzon, J. & Dudareva, N. The function of terpene natural products in the natural world. *Nat. Chem. Biol.* **3**, 408–414 (2007).

8. Larsen, E. M., Wilson, M. R. & Taylor, R. E. Conformation-activity relationships of polyketide natural products. *Nat. Prod. Rep.* **32**, 1183–1206 (2015).

9. Kamatou, G. P. P., Vermaak, I., Viljoen, A. M. & Lawrence, B. M. Menthol: a simple monoterpene with remarkable biological properties. *Phytochemistry* **96**, 15–25 (2013).

10. Zhang, Y. *et al.* An overview on the biosynthesis and metabolic regulation of monacolin K/lovastatin. *Food Funct.* **11**, 5738–5748 (2020).

11. Hotta, K. & Kondo, S. Kanamycin and its derivative, arbekacin: significance and impact. *J. Antibiot. 2018 714* **71**, 417–424 (2018).

12. Fujisaki, F., Oishi, M. & Sumoto, K. A Conventional New Procedure for N-Acylation of Unprotected Amino Acids. *Chem. Pharm. Bull.* **55**, 124–127 (2007).

13. Yan, T., Feringa, B. L. & Barta, K. Direct N-alkylation of unprotected amino acids with alcohols. *Sci. Adv.* **3**, 6494 (2017).

14. Sabatini, M. T. *et al.* Protecting-Group-Free Amidation of Amino Acids using Lewis Acid

Catalysts. *Chemistry* **24**, 7033 (2018).

15. Dumas, A., Lercher, L., Spicer, C. D. & Davis, B. G. Designing logical codon reassignment – Expanding the chemistry in biology. *Chem. Sci.* **6**, 50–69 (2014).

16. Han, J., Konno, H., Sato, T., Soloshonok, V. A. & Izawa, K. Tailor-made amino acids in the design of small-molecule blockbuster drugs. *Eur. J. Med. Chem.* **220**, (2021).

17. Liu, A. *et al.* New pharmaceuticals approved by FDA in 2020: Small-molecule drugs derived from amino acids and related compounds. *Chirality* **34**, 86–103 (2022).

18. Goldstein, D. S. L-Dihydroxyphenylserine (L-DOPS): A Norepinephrine Prodrug. *Cardiovasc. Drug Rev.* **24**, 189–203 (2006).

19. Mandel, S. J., Brent, G. A. & Larsen, P. R. Levothyroxine therapy in patients with thyroid disease. *Ann. Intern. Med.* **119**, 492–502 (1993).

20. Larden, D. W. & Cheung, H. T. A. Synthesis of N-α-aminoacyl derivatives of melphalan for potential use in drug targeting. *Tetrahedron Lett.* **37**, 7581–7582 (1996).

21. Weigert, W. M., Offermanns, H. & Degussa, P. S. D-Penicillamine—Production and Properties. *Angew. Chemie Int. Ed. English* **14**, 330–336 (1975).

22. Iwata, R., Furumoto, S., Pascali, C., Bogni, A. & Ishiwata, K. Radiosynthesis of O-[11C]methyl-L-tyrosine and O-[18F]Fluoromethyl-L-tyrosine as potential PET tracers for imaging amino acid transport. *J. Label. Compd. Radiopharm.* **46**, 555–566 (2003).

23. Perucca, E., Yasothan, U., Clincke, G. & Kirkpatrick, P. Lacosamide. *Nat. Rev. Drug Discov.* **7**, 973–974 (2008).

24. Evans, C. G., Smith, M. C., Carolan, J. P. & Gestwicki, J. E. Improved synthesis of 15-deoxyspergualin analogs using the Ugi multi-component reaction. *Bioorg. Med. Chem. Lett.* **21**, 2587–2590 (2011).

25. Ota, K. Review of ubenimex (Bestatin): clinical research. *Biomed. Pharmacother.* **45**, 55–60 (1991).

26. Savile, C. K. *et al.* Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **329**, 305–309 (2010).

27. Raja, A., LaBonte, J., Lebbos, J. & Kirkpatrick, P. Daptomycin. *Nat. Rev. Drug Discov.* **2**,

943–944 (2003).

28. Choo, E. J. & Chambers, H. F. Treatment of Methicillin-Resistant Staphylococcus aureus Bacteremia. *Infect. Chemother.* **48**, 267 (2016).

29. Richard Feelders, P. A., Yasothan, U. & Kirkpatrick, P. Pasireotide. *Nat. Rev. Drug Discov. 2012 118* **11**, 597–598 (2012).

30. Tymiak, A. A., McCormick, T. J. & Unger, S. E. Structure determination of lysobactin, a macrocyclic peptide lactone antibiotic. *J. Org. Chem.* **54**, 1149–1157 (2002).

31. Boger, D. L., Colletti, S. L., Honda, T. & Menezes, R. F. Total Synthesis of Bleomycin A2 and Related Agents. 1. Synthesis and DNA Binding Properties of the Extended C-Terminus: Tripeptide S, Tetrapeptide S, Pentapeptide S, and Related Agents. *J. Am. Chem. Soc.* **116**, 5607–5618 (1994).

32. Kato, K., Takita, T. & Umezawa, H. Synthesis of cleonine, amino(1-hydroxycyclopropyl)acetic acid, a novel amino acid contained in cleomycin. *Tetrahedron Lett.* **21**, 4925–4926 (1980).

33. Morens, D. M. & Fauci, A. S. Emerging Pandemic Diseases: How We Got to COVID-19. *Cell* **182**, 1077–1092 (2020).

34. Wencewicz, T. A. Crossroads of Antibiotic Resistance and Biosynthesis. *J. Mol. Biol.* **431**, 3370–3399 (2019).

35. Hu, S., He, W. & Wu, G. Hydroxyproline in animal metabolism, nutrition, and cell signaling. *Amino Acids* (2021). doi:10.1007/S00726-021-03056-X

36. Recktenwald, J. *et al.* Nonribosomal biosynthesis of vancomycin-type antibiotics: A heptapeptide backbone and eight peptide synthetase modules. *Microbiology* **148**, 1105–1118 (2002).

37. Zweerink, M. M., Edison, A. M., Wells, G. B., Pinto, W. & Lester, R. L. Characterization of a novel, potent, and specific inhibitor of serine palmitoyltransferase. *J. Biol. Chem.* **267**, 25032–25038 (1992).

38. Feling, R. H. *et al.* Salinosporamide A: A Highly Cytotoxic Proteasome Inhibitor from a Novel Microbial Source, a Marine Bacterium of the New Genus Salinospora. *Angew. Chemie Int. Ed.* **42**, 355–357 (2003).

39. Shevach, E. M. The Effects of Cyclosporin A on the Immune System. *Annu. Rev. Immunol.* **3**, 397–423 (1985).

40. Scott, T. A., Heine, D., Qin, Z. & Wilkinson, B. An L-threonine transaldolase is required for L-threo-β-hydroxy-α-amino acid assembly during obafluorin biosynthesis. *Nat. Commun.* **8**, 15935 (2017).

41. Schaffer, J. E., Reck, M. R., Prasad, N. K. & Wencewicz, T. A. β-Lactone formation during product release from a nonribosomal peptide synthetase. *Nat. Chem. Biol.* **13**, 737–744 (2017).

42. Moore, M. J. *et al.* Next-Generation Total Synthesis of Vancomycin. *J. Am. Chem. Soc.* **142**, 16039–16050 (2020).

43. Miller, M. J. Hydroxamate Approach to the Synthesis of β-Lactam Antibiotics. *Acc. Chem. Res.* **19**, 49–56 (1986).

44. Tanner, D. Chiral Aziridines—Their Synthesis and Use in Stereoselective Transformations. *Angew. Chemie Int. Ed. English* **33**, 599–619 (1994).

45. Desimoni, G., Faita, G. & Quadrelli, P. Pyridine-2,6-bis(oxazolines), helpful ligands for asymmetric catalysts. *Chem. Rev.* **103**, 3119–3154 (2003).

46. Sugiyama, H., Shioiri, T. & Yokokawa, F. Syntheses of four unusual amino acids, constituents of cyclomarin A. *Tetrahedron Lett.* **43**, 3489–3492 (2002).

47. Singjunla, Y., Baudoux, J. & Rouden, J. Direct synthesis of β-hydroxy-α-amino acids via diastereoselective decarboxylative aldol reaction. *Org. Lett.* **15**, 5770–5773 (2013).

48. M. Trost, B. & Miege, F. Development of ProPhenol Ligands for the Diastereo- and Enantioselective Synthesis of β-Hydroxy-α-amino Esters. *J. Am. Chem. Soc.* **136**, 3016–3019 (2014).

49. Kolb, H. C., VanNieuwenhze, M. S. & Sharpless, K. B. Catalytic Asymmetric Dihydroxylation. *Chem. Rev.* **94**, 2483–2547 (2002).

50. Levitskiy, O. A., Grishin, Y. K. & Magdesieva, T. V. Stereoselective Electrosynthesis of β-Hydroxy-α-Amino Acids in the Form of NiII-Schiff-Base Complexes. *European J. Org. Chem.* **2019**, 3174–3182 (2019).

51. Jiang, W. *et al.* EcdGHK are three tailoring iron oxygenases for amino acid building blocks of the echinocandin scaffold. *J. Am. Chem. Soc.* **135**, 4457–4466 (2013).

52. Recktenwald, J. *et al.* Nonribosomal biosynthesis of vancomycin-type antibiotics: A heptapeptide backbone and eight peptide synthetase modules. *Microbiology* **148**, 1105–1118 (2002).

53. Puk, O. *et al.* Biosynthesis of chloro-β-hydroxytyrosine, a nonproteinogenic amino acid of the peptidic backbone of glycopeptide antibiotics. *J. Bacteriol.* **186**, 6093–6100 (2004).

54. Cryle, M. J., Meinhart, A. & Schlichting, I. Structural characterization of OxyD, a cytochrome P450 involved in β-hydroxytyrosine formation in vancomycin biosynthesis. *J. Biol. Chem.* **285**, 24562–24574 (2010).

55. Chen, H. *et al.* Aminoacyl-S-enzyme intermediates in β-hydroxylations and α,β-desaturations of amino acids in peptide antibiotics. *Biochemistry* **40**, 11651–11659 (2001).

56. Kimura, T., P. Vassilev, V., Shen, G.-J. & Wong, C.-H. Enzymatic Synthesis of β-Hydroxy-α-amino Acids Based on Recombinant d- and l-Threonine Aldolases. *J. Am. Chem. Soc.* **119**, 11734–11742 (1997).

57. Di Salvo, M. L. *et al.* On the catalytic mechanism and stereospecificity of Escherichia coli l -threonine aldolase. *FEBS J.* **281**, 129–145 (2014).

58. Fesko, K. Threonine aldolases: perspectives in engineering and screening the enzymes with enhanced substrate and stereo specificities. *Applied Microbiology and Biotechnology* **100**, 2579–2590 (2016).

59. Chen, Q. *et al.* Improving and Inverting C $_{\beta}$ -Stereoselectivity of Threonine Aldolase via Substrate-Binding-Guided Mutagenesis and a Stepwise Visual Screening. *ACS Catal.* **9**, 4462–4469 (2019).

60. Recombinant, D. *et al.* Enzymatic Synthesis of beta-Hydroxy-alfa-amino Acids Based on Recombinant D- and L-Threonine Aldolases. *J. Am. Chem. Soc.* **7863119**, 11734–11742 (1997).

61. Zheng, W. *et al.* Directed Evolution of L-Threonine Aldolase for the Diastereoselective Synthesis of β-Hydroxy-α-amino Acids. *ACS Catal.* **11**, 3198–3205 (2021).

62. L. Goldberg, S. *et al.* Preparation of β-hydroxy-α-amino Acid Using Recombinant d-

Threonine Aldolase. *Org. Process Res. & Dev.* **19**, 1308–1316 (2015).

63. Murphy, C. D., O'Hagan, D. & Schaffrath, C. Identification of a PLP-Dependent Threonine Transaldolase: A Novel Enzyme Involved in 4-Fluorothreonine Biosynthesis in Streptomyces cattleya. *Angew. Chemie Int. Ed.* **40**, 4479–4481 (2001).

64. Barnard-Britson, S. *et al.* Amalgamation of nucleosides and amino acids in antibiotic biosynthesis: Discovery of an l -threonine: Uridine-5′-aldehyde transaldolase. *J. Am. Chem. Soc.* **134**, 18514–18517 (2012).

65. Wang, S. & Deng, H. Peculiarities of promiscuous L-threonine transaldolases for enantioselective synthesis of β-hydroxy-α-amino acids. *Appl. Microbiol. Biotechnol.* **2**, 3507–3520 (2021).

66. Barnard-Britson, S. *et al.* Amalgamation of Nucleosides and Amino Acids in Antibiotic Biosynthesis: Discovery of an l-Threonine:Uridine-5′-Aldehyde Transaldolase. *J. Am. Chem. Soc.* **134**, 18514–18517 (2012).

67. Ushimaru, R. & Liu, H. Biosynthetic Origin of the Atypical Stereochemistry in the Thioheptose Core of Albomycin Nucleoside Antibiotics. *J. Am. Chem. Soc.* **141**, 2211–2214 (2019).

68. Wang, S. & Deng, H. Peculiarities of promiscuous l-threonine transaldolases for enantioselective synthesis of β-hydroxy-α-amino acids. *Appl. Microbiol. Biotechnol.* **105**, 3507–3520 (2021).

69. Scott, T. A. *et al.* Immunity-Guided Identification of Threonyl-tRNA Synthetase as the Molecular Target of Obafluorin, a β-Lactone Antibiotic. *ACS Chem. Biol.* **14**, 2663–2671 (2019).

Chapter 2

Mechanistic and structural characterization of ObiH, a model L-threonine transaldolase

Content in this chapter is adapted from published work:

Kumar P*, Meza A*, Ellis JM, Carlson GA, Bingman CA, Buller AR. “L-Threonine Transaldolase Activity Is Enabled by a Persistent Catalytic Intermediate” *ACS Chem. Biol.* **2021**, 16, 86 – 95.

*These authors contributed equally to this work

ObiH mechanistic experiments were led by Tony Meza. Crystallographic studies were done in collaboration with Prof. Andrew Buller and Dr. Craig Bingman. MD simulations were performed by Jon Ellis. Biochemical characterization of ObiH variants were done in collaboration with Grace Carlson.

Chapter 2: Mechanistic and structural characterization of ObiH, a model L-threonine transaldolase

2. 1. Introduction

L-Threonine transaldolases (LTTAs) are a poorly characterized class of pyridoxal-5'-phosphate (PLP) dependent enzymes responsible for the biosynthesis of diverse β-hydroxy amino acids.[1] LTTAs catalyze retroaldol cleavage of L-threonine (Thr) and a subsequent aldol-like addition into an aldehyde to form a new side chain, setting the stereochemistry of the Cβ-OH group (Figure 1a). The ability of LTTAs to form a new C-C bond while simultaneously setting the stereochemistry at both the carbon atoms is elegant and could be synthetically valuable. Out of the five reported LTTAs, ObiH was the first to be successfully expressed in *E. coli*.[1] ObiH reacts with Thr and *p*-nitrophenylacetaldehyde to form β-hydroxy-*p*-nitrohomophenylalanine en route to obafluorin, a β-lactone antibiotic (Figure 1b).[2–4] In addition, ObiH was reported to have activity with a non-native aldehyde suggesting that ObiH could be a promising catalyst to access β-hydroxy amino acids.[3]

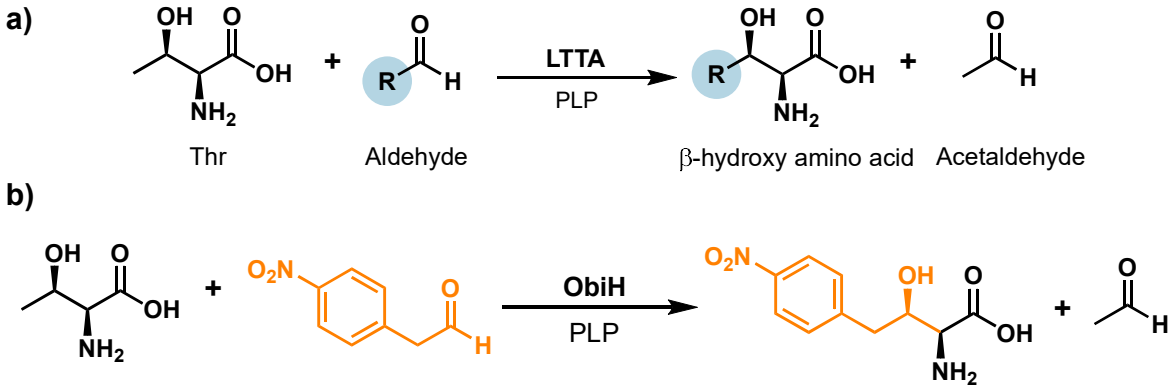


Figure 1. LTTA reactions. a) The general reaction catalyzed by LTTA using PLP as a cofactor. b) Native ObiH reaction. β-hydroxy-*p*-nitro-homophenylalanine is synthesized from *p*-nitrophenylacetaldehyde (orange) and Thr (black).

The mechanism of LTTA enzymes has not been studied in detail, and no structures are available. A mechanism has been proposed based on analogy to the homologous Thr aldolase (TA) and serine hydroxymethyltransferase (SHMT) enzymes, which perform related transformations.[5] TAs catalyze the thermodynamically favorable breakdown of Thr into glycine (Gly) and acetaldehyde (Figure 2). This reaction begins with covalent capture of Thr as an external aldimine with PLP, $E(Aex^{Thr})$.[6] The enzyme orients the sidechain of Thr such that it is periplanar to the π-system of the cofactor, in accordance with Dunathan's stereoelectronic hypothesis.[7,8] Retroaldol cleavage is initiated by deprotonation of the hydroxyl side chain by a conserved histidine (His) residue that π-stacks with the cofactor, either directly or through a proton relay with water.[6] Cleavage of the Cα-Cβ bond releases acetaldehyde and the resultant Cα carbanion is stabilized through resonance as a highly basic glycyl quinonoid intermediate, $E(Q^{Gly})$.[6] The glycyl quinonoid is subsequently protonated to form the external aldimine of glycine, $E(Aex^{Gly})$, which is released to complete the catalytic cycle. The TA reaction can be reversed to run in the synthetic direction in vitro by adding excess glycine and aldehyde. These enzymes have been shown to form diverse β-hydroxy amino acids, setting two stereocenters in the process.[6,9,10] However, the reversibility of the reaction and its relatively modest selectivity leads to scrambling of the stereochemistry at the β-position, limiting synthetic utility.[11–14]

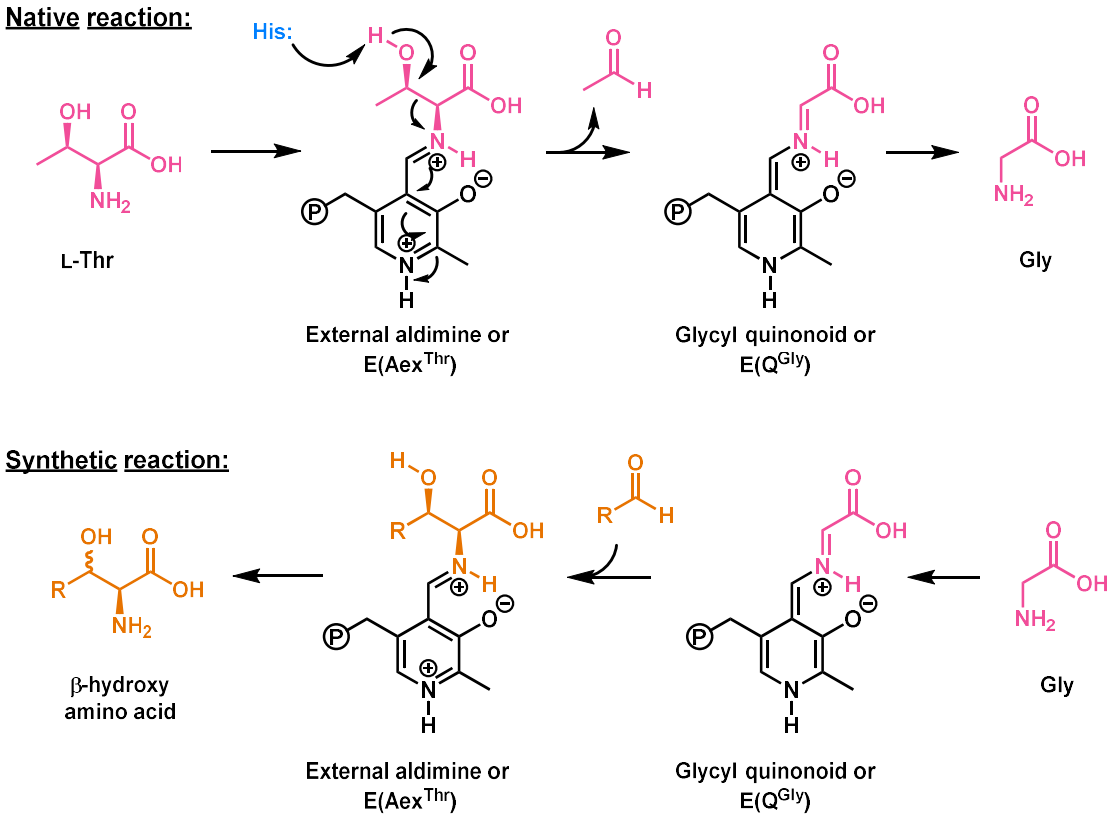


Figure 2. Native and synthetic TA reactions. TAs catalyzes the breakdown of Thr into glycine and acetaldehyde in vivo. This reaction can be reversed by addition of excess glycine and aldehyde to form diverse β-hydroxy amino acids. The PLP cofactor (black) is shown covalently bound either to the substrates/intermediates (pink) or to the product (orange). The catalytic base in shown in blue.

Alternatively, LTTAs cleanly form β-hydroxy amino acids in vivo using the same Thr starting material and PLP cofactor.[15] Although the molecular details have yet to be elucidated, it is believed that these enzymes function by intercepting a highly reactive E(Q^{Gly}) intermediate. Initial studies with the LTTAs demonstrate several promising features for synthetic applications.[16,17] However, the lack of structural information stymies our understanding of LTTA mechanism and hinders targeted engineering approaches. Here, we describe mechanistic and structural characterization of ObiH, a model LTTA, to understand the unique reactivity of these enzymes.

2. 2. Results and Discussion

2. 2. 1. ObiH forms a meta-stable PLP-glycyl quinonoid

ObiH was heterologously expressed as a *N*-His construct in *E. coli*, as was previously described.[2] This procedure reliably yields ~250 mg of protein per L culture (Figure 3a). Consistent with previous reports, this protein is pink in color whereas other PLP-dependent enzymes are yellow.[2,3,17] The UV-vis spectrum of ObiH shows a peak at 415 nm, characteristic of the covalently bound internal aldimine adduct E(Ain), as well as an additional peak at 515 nm, whose intensity varied between independent preparations of ObiH, that accounts for the pink color (Figure 3b).

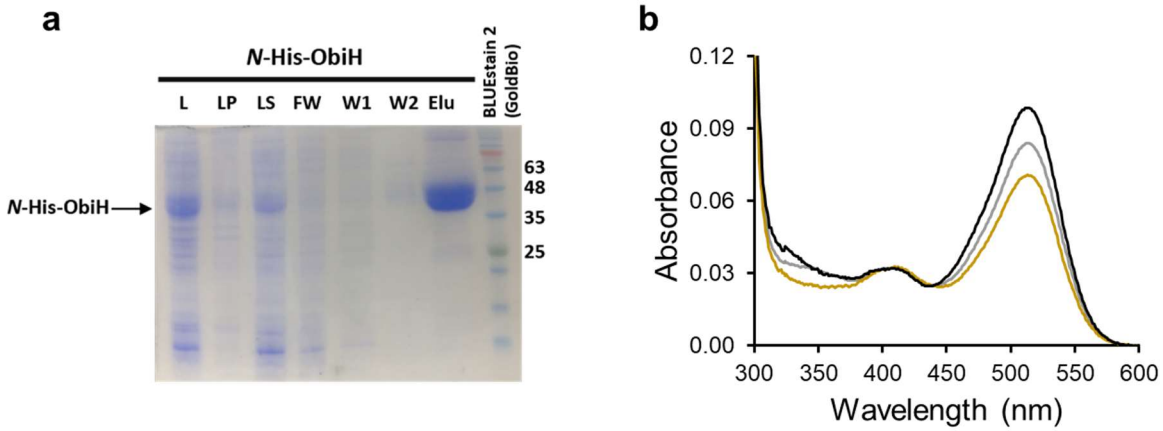


Figure 3. ObiH purification. a) Representative SDS-PAGE from Ni-NTA purification of ObiH. L is lysate, LP is lysate pellet, LS is lysate supernatant, FW is Flow through, W1 and W2 are washes with 20 and 40 mM Imidazole respectively, Elu is elution. The ladder used was GoldBio BLUEstain2 protein ladder. **b)** UV-vis spectra of as purified ObiH from three different preps following identical protocol. Variable amount of 515 nm species was observed between preps.

Previous steady-state kinetic analysis of ObiH established that the enzyme has a relatively high K_M for Thr, 40 mM.[3] We began our detailed mechanistic study by adding 100 mM Thr to ObiH and monitored the reaction by UV-vis spectroscopy. Addition of Thr resulted in an intense absorbance with λ_{max} = 493 nm, characteristic of a PLP quinonoid adduct, and distinct from the

515 nm peak (Figure 4a). We assign this species as E(Q^{Gly}), which is formed by retroaldol cleavage of a covalently bound PLP-Thr adduct, E(Aex^{Thr}) (Figure 5). The E(Q^{Gly}) absorbance band increased for several minutes before reaching a maximum and slowly decayed over the course of several minutes (Figure 4b). Notably, the species absorbing at 515 nm appeared not to enter the catalytic cycle under these conditions, suggesting the presence of a covalently inactivated PLP adduct.

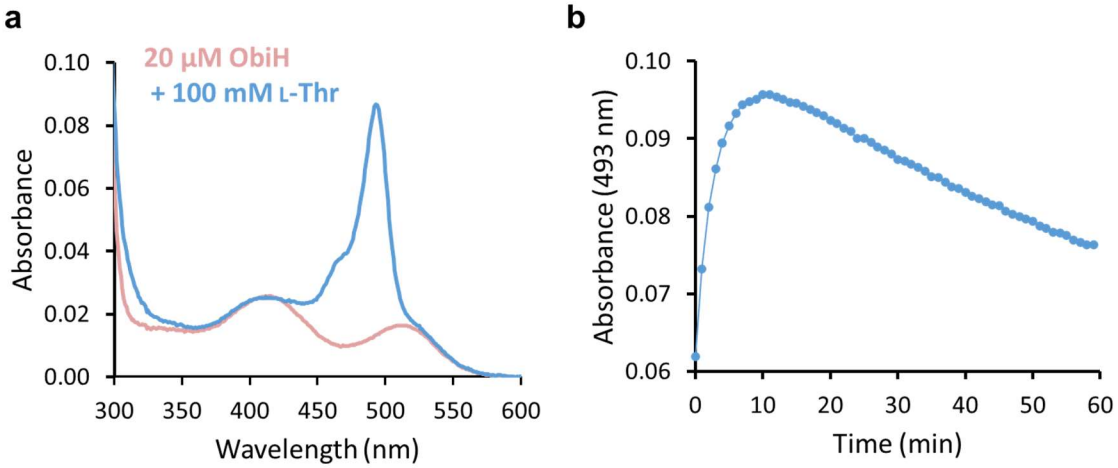


Figure 4. Spectroscopic characterization of ObiH catalytic intermediates. a) Absorbance spectra of ObiH (pink) and after addition of Thr (blue). Addition of Thr results in large peak at 493 nm, E(Q^{Gly}). **b)** Plot of absorbance at 493 nm vs time after addition of Thr. 493 nm peak increases rapidly before reaching a maximum with subsequent decay over several minutes.

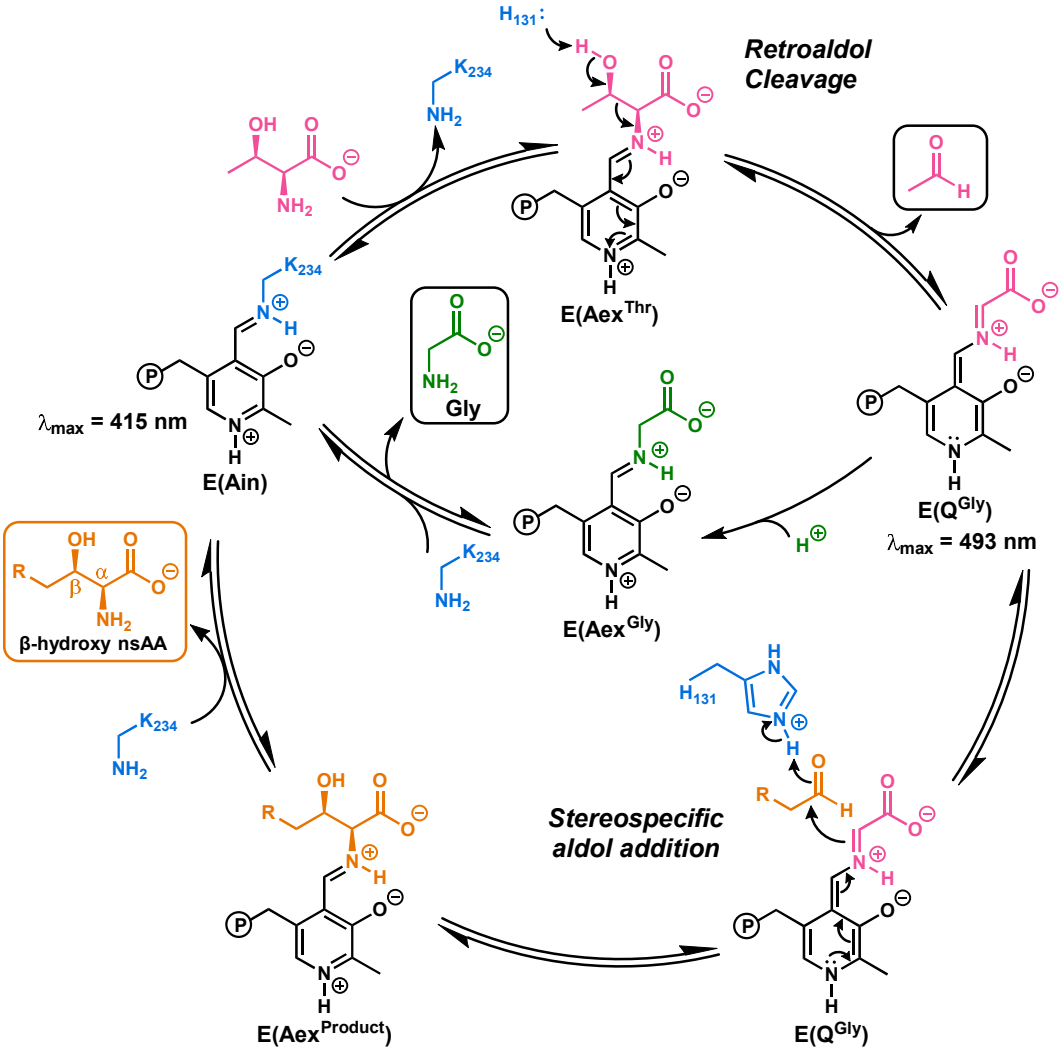


Figure 5. Catalytic mechanism of ObiH. The mechanism of ObiH catalysis with Thr and an aldehyde substrate (outer cycle) along with the disfavored shunt pathway (inner cycle). The PLP cofactor (black) is shown covalently bound either to the substrates/intermediates (pink) or to the relevant catalytic residues of the protein (blue).

A key insight into the nature of the 515 nm band came from Mr. Tony Meza, who observed that purified, pink ObiH seemed to lose its color when exposed to sunlight. Mr. Meza found that the species that absorbs at 515 nm can be converted into catalytic active E(Ain) species by green light (Figure 6). Phototreatment transformed the pink ObiH into a traditional yellow protein and afforded us the opportunity to cleanly assay the mechanistic properties of ObiH as a model LTTA. Further characterization of the 515 nm species, along with additional detailed experiments into the mechanism of ObiH were conducted by Mr. Meza and are described in further detail in Kumar et al.[18]

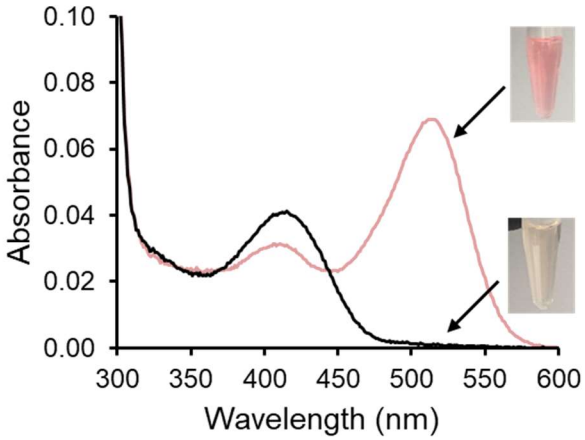


Figure 6. Phototreatment produce homogeneous ObiH. Absorbance spectra of natively purified ObiH (pink) and phototreated ObiH (black). Phototreatment results in the complete loss of the 515 nm species and increase in the E(Ain) peak at 415 nm. Images of ObiH stock solutions before and after phototreatment. This data was collected by Tony Meza.

Once the catalytic cycle is initiated by addition of Thr, many potential pathways may be responsible for the quenching of E(Q^{Gly}) on the h timescale. We hypothesized that, in analogy to the distantly-related TA enzymes, E(Q^{Gly}) reacted through simple protonation to form the glycyl external aldimine, E(Aex^{Gly}).[6] Consistent with this hypothesis, in the absence of an electrophile substrate, ObiH formed a small, but measurable amount of Gly, corresponding to <50 turnovers in 16 h (Figure 5, inner cycle). Furthermore, a higher amount of glycine was formed at pH: 7.0 compared to pH: 8.5, demonstrating the role of protonation in the quenching of E(Q^{Gly}) species.

To further confirm that Gly is formed through an ObiH-mediated process, we added Thr to ObiH to form a large population of E(Q^{Gly}), and then reductively trapped the decay product as a secondary amine via the addition of $NaBH_4$. This reaction was monitored both spectrophotometrically and via UPLC-MS. Spectroscopic experiments showed depletion of the absorbing species in the range of 400 – 420 nm, indicating that the imines present E(Aex^{Thr}), E(Aex^{Gly}), or unreacted E(Ain) were rapidly reduced by $NaBH_4$ (Figure 7a). A new absorbance band at ~340 nm appeared, consistent with formation of a reduced, secondary amine adduct.[19] However, the E(Q^{Gly}) species (493 nm), which is electron-rich, was not immediately reduced by $NaBH_4$ and decayed at a similar rate to reactions containing only Thr. The products of this reaction were monitored by UPLC-MS analysis and showed clear formation of a reduced glycyl adduct (Figure 7b).

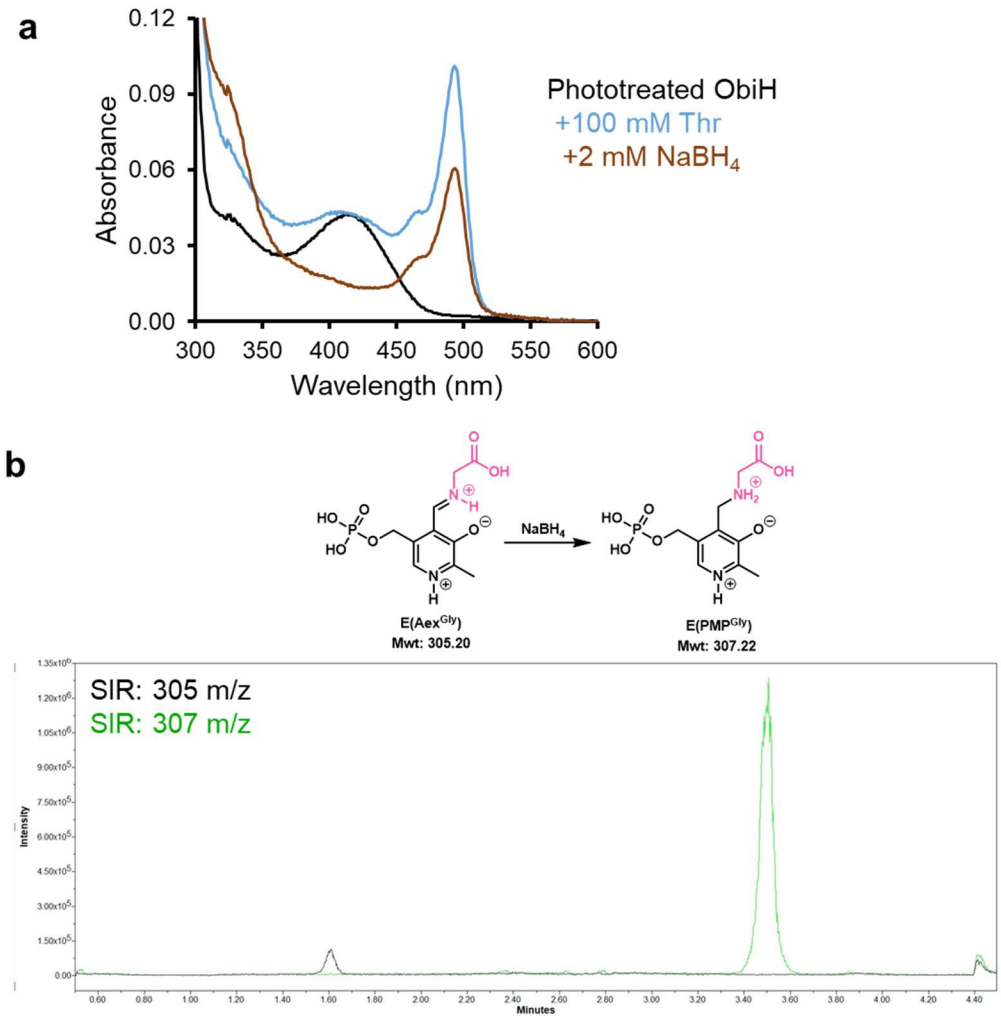


Figure 7. E(Q^{Gly}) and E(Q^{in}) are resistant to $NaBH_4$ reduction. a) Spectrum of phototreated ObiH (black). 100 mM Thr was added and E(Q^{Gly}) peak increased to a maximum at 10 min (blue). 2 mM $NaBH_4$ was then added and spectra were gathered every minute for 30 min. Spectra of reaction 25 min after the addition of $NaBH_4$ (brown) shows at increase at 340 nm and a decrease at 415 nm, as well as a large population of E(Q^{Gly}). This spectrum was collected by Tony Meza. **b)** Overlap of SIR corresponding to E(Aex^{Gly}) and E(PMP^{Gly}) after reduction of ObiH. Chemdraw figure corresponding to this transformation is shown on the top.

While the above data show that E(Q^{Gly}) is kinetically slow to react, they offer only indirect information on the thermodynamic stability of this intermediate. We therefore probed the effect of saturating Gly (1.0 M) on E(Ain) and observed no evidence of quinonoid formation, indicating that

population of E(Q^{Gly}) is not enabled by thermodynamic stabilization in the enzyme active site (Figure 8). Instead, these data establish that E(Q^{Gly}) species is a kinetically-trapped, high-energy intermediate.

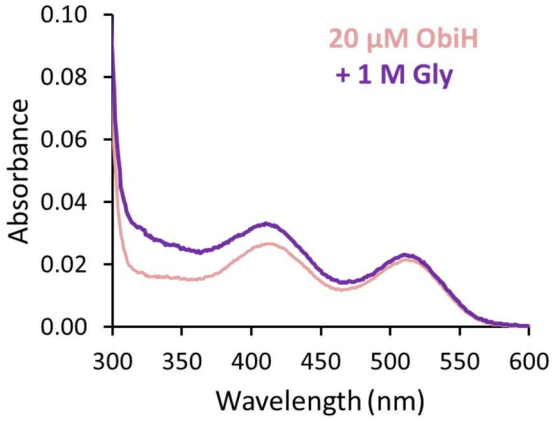


Figure 8. ObiH does not catalyze the retroaldol cleavage of Gly to form E(Q^{Gly}). Spectrum of 20 µM ObiH (pink). 1 M Gly was added and spectra were acquired every 1 minute for 15 minutes. 10 min after addition of 1 M Gly (purple) shows no evidence of a peak at 493 nm.

Quinonoid intermediates are nearly ubiquitous among PLP-dependent enzymes.[20] Some enzymes form thermodynamically stable quinonoids simply by binding substrates, products, or analogs thereof.[21–23] For example, addition of Gly to TA enzymes results in formation of a thermodynamically stable quinonoid.[6] Other enzymes only transiently form quinonoids and rapid kinetic analysis is needed to observe them.[24] ObiH is exceptional, in that it forms a large population of kinetically-trapped, thermodynamically unstable quinonoid. Were protonation and release of Gly to occur, this would be thermodynamically favored in vivo, precluding biosynthesis of new β-hydroxy amino acids. Hence, there is a selective pressure to kinetically shield E(Q^{Gly}) from protonation.

2. 2. 2. ObiH quinonoid rapidly reacts to form β-hydroxy amino acids

The native electrophile in the ObiH reaction is *p*-nitrophenylacetaldehyde, which is generated via a thiamine-dependent decarboxylation from the corresponding α-keto acid.[2,3] However, due to the inherent instability of α-aryl acetaldehydes, we sought an alternative electrophile for our mechanistic studies. Recent experiments using a biocatalytic cascade showed ObiH, as well as its downstream enzymes in obafluorin biosynthesis, can react with a range of aliphatic and benzylic aldehydes.[25] While the synthetic utility of ObiH with aromatic aldehydes has been recently reported,[17,26] we were drawn to the reaction with aliphatic aldehydes as mechanistic probes because they do not have confounding signals in their UV-vis spectra. We probed the on-path reactivity of ObiH via addition of reactive aldehyde to pre-formed E(Q^{Gly}). Addition of 20 mM proponal showed the rapid reaction of E(Q^{Gly}) within the mixing time of the experiment (< 20 s) and the persistence of a small, steady population of quinonoid (Figure 9). This data shows that E(Q^{Gly}) is thermodynamically unstable in ObiH active stie. These experiments were additionally repeated by Mr. Meza with photo-treated and spectroscopically clean ObiH, the results of which are consistent with those shown below.

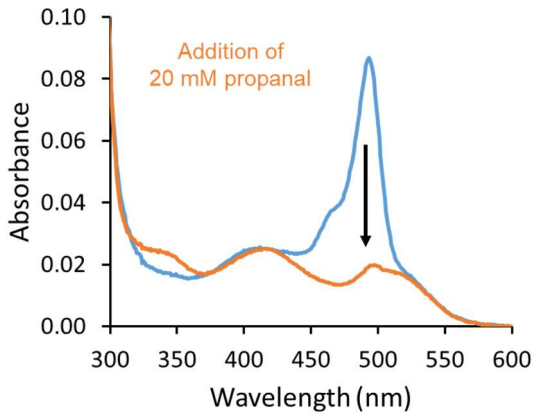


Figure 9. On cycle reactivity of E(Q^{Gly}) with propanal. Absorbance spectrum of 20 μM ObiH after addition of 100 mM Thr is shown in blue. Subsequent addition of propanal (orange) results in substantial reduction of the 493 nm peak.

Furthermore, Mr. Meza was able to isolate β-hydroxy-Leu from Thr and isobutyraldehyde using ObiH as the catalyst (described in Kumar et al).[18] This reactivity is enabled by the ObiH's ability to form a large population of reactive quinonoid (Figure 4). Were the rate of competing protonation pathway facile, formation of β-hydroxy-Leu would be severely limited. Hence, the long lifetime of the E(Q^{Gly}) intermediate enables reactivity with non-native aldehyde substrates. This reactivity of ObiH with non-native aldehydes, combined with the high expression titer of ObiH and its stability over months at -80 °C, make this enzyme highly attractive for future biocatalytic applications.

2. 2. 3. ObiH crystallizes with an unanticipated active site conformation

To unravel the structural properties underpinning the unique reactivity of ObiH, we crystallized *N*-His-ObiH to capture the enzyme in its internal aldimine state, E(Ain). High resolution, 1.66-Å, X-ray diffraction data were collected on crystals that retained the 515 nm species, as the pink color was maintained throughout the process (Table 1). The structure was solved by molecular replacement with a distantly related serine hydroxymethyl transferase (PDB ID: 4OT8, 28.2% identity).[27] The asymmetric unit of the ObiH crystal was comprised of four protamers (two dimers) for a total of four unique observations of the active site. ObiH crystallized as a domain swapped homodimer, with an extension of the C-terminus forming the dimer interface. Such dimerization is consistent with other members of the fold-type I superfamily of PLP-dependent enzymes (Figure 10a).[28] Structural superposition of all non-hydrogen atoms of the four ObiH protamers shows they are highly consistent (RMSD = 0.47 $\pm$ 0.06 Å^2).

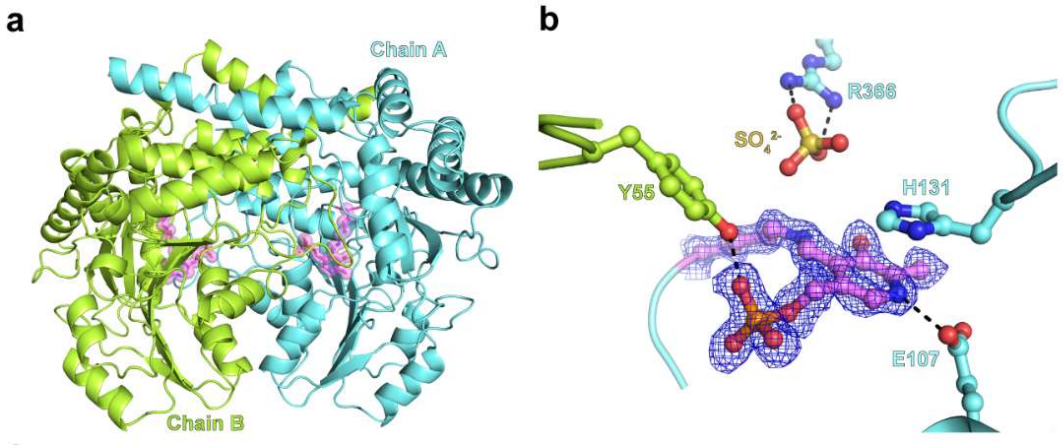


Figure 10. Structure of ObiH. a) Cartoon representation of the overall structure of ObiH. Individual monomers are colored cyan (chain A) and lime (chain B). E(Ain) is shown as semi-transparent pink spheres and sticks. **b)** ObiH active site residues superimposed on the *2MFo - DFc* electron density map (blue mesh, σ = 1.5) are shown as sticks and colored as in (**a**). Hydrogen bonds are shown as black dashes.

Table 1. X-Ray crystallographic data collection and refinement statistics

PDB ID	7K34
Protein	*N*-His-ObiH
Data Collection	
Space group	$P4_1$
Cell dimensions (Å)	a,b,c = 118.6, 118.6, 130.0
Cell angles	α = β = γ = 90°
Wavelength (Å)	1.03321
Beamline	APS 23-ID-D
Resolution (Å)	50– 1.66
Last bin (Å)	(1.70 – 1.66)
No. observations	2826075
Completeness (%)	100.0 (100.0)
R_{pim}	0.058 (0.827)
CC(1/2)	0.996 (0.450)
I/σI	6.9 (1.0)
Redundancy	13.4 (13.1)
Refinement	
Total no. of reflections	200,497
Total no. of atoms	14,557
Final bin (Å)	(1.70 – 1.66)
R_{work} (%)	21.2 (32.6)
R_{free} (%)	24.2 (34.2)
Average B factor (Å^2)	27.2
Ramachandran plot Favored, %	97.5
Allowed, %	100.0
Outliers, %	0.0

Values in parenthesis are for the highest resolution shell. R_{merge} is $\Sigma|Io - I| / \Sigma Io$, where Io is the intensity of an individual reflection, and I is the mean intensity for multiply recorded reflections. R_{work} is $\Sigma||Fo - Fc|| / Fo$, where Fo is an observed amplitude and Fc a calculated amplitude. R_{free} is the same statistic calculated with a 5% subset of the data that was excluded from refinement.

While the vast majority of the structure was rapidly built following common structural motifs, we were initially stymied by the density corresponding to a Cys-cisPro linkage in a π-bulge of an α-helix (Figure 11). The carbonyl oxygen of Cys262 is just 2.9 Å away from the carbonyl of the i-4 residue, an otherwise strained conformation that appears to be stabilized by hydrogen bonds from each carbonyl to Arg99.

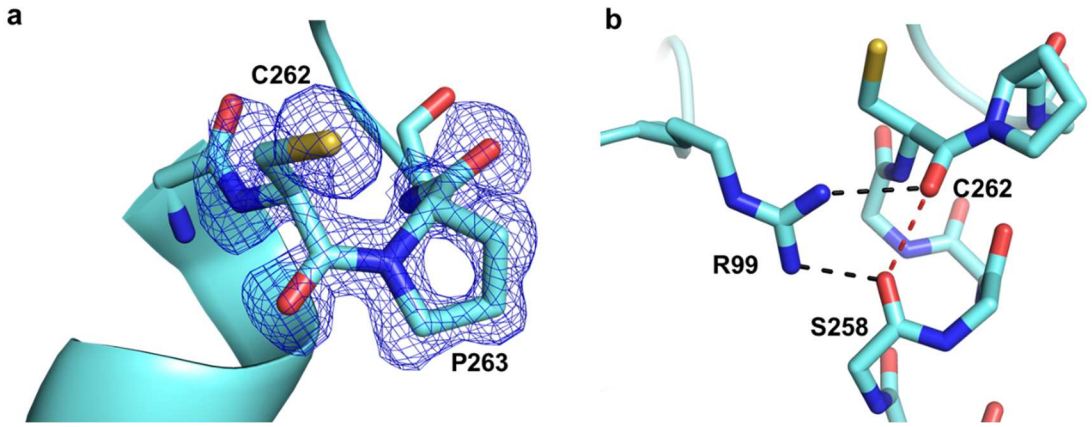


Figure 11. Uncommon structural features of ObiH. a) Cys-cisPro linkage (shown as sticks) in a π-buldge of an a-helix superimposed on the *2MFo - DFc* electron density map (blue mesh, σ = 1.5). **b)** Strained conformation due to the destabilizing interaction between carbonyl oxygens of C262 and S258 due to the π-buldge appears to be stabilized by the H-bonds from each carbonyls to Arg99. Hydrogen bonds are shown as black dashes. Destabilizing interaction is shown as red dashes.

The ObiH active site lies at the dimer interface, with most of the residues contributed from a single subunit. The electron density is consistent with a typical E(Ain) state of the enzyme where the conserved Lys234 forms a Schiff's base adduct with the PLP cofactor (Figure 10b).[29] A molecule of sulfate is bound in the active site with the same orientation in all four subunits and forms a salt bridge with Arg366. Such complexes with a sulfate or phosphate are common among PLP-dependent enzymes and often correspond to the carboxylate-binding motif within the active

site.[29,30,31,32] The pyridine ring of PLP is π-stacked with conserved His131 (Figure 10b, 12). Similar to the conserved His residue in TA enzymes that acts as the catalytic base, His 131 could be the catalytic base in ObiH which initiates the retroaldol cleavage of Thr. The PLP phosphate is buried with an intricate web of hydrogen bonds, including two to Tyr55 and Asn268 from the partner subunit. A multiple sequence alignment reveals that Lys234, Arg366, His131 and Tyr55 are highly conserved across biochemically characterized LTTAs (Figure 12).

Figure 12. Sequence alignment of example LTTAs and the evolutionarily related SHMT. Active site residues are highlighted with red boxes with numbering colors according to (b) and (c). GenBank ID of ObiH is ARJ35753, FTA is WP_014151017, Lipk is BAJ05887, AbmH is AVI57436 and SHMT is AFJ20773.

Another highly conserved residue in this family of enzymes is the residue that hydrogen bonds to the pyridine nitrogen of the cofactor, Asp204.[28] Ion pairing with the pyridinium moiety enforces protonation of the cofactor, thereby increasing its electrophilicity. However, the universally conserved Asp residue appeared to be tucked under the pyridine ring and, instead, Glu107 is in position to form a salt bridge with the cofactor (Figure 13). Mutation of the residue that H-bonds to the pyridine of PLP is known to occur within the fold type II PLP-dependent enzymes, but each residue is highly conserved among its family members.[29,33,34] However, Glu107 is not conserved between LTTAs and we speculated this residue might not be essential for ObiH function (Figure 12). To determine which residue was engaging the cofactor during catalysis, we made substantial efforts to trap an ObiH external aldimine. Despite successful

crystal growth and diffraction, the crystals were excessively twinned and no clear picture of the active site emerged. We therefore sought to use computational methods to build an atomic model of ObiH external aldimine, E(Aex^{Thr}), to assess whether a conformational rearrangement involving Glu107 and Asp204 might be occurring.

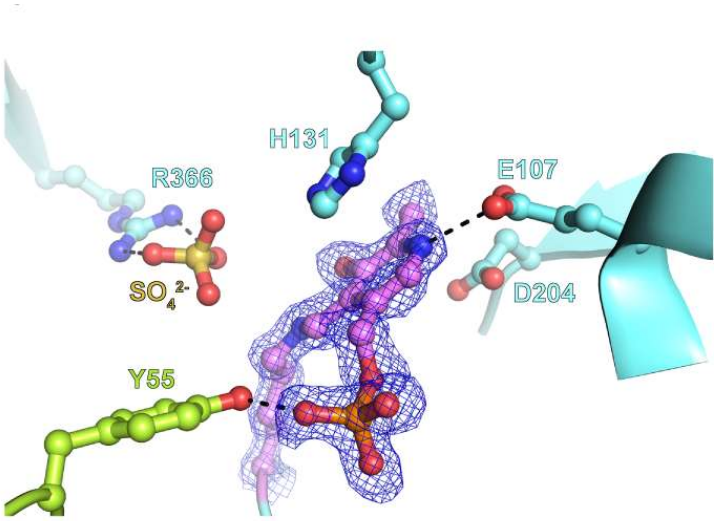


Figure 13. Cofactor and active site residue interactions in ObiH. ObiH active site residues superimposed on the *2MFo - DFc* electron density map (blue mesh, σ = 1.5) are shown as sticks. Individual monomers are colored cyan (chain A) and lime (chain B). E(Ain) is shown as semi-transparent pink spheres and sticks. Hydrogen bonds are shown as black dashes.

2. 2. 4. Asp204 is essential for ObiH catalysis

Prior literature has shown that PLP-dependent enzymes often undergo active site rearrangements upon substrate binding, including motion of the cofactor.[27,29,35] To simulate such a motion in ObiH, we conducted molecular dynamics (MD) simulations of the ObiH dimer in the E(Aex^{Thr}) state, which were performed by Mr. Jon Ellis and are described in detail in Kumar et al.[18]. We used the presence of H-bond between His131 and the sidechain hydroxyl group of E(Aex^{Thr}) as an indicator of a catalytic state, as His131 is the evolutionarily-conserved base for retroaldol cleavage (Figure 12). We found that His131 H-bonding to the sidechain hydroxyl group of E(Aex^{Thr}) predominantly coincided with Asp204 H-bonding to the PLP-N1 (Figure 14). This observation suggests that Asp204 not Glu107 is the main H-bonding acceptor in the catalytically active state of the enzyme.

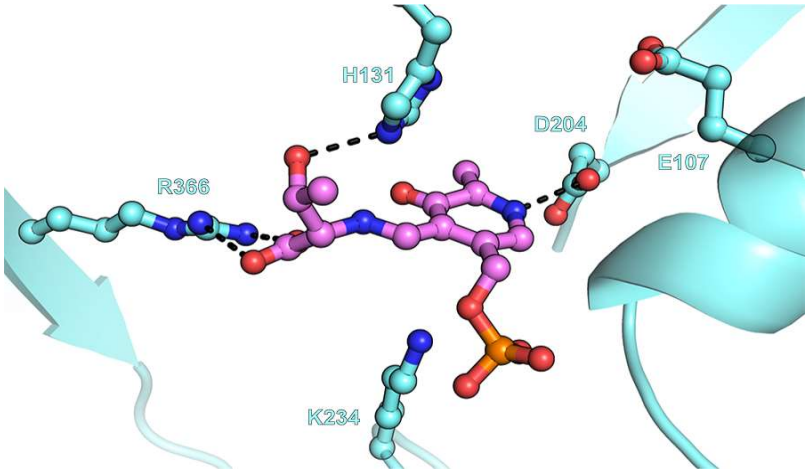


Figure 14. Molecular dynamics simulation of E(Aex^{Thr}). E(Aex^{Thr}) shown in pink. ObiH shown in cyan. Hydrogen bonds that are proposed to facilitate substrate binding and catalysis are shown as black dashes.

To experimentally validate the insights gained from crystallographic, bioinformatics analyses and molecular dynamic simulations, we performed biochemical characterization of His131, Glu107 and Asp204 variants. Rather than the traditional alanine scan approach to probe

the ‘importance’ of a residue, we screened a site-saturation mutagenesis library at His131 for retention of function with biphenyl-4-carboxaldehyde, as the products were well-behaved on UPLC-MS. The native ObiH enzyme, bearing His131, possessed the highest activity under the screening conditions (Figure 15a). The only residue that supported any catalytic function other than His was Gly, which may allow water to enter the active site and rescue function. This experiment supports our observation from crystallographic and bioinformatic analysis that His131 is the catalytic base.

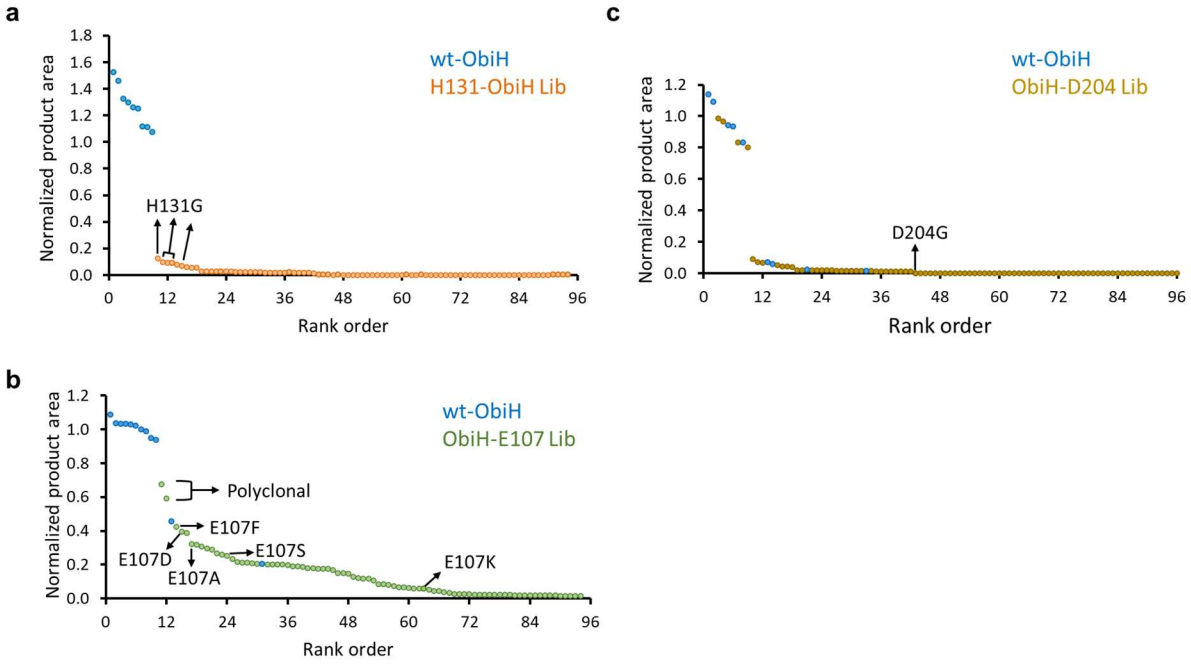


Figure 15. Activity profile of ObiH-H131, ObiH-E107 and ObiH-D204 site saturation mutagenesis libraries. For each of the libraries, lysate supernatants were incubated with 75 mM Thr and 15 mM Biphyenyl-4-carboxaldehyde at 37 °C for 16 h. Then, proteins in the reaction mixture were precipitated by addition of acetonitrile. After centrifugation, 1 uL of supernatant were injected into Acquity UPLC-MS (Waters) to quantitate the formation of b-hydroxy amino acid through UV@254 nm.

Next, we screened a site-saturation mutagenesis library at Glu107. This work was done in collaboration with Grace Carlson, a talented undergraduate researcher in the group. Similar to His131, the native ObiH enzyme, bearing Glu107, possessed the highest activity. Unlike His131, however, many other variants retained catalytic activity (Figure 15b). To further probe the role of Glu107, we expressed and purified the conservative E107Q variant, which has a similar steric profile to the native residue but cannot form an ion pair with the pyridinium nitrogen. This protein was pink in color and behaved similarly to wild type protein upon phototreatment (Figure 16a). Spectroscopic analysis affirmed that ObiH-E107Q binds PLP and, upon addition of Thr, enters the catalytic cycle to form a meta-stable quinonoid that goes on to react with propanal (Figure 16b).

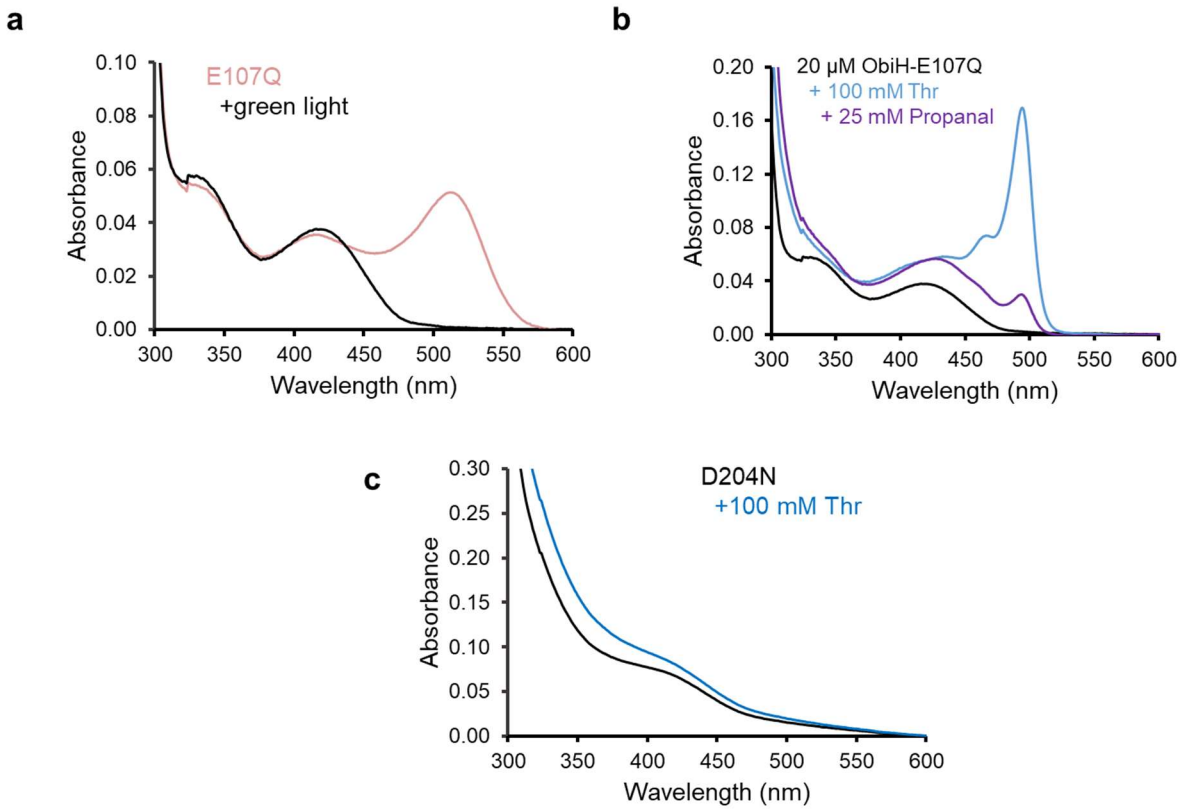


Figure 16. UV-vis spectra of ObiH-E107Q and ObiH-D204N. a) Absorbance spectra of purified E107Q-ObiH (pink) and phototreated E107Q-ObiH (black). **b)** Absorbance spectra of purified D204N-ObiH (black) and after addition of 100 mM Thr (black).

In stark contrast to mutation at E107, saturation mutagenesis at the evolutionarily conserved Asp204 position was catastrophic and only wild-type enzyme retained activity (Figure 15c). We attempted to further probe the contribution of Asp204 through study of the conservative D204N variant. This protein aggregated during purification and was stable only at low concentrations. The purified ObiH-D204N was colorless and UV-vis analysis revealed evidence of only trace PLP binding (Figure 16c). These experiments unambiguously demonstrate that Asp204, not Glu107, is essential for efficient cofactor binding and enforces protonation at PLP-N1 to enable the unique transaldolase activity of this enzyme (Figure 14). Given the conservation of Asp204 among LTTA enzymes, we anticipate its role in catalysis will be a general feature among this enzyme family (Figure 12).

2. 3. Conclusions

L-Threonine transaldolases (LTTAs) were a poorly characterized class of pyridoxal-5'-phosphate (PLP) dependent enzymes responsible for biosynthesis of diverse β-hydroxy amino acids. Here, we studied the catalytic mechanism of ObiH, an LTTA essential for biosynthesis of the β-lactone natural product obafluorin. Heterologously expressed ObiH purifies as a mixture of chemical states including a catalytically inactive form of the PLP cofactor. Photoexcitation of ObiH promotes the conversion of the inactive state of the enzyme to the active form. UV-vis spectroscopic analysis revealed that ObiH catalyzes the retro-aldol cleavage of L-threonine to form a remarkably persistent glycyl quinonoid intermediate. Protonation of this intermediate is kinetically disfavored, enabling on-cycle reactivity with aldehydes to form β-hydroxy amino acids. To further understand the structural features underpinning this desirable reactivity, we determined the crystal structure of ObiH bound to PLP as the Schiff's base at 1.66 Å resolution. This high-resolution model revealed a unique active site configuration wherein the evolutionarily conserved Asp that traditionally H-bonds to the cofactor is swapped for a neighboring Glu. Molecular dynamics simulations combined with mutagenesis studies indicate that a structural rearrangement is associated with L-threonine entry into the catalytic cycle. Together, these data explain the basis for the unique reactivity of LTTA enzymes and provide a foundation for future engineering and mechanistic analysis.

2. 4. Materials and Methods

General experimental procedures

Chemicals and reagents were purchased from commercial suppliers (Sigma-Aldrich, VWR, Chem-Impex International, Combi-blocks, Alfa Aesar, New England Biolabs, Zymo Research, Bio-Rad) and used without further purification unless otherwise noted. BL21 (DE3) *E. coli* cells were electroporated with a Bio-Rad MicroPulser electroporator at 2500 V. New Brunswick I26R, 120 V/60 Hz shaker incubators (Eppendorf) were used for cell growth. Optical density and UV-vis measurements were collected on a UV-2600 Shimadzu spectrophotometer (Shimadzu). UPLC-MS data were collected on an Acquity UHPLC with an Acquity QDa MS detector (Waters) using an ACQUITY UPLC CSH BEH C18 column (Waters) or an Intrada Amino Acid column (Imtakt).

Cloning, expression, and purification of ObiH

A codon-optimized copy of the ObiH gene was purchased as a gBlock from Integrated DNA Technologies. This DNA fragment was inserted into a pET-28b(+) vector by the Gibson Assembly method.[36] BL21 (DE3) *E. coli* cells were subsequently transformed with the resulting cyclized DNA product via electroporation. After 45 min of recovery in Luria-Burtani (LB) media containing 0.4% glucose at 37 °C, cells were plated onto LB plates with 50 μg/mL kanamycin (Kan) and incubated overnight. Single colonies were used to inoculate 5 mL LB + 50 μg/mL Kan, which were grown overnight at 37 °C, 200 rpm. Expression cultures, typically 1 L of Terrific Broth (TB) + 50 μg/mL Kan (TB-Kan), were inoculated from these starter cultures and shaken (180 rpm) at 37 °C. After 3 hours (OD_{600} = ~0.6), the expression cultures were chilled on ice. After 30 min on ice, ObiH expression was induced with 0.5 mM IPTG, and the cultures were expressed for 16 hours at 20 °C with shaking at 180 rpm. Cells were then harvested by centrifugation at 4,300×*g*

at 4 °C for 30 min. Cell pellets were pink in color and were frozen and stored at -20 °C until purification.

To purify ObiH, cell pellets were thawed on ice and then resuspended in lysis buffer (50 mM potassium phosphate buffer buffer (pH 8.0), 500 mM NaCl, 1 mg/ml Hen Egg White Lysozyme (GoldBio), 0.2 mg/ml DNaseI (GoldBio), 1 mM $MgCl_2$, 1 X BugBuster Protein extraction reagent (Novagen), and 400 μM pyridoxal 5′-phosphate (PLP)). A volume of 4 mL of lysis buffer per gram of wet cell pellet was used. After 45 min of shaking at 37 °C, the resulting lysate was then spun down at 75,600×*g* to pellet cell debris. The pellet was colorless whereas the supernatant was pink in color. Ni/NTA beads (GoldBio) were added to the supernatant and incubated on ice for 45 min prior to purification by Ni-affinity chromatography with a gravity column. The column was washed with 5 column volumes of 20 mM imidazole, 500 mM NaCl, 10% glycerol, 50 mM potassium phosphate buffer (pH 8.0). Washing with higher concentrations of imidazole resulted in slow protein elution. ObiH was eluted with 250 mM imidazole, 500 mM NaCl, 10% glycerol, 50 mM potassium phosphate buffer, pH 8.0. Elution of the desired protein product was monitored by the disappearance of its bright red color (resulting from the release of ObiH) from the column. The protein product was dialyzed to < 1 μM imidazole in 100 mM Tris buffer, pH 8.5 containing 2 mM DTT, dripped into liquid nitrogen to flash freeze, and stored at -80 °C before use. The concentration of protein was determined by Bradford assay. Generally, this procedure yielded 200 – 250 mg per L culture. E107Q and D204N variants were purified similar to that of wt-ObiH.

Preparation of phototreated ObiH

ObiH stock solutions (150 – 400 μM) or diluted samples in quartz cuvettes were placed on ice directly under an 8 Watt, green LED bulb for 10 min. The protein solutions were subsequently kept on ice or in the UV-spectrophotometer for 45 min, followed by a second round of green light treatment for 10 minutes which ensured complete abolishment of the 515 nm band.

Kinetics and UV-Vis Spectroscopy

Data were collected between 600 and 250 nm on a UV-2600 Shimadzu spectrophotometer (Shimadzu) with a semi-micro quartz cuvette (Starna Cells) at 25 °C (unless stated otherwise). ObiH stock solutions were diluted to 20 μM in 100 mM Tris-HCl, pH 8.5 and phototreated.

For the Thr titration, a 500 mM Thr solution was prepared in 100 mM Tris-HCl, pH 8.5 to ensure consistent pH between titration experiments. Time in seconds was recorded from addition of Thr to the time at which data collection at 493 nm occurred. Spectra from 600 – 250 nm were collected every 60 seconds for 1 h.

To monitor the aldol addition of aliphatic aldehyde to E(Q^{Gly}), 20 μM ObiH samples were prepared in 100 mM Tris-HCl, pH 8.5. 100 mM Thr was added and spectra were gathered while the E(Q^{Gly}) species reached a maximum (generally after ~10 min). Propanal was added to a final concentration of 20 mM and spectra collected.

To assess whether glycine can form E(Q^{Gly}), 2.0 M solution of glycine was prepared in 100 mM Tris-HCl, pH 8.5. 20 μM ObiH samples were prepared in 100 mM Tris-HCl, pH 8.5. Glycine was added to a final concentration of 1.0 M. Spectra from 600 – 250 nm were gathered every 1 minute.

$NaBH_4$ reduction experiments were conducted by Mr. Meza with phototreated ObiH. 20 μM samples of ObiH were prepared in 100 mM Tris-HCl, pH 8.5. 100 mM Thr was added and spectra were gathered while the E(Q^{Gly}) species reached a maximum. A fresh solution of 200 mM $NaBH_4$ was prepared in H_2O and added to the reaction mixture to a final concentration of 2 mM 10 minutes after the addition of Thr. Spectra were gathered from 600 – 250 nm every 1 minute. UV-vis experiments for E107Q and D204N variants were performed similar to that of the wild-type protein.

Crystallization and structural studies of ObiH

Crystallization conditions for ObiH were surveyed at both room temperature and 4 °C by the sitting drop method of vapor diffusion using JCSG -plus HT96 (Molecular Dimensions) and HT index HR2-134 (Hampton Research) sparse matrix screens with 17 and 34 mg/mL of ObiH. Crystals appeared in two of the wells that were screened at room temperature containing 0.1 M Bis-Tris pH 6.5, 2.0 M ammonium sulfate and 0.1 M Tris pH 8.5, 2.0 M ammonium sulfate, respectively, at both the concentrations of ObiH. After optimization, 0.1 M Bis-Tris pH 6.75 – 7.0, 1.8 – 2.0 M ammonium sulfate, 10 – 20 mg/mL of ObiH was found to be the best crystallization condition that resulted in reproducible crystal formation. Crystals grew over the course of three days and were pink in color. Crystals were stable in the dark over several months. For X-ray data collection, a cryoprotectant solution containing 30% glycerol, 0.1 M Bis-Tris pH 6.75 – 7.0, 1.8 – 2.0 M ammonium sulfate was added to crystal containing droplets and allowed to incubate for 2 – 30 minutes. After ~10 minutes in the cryoprotectant solution, the crystals slowly solubilized as observed by the loss of sharp edges in the crystal. To our surprise, the crystals retained most of their core structure, and were flash frozen in liquid N_2 until diffraction.

X-ray data were collected remotely at the Advanced Photon Source, Structural Biology Center Beamline 23-ID-D. The X-ray data were processed with xia2 and Aimless. We solved the structure via molecular replacement with PHASER, using the structure of Serine hydroxymethyltransferase [Protein Data Bank (PDB) ID: 4OT8] as a starting model. Unfortunately, ObiH crystals were highly twinned (>25%) and X-ray data from multiple crystals were processed to identify the crystal with least twinning. Iterative cycles of model building with COOT and refinement with REFMAC reduced Rwork and Rfree to 21.3% and 24.2%, respectively, from 50 to 1.66 - Å resolution. The crystals belonged to space group $P4_1$ with the following unit cell dimensions: a = 118.6 Å, b = 118.6 Å, c = 130.0 Å, α = 90°, β = 90° and γ = 90°. The asymmetric unit contained four subunits. Relevant X-ray data collection and model refinement statistics are

listed in Table 1. Crystallization with L-threonine, L-allo-threonine, L-serine, L-glycine or L-penicillamine did not result in high-quality diffraction data.

Design of site-saturation mutagenesis library

PCR was conducted using Phusion polymerase (New England Biolabs) according to the standard protocol. For the given site of mutagenesis, three primers were designed containing codons NDT (encoding for Ile, Asn, Ser, Gly, Asp, Val, Arg, His, Leu, Phe, Tyr, and Cys), VHG (encoding for Met, Thr, Lys, Glu, Ala, Val, Gln, Pro, and Leu), and TGG (Trp), respectively, thereby including all 20 natural amino acids. These three primers were mixed in a ratio 12:9:1 according to the previously described protocol.[37] Library plasmids were constructed by overlap extension PCR using the plasmid that contained the wt-ObiH gene in the pET28(b)+ vector as a template. DpnI was added to digest the template DNA (1 µL per 50 µL PCR reaction, incubated at 37 °C for 1 hour), and the amplified library genes were purified by preparative agarose gel then integrated into a linear pET28(b)+ plasmid via the Gibson Assembly method.[36]

For E107Q and D204N variants, a single primer containing codons CAA and AAT at the site of mutagenesis were used, respectively. Mutant plasmids were constructed by overlap extension PCR using the plasmid that contained the wt-ObiH gene in the pET28(b)+ vector as a template. The resulting amplified mutant genes were purified and integrated as above.

Library expression and screening

BL21 (DE3) *E. coli* cells carrying parent and variant plasmids were grown in 96-deep-well plates (600 µL/well TB-Kan) at 37 °C with shaking at 200 rpm. After shaking at 250 rpm overnight, 10 µL of the overnight cultures were transferred to new deep-well plates containing 600 µL/well TB-Kan, which were allowed to grow at 37 °C with shaking at 200 rpm. After 3 h, the plates were

chilled on ice for 30 min, then induced by the addition of IPTG in TB-Kan (1 mM final concentration). The cultures were expressed for 16 hours at 20 °C with shaking. After 20 hours, cells were harvested by centrifugation at 4,300×g for 20 min. The cell pellets were frozen at –20 °C for a minimum of 2 hours. For screening, cells were thawed at room temperature and then lysed by the addition of 250 µL of lysis buffer (50 mM potassium phosphate buffer buffer pH 8.0, 500 mM NaCl, 1 mg/ml Hen Egg White Lysozyme (GoldBio), 0.2 mg/ml DNaseI (GoldBio), 1 mM MgCl2, 1 X BugBuster Protein extraction reagent (Novagen) and 400 µM PLP). The plates were incubated for 1 h at 37 °C. The resulting lysate was spun down at 4,300×*g* for 20 min to pellet cell debris and supernatants were transferred to a new 96-well deep-well plate.

15 mM of Biphenyl-4-carboxaldehyde (from a 500 mM stock solution in DMSO) and 75 mM of Thr (from a 500 mM stock solution in water) were added to each well, followed by 50 mM potassium phosphate, pH 8.0 buffer such that the total volume in the wells was 250 µL. The plates were sealed with Teflon sealing mats, then incubated at 37 °C for 6 hours. After 6 hours, the reactions were quenched with 250 µL of acetonitrile. Quenched reaction mixtures were filtered using a filter plate to remove debris (by centrifugation at 4,300×*g* for 5 min) prior to UPLC-MS injection. Relative activity and diastereomeric ratio were calculated by comparing the product UV-vis absorbance peak areas at 254 nm (A_{254}) to the A_{254} absorbance from the wt-ObiH wells.

2. 5. References

1. Wang, S. & Deng, H. Peculiarities of promiscuous L-threonine transaldolases for enantioselective synthesis of β-hydroxy-α-amino acids. *Appl. Microbiol. Biotechnol.* **2**, 3507–3520 (2021).

2. Schaffer, J. E., Reck, M. R., Prasad, N. K. & Wencewicz, T. A. β-Lactone formation during product release from a nonribosomal peptide synthetase. *Nat. Chem. Biol.* **13**, 737–744 (2017).

3. Scott, T. A., Heine, D., Qin, Z. & Wilkinson, B. An L-threonine transaldolase is required for L-threo-β-hydroxy-α-amino acid assembly during obafluorin biosynthesis. *Nat. Commun.* **8**, 15935 (2017).

4. Scott, T. A. *et al.* Immunity-Guided Identification of Threonyl-tRNA Synthetase as the Molecular Target of Obafluorin, a β-Lactone Antibiotic. *ACS Chem. Biol.* **14**, 2663–2671 (2019).

5. Contestabile, R. *et al.* L-Threonine aldolase, serine hydroxymethyltransferase and fungal alanine racemase: A subgroup of strictly related enzymes specialized for different functions. *Eur. J. Biochem.* **268**, 6508–6525 (2001).

6. Di Salvo, M. L. *et al.* On the catalytic mechanism and stereospecificity of Escherichia coli l -threonine aldolase. *FEBS J.* **281**, 129–145 (2014).

7. Dunathan, H. C. Conformation and reaction specificity in pyridoxal phosphate enzymes. *Proc. Natl. Acad. Sci. U. S. A.* **55**, 712–6 (1966).

8. Kielkopf, C. L. & Burley, S. K. X-ray structures of threonine aldolase complexes: Structural basis of substrate recognition. *Biochemistry* **41**, 11711–11720 (2002).

9. Steinreiber, J. *et al.* Overcoming thermodynamic and kinetic limitations of aldolase-catalyzed reactions by applying multienzymatic dynamic kinetic asymmetric transformations. *Angew. Chemie - Int. Ed.* **46**, 1624–1626 (2007).

10. Steinreiber, J. *et al.* Synthesis of γ-halogenated and long-chain β-hydroxy-α-amino acids and 2-amino-1,3-diols using threonine aldolases. *Tetrahedron* **63**, 8088–8093 (2007).

11. Fesko, K., Strohmeier, G. A. & Breinbauer, R. Expanding the threonine aldolase toolbox for the asymmetric synthesis of tertiary α-amino acids. *Appl. Microbiol. Biotechnol.* **99**,

9651–9661 (2015).

12. Kimura, T., P. Vassilev, V., Shen, G.-J. & Wong, C.-H. Enzymatic Synthesis of β-Hydroxy-α-amino Acids Based on Recombinant d- and l-Threonine Aldolases. *J. Am. Chem. Soc.* **119**, 11734–11742 (1997).

13. Fesko, K. Threonine aldolases: perspectives in engineering and screening the enzymes with enhanced substrate and stereo specificities. *Applied Microbiology and Biotechnology* **100**, 2579–2590 (2016).

14. Chen, Q. *et al.* Improving and Inverting C $_{\beta}$ -Stereoselectivity of Threonine Aldolase via Substrate-Binding-Guided Mutagenesis and a Stepwise Visual Screening. *ACS Catal.* **9**, 4462–4469 (2019).

15. Murphy, C. D., O'Hagan, D. & Schaffrath, C. Identification of a PLP-dependent threonine transaldolase: A novel enzyme involved in 4-fluorothreonine biosynthesis in Streptomyces cattleya. *Angew. Chemie - Int. Ed.* **40**, 4479–4481 (2001).

16. Wu, L. *et al.* An unusual metal-bound 4-fluorothreonine transaldolase from Streptomyces sp. MA37 catalyses promiscuous transaldol reactions. *Appl. Microbiol. Biotechnol.* **104**, 3885–3896 (2020).

17. Xu, L., Wang, L. C., Xu, X. Q. & Lin, J. Characteristics of l-threonine transaldolase for asymmetric synthesis of β-hydroxy-α-amino acids. *Catal. Sci. Technol.* **9**, 5943–5952 (2019).

18. Kumar, P. *et al.* L -Threonine Transaldolase Activity Is Enabled by a Persistent Catalytic Intermediate. *ACS Chem. Biol.* **16**, 95 (2021).

19. Hill, M. P. *et al.* Light-enhanced catalysis by pyridoxal phosphate-dependent aspartate aminotransferase. *J. Am. Chem. Soc.* **132**, 16953–16961 (2010).

20. Toney, M. D. Reaction specificity in pyridoxal phosphate enzymes. *Archives of Biochemistry and Biophysics* **433**, 279–287 (2005).

21. Buller, A. R. *et al.* Directed evolution mimics allosteric activation by stepwise tuning of the conformational ensemble. *J. Am. Chem. Soc.* **140**, 7256–7266 (2018).

22. Milić, D. *et al.* Crystallographic snapshots of tyrosine phenol-lyase show that substrate strain plays a role in C-C bond cleavage. *J. Am. Chem. Soc.* **133**, 16468–16476 (2011).

23. Masuo, S. *et al.* Enzymatic Cascade in *Pseudomonas* that Produces Pyrazine from α-Amino Acids. *ChemBioChem* **21**, 353–359 (2020).

24. LANE, A. N. & KIRSCHNER, K. The Mechanism of Binding of L-Serine to Tryptophan Synthase from Escherichia coli. *Eur. J. Biochem.* **129**, 561–570 (1983).

25. Kreitler, D. F., Gemmell, E. M., Schaffer, J. E., Wencewicz, T. A. & Gulick, A. M. The structural basis of N-acyl-α-amino-β-lactone formation catalyzed by a nonribosomal peptide synthetase. *Nat. Commun.* **10**, 1–13 (2019).

26. Xu, L., Wang, L. C., Su, B. M., Xu, X. Q. & Lin, J. Multi-enzyme cascade for improving β-hydroxy-α-amino acids production by engineering L-threonine transaldolase and combining acetaldehyde elimination system. *Bioresour. Technol.* **310**, 123439 (2020).

27. Fairman, J.W., Jensen, M.M., Sullivan, A.H., Edwards, T.E., Lorimer, D. PDBID: 4OT8 - X-ray Crystal Structure of Serine Hydroxymethyl Transferase from Burkholderia cenocepacia bound to PLP and Serine. doi:10.2210/pdb4OT8/pdb

28. Eliot, A. C. & Kirsch, J. F. Pyridoxal Phosphate Enzymes: Mechanistic, Structural, and Evolutionary Considerations. *Annu. Rev. Biochem.* **73**, 383–415 (2004).

29. Buller, A. R. *et al.* Directed evolution of the tryptophan synthase β-subunit for stand-alone function recapitulates allosteric activation. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 14599–14604 (2015).

30. Argiriadi, M. A. *et al.* Creation of a S1P Lyase bacterial surrogate for structure-based drug design. *Bioorg. Med. Chem. Lett.* **26**, 2293–2296 (2016).

31. Tan, K., Bigelow, L., Bearden, J., Phillips Jr., G.N., Joachmiak, A. PDBID 5UID -The Crystal Structure of an Aminotranferase TlmJ from Streptoalloteichus.

32. Raj, I., Mazumder, M. & Gourinath, S. Molecular basis of ligand recognition by OASS from E. histolytica: Insights from structural and molecular dynamics simulation studies. *Biochim. Biophys. Acta - Gen. Subj.* **1830**, 4573–4583 (2013).

33. Tu, Y. *et al.* Crystal Structures of Cystathionine β-Synthase from Saccharomyces cerevisiae: One Enzymatic Step at a Time. *Biochemistry* **57**, 3134–3145 (2018).

34. Percudani, R. & Peracchi, A. The B6 database: A tool for the description and classification of vitamin B6-dependent enzymatic activities and of the corresponding protein families.

BMC Bioinformatics **10**, 273 (2009).

35. Angelucci, F. *et al.* The crystal structure of archaeal serine hydroxymethyltransferase reveals idiosyncratic features likely required to withstand high temperatures. *Proteins Struct. Funct. Bioinforma.* **82**, 3437–3449 (2014).

36. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

37. Kille, S. *et al.* Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth. Biol.* **2**, 83–92 (2013).

Chapter 3

Scalable and selective β-hydroxy amino acid synthesis catalyzed by ObiH, a promiscuous L-threonine transaldolase

Content in this chapter is adapted from published work:

Doyon T*, **Kumar P*,** Thein S, Kim M, Stitgen A, Grieger AM, Madigan C, Willoughby PH, Buller AR. "Scalable and selective β-hydroxy-α-amino acid synthesis catalyzed by promiscuous L-threonine transaldolase ObiH". ChemBioChem. **2021**, 22.

*These authors contributed equally to this work

Synthesis of β-hydroxy amino acids and mechanistic studies of ObiH were done in collaboration with Dr. Tyler Doyon. Prof. Patrick H. Willoughby and his undergraduate researchers at Ripon college synthesized many of the aldehyde substrates and performed initial analytical scale reactions with them.

Chapter 3: Scalable and selective β-hydroxy amino acid synthesis catalyzed by ObiH, a promiscuous L-threonine transaldolase

3. 1. Introduction

Enzymes can catalyze chemical transformations that are challenging to achieve using small molecule methods, enabling efficient access to valuable molecules.[1] The three-dimensional architecture of enzyme active sites enables exquisite control over the positioning of reactants, often leading to improved chemo-, site- and stereoselectivity profiles. As a result, a wide variety of new biocatalytic methods have been developed in recent years that seek to resolve selectivity challenges in complex molecule synthesis.[2] Functional group interconversions are the principle class of organic transformations for which high quality biocatalytic routes have been successfully developed.[2] In contrast, there are only a few biocatalytic methods for stereocontrolled C–C bond formation.[3,4] Novel biocatalytic approaches for C–C bond formation are highly desired as they possess unique potential for streamlining the construction of molecular building blocks, natural products, and pharmaceuticals in a sustainable fashion.[3,5]

We sought to develop ObiH as a C-C bond forming biocatalyst outside of its native metabolic context. ObiH is a pyridoxal-5'-phosphate (PLP) dependent L-threonine transaldolase (LTTA) responsible for the biosynthesis of β-hydroxy-*p*-nitrohomophenylalanine (**3**) en route to obafluorin, a β-lactone antibiotic (Figure 1).[6–8] ObiH catalyzes retro-aldol cleavage of L-threonine (Thr) to form a large population of kinetically-trapped PLP quinonoid intermediate, E(Q^{Gly}), in the absence of an aldehyde (See section 2. 2. 1). Mechanistic analysis showed that, upon addition of an aldehyde, E(Q^{Gly}) rapidly reacts with the aldehyde through an aldol-like addition to form a new side chain, setting the stereochemistry of the Cβ-OH group. The ability of ObiH to form a new C-C bond while simultaneously setting the stereochemistry at both the carbon atoms is powerful. In addition, ObiH was reported to have activity with non-native aldehydes to form β-hydroxy amino acids.[7] We also observed reactivity with non-native aldehydes during the characterization of ObiH

(See section 2. 2. 2). This reactivity of ObiH with non-native aldehydes, combined with its high expression titer and stability over months at -80 °C, makes this enzyme a promising biocatalyst.[9]

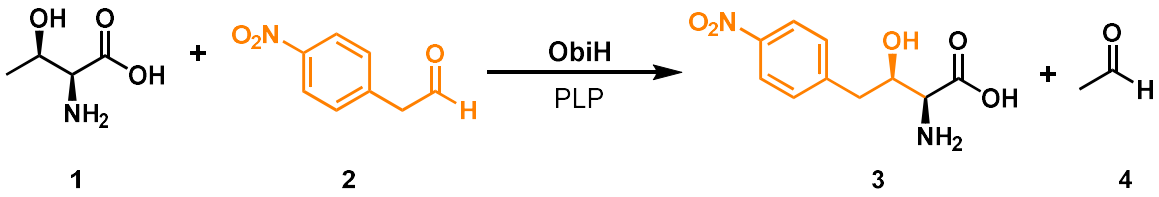


Figure 1. Native ObiH reaction. β-hydroxy-*p*-nitro-homophenylalanine is synthesized from *p*-nitrophenylacetaldehyde (orange) and Thr (black).

β-hydroxy amino acids are a medicinally valuable class of compounds found in several natural products and pharmaceutical drugs (See section 1. 3).[10] Existing enzymatic routes to access β-hydroxy amino acids predominantly use threonine aldolases (TAs). TAs suffers from poor diastereoselectivity profiles for non-native transformations.[11–14] These challenges have been overcome in exceptional cases, either through intensive directed evolution or through the use of diastereoselective crystallization. However, these routes are limited to one or a handful of substrates before selectivity is compromised.[12,15] Consequently, a scalable and generalizable biocatalytic route for the selective synthesis of β-hydroxy amino acids has yet to be realized.

Unlike TAs, LTTAs natively synthesizes β-hydroxy amino acids and operate through a slightly different mechanism.[9] Hence, LTTAs have been investigated as an alternative enzymatic route to access β-hydroxy amino acids.[16] Previous efforts to utilize LTTAs for synthetic purposes focused on optimization of analytical scale reactions with electron-deficient aromatic aldehydes using PsLTTA, an ObiH homolog, as the biocatalyst.[17,18] Protein engineering was also employed to increase reactivity with *p*-methylsulfonylbenzaldehyde (**5**) *en route* to the antibiotic thiamphenicol.[19] Mirroring Nature's approach, preparative scale LTTA transformations with this

substrate were driven to high yield by coupling to a downstream enzymatic reduction of acetaldehyde. This was followed by esterification to obtain ~250 mg of β-hydroxy-*p*-methylsulfonylphenylalanine ethyl ester (**7**) at 45% yield and 97.3% dr. (Figure 2).[17] Apart from this one example, the ability of PsLTTA as a biocatalyst for preparative scale reaction with other aldehyde is unknown. Furthermore, this methodology used three enzymes for synthesizing the β-hydroxy amino acid and required an additional esterification step for isolation which makes this strategy less appealing.

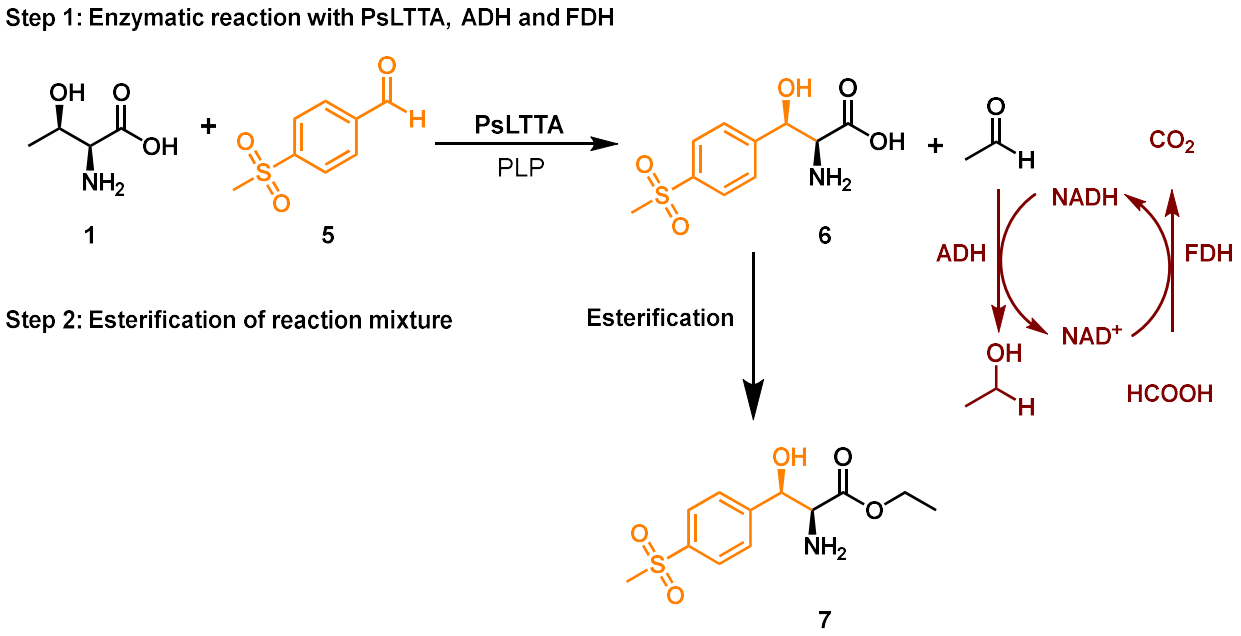


Figure 2. PsLTTA route to access β-hydroxy amino acids. Thr and *p*-methylsulfonylbenzaldehyde (orange) is used as substrate for the PsLTTA reaction. Acetaldehyde elimination system (maroon) enhanced the β-hydroxy amino acid production. The free amino acid was esterified before purification. ADH: Alcohol dehydrogenase, FDH: Formate dehydrogenase. Route shown here developed by Xu et al.[20]

For facile adoption by the synthetic community, an LTTA route to β-hydroxy amino acids must be deployed in the most operationally simple fashion and the limitations of the aldehyde

scope established. It has been previously observed that the diastereomeric ratio (dr) of the products of LTTA reactions decrease through an unknown mechanism, confounding the reproducibility of stereochemical outcomes.[21] Furthermore, biocatalytic methods often yield unprotected, hydrophilic products and demonstration of standardized methods for their isolation and downstream manipulation further increase the attractiveness of an enzymatic route. Here, we tackle these challenges and demonstrate how ObiH be efficiently prepared and utilized as an effective catalyst for the synthesis of structurally diverse β-hydroxy amino acids from inexpensive and readily available starting materials.

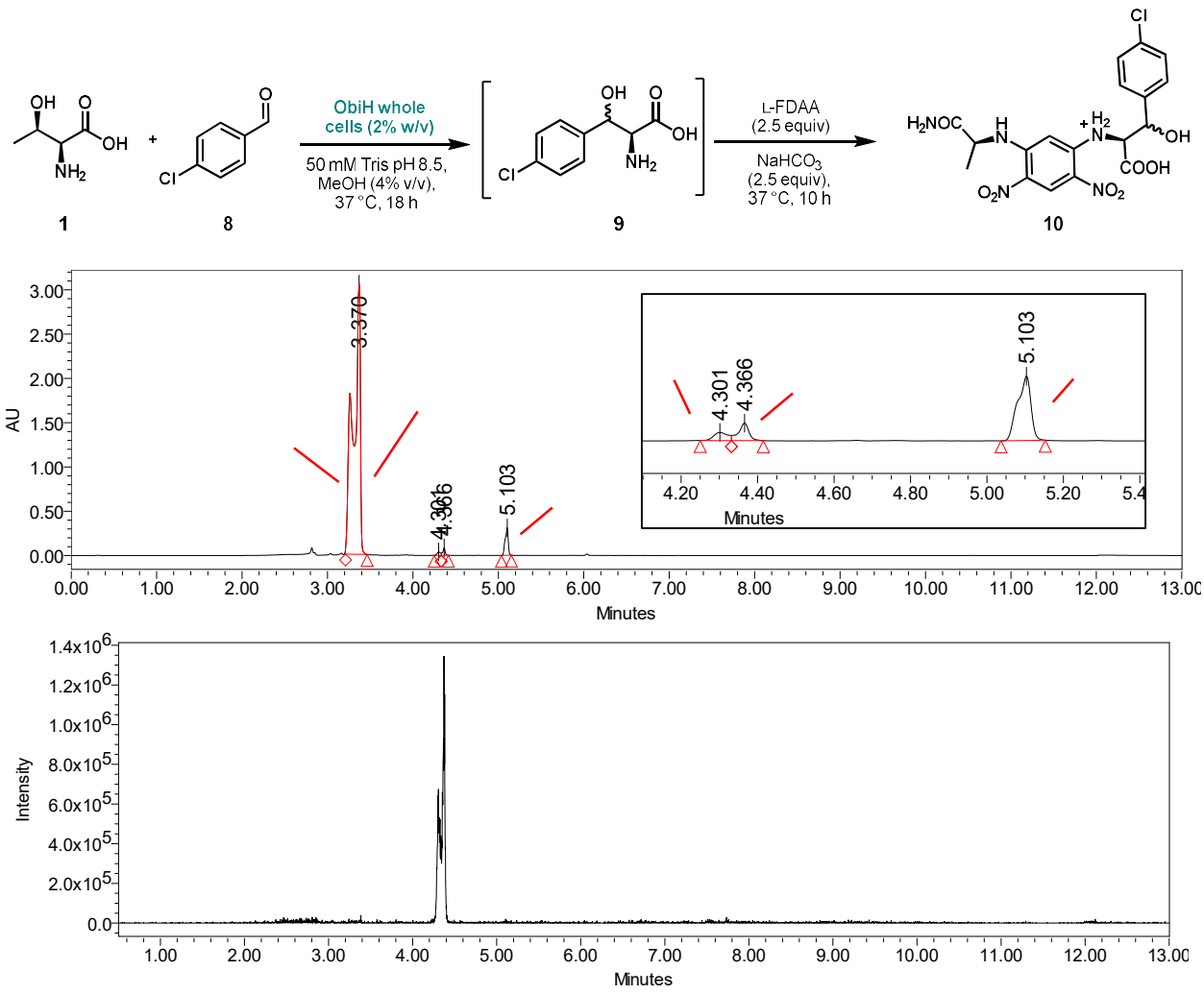
3. 2. Results and Discussion

3. 2. 1. Optimization of ObiH whole cell reactions

We first attempted reactions using wet, whole *E. coli* cells that contained heterologously-expressed *N*-His-ObiH as the biocatalyst. Biocatalytic transformations using whole cells have been demonstrated as an economical and efficient route to access valuable molecules without the need for time and resource-intensive enzyme purification procedures.[22–24] Reaction condition optimization experiments were performed by Dr. Tyler Doyon using *p*-chlorobenzaldehyde (**8**) as a model substrate. Dr. Doyon found that ObiH was catalytically active in the whole cell format. A cell loading of 1−2% wet whole cells and 4% (v/v) MeOH was sufficient to obtain high substrate conversions (>90%) without excessive catalyst and with an easy-to-remove co-solvent. Cell loadings of this magnitude are sufficiently low for process scale reactions and are often only achieved through extensive protein engineering. We attribute this excellent reactivity of ObiH to its high soluble expression in E. coli, boasting >200 mg protein per L culture.

3. 2. 2. Analytical and preparative scale synthesis of β-hydroxy amino acids

With these conditions in hand, we next explored the limits of ObiH-catalyzed reactions by probing the native substrate promiscuity of the catalyst, with the goal to selectively generate and isolate a diverse array of synthetically and pharmaceutically valuable β-hydroxy amino acids. Previous efforts with an ObiH homolog focused on the analytical scale synthesis of phenylserine derivatives, with a single example of product isolation.[17] To gain a more comprehensive view of ObiH reactivity, we chose to perform analytical and preparative scale reactions and compare the relative yields and observed selectivities. Under our standardized reaction conditions, unprotected β-hydroxy amino acids with halogen functionalities were swiftly generated in high yields on analytical scale, but with low to moderate diastereoselectivity as calculated by UPLC-PDA-MS analysis after functionalization with Marfey's reagent (Figure 3, 4). Analysis of analytical scale reactions using Marfey's reagent was performed by Dr. Doyon.



	Retention Time	Area	% Area	Height
1	3.370	16768271	94.31	3056364
2	4.301	100252	0.56	39190
3	4.366	158834	0.89	84523
4	5.103	752201	4.23	310883

Figure 3. Analytical scale ObiH reaction with 4-chlorobenzaldehyde analyzed via Marfey's analysis. UPLC traces at 340 nm and LC/MS traces of ESI M/Z 468.1 are shown above. Area under the UV peaks for the major and the minor product were calculated to determine the diastereomeric ratio of the products. Similar analyses were done for other aldehyde substrates. These data were collected by Dr. Doyon.

In comparison, preparative scale reactions with the same substrates delivered products with enhanced diastereoselectivity, albeit with reduced yields. These data suggest that ObiH is subject to a selectivity phenomenon observed with other LTTAs.[21,25] Namely, an inverse correlation has been noted between reaction yields and selectivity at Cβ, resulting in a precipitous drop in diastereoselectivity when reactions are run to high conversions. While there is a clear tradeoff between conversion and diastereoselectivity for LTTA-catalyzed reactions, we have demonstrated that a balance can be struck to deliver stereo-enriched products with a moderate sacrifice in isolated yields. Notably, the enantioselectivity of these transformations is quite high (>99% *ee*).

Phenylserine analogs generated under these conditions include *m*-bromo-phenylserine (**13**) and gram scale reactions to produce *p*-bromo-phenylserine (**14**) and *p*-chloro-phenylserine (**9**) (Figure 4). To unambiguously assign the absolute configuration of ObiH-generated product, amino acid **9** was recrystallized and characterized by small-molecule X-ray diffraction. This experiment confirmed that the major product of ObiH-catalyzed aldol addition was the (2*S*, 3*R), 'threo*' isomer, which matched the observed stereochemistry of the native product in obafluorin biosynthesis.[26,27]

We considered a reaction with 4-fluoro-3-nitrobenzaldehyde which would yield an intermediate in the recent total synthesis of vancomycin, a peptide antibiotic.[28] However, the highly electrophilic aldehyde starting material condensed with Thr, precluding effective entry into the active site. A recent directed evolution campaign generated TAs that can catalyze a selective aldol reaction with electron deficient benzaldehydes, and we were correspondingly less-motivated to study this reaction further.[14]

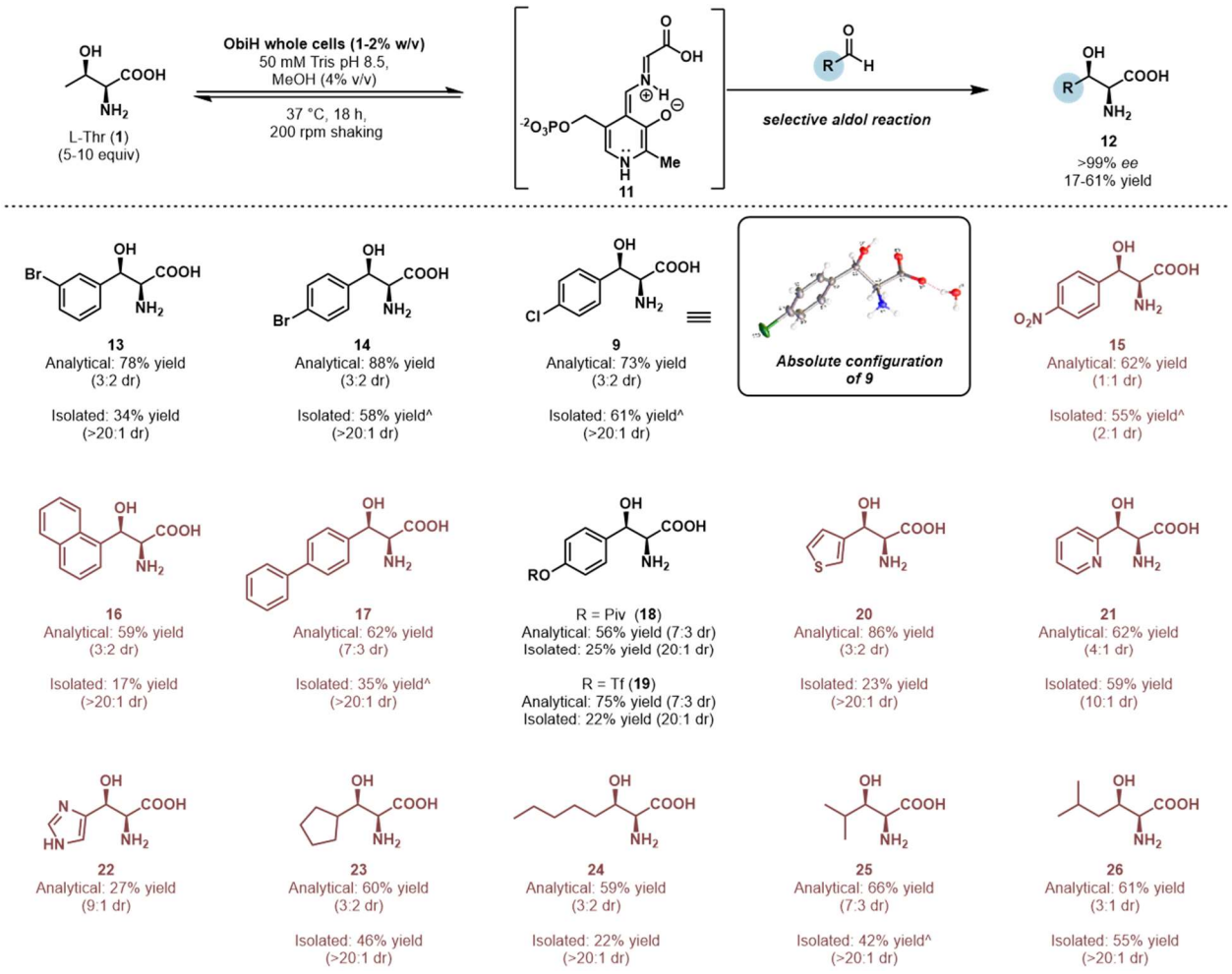


Figure 4. Analytical and preparative-scale synthesis of β-hydroxy amino acids by ObiH. *Products purified by Dr. Doyon are shown in brown.* Reactions were performed using 20 mM aldehyde, 100 mM L-Thr, 50 mM Tris pH 8.5 and 2% ObiH wet whole cells, with 4% (v/v) MeOH as co-solvent. Preparative scale reactions were incubated at 37 °C for 18 h before quench with 1 volume equivalent of MeCN, followed by freeze-thaw and centrifugation to remove cell debris. Purification was achieved using a Biotage purification system via reverse-phase chromatography. Yields are reported as isolated product mass after lyophilization. ^{1}H NMR hydration analysis was used to correct yield values for excess water. Analytical scale product yields determined by UPLC-PDA-MS following derivatization with Marfey's reagent. ^Reaction conducted on gram-scale.

In addition to the variety of phenylserine analogs that were produced using this method, we sought to generate β-hydroxy amino acids which could be used directly in the synthesis of natural products and pharmaceuticals. For example, β-hydroxytyrosine is an nsAA found in several cyclic peptide natural products, including the antibiotic hypeptin.[29] We initially attempted to directly generate β-hydroxytyrosine via an ObiH-catalyzed aldol reaction with 4-hydroxybenzaldehyde, but observed no product formation. This lack of reactivity has been mirrored in other TA systems, but no strategy to access these products has been put forth.[14] We reasoned that, because the pKa of the phenol is low relative to the pH of the reaction, the predominant form in solution is the highly electron rich phenolate. Consequently, this molecule is an intrinsically inert substrate for biocatalytic aldol reactions. We anticipated that protection of the phenolic group on this substrate with an electron-withdrawing substituent would prevent ionization and increase the electrophilicity of the benzaldehyde, enabling a productive catalytic reaction.

This hypothesis was confirmed in reactions with pivaloyl-protected 4-hydroxybenzaldehyde, generating the pivaloyl β-hydroxytyrosine **18** in 25% yield on milligram scale. As was previously observed, diastereoselectivity was improved in comparison to analytical scale conditions, owing to reduced yields in large scale reactions (Figure 4). This protection strategy was also successful with a trifluoromethylsulfonyl protected aldehyde, delivering the triflated β-hydroxytyrosine **19** in 22% yield. Through this simple substrate modification procedure, we generated a pair of aldehydes that could undergo productive catalysis with ObiH, enabling efficient and selective access to protected β-hydroxytyrosine. Furthermore, the protection as a triflate provides a useful functional handle compatible for downstream diversification through cross-coupling reactions.[30]

In parallel, Dr. Doyon purified several bulky, heterocyclic, and aliphatic amino acids at modest yields (25 – 60%) and good diastereomeric ratios (>10:1 dr), and is described in detail in Doyon et al (Figure 4, compounds colored in brown).[31] Several of these amino acids such as 3-

thienylserine (**20**) and pyridine-2-yl-serine (**21**) were previously inaccessible using TAs. The success of ObiH in generating a variety of β-hydroxy amino acids is a significant advance, as biocatalytic access to these motifs has generally been limited to a few examples of TAs with low to moderate diastereoselectivities.

Previously, a two-step multi-enzyme chemoenzymatic route was developed to isolate protected β-hydroxy amino acids using PsLTTA as the biocatalyst. The biocatalytic route reported here also results in similar yields and diastereomeric ratios but uses only one enzyme. In addition, we were able to successfully isolate β-hydroxy amino acids without any protecting groups. These results demonstrate that ObiH-catalyzed aldol reactions offer a streamlined enzymatic route to access diverse β-hydroxy amino acids.

3. 2. 3. Synthesis of a fluorescent β-hydroxy amino acid

Inspired by the success of ObiH in producing a variety of β-hydroxy amino acids, we sought to challenge the enzyme with a large, fluorescent aldehyde derived from the valuable probe molecule BODIPY.[32] An initial whole cell ObiH reaction was performed on milligram scale to generate BODIPY-containing β-hydroxy amino acid **28** (Figure 5a). BODIPY aldehyde **27** underwent productive catalysis with ObiH, generating amino acid **28** in poor yield (~1%). We improved the yields by using purified ObiH, which afforded higher catalyst loadings, and by the addition of 10% DMSO as co-solvent to increase the solubility of aldehyde **27.** These changes improved the yield to 7% with pristine diastereoselectivity (>20:1). While protein engineering would be needed to further increase the reactivity of ObiH with this aldehyde, such efforts are straightforward once initial activity is demonstrated, which we provide here. Fluorescence spectra showed that product **28** exhibits characteristic fluorescence after excitation at 501 nm and emission at 517 nm (Figure 5b).

We anticipate that this amino acid could be a valuable tool in chemical biology applications. This reaction showcases the potential of ObiH to react with a range of aldehydes to generate functional amino acid products. Despite some limitations in its native reactivity, we used ObiH to generate numerous benzylic, heterocyclic and aliphatic β-hydroxy amino acids on preparative scale. Together, these efforts demonstrate the native promiscuity of ObiH toward a variety of inexpensive aldehydes, selectively producing important synthetic precursors in one step. Additionally, this biocatalytic approach enables the use of Thr as an inexpensive substrate *without any pre-functionalization*, which is a considerable advantage over previous synthetic approaches.[33]

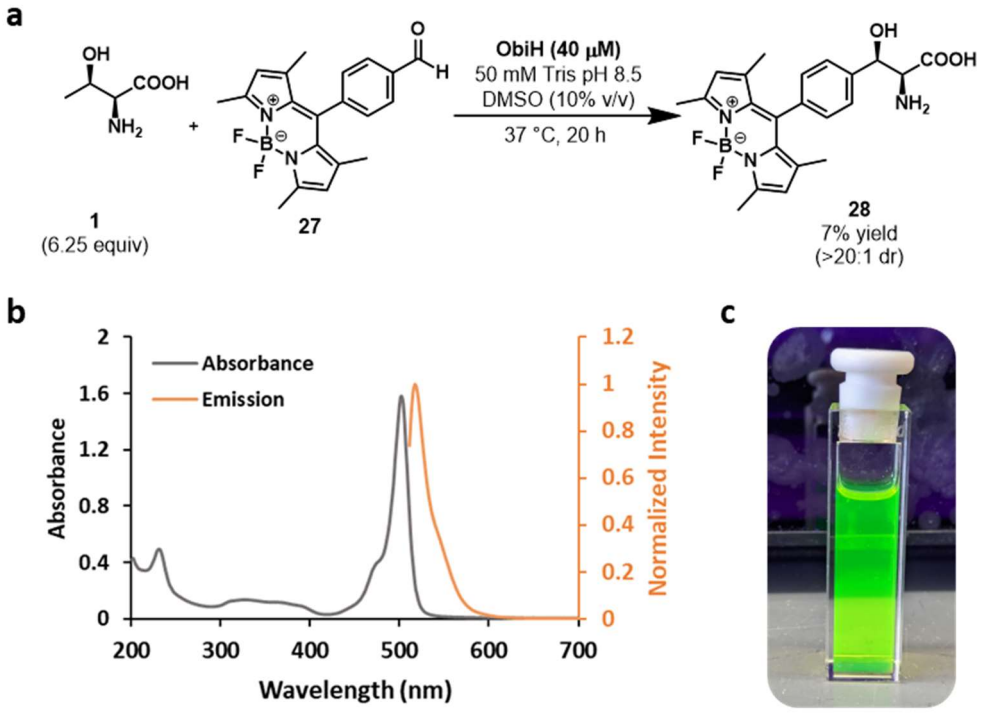


Figure 5. ObiH-catalyzed synthesis of BODIPY-containing β-hydroxy amino acid. **a)** Reactions were performed using 20 mM aldehyde, 125 mM L-Thr, 50 mM Tris pH 8.5 and 40 µM purified ObiH with 10% (v/v) DMSO as co-solvent. Reactions were incubated at 37 °C for 20 h before quenching with 1 volume equivalent of MeCN and centrifugation to remove protein debris. Purification was achieved using a Biotage purification system via reverse-phase chromatography. Yields are reported as isolated product mass after lyophilization. ^{1}H NMR hydration analysis was used to correct yield values for excess water. **b)** Absorbance and emission spectra of 20 µM amino acid **28** dissolved in MeOH. **c)** Photo of 20 µM of amino acid **28** in MeOH under UV light.

3. 2. 4. Product reentry erodes the diastereomeric ratio of the products

Observation of an inverse correlation between yield and diastereoselectivities in these transformations spurred further mechanistic questions about this common phenomenon in LTTAs. To characterize catalyst behavior throughout the reaction, we measured a time course of the ObiH reaction with 4-chlorobenzaldehyde. This reaction yielded two well-resolved diastereomers, enabling direct comparison (via UPLC) of the observed yield of *p*-chloro-phenylserine product **9** and diastereomeric ratios for each timepoint. At early timepoints, the diastereomeric excess was high, indicating that ObiH exhibits superb intrinsic diastereoselectivity. Analysis of later time points showed a notable reduction in the diasteromeric excess of the product **9**, particularly as yields increased beyond 50% (Figure 6). Based on our previous studies of the mechanism of ObiH transaldolase activity,[9] we hypothesized that this drop in the diastereomeric excess could be caused by the reversibility of the transaldolase reaction and product reentry into the catalytic cycle.

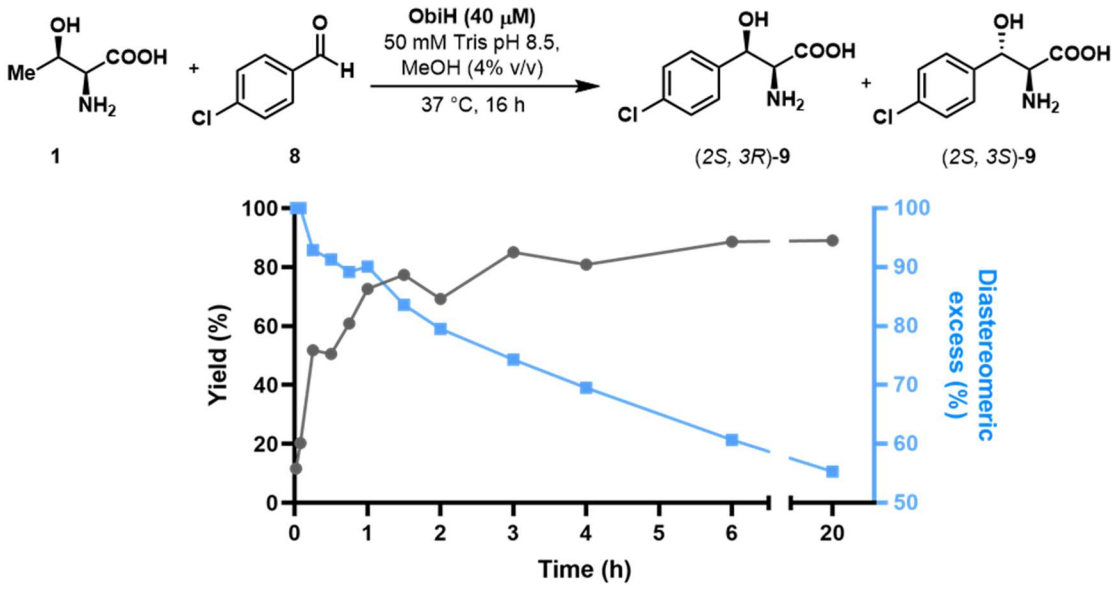


Figure 6. A. Plot of diastereoselectivity of ObiH-catalyzed reaction with 4-chlorobenzaldehyde versus conversion at varying timepoints.

To demonstrate that product can reenter the catalytic cycle and generate glycyl-quinonoid intermediate, Dr. Doyon performed a UV-vis experiment in which ObiH was mixed with *p*-chloro-phenylserine (**9**). In this experiment, following the addition of 5 mM of amino acid **9**, we observed a shift from the internal aldimine:PLP signal (405 nm) to a new peak at 494 nm, corresponding to stabilized glycyl-quinonoid intermediate **11** (Figure 7a).[9] This same intermediate was also observed upon titration of ObiH with native transaldolase substrate Thr, suggesting that the 4-chloro-phenylserine product can indeed reenter the catalytic cycle through retro-aldol cleavage.[9] For direct evidence of product reentry leading to diastereomeric erosion, Dr. Doyon monitored an ObiH reaction with *p*-chloro-phenylserine under turnover conditions in the presence of 4-chlorobenzaldehyde **8** and observed a slow decrease in diastereomeric excess over time, beginning with a *de* value of 82% and ending with a *de* of 64% after 18 h (Figure 7b). These experiments confirm that ObiH product **9** can indeed reenter the catalytic cycle to produce an on-cycle quinonoid intermediate and that this behavior leads directly to reduced diastereomeric excess for ObiH products. These data are consistent with observations that high yielding reactions with ObiH exhibit reduced diastereoselectivity for all tested substrates.

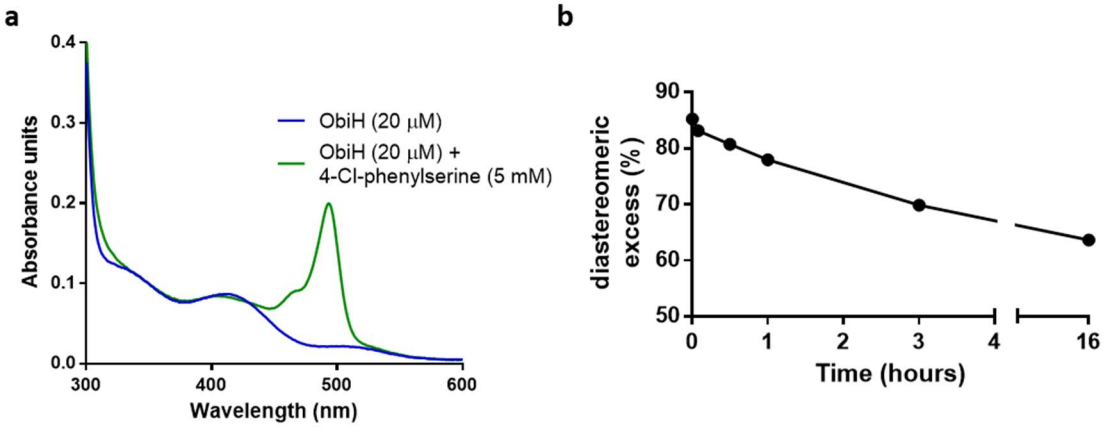


Figure 7. ObiH product reenters the catalytic cycle. a) UV-visible spectrum following titration of ObiH with 4-chloro-phenylserine demonstrating glycyl quinonoid formation (λ_{max} = 494 nm). **b)** ObiH-catalyzed racemization of diastereomerically-enriched 4-chloro-phenylserine. These data were collected by Dr. Doyon.

Based on these observations, we rationalized that product reentry leads to scrambling at the β-position through iterative cycles of retro-aldol product cleavage, followed by the forward aldol reaction with 4-chlorobenzaldehyde. Because the major isomer is at a higher concentration, it will preferentially re-enter the active site and be broken down, only to be re-formed as a mixture of *threo* and *erythro* isomers. In this way, the kinetically disfavored *erythro* isomer slowly accumulates during as the diastereomeric ratio shifts towards an equilibrium ratio. Notably, the stereochemistry at Cα is maintained within the limits of detection during all of the experiments described here.

To probe the structural basis for the diastereoselectivity of ObiH, we turned to the recently described structure at 1.66 Å (PDB ID: 7K34).[9] This structure is of the internal aldimine state of the enzyme and, despite extensive efforts, we were unable to determine a substrate-bound structure under these conditions. We previously turned to molecular dynamics, which provided a plausible structure of ObiH with Thr bound as the external aldimine.[9] Here, we considered how a β-arene would fit into the active site and found that it was a highly constricted environment. Only a single staggered rotamer could be formed without a significant steric clash with protein backbone (Figure 8). While it is possible that some extensive conformational rearrangement occurs upon substrate binding, these areas of contact were stable under the conditions of our previous molecular dynamics simulations. We therefore hypothesize that the bulky substrates described here bind in a single, common orientation extending towards Trp68, which is found on a highly flexible loop (Figure 8). In this model the preferred, on-pathway, mode of binding delivers the *si*-face of the electrophile to the E(Q^{Gly}) nucleophile and aldol addition gives rise to the *threo* product.

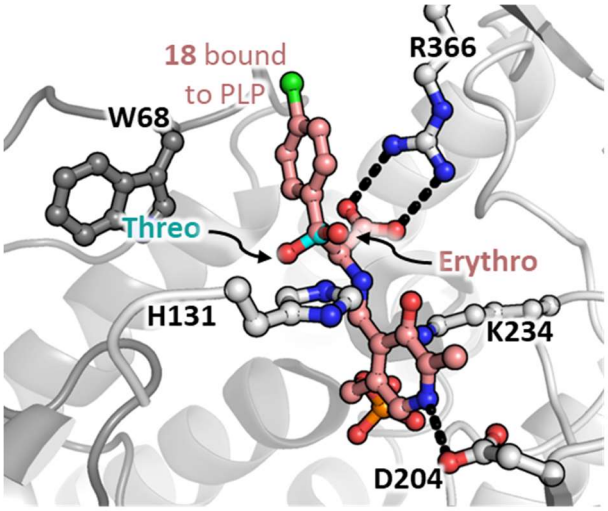


Figure 8. ObiH active site model with 4-chloro-phenylserine bound to PLP as the external aldimine (EAex). The active site is at the dimer interface and individual monomers are colored in light and dark grey. Active site residues are shown as sticks. The *erythro* isomer is colored salmon and the *threo* isomer is colored cyan. Hydrogen bonds are shown as black dashes.

3. 2. 5. Diversification of β-hydroxy amino acids

Finally, we aimed to demonstrate the utility of ObiH derived β-hydroxy amino acids as precursors to aliphatic α-ketoacids (Figure 9). The phenylserine dehydratase from *Ralstonia pickettii* was previously shown to react with a range of phenyl serine analogs.[34] Here we show this scope can be extended beyond aromatic amino acids by developing a one-pot, telescoped sequence for generating 4-methyl-2-oxopentanoic acid (**31**), an α-keto acid directly from Thr and isobuyraldehyde.[35] α-keto acid **31** was formed in 37% yield through this sequence. Such α-keto acid products are highly desirable and have found use as intermediates in biocatalytic cascade reactions, acylating agents in organic synthesis and as precursors for the synthesis of biofuels and pharmaceuticals.[36–38] Through this functionalization reaction, we have demonstrated how the β-hydroxy amino acids could be derivatized to enable access to a library of compounds.

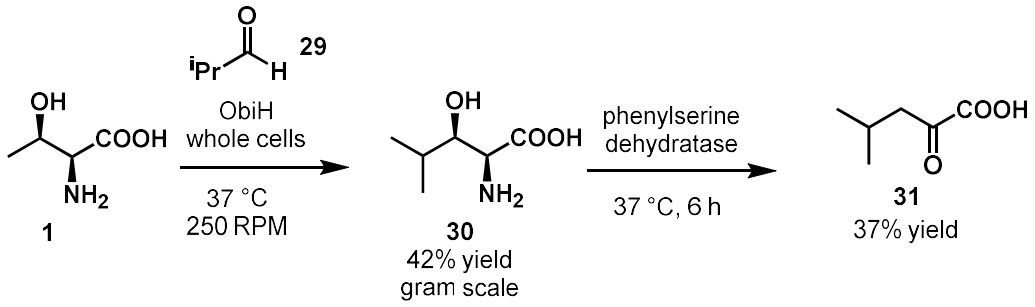
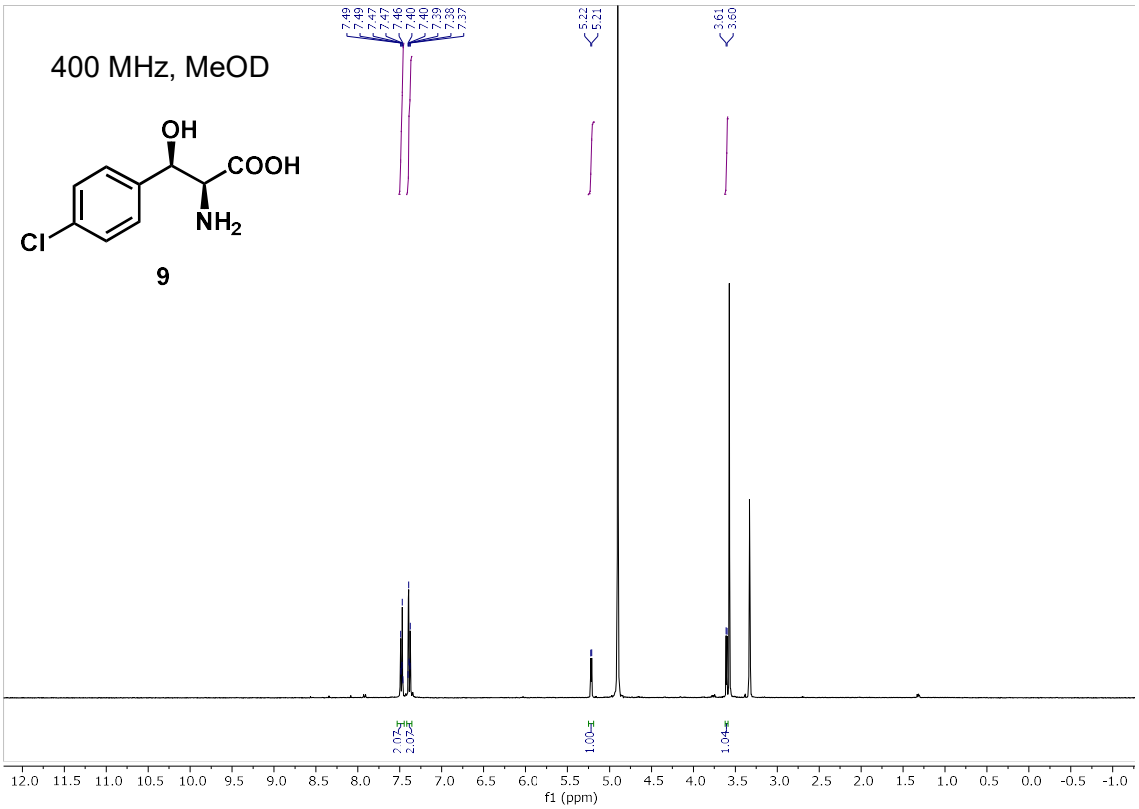


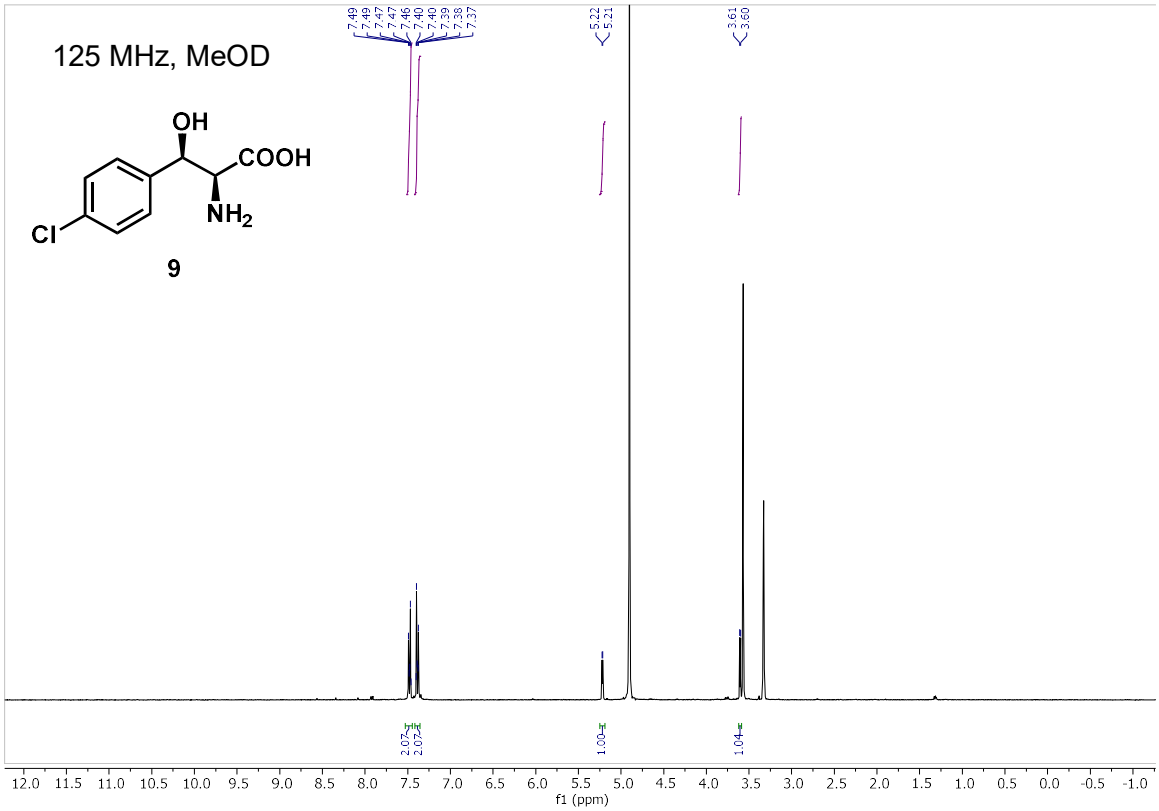
Figure 9. Downstream chemical modification of ObiH-generated β-hydroxyleucine. Analytic scale reactions were performed in triplicates using 20 mM aldehyde, 100 mM L-Thr, 50 mM Tris pH 8.5 and 1% ObiH wet whole cells, with 4% (v/v) MeOH as co-solvent at 37 °C incubator for 20 h. After centrifugation, supernatant was transferred to a new vial and 20 μM phenylserine dehydratase was added and incubated at 37 °C incubator for 6 h. Reaction mixture was quenched with ACN and analyzed through UPLC-MS analysis.

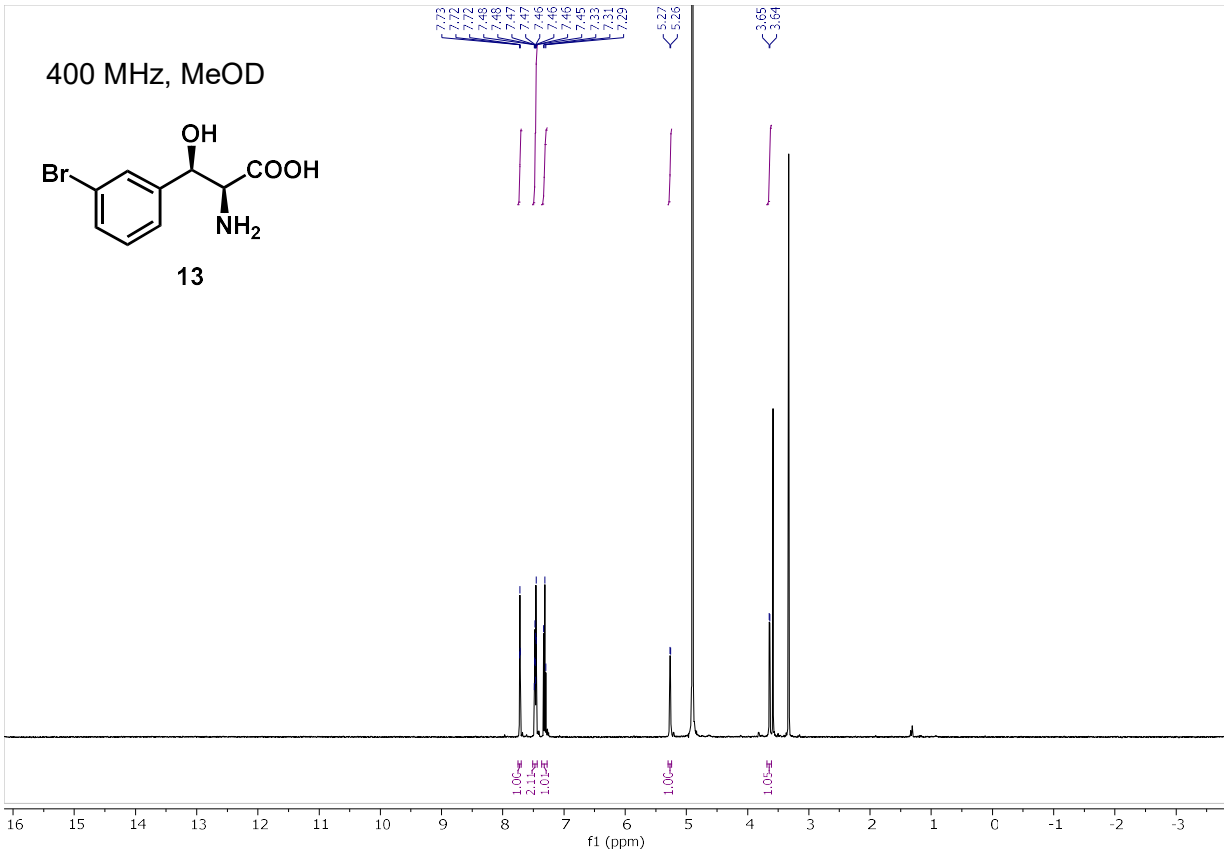
3. 3. Conclusions

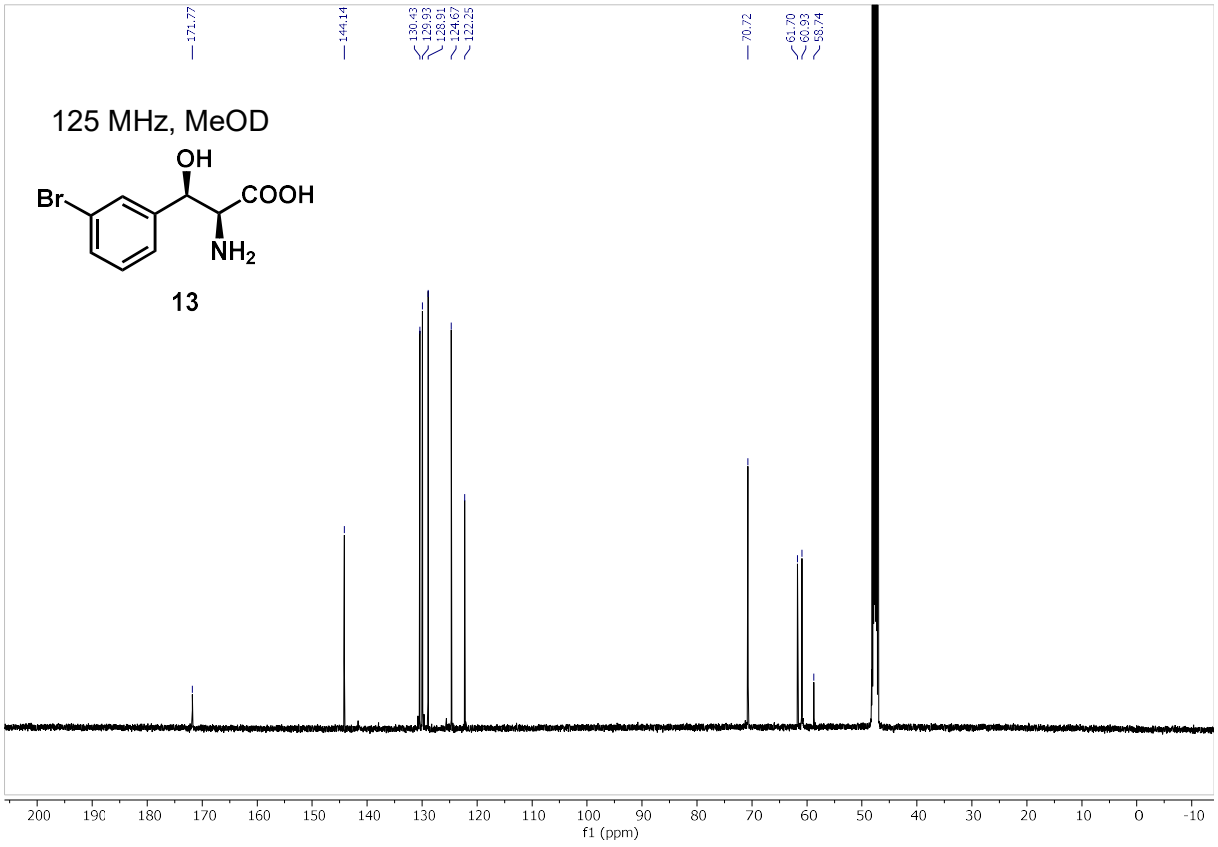
Here we have demonstrated the synthetic utility of ObiH for the diastereoselective production of diverse and valuable β-hydroxy amino acids from inexpensive starting materials. These molecules represent an important starting point for the synthesis of natural products and pharmaceuticals, and can be probes for biological systems. We sought to efficiently generate a variety of these materials through preparative-scale reactions using whole *E. coli* cells, including several transformations on gram scale. We have characterized the native reactivity of ObiH toward a panel of aldehydes, probing the abilities and limitations of ObiH-catalyzed aldolase reactions. This substrate scope analysis was supported by detailed characterization of ObiH diastereoselectivity and underpinning mechanistic implications of observed selectivity trends under analytical and preparative scale conditions. Through these efforts, we have developed an efficient route to accessing important organic building blocks, including downstream functionalization reactions to directly synthesize α-keto acids. Based on the simplicity and versatility of this reaction platform, we advance ObiH as a useful biocatalyst for selective C–C bond formation and we anticipate that this enzyme will serve as a new and effective implement in the organic chemist's toolbox.

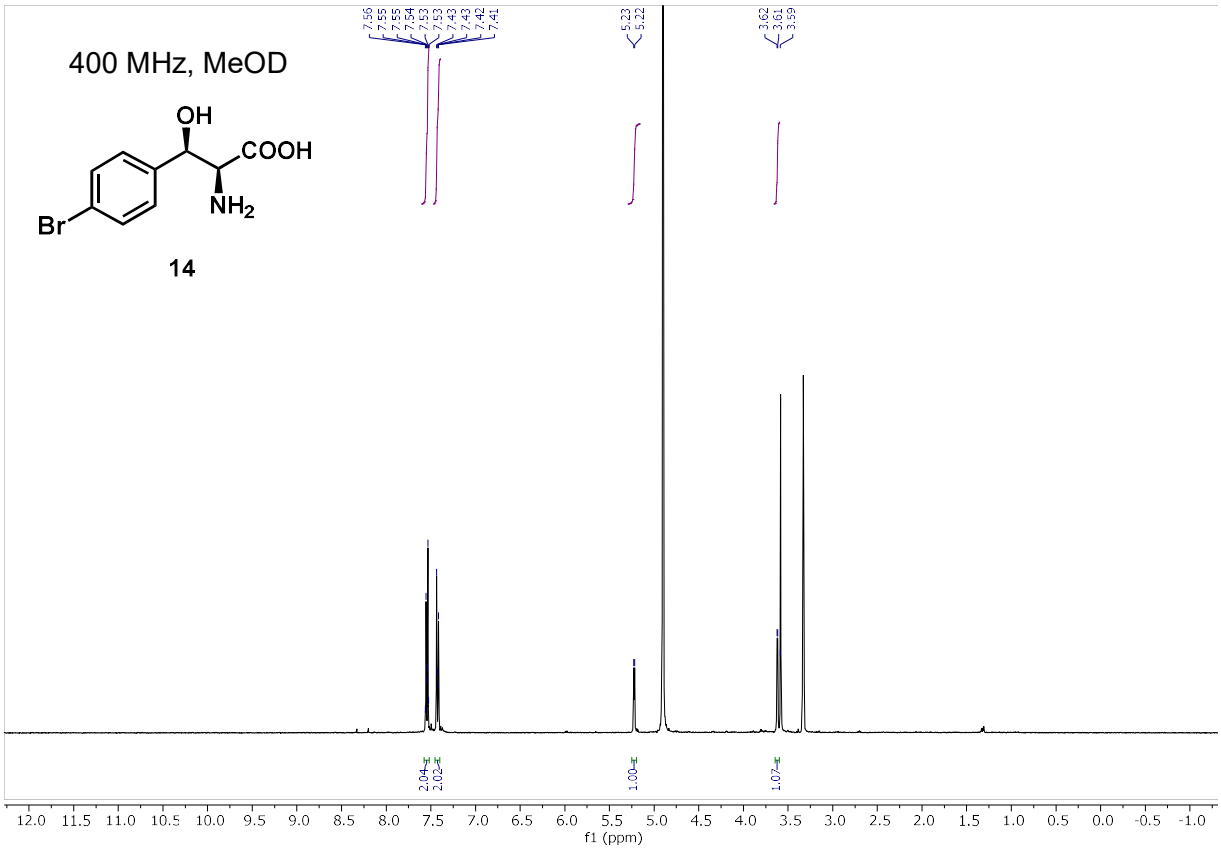
3. 4. NMR spectra of isolated products

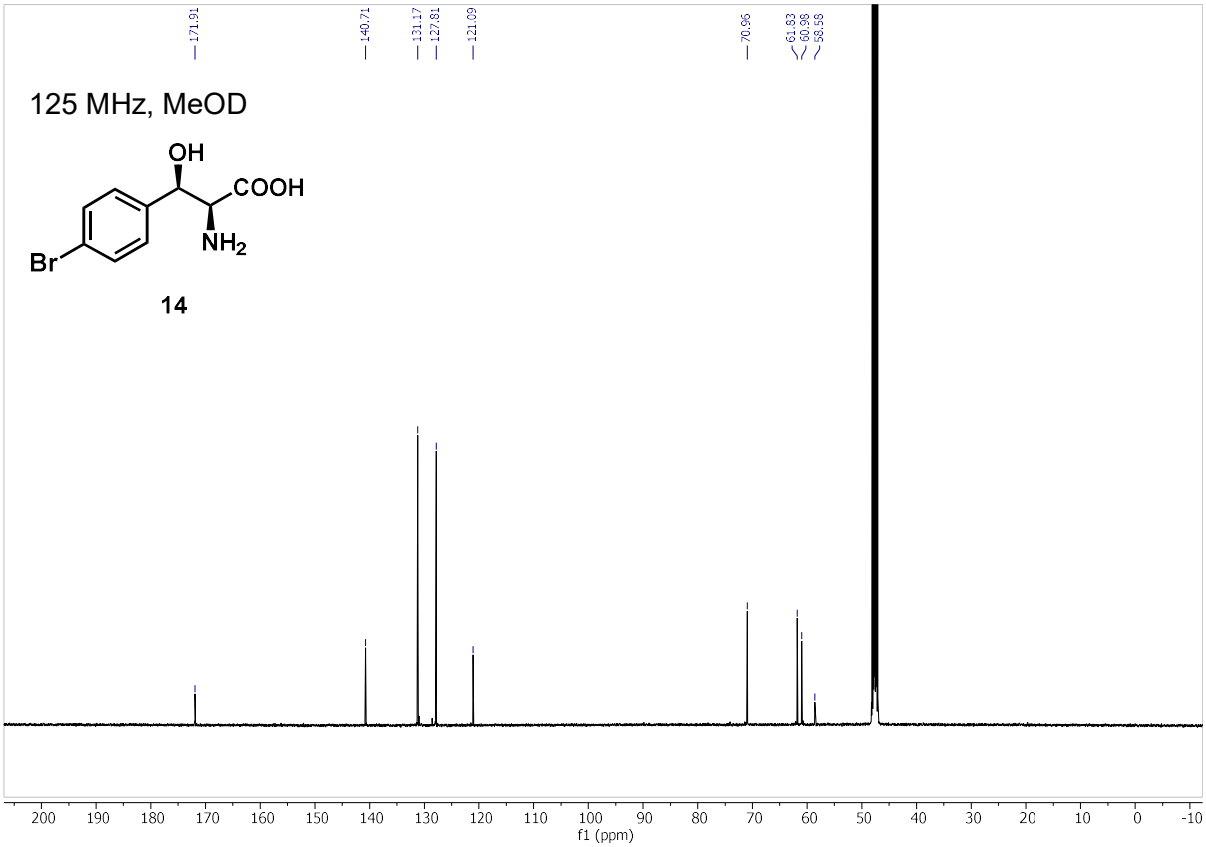


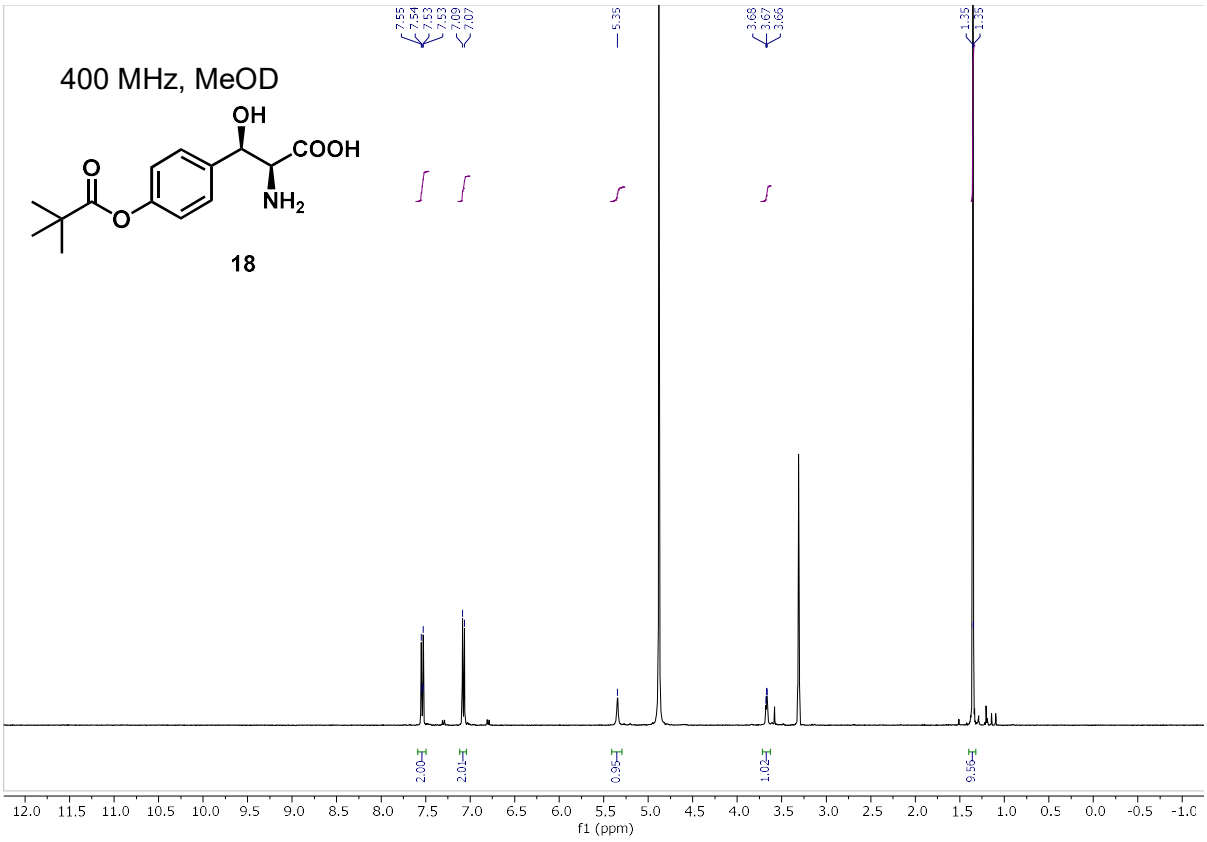


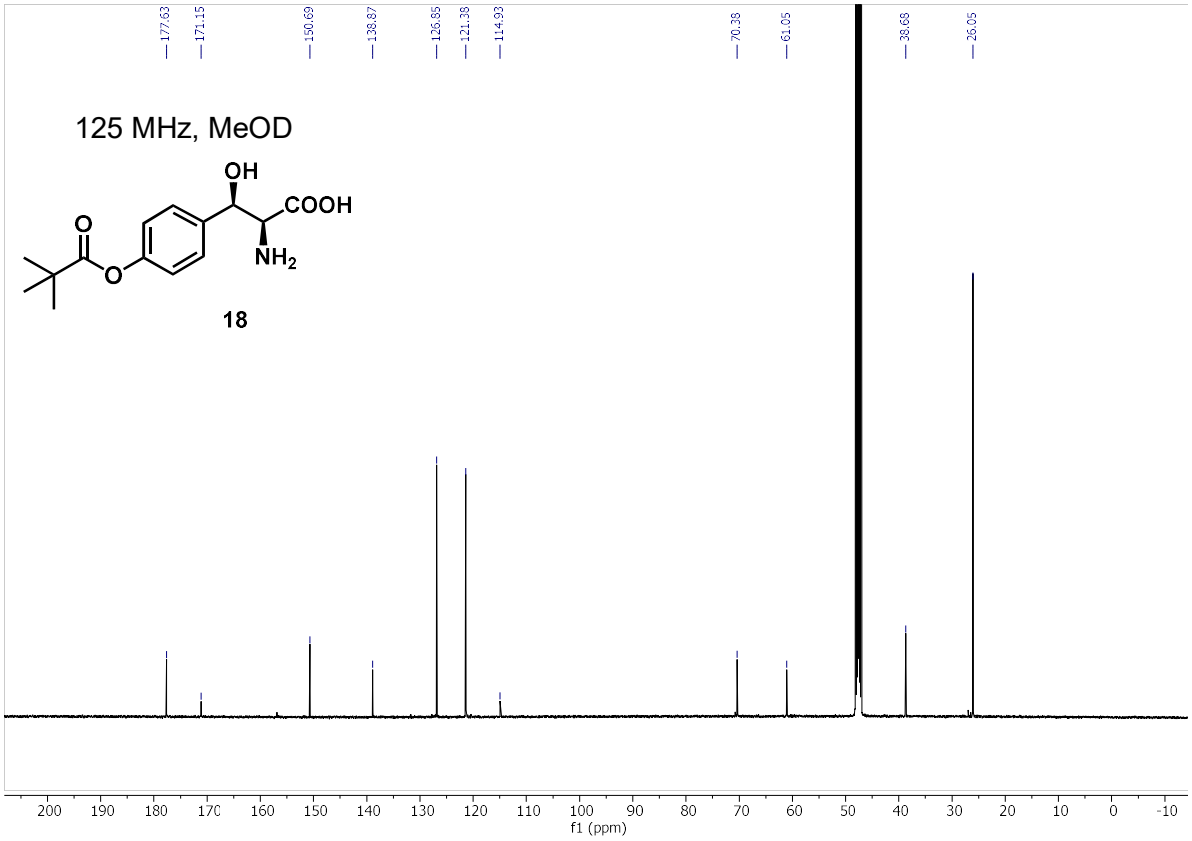

400 MHz, MeOD
OH
Br
COOH
NH2
13
7.73
7.72
7.72
7.48
7.48
7.47
7.47
7.46
7.46
7.46
7.45
7.33
7.31
7.29
5.27
5.26
3.65
3.64
1.00
2.11
1.01
1.00
1.05
f1 (ppm)

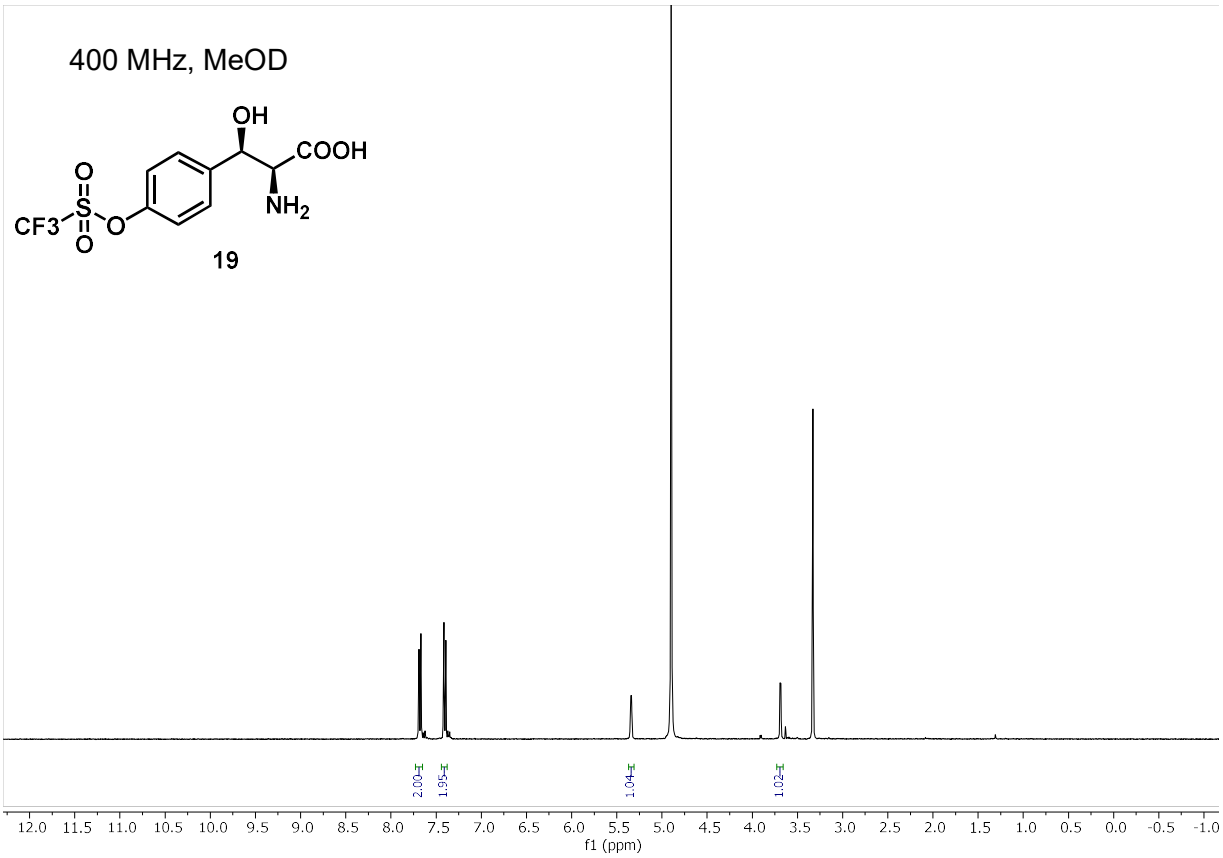

125 MHz, MeOD
OH
Br
COOH
NH2
13
171.77
144.14
130.43
129.93
128.91
124.67
122.25
70.72
61.70
60.93
58.74
f1 (ppm)

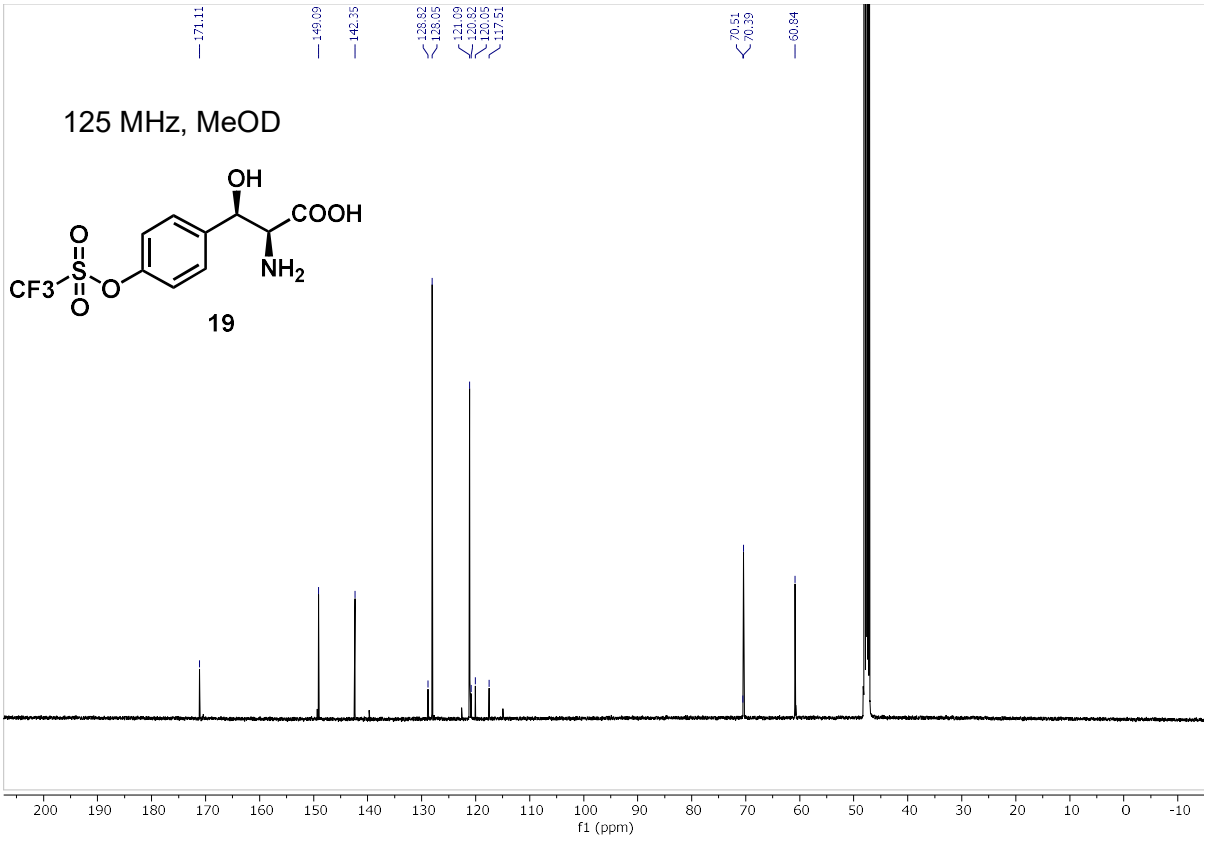

400 MHz, MeOD
OH
COOH
NH2
Br
14
7.56
7.55
7.55
7.54
7.53
7.53
7.43
7.42
7.41
5.23
5.22
3.62
3.61
3.59
2.04
2.02
1.00
1.07
f1 (ppm)

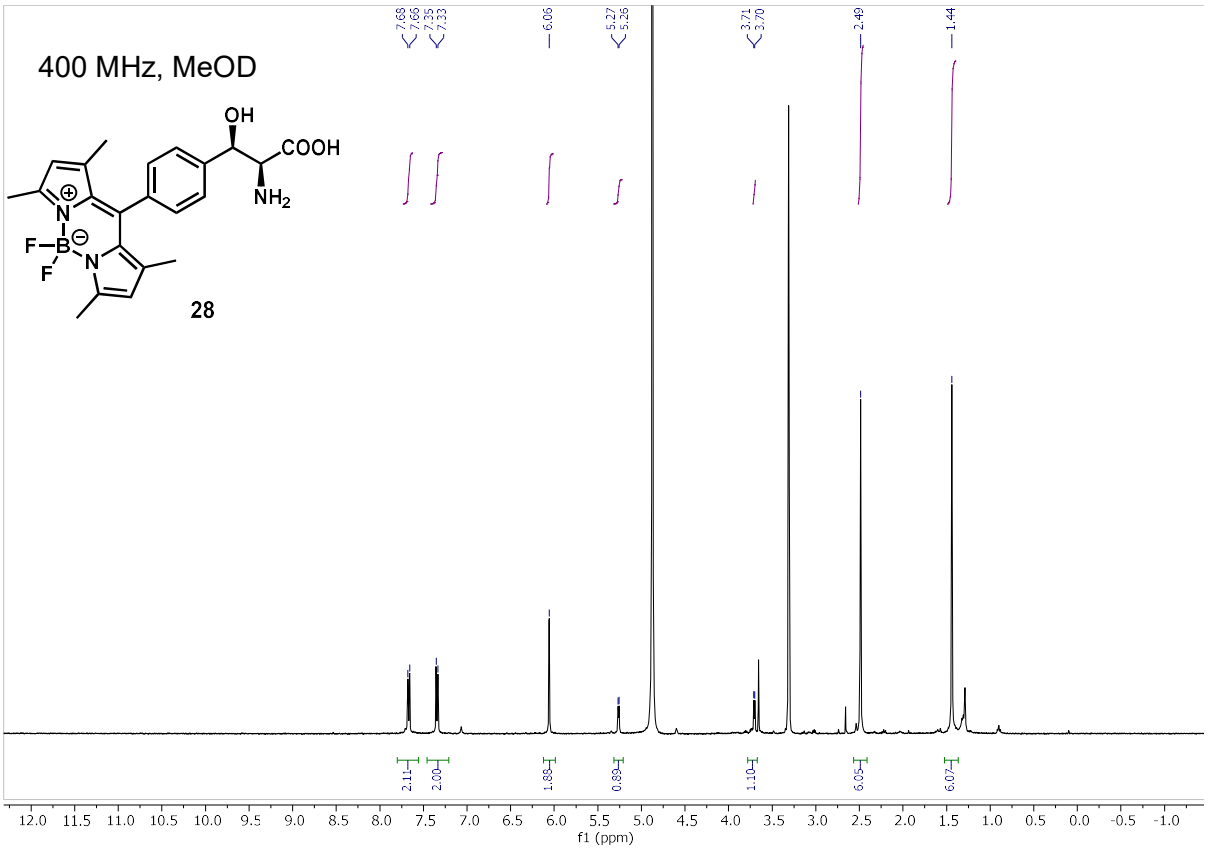

125 MHz, MeOD
OH
COOH
NH2
Br
14
171.91
140.71
131.17
127.81
121.09
70.96
61.83
60.98
58.58
f1 (ppm)

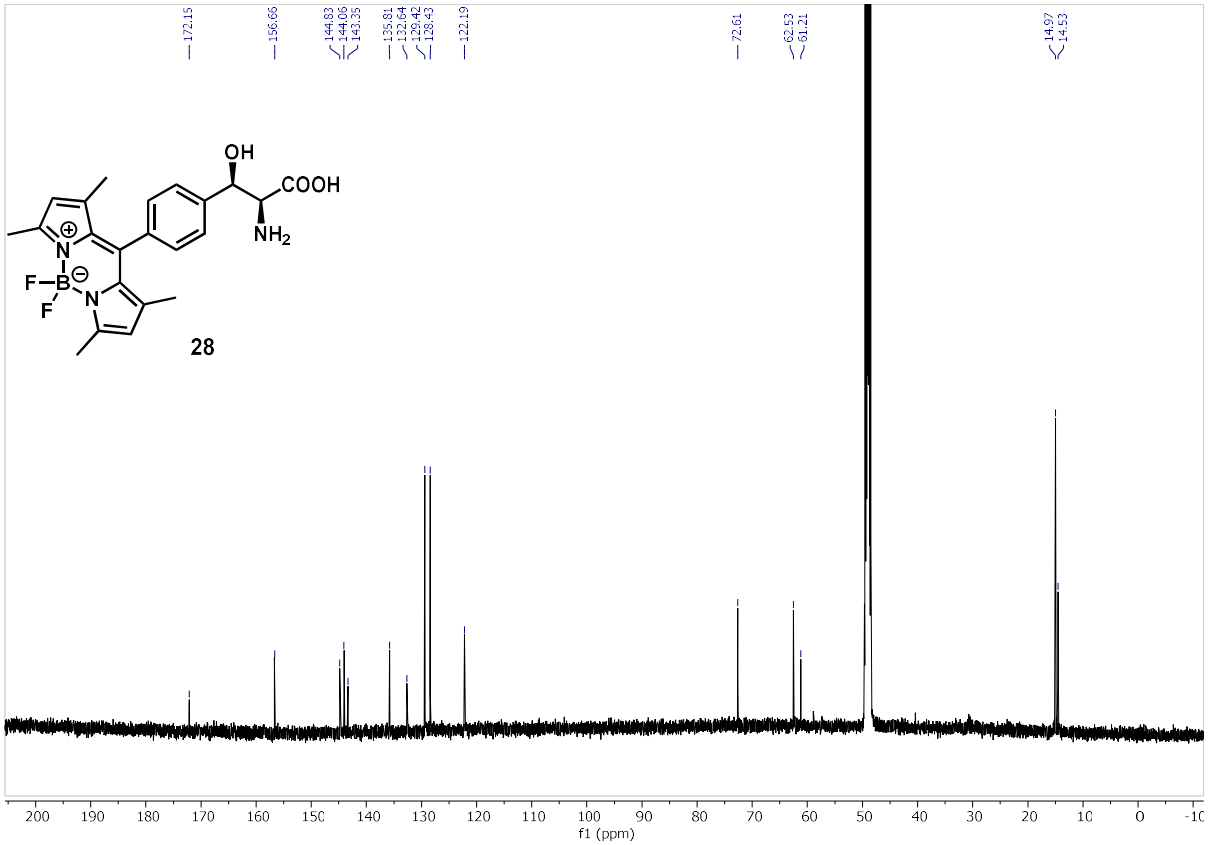

400 MHz, MeOD
OH
COOH
NH2
O
O
18
7.55
7.54
7.53
7.53
7.09
7.07
5.35
3.68
3.67
3.66
1.35
1.35
2.00
2.01
0.95
1.02
9.56
f1 (ppm)

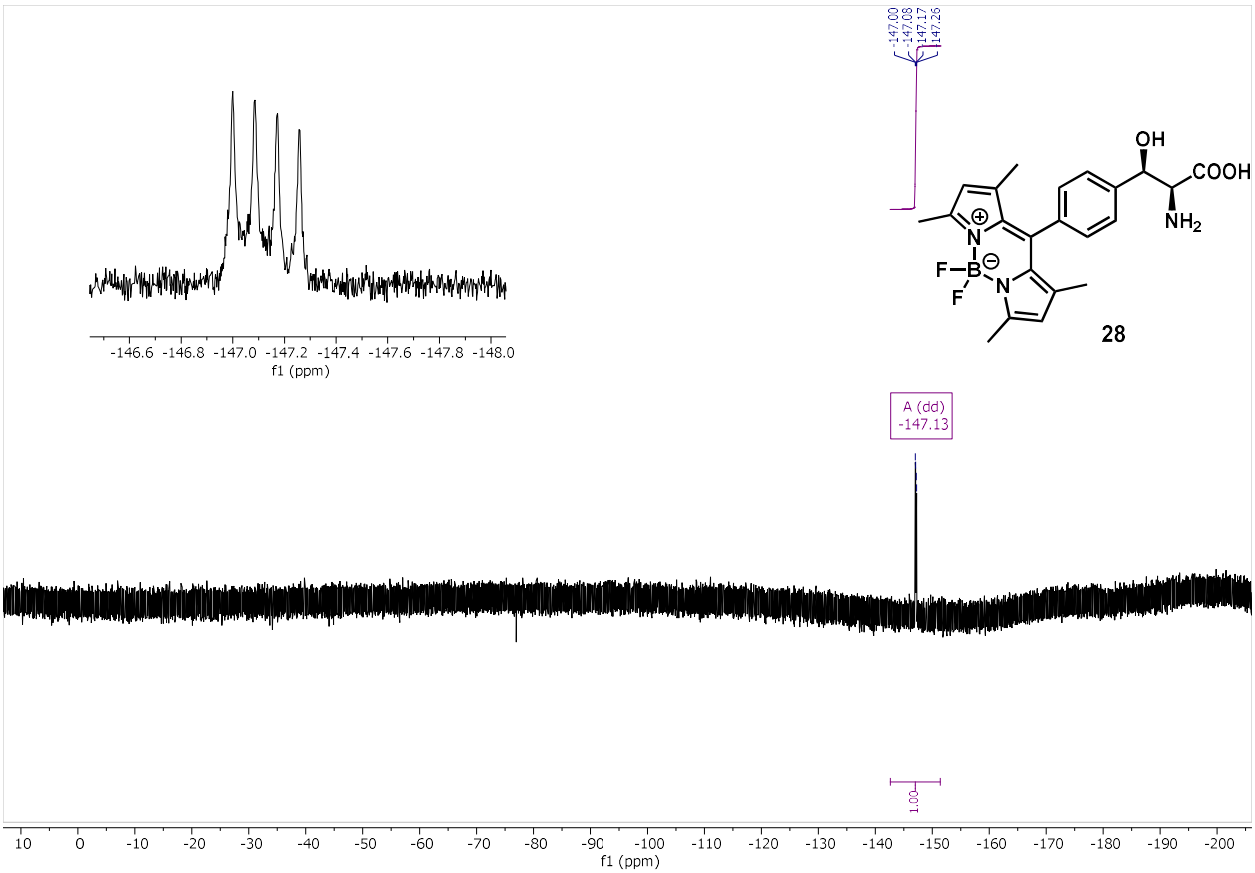

125 MHz, MeOD
OH
COOH
NH2
O
O
18
177.63
171.15
150.69
138.87
126.85
121.38
114.93
70.38
61.05
38.68
26.05
f1 (ppm)


400 MHz, MeOD
OH
COOH
NH2
CF3
O
S
O
O
19
2.00
1.95
1.04
1.02
12.0 11.5 11.0 10.5 10.0 9.5 9.0 8.5 8.0 7.5 7.0 6.5 6.0 5.5 5.0 4.5 4.0 3.5 3.0 2.5 2.0 1.5 1.0 0.5 0.0 -0.5 -1.0
f1 (ppm)


171.11
149.09
142.35
128.82
128.05
121.09
120.82
120.05
117.51
70.51
70.39
60.84
125 MHz, MeOD
OH
COOH
NH2
CF3
O
S
O
O
19
200 190 180 170 160 150 140 130 120 110 100 90 80 70 60 50 40 30 20 10 0 -10
f1 (ppm)


400 MHz, MeOD
OH
COOH
NH2
F
B
N
F
28
7.68
7.66
7.35
7.33
6.06
5.27
5.26
3.71
3.70
2.49
1.44
2.11
2.00
1.88
0.89
1.10
6.05
6.07
f1 (ppm)


172.15
156.66
144.83
144.06
143.35
135.81
132.64
129.42
128.43
122.19
72.61
62.53
61.21
14.97
14.53
OH
COOH
NH2
F
B
N
F
28
f1 (ppm)


-147.00
-147.08
-147.17
-147.26
OH
COOH
NH2
N
F–B
F
N
28
A (dd)
-147.13
1.00
-146.6 -146.8 -147.0 -147.2 -147.4 -147.6 -147.8 -148.0
f1 (ppm)
10 0 -10 -20 -30 -40 -50 -60 -70 -80 -90 -100 -110 -120 -130 -140 -150 -160 -170 -180 -190 -200
f1 (ppm)

3 .5. Materials and Methods

General Information: Chemicals and reagents were purchased from commercial suppliers (Sigma-Aldrich, VWR, Chem-Impex International, Combi-blocks, Alfa Aesar, New England Biolabs, Zymo Research, Bio-Rad) and used without further purification unless otherwise noted. All small molecules were commercially available with exception of 4-OPiv-benzaldehyde and 4-OTf-benzaldehyde, which were previously reported.[39,40] BL21 (DE3) E. coli cells were electroporated with a Bio-Rad MicroPulser electroporator at 2500 V. New Brunswick I26R, 120 V/60 Hz shaker incubators (Eppendorf) were used for cell growth. Optical density and UV-vis measurements were collected on a UV-2600 Shimadzu spectrophotometer (Shimadzu). UPLC-MS data were collected on an Acquity UHPLC with an Acquity QDa MS detector (Waters) using an ACQUITY UPLC CSH BEH C18 column (Waters) or an Intrada Amino Acid column (Imtakt). Preparative flash chromatographic separations were performed on an Isolera One Flash Purification system (Biotage). ^{1}H and ^{13}C NMR spectra were recorded on a Bruker AVANCE III-500 MHz spectrometer equipped with a DCH cryoprobe or a Bruker AVANCE-400 MHz spectrometer equipped with a BBFO probe. ^{1}H chemical shifts are reported in ppm (δ) relative to the solvent resonance (i.e., HOD, δ 4.79 ppm, δ DMSO 2.50 ppm, or δ MeOH 3.31 ppm). ^{13}C NMR data were acquired with ^{1}H decoupling and chemical shifts are reported in ppm (δ) relative to the solvent resonance (δ DMSO 39.52 ppm, δ MeOD 49.00 ppm, δ CD_3CN 1.32 ppm and 118.26 ppm). Data are reported as follows: chemical shift (multiplicity [singlet (s), doublet (d), doublet of doublets (dd), multiplet (m)], coupling constants [Hz], integration). All NMR spectra were recorded at ambient temperature (20–25 °C).

Protein sequence of *N-His-ObiH* (Uniprot accession code: A0A1X9LWZ7):

MGSSHHHHHHSSMSNVKQQTAQIVDWLSSTLGKDHQYREDSLSLTANENYPSALVRLTSGST
AGAFYHCSFPFEVPAGEWHFPEPGHMNAIADQVRDLGKTLIGAQAFDWRPNGGSTAEQALML
AACKPGEGFVHFAHRDGGHFALESLAQKMGIEIFHLPVNPTSLLIDVAKLDEMVRRNPHIRIVILD
QSFKLRWQPLAEIRSVLPDSCTLTYDMSHDGGLIMGGVFDSPLSCGADIVHGNTHKTIPGPQK
GYIGFKSAQHPLLVDTSLWVCPHLQSNCHAEQLPPMWVAFKEMELFGRDYAAQIVSNAKTLAR
HLHELGLDVTGESFGFTQTHQVHFAVGDLQKALDLCVNSLHAGGIRSTNIEIPGKPGVHGIRLG
VQAMTRRGMKEKDFEVVARFIADLYFKKTEPAKVAQQIKEFLQAFPLAPLAYSFDNYLDEELLA
AVYQGAQR

Protein sequence of *N-His-PSDH* (Phenylserine dehydratase; Uniprot accession code: Q10725)***:***

MGSSHHHHHHSSMTQLDTTTLPDLSAIAGLRARLKQWVRTTPVFDKTDFEPVPGTAVNFKLEL
LQASGTFKARGAFSNLLALDDDQRAAGVTCVSAGNHAVGVAYAAMRLGIPAKVVMIKTASPAR
VALCRQYGAEVVLAENGQTAFDTVHRIESEEGRFFVHPFNGYRTVLGTATLGHEWLEQAGALD
AVIVPIGGGGLMAGVSTAVKLLAPQCQVIGVEPEGADAMHRSFETGGPVKMGSMQSIADSLMA
PHTEQYSYELCRRNVDRLVKVSDDELRAAMRLLFDQLKLATEPACATATAALVGGLKAELAGK
RVGVLLCGTNTDAATFARHLGLG

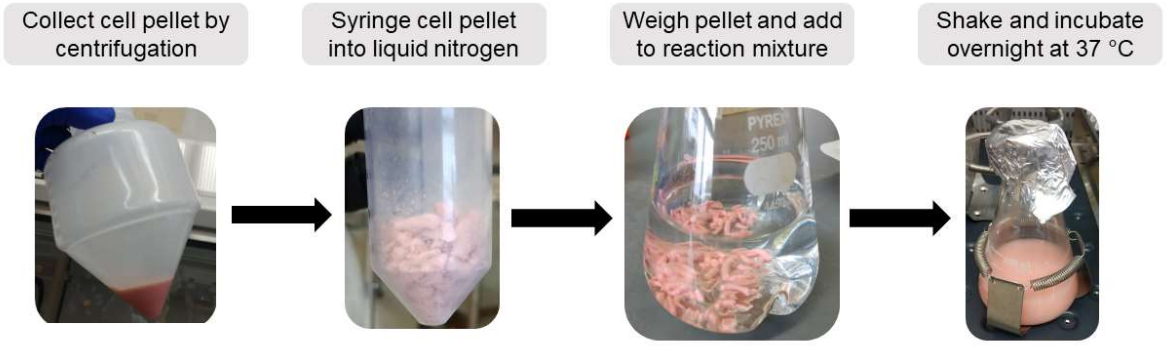
Cloning and expression of ObiH

A codon-optimized copy of the ObiH gene was purchased as a gBlock from Integrated DNA Technologies. This DNA fragment was inserted into a pET-28b(+) vector by the Gibson Assembly method.[41] BL21 (DE3) *E. coli* cells were subsequently transformed with the resulting cyclized DNA product via electroporation. After 45 min of recovery in Luria-Burtani (LB) media containing 0.4% glucose at 37 °C, cells were plated onto LB plates with 50 μg/mL kanamycin (Kan) and incubated overnight. Single colonies were used to inoculate 5 mL LB + 50 μg/mL Kan, which were grown overnight at 37 °C, 200 rpm. Expression cultures, typically 1 L of Terrific Broth (TB) + 50 μg/mL Kan (TB-Kan), were inoculated from these starter cultures and shaken (180 rpm) at 37 °C. After 3 hours (OD_{600} = ~0.6), the expression cultures were chilled on ice. After 30 min on ice, ObiH

expression was induced with 0.5 mM IPTG, and the cultures were expressed for 16 hours at 20 °C with shaking at 180 rpm. Cells were then harvested by centrifugation at 4,300×*g* at 4 °C for 10 min. Cell pellets were pink in color and were used immediately for preparation of freeze-dried cells. Cell pellets were frozen and stored at -20 °C if cells were used for protein purification.

Preparation of freeze-dried ObiH cells

Cell pellets were transferred into a 10 mL syringe and injected into liquid nitrogen to flash freeze cells as thin 'cell noodles' which will facilitate easy weighing of the biocatalyst when setting up reactions. The flash frozen cells were stored at -80 °C.



Photos courtesy of Dr. Tyler Doyon

Purification of ObiH

To purify ObiH, cell pellets were thawed on ice and then resuspended in lysis buffer (50 mM potassium phosphate buffer buffer (pH 8.0), 500 mM NaCl, 1 mg/ml Hen Egg White Lysozyme (GoldBio), 0.2 mg/ml DNaseI (GoldBio), 1 mM $MgCl_2$, 1 X BugBuster Protein extraction reagent (Novagen), and 400 µM pyridoxal 5′-phosphate (PLP)). A volume of 4 mL of lysis buffer per gram of wet cell pellet was used. After 45 min of shaking at 37 °C, the resulting lysate was then spun down at 75,600×*g* to pellet cell debris. The pellet was colorless whereas the supernatant was pink in color. Ni/NTA beads (GoldBio) were added to the supernatant and incubated on ice for 45 min

prior to purification by Ni-affinity chromatography with a gravity column. The column was washed with 5 column volumes of 20 mM imidazole, 500 mM NaCl, 10% glycerol, 50 mM potassium phosphate buffer (pH 8.0). Washing with higher concentrations of imidazole resulted in slow protein elution. ObiH was eluted with 250 mM imidazole, 500 mM NaCl, 10% glycerol, 50 mM potassium phosphate buffer, pH 8.0. Elution of the desired protein product was monitored by the disappearance of its bright red color (resulting from the release of ObiH) from the column. The protein product was dialyzed to < 1 μM imidazole in 100 mM Tris buffer, pH 8.5 containing 2 mM DTT. Purified enzyme was flash frozen in pellet form by pipetting enzyme dropwise into a crystallization dish filled with liquid nitrogen. The enzyme was transferred to a plastic conical and stored at -80 °C until further use. Frozen pellets were thawed at room temperature and centrifuged before use. The concentration of protein was determined by Bradford assay using bovine serum albumin for a standard concentration curve. Generally, this procedure yielded 200 – 250 mg per L culture. Protein purity was analyzed by sodium dodecyl sulfate-polyacrylamide (SDS-PAGE) gel electrophoresis using 12% polyacrylamide gels.

Cloning, expression, and purification of phenylserine dehydratase (PSDH)

A codon-optimized copy of the PSDH gene was purchased as a gBlock from Integrated DNA Technologies. This DNA fragment was inserted into a pET-28b(+) vector by the Gibson Assembly method.[41] BL21 (DE3) *E. coli* cells were subsequently transformed with the resulting cyclized DNA product via electroporation. After 45 min of recovery in Luria-Burtani (LB) media containing 0.4% glucose at 37 °C, cells were plated onto LB plates with 50 μg/mL kanamycin (Kan) and incubated overnight. Single colonies were used to inoculate 5 mL LB + 50 μg/mL Kan, which were grown overnight at 37 °C, 200 rpm. Expression cultures, typically 1 L of Terrific Broth (TB) + 50 μg/mL Kan (TB-Kan), were inoculated from these starter cultures and shaken (180 rpm) at 37 °C. After 3 hours (OD_{600} = ~0.6), the expression cultures were chilled on ice. After 30 min on ice, PSDH

expression was induced with 0.5 mM IPTG, and the cultures were expressed for 16 hours at 20 °C with shaking at 180 rpm. Cells were then harvested by centrifugation at 4,300×*g* at 4 °C for 10 min. Cell pellets were frozen and stored at -20 °C if cells were used for protein purification.

To purify PSDH, cell pellets were thawed on ice and then resuspended in lysis buffer (50 mM potassium phosphate buffer buffer (pH 8.0), 500 mM NaCl, 1 mg/ml Hen Egg White Lysozyme (GoldBio), 0.2 mg/ml DNaseI (GoldBio), 1 mM $MgCl_2$, 1 X BugBuster Protein extraction reagent (Novagen), and 400 μM pyridoxal 5′-phosphate (PLP)). A volume of 4 mL of lysis buffer per gram of wet cell pellet was used. After 45 min of shaking at 37 °C, the resulting lysate was then spun down at 75,600×*g* to pellet cell debris. The pellet was colorless whereas the supernatant was yellow in color. Ni/NTA beads (GoldBio) were added to the supernatant and incubated on ice for 45 min prior to purification by Ni-affinity chromatography with a gravity column. The column was washed with 5 column volumes of 20 mM imidazole, 500 mM NaCl, 10% glycerol, 50 mM potassium phosphate buffer (pH 8.0). PSDH was eluted with 250 mM imidazole, 500 mM NaCl, 10% glycerol, 50 mM potassium phosphate buffer, pH 8.0. Elution of the desired protein product was monitored by the disappearance of its bright yellow color (resulting from the release of PSDH) from the column. The protein product was dialyzed to < 1 μM imidazole in 100 mM Tris buffer, pH 8.5 containing 2 mM DTT. Purified enzyme was flash frozen in pellet form by pipetting enzyme dropwise into a crystallization dish filled with liquid nitrogen. The enzyme was transferred to a plastic conical and stored at -80 °C until further use. Frozen pellets were thawed at room temperature and centrifuged before use. The concentration of protein was determined by Bradford assay using bovine serum albumin for a standard concentration curve. Generally, this procedure yielded 125 – 150 mg per L culture. Protein purity was analyzed by sodium dodecyl sulfate-polyacrylamide (SDS-PAGE) gel electrophoresis using 12% polyacrylamide gels.

Preparation of phototreated ObiH

ObiH stock solutions (150 – 400 μM) or diluted samples in quartz cuvettes were placed on ice directly under an 8 Watt, green LED bulb for 10 min. The protein solutions were subsequently kept on ice or in the UV-spectrophotometer for 45 min, followed by a second round of green light treatment for 10 minutes which ensured complete abolishment of the 515 nm band.

Kinetics and UV-Vis Spectroscopy

Data were collected between 600 and 250 nm on a UV-2600 Shimadzu spectrophotometer (Shimadzu) with a semi-micro quartz cuvette (Starna Cells) at 25 °C (unless stated otherwise). ObiH stock solutions were diluted to 20 μM in 100 mM Tris-HCl, pH 8.5 and phototreated. To monitor product reentry, 20 μM ObiH samples were prepared in 100 mM Tris-HCl, pH 8.5 and phototreated. Stocks of 4-chlorophenylserine were prepared in at 100 mM concentration. Products were added to a final concentration of 5 mM and spectra were gathered after a two-minute incubation period at 30 °C.

Biocatalytic reaction conditions and products

Stock solutions: Stock solutions of each aldehyde (500 mM) were prepared by dissolving the substrate in MeOH or DMSO (analytical grade). 500 mM L-threonine (Thr) stock solutions were prepared in 100 mM Tris pH 8.5. Aliquots of purified ObiH were stored at -80 °C.

Calculation of UPLC yield using Marfey's Derivatization

General Procedure: All reactions were done in duplicate on analytical scale (50 □L). Stocks of Thr were made in 50 mM Tris pH 8.5 and aldehydes were prepared in MeOH. All samples were analyzed following Marfey's derivatization by Waters Acquity UPLC-PDA-MS using a BEH C18 column (Waters). Derivatized amino acid product quantitation was performed by PDA analysis, integrating the area under the product curve and correcting by dividing by the internal standard (tryptamine) peak area. To calculate product concentrations, a standard curve was generated by subjecting stock solutions of L-Phe in buffer using the identical procedure used to process and derivatize enzymatic reaction solutions, in duplicate. These curves were used to calculate the concentrations of β-hydroxy-α-amino acid product in solution.

General Marfey's Procedure: A Marfey's derivatization reaction was performed to assess UPLC yield of all the compounds in the substrate scope. In a microcentrifuge tube, 25 □L of quenched reaction mix (1 equiv., 1.0 mM final total amines from unreacted Thr and formed and β-hydroxy-α-amino acid product) was added to a solution of 125 □L of 15 mM $NaHCO_3$ (10 equiv., 5 mM final concentration) followed by addition of 150 µL 10 mM L-FDAA dissolved in ACN (4 equiv., 5 mM final concentration) to bring the total reaction volume to 300 □L. Each reaction was placed in a dark 37 □C incubator for 12 h, then quenched with 300 µL of 1:1 ACN:60 mM HCl (15 mM post-quench) before analyzing by UPLC-MS.

General procedure for whole cell biocatalytic reactions

A 500 mL glass bottle was charged with 100 mM Tris buffer (pH 8.5) and Methanol (4% v/v). Then, the corresponding aldehyde (1 equiv., 20 mM final concentration) and Thr (5.0 equiv., 100 mM final concentration) were added to the solution. Water was added to adjust the final reagent concentration to the appropriate amounts. The reaction was initiated upon the addition of freeze-dried *E. coli* cells harboring expressed ObiH (10-20 mg/mL, 1-2% w/v). Reaction vessel was placed in the shaking incubator at 37 ºC for 20 h. Product formation was monitored by UPLC-MS. After reaction completion, the reaction mixture was quenched with an equivalent volume of acetonitrile (ACN) and centrifuged (4,300×*g*, 15 min) to remove aggregated protein. Supernatant was transferred to a clean beaker and pellets were resuspended with 1 equivalent volume of ACN, centrifuged and supernatant collected. The decanted supernatant was then concentrated to ~10 mL by rotary evaporation and loaded onto a preparative reverse-phase C18 column pre-equilibrated with water. Purification was performed via gradient elution on an Isolera One Flash Purification system (Biotage). Fractions bearing product (confirmed by UPLC-MS sampling of fraction tubes) were pooled and dried by rotary evaporation. The product was then resuspended in a minimal quantity of water, transferred to a pre-weighed 20 mL vial, frozen, and lyophilized.

Analytical Scale Biocatalytic Reactions

Time course of ObiH reaction

Reactions were done in duplicate on analytical scale (250 µL) for each time point. Stocks of Thr were made in 50 mM Tris pH 8.5 and 4-chlorobenzaldhyde in MeOH. 0.5-dram glass vials were charged with 50.0 µL of 500 mM Thr, 10.0 µL of 500 mM 4-chlorobenzaldhyde, 2.5 µL of 20 mM PLP and 162.9 µL of 50 mM Tris pH:8.5. Reactions were initiated by the addition of 24.6 µL of 400 µM of purified ObiH (Final concentrations: 100 mM Thr, 25 µmol; 20 mM 4-chlorobenzaldhyde, 5 µmol; 40 µM ObiH, 0.2 mol% cat., 500 max TON; 20 µM PLP; 4%

Methanol). Reactions were allowed to proceed in a 37 ºC incubator for 0.02 h, 0.08 h, 0.25 h, 0.5 h, 0.75 h, 1 h, 1.5 h, 2 h, 3 h, 4 h, 6 h and 20 h prior to quenching with 250 µL ACN with 2.5 mM benzophenone as the internal standard. Quenched reactions were then centrifuged at 15,000 xg to remove aggregated protein, and 1:10 dilution of the quenched reaction mixture in water were made. Quantification was performed by UPLC-MS analysis on a BEH C18 column (Waters). Measurement of internal standard and product concentrations was done by measurement of the corresponding 254 nm UV peak areas and using positive mode single ion readout for the M+H mass peak. Variability in injection volumes were corrected by dividing peak areas by the observed internal standard peak area for each injection. To calculate product concentrations, a standard curve was generated by subjecting stock solutions of **9** (4-chlorobenzaldehyde product) (0.05 mM – 1 mM) to UPLC-MS analysis in triplicate, with internal standards. These curves were used to calculate the concentrations of both the *threo* and erythron *isomer of* **9**, and subsequently yields after dilution factor correction. Note: We observed significant stochasticity with respect to the stability of ObiH in solution over time between replicates. Consequently, when ObiH precipitated from solution, the diastereomeric excess ceased to change, resulting in larger apparent errors in results. Nevertheless, repetition always resulted in the same trend: As yields increased, diastereomeric excess decreased.

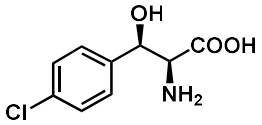
Hydration State NMR Experiments

Two 0.75 mL ampules of DMSO-d_6 were opened and combined in a 1-dram glass vial. Two 700 µL aliquots of appropriate solvent were added to a 0.5-dram glass vial containing product and a blank 0.5 dram glass vial, respectively. After the product dissolved fully, 700 µL of the product and blank solution were transferred to two oven dried NMR tubes. NMR spectra were acquired on either a Bruker Avance-400 or 500 MHz magnet, both equipped with a BBFO probe. T1 measurements were made on representative product containing solutions using a standard

inversion recovery pulse sequence, and subsequent experiment relaxation delays (60 s for DMSO-d_6 samples) were set to ≥5x the maximum measured T1 value to ensure reestablishment of equilibrium magnetism. For hydration state quantification, a standard ^{13}C-decoupled ^{1}H pulse sequence was used. Initially, the product-containing NMR sample was inserted into the NMR spectrometer and analyzed. For the blank sample, the same receiver gain value measured for the product-containing sample was used. The water content contributed by the dissolved product was measured by subtracting the blank spectra water peak integration from the product-containing sample water peak integration and comparing the resulting value to peaks with assigned proton counts from the product.

Biocatalytic reaction products and synthetic procedures

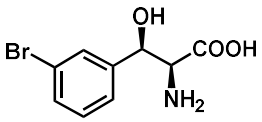
(2*S*,3*R*)- 2-amino-3-(4-chlorophenyl)-3-hydroxypropanoic acid (9)



9

Thr (2977.5 mg, 25.00 mmol, 100 mM final conc.), p-chloro-benzaldehyde (700 mg, 5.0 mmol, 20 mM final conc.), 10.00 mL MeOH (4% v/v), and 240.0 mL 100 mM Tris-HCl, pH 8.5 buffer were added to a 0.5 L glass bottle with a sealable lid. Whole *E. coli* cells containing overexpressed wild-type ObiH were added to a final concentration of 10 mg/mL (1% w/v). The total reaction volume was 250 mL. The reaction mixture was incubated at 37 °C for 18 h while shaking at 250 rpm to ensure thorough mixing. The reaction mixture was then quenched by the addition of 1 volume equivalent of acetonitrile, followed by freeze-thaw to improve material recovery from cells. The thawed mixture was transferred to 50 mL conical vials and centrifuged at 4,300×g for 10 minutes to pellet insoluble whole cells. The supernatant was transferred to a 1.0 L round bottom flask and concentrated by rotary evaporation. For purification, the resulting reaction mixture was loaded onto a Biotage SNAP Ultra 30g C18 column and purified on a Biotage flash purification system using a water/methanol gradient. Fractions were analyzed by UPLC-MS to identify product containing fractions. All product containing fractions from both rounds of purification were then pooled, concentrated by rotary evaporation, and dried via lyophilization, yielding a white solid (680.7 mg isolated). **Hydration analysis:** $C_9H_{10}ClNO_3$•$0.5H_2O$, 61% yield. **^{1}H NMR** (400 MHz, MeOD)) δ 7.51 – 7.45 (m, 2H), 7.42 – 7.36 (m, 2H), 5.22 (d, J = 3.8 Hz, 1H), 3.60 (d, J = 3.8 Hz, 1H); **^{13}C{^{1}H} NMR** (125 MHz, MeOD) δ 172.45, 140.31, 133.03, 130.50, 71.18, 62.09, 62.08, 61.09, 58.27; **HR-ESI-MS**: m/z calcd for $C_{13}H_{13}NO_3$ $[M-H]^-$: 214.0276, found: 214.0276.

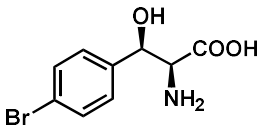
(2*S*,3*R*)- 2-amino-3-(3-bromophenyl)-3-hydroxypropanoic acid (13)



13

Thr (595.6 mg, 5.00 mmol, 100 mM final conc.), m-bromo-benzaldehyde (184 mg, 1.0 mmol, 20 mM final conc.), 2.00 mL MeOH (4% v/v), and 48.0 mL 100 mM Tris-HCl, pH 8.5 buffer were added to a 0.5 L glass bottle with a sealable lid. Whole *E. coli* cells containing overexpressed wild-type ObiH were added to a final concentration of 10 mg/mL (1% w/v). The total reaction volume was 50 mL. The reaction mixture was incubated at 37 °C for 18 h while shaking at 250 rpm to ensure thorough mixing. The reaction mixture was then quenched by the addition of 1 volume equivalent of acetonitrile, followed by freeze-thaw to improve material recovery from cells. The thawed mixture was transferred to 50 mL conical vials and centrifuged at 4,300×g for 10 minutes to pellet insoluble whole cells. The supernatant was transferred to a 1.0 L round bottom flask and concentrated by rotary evaporation. For purification, the resulting reaction mixture was loaded onto a Biotage SNAP Ultra 30g C18 column and purified on a Biotage flash purification system using a water/methanol gradient. Fractions were analyzed by UPLC-MS to identify product containing fractions. All product containing fractions from both rounds of purification were then pooled, concentrated by rotary evaporation, and dried via lyophilization, yielding a white solid (93.6 mg isolated). **Hydration analysis:** $C_9H_{10}BrNO_3 \bullet H_2O$, 34% yield. **^{1}H NMR** (400 MHz, MeOD) δ 7.72 (d, J = 2.0 Hz, 1H), 7.47 (dt, J = 7.7, 2.0 Hz, 2H), 7.31 (t, J = 7.8 Hz, 1H), 5.26 (d, J = 3.4 Hz, 1H), 3.64 (d, J = 3.5 Hz, 1H); **^{13}C{^{1}H} NMR** (125 MHz, MeOD) δ 171.77, 144.14, 130.43, 129.93, 128.91, 124.67, 122.25, 70.72, 61.70, 60.93, 58.74; **HR-ESI-MS**: m/z calcd for $C_{13}H_{13}NO_3$ $[M-H]^-$: 257.9771, found: 257.9771.

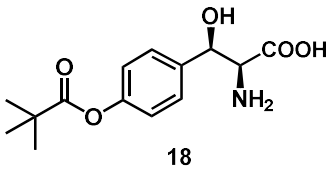
(2*S*,3*R*)- 2-amino-3-(4-bromophenyl)-3-hydroxypropanoic acid (14)



14

Thr (2977.5 mg, 25.00 mmol, 100 mM final conc.), p-bromo-benzaldehyde (920 mg, 5.0 mmol, 20 mM final conc.), 2.00 mL MeOH (4% v/v), and 48.0 mL 100 mM Tris-HCl, pH 8.5 buffer were added to a 0.5 L glass bottle with a sealable lid. Whole *E. coli* cells containing overexpressed wild-type ObiH were added to a final concentration of 10 mg/mL (1% w/v). The total reaction volume was 250 mL. The reaction mixture was incubated at 37 °C for 18 h while shaking at 250 rpm to ensure thorough mixing. The reaction mixture was then quenched by the addition of 1 volume equivalent of acetonitrile, followed by freeze-thaw to improve material recovery from cells. The thawed mixture was transferred to 50 mL conical vials and centrifuged at 4,300×g for 10 minutes to pellet insoluble whole cells. The supernatant was transferred to a 1.0 L round bottom flask and concentrated by rotary evaporation. For purification, the resulting reaction mixture was loaded onto a Biotage SNAP Ultra 30g C18 column and purified on a Biotage flash purification system using a water/methanol gradient. Fractions were analyzed by UPLC-MS to identify product containing fractions. All product containing fractions from both rounds of purification were then pooled, concentrated by rotary evaporation, and dried via lyophilization, yielding a white solid (777.9 mg isolated). **Hydration analysis:** $C_9H_{10}BrNO_3$•$0.5H_2O$, 58% yield. **^{1}H NMR** (400 MHz, MeOD) δ 7.60 – 7.51 (m, 2H), 7.47 – 7.39 (m, 2H), 5.22 (d, J = 3.7 Hz, 1H), 3.62 (d, J = 3.8 Hz, 1H); **^{13}C{^{1}H} NMR** (125 MHz, MeOD) δ 171.91, 140.71, 131.17, 127.81, 121.09, 70.96, 61.83, 60.98, 58.58; **HR-ESI-MS**: m/z calcd for $C_{13}H_{13}NO_3$ $[M-H]^-$: 257.9771, found: 257.9771.

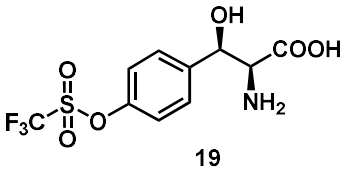
2-amino-3-hydroxy-3-(4-(pivaloyloxy)phenyl)propanoic acid (18)



18

Thr (595.6 mg, 5.00 mmol, 100 mM final conc.), 4-OPiv-benzaldehyde (184 mg, 1.0 mmol, 20 mM final conc.), 2.00 mL MeOH (4% v/v), and 48.0 mL 100 mM Tris-HCl, pH 8.5 buffer were added to a 0.5 L glass bottle with a sealable lid. Whole *E. coli* cells containing overexpressed wild-type ObiH were added to a final concentration of 10 mg/mL (1% w/v). The total reaction volume was 50 mL. The reaction mixture was incubated at 37 °C for 18 h while shaking at 220 rpm to ensure thorough mixing. The reaction mixture was then quenched by the addition of 1 volume equivalent of acetonitrile, followed by freeze-thaw to improve material recovery from cells. The thawed mixture was transferred to 50 mL conical vials and centrifuged at 4,300×g for 10 minutes to pellet insoluble whole cells. The supernatant was transferred to a 1.0 L round bottom flask and concentrated by rotary evaporation. For purification, the resulting reaction mixture was loaded onto a Biotage SNAP Ultra 30g C18 column and purified on a Biotage flash purification system using a water/methanol gradient. Fractions were analyzed by UPLC-MS to identify product containing fractions. All product containing fractions from both rounds of purification were then pooled, concentrated by rotary evaporation, and dried via lyophilization, yielding a white solid (75.8 mg isolated). **Hydration analysis:** $C_{14}H_{19}NO_5.H_2O$, 25% yield. **^{1}H NMR** (400 MHz, MeOD) δ 7.61 – 7.47 (m, 2H), 7.08 (d, J = 8.6 Hz, 2H), 5.35 (s, 0H), 3.73 – 3.62 (m, 1H), 1.35 (s, 9H); **^{13}C{^{1}H} NMR** (125 MHz, MeOD) δ 177.63, 171.15, 150.69, 138.87, 126.85, 121.38, 114.93, 70.38, 61.05, 38.68, 26.05; **HR-ESI-MS**: m/z calcd for $C_{13}H_{13}NO_3$ $[M-H]^-$: 280.1190, found: 280.1190.

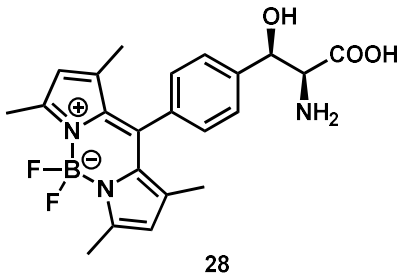
2-amino-3-hydroxy-3-(4-(((trifluoromethyl)sulfonyl)oxy)phenyl)propanoic acid (19)



19

Thr (595.6 mg, 5.00 mmol, 100 mM final conc.), 4-OTf-benzaldehyde (254 mg, 1.0 mmol, 20 mM final conc.), 2.00 mL MeOH (4% v/v), and 48.0 mL 100 mM Tris-HCl, pH 8.5 buffer were added to a 0.5 L glass bottle with a sealable lid. Whole *E. coli* cells containing overexpressed wild-type ObiH were added to a final concentration of 10 mg/mL (1% w/v). The total reaction volume was 50 mL. The reaction mixture was incubated at 37 °C for 18 h while shaking at 250 rpm to ensure thorough mixing. The reaction mixture was then quenched by the addition of 1 volume equivalent of acetonitrile, followed by freeze-thaw to improve material recovery from cells. The thawed mixture was transferred to 50 mL conical vials and centrifuged at 4,300×g for 10 minutes to pellet insoluble whole cells. The supernatant was transferred to a 1.0 L round bottom flask and concentrated by rotary evaporation. For purification, the resulting reaction mixture was loaded onto a Biotage SNAP Ultra 30g C18 column and purified on a Biotage flash purification system using a water/methanol gradient. Fractions were analyzed by UPLC-MS to identify product containing fractions. All product containing fractions from both rounds of purification were then pooled, concentrated by rotary evaporation, and dried via lyophilization, yielding a white solid (75.1 mg isolated). **Hydration analysis:** $C_{10}H_{10}F_3NO_6S.H_2O$, 22% yield. **^{1}H NMR** (400 MHz, MeOD) δ 7.72 – 7.64 (m, 2H), 7.44 – 7.36 (m, 2H), 5.34 (d, J = 3.6 Hz, 1H), 3.69 (d, J = 3.7 Hz, 1H); **^{13}C{^{1}H} NMR** (125 MHz, MeOD) δ 171.11, 149.09, 142.35, 128.82, 128.05, 121.09, 120.82, 120.05, 117.51, 70.51, 70.39, 60.84; **HR-ESI-MS**: m/z calcd for $C_{13}H_{13}NO_3$ $[M-H]^-$: 328.0108, found: 328.0109.

(2S,3R)-2-amino-3-(4-(5,5-difluoro-1,3,7,9-tetramethyl-5H-4l4,5l4-dipyrrolo[1,2-c:2',1'-f][1,3,2]diazaborinin-10-yl)phenyl)-3-hydroxypropanoic acid (28)



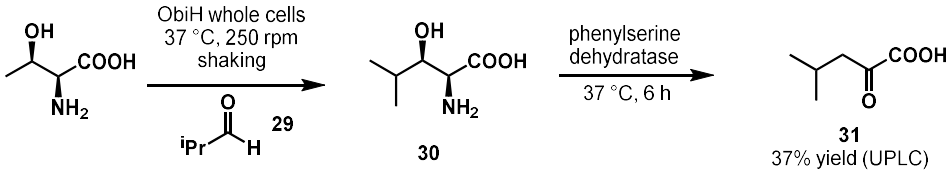
28

Thr (1116.6 mg, 9.37 mmol, 125 mM final conc.), Bodipy-aldehyde (258 mg, 1.5 mmol, 20 mM final conc.), 7.50 mL DMSO (10% v/v), and 59.6 mL 100 mM Tris-HCl, pH 8.5 buffer were added to a 0.5 L glass bottle with a sealable lid. 10 molar equivalents of pyridoxal-5'-phosphate (PLP) relative to final ObiH concentration were then added, followed by addition of ObiH (7.5 mL of 400 μM ObiH, 0.2% mol cat). The total reaction volume was 75 mL. The reaction flask was placed in the dark at 37 °C for 20 h. After reaction completion, the reaction mixture was quenched with an equivalent volume of acetonitrile (ACN) and centrifuged (4,000 rpm, 10 min) to remove aggregated protein. The decanted supernatant was transferred to a 1.0 L round bottom flask and concentrated by rotary evaporation. For purification, the resulting reaction mixture was loaded onto a Biotage SNAP Ultra 30g C18 column and purified on a Biotage flash purification system using a water/methanol gradient. Fractions were analyzed by UPLC-MS to identify product containing fractions. All product containing fractions were pooled, concentrated by rotary evaporation, and dried via lyophilization, yielding a red solid (46.0 mg isolated). **Hydration analysis:** $C_{22}H_{24}BF_2N_3O_3$•$0.25H_2O$, 7% yield. **^{1}H NMR** (400 MHz, MeOD) δ 7.67 (d, J = 7.9 Hz, 2H), 7.34 (d, J = 7.9 Hz, 2H), 6.06 (s, 2H), 5.26 (d, J = 4.7 Hz, 1H), 3.70 (d, J = 4.7 Hz, 1H), 2.49 (s, 6H), 1.44 (s, 6H); **^{13}C{^{1}H} NMR** (125 MHz, MeOD) δ 172.15, 156.66, 144.83, 144.06, 143.35, 135.81, 132.64, 129.42, 128.43, 122.19, 72.61, 62.53, 61.21, 14.97, 14.53; **^{19}F{^{1}H} NMR** (400

MHz, MeOD) δ 147.13 (dd, J = 64.9, 32.2 Hz). **HR-ESI-MS**: m/z calcd for $C_{13}H_{13}NO_3$ $[M-H]^-$: 426.1810, found: 426.1812.

Functionalization of ObiH-generated β-hydroxy amino acid

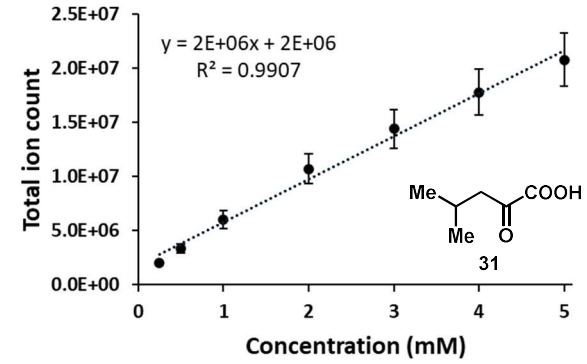
4-methyl-2-oxopentanoic acid (31)

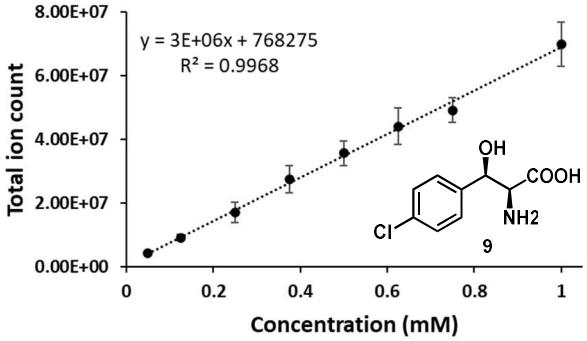


Reactions were done in triplicate on analytical scale (250 µL). Stocks of Thr were made in 50 mM Tris pH 8.5 and Isobutyraldehyde in MeOH. 0.5-dram glass vials were charged with 50.0 µL of 500 mM Thr, 10.0 µL of 500 mM Isobutyraldehyde, 2.5 µL of 20 mM PLP and 153.4 µL of 50 mM Tris pH:8.5. Reactions were initiated by the addition of 25.0 µL of 10% whole *E. coli* cells containing overexpressed wild-type ObiH (Final concentrations: 100 mM Thr, 25 µmol; 20 mM Isobutyraldehyde, 5 µmol; 1% ObiH cells, 200 µM PLP; 4% Methanol). Reactions were allowed to proceed in a 37 ºC incubator for 20 h and then centrifuged to remove cells. The supernatant was transferred to a new 0.5-dram glass vial and 9.1 µL of 550 µM phenylserine dehydratase (PSDH) was added and dehydration reactions were allowed to proceed in a 37 ºC incubator for 6 h (Final concentration: 20 µM PSDH, 0.1 mol% cat., 1000 max TON). Reaction mixture was quenched with 250 µL ACN with 2.5 mM benzophenone as the internal standard. Quenched reactions were then centrifuged at 15,000 xg to remove aggregated protein, amount of **38** were quantified by performing UPLC-MS analysis of the undiluted supernatant on a BEH C18 column (Waters). Measurement of internal standard and product concentrations was done by

measurement of the corresponding 254 nm UV peak areas and using negative mode single ion readout for the M-H mass peak. Variability in injection volumes were corrected by dividing peak areas by the observed internal standard peak area for each injection. To calculate product concentrations, a standard curve was generated by subjecting stock solutions of **38** (1 mM – 5 mM; **38** was purchased from from Sigma-Aldrich) to UPLC-MS analysis similar to enzymatic reaction, in triplicate, with internal standards. These curves were used to calculate the concentrations of 3**8**, and subsequently yields.

Standard curves used for analysis of UPLC-MS data





Small molecule crystallography

Data Collection

A colorless crystal with approximate dimensions 0.30 x 0.20 x 0.20 mm^3 was selected under oil under ambient conditions and attached to the tip of a MiTeGen MicroMount©. The crystal was mounted in a stream of cold nitrogen at 100(1) K and centered in the X-ray beam by using a video camera. The crystal evaluation and data collection were performed on a Bruker D8 VENTURE PhotonIII four-circle diffractometer with Cu Kα (λ = 1.54178 Å) radiation and the detector to crystal distance of 4.0 cm. The initial cell constants were obtained from a 180° φ scan conducted at a 2θ = 50° angle with the exposure time of 1 second per frame. The reflections were successfully indexed by an automated indexing routine built in the APEX3 program. The final cell constants were calculated from a set of 9890 strong reflections from the actual data collection. The data were collected by using the full sphere data collection routine to survey the reciprocal space to the extent of a full sphere to a resolution of 0.80 Å. A total of 25108 data were harvested by collecting 23 sets of frames with 0.5–1.0° scans in ω and φ with an exposure time 0.5–3 sec per frame. These highly redundant datasets were corrected for Lorentz and polarization effects. The absorption correction was based on fitting a function to the empirical transmission surface as sampled by multiple equivalent measurements.[42]

Structure Solution and Refinement

The systematic absences in the diffraction data were uniquely consistent for the space group $P2_12_12_1$ that yielded chemically reasonable and computationally stable results of refinement. A successful solution by the direct methods provided most non-hydrogen atoms from the E-map. The remaining non-hydrogen atoms were located in an alternating series of least-squares cycles and difference Fourier maps. All non-hydrogen atoms were refined with anisotropic displacement coefficients. All hydrogen atoms not participating in hydrogen-bonding interactions were included in the structure factor calculation at idealized positions and were allowed to ride on the

neighboring atoms with relative isotropic displacement coefficients. There is also one molecule of solvent water in the asymmetric unit. **The absolute configuration of the 2-amino-3-(4-chlorophenyl)-3-hydroxypropanoic acid was unequivocally established by anomalous dispersion effects: C2 – *S*, C3 – *R***. The crystal chosen for the experiment proved to be an inversion twin with the minor component contribution of 5(2)%. The final least-squares refinement of 155 parameters against 2298 data resulted in residuals R (based on F^2 for I≥2σ) and wR (based on F2 for all data) of 0.0316 and 0.0898, respectively. The final difference Fourier map was featureless. Data collected and model constructed by Dr. Ilia Guzei.

Summary

Crystal Data for $C_9H_{12}ClNO_4$ (M=233.65 g/mol): orthorhombic, space group $P2_12_12_1$ (no. 19), a = 5.5401(5) Å, b = 6.3588(7) Å, c = 30.308(3) Å, V = 1067.72(19) $Å^3$, Z = 4, T = 100.0 K, μ(CuKα) = 3.164 mm^{-1}, D_{calc} = 1.453 g/cm^3, 25108 reflections measured (5.832° ≤ 2Θ ≤ 158.884°), 2298 unique (R_{int} = 0.0478, R_{sigma} = 0.0206) which were used in all calculations. The final R_1 was 0.0316 (I > 2σ(I)) and wR_2 was 0.0898 (all data).

3. 6. References

1. Lin, C.-I., McCarty, R. M. & Liu, H. The Enzymology of Organic Transformations: A Survey of Name Reactions in Biological Systems. *Angew. Chemie Int. Ed.* **56**, 3446–3489 (2017).

2. Hönig, M., Sondermann, P., Turner, N. J. & Carreira, E. M. Enantioselective Chemo- and Biocatalysis: Partners in Retrosynthesis. *Angewandte Chemie - International Edition* **56**, 8942–8973 (2017).

3. Zetzsche, L. E. & Narayan, A. R. H. Broadening the scope of biocatalytic C–C bond formation. *Nature Reviews Chemistry* **4**, 334–346 (2020).

4. Watkins-Dulaney, E., Straathof, S. & Arnold, F. Tryptophan Synthase: Biocatalyst Extraordinaire. *ChemBioChem* **22**, 5–16 (2021).

5. Huffman, M. A. *et al.* Design of an in vitro biocatalytic cascade for the manufacture of islatravir. *Science (80-.).* **366**, 1255–1259 (2019).

6. Schaffer, J. E., Reck, M. R., Prasad, N. K. & Wencewicz, T. A. β-Lactone formation during product release from a nonribosomal peptide synthetase. *Nat. Chem. Biol.* **13**, 737–744 (2017).

7. Scott, T. A., Heine, D., Qin, Z. & Wilkinson, B. An L-threonine transaldolase is required for L-threo-β-hydroxy-α-amino acid assembly during obafluorin biosynthesis. *Nat. Commun.* **8**, 15935 (2017).

8. Scott, T. A. *et al.* Immunity-Guided Identification of Threonyl-tRNA Synthetase as the Molecular Target of Obafluorin, a β-Lactone Antibiotic. *ACS Chem. Biol.* **14**, 2663–2671 (2019).

9. Kumar, P. *et al.* L -Threonine Transaldolase Activity Is Enabled by a Persistent Catalytic Intermediate. *ACS Chem. Biol.* **16**, 95 (2021).

10. Blaskovich, M. A. T. Unusual Amino Acids in Medicinal Chemistry. *Journal of Medicinal Chemistry* **59**, 10807–10836 (2016).

11. Fesko, K. Threonine aldolases: perspectives in engineering and screening the enzymes with enhanced substrate and stereo specificities. *Applied Microbiology and Biotechnology* **100**, 2579–2590 (2016).

12. Chen, Q. *et al.* Improving and Inverting C β -Stereoselectivity of Threonine Aldolase via Substrate-Binding-Guided Mutagenesis and a Stepwise Visual Screening. *ACS Catal.* **9**, 4462–4469 (2019).

13. Recombinant, D. *et al.* Enzymatic Synthesis of beta-Hydroxy-alfa-amino Acids Based on Recombinant D- and L-Threonine Aldolases. *J. Am. Chem. Soc.* **7863119**, 11734–11742 (1997).

14. Zheng, W. *et al.* Directed Evolution of L-Threonine Aldolase for the Diastereoselective Synthesis of β-Hydroxy-α-amino Acids. *ACS Catal.* **11**, 3198–3205 (2021).

15. L. Goldberg, S. *et al.* Preparation of β-hydroxy-α-amino Acid Using Recombinant d-Threonine Aldolase. *Org. Process Res. & Dev.* **19**, 1308–1316 (2015).

16. Wang, S. & Deng, H. Peculiarities of promiscuous l-threonine transaldolases for enantioselective synthesis of β-hydroxy-α-amino acids. *Appl. Microbiol. Biotechnol.* **105**, 3507–3520 (2021).

17. Xu, L., Wang, L. C., Su, B. M., Xu, X. Q. & Lin, J. Multi-enzyme cascade for improving β-hydroxy-α-amino acids production by engineering L-threonine transaldolase and combining acetaldehyde elimination system. *Bioresour. Technol.* **310**, 123439 (2020).

18. Kreitler, D. F., Gemmell, E. M., Schaffer, J. E., Wencewicz, T. A. & Gulick, A. M. The structural basis of N-acyl-α-amino-β-lactone formation catalyzed by a nonribosomal peptide synthetase. *Nat. Commun.* **10**, 1–13 (2019).

19. Dückers, N., Baer, K., Simon, S., Gröger, H. & Hummel, W. Threonine aldolases-screening, properties and applications in the synthesis of non-proteinogenic β-hydroxy-α-amino acids. *Appl. Microbiol. Biotechnol.* **88**, 409–424 (2010).

20. Xu, L., Wang, L. C., Su, B. M., Xu, X. Q. & Lin, J. Multi-enzyme cascade for improving β-hydroxy-α-amino acids production by engineering L-threonine transaldolase and combining acetaldehyde elimination system. *Bioresour. Technol.* **310**, 123439 (2020).

21. Wang, S. & Deng, H. Peculiarities of promiscuous L-threonine transaldolases for enantioselective synthesis of β-hydroxy-α-amino acids. *Appl. Microbiol. Biotechnol.* **2**, 3507–3520 (2021).

22. Wachtmeister, J. & Rother, D. Recent advances in whole cell biocatalysis techniques

bridging from investigative to industrial scale. *Current Opinion in Biotechnology* **42**, 169–177 (2016).

23. Baker Dockrey, S. A., Doyon, T. J., Perkins, J. C. & Narayan, A. R. H. Whole-cell biocatalysis platform for gram-scale oxidative dearomatization of phenols. *Chem. Biol. Drug Des.* **93**, 1207–1213 (2019).

24. Garzón-Posse, F., Becerra-Figueroa, L., Hernández-Arias, J. & Gamba-Sánchez, D. Whole Cells as Biocatalysts in Organic Transformations. *Molecules* **23**, (2018).

25. Xu, L., Wang, L. C., Xu, X. Q. & Lin, J. Characteristics of l-threonine transaldolase for asymmetric synthesis of β-hydroxy-α-amino acids. *Catal. Sci. Technol.* **9**, 5943–5952 (2019).

26. Schaffer, J. E., Reck, M. R., Prasad, N. K. & Wencewicz, T. A. β-Lactone formation during product release from a nonribosomal peptide synthetase. *Nat. Chem. Biol.* **13**, 737–744 (2017).

27. Scott, T. A., Heine, D., Qin, Z. & Wilkinson, B. An L-threonine transaldolase is required for L-threo-β-hydroxy-α-amino acid assembly during obafluorin biosynthesis. *Nat. Commun.* **8**, 15935 (2017).

28. Moore, M. J. *et al.* Next-Generation Total Synthesis of Vancomycin. *J. Am. Chem. Soc.* **142**, 16039–16050 (2020).

29. Crüsemann, M. *et al.* Biosynthesis and mechanism of action of the cell wall targeting antibiotic hypeptin. *Angew. Chemie Int. Ed.* 1–9 (2021). doi:10.1002/anie.202102224

30. Willemse, T., Schepens, W., Van Vlijmen, H. W. T., Maes, B. U. W. & Ballet, S. The Suzuki-Miyaura cross-coupling as a versatile tool for peptide diversification and cyclization. *Catalysts* **7**, (2017).

31. Doyon, T. J. *et al.* Scalable and Selective β-Hydroxy-α-Amino Acid Synthesis Catalyzed by Promiscuous l-Threonine Transaldolase ObiH. *ChemBioChem* **23**, e202100577 (2022).

32. Lu, H. & Shen, Z. Editorial : BODIPYs and Their Derivatives : The Past , Present and Future. **8**, 1–2 (2020).

33. M. Trost, B. & Miege, F. Development of ProPhenol Ligands for the Diastereo- and Enantioselective Synthesis of β-Hydroxy-α-amino Esters. *J. Am. Chem. Soc.* **136**, 3016–

3019 (2014).

34. H, O., S, N. & H, M. Cloning, sequencing, and overexpression in Escherichia coli of a phenylserine dehydratase gene from Ralstonia pickettii PS22. *Biosci. Biotechnol. Biochem.* **66**, 2755–2758 (2002).

35. Heiwa, O., Shinji, N. & Haruo, M. Cloning, Sequencing, and Overexpression in Escherichiaï¿¿coli of a Phenylserine Dehydratase Gene from Ralstoniaï¿¿pickettii PS22. *Biotechnol. Biochem.* **66**, 2755–2758 (2002).

36. Xue, Y. P., Cao, C. H. & Zheng, Y. G. Enzymatic asymmetric synthesis of chiral amino acids. *Chem. Soc. Rev.* **47**, 1516–1561 (2018).

37. Penteado, F. *et al.* α-Keto Acids: Acylating Agents in Organic Synthesis. *Chem. Rev.* **119**, 7113–7278 (2019).

38. Atsumi, S., Hanai, T. & Liao, J. C. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. *Nature* **451**, 86–89 (2008).

39. Stauffer, C. S., Bhaket, P., Fothergill, A. W., Rinaldi, M. G. & Datta, A. Total synthesis and antifungal activity of a carbohydrate ring-expanded pyranosyl nucleoside analogue of nikkomycin B. *J. Org. Chem.* **72**, 9991–9997 (2007).

40. Zhao, Y. & Snieckus, V. A practical in situ generation of the schwartz reagent. reduction of tertiary amides to aldehydes and hydrozirconation. *Org. Lett.* **16**, 390–393 (2014).

41. Gibson, D. G. Enzymatic Assembly of Overlapping DNA Fragments. in 349–361 (2011). doi:10.1016/B978-0-12-385120-8.00015-2

42. Krause, L., Herbst-Irmer, R., Sheldrick, G. M. & Stalke, D. Comparison of silver and molybdenum microfocus X-ray sources for single-crystal structure determination. *urn:issn:1600-5767* **48**, 3–10 (2015).

Chapter 4

Expanding the substrate scope of C-C bond forming L-threonine transaldolase through directed evolution

ObiH library screening and validation of hits were done in collaboration with Sam Bruffy. Optimization of colorimetric assay were done in collaboration with Grace Carlson and Sam Bruffy.

Chapter 4: Expanding the substrate scope of C-C bond forming L-threonine transaldolase through directed evolution

4. 1. Introduction

Synthetic methods have the highest biomedical impact when they operate on diverse substrates to rapidly build molecular complexity.[1] Enzymes have a profound potential to contribute to green chemical synthesis.[2] However, their poor substrate scope often limits their synthetic utility.[3] Hence, enzymes that catalyze transformation on a range of substrates are highly desired, a property referred to as substrate promiscuity. Few enzymes, such as lipases and ketoreductases, perform transformations on a range of diverse substrates, which led to their widespread industrial use.[4] Discovery and characterization of promiscuous enzymes will increase the adoption of enzymes by the synthetic community. To this end, we have performed detailed structural and mechanistic characterization of ObiH, a promiscuous L-threonine transaldolase (Chapter 2).[5] Then, we demonstrated the synthetic utility of ObiH for the diastereoselective production of diverse valuable β-hydroxy amino acids from inexpensive starting materials (Chapter 3).[6]

Researchers often desire biocatalysts to perform a reaction that is not found in Nature.[7] In these cases, a common strategy is to first search for an enzyme that performs the desired reaction and has low levels of activity with the desired non-native substrate. Once an appropriate starting point is found, activity can be readily increased through traditional directed evolution.[8] This strategy mimics the divergent evolution that occurs in Nature and has been used by pharmaceutical companies for dozens of large-scale processes, such as the synthesis of sitagliptin and islatrarvir.[9,10] Despite the immense progress in the field of biocatalysis, there are still many desirable transformations that no enzyme is known to catalyze.[7]

We have shown that ObiH natively has reactivity with structurally diverse aromatic, heterocyclic, and aliphatic aldehydes to generate a palette of β-hydroxy amino acids (See section

3. 2. 2). This reactivity is enabled by the kinetically-trapped PLP quinonoid intermediate, E(Q^{Gly}) (See section 2. 2. 1). We hypothesized that the nucleophilic E(Q^{Gly}) could react with non-aldehyde electrophiles, enabling access to novel amino acids. Here, we describe our ongoing efforts to evaluate and engineer ObiH for reactivity with non-aldehyde electrophiles towards our ultimate goal of developing ObiH as a general enzymatic route for stereoselective C-C bond formation reaction.

4. 2. Results and Discussion

4. 2. 1. ObiH reacts with ketones to form quaternary β-hydroxy amino acids

To evaluate the reactivity of ObiH with non-aldehyde electrophiles, we performed analytical scale enzymatic reactions with a small panel of diverse electorophiles (Figure 1). These reactions were analyzed using UPLC-MS, and we found that ObiH had a low level of activity with two ketones: *p*-nitroacetophenone (**1**) and 1,1,1-trifluoro-3-phenyl-propan-2-one (**2**). In contrast, we did not see any evidence of product formation with Claisen substrates and Michael acceptors. Among the two ketones, the electron-deficient ketone **2** had much higher reactivity than *p*-nitroacetophenone, suggesting that electrophilicity is a key property of the substrate that determines reactivity. Michael acceptors chosen here are more reactive than ketones **1** and **2**, as determined by Mayr electrophilicity measurements.[11] However, they did not result in a productive catalysis. Together, these data suggest that ObiH possibly uses certain structural features of the electrophile to select them for stereoselective C-C bond formation.

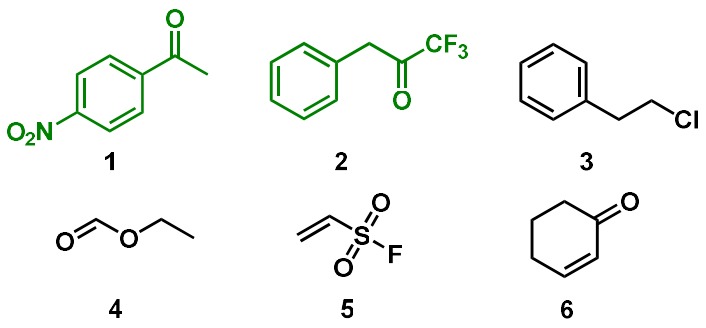


Figure 1. Evaluation of reactivity of ObiH with non-aldehyde electrophiles. Only the ketone substrates shown in green had reactivity with ObiH. Data for non-ketone substrates were collected by Sam Bruffy.

We are currently pursuing preparative scale reactions with the ketones to isolate the tertiary β-hydroxy amino acid product **9**. Upon isolation, we would use small molecule crystallography to determine the absolute configuration of the major product and the diastereomeric ratios. The ability of ObiH to react with ketones to form a new tetrasubstituted tertiary stereocenter is a synthetically challenging transformation and will further increase the appeal of ObiH as a biocatalyst.[12,13]

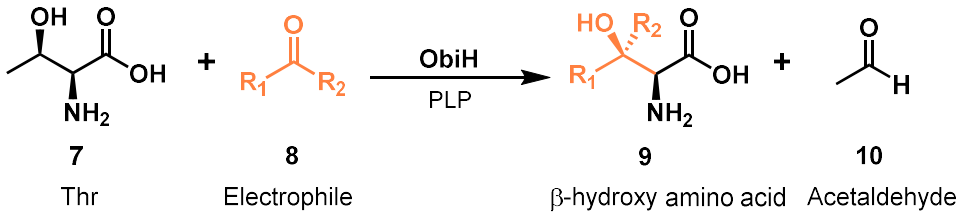


Figure 2. ObiH reaction with ketones. The general reaction catalyzed by ObiH to from quaternary β-hydroxy amino acids from Thr (black) and ketone (orange).

4. 2. 2. Colorimetric assay identified hotspots for improving reactivity

ObiH is not reactive with electron-rich aldehydes like *p*-hydroxybenzaldehyde but reacts with structurally similar electron-poor aldehydes such as *p*-chlorobenzaldehyde (See section 3. 2. 2). We also observed similar trends towards reactivity with ketones suggesting that electrophilicity, not sterics, determines the rate of reactivity towards the nucleophilic E(Q^{Gly}) intermediate. We hypothesize that if we increase the lifetime of the E(Q^{Gly}) intermediate, we will be able to improve the reactivity of ObiH. To test this hypothesis, we decided to employ protein engineering to improve the reactivity of ObiH by screening libraries at conditions that select for a more stable E(Q^{Gly}) intermediate. The rate of protonation determines the lifetime of the E(Q^{Gly}) intermediate. Predicting the residues in the enzyme that affect this protonation is challenging. Therefore, we screened a globally random mutagenesis library to identify variants with improved reactivity. We performed the screening at pH 7.0, which increases the rate of protonation of E(Q^{Gly}) to glycine compared to pH 8.5 by approximately two-fold (See section 2. 2. 1). The increase in the rate of protonation could act as a selective pressure enabling the identification of variants with more persistent E(Q^{Gly}).

To accelerate the screening process, we implemented a previously reported colorimetric assay for screening ObiH libraries.[14] This assay uses phenylserine dehydratase which catalyzes the dehydration reaction of phenylserine derivatives to form phenylpyruvate.[15] When phenylpyruvate forms a coplanar enol structure, it can react with Fe^{3+} to form a complex with a cyan color (Figure 3a). Since this assay does not work with ketones, we decided to use *p*-chlorobenzaldehyde, a model substrate for ObiH reactions, as the electrophile. Optimization of the colorimetric assay for ObiH reaction were done in collaboration with Grace Carlson and Sam Bruffy.

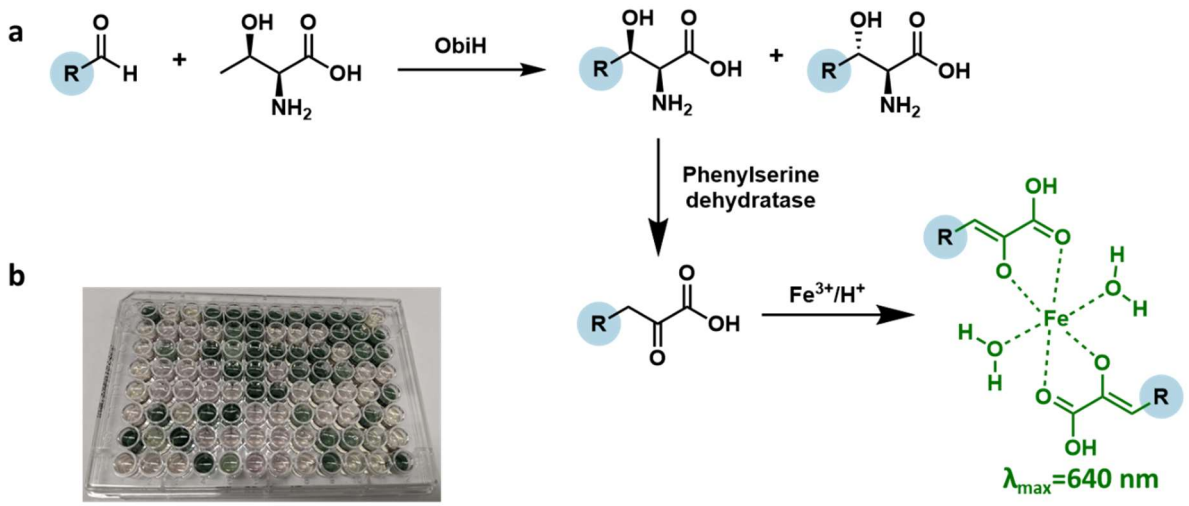


Figure 3. Development of colorimetric assay to accelerate ObiH library screening. a) Reactions of colorimetric assay. **b)** Photo of colorimetric assay from a representative library plate.

We screened a globally random mutagenesis library in a 96-well format. By comparing the absorbance of the variants to wt-ObiH, we were able to identify variants with higher product formation. In total, we screened ~1600 variants and found 15 wells that had significantly higher absorbance than wt-ObiH. These wells were sent for sequencing, and 6/15 turned out to be wt-ObiH. Even though a false positive rate of 40% is high, the colorimetric assay enabled much rapid screening of the libraries compared to traditional LC-MS-based screening. Also, no significant time was lost due to the false positives. The remainder of the hits identified in the colorimetric screen contained the following mutations:

F70S
D85G (observed twice)
F70L
S2P, H23Y, K88R, T105A
F70I, I91T
S136P
V341A
F181S, Y393C

These mutations are dispersed throughout the protein structure (Figure 4). Out of these eight unique variants, three of them turned out to be single-site mutations at Phe70, which attracted us to investigate this residue further. Phe70 is located on a 20-residue long loop (Tyr55 to Gly74) part of the active site.

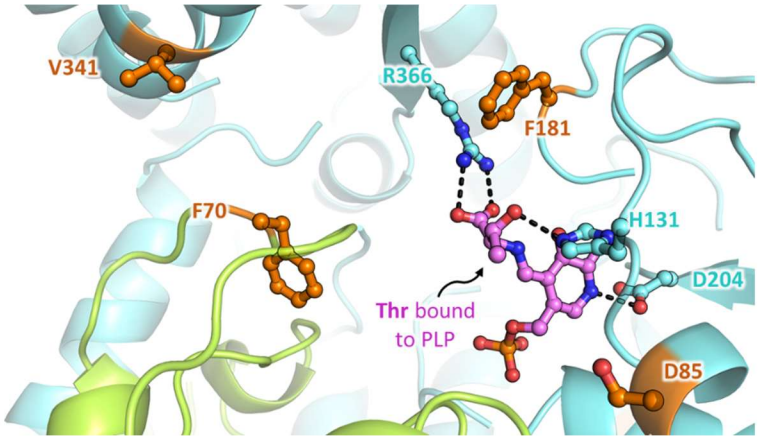


Figure 4. ObiH active site model with Thr bound to PLP as the external aldimine (EAex). The active site is at the dimer interface and individual monomers are colored in green and cyan. Active site residues are shown as cyan sticks. Variants identified from colorimetric assay are shown as orange sticks. Hydrogen bonds are shown as black dashes.

Previously, Mr. Jon Ellis observed this loop region to undergo large conformational changes during molecular dynamics simulations of ObiH external aldimine. This loop contains a six-residue insertion compared to a structurally similar SHMT (PDB: 1KL2), which may explain the high degree of mobility observed (Figure 5). A neighboring loop, Glu355-His363, is shorter compared to SHMT. These two loop regions directly interact with folate in SHMT, and the modifications present in ObiH may contribute to the aldehyde specificity of the enzyme.

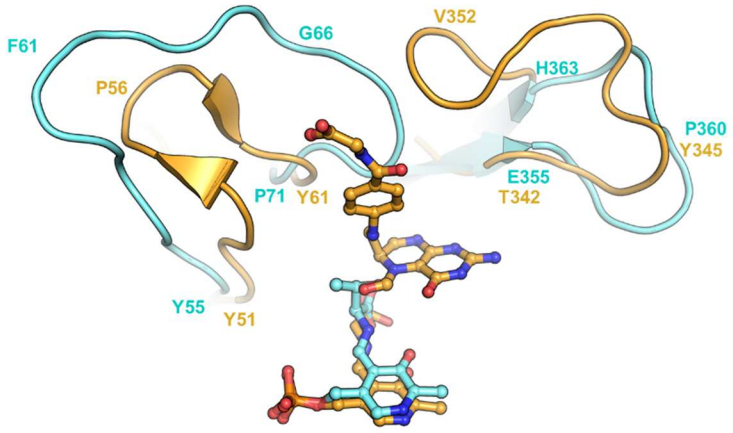


Figure 5. Superimposed view of the substrate access channel in the MD simulated ObiH-E(Aex^{Thr}) (cyan) and SHMT-E(Aex) crystal structure (PDB ID: 1KL2) (gold).

These experimental and computations results confirm that the active site Tyr55-Gly74 loop likely contributes to the aldehyde specificity and could be a hot spot for tuning ObiH activity. Therefore, we decide to screen site-saturation mutagenesis libraries at residues spanning these loop regions to identify additional improved variants. We are also validating other hits identified from the global random mutagenesis library. After validation, we will perform spectroscopic characterization to verify whether the improved variants form a persistent nucleophilic E(Q^{Gly}) intermediate.

4. 2. 3. SUMS directed evolution identified several improved variants

We decided to screen site-saturation mutagenesis libraries at residues in the Tyr55-Gly74 loop region to identify improved variants with either aldehydes or ketones or both. Though traditional protein engineering excels at increasing activity on a single substrate, it provides no selective pressure to improve activity across a broad substrate scope. Since we wanted to develop a promiscuous biocatalyst, we decided to screen libraries by letting multiple substrates compete for the enzyme active site (Figure 6). This methodology, referred to as substrate multiplexed screening (SUMS), was developed in our lab by leveraging the ability of a mass spectrometer to measure multiple compounds simultaneously and has facilitated the identification of enzymes with an increase in activity on multiple substrates.[16]

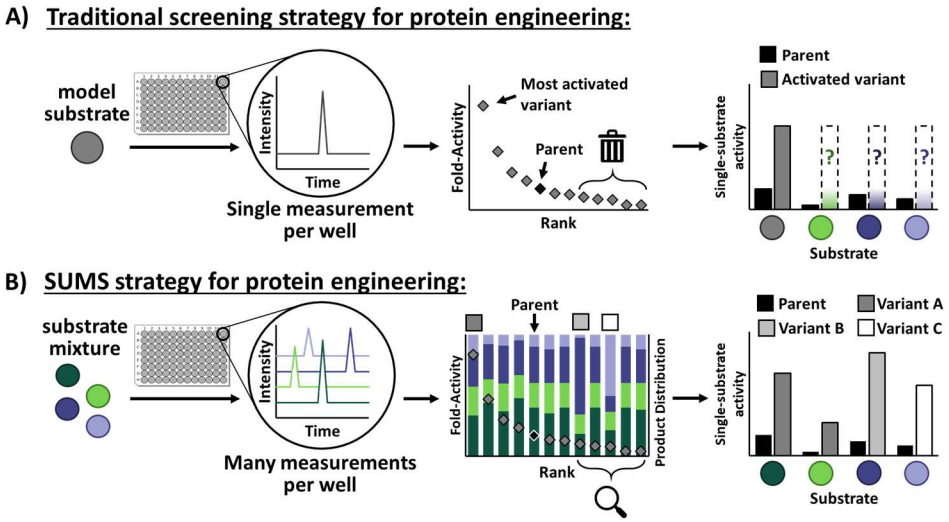


Figure 6. Comparison of traditional screening and SUMS. A) Traditional protein engineering workflow utilizes one model substrate for screening. The identified activated variant is tested against the desired substrate scope at the end of engineering. **B)** Substrate-multiplexed screening (SUMS) is compatible with a traditional workflow while providing information on enzyme activity as well as specificity. Different variants identified as activating for substrates of interest are verified in single-substrate reactions at the end of engineering. Figure reproduced from McDonald et al.[16]

We choose isovaleraldehyde, pivaloxybenzaldehyde, and 1,1,1-trifluoro-3-phenyl-propan-2-one (**2**) as the electrophiles for SUMS screening as they encompass two structurally diverse aldehydes and a ketone. We first screened F70 site-saturation mutagensis (F70X) library and found that there were several variants that had higher activity than wt-ObiH for both the aldehyde substrates. None of the variants had improved activity for the ketone substrate. If we have used only trifluorphenylpropanone as the electrophile for screening this library, we would not have identified any hits. These results demonstrate the strength of SUMS to identify variants that would have been lost in traditional proteon engineering methodolgoies.

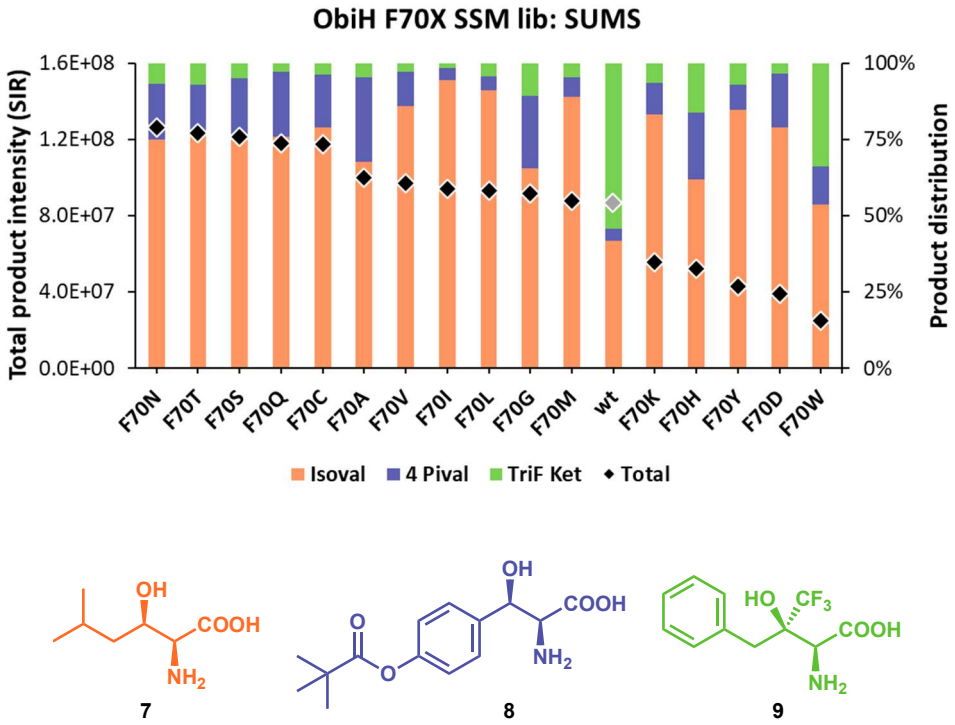


Figure 7. Substrate-multiplexed screening (SUMS) of F70X library. 5 mM isovaleraldehyde, 10 mM pivaloxybenzaldehyde, and 15 mM 1,1,1-trifluoro-3-phenyl-propan-2-one were used as the electrophiles. Colored bars indicate relative abundances of each product, and black diamonds indicate total intensity of single ion retention (SIR) for each product's unique m/z. Structures of the products are shown below the SUMS data and colored to match.

Encouragingly, the colorimetiric assay indicated F70I, L, and S mutations were all activating with the *p*-chlorobenzaldehyde substrate. Results of SUMS of the F70X library align with these observations and identfied the F70A and Q mutations as potentially more activating for benzaldehyde substrates. Although activity on the desired ketone **2** decreased, these results motivated us to screen H69X, W68X, and P71X libraries using the SUMS approach (Figure 8). These libraries have several hits that have improved activity with either aldehydes or ketones or both (Figure 9). For example, H69A has improved reactivity with both pivaloxybenzaldehyde and ketone **2**. F70S has improved reactivity with both the aldehydes screened, and W68M with ketone **2**.

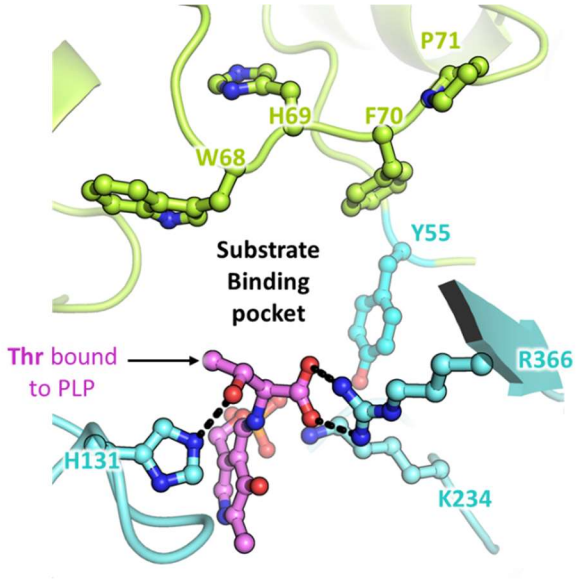


Figure 8. Active site of ObiH. Active site residues are shown as sticks. Residues in the active site loop Tyr55-Gly74 chosen for SUMS screening are shown in lime. Hydrogen bonds are shown as black dashes.

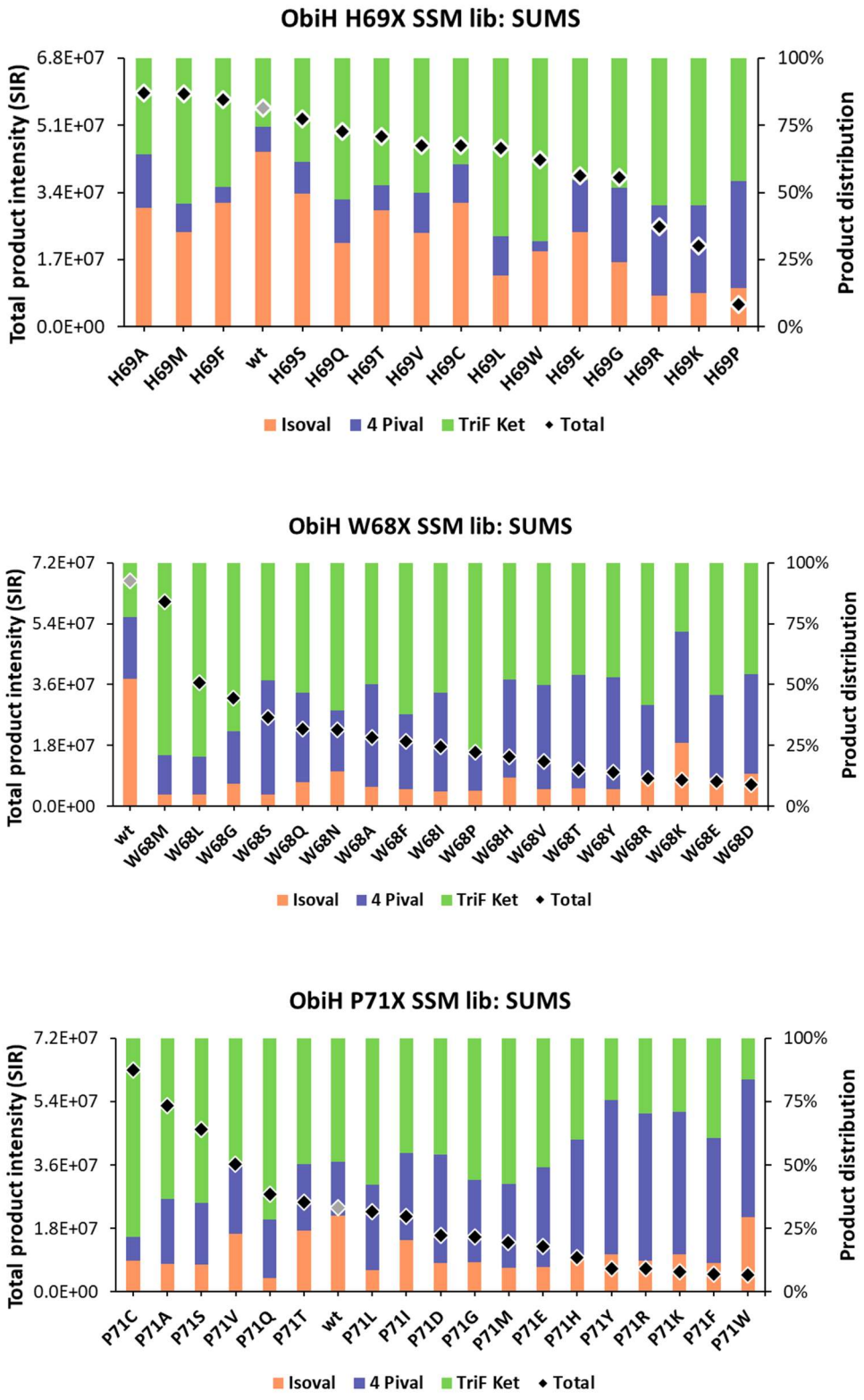


Figure 9. Substrate-multiplexed screening (SUMS) of H69X, W68X and P71X library. 5 mM isovaleraldehyde, 10 mM pivaloxybenzaldehyde, and 15 mM 1,1,1-trifluoro-3-phenyl-propan-2-

one were used as the electrophiles. Colored bars indicate relative abundances of each product, and black diamonds indicate total intensity of single ion retention (SIR) for each product's unique m/z.

We choose to evaluate two of the best variants from SUMS screening in single-substrate reactions: F70S (improved reactivity with both the aldehydes screened) and W68M (improved reactivity with ketone **2**). F70S had ~3 fold higher activity than wt-ObiH for both the aldehydes present in the SUMS screening (Figure 10, pink bars). Gratifyingly, F70S also had improved reactivity with aldehydes that were not present in the SUMS screen, such as the heterocyclic imidazole-4-carboxaldehyde. Similar to the library screening results, F70S had no improved activity with any of the ketone substrates. These results shows that results from SUMS screening translate well to single-substrate reactions.

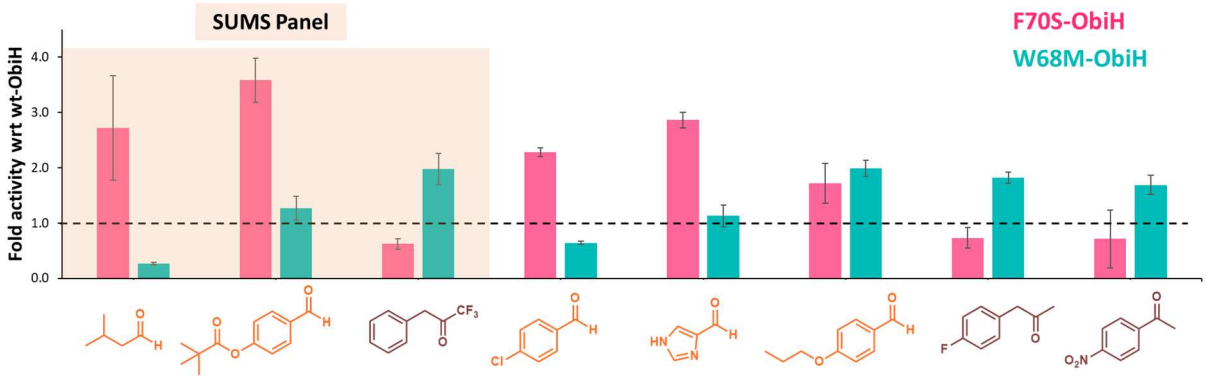


Figure 10. Product formation in single-substrate reactions. Activity of F70S (pink bars) and W68M (cyan bars) is shown relative to activity of wt-ObiH. Structures of the electrophile substrates are shown below.

W68M had ~2 fold higher activity with ketone **2** as well as two other ketones (*p*-nitroacetophenone and (4-fluorophenyl)acetone) that were not present in the SUMS screen (Figure 10, cyan bars). W68M had low reactivity with the aliphatic isovaleraldehyde and wt-ObiH

like activity for the rest of the aldehydes. These validation experiments show that W68M and F70S are broadly activating for ketones and aldehydes, respectively. Overall, these results demonstrate that SUMS is a powerful methodology for engineering promiscuous enzymes.

Because all the four site-saturation libraries yielded shifts in specificity and improved reactivity, we decided to pursue their cooperative effects by screening a multi-site combinational library comprising the top variants from each of these residues with more challenging aldehydes and ketone substrates. These experiments are in progress and could yield multiple variants with improved reactivity. These variants will also be monitored for serendipitous activity on the non-ketone electrophiles shown in Figure 1. We will also deploy kinetic, spectroscopic, and structural techniques to determine what features underly the new increases in activity.

4. 3. Conclusions

Here, we evaluated the ability of ObiH to react with a panel of non-aldehyde electrophiles and found that ObiH can react with ketones to form tertiary β-hydroxy amino acids. To improve the reactivity of ObiH, we screened random mutagenesis library through a colorimetric assay. Hits from this assay resulted in the identification of an active site loop that tunes the reactivity of ObiH. SUMS screening of four site-saturation mutagenesis libraries spanning the active site loop resulted in the identification of several variants with improved reactivity to aldehyde and ketone substrates. These variants also have improved reactivity with substrates that were not initially present in the screen. These improved variants will further increase the synthetic utility of ObiH and provide access to novel amino acid building blocks.

4. 4. Materials and Methods

General experimental procedures

Chemicals and reagents were purchased from commercial suppliers (Sigma-Aldrich, VWR, Chem-Impex International, Combi-blocks, Alfa Aesar, New England Biolabs, Zymo Research, Bio-Rad) and used without further purification unless otherwise noted. BL21 (DE3) *E. coli* cells were electroporated with a Bio-Rad MicroPulser electroporator at 2500 V. New Brunswick I26R, 120 V/60 Hz shaker incubators (Eppendorf) were used for cell growth. Optical density and UV-vis measurements were collected on a UV-2600 Shimadzu spectrophotometer (Shimadzu). UPLC-MS data were collected on an Acquity UHPLC with an Acquity QDa MS detector (Waters) using an ACQUITY UPLC CSH BEH C18 column (Waters) or an Intrada Amino Acid column (Imtakt).

Cloning, expression, and purification of ObiH

A codon-optimized copy of the ObiH gene was purchased as a gBlock from Integrated DNA Technologies. This DNA fragment was inserted into a pET-28b(+) vector by the Gibson Assembly method.[17] BL21 (DE3) *E. coli* cells were subsequently transformed with the resulting cyclized DNA product via electroporation. After 45 min of recovery in Luria-Burtani (LB) media containing 0.4% glucose at 37 °C, cells were plated onto LB plates with 50 μg/mL kanamycin (Kan) and incubated overnight. Single colonies were used to inoculate 5 mL LB + 50 μg/mL Kan, which were grown overnight at 37 °C, 200 rpm. Expression cultures, typically 1 L of Terrific Broth (TB) + 50 μg/mL Kan (TB-Kan), were inoculated from these starter cultures and shaken (180 rpm) at 37 °C. After 3 hours (OD_{600} = ~0.6), the expression cultures were chilled on ice. After 30 min on ice, ObiH expression was induced with 0.5 mM IPTG, and the cultures were expressed for 16 hours at 20 °C with shaking at 180 rpm. Cells were then harvested by centrifugation at 4,300×*g*

at 4 °C for 30 min. Cell pellets were pink in color and were frozen and stored at -20 °C until purification.

To purify ObiH, cell pellets were thawed on ice and then resuspended in lysis buffer (50 mM potassium phosphate buffer buffer (pH 8.0), 500 mM NaCl, 1 mg/ml Hen Egg White Lysozyme (GoldBio), 0.2 mg/ml DNaseI (GoldBio), 1 mM $MgCl_2$, 1 X BugBuster Protein extraction reagent (Novagen), and 400 μM pyridoxal 5′-phosphate (PLP)). A volume of 4 mL of lysis buffer per gram of wet cell pellet was used. After 45 min of shaking at 37 °C, the resulting lysate was then spun down at 75,600×*g* to pellet cell debris. The pellet was colorless whereas the supernatant was pink in color. Ni/NTA beads (GoldBio) were added to the supernatant and incubated on ice for 45 min prior to purification by Ni-affinity chromatography with a gravity column. The column was washed with 5 column volumes of 20 mM imidazole, 500 mM NaCl, 10% glycerol, 50 mM potassium phosphate buffer (pH 8.0). Washing with higher concentrations of imidazole resulted in slow protein elution. ObiH was eluted with 250 mM imidazole, 500 mM NaCl, 10% glycerol, 50 mM potassium phosphate buffer, pH 8.0. Elution of the desired protein product was monitored by the disappearance of its bright red color (resulting from the release of ObiH) from the column. The protein product was dialyzed to < 1 μM imidazole in 100 mM Tris buffer, pH 8.5 containing 2 mM DTT, dripped into liquid nitrogen to flash freeze, and stored at -80 °C before use. The concentration of protein was determined by Bradford assay. Generally, this procedure yielded 200 – 250 mg per L culture. F70S and W68M variants were purified similar to wt-ObiH.

Evaluation of ObiH substrate scope

Stocks of Thr were made in 100 mM Tris pH 8.5 and electrophile in MeOH. 0.5-dram glass vials were charged with 62.5 μL of 500 mM Thr, 10.0 μL of 500 mM electrophile, 5.0 μL of 20 mM PLP and 83.2 μL of 100 mM Tris pH 8.5. Reactions were initiated by the addition of 49.3 μL of 400 μM purified ObiH (Final concentrations: 125 mM Thr, 25 μmol; 20 mM electrophile, 5 μmol;

0.4 mol% cat., 400 μM PLP; 20% Methanol). Reactions were allowed to proceed in a 37 °C incubator for 16 h and then quenched with 250 μL ACN. Quenched reactions were then centrifuged at 15,000 xg to remove aggregated protein. Amount of amino acid product was quantified by performing UPLC-MS analysis of the undiluted supernatant on a BEH C18 column (Waters) using positive mode single ion readout for the M+H mass peak.

Production of ObiH random mutagenesis libraries

Random mutagenesis was carried out via error-prone PCR. Reaction conditions were optimized to generate 1-2 codon mutations per plasmid. Reactions were setup by adding the following to a PCR tube: 5 μL 10x Taq buffer (New England Biolabs), 1 μL 10 mM dNTP mix, 1 μL 10 μM 22b-intF, 1 μL 10 μM 22b-intR, 1 μL ~100 ng/μL parent plasmid, 5.5 μL 50 mM $MgCl_2$, 2.5 μL 1 mM $MnCl_2$, 1 μL DMSO, 0.5 μL Taq polymerase (New England Biolabs) and total volume was made up to 50 μL with H_2O. Reactions were carried out in a thermocycle according to the following scheme:

Thermocycler program

Step 1: 95 °C 2 min 30 s

Step 2: 95 °C 15 s

Step 3: 55 °C 20 s

Step 4: 68 °C 1 min 45 s

Step 5: 68 °C 5 min

Extension steps 2 – 4 were performed for 30 cycles.

The PCR product was purified using a preparative agarose gel. Purified DNA fragment was inserted into a pET-28b(+) vector by the Gibson Assembly method.[18] BL21 (DE3) E. coli cells were subsequently transformed with the resulting cyclized DNA product via electroporation. After

45 min of recovery in Luria-Burtani (LB) media containing 0.4% glucose at 37 °C, cells were plated onto LB plates with 100 μg/mL Ampicillin (Amp) and incubated overnight. Single colonies were used to inoculate 5 mL LB + 100 μg/mL amp (LB_{amp}), which were grown overnight at 37 °C, 200 rpm. Colonies were sequenced and there were 1 – 2 coding mutations.

Design of site-saturation mutagenesis library

PCR was conducted using Phusion polymerase (New England Biolabs) according to the standard protocol. For the given site of mutagenesis, three primers were designed containing codons NDT (encoding for Ile, Asn, Ser, Gly, Asp, Val, Arg, His, Leu, Phe, Tyr, and Cys), VHG (encoding for Met, Thr, Lys, Glu, Ala, Val, Gln, Pro, and Leu), and TGG (Trp), respectively, thereby including all 20 natural amino acids. These three primers were mixed in a ratio 12:9:1 according to the previously described protocol.[19] Library plasmids were constructed by overlap extension PCR using the plasmid that contained the wt-ObiH gene in the pET28(b)+ vector as a template. DpnI was added to digest the template DNA (1 μL per 50 μL PCR reaction, incubated at 37 °C for 1 hour), and the amplified library genes were purified by preparative agarose gel then integrated into a linear pET28(b)+ plasmid via the Gibson Assembly method.[17]

Expression of ObiH libraries

BL21 (DE3) *E. coli* cells carrying parent and variant plasmids were grown in 96-deep-well plates (600 μL/well TB-Kan) at 37 °C with shaking at 200 rpm. After shaking at 250 rpm overnight, 10 μL of the overnight cultures were transferred to new deep-well plates containing 600 μL/well TB-Kan, which were allowed to grow at 37 °C with shaking at 200 rpm. After 3 h, the plates were chilled on ice for 30 min, then induced by the addition of IPTG in TB-Kan (1 mM final concentration). The cultures were expressed for 16 hours at 20 °C with shaking. After 20 hours,

cells were harvested by centrifugation at 4,300×g for 20 min. The cell pellets were frozen at –20 °C for a minimum of 2 hours before screening.

Preparation of Fe^{3+} colorimetric reagent

The Fe^{3+} colorimetric agent (100 mL) was prepared by adding 0.05 g ferric chloride, 2 mL acetic acid, 38 mL of water and 60 mL DMSO.

Screening of ObiH libraries using colorimetric assay

ObiH pellets were re-suspended in Tris-HCl buffer (100 mM, pH 7.0) with 100 mM Thr, 25 mM 4-Cl-benzaldehyde, 100 μM PLP and 20% Methanol (v/v) in a final volume of 0.6 mL. The biotransformations were performed at 37°C for 20 h. After centrifugation, 500 μL of the was supernatant was transferred to a new 96 deep well plate and 50 uL of 75 uM PSDH was added. The dehydration reaction was carried out at 37°C for 4 h. After centrifugation, 40 uL of supernatant was mixed with 160 μL of chromogenic agent and absorbance measured at 600 nm after incubation for 30 min at r.t.

SUMS library screening

ObiH pellets were re-suspended in Tris-HCl buffer (100 mM, pH 7.0) with 75 mM Thr, 5 mM isovaleraldehyde, 10 mM Pivaloxybenzaldehyde, 15 mM 1,1,1-trifluoro-3-phenyl-propan-2-one, 100 μM PLP and 20% Methanol (v/v) in a final volume of 0.5 mL. The plates were sealed with Teflon sealing mats, then incubated at 37 °C for 1 hour. After 1 hour, the reactions were quenched with 500 μL of acetonitrile. Quenched reaction mixtures were filtered using a filter plate to remove debris (by centrifugation at 4,300×*g* for 5 min) prior to UPLC-MS injection. Product formation was quantified by integration of peaks on single ion retention (SIR) channels corresponding to each expected product.

Evaluation of ObiH variants

ObiH variants were thawed on ice from storage at -80 °C and then centrifuged at 15,000 x*g* for 5 min to pellet aggregated protein. Single substrate reactions were prepared in triplicate for each ObiH variant on a 96-well plate. Tris pH 7.0 buffer containing Thr (100 mM, Tris pH 7.0) was combined with aldehyde substrate (25 mM) and ObiH variants (10 μM, 2500 max TON). Aldehyde substrates were dissolved in methanol, and the final reaction mixture was composed of 20% Methanol. Reactions were run at 37 °C for 1 h. Reaction mixture (500 μL) were quenched with 500 μL ACN + 2.5 mM Benzophenone (as internal standard). Quenched reactions (200 μL) were transferred to a 96-well filter plate and spun down at 1000 x*g* for 10 min into a 96-well plate for UPLC-MS analysis. Amount of amino acid product was quantified by performing UPLC-MS analysis of the 1:10 supernatant (diluted with 50% ACN) on a BEH C18 column (Waters) using positive mode single ion readout for the M+H mass peak.

4. 5. References

1. Reetz, M. T. What are the Limitations of Enzymes in Synthetic Organic Chemistry? *Chem. Rec.* **16**, 2449–2459 (2016).

2. Sheldon, R. A. & Woodley, J. M. Role of Biocatalysis in Sustainable Chemistry. *Chem. Rev.* **118**, 801–838 (2017).

3. Goodwin, N. C., Morrison, J. P., Fuerst, D. E. & Hadi, T. Biocatalysis in Medicinal Chemistry: Challenges to Access and Drivers for Adoption. *ACS Med. Chem. Lett.* **10**, 1363–1366 (2019).

4. Nestl, B. M., Hammer, S. C., Nebel, B. A. & Hauer, B. New generation of biocatalysts for organic synthesis. *Angew. Chemie - Int. Ed.* **53**, 3070–3095 (2014).

5. Kumar, P. *et al.* L -Threonine Transaldolase Activity Is Enabled by a Persistent Catalytic Intermediate. *ACS Chem. Biol.* **16**, 95 (2021).

6. Doyon, T. J. *et al.* Scalable and Selective β-Hydroxy-α-Amino Acid Synthesis Catalyzed by Promiscuous l-Threonine Transaldolase ObiH. *ChemBioChem* **23**, e202100577 (2022).

7. Truppo, M. D. Biocatalysis in the Pharmaceutical Industry: The Need for Speed. *ACS Med. Chem. Lett.* **8**, 476–480 (2017).

8. Almhjell, P. J., Boville, C. E. & Arnold, F. H. Engineering enzymes for noncanonical amino acid synthesis. *Chem. Soc. Rev.* **47**, 8980–8997 (2018).

9. Savile, C. K. *et al.* Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **329**, 305–309 (2010).

10. Huffman, M. A. *et al.* Design of an in vitro biocatalytic cascade for the manufacture of islatravir. *Science (80-.).* **366**, 1255–1259 (2019).

11. Mayr, H. & Patz, M. Scales of Nucleophilicity and Electrophilicity: A System for Ordering Polar Organic and Organometallic Reactions. *Angew. Chemie Int. Ed. English* **33**, 938–957 (1994).

12. Roupnel, L. *et al.* Highly Stereoselective Aldol Reactions by Memory of Chirality: Synthesis of Quaternary β-Hydroxy α-Amino Acids. *Helv. Chim. Acta* **104**, e2100127 (2021).

13. Liu, Y., Han, S. J., Liu, W. B. & Stoltz, B. M. Catalytic Enantioselective Construction of

Quaternary Stereocenters: Assembly of Key Building Blocks for the Synthesis of Biologically Active Molecules. *Acc. Chem. Res.* **48**, 740–751 (2015).

14. Chen, Q. *et al.* Improving and Inverting C $_{\beta}$ -Stereoselectivity of Threonine Aldolase via Substrate-Binding-Guided Mutagenesis and a Stepwise Visual Screening. *ACS Catal.* **9**, 4462–4469 (2019).

15. H, O., S, N. & H, M. Cloning, sequencing, and overexpression in Escherichia coli of a phenylserine dehydratase gene from Ralstonia pickettii PS22. *Biosci. Biotechnol. Biochem.* **66**, 2755–2758 (2002).

16. Mcdonald, A. D., Higgins, P. M. & Buller, A. R. Substrate multiplexed protein engineering facilitates promiscuous biocatalytic synthesis. (2021). doi:10.26434/CHEMRXIV-2021-XL252

17. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

18. Gibson, D. G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

19. Kille, S. *et al.* Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth. Biol.* **2**, 83–92 (2013).

Chapter 5

How I used vitamins and proteins to synthesize new amino acids:

A chapter for the non-scientist

I believe that everyone will be able to enjoy and appreciate the fascinating world of chemical biology if the ideas are communicated in a simple and relatable language. This is why I wrote this chapter to describe the broader context and findings from my Ph.D. research to a wide, non-scientist audience. I want to thank the team at the Wisconsin Initiative for Science Literacy Program at the University of Wisconsin-Madison for their encouragement and support in the creation of this thesis chapter.

Chapter 5: How I used vitamins and proteins to synthesize new amino acids: A chapter for the non-scientist

5. 1. What did I do in my Ph.D.?

I am sure you have seen cereals at some point in your life, maybe even had them for breakfast this morning. If you read the nutrition facts on the side of the cereal box, you might have noticed they contain proteins and vitamins (see Figure 1). Today, I will tell a story of how we (my collaborators and I) used proteins and vitamins to build (synthesize) new amino acids. The first question that might come to your mind is: What are amino acids? Let's start with this question.

Figure 1. A picture of a multi-grain Cheerios cereal box. The red circle highlights the proteins and vitamin content.

5. 2. What are amino acids?

To build a sturdy house, we often use bricks as the building block. Similarly, cells in our bodies use 20 different amino acids as the building block to synthesize proteins (see Figure 2). When we eat foods that contain proteins, the protein gets broken down into individual amino acids, which are eventually used to build new proteins (similar to recycling plastic or paper). Our bodies can't make all 20 standard amino acids, so we need proteins in our diet.[1]

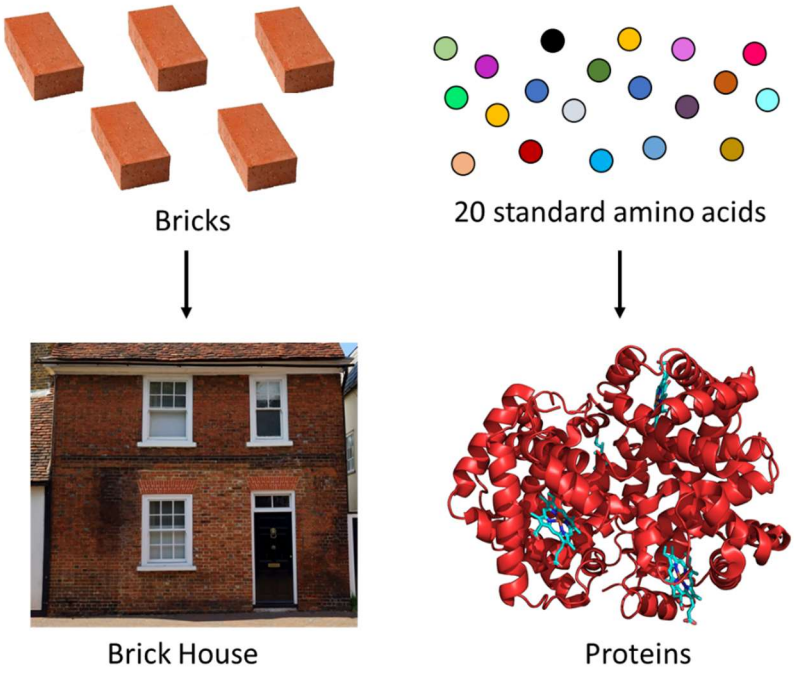


Figure 2. The emergence of complexity from simple building blocks around us and in Nature. The bottom right is a simplified representation of the Hemoglobin protein structure. The picture of the concrete home was downloaded with permission from pxhere.com.

The next question is: What are proteins? Proteins are the molecular machines in our body that carry out almost all bodily functions. For example, Hemoglobin is the protein that carries oxygen to our lungs, and Insulin is the protein that regulates sugar levels in the blood. Structurally,

proteins are polymers (a molecule composed of many simple repeating units) of the 20 standard amino acids, like beads on a necklace (see Figure 3). These amino acids differ in their side chain or R group, which brings unique properties to them.

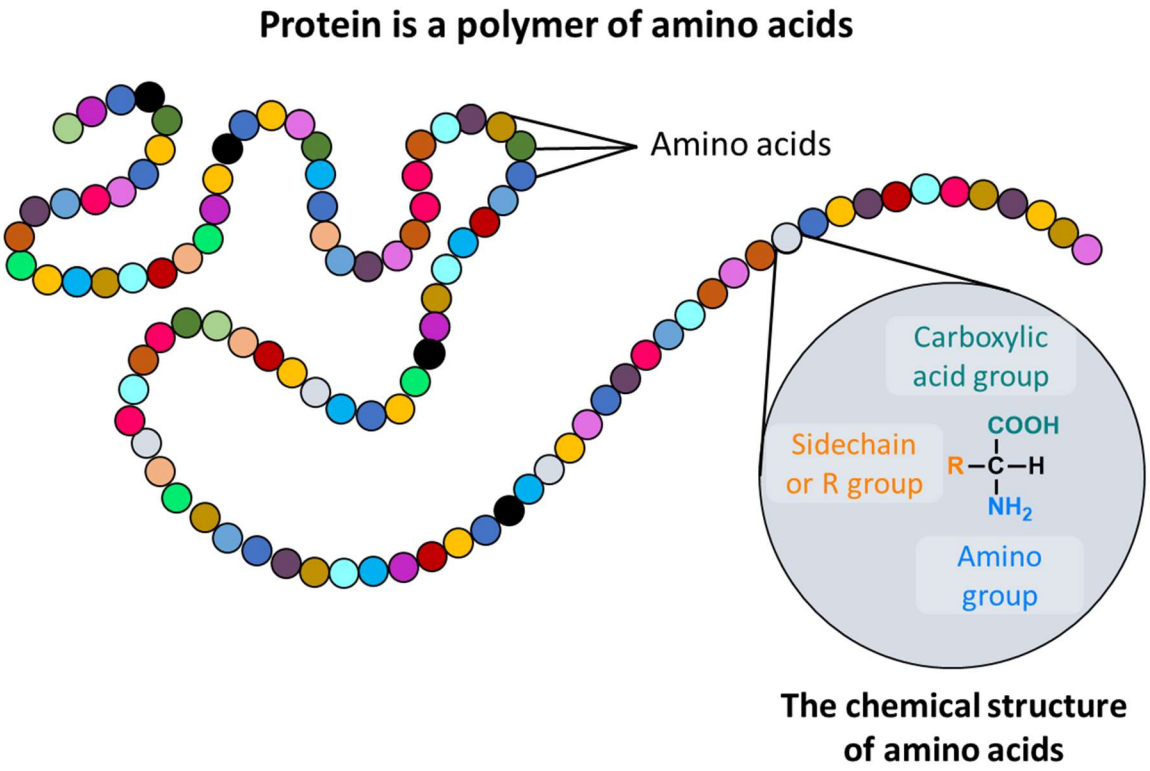


Figure 3. Primary structure of proteins. Proteins are polymers of the 20 standard amino acids. The general chemical structure of amino acids is shown in the grey circle.

5. 3. Why do we need new amino acids?

The next question that might come to your mind is: If cells only need 20 amino acids to make proteins, why do we need *new* amino acids? That's an excellent question! The answer is cells need more than 20 amino acids to build several biologically valuable compounds (see Figure 4).[2] For example, the first isolated antibiotic, Penicillin, is built from three different amino acids. Out of the three, two of the amino acids are not found in proteins. These amino acids are called non-standard or non-proteogenic amino acids. The hormone thyroxine (produced by the thyroid gland) and neurotransmitter serotonin are also built from such non-standard amino acids.

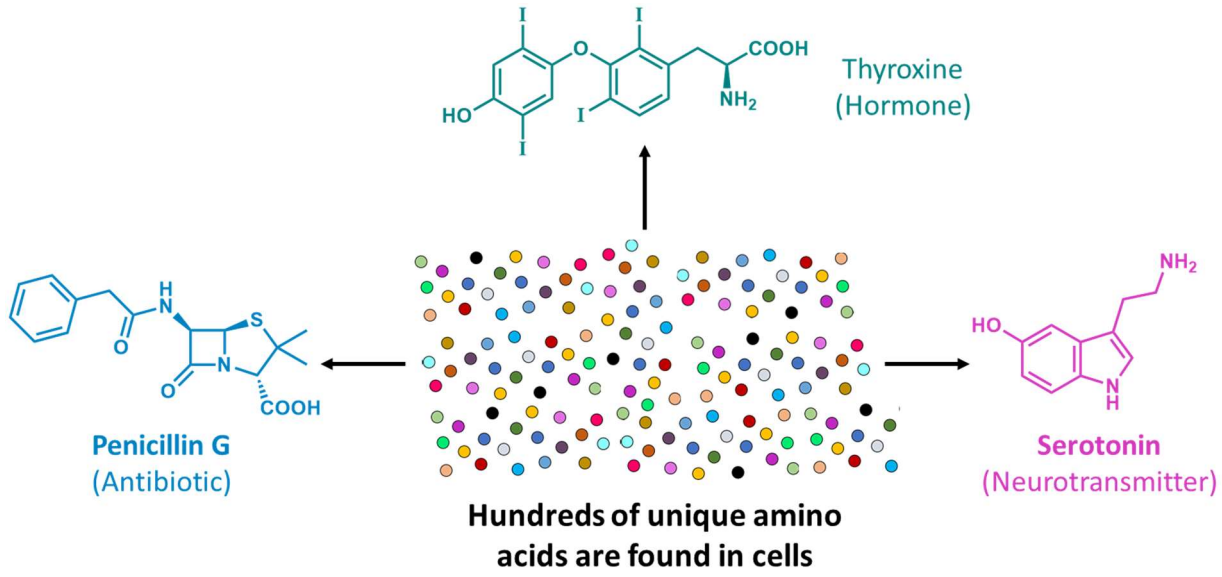


Figure 4. Several biologically valuable compounds are built from non-standard amino acids. Penicillin, thyroxine, and serotonin are shown in a format that is commonly used by chemists to depict chemical structures.

Inspired by the biological properties of non-standard amino acids, medicinal chemists (a chemist who develops and tests new drugs) have used non-standard amino acids to build several therapeutic drugs. These include drugs such as droxidopa (used in the treatment of Parkinson's

disease) and sitagliptin (antidiabetic drug), which have benefited millions of lives. With the emergence of new diseases and antibiotic resistance (where bacteria develop the ability to defeat the drugs designed to kill them), it is vital to build new drugs and improve existing drugs.

Similar to how distinct Lego building blocks help us build complex structures, access to new amino acids will help us create new molecules, which could be evaluated as therapeutics (see Figure 5). This is the reason we need new amino acids, which can be further developed into new therapeutics. So, how do we make them?

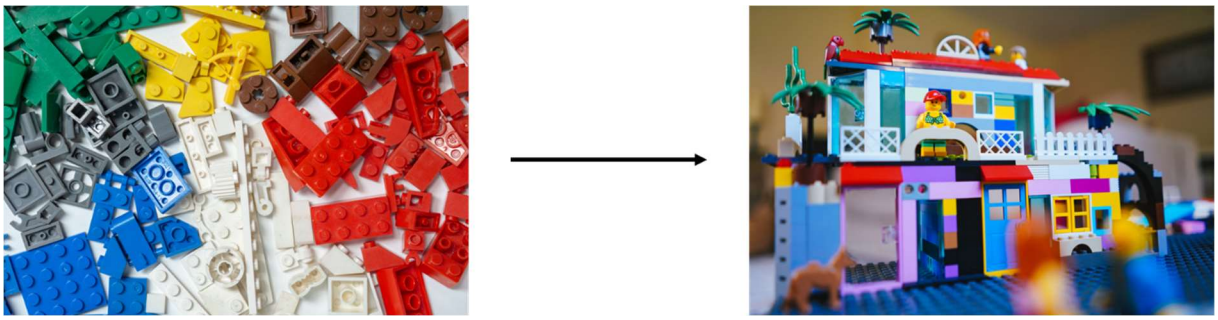


Figure 5. Building complex structures from simple LEGO building blocks. These pictures were downloaded with permission from unsplash.com.

There are several chemical methods to make amino acids. These methods usually require expensive starting materials (substrates), toxic solvents (a liquid used to dissolve substrates), and a catalyst (a substance that increases the rate of a chemical reaction) to build amino acids. Additionally, these methods sometimes require extreme temperatures or oxygen-free chambers. These challenges have motivated scientists to copy the strategy cells have been using to make amino acids for millions of years. Cells use enzymes as catalysts to build amino acids. What are enzymes, and what's special about them?

5. 4. What are enzymes?

Enzymes are a class of proteins that can perform chemical reactions. They are part of our everyday life. For example, the enzymes in the bacteria *Lactobacillus* (naturally found in the human gut) break down lactose in milk into lactic acid during the fermentation of milk into yogurt (see Figure 6). Similarly, the yeast enzymes break down glucose to ethanol and carbon dioxide in the bread-making process. The released carbon dioxide makes the bread rise and leads to soft fluffy bread.

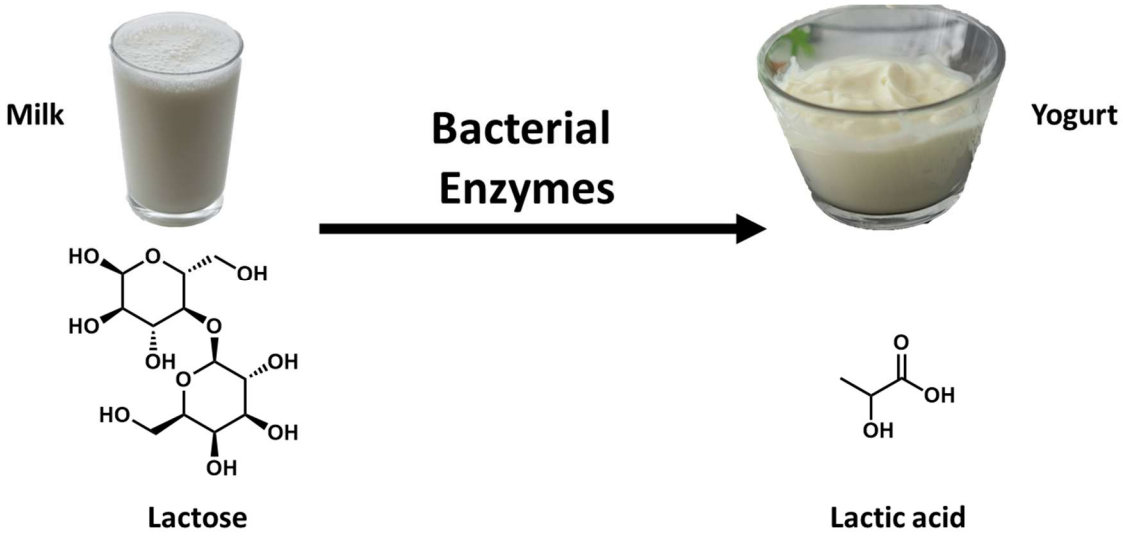


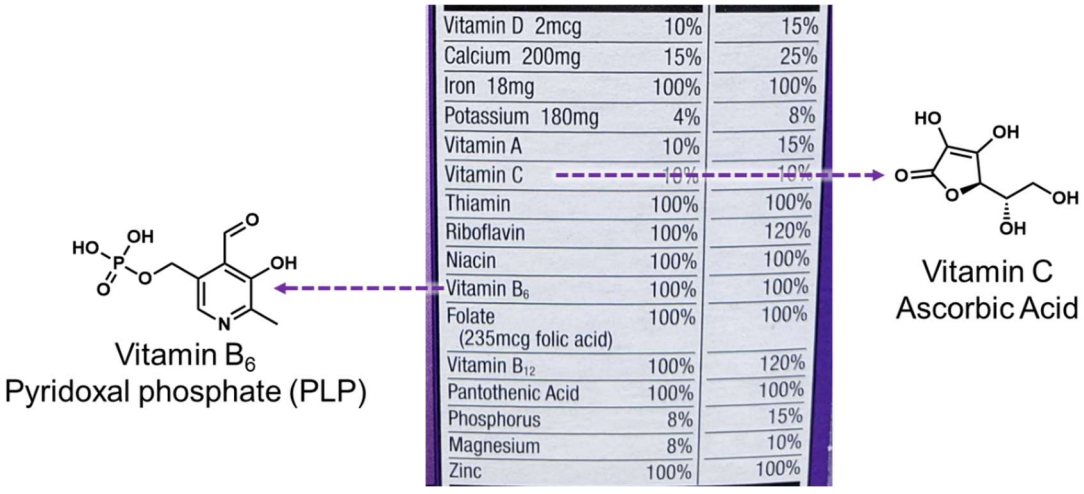
Figure 6. *Lactobacillus* enzymes break down lactose in milk to lactic acid during yogurt formation. Pictures of milk and yogurt were downloaded with permission from unsplash.com.

Several people, including me, have lost the ability to digest milk, a condition called lactose intolerance.[3] This is because as some of us age, our body stops producing lactase, the enzyme that breaks down lactose into glucose and galactose. This indigestion of lactose causes abdominal cramps, bloating, and diarrhea. Fortunately, I can buy lactase pills from a pharmacy. Once consumed, the lactase enzyme breaks down lactose in the digestive tract, and I can still enjoy dairy products.

Enzymes have been used by scientists for several applications due to their ability to work well outside our bodies. For example, lipases that break down fat stains are found in laundry detergents.[4] Enzymes are used to produce bioethanol (a commercially used biofuel) from the degradation of starchy plant materials.[5] Enzymes are also used in the textile, paper, and leather industries. Recently, enzymes that can break down plastics have been discovered.[6] Hopefully, in the coming years, plastic that was thought to take millions of years to break down could be broken down by enzymes in a few weeks.

5. 5. What are vitamins?

Some of the enzymes need helper molecules to perform their functions. These helper molecules are called cofactors. These could be inorganic metal ions such as Magnesium or organic compounds such as vitamins. There are 13 essential vitamins, and they also have roles beyond helping enzymes, such as preventing infections and maintaining strong bones. Our body cannot synthesize vitamins like vitamin B_6 and vitamin C, and we must acquire them through dietary sources. Like vitamins, minerals like iodine and fluoride also help our body function well. This is why we need to consume a variety of fruits and vegetables to make sure our body gets sufficient vitamins and minerals. Vitamins and minerals are also added to foods such as milk, cereals, fruit juices, and even salt to boost their nutritional value (see Figure 7).



Vitamin D 2mcg	10%	15%
Calcium 200mg	15%	25%
Iron 18mg	100%	100%
Potassium 180mg	4%	8%
Vitamin A	10%	15%
Vitamin C	10%	10%
Thiamin	100%	100%
Riboflavin	100%	120%
Niacin	100%	100%
Vitamin B_6	100%	100%
Folate (235mcg folic acid)	100%	100%
Vitamin B_{12}	100%	120%
Pantothenic Acid	100%	100%
Phosphorus	8%	15%
Magnesium	8%	10%
Zinc	100%	100%

Figure 7. Chemical structure of select vitamins found in the Cheerios cereal box.

With the help of cofactors, enzymes can catalyze (accelerate the rate of) several different chemical reactions in our bodies. Chemists have harnessed the power of naturally occurring enzymes from humans and microbes to convert simple starting materials into biologically valuable molecules. This field of research is called ‘biocatalysis’ and has revolutionized the synthesis of chemical building blocks.

5. 6. Why did we choose to synthesize β-hydroxy amino acids?

We chose to synthesize a class of amino acids called β-hydroxy amino acids (β is pronounced as beta). These β-hydroxy amino acids contain a hydroxy group (-OH group) at the β-carbon of the amino acid backbone. The R-group, or the side chain, determines the properties of these amino acids. For example, if the R-group is a hydrogen atom, the resulting amino acid is serine. Similarly, if the R-group is a methyl (-CH_3) group, the resulting amino acid is threonine (see Figure 8). Serine and threonine are two of the twenty amino acids found in proteins.

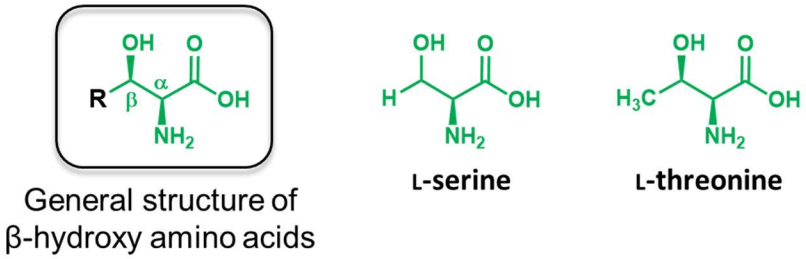


Figure 8. The chemical structure of β-hydroxy amino acids and two proteogenic β-hydroxy amino acids are shown on the right.

A number of microbes have enzymes that synthesize β-hydroxy amino acids with different R-groups, which are eventually used as building blocks for the synthesis of more complex biomolecules, also known as natural products.[7] These natural products include vancomycin (clinically used antibiotic) and cyclosporin (used as an immunosuppressant during organ transplantation). Inspired by these natural products, medicinal chemists have also designed several pharmaceutical drugs that contain β-hydroxy amino acid motifs, such as droxidopa (treatment of Parkinson's disease) and vilanterol (treatment of asthma). I hope you can see that β-hydroxy amino acids are biologically valuable compounds. However, existing chemical and enzymatic methods for the synthesis of β-hydroxy amino acids have limitations. Hence, we decided to search for a new enzyme to synthesize several β-hydroxy amino acids.

5. 7. How did we use enzymes and vitamins to build β-hydroxy amino acids?

We chose an enzyme called ObiH (pronounced as Oh-Bee-H), which is found in a soil bacteria called *Pseudomonas fluorescens*.[8] This bacterium uses the ObiH enzyme to make a β-hydroxy amino acid. This β-hydroxy amino acid is used as a building block to synthesize obafluorin, an antibiotic.

ObiH uses threonine and aldehyde as the two substrates (starting materials), with vitamin B_6 as the cofactor (see Figure 9, native reaction). Threonine is one of the 20 standard amino acids. Aldehydes are compounds in which a carbon atom shares a double bond with an oxygen atom, a single bond with a hydrogen atom, and a single bond with another atom or group of atoms. These two substrates are cheap and commercially available. Out of the two substrates, the aldehyde is the one that determines the side chain or the R-group of the resulting β-hydroxy amino acid. We asked the question: Could ObiH react with other aldehydes to make new β-hydroxy amino acids? (see Figure 9, desired reaction).

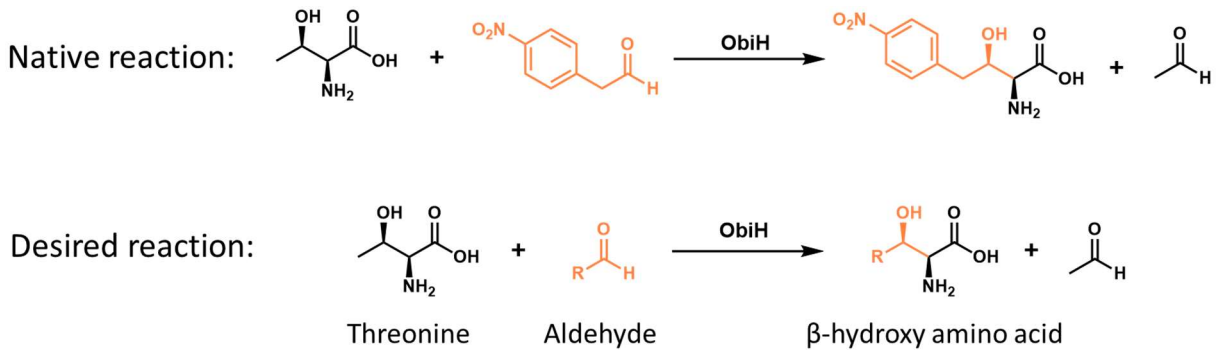


Figure 9. The native and the desired reaction catalyzed by ObiH. Aldehyde (orange) determines the side chain of the resulting β-hydroxy amino acid.

To answer our question, we bought a chemically synthesized ObiH gene (gene is the sequence of DNA alphabets that has the instructions for the cells to make a protein). We stitched this ObiH gene into a plasmid (a circular piece of DNA) and transferred it into *E. coli* (a bacteria

used in the labs to produce proteins). When we grow these *E. coli* bacteria in a liquid growth media (food for bacteria), they multiply rapidly and produce the ObiH enzyme inside them (see Figure 10). ObiH is a pink enzyme, so the bacteria turn pink in color.

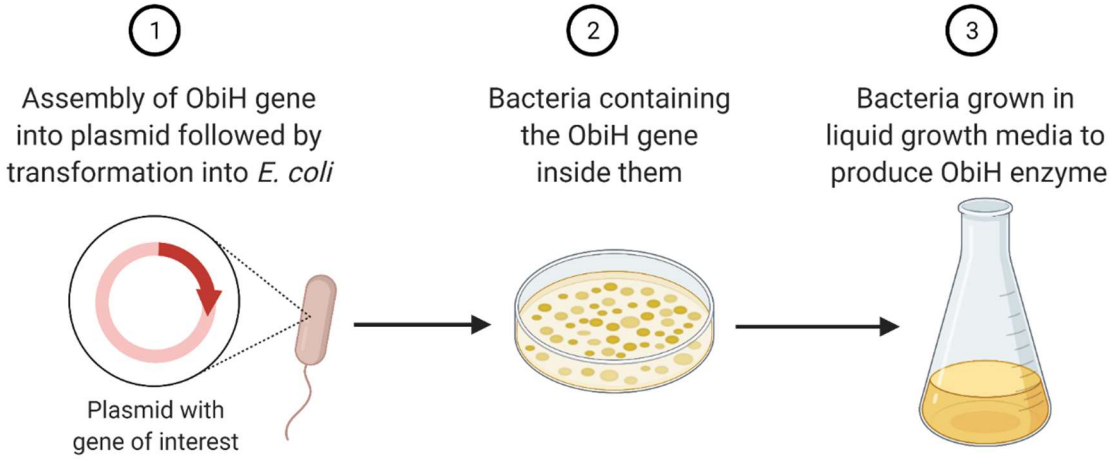


Figure 10. Production of the ObiH enzyme in *E. coli*. First, the commercially bought ObiH gene is assembled into a plasmid and transformed into *E. coli*. This *E. coli* can be grown in a liquid growth media to produce a large number of cells with the ObiH enzyme inside them. *This figure was created with BioRender.com from a pre-existing template.*

To isolate the cells from the growth media, we perform a centrifugation step. In centrifugation, liquid growth media containing bacteria is transferred to a specially designed tube and spun at high speeds using an instrument called a centrifuge. Spinning generates a centrifugal force that causes the heavy cells to settle down at the bottom of the tube. After centrifugation, we dump the liquid growth media into a biowaste container. Then, we transfer the remaining cells into a syringe and inject them into liquid nitrogen (see Figure 11). This causes cells to freeze rapidly, and these frozen cells can be stored in the freezer for a few months. Whenever we need to do experiments, we take a few of the cell noodles out of the freezer and add them to our reaction mixture.

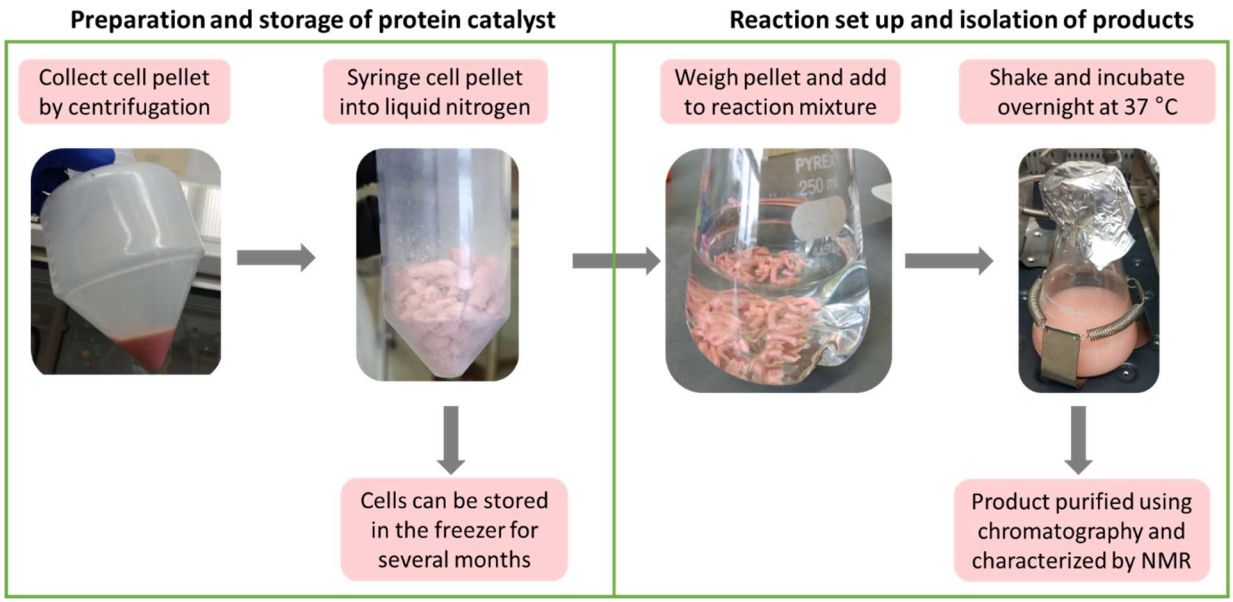


Figure 11. A flowchart describing the steps involved in the preparation of ObiH cells, the reaction set up, and isolation of the β-hydroxy amino acid.

The reaction mixture contains threonine, aldehyde, vitamin B_6, and ObiH cells dissolved in a buffer (a salt and water solution that keeps the cells happy). We shake this mixture in a flask at 37 °C overnight using an instrument called a shaking incubator. During this time, ObiH catalyzes the chemical reaction between threonine and aldehyde using vitamin B_6 as the cofactor to build the β-hydroxy amino acid. We use two techniques (chromatography in conjunction with Nuclear Magnetic Resonance or NMR) to isolate our desired β-hydroxy amino acid and determine its purity.

We found that ObiH could catalyze a reaction with almost every aldehyde we tested. We were able to successfully isolate 15 different β-hydroxy amino acids (in the range of 50 to 500 milligrams of the compound), a few of which were never made by anyone before (see Figure 12).[9] Our methodology used cheap and commercially available starting materials with a naturally occurring enzyme as the catalyst to isolate several β-hydroxy amino acids in an environmentally friendly fashion. This is an exciting result. We have been recently contacted by other researchers who have requested the compounds we made to do their experiments – even more exciting!

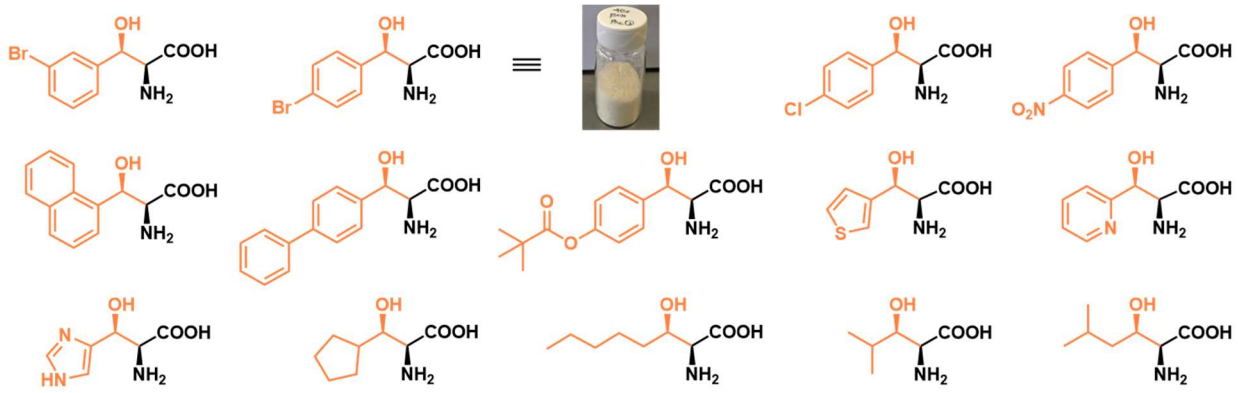


Figure 12. The structures of β-hydroxy amino acids isolated using ObiH as the catalyst are shown. A picture of the isolated compound is shown for one of the β-hydroxy amino acids.

5. 8. What else did we do with ObiH?

In addition to applying the ObiH enzyme for β-hydroxy amino acid synthesis, we were also interested in understanding how the enzyme interacts with the substrate and cofactor spatially to catalyze the chemical reaction. So, we decided to solve the three-dimensional (3D) structure of ObiH using a technique called X-ray diffraction. To do this, we crystallized the ObiH enzyme by slowly evaporating the buffer in which ObiH is stored (similar to making salt by evaporating seawater). We then shine powerful X-rays onto the crystals. When these X-rays pass through and exit the crystal, they produce a diffraction pattern that we can use to build a 3D model of the ObiH enzyme (see Figure 13). The ObiH structure gave us information about how the vitamin B_6 cofactor and substrates interact with the enzyme.[10] Most importantly, this structure gave us clues about how to modify the enzyme (also known as protein engineering) to enable the ObiH enzyme to react with more substrates.

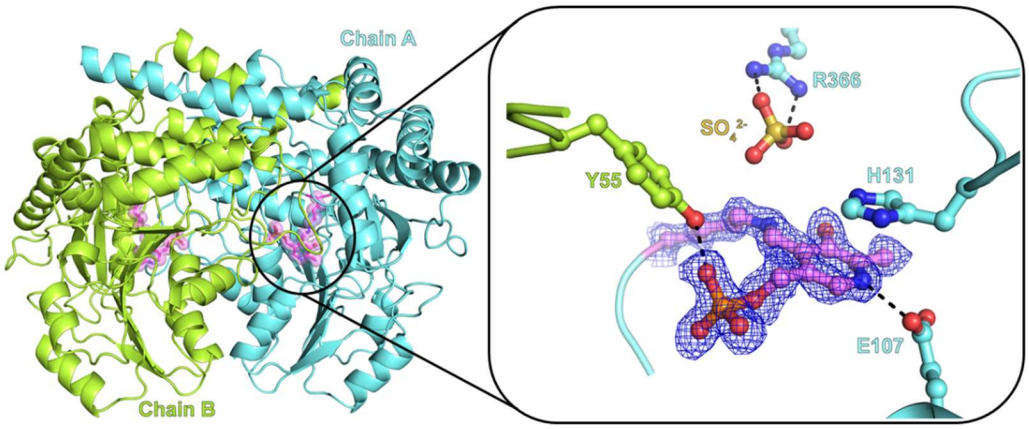


Figure 13. The 3D structure of ObiH. ObiH crystallized as a dimer (which means there are two copies of ObiH in contact with each other). Each of the monomers (one copy of ObiH) is shown in green and blue, respectively. The zoomed panel shows the enzyme's active site. The active site is the part of the enzyme where the starting materials and cofactor bind for the chemical reaction to take place.

5. 9. What if the ObiH enzyme does not accept specific substrates?

The ObiH enzyme accepts several different aldehydes as substrates to make β-hydroxy amino acids. But there are a few aldehydes that ObiH did not accept or accepted poorly as a substrate.[9] So, we decided to evolve the ObiH enzyme to accept even more aldehydes. What does it mean to evolve an enzyme?

The *E. coli* cells we have in the lab contain the ObiH gene (sequence of DNA alphabets that has the instructions to make the ObiH enzyme). We use a technique called mutagenesis to make small changes to the ObiH gene. This change in the gene changes the structure and function of the enzyme (see Figure 14). Some changes could be beneficial (reactivity with more substrates), whereas others could be harmful (no reactivity or reactivity with fewer substrates). Since it's challenging to predict what changes would be beneficial, we evaluate as many mutated enzymes as possible to identify an improved enzyme. This methodology of improving protein activity is called protein engineering.[11]

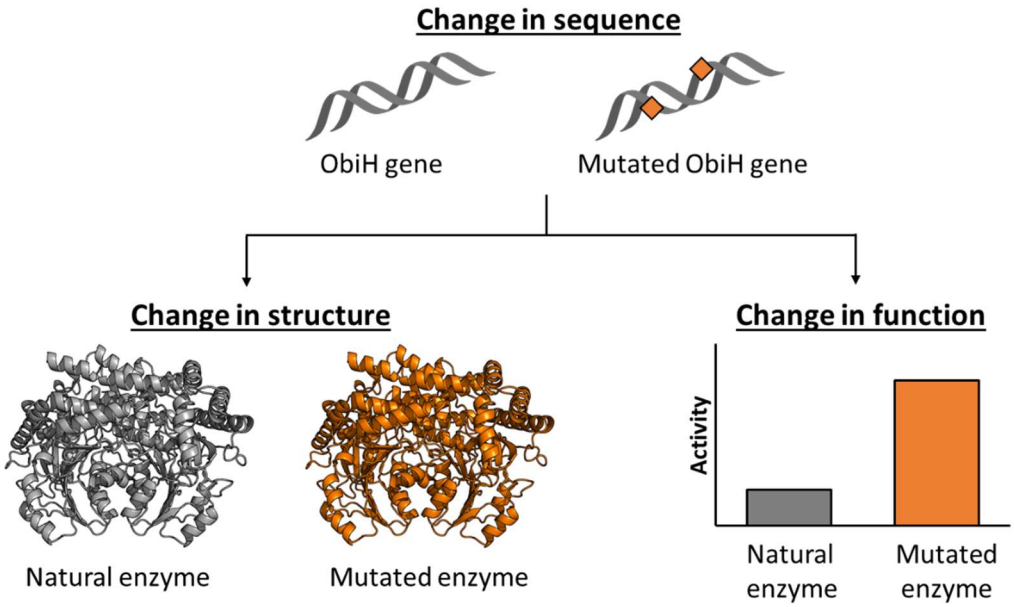


Figure 14. Protein engineering allows us to tailor enzymes for improved functionality in the lab.

Once we identify an improved enzyme, we subject it to additional changes to see if we can make the enzyme even better. We can repeat this process of small modifications as many times as we want. This process of gradually making changes to improve the enzyme activity is called directed evolution, and this methodology was awarded the Nobel Prize in Chemistry in 2018.[12] This process is similar to natural evolution and the concept of "survival of the fittest." In the natural world, evolution takes hundreds to millions of years. In the lab, we accelerate this evolution process over a few weeks.

We subjected the ObiH enzyme to directed evolution. The 3D structure of ObiH guided us on what changes to make. After evaluating hundreds of mutated enzymes, we have found a few mutated enzymes that accept more aldehydes as substrates. Now, we have evolved ObiH enzymes in the lab that are much better at producing a wide range of β-hydroxy amino acids than the naturally occurring ObiH enzyme.

5. 10. References

1. Watford, M. & Wu, G. Protein. *Adv. Nutr.* **9**, 651 (2018).

2. Blaskovich, M. A. T. Unusual Amino Acids in Medicinal Chemistry. *Journal of Medicinal Chemistry* **59**, 10807–10836 (2016).

3. Szilagyi, A. & Ishayek, N. Lactose Intolerance, Dairy Avoidance, and Treatment Options. *Nutrients* **10**, (2018).

4. Niyonzima, F. N. Detergent-compatible bacterial cellulases. *J. Basic Microbiol.* **59**, 134–147 (2019).

5. Vasić, K., Knez, Ž. & Leitgeb, M. Bioethanol Production by Enzymatic Hydrolysis from Different Lignocellulosic Sources. *Molecules* **26**, (2021).

6. Amobonye, A., Bhagwat, P., Singh, S. & Pillai, S. Plastic biodegradation: Frontline microbes and their enzymes. *Sci. Total Environ.* **759**, (2021).

7. Wang, S. & Deng, H. Peculiarities of promiscuous l-threonine transaldolases for enantioselective synthesis of β-hydroxy-α-amino acids. *Appl. Microbiol. Biotechnol.* **105**, 3507–3520 (2021).

8. Schaffer, J. E., Reck, M. R., Prasad, N. K. & Wencewicz, T. A. β-Lactone formation during product release from a nonribosomal peptide synthetase. *Nat. Chem. Biol.* **13**, 737–744 (2017).

9. Doyon, T. J. *et al.* Scalable and Selective β-Hydroxy-α-Amino Acid Synthesis Catalyzed by Promiscuous l-Threonine Transaldolase ObiH. *ChemBioChem* **23**, e202100577 (2022).

10. Kumar, P. *et al.* l-Threonine Transaldolase Activity Is Enabled by a Persistent Catalytic Intermediate. *ACS Chem. Biol.* **16**, 86–95 (2020).

11. Lutz, S. & Iamurri, S. M. Protein Engineering: Past, Present, and Future. *Methods Mol. Biol.* **1685**, 1–12 (2018).

12. Bloom, J. D. & Arnold, F. H. In the light of directed evolution: pathways of adaptive protein evolution. *Proc. Natl. Acad. Sci. U. S. A.* **106 Suppl 1**, 9995–10000 (2009).